Université de Strasbourg

UNIVERSITÉ DE STRASBOURG

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE

IGBMC, CNRS UMR 7104

THÈSE

présentée par

Anna BONHOURE

soutenue le : **21 octobre 2020**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline : Sciences de la Vie

Spécialité : Biophysique et Biochimie des protéines

Interférence de l'oncoprotéine E6 de HPV avec l'interactome humain

Étude quantitative et structurale

THÈSE dirigée par :

Dr. TRAVÉ Gilles	Directeur de recherches, IGBMC CNRS UMR 7104 (Illkirch-Graffenstaden, France)
Dr. NOMINÉ Yves	Maître de conférences, IGBMC CNRS UMR 7104 (Illkirch-Graffenstaden, France)

RAPPORTEURS :

Dr. ROBERTS Sally	Senior lecturer, University of Birmingham (Birmingham, Royaume-Uni)
Dr. WOLFF Nicolas	Directeur de recherches, Institut Pasteur (Paris, France)

AUTRES MEMBRES DU JURY :

Dr. ZANIER Katia	Chargée de recherches, Biotechnologie et Signalisation Cellulaire UMR 7242 (Illkirch-Graffenstaden, France)
Dr. BACHELERIE Françoise	Directrice de recherches, UMR-996 INSERM-Université Paris Sud (Paris, France)

"Success is going from failure to failure without losing your enthusiasm."
Winston Churchill

A mon grand-père Mariano:
Ya ves, terminaré como doctora en bioquímica y no presidente.

Remerciements

Je tiens en premier lieu à remercier les membres du jury qui ont accepté d'évaluer ma thèse : les rapporteur·trice **Sally Roberts** et **Nicolas Wolff** ainsi que les examinatrices **Françoise Bachelerie** et **Katia Zanier**. Merci pour votre disponibilité et votre bienveillance en lisant ce manuscrit et en assistant à la soutenance.

Je tiens également à remercier les nombreux collaborateurs qui m'ont permis d'aborder de nombreux projets aux thématiques et techniques variées. Merci à **Pilar Armisen** pour nous avoir fourni des résines d'affinité personnalisées, à **Murielle Masson** pour son expertise en GPCA, à **Ylva Ivarsson** pour m'avoir chaleureusement accueillie dans son équipe afin d'utiliser sa banque de phage display, aux spécialistes de l'évolution des HPV **Ignacio Bravo** et **Anouk Willemsen**, à **Claudia Simon** pour le challenge de CRPV et 31E6, à **Renaud Vincentelli** pour m'avoir expliqué le protocole robotisé du holdup et à **Pascal Villa** et **Sophie Gioria** pour l'avoir mis en place à la PCBIS. Je remercie particulièrement **Marcel Conrady**, pour sa bonne humeur contagieuse et ce, dès son arrivée au laboratoire ! Je tiens également à remercier **Pascal Eberling** du service de synthèse peptidique de l'IGBMC, pour sa bonne humeur et sa réactivité, ainsi que **Luc Negroni**, de la plateforme de Spectrométrie de Masse de l'IGBMC.

J'adresse un immense merci à mes directeurs de thèse **Gilles** et **Yves**. Merci à tous les deux pour m'avoir conjointement accompagnée pendant cette thèse, j'ai pu bénéficier de vos expertises complémentaires. Merci aussi de m'avoir donné l'occasion de participer à de nombreux congrès et d'être impliquée dans de nombreux projets, qui m'ont donné une bonne ouverture scientifique.

Gilles, merci pour m'avoir fait confiance pour mener ce projet et pour m'avoir épaulée avec bienveillance dans chacun de ses aspects, du design expérimental à la rédaction. Merci pour avoir partagé ta passion pour la belle science et tes idées créatives, j'ai énormément appris à tes côtés. Merci aussi pour les moments moins scientifiques, à base de découvertes musicales (He said Captain, I said wot!).

Yves, je te remercie pour ton accompagnement pendant cette thèse. Merci pour ta rigueur et pour tes relectures attentives lors de la rédaction des publications. Merci de m'avoir donné des bases solides de traitement de données et merci pour ta pédagogie et ton infinie patience quand tu m'as aidée à préparer des présentations orales.

Je remercie également les membres de l'équipe actuels mais également ceux qui sont partis vivre de nouvelles aventures. Je remercie en particulier **Alex**, pour sa gentillesse et sa bienveillance, **Auguste**, le génial imprimeur 3D, **Alberto**, pour ses discussions philosophiques très argentines, **André**, pour avoir partagé sa passion de la cristallographie, **Danièle**, pour m'avoir initiée à la SPR, **Gergő**, pour sa gentillesse et ses discussions scientifiques, **Elodie**, pour son aide précieuse au moment de préparer la soutenance, **Eduardo**, **Irina**, **Goran**, **Florine**, **Alison**.

Je remercie mes amies **Syrine**, pour son pilotage de l'extrême en toute situation, **Pernelle**, pour sa licornerie diplomate en toutes circonstances, et **Pauline**, pour sa gentillesse et sa logique imparable. Je ne remercie pas **Christophe** pour sa propreté dans le laboratoire ni pour son rire à la Lambert. Je remercie également **Claire**, pour sa bonne humeur et pour nous avoir fait le meilleur fraisier du monde, **Arnaud**, pour les bons moments et pour son aide scientifique, **Valentin**, **Paola**, **Léa**, **Judit**, **Nicolas**, **Benoît** et **Adam**.

J'adresse un grand merci à **Valérie Lamour** et **Sylvain Robin**, qui ont très gentiment accepté d'examiner mon travail lors de deux mi-thèses. Merci pour votre bienveillance et vos retours.

Je remercie également les associations qui ont financé ma thèse : La Ligue Contre le Cancer pour les trois premières années et la Fondation Recherche Médicale pour ma quatrième année.

Merci à ma famille et en particulier à mes parents, qui m'ont toujours soutenue dans mes projets et qui ne manquent pas une occasion de traverser le pays pour venir me voir à Strasbourg.

Enfin, je remercie celle qui a changé ma vie et qui m'a courageusement épaulée durant cette thèse. Merci à toi, **Camille**, de partager ma vie et de m'avoir supportée, surtout pendant cette longue période de rédaction.

Merci à **Hélène Terzian**, **Stéphane Bodin** et **Hend** pour m'avoir donné le goût de la science.

TABLE OF CONTENTS

Foreword

This manuscript presents the work achieved during the four years of my PhD under the direction of Gilles Travé and Yves Nominé. I worked in the team directed by Gilles Travé, hosted in the Center of Integrative Biology (CBI) in IGBMC, CNRS UMR 7104 - Inserm U 1258, Illkirch-Graffenstaden.

The thesis project aimed at deciphering the interactome of HPV E6 oncoproteins, with a particular focus on their LXXLL motif interaction preferences. We addressed this subject from the perspective of biochemistry, biophysics and structural biology. The difficulty inherent to the purification of recombinant E6 oncoprotein led us to design adapted experimental procedures, in line with the work carried out by Gilles Travé since he started studying E6 proteins. We analyzed the interaction preferences of various E6 proteins, representative of the phylogenetic and phenotypic diversity of existing HPVs. The comparative analysis has proven to be a potent strategy to highlight what makes the difference between an oncogenic and a low-risk HPV.

The results section of this manuscript is splitted into three parts. The first part gathers the methodological developments designed during the thesis work, entailing two publications. The second part details the LXXLL interaction features of various E6 oncoproteins, including a study of seven E6 proteins from diverse HPV types and a project performed in collaboration with Ignacio Bravo and Anouk Willemsen (MiVEGEC, IRD, Montpellier) focusing on ancestral E6 proteins from the genus alpha. The third part is an adjunct project aiming at deciphering the interaction of E6AP, a major target of oncogenic HPV E6 proteins, with another ubiquitin ligase named HERC2.

A part of the general introduction has been translated in French and added to this manuscript as an appendix.

I sincerely hope you enjoy reading this manuscript.

Abbreviations

AS: Angelman Syndrome

ATP: Adenosine triphosphate

DNA: Deoxyribonucleic Acid

RNA: Ribonucleic Acid

BPV: Bovine Papillomavirus

CIN: Cervical Intraepithelial Neoplasia

CBP: CREB-binding protein

CKII: Casein Kinase II

CRPV: Cottontail Rabbit Papillomavirus

DTT: Dithiothreitol

DYRK1A: Dual-specificity Tyrosine phosphorylation-regulated Kinase 1A

E6AP: E6-Associated Protein

EDC: 1-éthyl-3-(3-diméthylaminopropyl) carbodiimide

EV: Epidermodysplasia verruciformis

FAK: Focal Adhesion Kinase

HECT: Homologous to E6AP C-terminal

HERC2: HECT domain and RCC1-like domain-containing protein 2

HPV: Human Papillomavirus

IRF3: Interferon-Regulatory Factor 3

ITC: Isothermal Titration Calorimetry

MAML1: Mastermind-like protein 1

MBP: Maltose-Binding Protein

MST: Microscale Thermophoresis

NHS: N-hydroxysuccinimide

NMR: Nuclear Magnetic Resonance

PBM: PDZ-Binding Motif

PDZ: PSD-95, Dlg1, ZO-1

PTPN: Tyrosine-Protein Phosphatase Non-receptor

Rb: Retinoblastoma protein

SEC: Size-exclusion chromatography

SPR: Surface Plasmon Resonance

TAP-MS: Tandem Affinity Purification followed by Mass Spectrometry analysis

TCEP: Tris(2-carboxyethyl)phosphine

VLP: Virus-Like Particles

URR: Upstream Regulatory Region

WHIM: Warts Hypogammaglobulinemia Infections Myelokathexis

Y2H: Yeast Two-Hybrid

Amino acids

Alanine	Ala	A
Arginine	Arg	R
Aspartic acid	Asp	D
Asparagine	Asn	N
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Résumé de thèse en français

Introduction

Les papillomavirus humains (HPV) sont de petits virus à ADN infectant les épithéliums. Une grande diversité de papillomavirus a été identifiée. On dénombre à ce jour 228 types (hpvcenter.se et pave.niaid.nih.gov), classés en 5 genres phylogénétiques : alpha, bêta, gamma, mu et nu, se différenciant par leur tropisme (muqueux/cutané) et leurs manifestations cliniques. Si la plupart des HPV, dits à bas risque, ne causent que des proliférations bénignes (verrues, condylomes), certains HPV, dits à haut-risque, peuvent provoquer différents cancers. Ainsi, les HPV sont les principaux agents étiologiques du cancer du col de l'utérus, quatrième cancer le plus fréquemment diagnostiqué dans le monde, et responsable de 250.000 décès annuels (de Martel *et al.*, 2017). Les HPV 16 et 18 (haut risque muqueux) génèrent respectivement 61 et 11 % des cancers du col de l'utérus (Serrano *et al.*, 2015). D'autre part, un nombre croissant de cancers de l'oropharynx, de la cavité orale et du larynx est associé à une infection par un HPV. Parmi ces cancers HPV-positifs, 71 % sont attribués au HPV16 (Castellsagué *et al.*, 2016). Certains HPV du genre beta pourraient contribuer à l'apparition de carcinomes cutanés, en combinaison avec une exposition de la peau aux rayons ultraviolets (UV). De plus, certains HPV bêta provoquent des carcinomes spinocellulaires chez les patients atteints d'épidermodysplasie verruciforme, une affection cutanée rare d'origine génétique (Accardi and Gheit, 2014).

Les oncoprotéines virales E6 et E7 sont à l'origine de la cancérogenèse : elles favorisent la réplication virale en stimulant la prolifération cellulaire. En particulier, E6 perturbe le fonctionnement de nombreuses protéines cellulaires en les capturant via des motifs peptidiques de consensus "LXXLL". Ainsi, les E6 des HPV 16 et 18 capturent le motif LXXLL de l'ubiquitine ligase E6AP (E6 Associated Protein). Une fois liée à E6AP, HPV 16 E6 recrute le suppresseur de tumeurs p53 pour induire sa dégradation via le protéasome (Huibregtse *et al.*, 1991, 1993; Scheffner *et al.*, 1990, 1993). Les structures cristallographiques de E6 HPV 16 en complexe avec le motif LXXLL de E6AP et le domaine central de p53 ont été résolues par l'équipe (Martinez-Zapien *et al.*, 2016; Zanier *et al.*, 2013). Chaque protéine E6 produite par un type de HPV donné cible un certain ensemble de protéines hôtes, et ces préférences d'interactions déterminent la variabilité d'effets pathologiques provoqués par le virus. L'objectif de la thèse était d'explorer les préférences d'interactions des oncoprotéines

E6 issues de différents types de HPV pour un ensemble de motifs peptidiques LXXLL. Cette analyse a été accompagnée d'une quantification de la force de chaque interaction par des approches biophysiques.

Résultats

Développements métholodogiques

Les protéines E6 ont une tendance à l'auto-association : elles forment des oligomères inactifs sur le plan biologique, compliquant les tests d'interaction biophysiques. J'ai développé une tactique de purification selon un protocole de "batch" pouvant être effectué en parallèle sur plusieurs protéines E6, et qui permet de maximiser la proportion de protéine soluble et monomérique. En effet, chaque E6 contient au minimum 8 cystéines requises pour la coordination de deux atomes de zinc : elles sont donc susceptibles de s'auto-associer sous l'effet de l'oxydation et les oligomères inactifs ainsi formés peuvent biaiser les tests d'interaction. En se basant sur la taille et sur la diffusion des monomères de E6, j'ai pu optimiser un protocole de purification rapide testé sur des E6 de HPV 08 (genre bêta) et de HPV 16 (genre alpha). Pour évaluer la qualité de la protéine E6 de HPV 16 purifiée d'après la nouvelle approche, j'ai testé par résonance plasmonique de surface (SPR) l'interaction HPV 16 E6 – LXXLL de E6AP, déjà caractérisée au préalable par l'équipe (Zanier *et al.*, 2005). Le fait que les constantes de dissociation (K_D) évaluées sont cohérentes avec les données précédemment publiées, valide ce protocole qui permet donc de purifier une protéine active, utilisable pour des mesures biophysiques *in vitro* d'interaction protéine-motif. Cette approche a fait l'objet d'un article publié dans la revue *Microbial Cell Factories* (**a**). J'ai pu ensuite mettre cette stratégie à profit pour purifier en parallèle différentes E6 et comparer par SPR leurs préférences d'interactions LXXLL dans le cadre de mon projet principal de thèse.

Afin d'étudier les interactions entre les protéines E6 et les motifs LXXLL, j'utilise la technique holdup, une approche chromatographique de rétention développée au sein de l'équipe, permettant de détecter et de quantifier l'affinité de nombreuses interactions protéine-motif en parallèle (Charbonnier *et al.*, 2006). Cette méthode consiste à immobiliser un peptide sur résine avant de l'incuber avec le partenaire d'interaction (protéine) en solution. La fraction liquide est récupérée par filtration : elle

contient la part de protéine qui n'a pas interagi avec le peptide présenté sur résine. La quantité de protéine restant dans la fraction liquide est mesurée par électrophorèse capillaire avant d'être comparée avec un contrôle négatif. La déplétion dans la fraction liquide est d'autant plus importante (de par une quantité d'autant plus importante retenue sur la résine avec le peptide d'intérêt) que la force d'interaction est grande. Pour chaque motif, un profil d'interaction est établi pour un panel de domaines. Cette méthode est automatisable et applicable à haut-débit ; elle a d'ores et déjà permis une étude approfondie des interactions des PBM des E6 de HPV 16 et 18 avec l'ensemble des 266 domaines PDZ existant dans le protéome humain (Vincentelli *et al.*, 2015).

En complément de la structure de HPV16 E6 en complexe avec le peptide LXXLL de E6AP précédemment publiée par l'équipe (Zanier *et al.*, 2013), j'ai déterminé par méthode holdup les préférences fines d'interaction de HPV16 E6 pour une banque de 45 mutants ponctuels du motif LXXLL de E6AP. Ces données ont été confirmées par des tests d'interaction avec des membranes de peptides SPOT, une approche semi-quantitative souvent utilisée pour comparer les propriétés d'interactions de nombreux peptides (Mancilla and Volkmer, 2016). En parallèle, la technique holdup a été optimisée : l'étape d'électrophorèse capillaire, permettant de quantifier la fraction de protéine présente dans la fraction liquide, a été remplacée par une approche de spectroscopie de fluorescence, qui présente l'avantage d'être plus rapide et moins onéreuse pour des résultats de qualité équivalente. Enfin, les constantes de dissociation de l'ensemble des peptides mutés ont pu être estimées à partir des résultats issus du holdup et des mesures de SPR. Ces données d'interaction permettent d'identifier les résidus-clés de l'interaction et de concevoir des inhibiteurs de haute affinité : elles font l'objet d'une publication (**b**).

Interactions de différentes oncoprotéines E6 avec des motifs LXXLL

J'ai comparé la spécificité d'interaction de 6 oncoprotéines E6 pour une banque de peptides LXXLL issus de diverses protéines cellulaires sélectionnées parmi les cibles de HPV E6 les plus documentées dans la littérature. Ces E6 sont issues de différents types de HPV, représentatifs de la diversité phylogénétique et phénotypique : HPV16, 18 et 31 (genre alpha) sont fortement impliqués dans les cancers du col de l'utérus, HPV197 (gamma) est un HPV de tropisme cutané, HPV38 et 49 (beta) sont potentiellement oncogéniques avec respectivement un tropisme cutané et muqueux

(Viarisio *et al.*, 2016, 2018). Les données d'interaction protéine-motif sont en accord avec les interactions de protéines entières obtenues par purification d'affinité et spectrométrie de masse, confirmant la validité de notre approche fragmentaire (Grace and Münger, 2017; White *et al.*, 2012a). J'ai aussi participé à l'obtention de structures cristallographiques de plusieurs E6 en complexe avec leurs principaux ligands LXXLL : 16E6/IRF3$_{LXXLL}$ (PDB : 6SJA), 18E6/mutant E6AP$_{LXXLL}$ (PDB :6SJV) et 49E6/MAML1$_{LXXLL}$ (PDB : 6SMV). MAML1 est ciblé par des protéines E6 de HPV ayant un tropisme cutané : cette interaction mène à une répression de la transduction du signal de la voie Notch, qui joue un rôle central dans la différenciation des kératinocytes (Brimer *et al.*, 2012). IRF3 est un activateur transcriptionnel de la voie interféron, qui joue un rôle dans l'immunité innée en réponse à une infection virale (Ronco *et al.*, 1998; Suarez and Travé, 2018). Ces décryptages moléculaire et structural des variations de reconnaissance des différentes protéines E6 permettent de mieux comprendre la diversité de phénotypes provoqués par les différents types de HPV (manuscrits **c** et **d**).

Une collaboration avec Ignacio Bravo et Anouk Willemsen (Institut MiVEGEC, Montpellier) nous a permis d'étudier les spécificités d'interactions de plusieurs oncoprotéines E6 ancestrales reconstruites à partir d'alignements de séquences (Willemsen and Bravo, 2019). L'étude porte sur les E6 ancestrales d'un groupe qui inclut non seulement les papillomavirus humains du genre alpha mais également certains papillomavirus de singe phylogénétiquement proches. Quatre oncoprotéines E6 ancestrales ont été sélectionnées pour tester *in vitro* leurs préférences d'interaction avec des peptides à motif LXXLL. Grâce aux approches combinées de purifications en parallèle et de holdup, j'ai pu établir des profils d'interaction pour l'ensemble des E6 ancestrales et les comparer avec ceux des HPV E6 du genre alpha bien connues (HPV16 et 18). Les interactions de plus haute affinité ont été validées par SPR et ont permis d'identifier plusieurs complexes E6 ancestrale-LXXLL d'intérêt. Ces données d'interaction permettent une meilleure compréhension de la mise en place des interactions entre oncoprotéines E6 et motifs LXXLL cellulaires au cours de l'évolution des papillomavirus (manuscrit **e**).

Projet annexe

L'ubiquitine ligase E6AP est une cible d'interaction majeure de E6 de HPV 16, qui a un effet d'activateur allostérique tout comme HERC2, une ubiquitine ligase. Kühnle *et al.* ont observé une interaction directe entre les résidus 150-200 de E6AP et RLD2, le second domaine RLD (RCC1-like domain) de HERC2, pour lequel aucune structure n'est disponible à ce jour (Kühnle *et al.*, 2011). Dans un premier temps, j'ai confirmé par chromatographie d'exclusion stérique l'interaction entre le domaine RLD2 de HERC2 et la protéine E6AP entière. Dans un second temps, j'ai tenté de déterminer le motif d'interaction minimal de E6AP pouvant lier le domaine RLD2. La région de E6AP identifiée dans la littérature étant prédite comme désordonnée, j'ai utilisé des peptides chevauchants pour réduire l'interface aux résidus essentiels impliqués dans l'interaction. Par les approches de holdup et SPR, j'ai pu restreindre l'interface avec RLD2 à un peptide de 15 résidus. Ce projet fait aujourd'hui l'objet d'une thèse à part entière au laboratoire menée par Auguste Demenge, qui a résolu plusieurs structures cristallographiques du domaine RLD2 de HERC2 fusionné au motif court de E6AP. L'étude de cette interaction permettra de mieux comprendre comment la perte des fonctions de E6AP et HERC2 peut mener à des pathologies telles que le syndrome d'Angelman. Ce trouble du développement neurologique a pour origine la perte de fonction de E6AP et HERC2 par mutations du locus chromosomique où leurs gènes sont situés (locus 15q11-13) (Kalsner and Chamberlain, 2015). Il y a donc un double lien entre E6AP et HERC2 : l'interaction physique entre les deux protéines et leur implication dans le même trouble neurologique dû à des mutations dans leur séquence génique.

Conclusions

Au cours de ma thèse, j'ai pu acquérir des compétences pour les développements méthodologiques nécessaires à l'étude biophysique de protéines instables *in vitro*. Grâce à cette expérience, je suis en mesure de m'adapter à différents systèmes biochimiques et de concevoir des stratégies de purification et de tests d'interactions optimisés pour une protéine donnée. Ces optimisations sur le plan biochimique m'ont permis d'explorer les préférences de reconnaissance de motif LXXLL de plusieurs oncoprotéines E6 qui n'avaient pas pu être isolées *in vitro* jusque-là. Par des approches biochimiques, biophysiques et structurales, j'ai pu montrer que chaque

oncoprotéine E6 d'un type de HPV donné cible un ensemble distinct de protéines cellulaires via leur motif LXXLL. Ces variations de spécificité participent certainement aux variations de tropisme et d'effets pathologiques observées pour différents HPV. Les données d'interactions combinées aux structures des différents complexes E6/LXXLL pourront notamment servir de base au développement d'inhibiteurs spécifiques des protéines E6 à visée thérapeutique.

Thesis summary in English

Introduction

Human papillomaviruses (HPV) are small DNA viruses infecting epithelia. A wide variety of papillomaviruses have been identified. At the time of writing, there are 228 types (hpvcenter.se and pave.niaid.nih.gov), classified into 5 phylogenetic genera: alpha, beta, gamma, mu and nu, differentiated by their tropism (mucosal/cutaneous) and their clinical manifestations. While most HPVs, known as low-risk, cause only benign proliferations (warts, condylomas), some HPVs, known as high-risk, can cause various cancers. HPVs are the main etiological agents of cervical cancer, the fourth most frequently diagnosed cancer in the world and responsible for 250,000 deaths per year (de Martel *et al.*, 2017). HPV 16 and 18 (high mucosal risk) account for 61% and 11% of cervical cancers, respectively (Serrano *et al.*, 2015). In addition, an increasing number of head and neck cancers (oropharynx, oral cavity and larynx) are associated with HPV infection. Of these HPV-positive cancers, 71% are attributed to HPV16 (Castellsagué *et al.*, 2016). Some beta genus HPVs may contribute to the development of skin carcinomas in combination with exposure of the skin to ultraviolet (UV) radiation. In addition, some beta HPVs cause squamous cell carcinomas in patients with epidermodysplasia verruciformis, a rare skin condition of genetic origin (Accardi and Gheit, 2014).

The viral oncoproteins E6 and E7 are at the origin of carcinogenesis: they promote viral replication by stimulating cell proliferation. In particular, E6 disrupts the functioning of many cellular proteins by capturing them via "LXXLL" consensus peptide motifs. Thus, the E6 of HPV 16 and 18 capture the LXXLL motif of the ubiquitin ligase E6AP (E6 Associated Protein). Once linked to E6AP, HPV 16 E6 recruits the tumor suppressor p53 to induce its degradation via the proteasome (Huibregtse *et al.*, 1991, 1993; Scheffner *et al.*, 1990, 1993). The crystallographic structures of E6 HPV 16 in complex with the LXXLL motif of E6AP and the core domain of p53 have been solved by the team (Martinez-Zapien *et al.*, 2016; Zanier *et al.*, 2013). Each E6 protein produced by a given HPV type targets a certain set of host proteins, and these interaction preferences determine the variability of pathological effects caused by the virus. The aim of the thesis was to explore the interaction preferences of E6 oncoproteins from different HPV types for a set of LXXLL peptide motifs. This analysis

was accompanied by a quantification of the strength of each interaction using biophysical approaches.

Results

Methodological developments

E6 proteins have a strong tendency to self-association, complicating biophysical interaction tests. I have developed a purification strategy based on a "batch" protocol that can be performed in parallel on several E6 proteins. The approach allows to maximize the proportion of soluble and monomeric protein. Indeed, each E6 contains at least 8 cysteines required for the coordination of two zinc atoms: they are therefore likely to self-associate under the effect of oxidation and the resulting inactive oligomers may bias the interaction tests. Based on the size and diffusion of E6 monomers, I was able to optimize a rapid purification protocol tested on E6 from HPV 08 (beta genus) and HPV 16 (alpha genus). To evaluate the quality of the purified HPV 16 E6 protein according to the new approach, I tested by surface plasmon resonance (SPR) the HPV 16 E6 - LXXLL interaction of E6AP, previously characterized by the team (Zanier *et al.*, 2005). The estimated dissociation constants (K_D) are consistent with previously published data. The protein purified by this protocol is active and suitable for *in vitro* biophysical measurements of protein-motif interaction. This approach was the subject of an article published in the journal *Microbial Cell Factories* (**a**). I was then able to use this strategy to purify in parallel different E6 and compare by SPR their preferences of LXXLL interactions in the framework of my main thesis project.

In order to study the interactions between E6 proteins and LXXLL motifs, I used the holdup technique, a chromatographic retention approach developed within the team, to detect and quantify the affinity of many protein-motif interactions in parallel (Charbonnier *et al.*, 2006). This method consists of immobilizing a peptide on resin before incubating it with the interaction partner (protein) in solution. The liquid fraction is recovered by filtration: it contains the protein fraction that did not interact with the peptide presented on the resin. The amount of protein remaining in the liquid fraction is measured by capillary electrophoresis before being compared with a negative control. The greater the depletion in the liquid fraction, the more protein retained on the resin with the peptide of interest and the highest the affinity. For each motif, an

interaction profile is established for a panel of domains. This method can be automated and applied at high throughput; it has already enabled an in-depth study of the interactions of the PBMs of E6 of HPV 16 and 18 with all 266 PDZ domains existing in the human proteome (Vincentelli *et al.*, 2015).

To complement the structure of HPV16 E6 in complex with the LXXLL peptide of E6AP previously published by the team (Zanier *et al.*, 2013), the precise interaction preferences of HPV16 E6 for a library of 45 single-point mutants of the LXXLL motif of E6AP were determined by the holdup method. These data were confirmed by SPOT peptide membrane interaction assays, a semi-quantitative approach often used to compare the interaction properties of many peptides (Mancilla and Volkmer, 2016). In parallel, the holdup technique was optimized: the capillary electrophoresis step, allowing to quantify the fraction of protein present in the liquid fraction, was replaced by a fluorescence approach, which has the advantage of being faster and less expensive for results of equivalent quality. Finally, the dissociation constants of all the mutated peptides could be estimated from the holdup results and SPR measurements. These interaction data allow to identify the key residues of the interaction and to design high affinity inhibitors: they are the subject of a manuscript currently under revision (**b**).

Interactions of different E6 oncoproteins with LXXLL motifs

I compared the interaction specificity of 6 E6 oncoproteins for a library of LXXLL peptides from various cellular proteins selected from the most documented HPV E6 targets in the literature. These E6 are derived from different types of HPV, representative of phylogenetic and phenotypic diversity: HPV16, 18 and 31 (alpha genus) are strongly involved in cervical cancer, HPV197 (gamma) is a cutaneous tropism HPV, HPV38 and 49 (beta) displayed oncogenic potential in murine models (Viarisio *et al.*, 2016, 2018). Our protein-motif interaction data are consistent with full-length protein interactions detected by affinity purification and mass spectrometry, confirming the validity of our fragmentary approach (Grace and Münger, 2017; White *et al.*, 2012a). I also took part in obtaining crystallographic structures of several E6 in complex with their main ligands LXXLL: 16E6/IRF3$_{LXXLL}$ (PDB : 6SJA), 18E6/mutant E6AP$_{LXXLL}$ (PDB :6SJV) and 49E6/MAML1$_{LXXLL}$ (PDB : 6SMV). MAML1 is targeted by HPV E6 proteins with cutaneous tropism: this interaction leads to the repression of signal transduction of the Notch pathway, which plays a central role in keratinocyte

differentiation (Brimer *et al.*, 2012). IRF3 is a transcriptional activator of the interferon pathway, which plays a role in innate immunity in response to viral infection (Ronco *et al.*, 1998; Suarez and Travé, 2018). These molecular and structural insights of the recognition variations of the different E6 proteins allow a better understanding of the diversity of phenotypes caused by the different types of HPV (manuscripts **c** and **d**).

A collaboration with Ignacio Bravo and Anouk Willemsen (Institut MiVEGEC, Montpellier) allowed us to study the specificities of interactions of several ancestral E6 oncoproteins reconstructed from sequence alignments (Willemsen and Bravo, 2019). The study focuses on ancestral E6 from a group including not only human papillomaviruses of the alpha genus but also some phylogenetically related monkey papillomaviruses. Four ancestral E6 oncoproteins were selected for *in vitro* testing of their interaction preferences with LXXLL peptides. Thanks to the parallel purification and holdup approaches, I was able to establish interaction profiles for all ancestral E6 and compare them with those of the well-known alpha genus HPV E6 (HPV16 and 18). The interactions with higher affinity were validated by SPR and identified several ancestral E6/LXXLL complexes of interest. These interaction data allow a better understanding of the development of interactions between E6 oncoproteins and cellular LXXLL motifs during the evolution of papillomaviruses (manuscript **e**).

Side project

The ubiquitin ligase E6AP is a major interaction target of HPV 16 E6. Both HPV16 E6 and HERC2, a ubiquitin ligase, have an allosteric activating on E6AP. Kühnle et al. observed a direct interaction between residues 150-200 of E6AP and RLD2, the second RLD (RCC1-like domain) of HERC2, for which no structure is currently available (Kühnle *et al.*, 2011). First, I confirmed by size exclusion chromatography the interaction between the RLD2 domain of HERC2 and the full-length E6AP protein. In a second step, I attempted to determine the minimal E6AP interaction motif that could bind the RLD2 domain. Since the E6AP region identified in the literature is predicted to be disordered, I used overlapping peptides to reduce the interface to the residues most involved in the interaction. Using the holdup and SPR approaches, I was able to restrict the interface with RLD2 to a 15-residue peptide. This project is now the subject of a full-fledged PhD thesis led by Auguste Demenge, who resolved several crystallographic structures of the RLD2 domain of HERC2 fused to the short motif of

E6AP. The study of this interaction will allow a better understanding of how the loss of E6AP and HERC2 functions can lead to pathologies such as Angelman syndrome. This neurodevelopmental disorder is caused by the loss of function of E6AP and HERC2 through mutations in the chromosomal locus where their genes are located (locus 15q11-13) (Kalsner and Chamberlain, 2015). There is therefore a double link between E6AP and HERC2: the physical interaction between the two proteins and their involvement in the same neurological disorder due to mutations in their gene sequence.

Conclusions

During my thesis, I could acquire skills for the methodological developments necessary for the biophysical study of unstable proteins *in vitro*. Thanks to this experience, I can now adapt to different biochemical systems and to design purification strategies and interaction tests optimized for a given protein. These biochemical optimizations allowed me to explore the LXXLL motif recognition preferences of several E6 oncoproteins that could not be isolated *in vitro* so far. Using biochemical, biophysical and structural approaches, I showed that each E6 oncoprotein of a given HPV type targets a certain set of cellular proteins via their LXXLL motif. These variations in specificity certainly contribute to the variations in tropism and pathological effects observed for different HPVs. The interaction data combined with the structures of the different E6/LXXLL complexes could be used as a basis for the development of specific inhibitors of E6 proteins for therapeutic purposes.

Publications et communications orales

1 Liste d'articles

(a) Bonhoure, A., Demenge, A., Kostmann, C., San José, L., De la Cal, E., Armisen, P., Nominé, Y., and Travé, G. (2018). One-step affinity purification of fusion proteins with optimal monodispersity and biological activity: application to aggregation-prone HPV E6 proteins. Microbial Cell Factories *17*.

(b) Bonhoure, A., Forster, A., Babah, K.O., Gógl, G., Eberling, P., Kostmann, C., Volkmer, R., Tapia Mancilla, V., Travé, G., and Nominé, Y. (2020). Benchtop holdup assay for quantitative affinity-based analysis of sequence determinants of protein-motif interactions. Analytical Biochemistry 603, 113772.

(c) Bonhoure A., Cousido-Siah A., Kostmann C., Mitschler A., Podjarny A. and Travé G. Quantitative analysis of LXXLL recognition features of HPV E6 oncoproteins, en préparation.

(d) Conrady M.C., Suarez I.P., Gógl G., Frecot D.I., Cousido-Siah A., Jiawen L., Kostmann C., Bonhoure A., Mitschler A., Stubenrauch F., Iftner T., Masson M., Travé G. and Simon C. Structure of high-risk papillomavirus type 31 E6 oncogenic protein and characterization of E6/E6AP/p53 complex formation, en révision.

(e) Willemsen A., Bonhoure A., Suarez I. P., Kostmann C., Travé G. and Bravo I. Evolution of LXXLL-motif interaction preferences of alpha HPV E6, en préparation.

Publication annexe

Aviolat, H., Nominé, Y., Gioria, S., Bonhoure, A., Hoffmann, D., Ruhlmann, C., Nierengarten, H., Ruffenach, F., Villa, P., Trottier, Y., and Klein, F.A.C. (2018). SynAggreg: A Multifunctional High-Throughput Technology for Precision Study of Amyloid Aggregation and Systematic Discovery of Synergistic Inhibitor Compounds. Journal of Molecular Biology https://doi.org/10.1016/j.jmb.2018.09.009.

2 Communications orales

- **Journées scientifiques de l'Association Francophone pour l'étude des Infections à Papillomavirus et Polyomavirus (AFIPP)**

  - 28 septembre 2016, Loches: Hijacking of human interactome by HPV oncoproteins: a functional quantitative analysis; Bonhoure A., Forster A., Charbonnier S., Altschuh D., Suarez I. P., Cousido-Siah A., Mitschler A., Podjarny A., Nominé Y. and Travé G.
  - 2-4 octobre 2019, Malbuisson: Hijacking of LxxLL motifs in human proteins by ancestral HPV E6 oncoproteins; Bonhoure A., Willemsen A., Travé G and Bravo I. **Prix de présentation orale.**

- **Retraite du département de Biologie Structurale Intégrative**

  - 22 novembre 2011, Strasbourg: Holdup assay for measuring protein-ligand affinities at high throughput: current status, applications to the team's thematics and future developments; Bonhoure A., Forster A., Charbonnier S., Altschuh D., Suarez I. P., Cousido-Siah A., Mitschler A., Podjarny A., Nominé Y. and Travé G.

- **Séminaire transversal IGBMC Integrated analyses of Gene Expression and the Epigenome**

  - 28 novembre 2017, Strasbourg: Deciphering HPV E6 protein-motifs interactions; Bonhoure A., Nominé Y. and Travé G.
  - 5 décembre 2018, Strasbourg: Structure and Interactomics of HPV E6 oncoproteins involved in various epithelial cancers; Bonhoure A., Nominé Y. and Travé G.

- **Symposium Israel-Unistra-Académie des Sciences**

  - 24 octobre 2018, Strasbourg: Linear motif-mediated interactions allow E6 oncoprotein from HPV to highjack host proteins; Bonhoure A., Nominé Y. and Travé G.

- **Conférence de la Société Française de Biophysique**

  - 5-8 novembre 2018, Carry-le-Rouet : Comprehensive quantitative study of LXXLL motif capture by E6 oncoproteins from HPV ; Bonhoure A., Nominé Y. and Travé G.

- **LMB-IGBMC Graduate Life Sciences Symposium**

  - 12 juillet 2018, Cambridge (Royaume-Uni): A fragmental approach to study protein interactions; Bonhoure A., Nominé Y. and Travé G.

- **DNA Tumour Virus Meeting**

  - 9-14 juillet 2019, Trieste (Italie) : Quantifying the affinity of LXXLL motif capture by E6 oncoproteins from Papillomaviruses ; Bonhoure A., Forster A., Ould Babah K., Suarez I.P., Cousido-Siah A., Mitschler A., Kostmann C., Demenge A., Podjarny A., Nominé Y. and Travé G.

3 Posters

- **Conférence Instruct Biennial Structural Biology**

  - 24 mai 2017, Brno (République Tchèque): Hijacking of human interactome by HPV oncoproteins: a functional quantitative analysis; Bonhoure A., Forster A., Charbonnier S., Altschuh D., Suarez I. P., Cousido-Siah A., Mitschler A., Podjarny A., Nominé Y. and Travé G.

- **Advances in neurodevelopmental and neurodegenerative disorders**

  - 7 juin 2018, Illkirch: Mapping of the residues involved in E6AP - HERC2 ubiquitin ligases interaction; Bonhoure A., Demenge A., Suarez I., Howard E., Cousido-Siah A., Mitschler A., Kostmann C., Roth A., Podjarny A. and Travé G.

- **Journées Jeunes Chercheurs en cancérologie de la Fondation ARC**

- 28 novembre 2018, Paris : La capture de motifs LXXLL par les oncoprotéines E6 de HPV; Bonhoure A., Nominé Y. and Travé G.

- **Protein-Protein Interactions Conference**

- 4 avril 2019, Leeds (Royaume-Uni): Hijacking of LXXLL motifs in human proteins by HPV E6 oncoprotein; Bonhoure A., Forster A., Ould Babah K., Altschuh D., Kostmann C., Suarez I. P., Cousido-Siah A., Mitschler A., Podjarny A., Nominé Y. and Travé G.

- **Next-Generation Protein Analysis and Detection**

- 2-3 décembre 2019, Gand (Belgique): Hijacking of LXXLL motifs in human proteins by HPV E6 oncoproteins; Bonhoure A., Forster A., Ould Babah K., Altschuh D., Kostmann C., Suarez I. P., Cousido-Siah A., Mitschler A., Podjarny A., Nominé Y. and Travé G.

Introduction

1 Biology of Papillomaviruses

1.1 Generalities

Papillomaviruses are small, non-enveloped viruses with double-stranded circular DNA. They infect the epithelia of mammals, but also birds, fish and reptiles. Some of them cause only warts, condylomas and benign proliferations called papillomas. Others can cause different types of cancer (cervix and other ano-genital cancers, head and neck cancers, skin cancers).

1.1.1 Historical background

Genital warts in humans have been the subject of writings, the oldest dating back to antiquity (Oriel, 1971). However, the causal link between papillomaviruses and warts could only be made from the twentieth century onwards, thanks to advances in scientific method. The papillomavirus was first detected in animals, particularly through the work of Richard Shope on rabbits (Shope and Hurst, 1933). In rabbits, the involvement of papillomavirus in skin cancers was quickly identified (Rous and Beard, 1935), whereas human papillomavirus (HPV) was considered to cause only benign warts and proliferations (papillomas). In 1950, HPV viral particles were visualized for the first time by electron microscopy (Strauss *et al.*, 1950). Studies of oncogenic viruses gained momentum in the 1970s with the first attempts to isolate viral DNA from tumor samples (Wolf *et al.*, 1975). Through his studies on warts and skin lesions, the French virologist Gérard Orth identified the first types of HPV (Orth *et al.*, 1977, 1978). The work carried out by the German physician Harald zur Hausen marked a turning point in the characterization of human papillomaviruses: after identifying a new type of HPV present in condylomas (Gissmann and zur Hausen, 1980), he was the first to isolate the genomes of HPV16 and HPV18 from cervical cancer biopsies (Boshart *et al.*, 1984; Durst *et al.*, 1983). These major discoveries as well as his articles highlighting the involvement of HPV in various human cancers (zur Hausen, 1991, 2002) earned him the Nobel Prize in Medicine in 2008.

In parallel, studies were carried out to characterize the properties of a papillomavirus infecting cattle: bovine papillomavirus (BPV) (Cheville and Olson, 1964; Pamukcu *et al.*, 1959), from which the genomes of the first two types were isolated in 1978 (Lancaster and Olson, 1978). It later became apparent that some of the characteristics

of BPV also apply to HPV (Munday, 2014). Studies have shown that BPV is capable of infecting not only cattle but also hamsters, donkeys and horses (Koller and Olson, 1972). BPV is particularly prevalent in horses: it causes invasive skin tumors called sarcoids, which can cause ulceration and impede walking depending on their location (Bogaert *et al.*, 2008; Taylor and Haldorson, 2013). To date, no HPV infections in other animals have been reported, with the exception of one study reporting amplification of HPV9-derived DNA in feline papillomavirus samples (Munday *et al.*, 2007). Since this result has not been reproduced since its publication, it is now probably attributed to contamination of the samples (Munday *et al.*, 2019). The similarities observed between HPV and BPV have made the latter a model organism for the study of HPV, both for the understanding of infectious and carcinogenic mechanisms and for the development of vaccines (Campo, 2006).

1.1.2 Classification

There is a great diversity of papillomaviruses, infecting not only humans but vertebrates in the broadest sense (mammals, birds, snakes, turtles, fish). Papillomaviruses are partly designated by the name of their host (*Felis domesticus* PV; Human Papillomavirus; Bovine Papillomavirus...).

Papillomaviruses were first classified in the family *Papovaviridae*, which included Papillomaviruses, Polyomaviruses including the simian virus 40 (SV-40). The characteristics at the origin of the creation of this taxon are the absence of an envelope as well as the genome in the form of double-stranded circular DNA. However, these viruses present quite distinct genomic organizations. In 2002, the International Committee for Virus Taxonomy (ICTV) decided to split the *Papovaviridae* group into two groups: *Papillomaviridae* and *Polyomaviridae* (de Villiers *et al.*, 2004). However, it should be noted that Papillomaviruses and Polyomaviruses have significant protein sequence homologies and the identification of papilloma-polyomavirus hybrids BPCV1 and 2 supports the hypothesis of a common origin (Rector and Van Ranst, 2013).

This example illustrates the positive impact that technological developments have had on the study of papillomaviruses since their identification, allowing a better understanding of their properties. The development of cell lines allowing the culture of HPV *in vitro* has also facilitated the identification of new viral types (de Villiers *et al.*, 2004). Finally, the first DNA sequencing methods made it possible to clone and then

sequence complete HPV genomes, starting with HPV1 in the early 1980s (Danos *et al.*, 1980, 1982). Since the 2010s, new generation sequencing at high throughput has allowed a tenfold increase in the number of HPV types identified (Arroyo *et al.*, 2013), in particular HPVs with skin tropism (Ekström *et al.*, 2011).

The current classification is based on the sequence homology of the gene coding for the L1 capsid protein, since it is the best conserved gene in the viral genome. Papillomaviruses are grouped in the family *Papillomaviridae*, divided into genera designated by Greek letters (alpha, beta, gamma, mu, nu...), which are further subdivided into species and types. HPVs belonging to the same genus have at least 60% nucleotide sequence identity in the open reading frame of L1, compared to 70 to 80% for HPVs of the same species. A new HPV type is defined by at least 10% sequence variation from any other known type. Finally, genomes with less than 10% L1 sequence difference are called variants (de Villiers, 2013).

Each isolate of a potential new type of HPV is sent to the HPV International Reference Centre, currently housed at the Karolinska Institute in Stockholm, Sweden. The role of this centre is to confirm, record and reference all new HPV types in the phylogenetic tree and to store DNA samples of all known HPV genomes (Bzhalava *et al.*, 2015). Recently, a reference center for Papillomaviruses infecting animals has been established at the University of Arizona, Tucson, USA (Van Doorslaer and Dillner, 2019). In addition, the Papillomavirus Episteme (PaVE) is a database that was created to compile all annotated genomes and to provide tools for sequence alignment and construction of phylogenetic trees specifically dedicated to papillomaviruses (Van Doorslaer *et al.*, 2017). To date, 227 reference HPV types have been identified and 215 papillomaviruses targeting animals (Papillomavirus Episteme, pave.niaid.nih.gov, accessed 25 March 2020). **Figure 1** shows the phylogenetic tree of HPV identified to date.

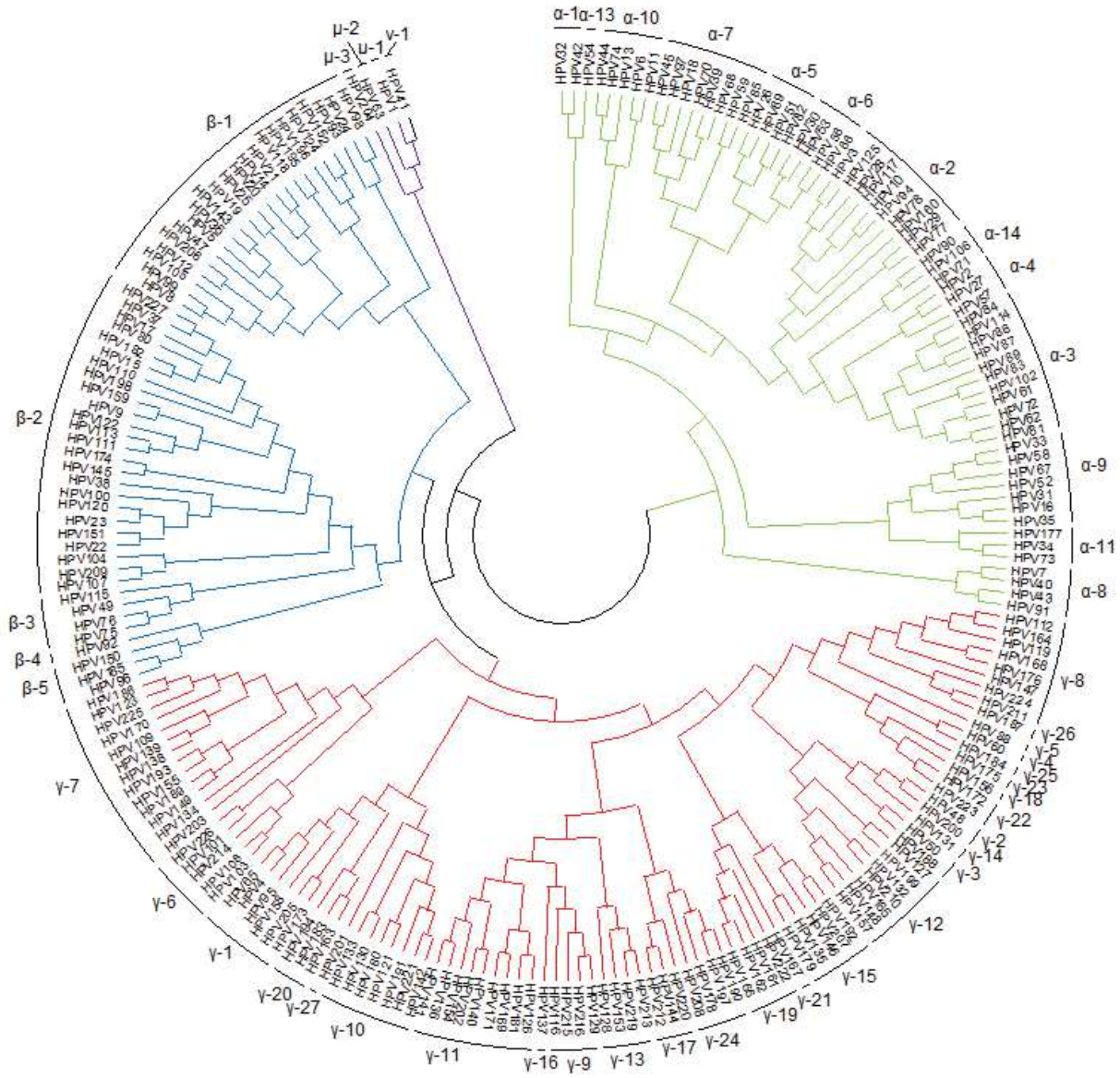


Figure 1: Phylogenetic tree of HPV, according to the International HPV Reference Centre (hpvcenter.se, accessed 25 March 2020). The genera alpha, beta, gamma, mu and nu are shown in green, blue, red and purple respectively. Species are indicated by numbers.

Given the wide variety of pathological effects caused by HPV, other classifications based on their oncogenic risk as well as on their tropism exist on the fringes of the taxonomic classification.

The so-called "high risk" HPVs can cause cancers while the "low risk" HPVs cause only benign proliferations (condylomas, warts). **Table 1** summarizes all HPVs considered to be carcinogenic, including the types recognized in 2012 by the World Health Organization as carcinogenic substances.

Oncogenic risk	Genus	Species	HPV types
High	Alpha	5	51
		6	56
		7	18 ; 39 ; 45 ; 59
		9	16 ; 31 ; 33 ; 35 ; 52 ; 58
Probably high		7	68
Possibly high		5	26 ; 69 ; 82
		6	30 ; 53 ; 66
		7	70 ; 85 ; 97
		9	67
		11	34; 73
	Beta	1	5*; 8*
Low	Alpha	1	54
		8	40;43
		1	42
		13	54

Table 1: Classification of HPV by oncogenic risk. Based on (Muñoz *et al.*, 2006) and data from the International Agency for Research on Cancer (monographs.iarc.fr, accessed 25 March 2020). * HPV5 and 8 are known to be carcinogenic only in patients with Epidermodysplasia verrucciformis.

On the other hand, HPVs are distinguished according to their tissue tropism. HPVs with skin tropism infect keratinized epithelia, such as the skin. HPVs with mucosal tropism target non-keratinized epithelia, such as the uterine mucosa or the oropharynx. It is noteworthy that most mucosal HPVs belong to the phylogenetic genus alpha, while the beta, gamma, mu and nu genera comprise a majority of cutaneous HPVs. However, this is not an absolute rule. Some alpha HPVs have skin tropism (HPV2, 3, 7, 10, 27, 28, 57) and some types such as beta 3 seem able to localize in both types of epithelia (Gheit, 2019; Hampras *et al.*, 2017).

1.1.3 Capsid, genome and infectious cycle

1.1.3.1 Capsid

Papillomaviruses are non-enveloped viruses with a capsid diameter of about 60 nm. The atomic structure of the capsids of HPV1 and BPV1 was first determined in 1991

by cryo electron microscopy (Baker *et al.*, 1991). The capsid is icosahedral and has a symmetry T = 7. It consists of 360 copies of the L1 capsid protein and up to 72 copies of the L2 protein, organized into 72 pentamers (Li *et al.*, 2016).

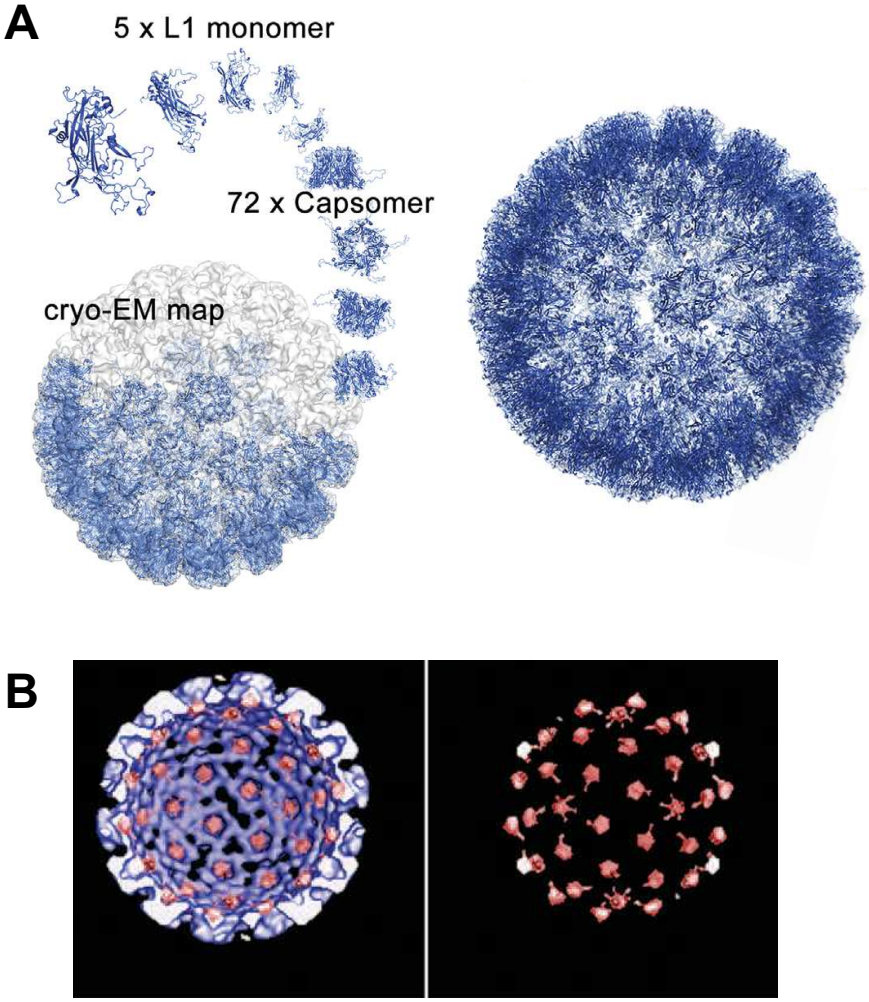


Figure 2 : HPV capsid structure and positioning of L1 and L2 proteins. **A**. HPV59 capsid structure reconstituted with the major L1 protein only. The capsid is composed of 72 pentamers of L1. The structure was obtained by combining crystallography data with a cryo electron microscopy map. Adapted from (Li *et al.*, 2016). **B**. Position of the L2 minor capsid protein in an HPV16 capsid. The L2 protein units are shown in red, superimposed with the inside of a capsid composed of L1 protein only (left) or showing only L2 (right). Adapted from (Buck *et al.*, 2008)

The capsid can be reconstituted *in vitro* by expressing only the major L1 protein or by co-expressing the L1 and L2 proteins (Hagensee *et al.*, 1993). The L2 protein is mostly buried in the capsid, with the exception of its N-terminus (Buck *et al.*, 2008; Guan *et al.*, 2017).

1.1.3.2 Genome

Papillomaviruses have a circular double-stranded DNA genome, about 7-8 kb. There are three distinct zones in this genome: the Upstream Regulatory Region (URR), the Early Region (Early) and the Late Region (Late) (**Figure 3**).

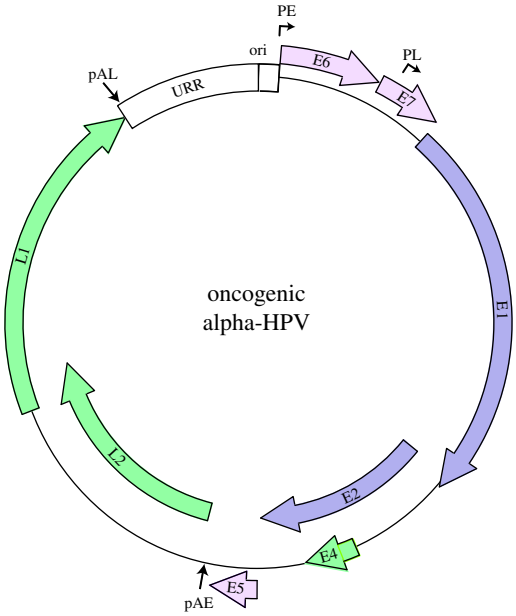


Figure 3: Organization of the genome of oncogenic HPV genus alpha. The URR regulatory region is shown in white, the early region in purple and the late region in green. The polyadenylation signals of the early and late regions are denoted pAE and pAL respectively. Similarly, promoters that control the expression of the early (PE) and late (PL) regions are indicated. The origin of replication (ori) is included in the regulatory region. The organization of the genome varies according to the phylogenetic genus of HPV (Doorbar, 2018). Adapted from (McBride, 2017).

The three regions are separated by polyadenylation signals: one at the end of the early region (pAE), the second at the end of the late region (pAL). The genome is very compact: several open reading frames intersect and the viral genome produces polycistronic transcripts that undergo alternative splicing (Johansson and Schwartz, 2013). The various viral proteins encoded by the early and late regions are indicated on **Table 2**.

Protein	Best known activity
E1	Acts in synergy with E2 to form the initiation complex of viral DNA replication. Helicase binding to the origin of genome replication.
E2	DNA binding protein, enables loading of E1 helicase, binds the viral genome to the host cell chromosomes, transcriptional regulator.
E8^E2	DNA binding protein, repressor which limits productive replication.
E1^E4 or E4	Late protein, reorganizes the keratin network in differentiated cells to facilitate viral transmission.
E5	A protein encoded only by HPV genus alpha, promotes cell proliferation by activating growth factor receptors and allows the virus to escape the immune system.
E6	An oncoprotein that disrupts cellular functions through protein-motif interactions, either by capturing the LXXLL motifs of host proteins or by interacting with PDZ-domain proteins in mucosal high-risk HPVs. May lead some proteins to their degradation, such as p53 in the case of HPV16.
E7	An oncoprotein promoting keratinocyte proliferation in synergy with E6, equipped with a LXCXE motif targeting proteins of the Rb family, and a conserved region CR3 mediating interaction with PTPN14 and 21.
L1	Major capsid protein, 360 copies per virion.
L2	Minor capsid protein, up to 72 copies per virion. L2 binds to DNA and participates in transporting the pseudogenome to the nucleus as well as compacting the genome in the viral capsid.

Table 2: Presentation of the proteins encoded by the HPV genome. Adapted from (McBride, 2017).

The genome also contains two promoters, one of which is located before the sequence coding for E6 and controlling the expression of almost all genes in the early region (PE) (Zheng and Baker, 2006). The second promoter is located in the open reading frame of E7 and controls the expression of the late region (PL). This PL promoter is activated only in differentiated keratinocytes while the PE promoter is activated in basal epithelial cells, corresponding to viral expression during the infection cycle as shown in **Figure 4**.

1.1.3.3 Infectious cycle

1.1.3.3.1 Entry into the host cell

Papillomaviruses infect stratified, mucosal or cutaneous epithelia depending on their tropism. The virus infects the cells of the basal layer by infiltrating through micro-abrasions (Roberts *et al.*, 2007). Through the process of healing and tissue renewal, these basal cells divide before migrating to the surface of the epithelium and differentiating. The HPV infection cycle follows the progression of these cells: their

differentiation induces strong viral replication as well as the expression of viral genes. The virions are finally assembled in the upper layers before being released into the environment as squames (Graham, 2017). In the case of infections leading to cervical cancer, the high-risk mucosal HPVs may target a very specific cell population: the squamous cells of the junction between the endocervix and the ectocervix (Herfs *et al.*, 2012).

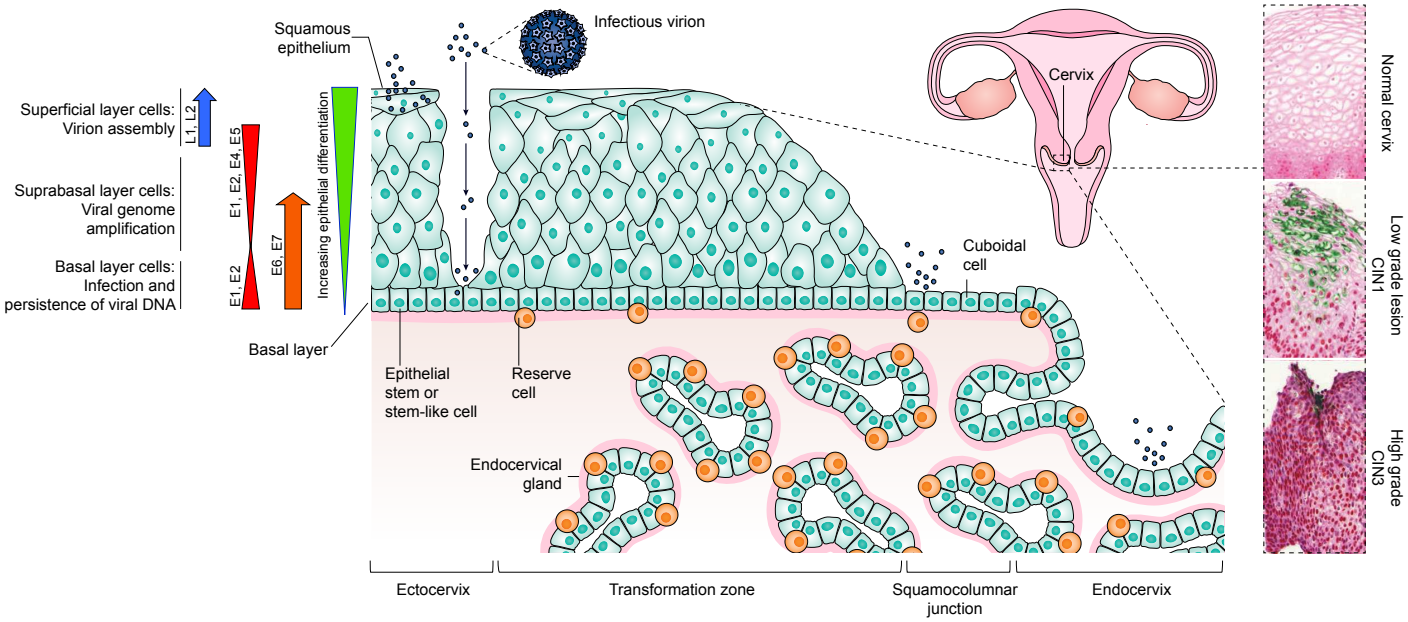


Figure 4: Infectious cycle of HPV on the cervix epithelium. Figure modified from (Graham, 2017) and (Schiffman *et al.*, 2016). HPV infects epithelia at the level of microabrasions or, in the case of the cervix, by targeting cuboid cells and reserve cells (squamo-cellular junction) or cells of the endocervix. By infecting these poorly differentiated cells that are in active cell division to ensure epithelial renewal, the virus ensures persistence of viral DNA by maintaining a low number of copies of its genome in a set of cells close to the basement membrane. Thanks to the action of oncoproteins E6 and E7, viral amplification is activated in cells that have migrated to the suprabasal layer. Finally, the expression of the capsid proteins L1 and L2 in the superficial layers of the epithelium allows the assembly of virions which are then released into the environment by desquamation. The immunohistochemistry images on the right side of the image show a healthy cervix, as well as cervical intraepithelial neoplasia (CIN) grade 1 and 3. The HPV E4 protein is stained green and the red staining indicates a cell cycle marker (minichromosome maintenance protein complex).

On the surface of the viral capsid, the L1 protein is exposed to the solvent and interacts with heparan sulfate proteoglycans (HSPG) present on the basement membrane. Lysine residues 278 and 361 of the L1 protein are the points of attachment to the HSPG (Raff *et al.*, 2013). Cyclophilin B is an isomerase related to HSPG, in particular syndecan-1. It induces a conformational change in the viral capsid protein L2, which facilitates cleavage of the N-terminus of L2 by a cellular endoprotease, furin and/or PC5/6 (Bienkowska-Haba *et al.*, 2009; Kines *et al.*, 2009; Richards *et al.*, 2006). To date, no model describing the signalling pathways leading to internalization of HPV has been unanimously established, but several receptors potentially involved have been identified: integrin α6, epidermal growth factor and keratinocyte receptors, tetraspanins, S100A10 subunit of the heterotetramer Annexin A2 (Raff *et al.*, 2013; Woodham *et al.*, 2012). The internalization pathway appears to involve caveoline 1 and dynamine 2, as demonstrated for HPV31 (Smith *et al.*, 2007). The viral particle is then transported by the endosomal route. During this transport, the viral capsid dissociates under the effect of acid pH and cell proteases and chaperones. The L2 protein having the ability to bind to a DNA molecule without recognizing a specific sequence (Zhou *et al.*, 1994), it forms a complex with the pseudogenome which is then transported to the nucleus (Aksoy *et al.*, 2017). The L2-pseudogenome complex enters the nucleus while the host cell is in mitosis, during nuclear envelope breakdown (Aydin *et al.*, 2014).

1.1.3.3.2 Genome maintenance

The viral genome is maintained in the nucleus of basal epithelial cells as an episome, defined as circular extrachromosomal DNA capable of autonomously replicating and integrating into the host genome (Mellin *et al.*, 2002).

The HPV synchronizes the replication of its genome with that of the infected cell, which in the case of a basal cell is poorly differentiated and actively divides. The viral proteins E1 and E2 initiate viral replication: the early protein E2 recruits and positions the helicase E1 on the origin of replication. The E1 protein unwinds the DNA and recruits the cell replication machinery (topoisomerase I, DNA polymerase α primase, DNA polymerase δ ...). The E2 protein binds the viral genome to the chromosomes of the two daughter cells, which allows the distribution of copies of the viral genome in the daughter cells, thus leading to persistent infection in an increasing number of cells

(McBride, 2008). Initially, amplification is limited to 10 to 50 copies per cell to avoid activating the immune response in the basal cells. The E2 protein acts as a repressor of the early PE early region promoter by blocking access to transcriptional factors and altering the conformation of the chromatin (Graham, 2017).

1.1.3.3.3 Genome amplification

The papillomavirus activates the multiplication of the host cell through the action of oncoproteins E6 and E7. Their expression is activated in the basal and suprabasal layers of the epithelium, then it decreases in the superficial layers (**Figure 4**). This mechanism compensates for the progressive decrease in cell multiplication as cells differentiate.

On the one hand, the E6 oncoprotein from alpha HPV (including high-risk HPV) captures the host protein E6AP (E6-Associated Protein) by binding its LXXLL motif (Huibregtse *et al.*, 1991, 1993). The E6AP protein is an E3 ubiquitin ligase, it is involved in the last step of the ubiquitination reaction and is responsible for the recognition of the protein substrate to which a ubiquitin molecule will be attached. Its interaction with the viral protein E6 modifies its interaction specificities. Thus, the heterodimer E6/E6AP binds the tumor suppressor p53 which is then ubiquitinated, causing its degradation by the proteasome (Scheffner *et al.*, 1990, 1993). E6 proteins have other activities that are will be detailed in section 2.1.

In high-risk alpha HPV, the E7 oncoprotein binds to the tumor suppressor Rb (Retinoblastoma Protein) and disrupts its interaction with E2F transcription factors (Liu *et al.*, 2006; Slebos *et al.*, 1994). In differentiating keratinocytes, cell division is suppressed by the inhibitory action of hypophosphorylated Rb associated with E2F. The interaction of the E7 protein removes the inhibition exerted by Rb and allows the cell to move from the G1 phase to the S phase of the mitotic cycle (Giarrè *et al.*, 2001). The additional activities mediated by E7 oncoprotein will be addressed in section 2.2.

By keeping infected cells in a state of active division, HPV drastically amplifies its viral genome, reaching several thousand copies per cell (Doorbar *et al.*, 2015). In order to carry out viral replication, the E1 and E2 proteins are highly expressed, as indicated on **Figure 4** (Graham, 2017).

1.1.3.3.4 Viral particle assembly and release

When infected cells migrate to the superficial layers of the epithelium, the phase of productive viral replication is initiated. Through changes in splice sites, transcripts produced from the PL promoter to the pAL polyadenylation signal result in the expression of the capsid proteins L1 and L2 (**Figure 3**). The L2 protein is recruited at the replication regions by the E2 protein, which allows the encapsidation of the genome. Viral maturation takes place in the senescent keratinocytes closest to the surface, where the environment is highly oxidative. In this context, numerous disulfide bridges are formed between L1 proteins which assemble to form extremely stable capsids. The E4 protein facilitates the release of infectious virions by disrupting the keratin network (Doorbar *et al.*, 2012).

1.1.3.3.5 Genome integration

Integration of viral DNA into the host genome is not a normal part of the HPV life cycle. On the contrary, such an event marks the end of the infectious cycle, as the linearized viral genome integrated into the chromosomal DNA is no longer compact enough to be encapsulated and transmitted to a new host. Chromosomal DNA integration appears to be a rare event that occurs randomly at so-called fragile sites with some homology to sequences in the viral genome (Schmitz *et al.*, 2012; Wentzensen *et al.*, 2004).

In most invasive HPV-related cancers, viral DNA is integrated into the human genome. This event usually causes increased expression of the E6 and E7 oncoproteins, which then stimulate long-term cell proliferation and lead to carcinogenesis. The most commonly proposed mechanism to explain this deregulation is the altered inhibitory function of E2 on E6 and E7 coding regions. A defect in the E2 gene sequence or alteration of the binding sites recognized by E2 may reverse the transcriptional inhibition exerted by this protein (**Figure 5**).

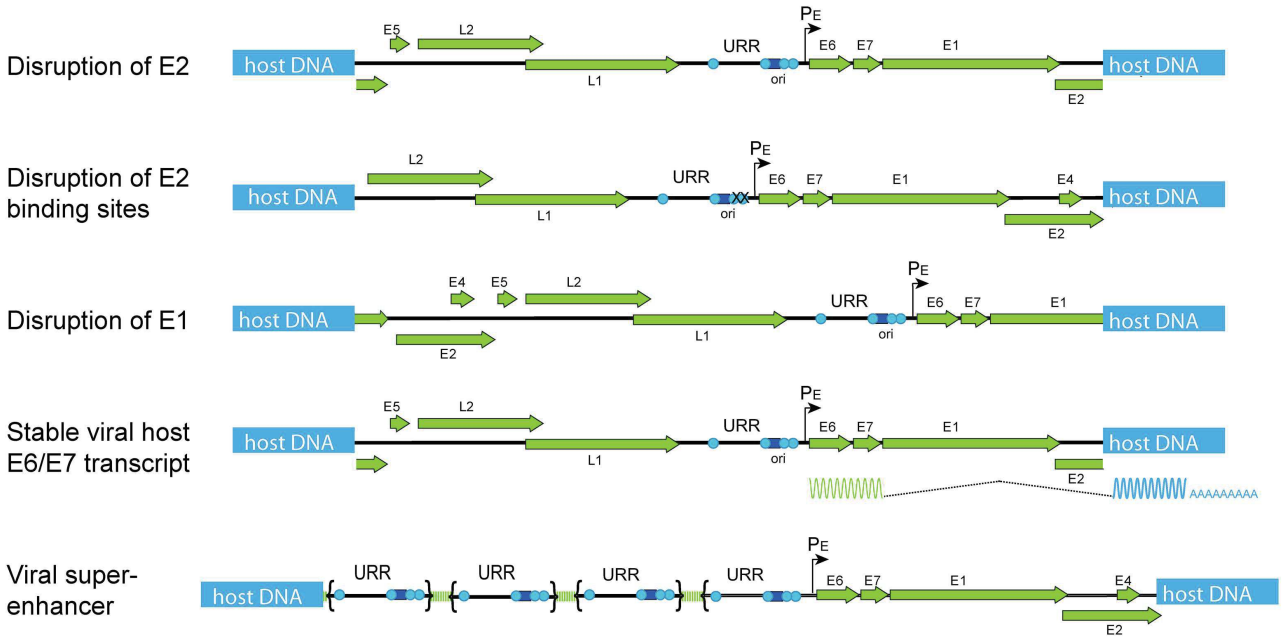


Figure 5: Models of the various modes of integration in the host genome leading to carcinogenesis. The regulatory region (URR) contains the E2 binding sites (indicated by light blue circles) and the origin of replication of the viral genome (indicated by a dark blue square). The promoter that regulates the early region is labelled PE. Adapted from (McBride and Warburton, 2017).

However, this is not the only mechanism that can lead to carcinogenesis through overexpression of oncoproteins. Alteration of the E1 gene can lead to DNA damage and increased genomic instability at the locus of integration. The formation of a transcript that fuses the E6 and E7 sequences with sequences from the host often stabilizes the RNA and promotes its translation. Duplication of the URR regulatory region upstream of the E6 and E7 genes can form an enhancer that amplifies the expression of both oncogenes (McBride and Warburton, 2017).

Furthermore, it is noted that not all cancers caused by HPV and notably by the oncogenes E6 and E7 are systematically associated with events of integration of viral DNA into the host genome. A recent study carried out on mouse models has suggested a hit-and-run mechanism of the E6 and E7 oncoproteins of HPV38 of the beta genus: after a transient expression of the oncogenes in the basal layer of the cutaneous epithelium coupled with ultraviolet exposure, a cancerous phenotype is established which is not attenuated by the deletion of the E6 and E7 genes (Viarisio *et al.*, 2018).

1.2 HPV pathologies and treatment

1.2.1 Epidemiology

Each year, more than 311 000 women die from cervix cancer, which is the fourth most common female cancer after breast cancer (627 000 deaths per year worldwide), lung cancer (576 000 deaths) and colorectal cancer (397 000 deaths) (gco.iarc.fr). Mucosal high-risk HPVs are responsible for 8.6% and 0.8% respectively of cancers diagnosed in women and men worldwide. The most affected areas in the world are developing countries, where 85% of HPV cancer deaths occur (**Figure 6**). This difference is partly due to limited access to preventive measures such as prophylactic vaccines and screening smears. In developed countries, these measures can prevent up to 80% of cervical cancers. Similarly, access to treatments that apply when the cancer is at a very advanced stage is limited in developing countries where the death rate is very high (World Health Organization).

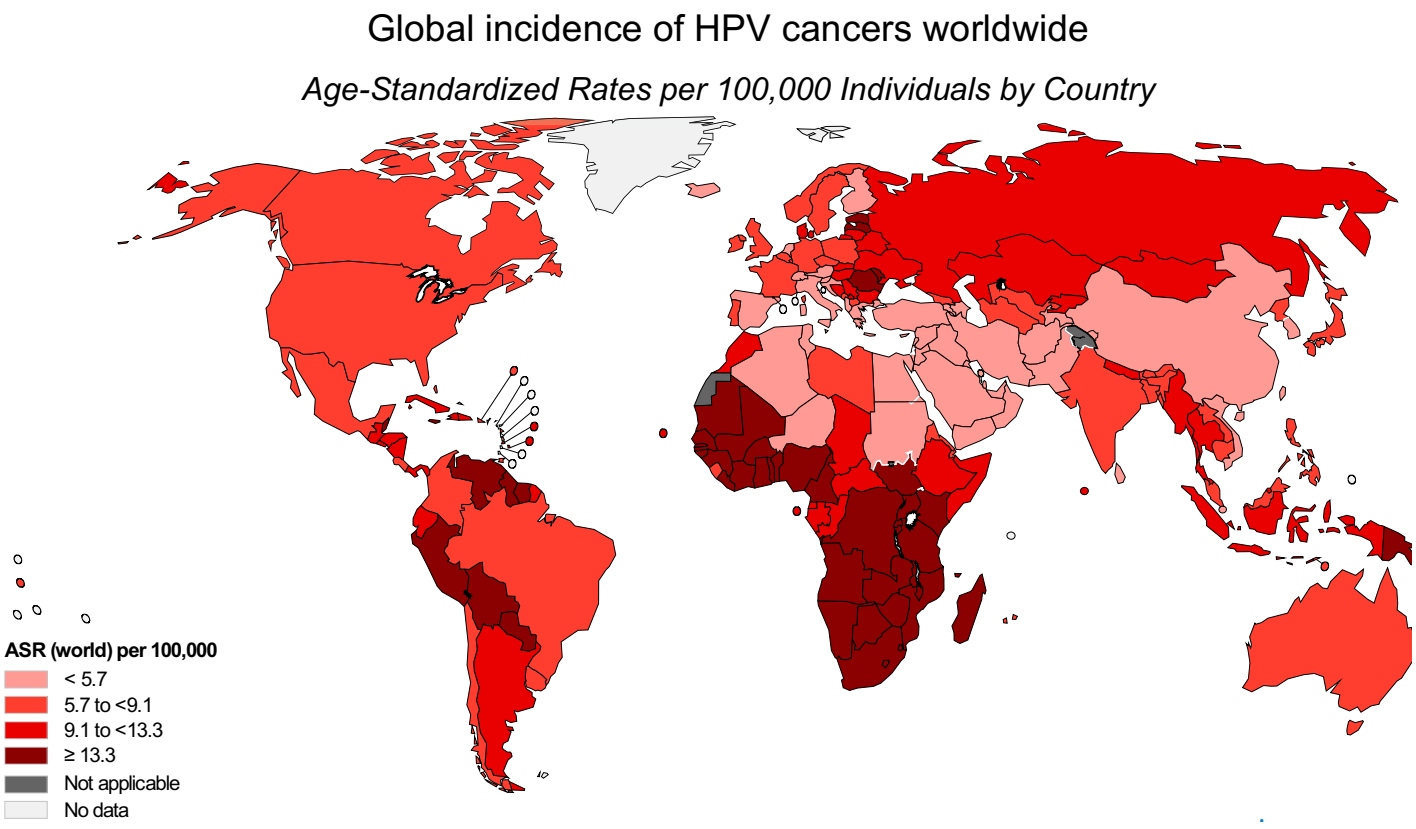


Figure 6 : Incidence of cancers attributed to HPV worldwide. Data from (de Martel *et al.*, 2020), formatting of the map on the site of the World Cancer Observatory managed by the International Agency for Research on Cancer (gco.iarc.fr, accessed 9 April 2020).

High-risk mucosal HPVs are the main etiological agents of cervical cancer, which is the fourth most common cancer in women. They are also responsible for 88% of anal cancers, 25% of vulvar cancers, 78% of vaginal cancers, 50% of penile cancers and 35% of head and neck cancers (oropharynx, oral cavity, larynx) (de Martel *et al.*, 2017). For these different cancers, we note that certain types of HPV have a higher prevalence, even among HPVs classified as high mucosal risk. HPV16 and HPV18 are responsible for 61% and 10% of cases of uterine cancer, respectively (**Figure 7**). HPV16 is the HPV type with the highest prevalence for anogenital cancers (vulva, vagina, anus, penis). It is also responsible for 71% of head and neck cancers, which includes the oral cavity, pharynx and larynx.

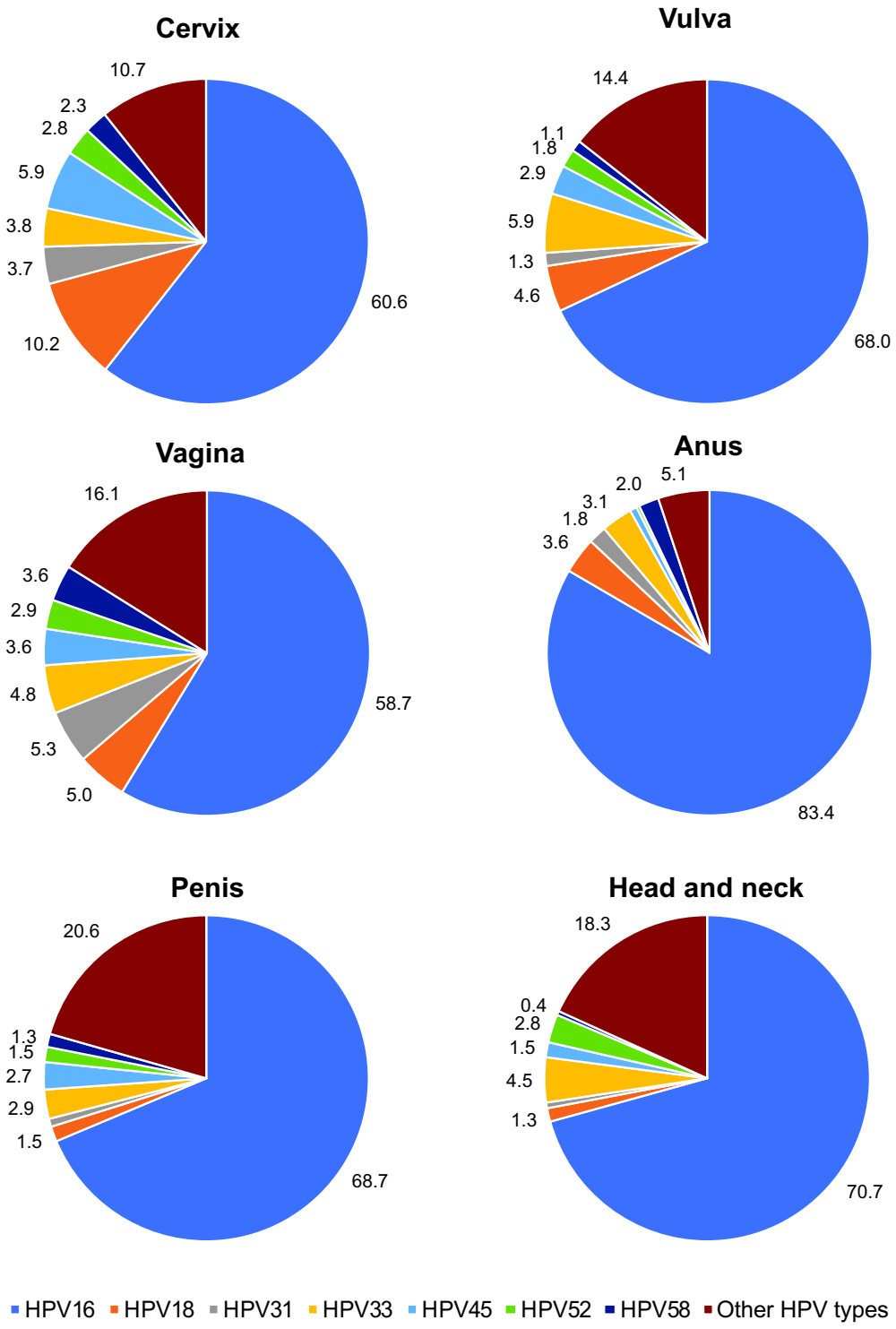


Figure 7 : Prevalence of different types of high-risk mucosal HPV in cervical, vaginal, vulvar, penile, anal and head and neck cancers, expressed as percentages. The data for anal cancer corresponds exclusively to female population. Worldwide data from (Alemany *et al.*, 2016; Castellsagué *et al.*, 2016; Serrano *et al.*, 2015).

On the other hand, cutaneous HPVs cause warts, which can be located on the face, hands, feet, elbows, knees and genitals. HPV2 is the cause of most vulgar warts: located on the hands, they disappear spontaneously in 65% of cases. Plantar warts

are caused by either HPV1 or HPV2 (mosaic wart). Butcher's warts are a type of vulgar wart related to HPV7 that particularly affects the hands of people handling meat or fish. Treatment is often difficult for them because of the 50% recurrence rate (Aubin and Guerrini, 2007). Concerning genital warts or condylomas, 75 to 90% are caused by the low-risk mucosal HPVs HPV6 and HPV11. With 1% of the sexually active American population affected, these warts represent a real public health problem. If left untreated, HPV infection can be linked to the development of warty carcinoma, a low-grade, well-differentiated cancer that rarely metastasizes. Buschke and Löwenstein's tumor or giant condyloma acuminata is a type of verrucous carcinoma located in the anogenital region and is often associated with HPV6 and HPV11 infection (Yanofsky *et al.*, 2012). In addition, HPV types 6 and 11 are associated with recurrent respiratory papillomatosis: this rare respiratory disease is characterized by the presence of papillomas (small benign tumors) in the mucous membranes of the upper aerodigestive tract, particularly the larynx. Usually benign, in children it can reach the trachea and bronchial tubes, which can be fatal if the airways become blocked. In its infantile form, the HPV infection that causes HPV occurs through placental transmission from mother to fetus: the risk of developing the disease a few months or years after birth is 231 fold higher if the mother had anogenital warts during pregnancy (Fusconi *et al.*, 2014).

Epidermodysplasia verrucciformis (EV), also known as Lutz-Lewandowsky syndrome, is a rare skin disease that manifests itself as a defect in primary immunity and increased susceptibility to beta HPV infections. It is an autosomal recessive genetic disease. In 75% of cases, it is caused by mutations in the EVER1 and EVER2 genes located on chromosome 17: these genes code for membrane proteins that play a role in regulating intracellular zinc levels and whose loss of function drastically increases the sensitivity of the skin to HPV (Cardoso and Calonje, 2011). The HPV types implicated are mainly HPV beta 1 species: HPV5 and HPV8 are most often involved, but HPV14, 20, 47 as well. Similarly, some HPVs of beta 2 species are involved, e.g. HPV38, as well as HPVs of beta 3 species, such as HPV49. Finally, some HPVs of the alpha genus may also be involved, such as HPV3 and HPV10. The disease usually manifests itself before the age of ten years with the appearance of flat warts and scaly macules that persist throughout the patient's life. The appearance of these scaly, scaly

lesions is reminiscent of the bark of a tree, giving EV a nickname: "the man-tree disease".

Figure 8 : Abul Bajandar, EV patient nicknamed "the tree man". Originally from Bangladesh and then aged 26, Abul Bajandar underwent a total of 16 surgeries in 2015 to remove about 5 kg of skin growths from his hands. Unfortunately, the disease recurred in the years that followed, despite the multiple operations performed. Source of the image maxisciences.com, accessed 13 April 2020.

People with EV have an increased risk of developing squamous cell carcinoma, particularly in areas of skin exposed to the sun. Each tumor cell contains multiple copies of the viral genome in the form of episomes. In 90% of patients with squamous cell carcinoma and EV, the HPV5 and 8 genomes have been identified (Accardi and Gheit, 2014). These two types of HPV have been classified as possible carcinogens in patients with EV (**Table 1**). It was reported that, while patients with EV are susceptible only to HPV infections, they are not particularly susceptible to other pathogens. Humoral immunity is preserved: antibodies to E6 and E7 proteins are detected in 70% of patients, which also indicates that the expression of these oncoproteins is required for the cancerous progression of the disease (Orth, 2010).

First described 1964 (Zuelzer *et al.*, 1964), the WHIM syndrome is an extremely rare congenital immune deficiency. This syndrome is named after its main clinical manifestations, which are Warts, Hypogammaglobulinemia, bacterial Infections and Myelokathexis. Hypogammaglobulinemia is defined by an insufficient amount of gamma globulin produced in the blood and myelokathexis is the retention and apoptosis of mature neutrophils in the bone marrow. The WHIM syndrome is related to heterozygous autosomal dominant mutations in the gene *CXCR4*, which encodes for the chemokine receptor type 4 CXCR4. CXCR4 is the receptor of CXCL12

chemokine, which is constitutively expressed in the bone marrow. CXCL12 regulates the release and clearance of bone marrow neutrophils into the blood (Bachelerie, 2010). The mutations causing the WHIM syndrome mainly affect the C-terminus of CXCR4 receptor and result in impaired desensitization and internalization of CXCR4 in response to CXCL12 (Balabanian *et al.*, 2005). Patients suffering from this disease have an increased susceptibility to HPV infections, which cause warts, skin and mucosal lesions. Depending on their severity, some lesions can evolve in metastatic carcinomas. As for EV patients, these lesions can be recurrent and require reinforced medical monitoring and surgical ablations (Kawai and Malech, 2009). A study not yet published reported the identification of HPV genomes from skin samples originating from a patient with WHIM syndrome (Molet, 2018). A total of 16 distinct HPV types were identified, including 9 β-2 HPV types with a variant of HPV23, named $HPV23_{WHIM}$, being the more represented in all biopsy samples. The sequences encoding for the oncoproteins E6 and E7 of $HPV23_{WHIM}$ were identified and conserved in samples collected several years apart. This result suggests that the defect in CXCR4/CXCL12 axis may facilitate the infection and long-term persistence of HPVs, in particular $HPV23_{WHIM}$.

1.2.2 Cervical cancer: Infection, persistence and cancer progression

Cutaneous HPV infection is spread by skin contact while mucosal HPVs infection is transmitted through sexual intercourse. Hence, the modes of contact of these two types of HPV imply distinct natural histories. Infection by mucosal HPV is relatively common: it is estimated that more than 70% of sexually active men and women will come into contact with HPV at least once in their lifetime. In most cases of mucosal tropism HPV, the infection occurs during the first sexual intercourse and remains asymptomatic. Thereafter, the viral load decreases until the virus becomes undetectable 5 years after infection. In 90% of cases, this decrease is a sign of viral clearance, i.e. the total elimination of viral particles from the tissue. For the remaining 10%, the virus is still present at a low level in the latent phase (papillomavirus.fr, consulted on 13 April 2020). The likelihood of virus persistence and precancerous progression depends in part on the type of HPV: the risk of cancer associated with HPV16 is much higher than for other high-risk HPVs (**Figure 7**). The host immune response is also a determining factor: patients infected with the human immunodeficiency virus (HIV) have a higher incidence of HPV-induced cancers (de

Martel *et al.*, 2015). Finally, behavioral factors such as smoking, multiparity and long-term use of hormonal contraceptives may have a moderate influence on cancer risk (Schiffman *et al.*, 2016).

The progression from healthy cervical epithelium to squamous cell cancer goes through intermediate stages of precancerous lesions called cervical intraepithelial neoplasia (CIN). These lesions are characterized by an increasing number of cells with morphological abnormalities such as epithelial stratification defects, nuclear abnormalities or differentiation defects. The different grades of CIN can be defined by the proportion of dysplastic cells in the cervical epithelium. In the case of a low-grade CIN1 lesion, dysplastic cells are present in the lower third of the epithelium, whereas they occupy two thirds and up to the entire epithelium for high-grade CIN2 and CIN3 lesions (**Figure 9**).

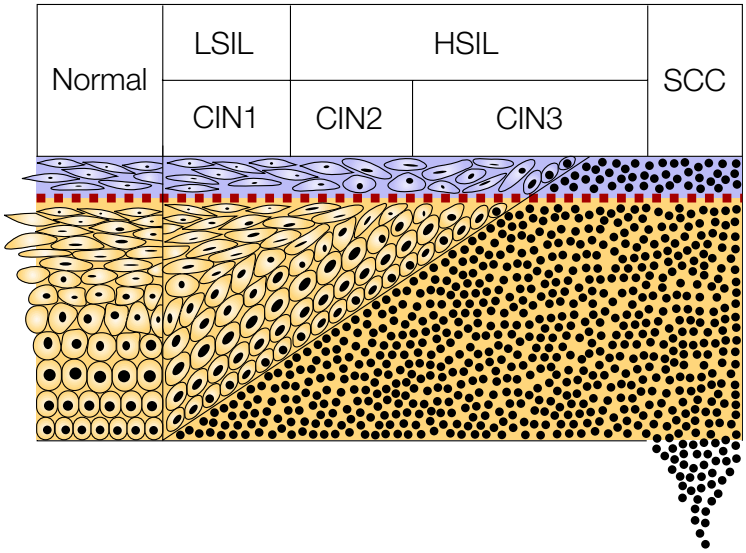


Figure 9: Intraepithelial neoplastic lesions of the cervix: classification by third party system. Progression from the epithelium of the normal cervix to squamous cell carcinoma (SCC) is represented by the increasing number of dysplastic cells. For low grade squamous intraepithelial lesions (LSIL) or grade 1 intraepithelial neoplasia (CIN1), dysplastic cells are located in the lower third of the epithelium. High grade lesions (HSIL) are characterized by dysplastic cells occupying two-thirds (CIN2) or more (CIN3) of the epithelium. Adapted from (Baldwin *et al.*, 2003).

Low grade CIN1 lesions are associated with spontaneous regression that does not require any particular treatment. The rate of CIN1 regression ranges from 70-80% for all women, increasing to 90% for patients under 25 years of age. In contrast, 0.2-4% of CIN3 progress to invasive cancer within 12 months. It is important to note that an

average of 25 to 30 years separates HPV infection from the development of invasive cancer (Martin and O'Leary, 2011).

1.2.3 Preventive and curative measures

There are a number of measures that can be taken to limit a woman's risk of cervical cancer. First, preventive measures can limit HPV infection to reduce the risk of developing cancer. These include condom use, which appears to reduce the risk of cervical and genital HPV infection in women (Winer *et al.*, 2006) and men (Pierce Campbell *et al.*, 2013). On the other hand, prophylactic vaccines have been developed to prevent the infection of different types of HPV. The bivalent Cervarix vaccine was marketed in 2008 to prevent HPV16 and 18 infections. In 2007, the Merck Group developed a tetravalent vaccine called Gardasil, containing L1 viral particles (VLPs) of HPV6, 11, 16, 18. In addition to these 4 HPV types, HPV31, 33, 45, 52 and 58 are included in the nonavalent Gardasil 9 circulating since 2018. The latter is intended to replace its predecessors by providing better coverage against the main high-risk mucosal HPVs (HPV16, 18, 31, 33, 45, 52, 58) and the two main low-risk HPVs responsible for genital warts (HPV6 and 11). Because HPV infection usually occurs early in sexual life, vaccination is recommended during preadolescence (11-14 years) and up to 19 years for girls who have not yet had sex. Vaccination against HPV started about 10 years ago and the prevalence of mucosal high-risk HPV has declined significantly in countries with the highest vaccination rates, including Australia (Machalek *et al.*, 2018). Despite these positive results, the vaccination rate remains low and vaccination campaigns are poorly monitored in the United States (Walling *et al.*, 2016) and in France, where part of the population remains distrustful (Lefèvre *et al.*, 2018).

In addition to these preventive measures, regular screenings are carried out to detect and treat any precancerous lesion before it develops into cancer. The cervical-uterine smear or Papanicolaou test is a cytological sample to check for the presence of dysplastic cells in the cervical epithelium. The result of the smear is classified according to the Bethesda system. The HPV test is a test that complements the smear and provides more reliable results: it can detect the presence of HPV with a sensitivity of more than 95%, and identify HPV types with a specificity of more than 90%. Eventually, HPV testing could replace the Pap smear as a screening approach for

cervical cancer (Hill *et al.*, 2017; Ronco *et al.*, 2010). If these initial examinations indicate the presence of dysplasia and high-risk mucosal HPV, a colposcopy is carried out to observe the cervix, vagina and vulva under the microscope for possible precancerous lesions. In case of abnormality, the treatments used are initially cryotherapy or laser for precancerous lesions. Depending on the severity of the lesion, conization may be considered: this is a surgical operation that consists of removing all or part of the cervix. These treatments to excise CIN are 90-95% effective. In the case of cervical cancer, the treatments applied are chemotherapy, radiotherapy and surgery. No specific treatment for HPV is available to date, although clinical trials are underway for molecules that inhibit the binding of E1 and E2 to DNA (Bosch *et al.*, 2013).

2 The E6 and E7 oncoproteins

E6 and E7 oncoproteins are expressed simultaneously by the virus. Both proteins promote viral replication by triggering cell proliferation. Neither E6 nor E7 has an intrinsic enzymatic activity, however their ability to interact with many host proteins can stimulate the enzymatic activity of the targeted proteins. Since the expression of the two oncoproteins E6 and E7 is maintained in tumor cells, they should be considered as acting in synergy to induce carcinogenesis. The following sections describe the biochemical properties, most described host protein targets and subsequent effects on cell functions for both E6 and E7.

2.1 The E6 oncoprotein

2.1.1 Structure and biochemical properties

The oncoprotein E6 synergistically participates with E7 in carcinogenesis induced by HPV infections: this small protein of about 151 residues is capable of interacting with a large number of host proteins, thus disrupting pre-existing interaction networks in the cell and compromising the regulatory functions of the cell cycle. In fact, the structural characterization of E6 proteins is essential for the study of their molecular mechanisms of interaction and allows the development of inhibitors that specifically block E6 interactions. However, while the E6 protein was identified as an oncogene in the late 1980s (Androphy *et al.*, 1987), it was not until 2006 that the structure of the C-terminal zinc binding domain was resolved by Nuclear Magnetic Resonance (Nominé *et al.*, 2006) and 2013 for the crystallographic structure of the full-length E6 protein (Zanier *et al.*, 2013). This delay is largely due to the difficulty of expressing the recombinant E6 protein and isolating it in its active form. Produced in bacteria, the E6 protein is found in inclusion bodies when it is not fused to a protein that increases its solubility (Nominé *et al.*, 2001). Maltose-Binding Protein (MBP) greatly improves the solubility of the E6 oncoprotein, however the MBP-E6 fusion is partially present as micellar oligomers comprising the misfolded buried E6 polypeptide and the soluble and solvent-exposed MBP units (**Figure 10**).

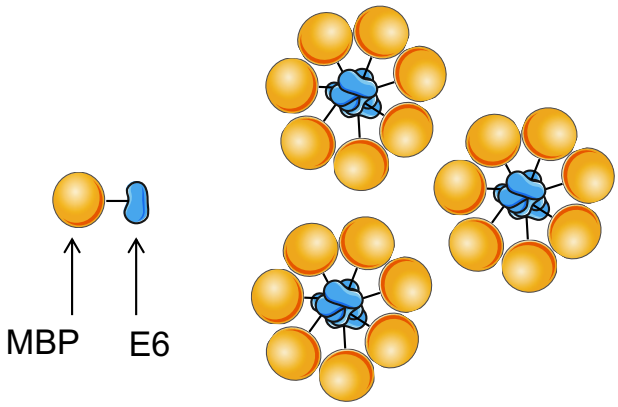


Figure 10 : Micellar oligomers of oncoprotein E6 fused to MBP. The MBP protein is fused to the N-terminus of oncoprotein E6. A portion of the MBP-E6 fusions tends to form oligomers where misfolded E6 polypeptides are grouped together in the center of the micelle while the MBP units remain soluble and are exposed to the solvent.

These difficulties in isolating the E6 oncoprotein in soluble and folded form have led some HPV specialists to consider it as a natively unfolded protein whose intrinsic disorder participates in the oncogenic activity of high-risk HPVs (Uversky *et al.*, 2006).

To understand why the E6 protein has such a strong tendency to self-association, we need to go back to its structural organization and amino acid composition (**Figure 11**).

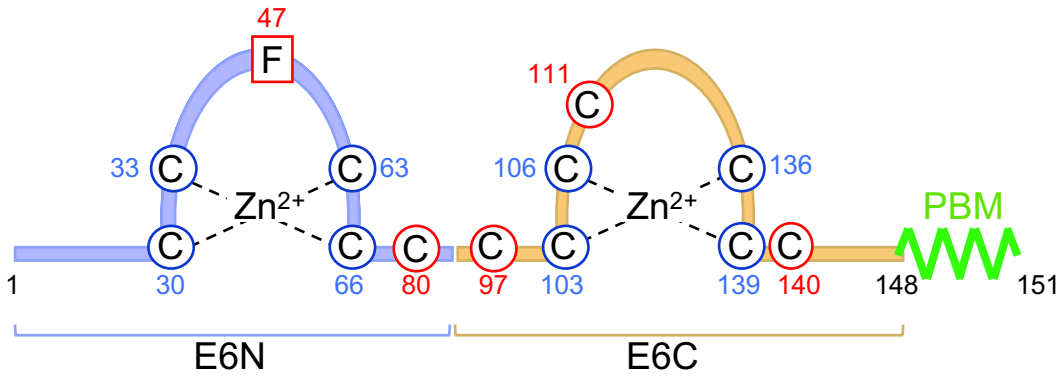


Figure 11 : Organization of HPV16 E6 oncoprotein. The two zinc binding domains are denoted E6N and E6C and are shown in purple and orange respectively. Each zinc ion is coordinated by 4 cysteines, shown in blue circles. Non-conserved cysteines that have been mutated to serine in order to minimize protein aggregation are annotated in red circles. The phenylalanine residue at position 47 has been mutated to arginine to avoid dimerization of the E6 protein. The PDZ domain binding motif (PBM) is present at the C-terminus. Figure adapted from (Poirson, 2016).

High-risk mucosal HPV E6s have a PDZ Domain Binding Motif (PBM). Mammalian HPV E6 proteins consist of two zinc-binding domains, one located in the N-terminal (E6N) and the second in the C-terminal (E6C) (Suarez and Travé, 2018). Each zinc ion requires four cysteines for its coordination: these residues are highly conserved because they are essential for the proper folding of the protein. To these eight

cysteines are added non-conserved cysteines, some of which are exposed on the surface of the protein: under oxidizing conditions, these residues are likely to form intermolecular disulfide bridges which strongly participate in the self-association phenomenon of E6. The E6 oncoprotein of HPV16 contains a total of 14 cysteine residues. Among the six residues not involved in the coordination of zinc ions, four are exposed to the solvent: these have been mutated into serine in order to limit the formation of artefactual disulfide bridges between several molecules. On the other hand, the phenylalanine residue in position 47 promotes dimerization of the E6N domain (Zanier *et al.*, 2012). In order to avoid auto-association during the interaction tests, this residue was mutated into arginine to prevent the dimerization of the protein.

The HPV16 E6 mutant combining the F47R mutations and the four serine mutated cysteines is the one whose structure could be resolved by X-ray crystallography at a resolution of 2.6 Å (Zanier *et al.*, 2013). The structure presents the HPV16 E6 mutant in complex with its prototypical target, the LXXLL motif of the ubiquitin ligase E6AP of sequence $E^{1}L^{2}T^{3}L^{4}Q^{5}E^{6}L^{7}L^{8}G^{9}E^{10}E^{11}R^{12}$. Visualization of this complex at atomic resolution allows us to see that the motif folds into an alpha helix and lodges in a hydrophobic pocket formed by the two zinc binding domains. The structure also reveals that the glutamate residues upstream and downstream of the LXXLL consensus appear to participate in the interaction through hydrogen bonds, suggesting that the consensus recognized by E6 of HPV16 is not limited to the three leucine residues but includes an acidic environment on either side of the motif (**Figure 12**).

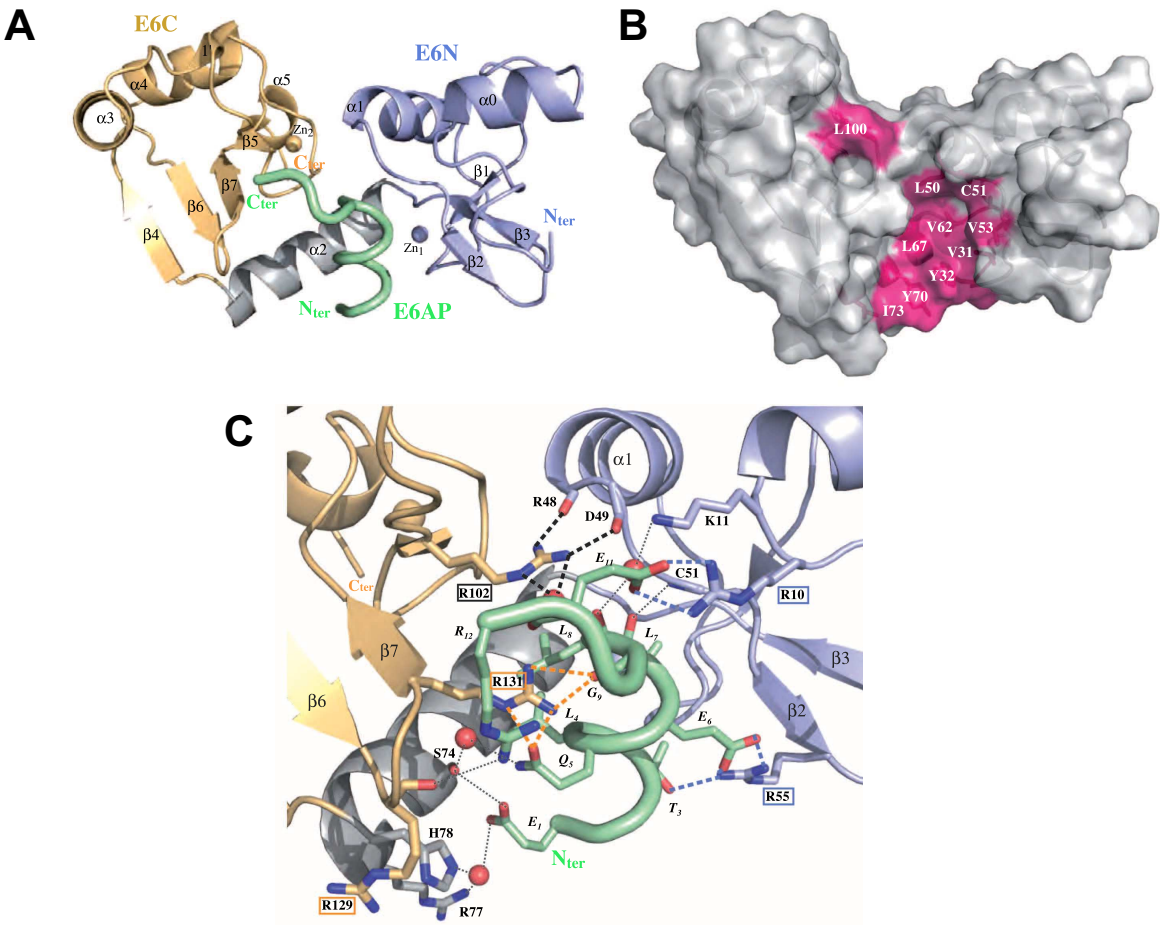


Figure 12 : Structure of E6 of HPV16 complexed with the LXXLL motif of ubiquitin ligase E6AP, of peptide sequence $E^1L^2T^3L^4Q^5E^6L^7L^8G^9E^{10}E^{11}R^{12}$. **A**. Structure of the E6 HPV16 / E6AP complex. The E6N and E6C domains and the two corresponding zinc ions are shown in violet and orange respectively. The linker, folded in the form of an alpha helix, is shown in grey. The LXXLL motif of E6AP (shown in green) adopts an alpha-helix conformation when complexed with the E6 protein. **B**. The binding site of the LXXLL motif is a hydrophobic pocket formed between the two zinc binding domains, the residues of which are shown in pink. **C.** Hydrogen bonds between the LXXLL unit of E6AP and the E6 protein. The interactions are represented by dashes and the water molecules involved in these interactions are represented by red spheres. Adapted from (Zanier *et al.*, 2013)

Once bound to the LXXLL motif of the ubiquitin ligase E6AP, E6 from HPV16 is able to bind the core domain of the tumor suppressor p53. As detailed in the section 2.1.2.1.1, this interaction leads to the proteasomal degradation of p53, which maintains the cell in a proliferative state. The structure of the ternary complex entailing HPV16 E6, the LXXLL motif from E6AP and the core domain of p53 was solved by X-ray crystallography at a resolution of 2.25 Å (Martinez-Zapien *et al.*, 2016).

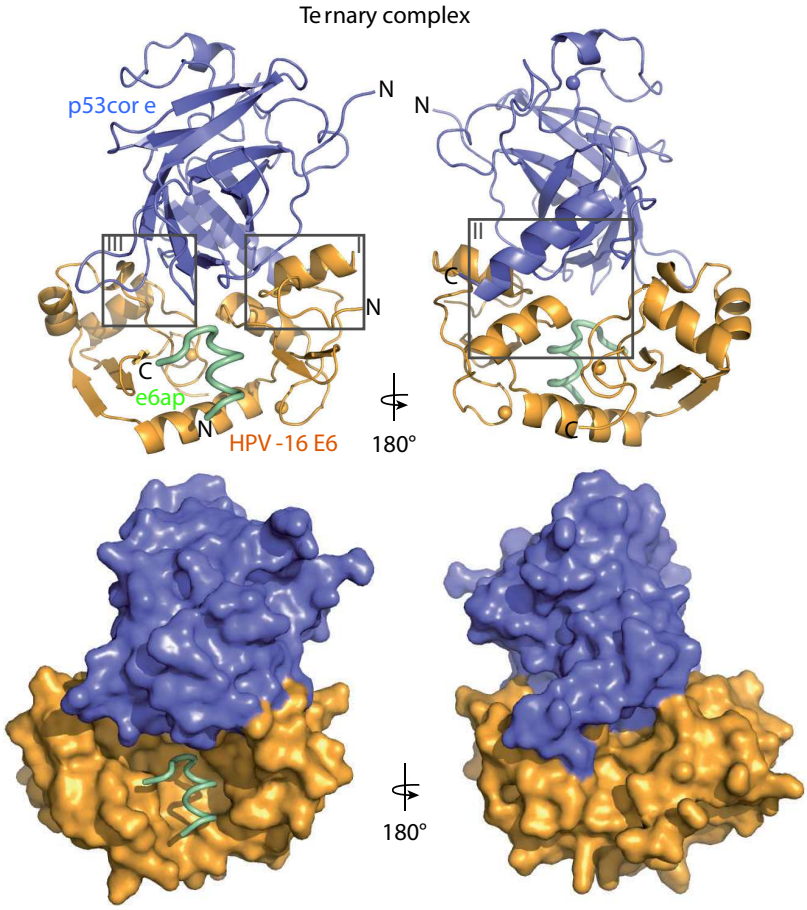


Figure 13 : Structure of 16E6 (in orange) in complex with E6AP LXXLL motif (green α-helix) and p53 core domain (purple). Adapted from (Martinez-Zapien *et al.*, 2016).

This structure allowed the identification of the HPV16 E6 / p53 interface, which was a great achievement for the characterization of E6 interaction properties. The structure displays the minimal interaction domains or motifs, namely the core domain of p53 and the LXXLL motif of E6AP. They are likely to drive the interaction but additional effects observed with full-length proteins can enhance the affinity and stabilize the complex. For instance, p53 is tetrameric and E6AP forms trimers. Their quaternary structure increases the local concentration of the interaction partners, which makes them more likely to form a complex. Indeed, a protein complex associates and dissociates at a certain rate. If one of the partners is dimeric, two interfaces are available for binding at a shorter distance than two independent molecules diffusing in the cell. Thus, the complex is likely to re-associate more rapidly with one of the two binding units. This phenomenon is called avidity: it completes domain-motif interaction and increases their affinity *in vivo*.

Finally, the structures of the C-terminal PDZ-binding motif (PBM) from HPV18 E6 in complex with three PDZ domains (PDZ1 domain from MAGI-1; PDZ2 and PDZ3 domains from Dlg1) were first solved in 2007 by X-ray crystallography (**Figure 14**) (Zhang *et al.*, 2007).

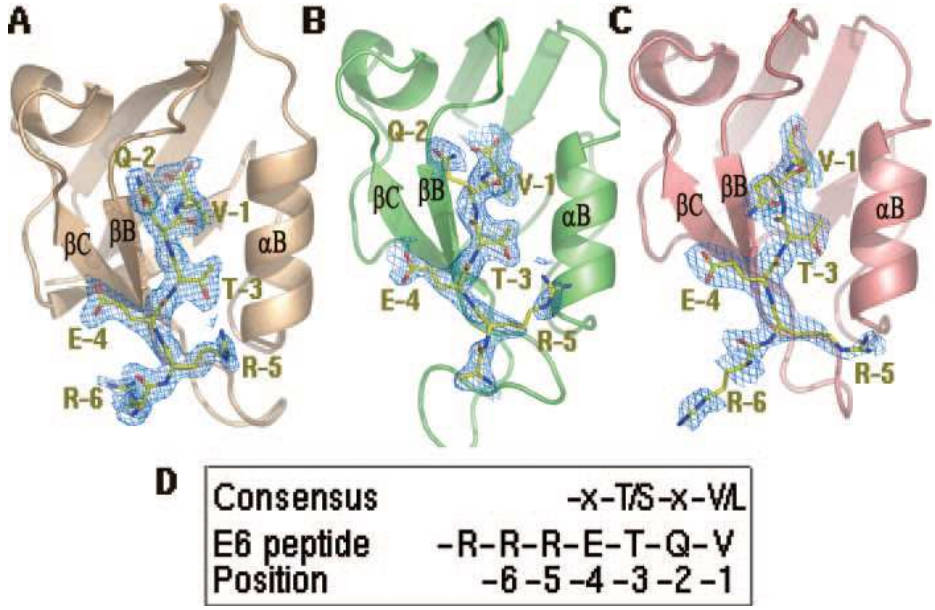


Figure 14 : Structure of PDZ-binding motif (PBM) from HPV18 E6 in complex with three PDZ domains. HPV18 E6 PBM is shown in complex with **A** MAGI-1 PDZ1; **B** Dlg1 PDZ2; **C** Dlg1 PDZ2. The electron density of the PBM is displayed, as well as a stick model represented in yellow. The sequence of C-terminal PBM motif of HPV18 E6 is indicated in **D**. The structure was solved by X-ray crystallography, at a resolution of 2.1 Å for complex A; 2.3 Å for complex B and 2.8 Å for complex C. Adapted from (Zhang *et al.*, 2007).

The structural study reported in (Zhang *et al.*, 2007) highlighted the important contribution of the threonine/serine in position -3 of E6 PBM, which tightly fits into the PDZ groove. The crystal structures revealed that the hydroxyl group from threonine/serine forms a hydrogen bond with a highly conserved PDZ histidine residue. The author demonstrated that mutating this residue to glutamic acid abolishes E6/Dlg binding and reduces Dlg degradation. Interestingly, this interaction between a PBM -3 threonine and a PDZ histidine is common to other PDZ-binding proteins, such as K^+ ion channel and PSD95 (Kim *et al.*, 1995). The protein kinase A-dependant phosphorylation of threonine -3 was reported to prevent PDZ binding (Kühne *et al.*, 2000). The structural data presented by Zhang *et al.* indicate that threonine -3 phosphorylation results in steric hindrance with the PDZ groove, which may abrogate the interaction with PDZ domain. Hence, this study provided a structural explanation

for the negative regulation exerted by protein kinase A on HPV18 E6 PBM (Delury *et al.*, 2013).

A structure of HPV16 E6 PBM in complex with PDZ1 domain from MAGI-1 was solved by Nuclear Magnetic Resonance in 2011 (Charbonnier *et al.*, 2011). In addition, several structures of HPV16 E6 PBM in complex with various PDZ domains were solved, for instance with the PDZ2 domain from MAGI-1 (Gógl *et al.*, 2020a); CTRF-associated ligand (Amacher *et al.*, 2014) and PTPN3 (Genera *et al.*, 2019).

2.1.2 Disruption of cellular functions by protein-protein interactions

2.1.2.1 Capture of LXXLL motifs

2.1.2.1.1 E6AP-mediated degradation of the tumor suppressor p53

In normal cellular conditions, the tumor suppressor p53 is present in low protein levels, regulated by the E3 ubiquitin ligase MDM2, and transcriptionally inactive. Under a number of stress conditions including DNA damage or viral infection, MDM2-mediated degradation is inhibited and the amount of p53 protein increases. A series of post-translational modifications activates p53, which can either stimulate cell survival or apoptosis, depending on the extent of the damage. p53 can upregulate genes involved in DNA repair and induce cell cycle arrest by transcriptionally activating the cyclin-dependent kinase inhibitor p21, which stops the cell cycle in G_1 phase. In case of irreparable damage, p53 triggers apoptosis by activating the expression of pro-apoptotic BCL-2 family proteins, such as BCL-2 associated X protein (BAX), or by direct interaction with pro-apoptotic cytoplasmic and mitochondrial proteins (Kruiswijk *et al.*, 2015). HPV replication occurs in differentiating cells, which usually exit the cell cycle and inhibit DNA synthesis. In order to maximize their replication, HPVs must maintain the DNA machinery in an activated state and inhibit p53 response, which would reduce the number of viral copies by provoking cell cycle arrest or apoptosis (Howie *et al.*, 2009).

The E6-mediated inactivation of p53 is also a mechanism that complements E7-mediated perturbation of the retinoblastoma protein (Rb). HPV E7 binds Rb through its LXCXE motif and was suggested to induce Rb degradation (Giarrè *et al.*, 2001). The consequence of this interaction is the disruption of E2F/Rb complex and the deregulation of entrance into S phase. As a result, p53 protein levels increase in E7-

containing cells, which may cause cell cycle arrest and apoptosis (Münger *et al.*, 2001). The inactivation of p53 excess mediated by E6 oncoprotein thwarts this cellular response to E7-mediated Rb inactivation.

The most studied mechanism for inhibiting p53 response by E6 is its proteasomal degradation, mediated by an E3 ubiquitin ligase. The first study reporting the interaction between the tumor suppressor p53 and the E6 oncoproteins from HPV 16 and 18 was published in 1990 (Werness *et al.*, 1990). The same E6 proteins were then demonstrated to trigger p53 degradation by an ATP-dependent mechanism involving the Ubiquitin-Proteasome System (UPS) (Scheffner *et al.*, 1990). The so-called E6-associated protein (E6AP, also called UBE3A) was then identified as a cellular protein required for the formation of a stable ternary complex with E6 and p53 (Huibregtse *et al.*, 1991). E6AP contains a specific sequence recognized by HPV16 E6 oncoprotein, containing the consensus sequence LXXLL: ^{402}PESSELT**L**QE**LL**GEERR^{418}. In addition, E6AP has a function of E3 ubiquitin ligase (Huibregtse *et al.*, 1993). In the ubiquitylation cascade, the E3 enzymes are responsible for the recognition of the substrate protein (**Figure 15**). The C-terminal catalytic domain of E6AP gave its name to a class of E3 ubiquitin ligases: the HECT-E3 (Homologous to E6AP C-Terminal).

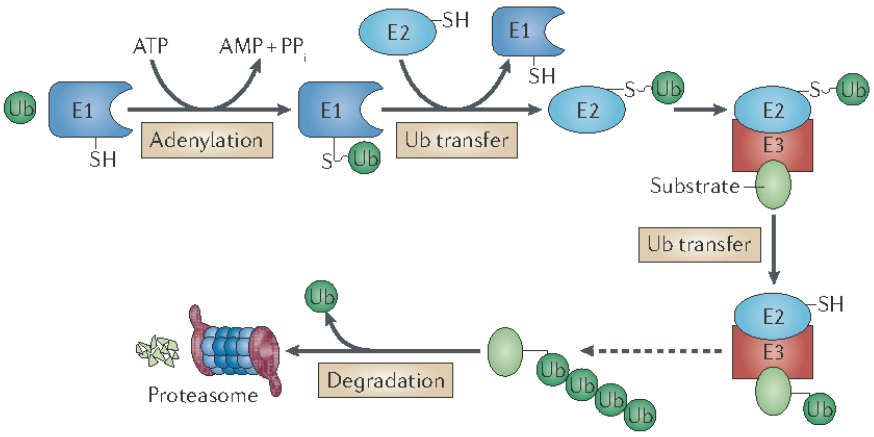


Figure 15: Ubiquitylation mechanism involving the E1, E2 and E3 enzymes. E1 first activates ubiquitin (Ub) by adenylating its carboxy-terminal carboxylate, which results in a thioester intermediate (Ub-E1). Ub is then transferred to the conjugating enzyme E2, and then to the substrate protein (shown in light green) recognized by E3 ubiquitin ligase. Ub is usually bound to the amino-group of a lysine side chain or to the ubiquitin chain formed on the substrate. Ubiquitylation can lead to various outcomes, one of them being the proteasomal degradation of the ubiquitinated protein. Once recognized by the proteasome, the substrate protein is degraded and Ub is recycled. Adapted from (Maupin-Furlow, 2012)

In the absence of HPV E6 oncoprotein, E6AP ubiquitin ligase does not recognize p53 as its prototypical target for ubiquitylation. Its substrate recognition preferences are

altered upon binding to HPV16 E6. The formation of the ternary complex E6/E6AP/p53 leads to the ubiquitination and proteasomal degradation of p53. The ability to induce p53 degradation is shared by most E6 oncoproteins from high risk HPVs (Mesplède *et al.*, 2012).

While high-risk HPV E6 oncoproteins induce p53 degradation by recruiting E6AP, some β-HPV E6 (such as HPV38 and HPV92) were reported to interact with p53 in an E6AP-independent interaction, which stabilizes p53 protein levels. In the same study, some β-HPV E6 (HPV17; 38; 76 and 92) were able to reduce the expression levels of p53-targeted genes (White *et al.*, 2014).

Some E6 oncoproteins can also alter p53 transcriptional transactivation without inducing its degradation, as demonstrated in a study involving E6 mutants (Zimmermann *et al.*, 1999). In addition, E6 binds to and prevents the acetyltransferases p300 and CBP (CREB-binding protein) from stabilizing p53 by means of acetylations. E6 proteins also lead the histone acetyltransferase ADA3 to degradation, with a similar effect on p53 (Moody and Laimins, 2010).

2.1.2.1.2 Repression of keratinocyte differentiation by interaction with the transcriptional co-activator MAML1

The Notch signaling pathways drives cell differentiation in keratinocytes. The gene *mastermind* was first identified in *Drosophila melanogaster* as involved in Notch pathway. The human homologue of this gene is Mastermind-Like 1 (MAML1) and encodes for a transcriptional co-activator for Notch signaling. MAML1 is part of the Notch enhancer complex, which consists of the intracellular domain of Notch (ICN) and transcription factors from the CSL family (Fryer *et al.*, 2002). MAML1 binds to ankyrin repeats of the intracellular domain of Notch (ICN) and induces the redistribution of ICN and recombination signal binding protein Jκ (RBP-Jκ) to nuclear foci (Wu *et al.*, 2000). MAML1 was reported to induce CBP/p300 phosphorylation and accumulation in nuclear foci (Fryer *et al.*, 2002). MAML1 was also reported as a p53 co-activator, which stabilizes p53 and promotes p53-dependant apoptosis (Zhao *et al.*, 2007).

Three independent proteomic studies published in 2012 revealed the interaction between the acidic C-terminus of MAML1 and E6 oncoproteins from Bovine Papillomavirus (BPV) and HPV with cutaneous tropism (Brimer *et al.*, 2012; Rozenblatt-Rosen *et al.*, 2012; Tan *et al.*, 2012). This interaction represses the

transcriptional activation exerted by MAML1 on Notch-responsive genes, as demonstrated by a reporter gene coupled to HES1 promoter. Another study showed that the interaction was mediated by an acidic LXXLL motif located at the C-terminus of MAML1: ^{1006}MSDLDDLLGSQ^{1016} (Brimer *et al.*, 2012).

In a study aiming at deciphering the interaction preferences of various E6 oncoproteins for different LXXLL motifs, the authors raised the point that most α-HPV E6 bind to E6AP ubiquitin ligase LXXLL motif while β, ɣ, ν and μ-HPVs preferentially target the LXXLL motif from MAML1 transcriptional activator (Brimer *et al.*, 2017). The resulting biological effects seem to confirms this trend, as α-HPV E6 proteins induce the proteasomal degradation of a target protein and E6 oncoproteins from other HPV genera repress MAML1 transcriptional activation. The authors argue that this association of E6 proteins with either E6AP or MAML1 LXXLL motif is the sign of an evolutionary split as none of the tested E6 proteins was able to bind both LXXLL motifs. Although their observations about E6AP and MAML1 are valid, one cannot rule out that the repartition of LXXLL binding preferences among all existing E6 proteins might be more complex and may involve additional LXXLL motif-containing protein targets.

2.1.2.1.3 Disruption of the cytoskeleton by interaction with the focal adhesion protein paxillin

Paxillin is a cytoskeletal protein encoded by *PXN* gene and located at sites of cell adhesion to the extracellular matrix (focal adhesion sites). It is involved in signal transduction from the plasma membrane to actin cytoskeleton and focal adhesions. Mutations in *PXN* gene and abnormal expression of paxillin are associated with tumor growth and metastasis in various cancers, such as lung cancer (Jagadeeswaran *et al.*, 2008). Due to this major function in cell signaling, paxillin is equipped with a number a protein- interacting modules. Paxillin is anchored at the plasma membrane by C-terminal LIM domains, which also interact with α-tubulin, kinases and phosphatases such as protein tyrosine phosphatase-PEST. At its N-terminus, paxillin is equipped with SH3 and SH2 binding domains mediating the localization of the tyrosine-protein kinase Src at focal adhesion sites. The N-terminus of paxillin is also targeted for phosphorylations regulating the dynamics of focal adhesion structures and actin network. Paxillin is also known for its five conserved LD motifs: LD1: ^{1}MDD**L**DA**LL**ADLE^{12}; LD2: ^{142}LSE**L**DR**LL**LELN^{153}; LD3: ^{216}SVESLLDELESSV^{228}; LD4: ^{263}TRE**L**DE**LM**ASLS^{274} and LD5: ^{333}QLDSMLGSLQSDL^{345}. These motifs are

bound by a paxillin-binding sequence located on the structural protein vinculin and Focal Adhesion Kinases (FAK) (Turner, 2000). One can note here that LD1, LD2 and LD4 motifs also contain LXXLΦ, which are also recognized by E6 proteins.

A study published in 1997 reported the interaction of paxillin with E6 from Bovine Papillomavirus disrupting the formation of actin fibers in mammalian cells. The authors also noted *in vitro* interaction between paxillin and HPV16 E6, at a lower extent that with BPV E6 (Tong and Howley, 1997). The LD1 and an additional binding site located between amino acids 114-313 were then identified as the main interaction sites recognized by BPV E6. The association between BPV E6/paxillin interaction and transformation was demonstrated and the authors noted that paxillin phosphorylation was altered in presence of BPV E6 (Vande Pol *et al.*, 1998). With the publication of the crystallographic structure of BPV E6 in complex with paxillin LD1 motif in 2013, the molecular basis of this protein-motif recognition was revealed (Zanier *et al.*, 2013). A phage display selection in a library of random 12-mer peptides published in 2017 allowed the identification of a consensus sequence for BPV E6-interacting LXXLL motifs: [Φ-X-X-L-X-X-L-L/F] (Brimer *et al.*, 2017). This consensus includes paxillin and a closely related protein HIC5 (also named Transforming growth factor beta-1-induced transcript 1 protein) and MAML1, all three being targeted by BPV E6 through motif recognition.

2.1.2.1.4 Escaping the innate immune response by targeting the transcriptional factor IRF3

The innate immune response is the first defense against viral infection, it is activated in the infected cell before the immune system. The interferon (IFN) signaling pathway is a major player in the antiviral innate immune response. The transcription factor interferon regulatory factor 3 (IRF3) is a major activator of IFN pathway as antiviral response. The basal state of IRF3 is an autoinhibited monomer. Upon viral infection, cytosolic viral DNA is detected by a sensor, which leads to the oligomerization of the endoplasmic reticulum-localized protein stimulator of IFN genes (STING). IRF3 is then recruited at the C-terminus of STING and phosphorylated by the TANK-binding kinase 1 (TBK1) (Tanaka and Chen, 2012). Together with the phosphorylation of STING, this phosphorylation activates IRF3, which forms a homodimer that enters the nucleus (Liu *et al.*, 2015). IRF3 stimulates the expression of genes entailing an IFN-stimulated response element (ISRE) in their promoters, i.e. type I IFN genes, a type of cytokines

playing a major role in host protection, and IFN-stimulated genes (ISG) (Solis *et al.*, 2006). Targeting a transcriptional factor to delay or inhibit the antiviral innate immune response may promote viral replication, which is an advantage for HPVs.

IRF3 was identified as a cellular target of HPV16 E6 in a yeast two-hybrid screen (Au *et al.*, 1995). Another study confirmed the binding of HPV16 E6 to IRF3 in mammalian cells (Ronco *et al.*, 1998). The authors tested whether other HPV E6 proteins could bind IRF3 and observed that the low risk HPV6 and HPV11 E6 do not bind IRF3 nor does the high risk HPV18 E6. HPV16 E6 recognizes specifically IRF3, as it does not interact with any other IRF protein (IRF1, IRF2, ICSBP, ISGF3 ɣ). The interaction does not induce IRF3 proteasomal degradation, unlike p53. Finally, the use of an ISRE-containing promoter controlling the expression of a reporter gene demonstrated the inhibition of IRF3 transactivation in presence of HPV16 E6. These observations were confirmed in mammalian cells (*Gaussia princeps* luciferase complementation assay) and *in vitro* (Surface-plasmon resonance) by a recent study yet to be published (Poirson, 2016). It is interesting to note that IRF3 is equipped with an LXXLL motif ^{137}EDI**L**DE**LL**GNMV^{148} which is the binding motif recognized by HPV16 E6, as shown on the crystallographic structure deposited on the PDB (6SJA).

2.1.2.2 Interaction with PDZ domains

Some E6 oncoproteins have the ability to interact with PDZ domains. The PDZ domains were named after the three first described PDZ-containing proteins: Post Synaptic Density 95 (PSD95), Discs Large (Dlg) and the Zona Occludens 1 (ZO-1) proteins. They are usually 90 residues-long and specifically recognize a certain group of interaction motifs, the so-called PDZ-binding motifs (PBM). These motifs are generally located at the C-terminus of proteins, although some of them are present at internal sites (Songyang *et al.*, 1997). There are three major classes of PBM defined by distinct consensus sequences: type I PBM [X-S/T-X-Φ_{COOH}], type II PBM [X-Φ-X-Φ_{COOH}] and type III PBM [X-D/E-X-Φ_{COOH}], where X is any residue and Φ any hydrophobic residue (Ganti *et al.*, 2015). The PDZ domains are necessary and sufficient for binding a PBM, which means that they can be added to a heterologous protein without losing their protein-motif interaction ability. In addition, the PDZ-containing proteins often harbor several PDZ domains, attracting different interaction partners at a certain subcellular location. Hence, the PDZ domains are protein

interaction modules conferring a scaffolding function to the PDZ-containing proteins (Sheng and Sala, 2001).

E6 oncoproteins from various α-HPVs entail a C-terminal type I PDZ-binding motif (PBM). The HPV types expressing an E6 protein that harbors a PBM include, but are not restricted to, high-risk mucosal α7 and α9 HPV types. Some low risk cutaneous HPV types also produce an E6 protein equipped with a C-terminal PBM, for instance HPV7, which causes Butchers'Warts (Gógl *et al.*, 2020b). Similarly to p53, the degradation of PDZ-containing proteins is triggered by HPV E6 binding. For instance, the human homolog of the *Drosophila melanogaster* Scribble (hScribble) may be targeted for E6AP-mediated degradation *in vivo* (Nakagawa and Huibregtse, 2000). The ability of HPV31 E6 to bind PDZ domains has been shown to play a role in the viral cycle, in particular for the proliferation and maintenance of the viral genome in undifferentiated cells (Lee and Laimins, 2004).

Since all E6 oncoproteins associated with a high oncogenic risk are harboring a PBM, several studies aimed at investigating the role of the PBM-PDZ interaction in carcinogenesis. In a proteomics study aiming at identifying the PDZ targets of various E6 oncoproteins, the authors postulated that the binding to the tumor suppressor hScribble has a critical role in conferring increased oncogenic potential (Thomas *et al.*, 2016). In a more recent paper, the same authors highlighted the importance of Na^+/H^+ exchange regulatory factor 2 (NHERF-2) as a target for HPV16 E6 and HPV18 E6. The interaction of NHERF-2 with either of these E6 oncoproteins leads to NHERF-2 proteasomal degradation, which indirectly stimulates cell proliferation (Saidu *et al.*, 2019). A recent study showed that the interaction of HPV E6 oncoprotein with PDZ domains induces the nuclear translocation of the oncogenic transcriptional co-factor YAP (Webb Strickland *et al.*, 2018). YAP and its paralog TAZ are the main effectors of Hippo signaling pathway, which plays an important role in the regulation of cell proliferation and as such, is often targeted by oncogenic viruses (Banks *et al.*, 2012).

Both LXXLL and PDZ-mediated interactions appear independently to participate to carcinogenesis. They also act in cooperation to promote the maintenance of the viral genome. The degradation of p53 is mediated by the capture of E6AP LXXLL motif. Despite this mechanism for suppressing p53 activity, some residual undegraded p53 may still be active and impair viral replication. Studies reported that the deletion of the

PBM within E6 sequence resulted in loss of viral genome. The maintenance of the viral genome could be restored by inactivating cellular p53, which suggests a possible role for the E6 PBM-PDZ interactions in neutralizing the activity of residual undegraded p53 (Brimer and Vande Pol, 2014; James and Roberts, 2016; Lorenz *et al.*, 2013).

2.1.3 Development of E6 inhibitors

To date, no specific curative treatment for HPV-induced cancers is available on the market. A prophylactic vaccine is available; however, it is not efficient after infection by an oncogenic HPV and some parts of the world do not have access to effective vaccination campaigns. E6 oncoproteins induce carcinogenesis by disturbing cell functions related to proliferation and cell cycle control. Their expression in cells is required for maintaining the malignant phenotype, as RNA interference with E6 expression results in apoptosis of HPV-positive cancer cells (Butz *et al.*, 2003). Thus, E6 is a therapeutic target of choice for the inhibition of its protein-protein interactions and the targeted apoptosis of HPV-positive cancer cells.

E6 interacts with a high number of host proteins via different binding interfaces, which can be targeted by anti-E6 inhibitors: its LXXLL-binding pocket for the interaction with E6AP ubiquitin ligase, the PBM which is recognized by PDZ domains and the interface with the tumor suppressor p53.

2.1.3.1 Peptides and small molecules targeting the hydrophobic LXXLL-binding pocket

The LXXLL-binding hydrophobic pocket is an interesting target for a therapeutic inhibitor, since blocking it would prevent E6 from recruiting the E3 ubiquitin ligase E6AP and thus from inducing the proteasomal degradation of the tumor suppressor p53. Since the first attempts were conducted before the release of the crystallographic structure of HPV16 E6, the design was based on the structure of E6-binding motifs. In a first study, the authors grafted the E6-binding motif within E6AP sequence to molecular scaffolds including an exposed α-helix (Liu *et al.*, 2004). The selected peptides were validated for inhibiting E6-E6AP interaction *in vitro*. Despite the absence of *in vivo* validation, this study paved the way to mutagenesis studies for the design of inhibitory peptides. The same authors reported in 2006 the design of pharmacophores mimicking the helical fold of E6-binding LXXLL motifs (Baleja *et al.*, 2006). The selected molecules effectively inhibited E6-E6AP interaction and p53 degradation as

assessed *in vitro* and in cell culture. In 2013, the release of the crystallographic structure of HPV16E6 in complex with E6AP LXXLL peptide allowed the identification of E6 interfaces targeted by the inhibitor. In a study published the same year, flavonoids were reported for inhibiting E6/E6AP interaction *in vitro* with SPR data showing the slow dissociation time of the compound from E6 and the stabilization of p53 and p21 was assessed in various cell lines (Cherry *et al.*, 2013).

Screening approaches can lead to the identification of novel binding peptide whose sequence is not related to any known E6-binding LXXLL motif. In 2009, a yeast two-hybrid screening in a library of randomized linear motifs allowed the identification of a HPV16 E6-binding peptide. After optimization by mutagenesis, the resulting peptide (named pep11**) binds specifically HPV16 E6 at high affinity (K_D = 118 nM), restored p53 protein levels and induced apoptosis in HPV16-positive cancer cells (Dymalla *et al.*, 2009). Further biophysical and structural characterization demonstrated that pep11** targets the LXXLL-binding hydrophobic pocket with slower dissociation time than E6AP LXXLL motif (Zanier *et al.*, 2014).

An ELISA-based screening enabled the selection of compounds inhibiting E6/E6AP interaction. 30 identified compounds blocked p53 degradation in a cell-free system and 7 of them efficiently diffused through the plasma membrane and prevented p53 degradation, including 2 flavonoid compounds (Malecka *et al.*, 2014).

2.1.3.2 Inhibitors targeting p53 interface with E6/E6AP

The publication of the structure of the ternary complex E6/E6AP/p53 enabled *in silico* screenings aiming at identifying small-molecule inhibitors of E6/p53 interaction. In a recent study, a compound selected by *in silico* screening could block p53 degradation both in HPV-negative cells transiently overexpressing p53 and/or HPV16 E6 and in HPV-positive cancer cells. Its ability to disrupt the ternary complex E6/E6AP/p53 was demonstrated *in vitro* (Celegato *et al.*, 2020).

To block E6-mediated p53 degradation, one can target E6AP/p53 interaction instead of E6/p53 interface. In 2010, a small-molecule previously reported for inducing apoptosis in different types of cancer could induce cell death in HPV-positive cancer cell lines. The molecule prevented p53 proteasomal degradation by destabilizing p53/E6AP interaction (Zhao *et al.*, 2010).

2.1.3.3 Multivalent inhibitor

Designing an inhibitor targeting E6 protein-protein interactions can be challenging since the inhibitor should have high affinity with E6 in order to efficiently block its interaction with host target proteins. One strategy for maximizing the affinity of a compound is its optimization by mutagenesis for a peptide or structure-based design for a compound. Another way to enhance the affinity of an inhibitor is the design of a bivalent inhibitor targeting to distinct binding sites on E6 oncoprotein. In a study published in 2015, a chimeric fusion protein combining E6AP LXXLL motif and the PDZ2 domain of MAGI1 was designed in order to block both hydrophobic pocket and PBM of HPV16 E6 (Ramirez *et al.*, 2015). The dissociation constants of each moiety were in the micromolar range (K_D = 42 µM for E6AP LXXLL motif and K_D = 3.5 µM for MAGI1 PDZ2 domain) while the chimera bound E6 at nanomolar affinity (K_D = 10 nM). Upon expression in HPV-positive cancer cell lines, the chimera rescued p53 from proteasomal degradation and increased the protein levels of the cyclin-dependent kinase inhibitor p21 and the apoptosis marker caspase 3. The advantage of such chimeric protein is that it should not be targeted by the immune system since it is made of fragments of endogenous proteins. However, inducing its expression in cancer cells or administrating the exogenous protein to enter the tumor cells may represent a challenge for the development of a therapy.

2.2 The E7 oncoprotein

2.2.1 Structure and domain organization

The E7 oncoprotein is a 100 residue-long protein that promotes viral replication by triggering cell proliferation. Its oncogenic properties were discovered in the late 1980s, as HPV E7 could induce transformation of rodent fibroblast cell lines (Bedell *et al.*, 1989; Yasumoto *et al.*, 1986). Together with E6, E7 expression is maintained in HPV-transformed cancer cells, as previously described in section 1.1.3.3.5.

The protein contains three conserved regions, named CR1, CR2 and CR3 (**Figure 16**). CR1 and CR2 display sequence homology to the conserved regions 1 and 2 of adenovirus E1A protein (Phelps *et al.*, 1988). Both are intrinsically disordered while the CR3 region is a structured zinc-binding domain. Since E6 proteins contain a tandem repeat of a similar zinc-binding domain, it was speculated that E6 and E7 evolved from a common ancestor (Cole and Danos, 1987).

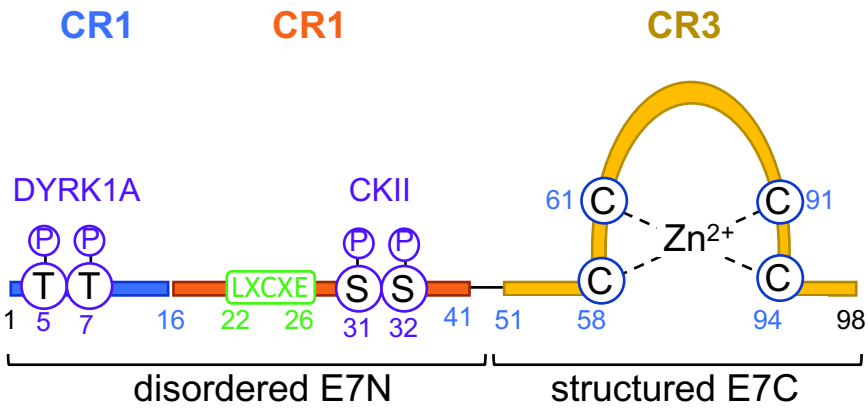


Figure 16: Organization of HPV16 E7 oncoprotein. The E7 protein is made of an intrinsically disordered N-terminus and a structured C-terminus. The three conserved regions CR1; CR2 and CR3 are indicated in blue, orange and yellow, respectively. The CR2 contains a conserved LXCXE motif involved in protein interaction. The CR3 is a zinc-binding domain: the four cysteine residues involved in zinc coordination are indicated in blue circles. The protein also entails phosphorylation sites, indicated in purple. The threonine residues in CR1 are phosphorylated by Dual-specificity Tyrosine phosphorylation-regulated Kinase 1A (DYRK1A) and the serine residues in CR2 are phosphorylated by casein kinase II (CKII). Adapted from (Basukala *et al.*, 2020; Poirson, 2016).

The CR2 region of HPV E7 contains consensus phosphorylation sites for casein kinase II (CKII). The serine residues S31 and S32 are phosphorylated by CKII (Barbosa *et al.*, 1990). Recently, a variant of HPV16 E7 with an additional phospho-acceptor site in position S29 was found to be phosphorylated by CKII on this same serine residue (Zine El Abidine *et al.*, 2017). This additional phosphorylation appears to confer higher potential to transform primary rodent cells. A study performed by NMR suggested that CKII phosphorylation of S31 and S32 could modulate the ability of HPV E7 to interact with its target proteins by inducing minor changes in the residues 26-29, which are contiguous to LXCXE binding motif (Nogueira *et al.*, 2017). Another study confirmed that phosphorylation at CKII sites increases the affinity of E7 for Rb (Chemes *et al.*, 2010). HPV16 E7 was also shown to be phosphorylated by dual-specificity tyrosine phosphorylation-regulated kinase 1A (DYRK1A) on its threonine residues T5 and T7, located in CR1 region (Liang *et al.*, 2008). This DYRK1A phosphorylation appeared to extend HPV16 E7 half-life, which results in improved transforming properties of E7. This can be explained by a poly proline type II (PII) structure of the CR1 and CR2: phosphorylation of the N-terminus may stabilize this PII structure and increase E7 half-life (García-Alai *et al.*, 2007).

E7 contains an LXCXE motif located in the CR2 region, which allows binding to the Retinoblastoma (Rb) tumor suppressor. The structure of the pocket domain of Rb in

complex with the LXCXE motif from HPV16 E7 was first solved by crystallography in 1998, at a resolution of 1.85 Å (Lee *et al.*, 1998). As shown on **Figure 17**, Rb pocket domain consists of A and B boxes, which contain a cyclin fold. The LXCXE peptide binds to a conserved region of the B box, through the intermolecular contacts mediated by the side chains of its residues L22, C24, E26 and L28. Similarly, HPV16 E7 LXCXE motif binds the Rb paralog p107 on the B box of its pocket domain, as demonstrated by a crystal structure published in 2015 (Guiley *et al.*, 2015).

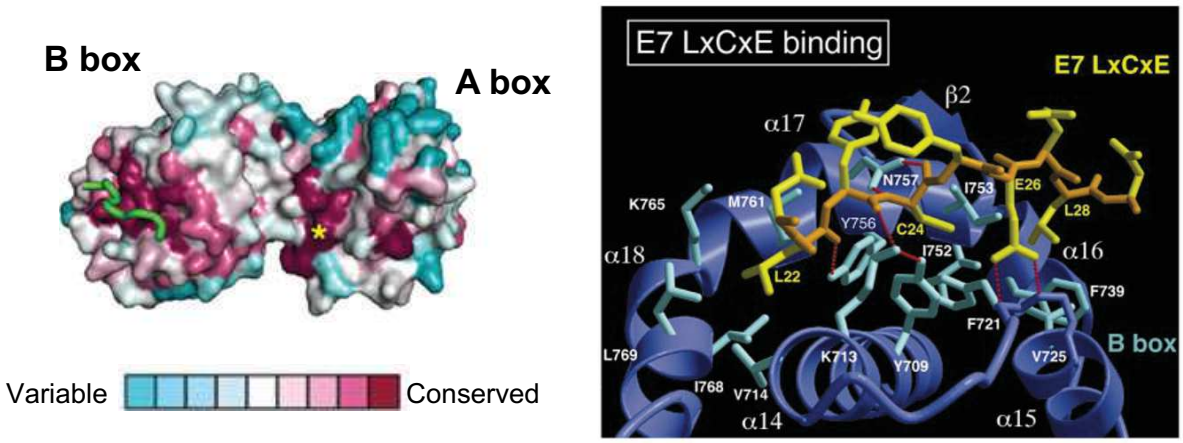


Figure 17: HPV16 E7 LXCXE motif in complex with pRb. The left panel shows the binding of LXCXE peptide (shown in green) on a conserved region of the B box of pRb. From (Chemes *et al.*, 2010). The right panel shows the residues involved in pRb-LXCXE interaction. The complete sequence of HPV16 E7 LXCXE peptide is $D^{21}L^{22}Y^{23}C^{24}Y^{25}E^{26}Q^{27}L^{28}N^{29}$. From (Lee *et al.*, 1998).

Like E6, the E7 oncoprotein contains a high number of cysteine residues. Four of them are required for zinc coordination while noncanonical cysteine residues within the CR3 region of HPV16 E7 may have a role in redox regulation, as deciphered by Chemes *et al.* (Chemes *et al.*, 2014). During oxidative stress, a disulfide bridge is formed between C59 and 68: it induces a conformational rearrangement without altering the zinc binding ability of HPV16 E7. The study also suggests an interaction between N-terminal and C-terminal domains of HPV16 E7 that would protect C24 from glutathionylation and maintain the ability of E7 to bind Rb.

In 2006, two groups published structures of the CR3 region of HPV E7: a structure of HPV45 E7 CR3 in solution was solved by NMR (Ohlenschläger *et al.*, 2006) while a crystal structure of HPV1a CR3 was released the same year (Liu *et al.*, 2006). Both structures revealed that the CR3 domain folds into homodimers and Liu et al highlighted the binding of CR3 to Rb and E2F.

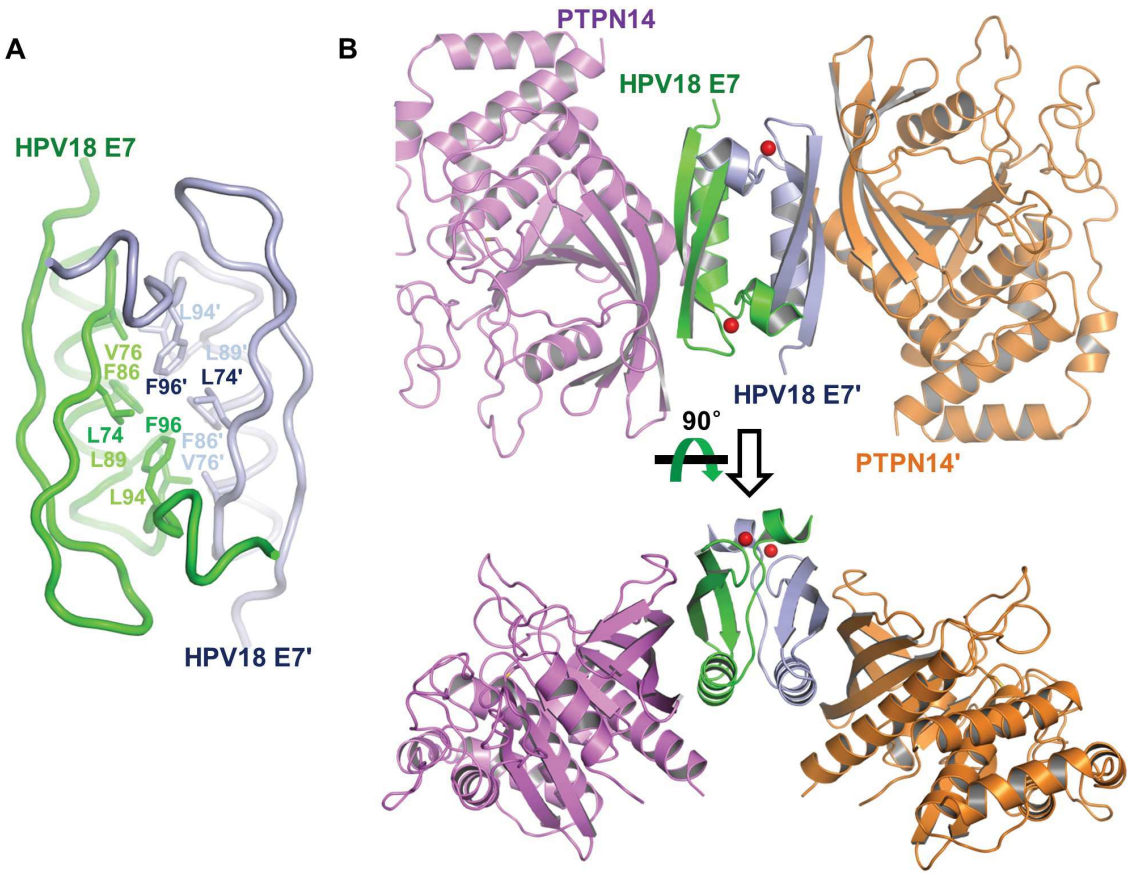


Figure 18: Structure of the CR3 domain from HPV18 E7 in complex with the catalytic domain of PTPN14. Since HPV18 E7 CR3 forms a homodimer, the present figure shows a 2:2 dimeric structure. From (Yun *et al.*, 2019).

Recently, a structural study investigated the interaction of HPV18 E7 CR3 region with the tyrosine-protein phosphatase non-receptor types 14 and 21 (PTPN14 and PTPN21) (Yun *et al.*, 2019). The crystal structure of HPV18 E7 CR3 in complex with the PTP domain of PTPN14 was solved at a resolution of 1.80 Å (**Figure 18**). The study revealed that HPV18 E7 CR3 specifically interacts with the catalytic domains of PTPN14 and PTPN21 but not with other PTP proteins such as PTP1B and PTPN3.

2.2.2 Host proteins targeted by E7 oncoprotein

2.2.2.1 Targeting the pocket domain of retinoblastoma proteins

The retinoblastoma family is a group of proteins characterized by two subdomains A and B forming the pocket domain, which recognizes LXCXE peptide motifs from host or viral proteins. The family entails Rb protein, p107/RBL1 and p130/RBL2 (Paggi *et al.*, 1996). The retinoblastoma proteins (Rb) are nuclear proteins involved in cell-cycle-dependent transcriptional regulation. Rb interacts with members of the E2F family of

transcription factors. E2F proteins form a heterodimer with DP proteins, which enhances their ability to bind DNA and stabilize their interaction with Rb. Once bound to DNA, the E2F / Rb complex recruits additional factors notably by means of LXCXE motif recognition by Rb, which results in transcriptional repression of genes involved in DNA replication and S phase entry. Rb is regulated by cyclin-dependent kinases (CDK). It is able to bind E2F when un(der)phosphorylated in G0 and G1 phase while hyperphosphorylated Rb cannot bind E2F at G1/S phase transition (Kaelin, 1999). The dissociation of Rb from E2F enables E2F to activate the transcription of genes involved in S phase entry (Münger and Howley, 2002).

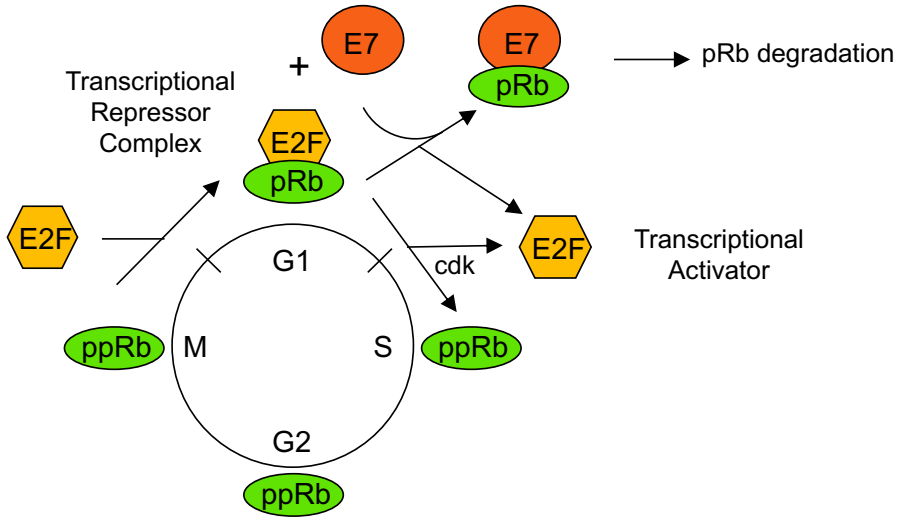


Figure 19: Targeting of Rb by HPV E7 oncoprotein. Rb is un(der)phosphorylated in G1 phase (pRb), which allows the formation of Rb/E2F complex that inhibits the transcription of genes involved in S phase entry. When Rb is hyperphosphorylated by cyclin-dependent kinases (ppRb), it dissociates from E2F, leading to E2F-mediated transcriptional activation of genes ruling S-phase entry. During M phase, Rb is dephosphorylated and able to reform the transcriptional repressor complex Rb/E2F. Upon infection by high-risk HPV, the E7 oncoprotein binds Rb through LXCXE motif recognition, which induces Rb proteasomal degradation. Adapted from (Münger and Howley, 2002).

The pocket domain of retinoblastoma proteins is targeted by several viral proteins equipped with an LXCXE motif: in addition to HPV E7 oncoprotein (Dyson *et al.*, 1989), the large T antigen from SV-40 (DeCaprio *et al.*, 1988) and E1A proteins from adenovirus (Whyte *et al.*, 1989) also target Rb by a similar mechanism involving LXCXE motif recognition. E7 oncoproteins from high-risk HPVs bind the active, un(der)phosphorylated form of Rb and may induce its proteasomal degradation (Boyer *et al.*, 1996). As a result, the abundance of active Rb is decreased, which removes the inhibition by transcriptional repressor complex Rb/E2F. The transcriptional activator E2F stimulates genes involved in S phase entry, which include inter alia DNA

polymerase alpha, dihydrofolate reductase and thymidine kinase. Hence, the E7-mediated proteasomal degradation of Rb leads to deregulated G1/S phase transition (Münger and Howley, 2002). HPV16 E7 protein was reported to associate with the Cullin 2 Ubiquitin Ligase Complex to induce Rb degradation (Huh *et al.*, 2007).

All three members of the Rb protein family, namely Rb, p107 and p130, are targeted by high-risk HPV E7 proteins for proteasomal degradation. This ability to degrade Rb family members was postulated as necessary for E7-mediated malignant transformation (zur Hausen, 2000). In comparison, E7 protein from low-risk HPVs such as HPV6 was reported to target p130 for degradation but not Rb nor p107 (Barrow-Laing *et al.*, 2010).

Some HPV E7 oncoproteins associate with Rb without inducing its degradation, but their effect on disturbing Rb/E2F complex may be sufficient to deregulate S phase entry. Indeed, E7 may destabilize Rb/E2F interaction by disrupting E2F secondary binding sites (Chemes *et al.*, 2010). The increase of MDM2 protein levels in HPV E7-positive cells (Eichten *et al.*, 2002) may also participate to the disruption of Rb/E2F complex since MDM2 binds to E2F and displaces Rb (Martin *et al.*, 1995).

2.2.2.2 High-risk HPV E7 proteins target the tumor suppressor PTPN14 for proteasomal degradation by recruiting UBR4 ubiquitin ligase

PTPN14 is a cytosolic non-transmembrane protein tyrosine phosphatase. PTPN14 is considered as a putative tumor suppressor since *PTPN14* is mutated in several human cancers (colorectal cancer, basal cell carcinoma...) (Bonilla *et al.*, 2016; Wang *et al.*, 2004).

It contains PPXY motifs allowing interaction with YAP, a major effector of the Hippo pathway. The direct interaction between PTPN14 and YAP results in the inhibition of YAP transcriptional co-activator function (Liu *et al.*, 2013). The PPXY motifs within PTPN14 sequence also mediate the interaction with LATS1 proteins, which are upstream negative regulators of YAP. PTPN14 activates LATS1, which negatively regulate the oncogenic function of YAP (Wilson *et al.*, 2014).

E7 oncoprotein from high-risk HPVs recruits the E3 ubiquitin ligase UBR4 to lead PTPN14 to its proteasomal degradation. A tripartite complex similar to HPV16 E6 / E6AP / p53 is presumably formed between HPV16 E7 / PTPN14 / UBR4 (White *et al.*,

2016). Although both high-risk and low-risk HPV E7 bind PTPN14 and UBR4, a more stable complex formation with high-risk HPV E7 might explain their enhanced ability to induce PTPN14 degradation (White *et al.*, 2012b).

To further describe HPV16 E7-mediated degradation of PTPN14, Hatterschide *et al.* identified the variant HPV16 E7 E10K as unable to target PTPN14 due to its inability to interact with UBR4. The authors compared the variations of transcriptome in keratinocytes expressing either HPV16 E7 wild-type or HPV16 E7 E10K. They observed that HPV16 E7 represses genes involved in keratinocyte differentiation. This E7-mediated gene repression is dependent on PTPN14 degradation but appears to be independent of RB1 binding. In addition, full immortalization of primary keratinocytes could be achieved by HPV16 E7 wild-type but not HPV16 E7 E10K. Finally, the transcription profiles of human head and neck samples were analyzed. Repression of differentiation markers consistent with PTPN14 degradation was observed in HPV+ samples but not in HPV- samples, which indicates the involvement of E7-mediated PTPN14 degradation in the development of HPV+ head and neck cancers (Hatterschide *et al.*, 2019).

3 Approaches for the study of protein-protein interactions

3.1 Principles and three examples of biophysical quantitative methods

3.1.1 Basic principles for the quantitative study of protein-protein interactions

Protein-protein interactions can be described by different constants, which reflect the strength of the interaction or the association and dissociation kinetics of a complex. These constants and the strategies for determining them are common to most biophysical methods for affinity quantification.

The dissociation constant, referred to as K_D, is the value most often used to quantify the affinity of an interaction. In the case of the interaction of two molecules A and B forming an AB complex, the reaction is described according to **Equation 1**.

$$A + B \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} AB$$

Equation 1

In this example, K_D is the reaction constant associated with the dissociation of the AB complex at equilibrium. It depends on the concentrations of free compounds A and B in solution and the concentration of AB complex at equilibrium (**Equation 2**).

$$K_D = \frac{[A] \times [B]}{[AB]} = \frac{k_{off}}{k_{on}}$$

Equation 2

In practice, the most common strategy for determining the K_D of an interaction is to incubate both partners at different concentrations: partner A remains at a fixed concentration while partner B is added at increasing concentrations. The A + B mixture is incubated until equilibrium is reached and then the AB complex is quantified using different biophysical approaches. A signal proportional to the concentration of the complex formed is plotted as a function of the concentration of one of the two partners. **Figure 20** shows an example of K_D determination for surface plasmon resonance (SPR) data. The data present the interaction between E6 HPV16 fused to MBP (labelled MBP-16E6) and the LXXLL motif of E6AP (labelled $E6AP_{LXXLL}$). Several

methods allow monitoring a signal proportional to the complex concentration: surface-plasmon resonance, fluorescence anisotropy, fluorescence polarization, nuclear magnetic resonance...

An interaction can be described according to several criteria: affinity, kinetics or thermodynamics.

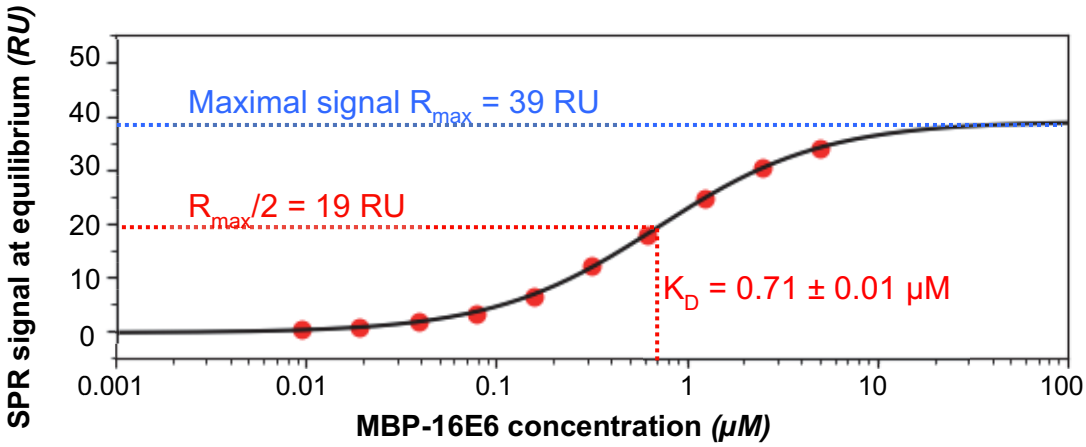


Figure 20 : Example of graphical K_D determination. The above data are from an assay performed by Surface Plasmon Resonance Plasma (SPR) for the interaction between the MBP-fused 16E6 F47R 4C/4S protein (noted MBP-16E6) and the biotinylated peptide of the LXXLL motif of E6AP (sequence PESSELTLQELLGEER). The SPR signal at equilibrium is normalized by the immobilized E6AP peptide signal. The MBP-16E6 protein was incubated at 2-fold serial dilutions whose concentration range between 9.6 nM and 5 µM. Experimental data are represented by red dots and fitting according to a sigmoidal model is indicated by a black line. The K_D of the interaction is equal to the concentration of MBP-16E6 at which the signal reaches half of the maximum signal.

By immobilizing a constant amount of $E6AP_{LXXLL}$ peptide, MBP-16E6 protein was injected successively at increasing concentrations. For each injection, the equilibrium signal, which is representative of the amount of MBP-16E6/$E6AP_{LXXLL}$ complex, is extracted. On a graph representing this equilibrium signal as a function of the concentration of MBP-16E6, non-linear regression following a Langmuir function assuming a 1:1 interaction model is performed, which will indicate an estimation of the K_D. The K_D can also be read on the graph: it is the protein concentration for which the signal is equal to half of the maximum signal visible in the upper plateau (**Figure 20**). In order to perform this data interpretation, it is important to adapt the concentration ranges of the two partners so as to frame the inflection point where the K_D is located and to see a plateau at the highest and lowest concentrations. In addition, the time required to reach equilibrium is specific to each interaction and is one of the parameters to be optimized in order to quantify the complex at equilibrium.

The dissociation constant is also related to the kinetic constants of association (k_{on}) and dissociation (k_{off}). Beyond affinity, the interaction kinetics are very informative to describe the dynamic aspect of an interaction in a living system that is rarely at equilibrium. In pharmacokinetics, these data are essential to determine the time of action of a molecule and therefore the dose and frequency of drug administration. Determining these kinetic constants requires a method for continuously measuring the complex formation from the moment the two partners are brought together until equilibrium is reached and then the dissociation phase. SPR is commonly used to study the interaction kinetics of biomolecules.

Finally, thermodynamic parameters describe the energy of the complex formation reaction. The free energy variation is noted ΔG, it is related to the enthalpy (ΔH), entropy (ΔS) and temperature T variations (**Equation 3**).

$$\Delta G = \Delta H - T\Delta S$$

Equation 3

The variation of free energy reflects the stability of a complex: if this variation is negative, the system is favorable to the formation of the AB complex which is more stable than the unbound A and B partners. To date, the only approach to determine the contribution of the thermodynamic entropy and enthalpy parameters of an interaction in a single measurement is isothermal calorimetric titration (ITC). Other approaches such as surface plasmon resonance (SPR), microscale thermophoresis (MST) and nuclear magnetic resonance (NMR) allow to estimate thermodynamic parameters only by performing measurements at different temperatures.

Finally, the variation of free energy associated with the association of two partners is related to the K_D, according to **Equation 4**, where T stands for the temperature in Kelvin and R is the gas constant.

$$\Delta G = RT \times ln(K_D)$$

Equation 4

Thus, the affinity of an interaction can be expressed as molar concentration or energy release per mole.

3.1.2 Surface plasmon resonance

Surface plasmon resonance (SPR) is an optical technique for detecting interactions between two partners. It can be applied to proteins as well as nucleic acids, lipids, carbohydrates, small molecules, etc. Since the publication of the phenomenon of excitation of surface plasmons by reflection of a light beam in 1968 (Kretschmann and Raether, 1968), SPR has evolved and found many applications, particularly in the study of biomolecules.

One of the interaction partners is called the analyte it is injected at constant flow onto the surface of a chip with a gold layer functionalized by a dextran matrix in which the second interaction partner, called the ligand, is immobilized. Several channels are available on the surface of the chip. One of the channels serves as a reference while the ligand(s) is/are immobilized on the remaining channels. The analyte is injected on all channels of the chip: a signal is emitted when the analyte is bound. The signal of the reference channel results exclusively from the possible binding of the analyte on the surface, while the signal of the active channels is the sum of the binding to the surface and the ligand. Thus, the signal due to the specific recognition of analyte and ligand can be deduced by subtraction.

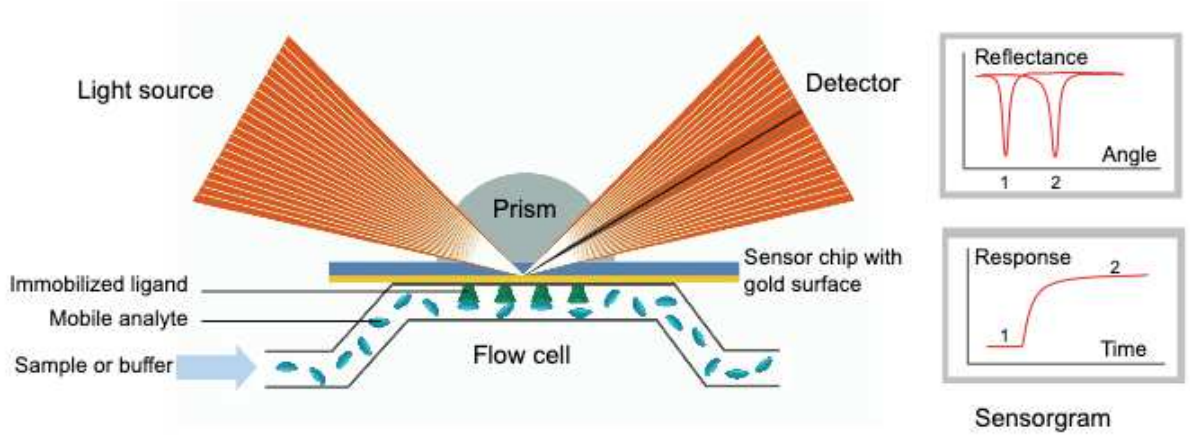


Figure 21 : Principle of surface plasmon resonance. The ligand is immobilized on the gold surface of a detection chip and the analyte is injected by a flow system. A beam of polarized light passes through a prism and reaches the gold surface of a detection chip: the incident light beam has an angular width covering the angle for which the condition of total internal reflection will lead to an extinction by absorption of a very fine line in the reflected beam. A detector makes it possible to track the position, and thus the angle of reflectance, of this shadow cone which differs according to the density at the surface. Thus, this technique makes it possible to distinguish between free ligand (1) and ligand bound to the analyte (2). The sensorgram is the result that shows the association and dissociation of the complex as a function of time. Adapted from (Gueneau and Dufour, 2019).

The principle of SPR is illustrated in **Figure 21**. A beam of polarized light passes through a prism at an angle that allows the beam to be totally internally reflected through the prism. Under these conditions, the electric field of the photons extends to the gold surface and interacts with the free electrons there. The photons are then converted into plasmons, i.e. surface electron density, and the light corresponding to a very fine angular condition is absorbed, resulting in a shadow cone in the reflected beam). When the ligand immobilized on the dextran layer is bound by an analyte, this induces a change in density, and thus the refractive index at the surface, which displaces the shadow cone of the reflected beam. This shift from angle 1 to angle 2 is shown in the upper right part of the figure and can be followed in real time. The intensity of the response as a function of time is shown on the sensorgram, which serves as the basis for the interpretation of the results. The transition between the states 1 and 2 corresponds to a bimolecular association kinetics. Thus, the SPR allows the association of two molecules to be detected at high sensitivity and in real time, allowing the kinetic parameters of the interaction to be determined. On the other hand, this approach is sometimes used to estimate the concentration of biologically active analyte (Drescher *et al.*, 2018).

Ligand immobilization can be done by different approaches: it is a determining parameter to be optimized for the smooth running of the interaction measurement. **Figure 22** shows different ligand immobilization strategies commonly used in SPR and which can also be applied to other methods involving the immobilization of a protein while preserving its folding and activity.

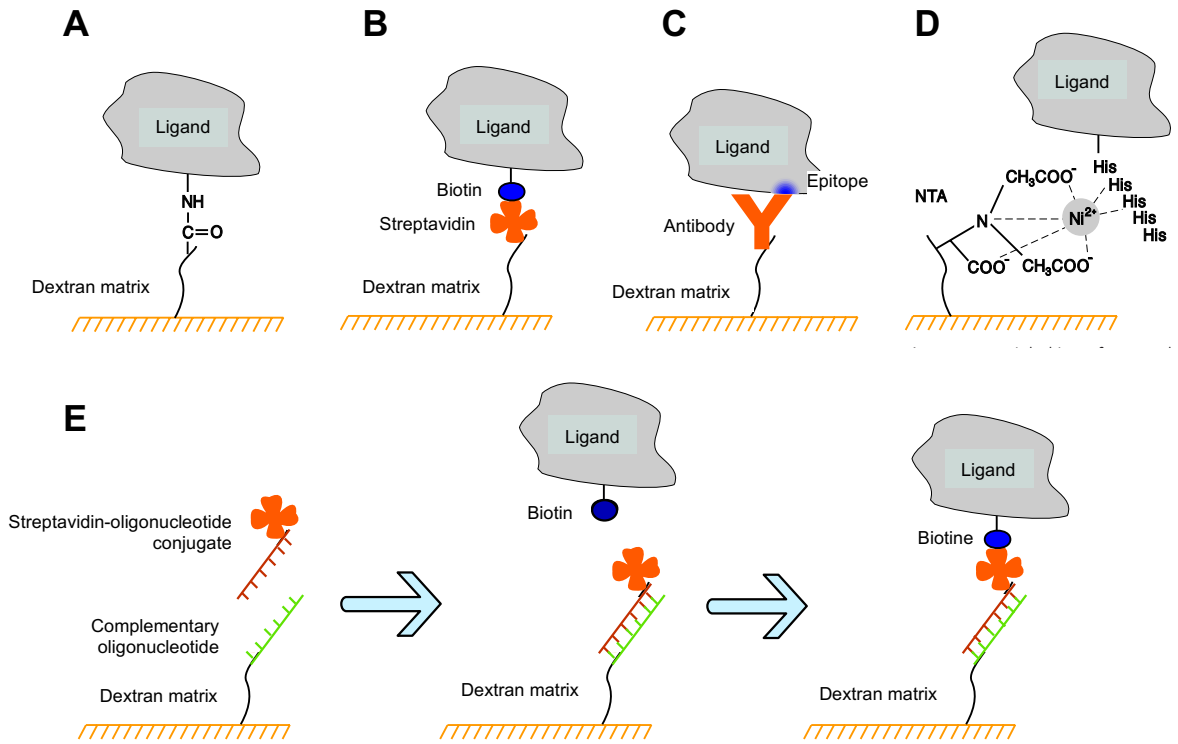


Figure 22: Ligand immobilization on the surface of the SPR sensor chip. All surface types include a dextran matrix that ensures better accessibility of the immobilized ligand. **A**. Covalent immobilization by amine coupling. **B**. High affinity capture of streptavidin biotinylated ligand immobilized on the surface. **C**. Ligand capture by specific antibody. **D**. Capture of poly-histidine labeled ligand by Ni-NTA (Nickel ion immobilized on surface by nitrilotriacetic acid). **E**. CAP chip system for reversible immobilization of biotinylated ligand. An oligonucleotide is immobilized on the surface and hybridizes with a complementary oligonucleotide strand conjugated with streptavidin. The biotinylated ligand binds to streptavidin and can be washed off the surface by dehybridization of the oligonucleotides. Figure adapted from the Biacore user manual, GE Healthcare.

The ligand can be immobilized by covalent coupling or by capture. Covalent coupling is carried out by binding reactive functional groups: in the case of proteins, these are most often amine groups, but there are also coupling reactions involving thiol or aldehyde groups. Covalent immobilization by amines does not require any specific sequence and allows the N-terminal end as well as the residues carrying an amine group to be linked to their side chain (asparagine, glutamine and lysine). However, the ligand can be immobilized in an orientation that renders its site of interaction inaccessible. In addition, the coupling reaction involves a step of pre-concentration of the ligand on the surface by electrostatic attraction. The surface carries negative charges, so the pH conditions are chosen to allow the ligand to carry positive charges: the pH is generally between the logarithmic acidity constant of the surface (noted pKa and equal to 3.5) and the isoelectric point of the protein. This step may be limiting for

proteins subject to aggregation under acidic conditions and excludes proteins with an isoelectric point lower than 3.5 (Gueneau and Dufour, 2019).

Capture is done through an intermediate molecule immobilized on the surface. It allows an oriented immobilization of the ligand for an optimal accessibility of the binding site to the analyte. On the other hand, the affinity of the interaction between the ligand and the capture molecule must be high enough to limit ligand leakage during all cycles of interaction with the analyte. The different modes of capture include the use of a biotin-streptavidin couple, an antibody specifically recognizing an epitope of the ligand, or the immobilization of nickel ions on the surface in order to capture proteins with a poly-histidine label. Finally, a capture system combining the hybridization of two complementary oligonucleotides and the biotin-streptavidin interaction makes it possible to reversibly immobilize a biotinylated ligand. This strategy is advantageous when many ligands are to be tested using a single detection chip. It also avoids the loss of activity of a ligand immobilized on a surface for a long period of time and thus considerably extends the useful life of the same detection chip.

Since SPR is an interaction technique between an immobilized ligand and an analyte in solution, the success of an experiment depends in part on the efficiency of mass transport on the surface of the chip. In the configuration used in SPR, the analyte in solution is brought into contact with a dextran surface by laminar flow. In the absence of any interaction between the analyte and the surface, the analyte concentration is homogeneous. When the analyte is recruited to interact with the ligand, a concentration gradient is created between the surface and the rest of the volume. During injection of the analyte solution, the analyte can be locally depleted in the area near the surface. Conversely, during buffer injection, the analyte can be retained on the surface. Diffusion through this gradient makes it possible to renew the quantity of analyte present on the surface of the chip: this phenomenon is essential for the phases of attachment and detachment of analyte on the ligand molecules. In the case where the mass transport is slower than the association and dissociation kinetics of the analyte-ligand complex, the measurement recorded in SPR would no longer be representative of the interaction, but of the mass transport and therefore of the analyte concentration. The association phase would be slower because the analyte on the surface would not be renewed quickly enough and the dissociation phase would also be slower because the analyte could re-bind on the surface (Schuck and Zhao, 2010). In practice, the

more interaction sites available on the surface, the more efficient mass transport must be to ensure analyte turnover. Immobilizing the ligand in small quantities and at high flow rates ensures a homogeneous distribution of binding sites over the entire surface and limits re-binding of the analyte during the dissociation phase. The choice of a high flux for the whole experiment also provides optimal mass transport conditions (Gueneau and Dufour, 2019). On the contrary, when mass transport conditions are limiting, the signal is related to the analyte concentration, allowing an original measurement of the active concentration of the analyte.

Finally, the bio-layer interferometry method has some similarities with SPR: it is also an optical technique for detecting bimolecular interactions, making it possible to measure the association and dissociation kinetics of two molecules. The interference profile of white light is first measured in buffer and then when the two partners are brought together. The change between these two states produces a signal that quantifies the association and dissociation of the two partners in real time.

3.1.3 Isothermal Calorimetric Titration

Isothermal calorimetric titration is currently the most commonly used approach to determine the thermodynamic parameters of an interaction. It is applicable to proteins as well as nucleic acids, lipids and small molecules (Falconer, 2016). This technique does not require a fluorescent probe or immobilization of either partner and allows the thermodynamic study of the interaction between two molecules in solution. In a single experiment, it allows the determination of affinity (K_D), enthalpy variation (ΔH), entropy variation (ΔS), free energy variation (ΔG) and stoichiometry of an interaction.

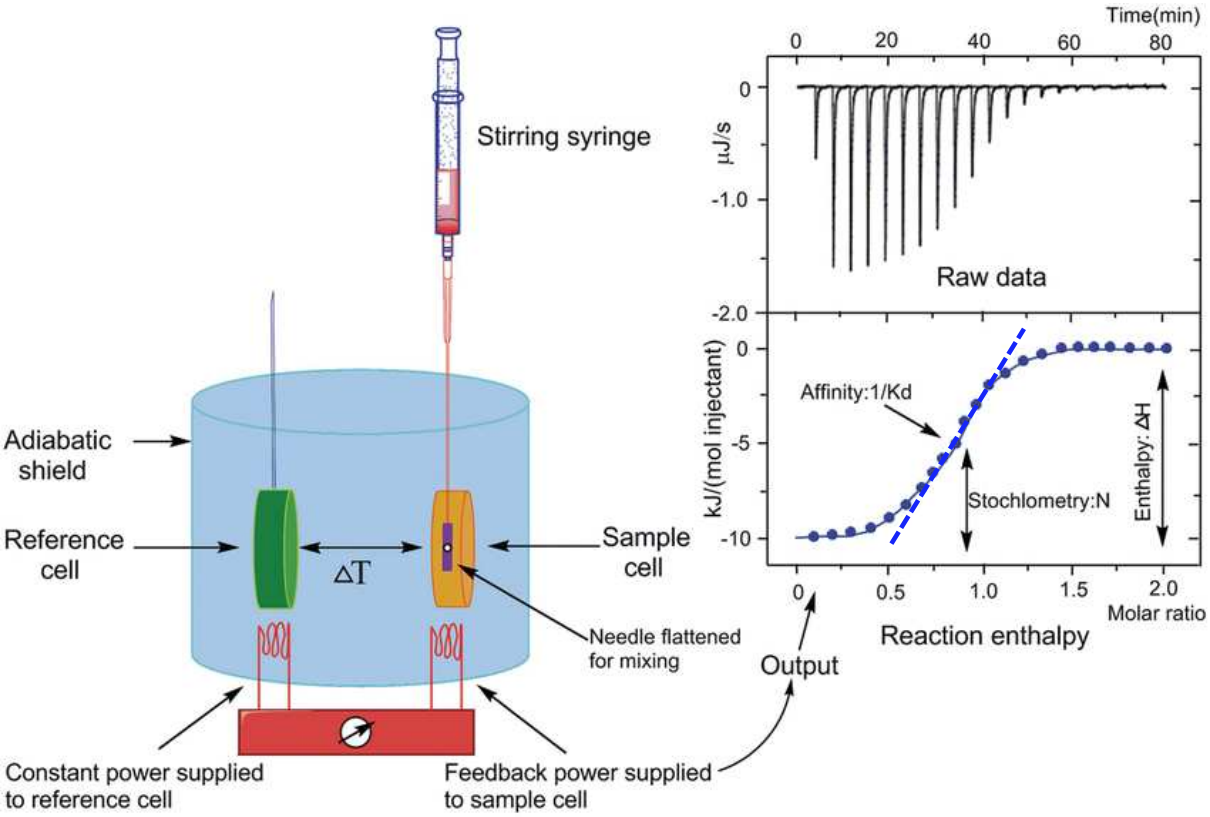


Figure 23 : Principle of isothermal calorimetric titration. The apparatus contains two cells each surrounded by an adiabatic envelope that isolates them from external temperature changes. One of the cells contains working buffer or water (reference cell) and the two interaction partners are brought together in the cell containing the sample by means of a syringe which injects the ligand solution. A constant temperature is maintained in both cells thanks to a device that compensates for heat releases (exothermic reaction) or absorptions (endothermic reaction) due to the formation of the complex in the cell containing the sample. The feedback power required for each ligand injection is recorded. The final result is called the isotherm: this graph is used to determine the enthalpy variation, the dissociation constant and the stoichiometry of the reaction. Adapted from (Song *et al.*, 2015).

The discovery of calorimetry principle predates 1895, based on the publication of patents reporting improvements on calorimeters (Carpenter, 1895). The device consists of two identical cells surrounded by an adiabatic envelope (**Figure 23**). One of the cells contains the buffer alone or water (reference cell) and the second cell contains the sample which will be gradually brought into contact with the ligand injected using a syringe. Constant power is applied to the reference cell. In order to maintain the two cells at an equal temperature, a feedback current applied to the cell containing the sample compensates for the heat variations caused by the complex-forming reaction. This reaction can be exothermic or endothermic, depending on whether it generates or consumes heat. In the case of an exothermic reaction, the heat release from the complexing reaction is captured by the instrument and the feedback current is reduced to keep the temperature constant (Ramesh, 2019). The instrument records changes in the intensity of the feedback current as a function of time, which are

represented by peaks of negative values for each ligand injection in the case of an exothermic reaction (see raw data for **Figure 23**). The amount of ligand gradually increases until the analyte is fully titrated by the ligand; the reaction then reaches a saturation plateau and the cell contains excess ligand. The peaks are then integrated and plotted according to the amount of ligand injected: the resulting graph is called an isotherm. The stoichiometry of the reaction as well as the enthalpy variation (ΔH) can be directly read on the isotherm while the dissociation constant (K_D) can be determined by a binding model.

ITC is a powerful method for the study of interactions because it allows the determination of energy contributions and access to the intrinsic parameters of the interaction provided they have a thermodynamic effect. Thus, this approach can even allow the identification of any conformational or protonation changes in either partner required for the interaction. Such elements greatly improve the understanding of the mechanisms by which two molecules interact. On the other hand, ITC allows high-precision quantification of affinities over a very wide range, reaching extremely high affinities as in the case of the biotin-streptavidin couple (10^{-15} M).

One of the limitations of ITC is that it requires large quantities of each interacting partner, both of which must be soluble and stable at working concentrations (of the order of 10 to 100 µM). In addition, no reactions other than complex formation should take place during the measurement in order not to interfere with the measured heat releases (e.g. oxidation-reduction reaction between oxygen and reducing agents in the buffer) (Baranauskiene *et al.*, 2019).

In 2012, a method enabling the determination of both kinetic and thermodynamic data by ITC was published (Burnouf *et al.*, 2012). So far, this is the first method giving access to affinity, kinetic and thermodynamic parameters of an interaction within a single experiment.

3.1.4 Microscale thermophoresis

Microscale thermophoresis (MST) is an approach for the study of bimolecular interactions that combines fluorescence detection and thermophoresis. It can be used on molecules carrying a fluorescent probe or using only the intrinsic fluorescence of tryptophan.

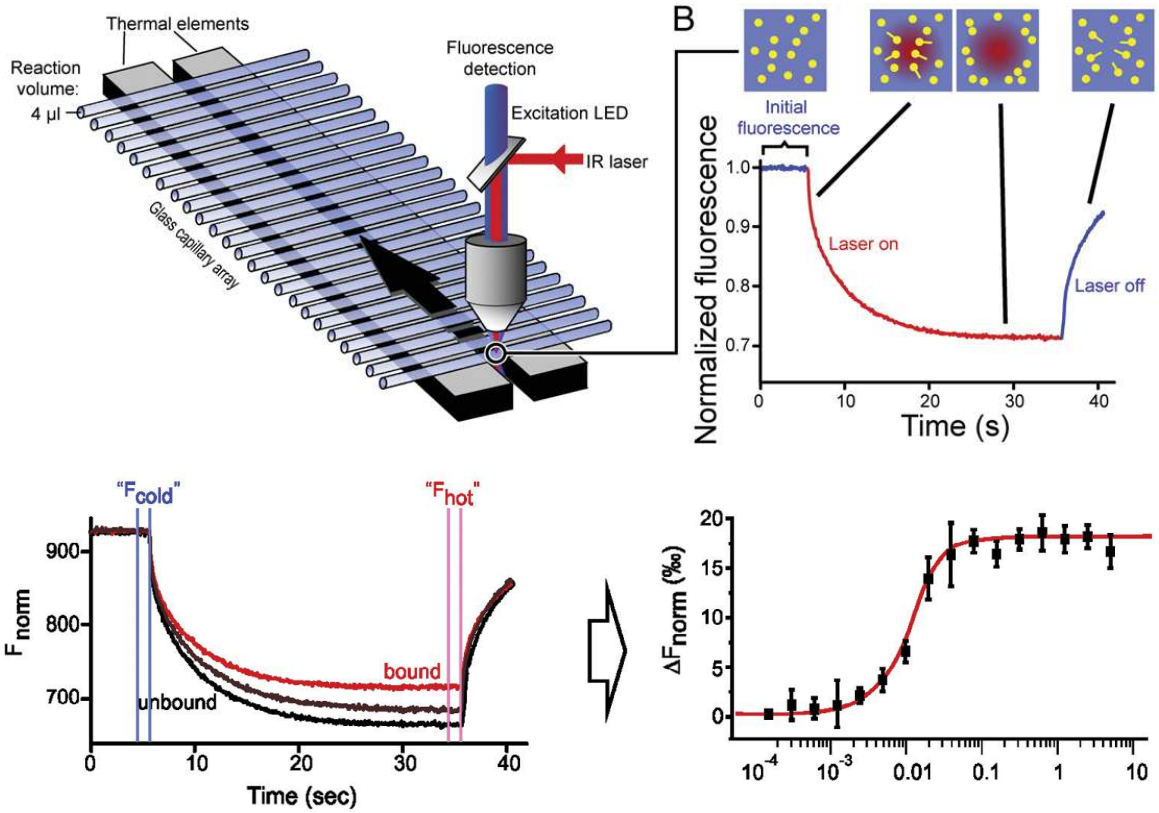


Figure 24 : Principle of microscale thermophoresis. The experiment requires different mixtures of the two interaction partners, with a concentration range of the ligand and a constant concentration of the interaction partner emitting fluorescence. The samples containing the two interaction partners are in glass capillaries. An IR laser induces a temperature gradient in the sample, which activates thermophoretic movement and fluorophore excitation. The initial fluorescence is called "F_{cold}" and the fluorescence after fluorophore excitation is called "F_{hot}". The variation of normalized fluorescence (ΔF_{norm}) for each ligand concentration allows the determination of the dissociation constant. Adapted from (Alexander *et al.*, 2014) and (Jerabek-Willemsen *et al.*, 2014)

The principle of thermophoresis was first published in 1856, with a set-up that did not yet include a laser to induce a temperature gradient.(Ludwig *et al.*, 1856). The sample is in glass capillaries with a capacity of approximately 4 µL (**Figure 24**). An infrared laser induces a microscopic temperature gradient at the sample (of the order of 2 to 6 K). This gradient activates the molecular thermophoretic movement, which causes the excitation of the fluorophores. The fluorescence emitted is captured by the same lens as the infrared laser. Thus, the camera records the depletion or accumulation of fluorophores in the temperature gradient induced by the laser. The dissociation constant is determined by a range of ligand concentration incubated with a constant amount of partner emitting a fluorescent signal. These solutions are placed in different capillaries that can be scanned in a single experiment. The variation in fluorescence measured for each concentration point allows the dissociation constant of the complex

to be determined, following a model similar to that shown in **Figure 24** (Jerabek-Willemsen *et al.*, 2014).

MST can be applied to the hybridization of two DNA molecules, DNA-protein interactions, and protein-protein interactions. Dissociation constants can be estimated for high affinities, of the order of nM and pM). This approach has the advantage of consuming small amounts of sample and delivering a result in a reduced measurement time. The proteins studied can be used in purified form or expressed in fusion with a fluorescent protein in cell lysates. In addition, it can be used to study protein denaturation steps in the presence of increasing concentrations of chaotropic agents (e.g. urea, guanidine). The main limitation of this approach is that although measurement is still possible in the absence of fluorophores, i.e. based solely on the intrinsic fluorescence of tryptophan residues, a majority of proteins contain a large number of aromatic residues that may cause interference with the signal (Sparks and Fratti, 2019). It is therefore often preferable to use fluorescent or fluorophore-coupled fused proteins, with the drawback that their presence does not always preserve the folding of the partners and may generate non-specific interactions due to the presence of the fluorophore.

3.2 High-throughput qualitative methods

While biophysical approaches aim at quantifying the parameters of an interaction (K_D, ΔG), they have most often a rather low throughput and imply the proper isolation of at least one of the interaction partners for an *in vitro* assay. For a binary screen aiming at identifying binders in a large library of putative partners, the use of approaches with a binary output (bind/does not bind) can be adapted.

Phage display is a powerful technique for selecting proteins binding to a bait immobilized in plate or beads. The team of Ylva Ivarsson (Uppsala University, Sweden) developed a proteomic peptide phage display library (ProP-PD), which enables the screening of binding motifs from the whole proteome (Davey *et al.*, 2017). The library and experimental procedures were optimized for low-affinity domain-motif interactions, as different categories of motifs reported in the literature could be identified (for instance, PDZ-binding motif and LXXLL motifs binding to nuclear receptors). During the third year of my PhD, I attempted to identify LXXLL motifs binding to different purified E6 proteins immobilized on Maxisorp™ plates. However, I could not identify

any known nor novel LXXLL binding motif during the two months that I spent in Uppsala. The difficulty to maintain immobilized E6 folded and active during the four days of panning cycles can explain the absence of selected clones. Further optimization of the immobilization method would have been necessary for maximizing the chances of selecting a binding motif.

The screening of interacting proteins within a cellular sample is often achieved by tandem-affinity purification followed by mass spectrometry analysis (TAP-MS)(Rigaut *et al.*, 1999). The approach is very robust for detecting interactions with full-length proteins, as the tandem affinity allows a stringent selection of specific binders. However, stable multi-protein complexes that do not dissociate during the two purification steps may complicate the data interpretation, since all of their components can be detected without particular enrichment of the direct interactant.

Protein complementation assays are commonly used for screening direct binary protein-protein interactions. Yeast-two hybrid (Y2H) is an approach in which the complementation of the transcription factor is measured indirectly by a reporter gene (Venkatesan *et al.*, 2009). Also, the complementation can be directly measured by the emission of a bioluminescent signal, such as split-GFP system (Finnigan *et al.*, 2016) and *Gaussia princeps* luciferase complementation assay (GPCA) (Cassonnet *et al.*, 2011). The advantage of these approaches is that the interaction measured occurs in living cells, which greatly reduces the bias due to the stability of isolated protein. However, the potential bias due to the overexpression of the two interacting proteins may in some cases result in false positives that would not interact at their physiologic concentration.

Very recently, two major studies aiming at mapping the human interactome were published. One was focused on direct binary protein interactions: more than 150 million pairwise interactions involving 17,400 protein coding genes in the human ORFeome were tested by Y2H (Luck *et al.*, 2020). The resulting data, carefully cross-validated by orthogonal methods, provide a reference map indexing all validated binary protein interactions. The second study investigated protein-protein interactions by affinity purification followed by mass spectrometry (Huttlin *et al.*, 2020). The human ORFeome was expressed in 293T cells and more than 10,000 human proteins underwent affinity purification followed by LC-MS (liquid chromatography and mass spectrometry). A

second interaction network was established by performing more than 5,500 immunoprecipitations in HCT116 cells. The comparison of the two interaction networks from two distinct cell lines allowed the identification of cell-specific remodeling of the interactome. These two studies are a major breakthrough in the comprehension of protein-protein interactions. The two approaches are complementary since Y2H allows identification of direct binary interactions while multiprotein complexes can be identified by AP-MS. These datasets pave the way for comprehensive studies in which the protein of interest is involved in an ensemble of interactions as opposed to a more reductive view in which only a few interaction partners are emphasized.

Introduction to the research project

E6 and E7 oncoproteins from HPV promote viral replication by triggering cell proliferation. In particular, E6 protein targets and disrupts the function of a wide range of host proteins by protein-protein interaction. Since E6 is a small protein (about 150 amino acids), it has few interfaces available for protein-protein interaction but they are well adapted for protein-motif interactions. They allow the recognition of a large number of proteins while requiring a reduced interface. E6 proteins evolved to hijack and mimic existing cellular protein-motif interaction networks. Its C-terminal motif mimicking a PBM was comprehensively described in the literature, as it allows E6 protein from high-risk mucosal HPVs to interact with PDZ domain-containing proteins. The present thesis focuses on LXXLL motif hijacking by HPV E6 oncoproteins.

Apart from a few well-documented targets, little is known about the diversity of LXXLL motif-containing proteins targeted by E6. The recruitment of E6AP ubiquitin ligase was widely described as it leads to the proteasomal degradation of several E6 protein targets, such as the tumor suppressor p53. E6AP-independent inhibition mediated by HPV E6 protein was also described for the transcription factors MAML1 and IRF3.

Our approach consists in quantifying the affinity of protein-motif interactions *in vitro*. We chose to quantify affinity as it can be used as a criterion for ranking target proteins. As affinity is an intrinsic parameter of protein interaction, it is a more reliable ranking criterion than the protein amount resulting from a pulldown, which can be biased by indirect interactions, non-specific binding and insoluble proteins. However, we are aware that the fraction of protein complex in the cell not only depends on the affinity but also on the relative abundancy of each interaction partner, which depends on the cell type and are modulated by cellular stress. We opted for a fragmental approach by studying the interactions between E6 proteins and LXXLL peptides whose sequence originate from cellular protein targets. This strategy allowed us to identify high-affinity binding peptides forming stable E6/LXXLL complexes suitable for structure solving by X-ray crystallography. Such structural data greatly facilitate the design of therapeutic inhibitors targeting specifically the LXXLL-binding pocket. As to whether these data are representative of an interaction between two full-length proteins, we can expect that secondary binding regions participate to the interaction and modulate the affinity. We focus on LXXLL motif binding because they are the main interface. We do not claim to provide an exhaustive study of viral-host interactions but we propose a reliable starting

point at the molecular level, by characterizing the binding affinities of the minimal interaction units.

To date, several proteomics studies were carried out for identifying the host proteins targeted by HPV E6 oncoproteins (Grace and Münger, 2017; Rozenblatt-Rosen *et al.*, 2012; White *et al.*, 2012a). Together with studies conducted on murine models (Viarisio *et al.*, 2016, 2018), these studies draw a picture of various E6 target proteins, as illustrated on **Figure 25**.

The trend that emerges from these data is the binding of high-risk mucosal α-HPV E6 to E6AP ubiquitin ligase. HERC2 (HECT domain and RCC1-like domain-containing protein 2) is a well-known interaction partner of E6AP (Kühnle *et al.*, 2011), its co-precipitation with E6AP-interacting α9-HPV E6 proteins is likely due to an indirect interaction via E6AP. Since E6AP is required for the proteasomal degradation of p53, the tumor suppressor was co-purified with most E6AP-interacting high-risk mucosal HPV E6. It is interesting to note that p53 was also co-purified with the β-HPV 92 and 38 E6 proteins, even though these two oncoproteins did not interact with E6AP. This observation confirms that some β-HPV E6 bind p53 and stabilize its protein level by an E6AP-independent mechanism (White *et al.*, 2014). The transcription factor MAML1 is bound by cutaneous β-HPV E6, which is consistent with the recent study reporting that cutaneous β-HPV E6 inhibit MAML1 transcriptional activity (Brimer *et al.*, 2017). The adhesion protein paxillin is bound by β2 and β4-HPV E6 proteins. We can note that the histone acetyltransferases p300 and CBP were co-purified with β1 and β2-HPV E6 proteins. According to a study published in 2011, E6 proteins from β1-HPV bind to p300 and induce its proteasomal degradation in an E6AP-independent manner (Howie *et al.*, 2011). Interestingly, E6 proteins from β2-HPV types 17a and 38 bind to the Ccr4-Not complex, a major regulator of gene expression and RNA metabolism in eukaryotes (Collart, 2016). Ccr4-Not complex has a deadenylase function allowing the removal of polyA tails from messenger RNAs targeted by micro RNAs (Braun *et al.*, 2011) and also a ubiquitin ligase function associated with the CNOT4 subunit (Albert *et al.*, 2002). In another proteomics study, the ɣ-genus HPV197 E6 was also reported to interact with Ccr4-Not complex (Grace and Münger, 2017). The subunit CNOT1 entails several LXXLL motifs mediating the interaction with the Estrogen Receptor α (ERα), which leads to the repression of ERα ligand-dependent transcriptional activation function

(Winkler *et al.*, 2006). To date, whether the Ccr4-Not/E6 interaction is mediated by these LXXLL motifs or by another interface has not been determined.

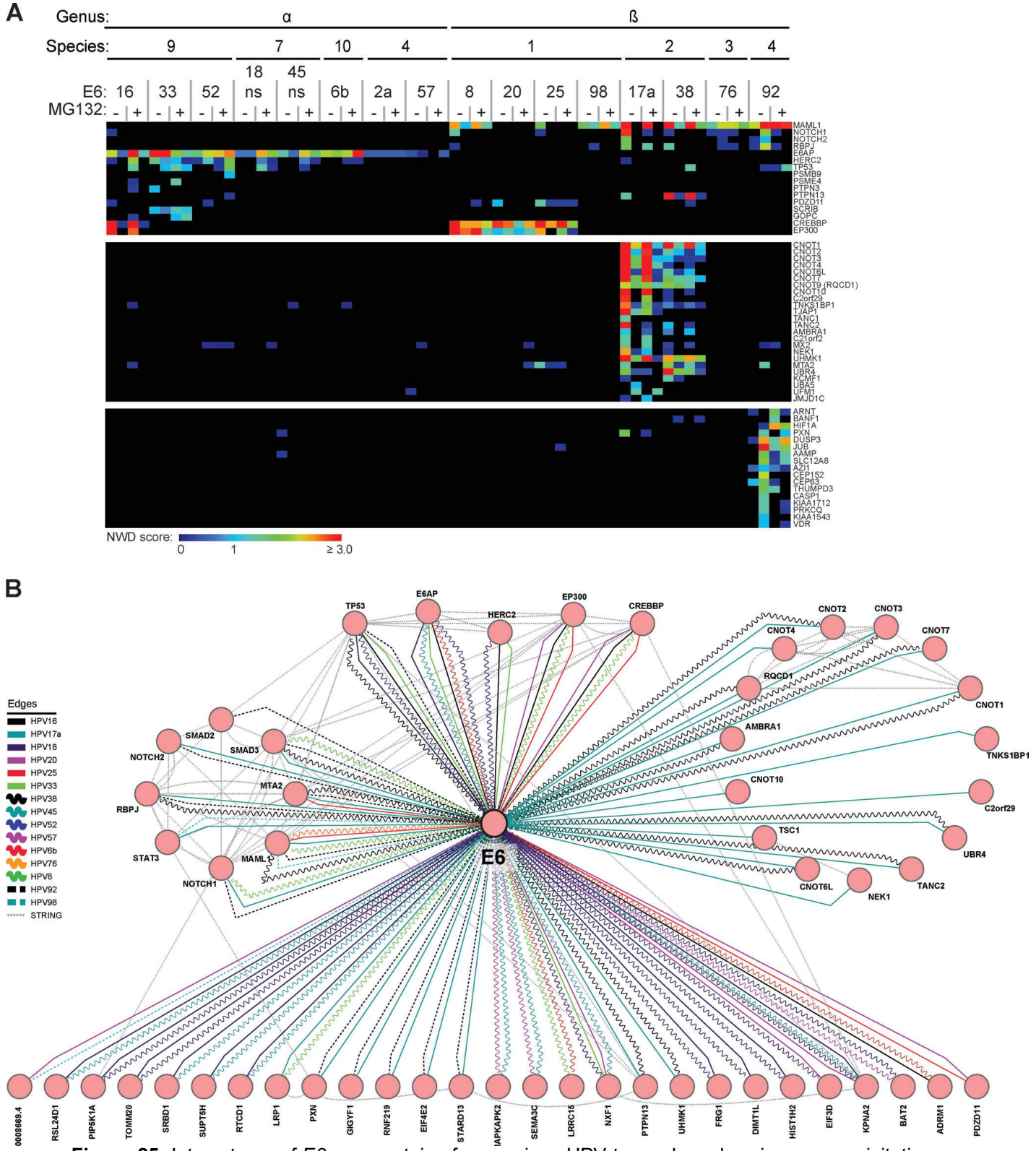


Figure 25: Interactome of E6 oncoproteins from various HPV types, based on immunoprecipitation-MS/MS data. NWD stands for Normalized weighted D-score, which defines a high confidence interaction when NWD-score ≥ 1. Cells were treated with the proteasome inhibitor MG132 (+) or with negative control DMSO (-) prior to harvest. From (White *et al.*, 2012a).

The present study aims at investigating the interaction preferences of E6 oncoprotein from various HPV types in terms of LXXLL motif recognition. We believe that E6 proteins target a multiplicity of host proteins by their LXXLL motifs. Hence, the repartition of LXXLL recognition features across the phylogenetic diversity of HPV E6 proteins might be more complex than the binary split between E6AP and MAML1 postulated by Brimer *et al.* (Brimer *et al.*, 2017).

To tackle the challenge of isolating biologically active E6 oncoproteins and quantifying low-affinity protein-motif interactions, we developed dedicated approaches. These methods were advantageously used for comparing the LXXLL recognition preferences of E6 oncoproteins from various HPV types.

The first part of this result section focuses on the methodological developments, which entail i) a purification protocol enabling parallel purification of soluble E6 proteins and ii) optimization of the previously published holdup comparative chromatographic retention assay. The second part presents the interaction profiles of multiple E6 proteins in terms of LXXLL motifs. This chapter includes a first study conducted on seven E6 proteins from various HPV representative of the phylogenetic diversity and a second project focusing on ancestral E6 oncoproteins from α-genus HPVs. Finally, the topic of the last part is a side project aiming at identifying the residues involved in the interaction between the E6 prototypical target E6AP and its main interaction partner HERC2.

Results

1 Methodological developments

1.1 One-step affinity purification of fusion proteins with optimal monodispersity and biological activity: application to aggregation-prone HPV E6 proteins

1.1.1 State of the art

My thesis project is focused on the *in vitro* characterization of the protein-protein interaction properties of various E6 oncoproteins from different HPV types. To date, only few studies involving high-precision biophysical data of E6 oncoproteins have been reported. Most of the published biophysical data focus on HPV16 E6 and BPV E6 whose crystallographic structures were solved in 2013, which indicates that substantial amounts of protein in a homogeneous and soluble state could be isolated for these two E6 oncoproteins. Isolating biologically active E6 proteins can be challenging, since they rapidly lose their activity by forming inactive oligomers under the effect of oxidation (Zanier *et al.*, 2010). This propensity to oligomerize in oxidative conditions is shared by all E6 oncoproteins at varying degrees, due to their high content in cysteine residue forming artifactual disulfide bridges. At least 8 conserved cysteine residues are required for the coordination of two zinc ions and non-conserved cysteine residues may increase the risk of intermolecular disulfide bridges, in particular when they are solvent-exposed. To tackle the challenge of isolating active E6 proteins, several strategies can be implemented. First, the solvent-exposed cysteine residues can be mutated into amino acids that do not alter the intrinsic fold of the protein. In addition, the recombinant E6 proteins can be fused to the solubility-enhancing Maltose-Binding Protein (MBP). Second, the purification can be performed under reductive conditions, i.e. in presence of a reducing agent (DTT, β-mercaptoethanol, TCEP). The purification buffers can be degassed under a vacuum pump and equilibrated with an inert gas (Argon or Nitrogen) in order to remove dissolved oxygen from water (Butler *et al.*, 1994). Finally, as detailed in the publication reporting the crystallographic structure of HPV16 E6, an overnight ultracentrifugation step included in the purification process can participate to the elimination of a fraction of oligomer species (Zanier *et al.*, 2013). Although well-adapted for crystallographic purpose, this ultracentrifugation step can be limiting as is implies a 2-day long purification of only one protein at a time. When the purpose is the comparison of the interaction preferences of various HPV E6

proteins by biophysical approaches, less protein is needed than in crystallography but the same high quality is required.

1.1.2 Objective

This study aimed at developing a fast, parallelizable purification protocol for isolating in one single experiment several biologically active MBP-E6 protein samples. By testing different types of nickel affinity resins and optimizing the incubation time, we were able to isolate monomeric, active E6 samples as assessed by SPR interaction assay with their prototypical LXXLL target motif.

The designed purification strategy will be of great use for i) screening several E6 from various HPV types and assess whether they can be isolated as biologically active samples and ii) test the interaction preferences of several E6 proteins for an array of LXXLL peptide motifs in one high-throughput assay (by SPR or holdup assay).

1.1.3 My contribution

For this project, I designed and executed the recombinant protein production and purification experiments. I also performed SPR assays for assessing the binding of the purified proteins. I processed the data and produced the figures presented in the article. I wrote the initial manuscript and made the revisions with my supervisors Gilles Travé and Yves Nominé.

Bonhoure *et al. Microb Cell Fact* (2018) 17:191
https://doi.org/10.1186/s12934-018-1039-z

Microbial Cell Factories

RESEARCH

Open Access

One-step affinity purification of fusion proteins with optimal monodispersity and biological activity: application to aggregation-prone HPV E6 proteins

Anna Bonhoure[1], Auguste Demenge[1], Camille Kostmann[1], Leticia San José[2], Eva De la Cal[2], Pilar Armisen[2], Yves Nominé[1] and Gilles Travé[1*]

Abstract

Background: Bacterial expression and purification of recombinant proteins under homogeneous active form is often challenging. Fusion to highly soluble carrier proteins such as Maltose Binding Protein (MBP) often improves their folding and solubility, but self-association may still occur. For instance, HPV E6 oncoproteins, when produced as MBP-E6 fusions, are expressed as mixtures of biologically inactive oligomers and active monomers. While a protocol was previously developed to isolate MBP-E6 monomers for structural studies, it allows the purification of only one MBP-E6 construct at the time. Here, we explored a parallelizable strategy more adapted for biophysical assays aiming at comparing different E6 proteins.

Results: In this study, we took advantage of the distinct size and diffusion properties of MBP-E6 monomers and oligomers to separate these two species using a rapid batch preparation protocol on affinity resins. We optimized resin reticulation, contact time and elution method in order to maximize the proportion of monomeric MBP-E6 in the final sample. Analytical size-exclusion chromatography was used to quantify the different protein species after purification. Thus, we developed a rapid, single-step protocol for the parallel purification of highly monomeric MBP-E6 samples. MBP-fused HPV16 E6 samples obtained by this approach were validated by testing the binding to their prototypical peptide targets (the LXXLL motif from ubiquitine ligase E6AP) by BIAcore-SPR assay.

Conclusions: We have designed a rapid single-step batch affinity purification approach to isolate biologically active monomers of MBP-fused E6 proteins. This protocol should be generalizable to isolate the monomer (or the minimal biologically active oligomer) of other proteins prone to self-association.

Keywords: Protein aggregation, Protein solubility, MBP fusion, IMAC, One-step purification

*Correspondence: gilles.trave@igbmc.fr
[1] Équipe Labellisée Ligue 2015, Department of Integrated Structural Biology, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U1258/CNRS UMR 7104/Université de Strasbourg, 1 rue Laurent Fries, BP 10142, 67404 Illkirch, France
Full list of author information is available at the end of the article

Background

Protein solubility is a major issue in recombinant protein purification from *Escherichia coli*. It is influenced by different parameters, such as proper folding and/or aggregation. Under the stress induced by high rates of heterologous protein expression, inactive misfolded polypeptides accumulate as inclusion bodies. They can be solubilized by using chaotropic agents (urea or guanidium hydrochloride) and then refolded in vitro (by dilution or dialysis). However, such an approach can be time-consuming since it requires optimization adapted to every protein, and the yield of final soluble product can be low. Another strategy consists of fusing the protein of interest (also called "passenger protein") to a solubility-enhancing protein (also called "carrier protein"), such as Thioredoxin, glutathione-S-transferase (GST) or maltose-binding protein (MBP). MBP was reported to be particularly efficient in improving the solubility of its fusion partners [1, 2]. In addition to its ability to stabilize and solubilize the passenger protein, MBP can be used as an affinity tag for purification on amylose resin [3, 4]. However, an MBP-fused protein can sometimes be solubilized in the form of a mixture of properly folded monomers (or minimal biologically active oligomers) and of large oligomers in which the passenger protein is self-associated and its folding and/or biochemical activity may be altered [5–13]. In such situations, the challenge consists in exploring conditions of expression and purification favoring the biologically active monomeric -or minimally oligomeric- MBP-fused samples [9, 11, 14, 15].

E6 proteins produced by oncogenic human papillomaviruses (HPV) are a prototypical case of proteins that display unstable biochemical behavior when produced recombinantly [16–18]. While unfused E6 proteins are expressed in insoluble form, MBP-fused E6 proteins are soluble yet prone to self-association, leading to soluble oligomers, which in turn are unable to specifically interact with protein partners [9, 10, 19]. Furthermore, E6 proteins are generally cysteine-rich. E6 proteins contain two zinc-binding domains, each involving four cysteine residues to coordinate one zinc ion. In addition to these eight cysteine residues highly conserved for structural purposes, E6 proteins contain additional non-conserved cysteine amino acids. For instance, wild-type HPV16 E6 and HPV8 E6 have a total of 14 and 16 cysteine residues, respectively. Oxidation promotes the formation of intermolecular disulfide bridges, and thus increases the aggregative propensity of E6 proteins.

Papillomaviruses E6 oncoproteins establish numerous interactions with host proteins [16–18, 20]. For instance, E6 from "high-risk" mucosal HPV was reported to hijack the ubiquitin ligase E6AP (E6-associated protein), resulting in the proteasome-mediated degradation of the tumor suppressor p53 [21]. Crystal structures of the HPV16 E6 oncoprotein from high-risk mucosal HPV16 in complex with a minimal target fragment from E6AP and the core domain of p53 were solved by X-ray crystallography, using solubility-enhanced mutants of E6 fused to MBP [22, 23].

There are more than 200 HPV types, with different tropisms (mucosal, cutaneous) causing a large variety of phenotypes ranging from warts and condylomas for low-risk HPVs to malignant tumors for high-risk HPVs [24–26]. The present work will focus on E6 oncoproteins from HPV16 and HPV8. HPV16 is the highest-risk mucosal HPV, responsible for more than 50% of cervical cancers and more than 90% of HPV-positive oropharyngeal cancers [27, 28]; whereas HPV8 is a cutaneous HPV type that can generate skin cancers, in particular in individuals with genetic or immunological diseases [29].

Since E6 proteins from different HPV types were reported to have distinct protein interaction preferences by proteomic studies [20], it would be interesting to further decipher their binding properties by quantitative biophysical and structural approaches. The protocol used in our former publications [9–12] allows the isolation of soluble, monomeric E6 proteins by amylose affinity chromatography followed by overnight ultracentrifugation and size-exclusion chromatography (SEC) for the elimination of protein oligomers. Such a strategy is appropriate for structural studies that require relatively large amounts of high quality protein material. However, it allows the purification of only one protein at a time. To decipher the protein–protein interactions of a panel of E6 oncoproteins, parallel purifications would be more appropriate. In addition, a faster purification protocol would reduce the risk of protein aggregation over time.

In the present work, we exploited the distinct size and diffusion properties of monomers and oligomers of MBP-E6 to separate these two species using a rapid batch affinity preparation approach. This led us to obtain a fast, single-step protocol for the preparation of two test E6 oncoproteins under monomeric and biologically active form: HPV16 F47R 4C/4S E6 (a solubility-enhanced mutant of HPV16 E6) and HPV8 E6 (prone to rapid self-association). The long-lasting and non-parallelizable steps of the previous protocol (ultracentrifugation and preparative gel-filtration) were substituted by a fast batch affinity chromatography on reticulated nickel resins. The reticulation state of the resin, the elution protocol and the contact time were adjusted for optimal separation of monomeric and oligomeric MBP-fused E6 proteins. We show that this new and fast protocol is effective enough to obtain soluble and active E6 protein samples suitable for protein–peptide interaction assay as revealed by surface plasmon resonance (SPR). This protocol is amenable

to robotization, offering the ability to prepare in a parallel fashion multiple protein samples displaying optimal monodispersity.

Results

Protein constructs used for this study

To develop and evaluate a new E6 purification protocol, we used two distinct HPV E6 oncoproteins. On the one hand, HPV16 E6 F47R 4C/4S (thereafter named 16E6mut) is a solubility-optimized mutant of HPV16 E6. The structure of this mutant has been solved by X-ray crystallography and several of its interactions have been precisely characterized by surface plasmon resonance (SPR) and Isothermal Titration Calorimetry [22]. On the other hand, HPV8 E6 (thereafter named 8E6) has a strong tendency to self-associate and is more challenging to purify as a monomer (unpublished observations).

As depicted in Fig. 1, HPV E6 constructs were overexpressed as fusions to the C-terminus of bacterial MBP. MBP is known to favor solubilization of recalcitrant recombinant proteins [1, 2], and can be used for affinity purification on amylose resin. In addition, the constructs include a 6-His tag for immobilized-metal affinity chromatography (IMAC) on Nickel resin and a TEV (Tobacco Etch Virus) protease cleavage site allowing elimination of N-ter purification tags.

Impact of the reticulation level of affinity resins on the aggregation level of purified MBP-E6 samples

In the protocol published previously for the purification of soluble MBP-E6 protein [9, 10], the initial affinity step was performed on amylose resin. In such conditions, a fraction of the MBP-E6 sample formed oligomers that were detected by static light scattering in a fluorimeter. These soluble oligomers can be eliminated by means of overnight ultracentrifugation followed by preparative gel-filtration [19]. However, this approach requires substantial amounts of protein in order to be efficient and is not suitable for parallel purification of HPV E6 proteins. We hypothesized that testing resins of different reticulation degree might be an alternative strategy to reduce the binding of the oligomeric proteins to the affinity resin since large particles cannot interact with a highly reticulated resin, thus promoting the preferential interaction of small molecules with the affinity resin.

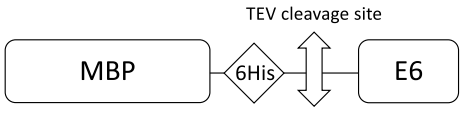


Fig. 1 Construct for the production of recombinant HPV E6 protein: MBP is used as a solubilizing fusion while 6His tag allows IMAC on Nickel resin. Purification tags can be removed by TEV protease cleavage

Here, we compared the aggregation levels of purified MBP-E6 protein obtained either with the same amylose resin than the one used in the initial protocol (thereafter called A), or with six different reticulated nickel resins (numbered from N1 to N6, as detailed in Table 1). In order to test the preferential binding of the monomer to the different resins, elution was performed by centrifugal filtration on a filter plate to maximize the recovered liquid fraction and the detection of all protein species eluted from the resin. The purity of the final fractions was satisfactory, as checked by microfluidics capillary electrophoresis in denaturing conditions (Fig. 2). The bands corresponding to the protein of interest are visible between the 44 and 72 kDa markers for MBP-16E6mut (expected MW = 61.4 kDa) and between the 48 and 69 kDa markers for MBP-8E6 (expected MW = 60.9 kDa).

The oligomeric state of the purified protein samples was assessed by analytical SEC, based on absorbance at 280 nm. Figure 3 shows chromatograms of MBP-16E6mut (A) and MBP-8E6 (B) purified using the initial protocol in which the bacterial lysate of overexpressed E6 protein was incubated 2 h with amylose resin. After washing the resin, protein was eluted by maltose-containing buffer and recovered by centrifugal filtration.

According to the column calibration based on (Eqs. 1 and 2), the elution volume for MBP-16E6mut purified on amylose resin ($V_e = 1.90$ mL) corresponds to a molecular weight MW of 75 kDa and a hydrodynamic radius R_h of 37 Å. For MBP-8E6, a similar peak appears at an elution volume V_e of 1.91 mL, which would correspond to a molecular weight MW = 72 kDa and a hydrodynamic radius $R_h = 37$ Å. Considering the accuracy of the method and the fact that calibration is based on the diffusion properties of globular proteins, we assume that both peaks refer to monomeric MBP-E6 (theoretical molecular weights: MW (MBP-16E6mut) = 61.4 kDa, MW (MBP-8E6) = 60.9 kDa). Since MBP-E6 fusions are not globular [22], the experimental elution volumes are expected to be slightly lower than the calculated elution volume obtained from the calibration ($V_{e\,theo} = 1.95$ mL).

For each tested resin, analytical SEC was performed. Based on absorbance at 280 nm, the monomeric and oligomeric protein were detected at a given elution volume. We measured the following elution volumes for monomeric MBP-16E6mut and MBP-8E6 at $V_e = 1.92 \pm 0.01$ mL and MBP-8E6: $V_e = 1.99 \pm 0.04$ mL, respectively. All peaks with $V_e < 1.90$ mL correspond to particles larger than monomer, that were therefore considered as oligomeric proteins (Fig. 3, Additional file 1). Elution peaks were integrated based on the signal of absorbance at 280 nm. Their areas were utilized

Table 1 Panel of resins used for affinity purification, gathering technical data provided by the manufacturers

	Manufacturer	Catalogue ref	Reference	Affinity	Nickel ligand	Features	Exclusion limit for globular proteins	Metal ion capacity/ binding capacity	Max. pressure
N 1	GE Healthcare	17-5318-06	Ni Sepharose 6 Fast Flow	Nickel	NTA	6% cross-linked agarose	4000 kDa	Approx 15 µmol Ni^{2+}/mL gel	0.1 MPa
N 2	ABT Agarose Beads Biotechnologies	10BCL-QHNi-5	High Density Nickel 10 BCL-QHNi-5	Nickel	IDA	10% cross-linked agarose	500 kDa	20–40 µmol Ni^{2+}/mL gel	Low pressure
N 3	ABT Agarose Beads Biotechnologies	6BCL-QHNi-2	High Density Nickel 6 BCL-QHNi-2	Nickel	IDA	6% cross-linked agarose	4000 kDa	20–40 µmol Ni^{2+}/mL gel	Low pressure
N 4	ABT Agarose Beads Biotechnologies	6BCL-NTANi-2	Nickel NTA agarose resin 6 BCL-NTANi-2	Nickel	NTA	6% cross-linked agarose	4000 kDa	5–19 µmol Ni^{2+}/mL gel	Low pressure
N 5	ABT Agarose Beads Biotechnologies	6NiRR-2	Nickel Rapid Run 6 NiRR-2	Nickel	IDA	6% agarose	4000 kDa	5–19 µmol Ni^{2+}/mL gel	0.3 MPa
N 6	ABT Agarose Beads Biotechnologies	6NTANiRR-2	Nickel NTA rapid run 6 RR-NTANi-2	Nickel	NTA	6% agarose	4000 kDa	5–19 µmol Ni^{2+}/mL gel	0.3 MPa
A	NEB New England Biolabs	E8022S	Amylose resin high flow	Maltose		Cross-linked agarose	Unspecified	7 mg MBP-protein/mL gel	0.5 MPa

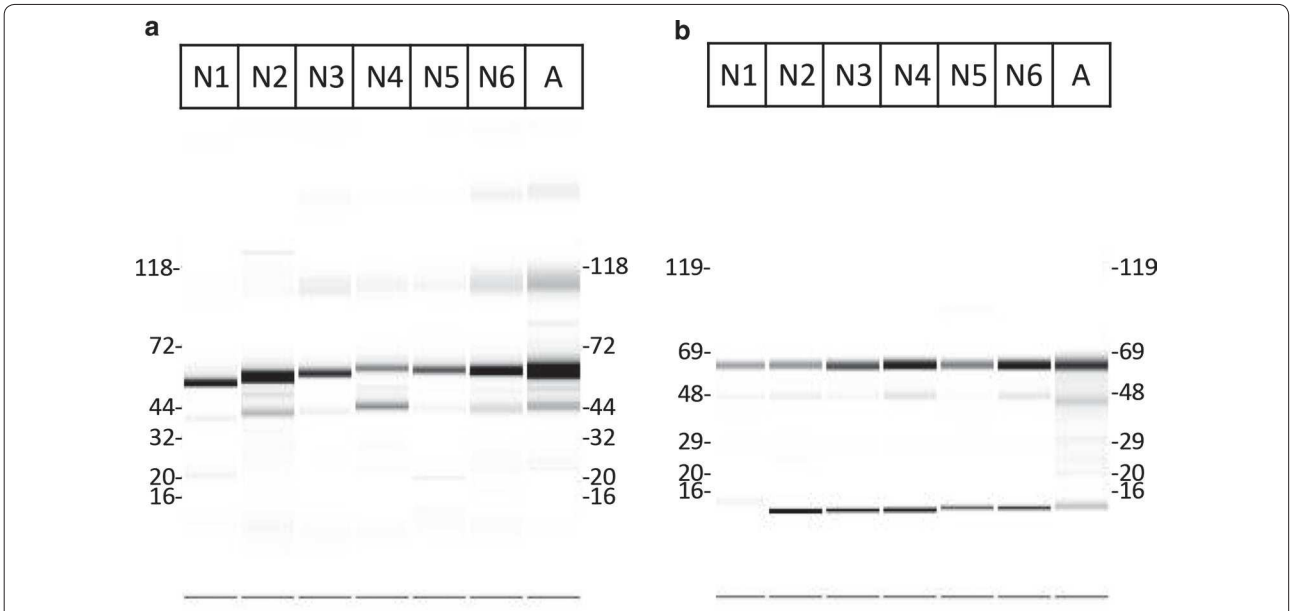


Fig. 2 Example of analysis of purified MBP-E6 fusions by microfluidics capillary electrophoresis in denaturing conditions. **a** MBP-16E6mut, **b** MBP-8E6. The data, which normally consist in electropherograms (e.g., curves describing migrating protein signal intensity as a function of molecular size), are plotted as an intensity-based projection to mimic a conventional SDS PAGE display. Molecular weights, obtained from internal standards, are indicated in kDa. A slight shift may sometimes be observed from one lane to another, as observed in (**a**). These eluted samples were obtained after purification on the different resins (N1–6 for the nickel resins; A for amylose resin), 2 h contact time between bacterial lysate and resin, filter elution

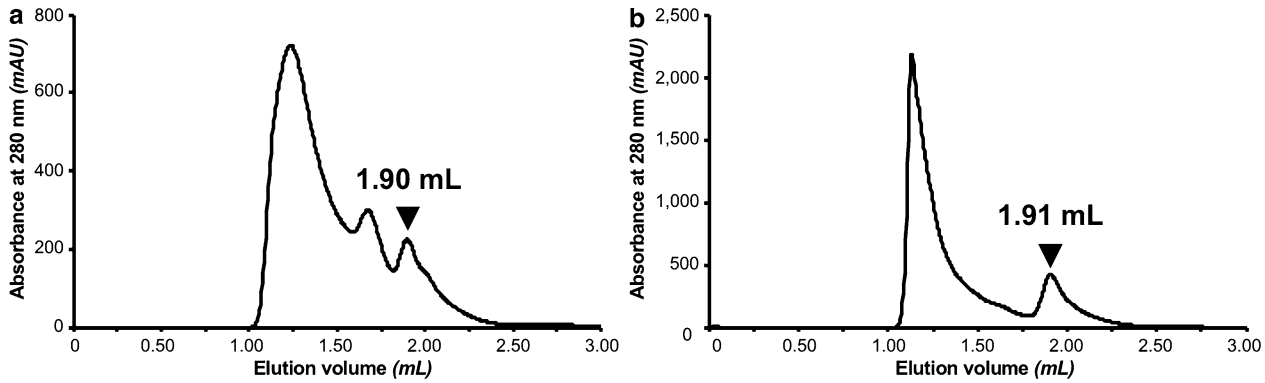


Fig. 3 SEC chromatograms of purified MBP-E6 proteins with the initial protocol (amylose resin; 2 h contact time between bacterial lysate and resin; filter elution). The elution peak corresponding to the monomeric protein is indicated by an arrow, according to column calibration using standard proteins. **a** MBP-16E6mut, **b** MBP-8E6

to quantify monomeric and oligomeric species. As summarized in Table 2, the fraction of monomer in the same purification conditions is generally higher for MBP-16E6mut than for MBP-8E6. This result confirms that MBP-8E6 is highly prone to self-association, consistent with our previous observations.

The difference between the MBP-16E6mut reference protein and the challenging MBP-8E6 protein is particularly visible in Fig. 4, since the proportion of oligomers over total purified protein is much higher for MBP-8E6 (values comprised between 70 and 93% for the different tested nickel resins, Table 2) than for MBP-16E6mut (between 5 and 70%). It also clearly appears that amylose resin retains a higher proportion (90%) of MBP-16E6mut oligomers than nickel resins, confirming the hypothesis of oligomer exclusion by reticulated nickel resins. A standard deviation of 15% unit was inferred from the duplication of several experiments, warranting significance. Three nickel resins allow to purify 85% or more of monomeric MBP-16E6mut, namely N1 (95%), N2 (87%) and N4 (85%). In addition, quantification of the total protein amount (Table 2, see "Total A280 GF Peak Area") shows that N4 resin has the highest recovery capacity of total purified protein. Regarding the purified MBP-8E6 sample, the same three resins N1, N2 and N4 show the maximal ratio of monomeric state (19%, 30% and 22%, respectively), although the monomer fraction obtained with the amylose resin (16%) is within the same range as that obtained with nickel resins (7 to 30%).

The present assays showed the usefulness of using a reticulated nickel resin for the purification of MBP-16E6mut, since purification trials allowed to reach up to 95% monomer in a single purification step. However, MBP-8E6, being more prone to oligomerization, could not yet be purified as a predominantly monomeric state at this step.

Optimizing the elution protocol to maximize the proportion of monomeric protein

At the previous stage, resins were thoroughly dried by centrifugal filtration in order to recover maximal liquid fraction and to quantify all oligomer species that were bound (filter elution protocol). After probing the efficiency of nickel reticulated resins, we implemented another elution method aimed at reducing the fraction of oligomers in the final purified fraction. Instead of being filtrated, the resin/elution buffer mixture was centrifuged at low speed (5 min at 500×*g*) and the supernatant was collected immediately after centrifugation. The oligomers are expected to diffuse out of the resin slower than the monomers, thereby increasing the proportion of monomers in the final pipetted material.

Using this adapted elution protocol, the proportion of monomeric MBP-8E6 was significantly improved and reached up to 71% for resin N1 (Fig. 5). Resins N2 and N4 allowed to purify up to 56% of monomeric protein, but N1 remains the only resin allowing to purify the highest total protein concentration and the largest monomer ratio (as defined by the proportion of monomer over the total amount of protein in the sample) for both MBP-8E6 and MBP-16E6mut samples. For MBP-16E6mut, the two resins reaching highest monomer ratio with the new elution method are N1 and N2, with approximately 86% and 88%, respectively. Although the monomer fraction was significantly increased with this "decant and take up" elution procedure, it is noticeable that the total protein amount was reduced by two-fold for most of the resins. By adapting the elution method, we have shown that it is possible to significantly reduce the diffusion of

Table 2 Summary table for purification results: Monomer are defined based on their elution volume, as it follows: MBP-16E6mut: $V_e = 1.92 \pm 0.01$ mL; MBP-8E6: $V_e = 1.99 \pm 0.04$ mL Oligomers are defined as all protein species eluted between void volume (V_0) (included) and monomer elution volume (excluded)

	Monomer (%)	Oligomer (%)	Total A280 GF peak area (mAU x mL)
A: MBP-16E6mut			
(1) Filter elution			
N1	95	5	256
N2	87	13	243
N3	52	48	167
N4	85	15	480
N5	30	70	116
N6	68	32	345
A	10	90	263
(2) Pipetting elution			
N1	86	14	366
N2	88	12	125
N3	60	40	125
N4	72	28	128
N5	39	61	39
N6	69	31	155
B: MBP-8E6			
(1) Filter elution			
N1	19	81	467
N2	30	70	113
N3	12	88	230
N4	22	78	256
N5	7	93	194
N6	9	91	404
A	16	84	486
(2) Pipetting elution			
N1	71	29	301
N2	56	44	117
N3	28	72	227
N4	56	44	121
N5	16	84	156
N6	74	26	89
(3) Optimization of lysate incubation time			
N1			
5 min	74	26	185
30 min	77	23	160
2 h	56	44	220
N2			
5 min	92	8	25
30 min	71	29	38
2 h	77	23	138

Quantification of monomeric and oligomeric species is based on the integration of elution peak (absorbance at 280 nm), with estimated uncertainty of ± 15% for monomer/oligomer ratio and ± 57 mAU × mL for total A80 peak area based on duplicated experiments

oligomeric species in solution, and subsequently to maximize the ratio of monomeric purified protein.

Reducing the contact time of bacterial extract with resin during the initial binding step limits protein oligomer binding

Contact time between affinity resin and cleared bacterial lysate during the initial binding step is another critical parameter for protein purification since it determines the amount and the nature of protein species bound to the resin. On the one hand, a long incubation time should increase the total amount of protein that will be recovered. On the other hand, a shorter incubation time might restrict the size of the proteins interacting with the resin. As large particles such as soluble oligomers diffuse slower in solution, their binding on affinity resin should thus be limited when incubation time is short.

To assess this hypothesis, we selected two resins showing the highest monomer ratio for both the reference MBP-16E6mut and the challenging MBP-8E6 proteins, namely N1 and N2. Since the "decant and take up" elution method already improved the fraction of monomeric MBP-16E6mut by up to 70–90%, this third step in the optimization protocol was focused on MBP-8E6, the protein most prone to self-association in our study. Incubation time was set at either 30 min or 5 min and systematically compared to the 2-h reference time used in the previous trials. Elution was done by pipetting the supernatant after low-speed centrifugation of the resin resuspended in elution buffer. The results presented in Fig. 6 show that the shorter the contact time between the cleared lysate and the resin, the higher the proportion of monomer in the final purified fraction, confirming our assumption. With the N1 resin, about 75% of monomers were purified with either 30 min or 5 min contact time whereas this ratio dropped down to 56% for the 2-h incubation delay. The best monomer ratio for monomeric MBP-8E6 (92%) was obtained by using the N2 resin for purification, with a 5 min contact time and the "decant and take up" elution method. Considering that only 30% of monomers were obtained using this same resin and applying the initial protocol (i.e. 2-h contact time and elution by filtration), this result shows that our optimization brought a significant improvement to purify a biochemically problematic HPV E6 protein. One should note that this quality refinement of the purified protein is at the expense of a decrease of the total amount of the MBP-8E6 protein. It is particularly visible from the purifications with the N2 resin, for which a fivefold decrease of the protein quantity was observed between 2-h and 5-min contact time.

For MBP-16E6mut, the absolute amount of monomeric protein was further increased by applying the above

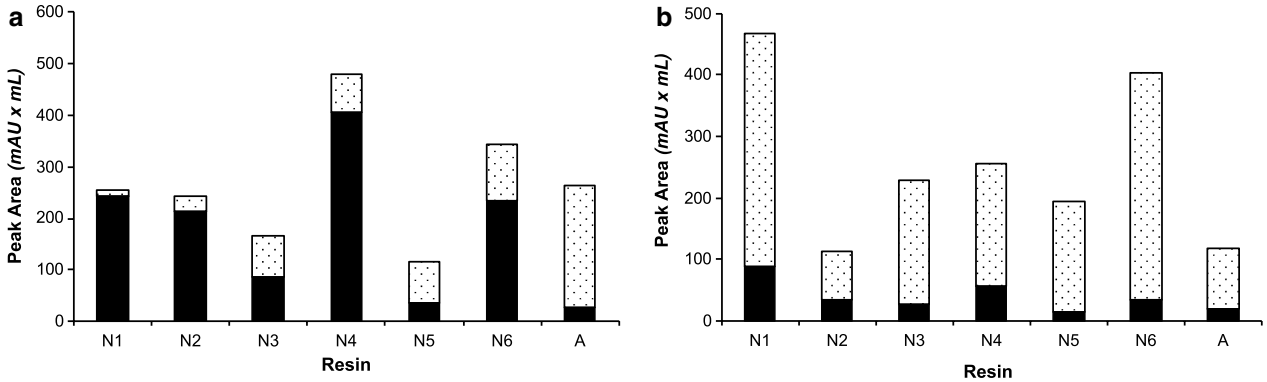


Fig. 4 Tests of various reticulated affinity resins (N1–6: Nickel resins; A: amylose resin, as detailed in Table 1). Elution was recovered by centrifugal filtration (5 min at 500×*g*) of the elution buffer–resin mixture. The ratio between monomer and total protein amount [monomer (black bars) + oligomers (dotted bar)] were calculated from the peak areas of the corresponding elution volumes. **a** MBP-16E6mut. **b** MBP-8E6

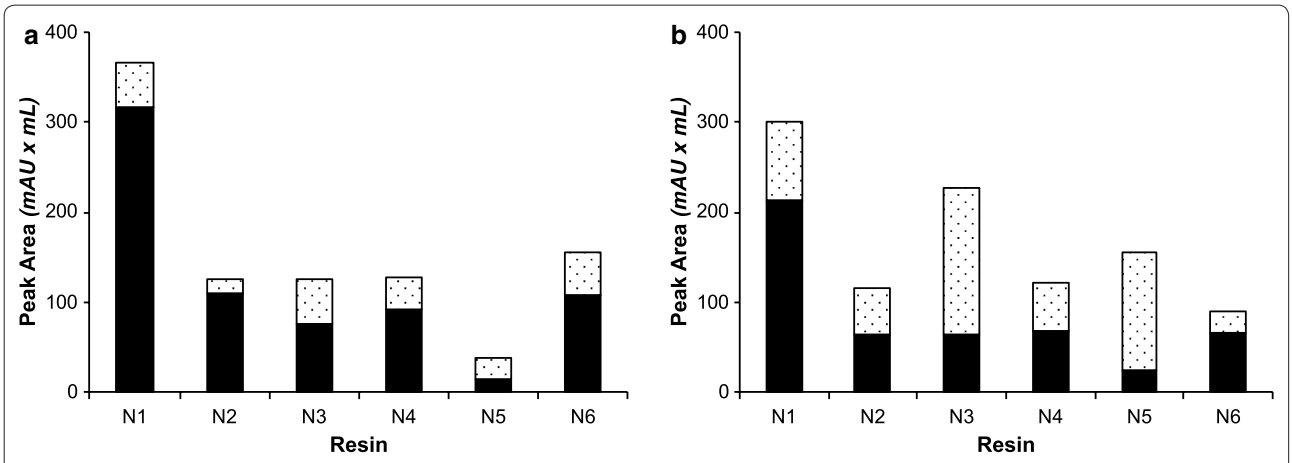


Fig. 5 Test of the "decant and take up" elution method on nickel resins (N1–6: Nickel; A: amylose, as detailed in Table 1). Elution was recovered by pipetting the supernatant after gentle centrifugation (5 min at 500×*g*) of the elution buffer–resin mixture. Monomer: black bars; oligomers: dotted bars. **a** MBP-16E6mut. **b** MBP-8E6

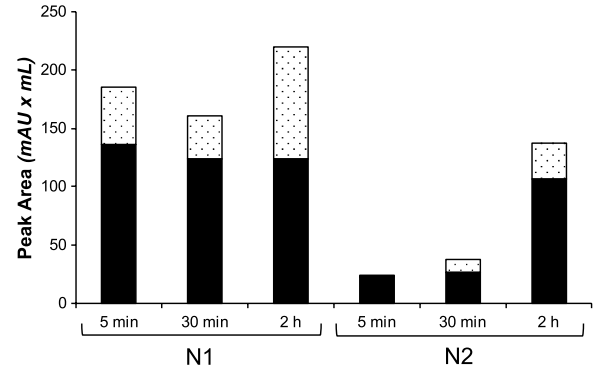


Fig. 6 Optimization of bacterial extract contact time for the purification of MBP-8E6 on nickel resins N1 and N2: three contact times were tested (5 min; 30 min; 2 h). Monomer: black bars; oligomers: dotted bars

refined protocol (Fig. 7). In contrast, the amount of monomeric MBP-8E6 remained approximately the same with both purification strategies. The increase of the ratio of monomeric protein with the optimized protocol can be explained by the drastic elimination of oligomeric species (in particular those from the void volume at 1.18 mL).

Probing the activity of the purified protein by SPR

In order to assess the activity of protein samples purified according to the optimized protocol, we performed a protein–peptide interaction assay by SPR. This was achieved using 16E6mut in order to compare with SPR data previously obtained on a related 16E6 mutant using the original large-scale approach [30].

We purified MBP-16E6mut according to our batch protocol: 5 min contact time between the bacterial

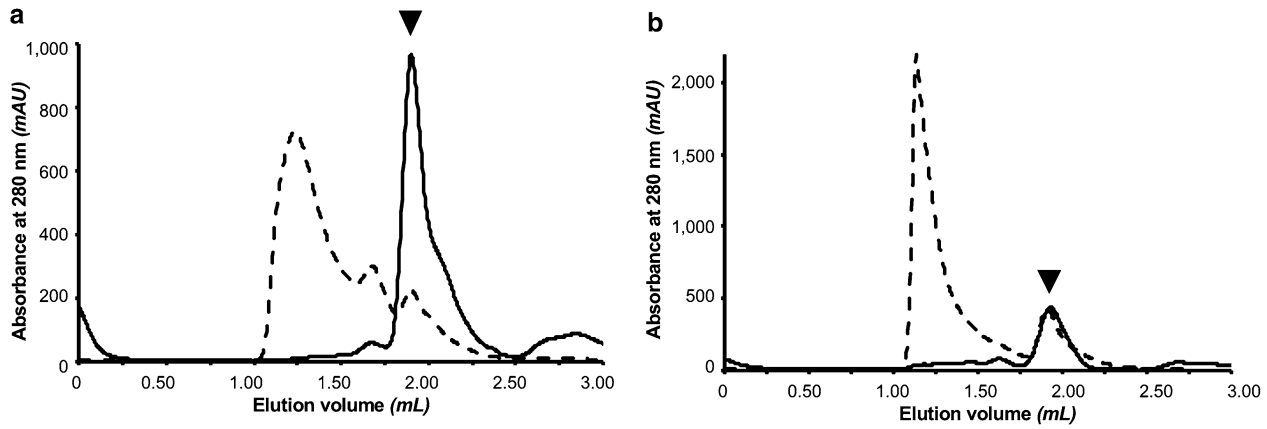


Fig. 7 SEC chromatograms for purified MBP-E6 proteins with initial protocol, represented with dashed line (amylose resin; 2 h contact time between bacterial lysate and resin; filter elution) versus optimized protocol, represented with continuous line (N2 nickel resin; 5 min contact time; "decant and take up" elution). The elution peak corresponding to monomeric protein is indicated by the arrow. **a** MBP-16E6mut. **b** MBP-8E6

lysate and the affinity resin, followed by "decant and take up" elution by pipetting the supernatant after short centrifugation in Eppendorf tube. We performed this protocol with N1 and N2, the two resins that gave the highest monomer ratio after purification with the optimized elution protocol. In order to remove imidazole from protein solution, the final fractions were desalted in the SPR running buffer by bench desalting column.

Following this protocol, we purified the equivalent of 8.5 µmol of MBP-16E6mut per liter of bacterial culture, which is largely sufficient to perform parallel SPR assays (in our setup, the minimum protein amount required for one SPR assay including 10 cycles is 2 nmole).

MBP-E6 was tested against its prototypical target peptide, the 16-mer LXXLL motif of E6AP (PESSELTLQELLGEER, called thereafter *E6APwt*) [21]. The CAPture kit system (GE Healthcare) was used to reversibly immobilize biotinylated peptides on the chip. Each cycle consists of an immobilization of the peptide, followed by the injection of the protein at a given concentration, and a regeneration step to wash out both the remaining protein and the immobilized peptide of the surface. Figure 8b shows the superimposition of time-aligned sensorgrams

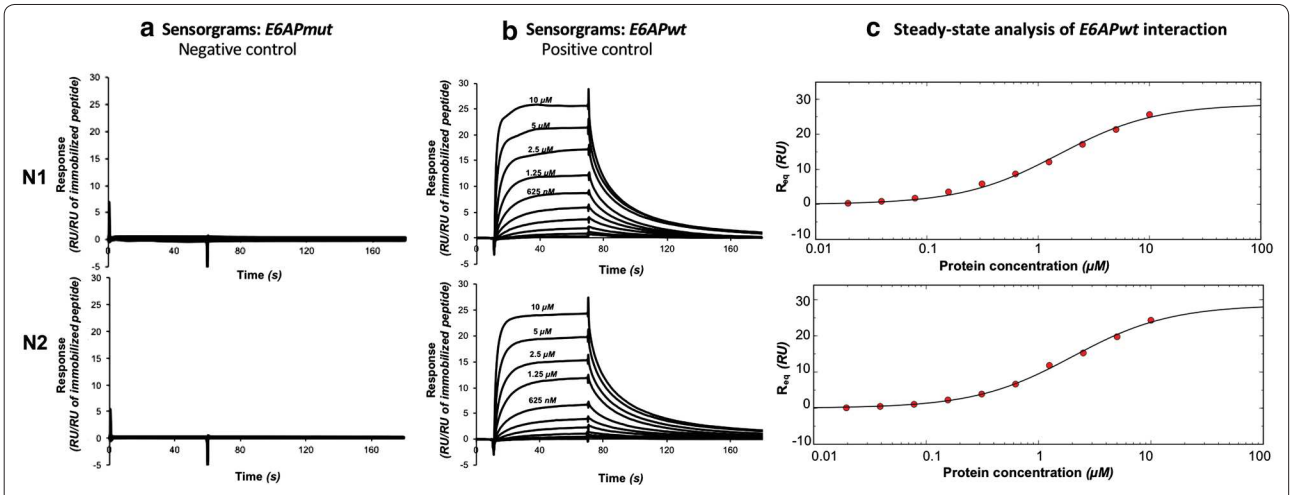


Fig. 8 SPR data for the assessment of MBP-16E6mut interaction properties. The protein was purified on nickel resins N1 (top) and N2 (bottom) according to the optimized protocol (5 min contact time between the bacterial extract and the affinity resin, "decant and take up" elution). **a** Sensorgrams for the interaction with E6APmut, plotting the response normalized by the peptide immobilization level versus time. **b** Sensorgrams for the interaction with peptide E6APwt, same axes. **c** Steady-state analysis of interaction with E6APwt, plotting the equilibrium response R_{eq} obtained in a 5-s. Window versus protein concentration

obtained for different concentrations after reference substraction and normalization according to the peptide immobilization level. As a negative control, we tested the interaction against an E6AP mutant (PESSELT**A**QELLGEER, called thereafter *E6APmut*) abolishing the binding (Fig. 8a). No significant signal was observed for this control. In particular, the association phase does not display any concentration-dependent signal.

Steady-state analysis was performed to estimate dissociation constants (K_D) of the MBP-16E6mut interacting with *E6APwt* (Fig. 8c). For the protein samples purified on resins N1 and N2, the estimated K_D were 1.6 ± 0.3 µM and 2.0 ± 0.3 µM, respectively. These data are in close agreement with a previously published SPR analysis of the binding of purified HPV16 E6 to the LXXLL motif of E6AP [30]. In that study, the obtained K_D was 4.15 ± 0.52 µM. The moderate difference in the absolute affinity constant measured in that previous work might be due to several changes in the experimental setup. In the former publication, the recombinant peptide was fused to GST and further captured on CM5 chip via an anti-GST antibody, whereas biotinylated synthetic peptides used in the present study were reversibly captured on CAP chip. In addition, the previous publication used a distinct mutant of HPV16 E6, called E6 6C/6S. This former mutant included two additional cys/ser mutations that were later shown to slightly destabilize the N-terminal domain of HPV16 E6 [31]. Despite these experimental differences, the good agreement between affinity measurements indicates that the optimized single-step purification protocol is efficient to obtain protein sample in an active form.

Discussion

Quality of recombinant protein sample can be improved by optimizing either the expression or the purification process. The present work mainly focused on the purification process to maximize the yield of soluble overexpressed protein. Our purpose was to obtain an MBP-E6 sample as homogeneous as possible from the available material based on the expression conditions in *E. coli* that we previously published [9].

E6 oncoproteins from papillomaviruses contain two zinc-binding domains and have been found to interact with some host proteins through LXXLL motif recognition, such as E6AP ubiquitin ligase [32]. After overexpression in *E. coli* and purification, it was reported that MBP-E6 forms mixtures of soluble active monomers and soluble oligomers comprising misfolded E6 moieties (thus unable to specifically bind LXXLL motifs) [10]. These oligomers can be eliminated by overnight ultracentrifugation followed by size-exclusion chromatography [22]. Although this approach is appropriate for X-ray crystallography, a faster and easier protocol was designed here for the purpose of performing protein–peptide interaction assays in parallel. As compared to structural studies, binding assays generally need less material, yet high homogeneity and solubility are required. Thus, our ultimate quality criterion was the percentage of active protein rather than the overall purification yield. We designed a single-step batch purification process maximizing the proportion of monomeric and active MBP-E6 protein. This approach is compatible with parallel purifications and generates sufficient amounts of protein for our biophysical assays. In the present study, we investigated the purification of two HPV E6 oncoproteins fused to MBP: the solubility-enhanced mutant HPV16 E6 F47R 4C/4S (16E6mut) whose structure was solved by X-ray crystallography, and the HPV8 E6 (8E6) which, in our experience, was particularly recalcitrant to purify as a monomer.

This study aimed at the preferential selection of monomeric protein species within a single purification step. Oligomers can be distinguished from monomers based on their size, more precisely their volume in solution. The radius of a particle in solution is correlated with the diffusion coefficient D according to the Stokes–Einstein formula [33]:

$$D = kT/6\pi\eta r$$

with k the Boltzmann constant, T the temperature, η the viscosity coefficient of the medium and r the radius. Thus, larger protein species diffuse slower than monomers in solution. Based on these size and diffusion properties, we optimized three parameters in the purification protocol: (i) reticulation of affinity resin, (ii) contact time between the bacterial lysate and the resin and (iii) elution by "decant and take up" method. Adapting the resin reticulation is a way to select particles based on their size in solution that is faster and easier to parallelize than SEC. The two other optimization lines rely on the distinct diffusion properties of oligomers and monomers. First, a reduced contact time limits the binding of oligomeric species on the resin. Second, a fast elution method by pipetting only the supernatant fraction enables the preferential recovery of monomeric species.

To discriminate particles based on their volume in solution, reticulated resins can be selected to set a specific exclusion limit by adapting the pore size and the agarose concentration. Protein species exceeding a threshold hydrodynamic volume in solution will not be able to enter the affinity resin pores. This property allows a double selection based on the protein size and the affinity tag, within a single step.

We compared the size of particles retained on different types of resin, amylose resin and nickel resins with

distinct reticulation levels. In particular, resin N2 was a 10% cross-linked agarose resin excluding proteins above 500 kDa (10% BCL Agarose Bead Standard, ABT) that was functionalized to bind Nickel, whereas all other nickel-chelating resins have an exclusion limit of about 4000 kDa. Our results (notably with MBP-16E6mut) showed that amylose resin leads to the purification of a larger fraction of oligomers than nickel resins. N2 resin was the most efficient resin to purify the monomeric form of both MBP-16E6mut and MBP-8E6, which confirms the efficiency of adapting affinity resin pore size to limit the purification of aggregates. N1 resin was the second most efficient resin to maximize the ratio of monomeric protein. In addition, the highest total protein amount (including both monomer and oligomers) was purified with N1 resin.

The proportion of purified unwanted oligomers can be minimized by reducing incubation times at key steps of the process. First, the contact time between the bacterial crude extract and the affinity resin can be reduced in order to minimize the binding of oligomers, as demonstrated with the recalcitrant MBP-8E6 construct for which the highest monomer fraction was obtained for the minimum contact time. Second, the elution method can be designed to recover a predominantly monomeric species capable to rapidly diffuse to the supernatant. This approach increased by two-fold the ratio of purified monomeric MBP-8E6, despite its natural tendency to form a majority of oligomeric species.

Finally, we were able to confirm by SPR the binding affinity between *E6APwt* LXXLL peptide and MBP-16E6mut purified according to our optimized protocol and using the two nickel reticulated resins that allowed to purify highest proportion of monomeric protein (N1 and N2). The protein did not interact with the mutated motif *E6APmut* (AXXLL), which rules out the possibility of unspecific interactions that might happen with misfolded proteins and/or aggregated proteins.

Conclusions

In the present work, we designed a single-step purification strategy optimized for maximal recovery of monomeric MBP-E6 that we accurately quantified by means of analytical SEC. By using a customized reticulated resin, adapting the contact time and performing fast elution, we were able to reach high monomer ratios in the final purified fraction. This method was particularly efficient for the solubility-enhanced mutant MBP-16E6mut and the highly aggregation-prone MBP-8E6. As assessed by SPR, the purified MBP-16E6mut protein interacts with its known partner within the expected range of affinity. This method is promising for forthcoming biophysical studies with recalcitrant proteins, while allowing to perform parallel purifications. The approach should also be easily adaptable to robotization. It would be interesting to automatize this protocol for purifying at high throughput families of proteins, which, like E6 oncoproteins, are difficult to fold except when fused to MBP or other soluble carrier proteins. The overexpression of such constructs leads to a limited proportion of properly folded and active monomers that require to be separated from non-active oligomers, which is now possible using the protocol presented here.

Methods

Customized Nickel resin N2

Preparation

High Density Nickel 10BCL has been manufactured as a customized product by Agarose Bead Technologies. The resin consists of crosslinked 10% agarose beads to which a Nickel chelating group has been immobilized. This chelating group has been obtained by a modification of the procedures described elsewhere [34, 35] using 10% BCL Agarose Bead Standard as raw material. The epoxide generated has been treated with iminodiacetic acid and charged with Ni^{2+} ions according to a modification of the method previously described.

Characterization

The characteristics of 10% BCL Agarose Bead Standard for protein separation based on their molecular weight are the following: the fractionation range for globular proteins is 1×10^4–5×10^5 Da and the exclusion limit is $> 5 \times 10^5$ Da. Nickel was quantified by spectrophotometric assay, resulting in 34 µmol Ni^{2+}/mL gel.

Expression and fast batch purification procedure

The sequences encoding for 16E6mut (already described in [22]) and for 8E6 were cloned into a pETM-41 vector by using NcoI and Acc65I sites, allowing the production of MBP-6His-TEV site-E6 fusions (Fig. 1). The sequence encoding for 8E6 was kindly provided by Yves Jacob from Pasteur Institute (Paris, France) and corresponds to a variant with three substitutions as compared to the reference protein sequence (Uniprot ID: P06428), namely Y9N, A26E and S36L.

Both constructs were expressed in *Escherichia coli* BL21 (DE3) cells, grown in LB medium supplemented with kanamycine 50 µg/mL at 37 °C until $OD_{600} \approx 0.7$. Expression was induced by adding 0.5 mM IPTG (Isopropyl β-D-1-thiogalactopyranoside) and 100 µM $ZnSO_4$. Cells were grown overnight at 16 °C and harvested by centrifugation. Pellets were stored at − 20 °C.

Purification buffers were thoroughly degassed and equilibrated with argon before adding 2 mM DTT. Culture pellets were then resuspended in lysis buffer, which

was Buffer A (Tris 50 mM pH 8; NaCl 400 mM; DTT 2 mM) supplemented with 5% (w/v) glycerol, 0.25 µg/mL RNase I, 0.25 µg/mL DNase I, lysozyme at approximately 1 µg/mL and anti-protease cocktail EDTA-free (Roche) according to the manufacturer instruction. This buffer was supplemented with 10 mM imidazole for purification on Nickel resins. Pellet equivalent to 250 mL culture volume was resuspended in 10 mL lysis buffer. Cells were lysed by sonication on ice and then centrifuged at 100,000×*g* at 4 °C for 45 min. The supernatant was then incubated at 4 °C with either nickel (100 µL resin for 250 mL culture pellet) or amylose resins (300 µL resin for 250 mL culture pellet), equilibrated in Buffer A supplemented with 20% of the recommended concentration of anti-protease cocktail. The incubation time of the cleared lysate with the resin was 2 h in the initial purification tests, before to be evaluated as a variable parameter (5 min, 30 min or 2 h) and reduced to 5 min in the final optimized protocol. The resins were pelleted by 5 min-centrifugation at 500×*g* at 4 °C; the supernatant was discarded and the resins were first washed in Buffer A (30 min incubation at 4 °C), then in Buffer B (Tris 50 mM pH 8, NaCl 1 M, DTT 2 mM, 20% of the recommended concentration of anti-protease cocktail and supplemented with 10 mM imidazole for Nickel resins) with same incubation time and temperature. Resins were re-equilibrated in Buffer A during 10 min at 4 °C before elution. Resins were then transferred in Eppendorf tubes, pelleted by centrifugation for 5 min at 500×*g*, 4 °C and supernatants were discarded. Finally, resins were resuspended in elution buffer (Buffer A supplemented with 20% anti-protease cocktail and, for Nickel resins: imidazole 600 mM; for Amylose resin: maltose 15 mM). 100 µL Nickel resin were eluted with 140 µL imidazole elution buffer while 300 µL amylose resin were eluted with 200 µL maltose elution buffer.

For standard elution, resins were transferred in 96-well filter plate and centrifuged 5 min at 500×*g* and 4 °C for maximal recovery of liquid phase by filtration. For the optimized elution protocol, the tube content was then rapidly mixed by gentle vortex and immediately centrifuged during 5 min at 500×*g* and 4 °C. The supernatant was recovered and used as final purified protein solution. Protein concentration was determined by Nanodrop measurement, based on absorbance at 280 nm. Final protein samples were systematically analyzed either by SDS-PAGE or by microfluidics capillary gel electrophoresis in order to check the presence of protein contaminants in denaturing conditions. Prior to SPR assays, protein samples were desalted on Illustra NAP-5 columns (GE Healthcare) equilibrated in Buffer A allowing to eliminate imidazole or maltose from the solution.

Microfluidics capillary gel electrophoresis

Purification samples (from the washing steps and elution) were transferred in 96-well plates, mixed with the Caliper sample buffer and boiled according to the manufacturer instructions. The plates were loaded and measured on LabChip GX II device (Caliper, Perkin Elmer), with HT Protein Express 100 High Sensitivity protocol (10–100 kDa). The kit allows to label the proteins with a fluorescent dye. This dye is excited by a laser during protein capillary electrophoresis in denaturing conditions. The emitted fluorescence signal is plotted versus migration time, leading to the protein separation according to their molecular weight. Data processing is performed with LabChip GX II software. Migration time is first converted into molecular weight using standard protein markers before a quantitative analysis is performed thanks to empirical linear correlation between fluorescence intensity and protein amount.

Analytical size-exclusion chromatography

Samples used for analytical SEC were centrifuged 5 min at 13,700×*g*, 4 °C and only the supernatant was loaded on the column Superdex 200 Increase 5/150 GL (GE Healthcare) previously equilibrated in Buffer A. SEC run was performed on ÄKTA purifier (GE Healthcare), by injecting 50 µL per sample, with flow set at 0.150 mL/min. Spikes of absorbance 280 nm signal due to air bubbles were excluded from the chromatogram before performing data analysis. The original chromatograms are provided in Additional file 2.

We calibrated the column with Gel Filtration Calibration Kit High Molecular Weight (GE Healthcare), using a solution of dextran Blue (0.1 mg/mL) and two mixes. Mix 1 was composed of ovoalbumine (4 mg/mL), aldolase (4 mg/mL) and thyroglobuline (5 mg/mL). Mix 2 contained ferritine (0.3 mg/mL) and conalbumine (3 mg/mL). Each of the three calibration solution was loaded in a 50 µL loop and injected separately at a flow rate of 0.150 mL/min on the column previously equilibrated in Buffer A. The resulting elution volumes were used to establish a calibration plot to determine the molecular weight *MW*, based on the following equation [33]:

$$K_{av} = \frac{V_e - V_o}{V_t - V_o}$$

where K_{av} stands for the available distribution coefficient, V_e corresponds to the elution volume of each injected protein, V_o is the void volume of the column (determined by the elution volume of the dextran blue) and V_t is the total bed volume (3 mL with this specific column). Thus, the coefficients *a* and *b* were determined by linear regression using the following equation [33]:

$$K_{av} = a \times \log(MW) + b \quad (1)$$

Moreover, the hydrodynamic radius R_h was determined according to:

$$\sqrt{-\log_{10}(K_{av})} = c \times R_h + d \quad (2)$$

where c and d coefficients are determined by linear regression [33].

The area of elution peaks (based on absorbance at 280 nm) was calculated using the evaluation module of Unicorn 5.31 software.

According to the column calibration, we inferred the elution volume of monomeric (approximately 1.92 mL) and oligomeric MBP-E6 ($V_e < 1.90$ mL). For each assay, the area of the elution peak corresponding to the monomer was divided by the area of peaks corresponding to all eluted protein species (monomeric and oligomeric), allowing us to measure the ratio of monomer and oligomer in the purified protein sample.

Regarding the ratios of monomeric protein, we observed that the duplication of the data leads to standard deviations never exceeding 15% unit. This value has consequently been considered as the uncertainty of monomer/oligomer ratios in this work. This uncertainty, although given as a percentage, corresponds to the absolute variation of the ratio of monomeric protein and not as a relative variation coefficient.

Peptide synthesis

To test the ability of the purified protein to interact with LXXLL peptides, we used two synthetic peptides: *E6APwt* (PESSELTLQELLGEER) and *E6APmut* (PESSELTAQELLGEER). Both peptides were N-ter biotinylated with a TTDS spacer (*N*-(13-amino-4,7,10-trioxa-tridecayl)-succinamic acid) (Iris Biotech GMBH) and synthesized either by JPT Innovative Peptide Solutions or by the peptide synthesis service at IGBMC with 70–80% purity. The lyophilized peptides were resuspended in water at a final stock concentration of 5 mM and stored at − 20 °C.

Surface plasmon resonance

Peptide–protein interaction assays were performed by surface plasmon resonance (SPR) on a Biacore T200 instrument (GE Healthcare—Biacore) at 25 °C. The running buffer was Tris 50 mM pH 8, NaCl 400 mM, DTT 2 mM and 0.005% (v/v) surfactant polysorbate 20 (GE Healthcare). We used Biotin CAPture kit (GE Healthcare—Biacore) for the reversible capture of biotinylated peptides on a chip. Briefly, the chip is coated with a deoxyribooligonucleotide that hybridizes with the complementary strand bound to streptavidin. The biotinylated peptide binds to streptavidin and can be washed by dehybridizing the two oligonucleotides. An empty control surface was systematically included on every cycle to serve as a reference for non-specific binding of the analyte to the matrix and for monitoring changes in solution refractive index. This reference surface was treated as the peptide surfaces except that peptide injection was omitted. Kinetic runs were performed by injecting series of two-fold cascade dilutions of the analyte MBP-E6 samples, starting from 10 µM.

Thus, at the beginning of each binding cycle, CAPture reagent (diluted 5 times in running buffer) was injected on all channels during 300 s at 2 µL/min. For peptide surfaces, peptide solution (50 nM) was injected during 15–30 s at 10 µL/min in order to reach 4–15 RU immobilization level. Protein sample was then injected during 60 s at 30 µL/min, followed by 120 s buffer flow. The surfaces were then regenerated by injecting regeneration solution as indicated by the manufacturer (guanidine hydrochloride 6 M; sodium hydroxide 250 mM) for 60 s at 5 µL/min.

The SPR signals from the regions corresponding to the protein injection and post-injection phases were plotted as RU versus time. Data were first processed using the Biacore T200 Evaluation 1.0 software (GE Healthcare, Biacore Life Science, Uppsala, Sweden). Sensorgrams obtained for the different protein concentrations were corrected for buffer effects and bulk refractive index changes by subtracting the empty cell signal, and subsequently normalized according to the peptide levels which can slightly differ from cycle to cycle due to the immobilization process. The steady-state binding signal was derived by averaging the signals at equilibrium within a five second-window (R_{eq}). Steady-state analysis was performed using an in-house Python script by fitting the average and normalized signal R_{eq} as a function of total analyte concentration, and assuming a simple 1:1 interaction binding isotherm model. Uncertainties of the derived parameters were estimated using Monte-Carlo approach, considering an experimental uncertainty of 5 RU. This value was estimated by duplicating the full experimental cycle corresponding to a single protein concentration.

Additional files

Additional file 1. Corrected size-exclusion chromatograms. Prior to protein peak integration, spikes due to air bubbles were excluded. **A**: Chromatograms for analytical SEC of MBP-16E6mut. In order to control to oligomeric state of the purified protein, we performed systematic analytical SEC on the final protein sample. The calibration of the column allowed

us to estimate the size of the particles at different elution volumes, thus the elution peak of monomeric protein is indicated by an arrow. Note the increase of the monomer fraction compared to the total amount of purified protein when shifting from the filter elution to the "decant and take up" elution method. **B**: Chromatograms for analytical SEC of MBP-8E6. On this protein challenging to purify, the increase of the monomer fraction (indicated by an arrow) between the two elution methods is dramatic. **C**: Chromatograms for analytical SEC of MBP-8E6: optimization of bacterial extract contact time. The elution peak of monomeric protein is indicated by an arrow. The decrease of the oligomeric fraction (eluted between the void volume V_0 and the monomer fraction) is particularly visible for nickel resin N1.

Additional file 2. Raw size-exclusion chromatograms. We provide in this file the raw data of size-exclusion chromatograms used in the study (absorbance at 280 nm versus elution volume). The protein (MBP-16E6mut, MBP-8E6), resin (N1…N6, A), elution method (filter, decant) and contact time for optimization assays (2 h, 30 min, 5 min) are indicated in the worksheet names. The first worksheet entitled Readme contains explanations on the nomenclature used for worksheet titles.

Abbreviations
DTT: dithiothreitol; HPV: human papillomavirus; IMAC: immobilized-metal affinity chromatography; IPTG: isopropyl β-D-1-thiogalactopyranoside; MBP: maltose-binding protein; SEC: size-exclusion chromatography; SPR: surface plasmon resonance.

Authors' contributions
GT and AB designed the experiments; AD, AB and CK performed the experiments, SPR; YN and AB analyzed data and generated figures (for SPR and analytical SEC respectively); PA, LSJ and EDC designed and produced custom affinity resins; AB; YN and GT wrote the manuscript. All authors read and approved the final manuscript.

Author details
[1] Équipe Labellisée Ligue 2015, Department of Integrated Structural Biology, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U1258/CNRS UMR 7104/Université de Strasbourg, 1 rue Laurent Fries, BP 10142, 67404 Illkirch, France. [2] ABT-Agarose Bead Technologies, C/La Forja, 9, Torrejón de Ardoz, 28850 Madrid, Spain.

Acknowledgements
The authors thank D. Altschuh for the optimization of Biacore assays, P. Eberling for peptide synthesis, A. Vanden Broeck for his valuable help on size-exclusion chromatography columns, P. Klein for careful reading and colleagues of the Travé team for helpful discussion and support.

Competing interests
L.S.J., E.D.C and P. A. declare competing financial interests since ABT (Agarose Bead Technologies), the company they belong to, provided affinity beads. Other authors declare that they have no competing interests.

Availability of data and materials
The datasets analyzed during the current study are included within the article and additional files.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Funding
This work received institutional support from le Centre National de la Recherche Scientifique (CNRS), Université de Strasbourg, Institut National de la Santé et de la Recherche Médicale (INSERM) and Région Alsace. The work was supported in part by grants from Ligue contre le Cancer (équipe labellisée 2015 and fellowship to A.B.), National Institutes of Health (Grant R01CA134737), Instruct (ESFRI), and the French Infrastructure for Integrated Structural Biology (FRISBI). The authors declare that the content is solely their responsibility and does not represent the official views of the National Institutes of Health.

Publisher's Note

Received: 10 October 2018 Accepted: 23 November 2018
Published online: 01 December 2018

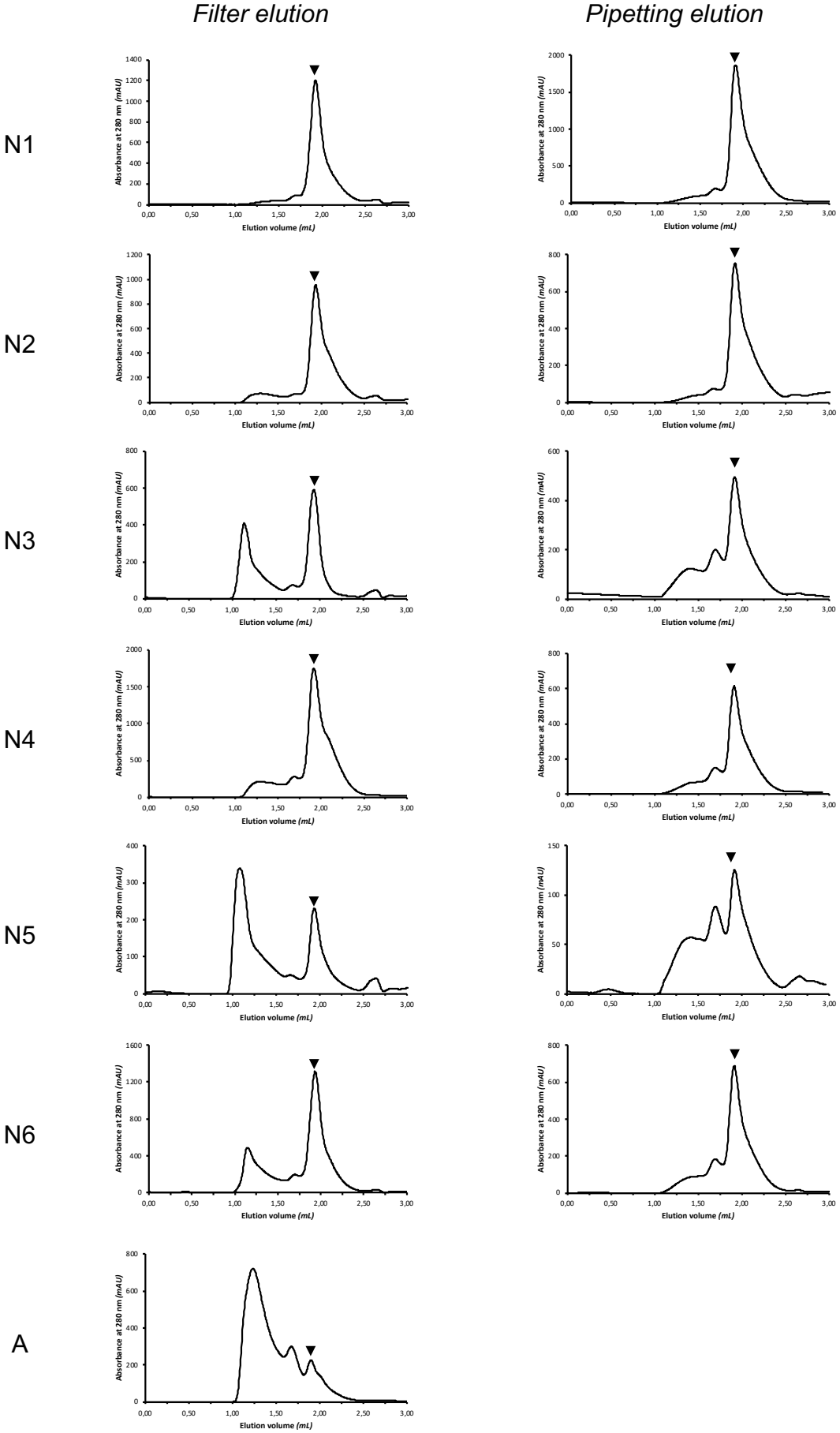
References
1. Waugh DS. The remarkable solubility-enhancing power of *Escherichia coli* maltose-binding protein. Postepy Biochem. 2016;62:377–82.
2. Berrow NS, Büssow K, Coutard B, Diprose J, Ekberg M, Folkers GE, et al. Recombinant protein expression and solubility screening in *Escherichia coli*: a comparative study. Acta Crystallogr D Biol Crystallogr. 2006;62(10):1218–26.
3. Riggs P. Expression and purification of recombinant proteins by fusion to maltose-binding protein. Mol Biotechnol. 2000;15(1):51–63.
4. Riggs P. Expression and purification of maltose-binding protein fusions. In: Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, et al., editors. Current protocols in molecular biology. Hoboken: John Wiley & Sons, Inc.; 2001 **(Chapter 16:Unit16.6)**.
5. Sachdev D, Chirgwin JM. [20] Fusions to maltose-binding protein: control of folding and solubility in protein purification. In: Thorner J, Emr SD, Abelson JN, editors. Applications of chimeric genes and hybrid proteins part A: gene expression and protein purification. Cambridge: Academic Press; 2000. p. 312–21 **(Methods Enzymol; vol. 326)**.
6. Sachdev D, Chirgwin JM. Properties of soluble fusions between mammalian aspartic proteinases and bacterial maltose-binding protein. J Protein Chem. 1999;18(1):127–36.
7. Sachdev D, Chirgwin JM. Order of fusions between bacterial and mammalian proteins can determine solubility in *Escherichia coli*. Biochem Biophys Res Commun. 1998;244(3):933–7.
8. Sachdev D, Chirgwin JM. Solubility of proteins isolated from inclusion bodies is enhanced by fusion to maltose-binding protein or thioredoxin. Protein Expr Purif. 1998;12(1):122–32.
9. Nominé Y, Ristriani T, Laurent C, Lefèvre J-F, Weiss É, Travé G. A strategy for optimizing the monodispersity of fusion proteins: application to purification of recombinant HPV E6 oncoprotein. Protein Eng. 2001;14(4):297–305.
10. Nominé Y, Ristriani T, Laurent C, Lefèvre J-F, Weiss É, Travé G. Formation of soluble inclusion bodies by HPV E6 oncoprotein fused to maltose-binding protein. Protein Expr Purif. 2001;23(1):22–32.
11. Zanier K, Nominé Y, Charbonnier S, Ruhlmann C, Schultz P, Schweizer J, et al. Formation of well-defined soluble aggregates upon fusion to MBP is a generic property of E6 proteins from various human papillomavirus species. Protein Expr Purif. 2007;51(1):59–70.
12. Zanier K, Ruhlmann C, Melin F, Masson M, Ould M'hamed Ould Sidi AO, Bernard X, et al. E6 proteins from diverse papillomaviruses self-associate both in vitro and in vivo. J Mol Biol. 2010;396(1):90–104.
13. Raran-Kurussi S, Waugh DS. The ability to enhance the solubility of its fusion partners is an intrinsic property of maltose-binding protein but their folding is either spontaneous or chaperone-mediated. PLoS ONE. 2012;7(11):e49589.
14. de Marco A. Recombinant polypeptide production in *E. coli*: towards a rational approach to improve the yields of functional proteins. Microb Cell Fact. 2013;12(1):101.
15. de Marco A. Optimization of purification protocols based on the step-by-step monitoring of the protein aggregates in soluble fractions. In: Lorence A, editor. Recombinant gene expression. Totowa: Humana Press; 2012. p. 145–54 **(Methods Mol Biol; vol 824)**.
16. Suarez I, Trave G. Structural insights in multifunctional papillomavirus oncoproteins. Viruses. 2018;10(1):37.
17. Wallace NA, Galloway DA. Novel functions of the human papillomavirus E6 oncoproteins. Annu Rev Virol. 2015;2(1):403–23.

18. Vande Pol SB, Klingelhutz AJ. Papillomavirus E6 oncoproteins. Virology. 2013;445(1–2):115–37.
19. Ould M'hamed Ould Sidi A, Ould Babah K, Brimer N, Nominé Y, Romier C, Kieffer B, et al. Strategies for bacterial expression of protein–peptide complexes: application to solubilization of papillomavirus E6. Protein Expr Purif. 2011;80(1):8–16.
20. White EA, Howley PM. Proteomic approaches to the study of papillomavirus–host interactions. Virology. 2013;435(1):57–69.
21. Scheffner M, Huibregtse JM, Vierstra RD, Howley PM. The HPV-16 E6 and E6-AP complex functions as a ubiquitin–protein ligase in the ubiquitination of p53. Cell. 1993;75(3):495–505.
22. Zanier K, Charbonnier S, Sidi AOMO, McEwen AG, Ferrario MG, Poussin-Courmontagne P, et al. Structural basis for hijacking of cellular LxxLL motifs by papillomavirus E6 oncoproteins. Science. 2013;339(6120):694–8.
23. Martinez-Zapien D, Ruiz FX, Poirson J, Mitschler A, Ramirez J, Forster A, et al. Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53. Nature. 2016;529(7587):541–5.
24. Tommasino M. The human papillomavirus family and its role in carcinogenesis. Semin Cancer Biol. 2014;26:13–21.
25. Bzhalava D, Eklund C, Dillner J. International standardization and classification of human papillomavirus types. Virology. 2015;476:341–4.
26. Van Doorslaer K, Tan Q, Xirasagar S, Bandaru S, Gopalan V, Mohamoud Y, et al. The Papillomavirus Episteme: a central resource for papillomavirus sequence data and analysis. Nucleic Acids Res. 2012;41(D1):D571–8.
27. de Sanjose S, Quint WG, Alemany L, Geraets DT, Klaustermeier JE, Lloveras B, et al. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. Lancet Oncol. 2010;11(11):1048–56.
28. Gillison ML, Alemany L, Snijders PJF, Chaturvedi A, Steinberg BM, Schwartz S, et al. Human papillomavirus and diseases of the upper airway: head and neck cancer and respiratory papillomatosis. Vaccine. 2012;30:F34–54.
29. Tommasino M. The biology of beta human papillomaviruses. Virus Res. 2017;231:128–38.
30. Zanier K, Charbonnier S, Baltzinger M, Nominé Y, Altschuh D, Travé G. Kinetic analysis of the interactions of human papillomavirus E6 oncoproteins with the ubiquitin ligase E6AP using surface plasmon resonance. J Mol Biol. 2005;349(2):401–12.
31. Zanier K, Ould M'hamed Ould Sidi A, Boulade-Ladame C, Rybin V, Chappelle A, Atkinson A, et al. Solution structure analysis of the HPV16 E6 oncoprotein reveals a self-association mechanism required for E6-mediated degradation of p53. Structure. 2012;20(4):604–17.
32. Brimer N, Drews CM, Vande Pol SB. Association of papillomavirus E6 proteins with either MAML1 or E6AP clusters E6 proteins by structure, function, and evolutionary relatedness. PLoS Pathog. 2017;13(12):e1006781.
33. Tayyab S, Qamar S, Islam M. Size exclusion chromatography and size exclusion HPLC of proteins. Biochem Educ. 1991;19(3):149–52.
34. Hubert P, Porath J. Metal chelate affinity chromatography: I. Influence of various parameters on the retention of nucleotides and related compounds. J Chromatogr A. 1980;198(3):247–55.
35. Sundberg L, Porath J. Preparation of adsorbents for biospecific affinity chromatography: I. Attachment of group-containing ligands to insoluble polymers by means of bifunctional oxiranes. J Chromatogr A. 1974;90(1):87–98.

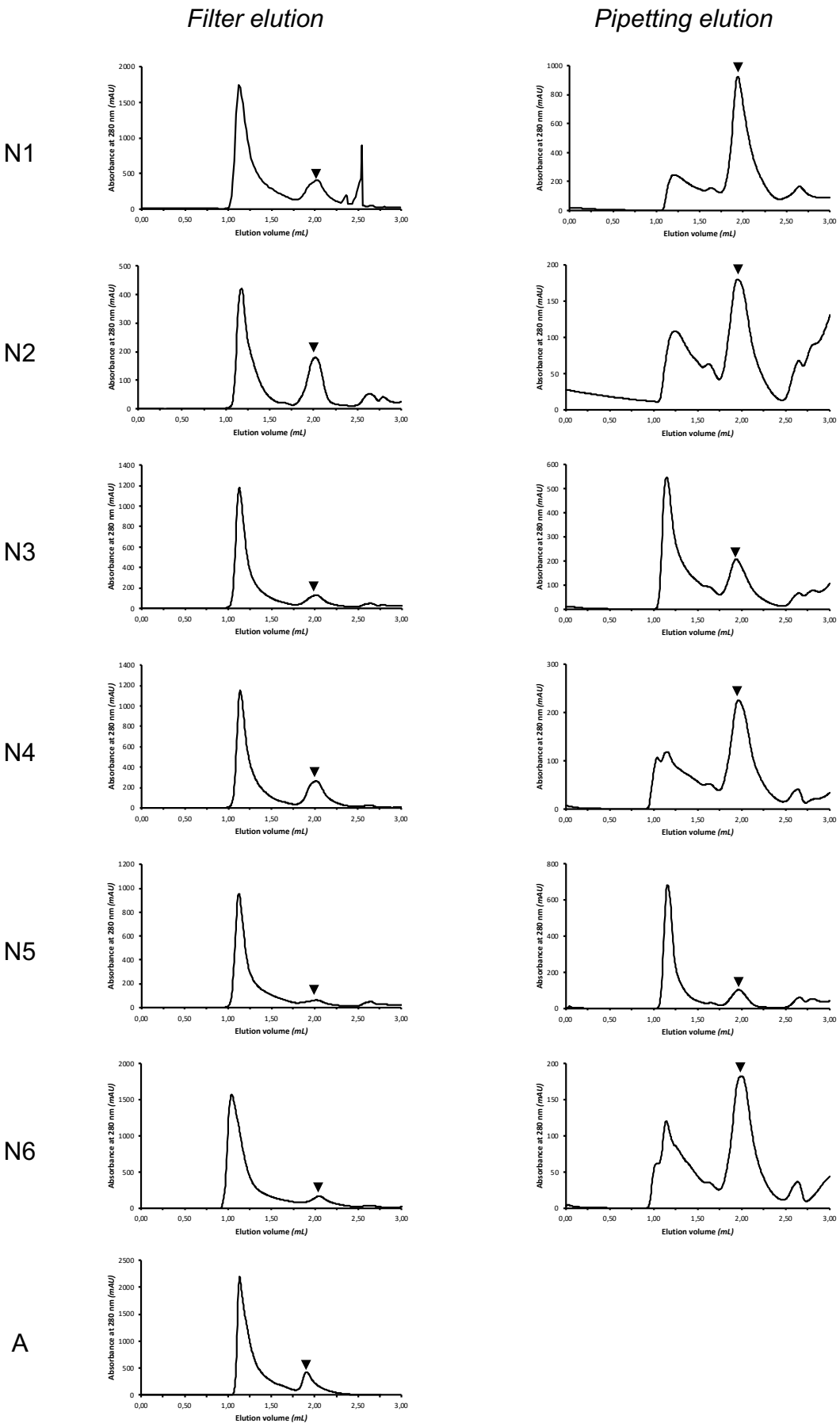
Additional file 1.

Corrected size-exclusion chromatograms. Prior to protein peak integration, spikes due to air bubbles were excluded. **A**: Chromatograms for analytical SEC of MBP-16E6mut. In order to control to oligomeric state of the purified protein, we performed systematic analytical SEC on the final protein sample. The calibration of the column allowed us to estimate the size of the particles at different elution volumes, thus the elution peak of monomeric protein is indicated by an arrow. Note the increase of the monomer fraction compared to the total amount of purified protein when shifting from the filter elution to the “decant and take up” elution method. **B**: Chromatograms for analytical SEC of MBP-8E6. On this protein challenging to purify, the increase of the monomer fraction (indicated by an arrow) between the two elution methods is dramatic. **C**: Chromatograms for analytical SEC of MBP-8E6: optimization of bacterial extract contact time. The elution peak of monomeric protein is indicated by an arrow. The decrease of the oligomeric fraction (eluted between the void volume V0 and the monomer fraction) is particularly visible for nickel resin N1.

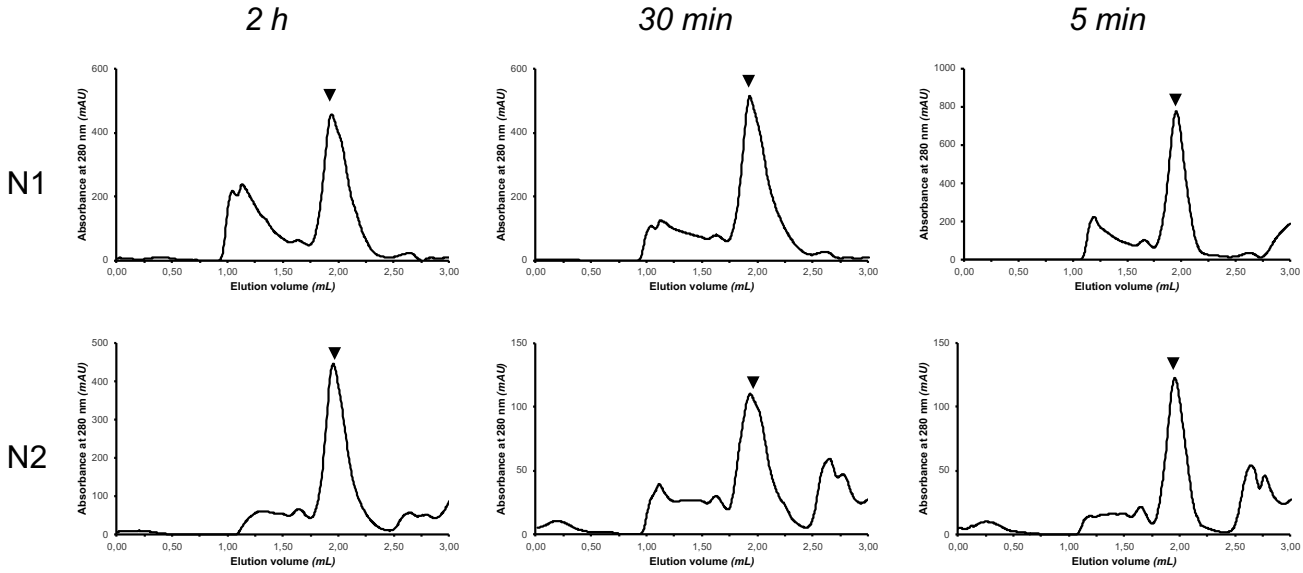
A. Chromatograms for analytical SEC of 16E6mut



B. Chromatograms for analytical SEC of 8E6



C. Chromatograms for analytical SEC of 8E6:
Optimization of bacterial extract contact time



1.2 Benchtop holdup assay for quantitative affinity-based analysis of sequence determinants of protein-motif interactions

1.2.1 State of the art

The study of protein-motif interaction requires an accurate approach allowing sensitive detection of transient interactions and low to medium affinity (in the micromolar range). The quantification of binding affinity is a great advantage for ranking the binding preferences of a given HPV E6 oncoprotein. Hence, the most adapted strategy for comparing E6 interaction preferences is an approach that combines high-throughput and high sensitivity for the estimation of binding affinities.

To meet these needs, the principle of the holdup chromatographic retention assay was published in 2006 (Charbonnier *et al.*, 2006). The assay consists in testing the interaction between an analyte in solution and a biotinylated peptide ligand immobilized on streptavidin resin, as illustrated on **Figure 26**.

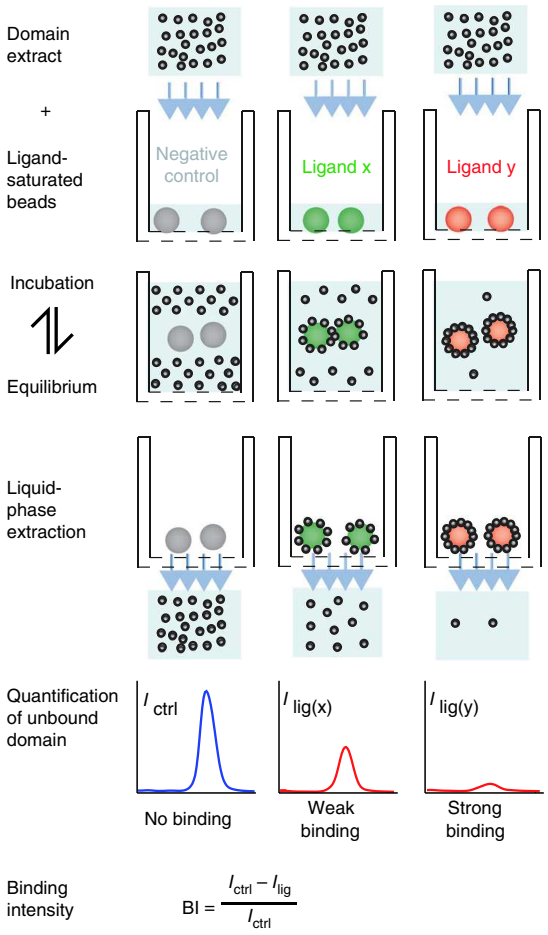


Figure 26: Principle of the holdup comparative chromatographic retention assay. In a filter plate, the domain extract (or purified MBP-E6) is incubated with biotinylated ligand immobilized on beads (in our case, LXXLL peptides). After 15 min incubation time, the liquid fraction is recovered by centrifugal filtration and the unbound protein is quantified by capillary electrophoresis, SDS-PAGE or intrinsic tryptophan fluorescence. The protein depletion as compared to a negative control is represented by a value, called Binding Intensity (BI), which is correlated to the K_D of the interaction. From (Vincentelli *et al.*, 2015).

In a filter plate, the analyte is incubated with either negative control (biotin-saturated resin) or with a peptide ligand. After 15 min incubation time, the filter place is centrifuged for extracting the liquid fraction of all wells. By doing so, the analyte/ligand complex is trapped on the filter with the resin while the liquid fraction contains all analyte molecules that did not interact with the ligand, which are quantified for affinity estimation. In this setup, a high analyte/ligand affinity results in low amounts of analyte in the liquid fraction and conversely, a low affinity results in high amounts of analyte in the liquid fraction. The concentration of analyte after incubation with the ligand is systematically normalized by the amount of analyte in the negative control. The analyte is not expected to interact with biotin-saturated resin, however this negative control includes the loss of analyte due to non-specific interactions with the resin, the plastic

well or the filter. For this reason, the holdup assay is more accurate than the very popular pulldown assay since it discriminates specific from non-specific binding in addition to affinity estimation.

In the proof of concept published in 2006, the quantification of unbound analyte in the liquid fraction was achieved by measuring band intensity on SDS-PAGE analysis (Charbonnier *et al.*, 2006). In a study aiming at estimating the affinities of the 266 existing PDZ domains (the "PDZome") for the PBM of HPV18 and HPV16 E6, the setup and readout of the holdup assay were adapted for a high-throughput study (Vincentelli *et al.*, 2015). The PDZ domains displayed the same binding activity when purified or in bacterial lysates, thus the latter were used for the holdup interaction assay to facilitate the sample preparation. The holdup assay was conducted using pipetting robots for the resuspension and transfer of overexpressed PDZ domains in bacterial extracts. In the final liquid fractions, the unbound analyte was present in a mixture containing bacterial proteins, hence the need of a size-specific protein quantification method. To this aim, capillary electrophoresis was adapted as it allows size separation of up to 384 protein samples in one run. However, apart from this high-throughput project, the use of pipetting robots and capillary electrophoresis is expensive and difficult to implement to another study at lower throughput such as E6-LXXLL interactions. In particular, capillary electrophoresis is time-consuming and not very robust as the capillary often gets clogged and needs extensive maintenance from the user. In addition, E6 proteins must be purified prior to interaction assays for isolating active species. In this case, capillary electrophoresis no longer provides a major advantage since the analyte are pure from any bacterial contaminant.

1.2.2 Objective

In the present publication, we developed a manual benchtop holdup assay that can be easily adapted to a wide range of protein interactions without the need of specific equipment such as pipetting robots and capillary electrophoresis. As a readout for quantifying the unbound analyte, we propose intrinsic tryptophan fluorescence, which is faster and more adapted to purified MBP-E6 proteins. The binding intensities from holdup assay could be converted into ΔG using K_D estimated by SPR. Finally, the holdup assay was compared to SPOT peptide array and displayed higher sensitivity to low binding affinities. This approach was applied to the study of sequence

determinants of HPV16 E6/E6AP LXXLL motif interaction. We compared the interaction of HPV16 E6 within a library of single-point mutants of E6AP LXXLL motif. The resulting interaction data complement structural information and allow a better identification of the sequence determinants within E6AP LXXLL motif.

1.2.3 My contribution

I produced and purified the recombinant protein MBP-16E6 F47R 4C/4S following protocols previously established. I performed SPR, SPOT and holdup assay with capillary electrophoresis and fluorescence readout. I designed and optimized the holdup assay with fluorescence readout by testing and comparing two standard molecules. I processed the data, made the figures, wrote the initial manuscript and made the revisions with my supervisors.

Analytical Biochemistry 603 (2020) 113772

Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

Benchtop holdup assay for quantitative affinity-based analysis of sequence determinants of protein-motif interactions

Anna Bonhoure[a], Anne Forster[b], Khaled Ould Babah[b], Gergő Gógl[a], Pascal Eberling[a], Camille Kostmann[a], Rudolf Volkmer[c], Victor Tapia Mancilla[c], Gilles Travé[a,**], Yves Nominé[a,*]

[a] (Équipe Labellisée Ligue, 2015) Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U1258, CNRS UMR 7104, Université de Strasbourg, Illkirch, France
[b] UMR 7242, Biotechnologie et Signalisation Cellulaire, École Supérieure de Biotechnologie de Strasbourg, Université de Strasbourg, Illkirch, France
[c] Institute of Medical Immunology Charité-Universitätsmedizin, Berlin, Germany

ARTICLE INFO

Keywords:
Protein-protein interaction
Papillomavirus
Oncoprotein
Linear motif
Conservative replacement
Peptide array

ABSTRACT

Many protein-protein interactions are mediated by short linear peptide motifs binding to cognate proteins or protein domains. Such interactions often display affinities in the mid-micromolar range that are challenging to quantify accurately, especially when the motifs harbor single-point mutations. Here, we present a manual benchtop assay for determining affinities of weak interactions between a purified protein and a peptide array representing mutants of a target motif. The assay is based on the "holdup" principle, a chromatographic approach allowing sensitive detection of weak interactions at equilibrium and accurate estimation of their binding free energy. We tested two alternative setups using, as a readout, either capillary electrophoresis or fluorescence. Using this approach, we studied the amino acid sequence determinants of the interactions between HPV16 E6 viral oncoprotein and single-point mutants of its prototypical target LXXLL motif from the E3 ubiquitin ligase E6AP. Comparing SPOT peptide array and holdup approaches revealed a good agreement for most interactions except the weakest ones, which were only detected by holdup assay. In addition, the strongest interactions were validated by Surface-Plasmon Resonance. The manual holdup procedure proposed here can be readily adapted for accurate evaluation of a wide variety of protein-motif interactions displaying low to medium affinities.

1. Introduction

Many protein-protein interactions are determined by the recognition of short linear motifs (SLiMs) by proteins or folded domains [1]. Binding motifs are located within intrinsically disordered regions of proteins and mediate transient, low-affinity interactions in the micromolar range [2]. The involvement of each residue in protein-motif interaction can be deciphered by generating single-point mutants and testing their interaction properties. However, in order to discriminate the binding features of various mutants, a binding assay with high sensitivity, accuracy and reproducibility is required.

A variety of methods can be used to assess protein-protein interactions. Phage display, yeast two-hybrid and protein complementation assays are powerful high-throughput methods. Proteomic peptide phage display allows screening binding peptides in a phage library entailing the disordered regions of the human proteome [3]. Recently, a NanoLuc Two-hybrid assay was reported for high-throughput proteome-scale mapping of protein-protein binding [4]. However, all above-listed techniques have a binary output: "binding or no binding". To complement these qualitative screening approaches, several biophysical technologies allow affinity quantification (e.g. Isothermal Titration Calorimetry (ITC), Surface Plasmon Resonance (SPR), …). Although these approaches can lead to the accurate quantification of dissociations constants, they generally have a low-throughput and require a certain level of know-how for handling specific devices and analyzing the results. Recently, innovative high-throughput methods have been developed for the detection and quantification of protein-protein interactions. A Multireporter Bacterial 2-Hybrid strategy was reported for high-throughput identification of *in vivo* binding partners [5]. The binding partners are identified by deep sequencing and a fluorescence output allows the approximation of their relative binding affinity. A novel technology called MRBLE-pep uses peptides

* Corresponding author.
** Corresponding author. Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U1258, CNRS UMR 7104, Université; de Strasbourg, 1 Rue Laurent Fries, BP 10142, 67404, Illkirch, France.
E-mail addresses: traveg@igbmc.fr (G. Travé), nominey@igbmc.fr (Y. Nominé).

https://doi.org/10.1016/j.ab.2020.113772
Received 2 April 2020; Received in revised form 6 May 2020; Accepted 9 May 2020
Available online 16 May 2020

Abbreviations

BSA	Bovine Serum Albumin
DTT	Dithiothreitol
HPV	Human Papillomavirus
IPTG	Isopropyl β-D-1-thiogalactopyranoside
MBP	Maltose-Binding Protein
SEC	Size-Exclusion Chromatography
SPR	Surface Plasmon Resonance

synthesized on spectrally encoded beads for quantifying the protein binding affinity to a peptide library [6]. Despite the high potency and robustness of these approaches, their use requires specific equipment and expertise, in particular for the peptide synthesis of the MRBLE-pep technology.

To fill the gap between high-throughput qualitative and low-throughput quantitative binding detection methods, we previously developed the holdup comparative chromatographic retention assay, which allows evaluating the affinity of hundreds of domain-motif interactions in a single experiment. This assay involves testing the interaction between an analyte (protein) in solution and a ligand (biotinylated peptide) immobilized on resin (Fig. 1). This approach was originally demonstrated for two viral PDZ-binding motifs tested against a large library of more than 200 PDZ domains [7]. However, in this previous work the holdup assay was performed in an automated setup requiring specific equipment (mainly, a multifunction liquid-handling robot and a microcapillary electrophoresis). Here, we aimed to develop a streamlined holdup protocol that can be performed manually with standard laboratory equipment. Furthermore, we tested whether the assay would be applicable to quantitatively assay the binding of a given protein to a large array of peptides, rather than the binding of a given peptide to a large array of proteins.

Human Papillomaviruses (HPV) are small DNA viruses infecting epithelia, of which a subset causes cancer. Responsible for 61% of cervix [8] and 90% of oropharyngeal cancers [9], HPV type 16 is the most prevalent carcinogenic HPV. Once expressed in the host cell, the viral oncoproteins E6 and E7 trigger viral replication by activating cell proliferation. In particular, E6 protein was reported to hijack several human proteins by interacting with a leucine-rich consensus sequence LXXLL motif [10,11]. Our team published the structure of E6 oncoprotein from HPV16 in complex with the LXXLL motif from the E3 ubiquitin ligase E6AP (E6-Associated Protein) (thereafter named $E6AP_{LXXLL}$) [12]. Once formed, the heterodimer E6/E6AP recruits the tumor suppressor p53 and induces its proteasomal degradation [13–15].

In the present work, we investigated the critical positions within the LXXLL peptide motif that govern the affinity of HPV16 E6/$E6AP_{LXXLL}$ interaction, by means of systematic measurements of the binding affinity of purified HPV16 E6 towards a peptide array of 45 single points mutants of the $E6AP_{LXXLL}$ motif. Since the affinity of the wild-type complex is relatively weak (in the micromolar range) we needed an approach combining high throughput and high sensitivity and accuracy for measuring affinities in the mid-micromolar range. To this aim, we designed a manual holdup assay protocol, which used chemically synthesized biotinylated peptides and freshly purified monomeric HPV16 E6 F47R 4C/4S (thereafter named 16E6). As the proper quantification of the unbound protein fraction is critical in a holdup assay, we present and compare here two options: capillary electrophoresis and fluorescence spectroscopy. This holdup strategy has proven to be reliable as confirmed by cross-validating the results with i) SPOT peptide arrays for all peptide mutants and ii) SPR for the highest affinity peptide mutants. This assay can be adapted to any protein-motif interaction including those involving aggregation-prone proteins, as demonstrated here with HPV E6 oncoprotein.

2. Material and methods

2.1. Protein expression and purification

The solubility-enhanced mutant of HPV16 E6 protein F47R C80S C97S C111S C140S was fused to a mutant of Maltose-Binding Protein (MBP) as previously described for crystallographic purposes [12]. For SPOT assays requiring a His-tag for the detection of MBP-16E6, an MBP-6His-TEVsite-E6 fusion was constructed by cloning the 16E6 sequence into pETM-41 vector by using NcoI and Acc65I restriction sites.

The protein was recombinantly expressed in *Escherichia coli* BL21 (DE3) grown in LB medium supplemented with glucose 0.2% (w/v) and kanamycin 50 μg/mL. Bacteria were grown at 37 °C until $OD_{600} \approx 0.7$, induced by adding 0.5 mM IPTG (Isopropyl β-D-1-thiogalactopyranoside) and $ZnSO_4$ 100 μM, incubated overnight at 16 °C and harvested by centrifugation.

To avoid the formation of intermolecular disulfide bridges, all purification buffers were extensively degassed and equilibrated with argon before adding reducing agent.

16E6 was purified as previously described [12]. Briefly, bacterial pellets were resuspended in Buffer A (Tris HCl 50 mM pH 8.00, NaCl 300 mM and 2 mM DTT) supplemented with glycerol 5% (w/v), RNAse 0.25 μg/mL, DNAse 0.25 μg/mL, lysozyme at approximately 1 μg/mL and EDTA-free protease inhibitor cocktail (Roche) at the recommended concentration. Cells were disrupted by three Microfluidizer cycles and centrifuged 45 min at 150,000 × *g* at 4 °C. For MBP-mediated purification, supernatant was loaded on packed amylose resin (New England BioLabs), washed with buffer A and eluted with buffer B (Tris HCl 50 mM pH 8.00, NaCl 300 mM, DTT 2 mM, maltose 15 mM, protease inhibitor cocktail at 20% of the recommended concentration). To eliminate MBP-16E6 aggregates, the sample was then ultracentrifuged 15 h at 200,000 × *g* at 4 °C in swing rotor. As a final step to isolate monomers, the supernatant was purified by size-exclusion

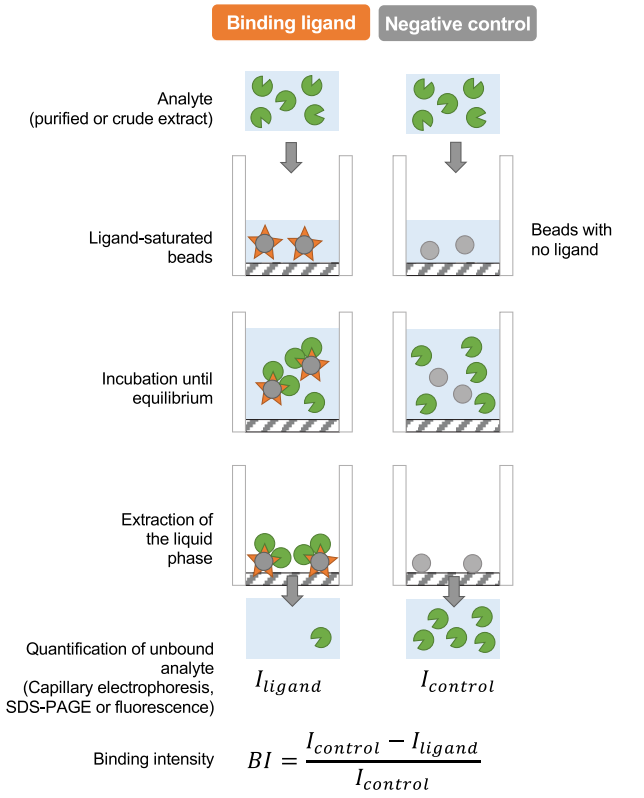


Fig. 1. Schematic representation of the holdup assay principle. Adapted from Refs. [7]. The analyte in solution is incubated with ligand immobilized on beads and then extracted by centrifugal filtration. A quantification of the unbound analyte allows the calculation of a binding intensity (BI), which can be further correlated with dissociation constant.

chromatography in Buffer A on HiLoad 16/600 or 26/600 Superdex 200 columns (GE Healthcare).

2.2. Peptide synthesis

Biotinylated peptides were synthesized by either JPT Innovative Peptide Solutions with > 70% purity or by peptide synthesis service from IGBMC with > 80% purity. They are all N-terminally conjugated to biotin via a TTDS spacer (N-(13-amino-4,7,10-trioxa-tridecayl)-succinamic acid) (Iris Biotech GMBH). The proper resuspension of the lyophilized peptides was ensured by overnight incubation of the lyophilized peptide in water at a final concentration of 5 mM. If particles in suspension were still visible, water was supplemented with ammonia 10% or sodium hydroxide until complete resuspension of the peptide. The resulting solution was stored at −20 °C. Their sequences correspond to the LXXLL motif from E6AP isoform II (uniprot ID: Q05086) residues 402 to 417, with the indicated mutations.

2.3. Holdup assay

2.3.1. Principle

Holdup is a comparative chromatographic retention assay [16] for which a robotized protocol was previously described [7]. In the present work, the assay was performed manually in 96-well plates using common eight-channel micropipettes. The analyte (in our setup, MBP-16E6) was incubated either with ligand-saturated beads (here, biotinylated LXXLL peptides) or with biotin-saturated beads as a negative control. After 15 min incubation of the analyte/beads mixture, the liquid fraction was recovered by centrifugal filtration. The liquid fraction contained the free analyte, which was directly quantified. By comparing the concentration of free analyte after incubation with the peptide ligand versus the negative control, the proportion of ligand-bound analyte was deduced.

2.3.2. Manual holdup assay protocol

For cysteine-rich proteins sensitive to oxidation: All buffers were degassed and extensively equilibrated with argon gas before adding reducing agent.

The experiment was performed at room temperature.

Remark: The temperature can be adjusted to the protein's stability as long as it is correctly indicated for conversion into binding free energy ΔG (see section 2.8 for further information on free energy calculation).

1) Resin dilution and transfer in the filter plate: 3.7 mL (50% slurry suspension in ethanol 20%) streptavidin Sepharose High-Performance beads (17-5113-01, GE Healthcare) were equilibrated in 20.3 mL buffer A. The bead suspension was homogenized by up and down pipetting with a multichannel pipette and dispensed in a 96-well filter plate (MSDVN65, Millipore). For each well, 200 µL suspension (corresponding to 15 µL beads) were transferred.
   Critical step: the resin should be properly homogenized before being distributed on the plate. Comparative analysis is only possible when an equivalent amount of resin is present in each well of the plate.
2) Resin equilibration in the filter plate: the liquid fraction was removed by Vacuum Manifold (NucleoVac 96, Macherey-Nagel). The resin was washed twice with 200 µL buffer A per well.
3) Peptide dilution: Biotinylated peptides and biotin were diluted at 62.5 µM in Buffer A in a V-shape 96-well plate.
4) Biotinylated peptide-streptavidin conjugation: 80 µL ligand (in our case, biotinylated peptide) or negative control (in our case, biotin, B4501, Sigma-Aldrich) at 62.5 µM were incubated with streptavidin beads for 15 min under 1,200 rpm agitation (IKA MS3 digital plate shaker). The liquid fraction was discarded from the filter plate by vacuum filtration.
   Remark: According to the manufacturer's specifications, these conditions should allow a maximal occupancy of streptavidin beads binding sites.
5) Saturation of streptavidin sites by an excess of biotin: beads were subsequently incubated with 1 mM biotin (80 µL per well) for 10 min under 1,200 rpm agitation.
6) Washing steps prior to analyte-ligand incubation: The beads in the filter plate were washed thrice (200 µL buffer A per well) with vacuum removal of the liquid phase between each washing step. The liquid from the last wash was removed by gentle centrifugation for 2 min at 200 × *g*. This step allows better efficiency of liquid phase elimination for all wells before adding the analyte.
7) Preparation of analyte mixture with internal standards:
   The analyte (MBP-16E6 in the present study) was diluted at 4 µM in buffer A with the internal standards, indicated as follows:
   a) For capillary electrophoresis readout: two internal markers of distinct molecular weight were added to the analyte solution, namely lysozyme (5933-B, Euromedex) at 0.05 mg/mL, and BSA (A0281, Sigma-Aldrich) at 4 µM.
   b) For fluorescence readout: we chose fluorescein and mCherry at 50 nM, whose absorption and emission spectra do not overlap with tryptophan spectra.
   Critical step: The mixture analyte-internal standards should be mixed by inverting the tube several times for reproducible distribution of the protein standards throughout the plate.
8) Analyte-ligand incubation: 30 µL of analyte mixture were added to each well of the filter plate, corresponding to twice the volume of beads. The plate was shaken at 1,200 rpm for 15 min during the ligand-analyte incubation step, allowing to reach the equilibrium of protein-peptide interaction.
   Critical step: the volume of analyte mixture as well as the time of incubation under stirring should be precisely respected for reproducible study of complex at equilibrium.
9) Collection and storage of liquid fraction after the interaction assay:
   The liquid fraction was collected in 96-well PCR plate by centrifugal filtration. Samples dedicated to capillary gel electrophoresis were prepared according to the manufacturer's recommendations. The remaining liquid fraction was blocked with SDS-PAGE sample buffer and boiled for 5 min at 95 °C before storage at −20 °C. Samples dedicated to fluorescence were immediately measured.
   Remark: in case of limited amount of analyte or ligand, the assay can be performed in 384-well plate for lower volumes and equal accuracy. In the 384-well setup, 14 µL of biotinylated peptide are immobilized on 2.5 µL resin per well and 5 µL analyte mixture are required for the interaction assay.

2.4. Microfluidics capillary gel electrophoresis for quantitative analysis

2.4.1. Data acquisition

The device LabChip GX II (Caliper, PerkinElmer) was used for High-Throughput Protein Express 100 High Sensitivity assay, for sizing of proteins in the 14–100 kDa range. The samples were mixed with sample buffer and denatured for 5 min at 95 °C. Prior to the injection in the microcapillary, the proteins are labelled with a fluorescent dye. During capillary electrophoresis course, a laser excites the dye attached to proteins. The resulting electropherogram is the signal of emitted fluorescence plotted versus migration time. By calibration with standard markers, the LabChip GX II software converts migration time into molecular weight. The emitted fluorescence signal being proportional to protein concentration, this analysis allows accurate quantification of proteins separated by size.

2.4.2. Curation of electropherograms

This step was performed on LabChip GX II software. Inappropriate molecular weight markers were corrected or excluded prior to samples alignment. Each electropherogram was visually inspected and invalid measurements were excluded according to the three following criteria. First, the dataset was rejected if the molecular weight markers provided by the manufacturer were not detected since their absence would impede the conversion of migration time into molecular weight. Second, the height of the peaks of interest (analyte or internal standards) should be above 50 AU. Below this threshold, we observed a higher variability of intensity, which may indicate that the capillary electrophoresis was reaching its lower

quantification limit due to signal uncertainty. Finally, some electropherograms contained spikes that were due to impurities present in the capillary. Those spikes could impede proper selection of the peaks of interest if they migrated in the same size range and/or conduct to incorrect estimations of peaks of interest, which led us to exclude the electropherograms containing spikes overlapped with the peaks of interest. Our samples contained three proteins: purified MBP-16E6, with lysozyme and BSA as internal markers for holdup normalization. The peaks corresponding to the three proteins were selected by using the following non-overlapping molecular weight windows: 50–70 kDa for MBP-16E6, 10–20 kDa for lysozyme and 70–90 kDa for BSA. For each sample, the height of the three peaks was calculated by LabChip GX II software and exported in csv table for the calculation of binding intensity, as detailed in section 2.6.

2.5. Detection of MBP-16E6 and standard proteins by fluorescence

Fluorescence was measured on Pherastar^{PLUS} microplate reader from BMG Labtech, using filter sets consisting of an excitation and emission bandpass filters with 485 ± 20 nm and 520 ± 20 nm cut-offs for the detection of fluorescein, 575 ± 20 nm and 620 ± 20 nm cut-offs for mCherry and 295 ± 20 nm and 350 ± 20 nm cut-offs for tryptophan. 10 µL of final holdup samples were pipetted twice in 384-well Greiner black plate (reference 784 900) for duplicated measurements.

2.6. Binding intensity calculation

For each sample, a normalized analyte intensity ratio R was calculated by comparing the intensity of the analyte (in our case, MBP-16E6) with the intensity of one of the markers (lysozyme, BSA, mCherry of fluorescein) present in the same well. R is defined as the ratio of analyte over standard intensities I (**Equation 1**).

$$R = \frac{I_{analyte}}{I_{standard}} \quad \text{[Equation 1]}$$

In holdup assay, the analyte is incubated in presence of resin saturated with either a ligand (here, biotinylated LXXLL peptide) or the negative control (biotin). R from wells containing biotin-saturated resin is thereafter named R_{biotin}. Similarly, R from wells containing resin saturated with biotinylated LXXLL peptides is thereafter named R_{LXXLL}.

In our setup using 96-well plate, biotin-saturated resin has been deposited in 6 to 8 wells randomly distributed over the whole plate. To identify and exclude outliers among these 6 to 8 negative control values, we performed a modified Thompson-Tau test on R_{biotin}.

After outlier exclusion, the mean of the remaining R_{biotin} values ($\overline{R_{biotin}}$) was used for the calculation of the binding intensity (*BI*) of each analyte-ligand pair according to **Equation 2**.

$$BI = 1 - {R_{LXXLL}}/{\overline{R_{biotin}}} \quad \text{[Equation 2]}$$

In order to check for consistency, BI values were calculated for each of the two markers corresponding to the quantification method: lysozyme and BSA for capillary electrophoresis, fluorescein and mCherry for fluorescence.

Final holdup results entail the mean and standard deviation of BI from duplicated experiments, performed with independent preparations of purified MBP-16E6.

2.7. Surface plasmon resonance

2.7.1. Experimental setup

SPR interaction assays were performed on a Biacore T200 instrument (GE Healthcare - Biacore) at 25 °C. We used as running buffer Tris HCl pH 8.00 50 mM, NaCl 300 mM, DTT 5 mM, surfactant polysorbate 20 (GE Healthcare) 0.005% (v/v). The biotin CAPture kit was used to immobilize the biotinylated LXXLL peptides in a reversible manner. Each cycle started by injecting CAPture reagent (an oligonucleotide coupled with streptavidin which binds to the CAPchip surface by complementarity with the oligonucleotide coating the surface) diluted five times in running buffer on all channels for 300 s at 2 µL/min. In each cycle, one channel (harboring only CAPture reagent and no biotinylated peptide) was defined as the reference flow cell to serve as a control for non-specific binding of the analyte. Biotinylated LXXLL peptides to be tested were immobilized by injecting a 50 nM solution on the remaining channels for 15–30 s at 10 µL/min. MBP-16E6 analyte was then injected on the four channels for 60 s at a flow rate of 30 µL/min. The post-injection phase was recorded for 120 additional seconds. At the end of each cycle, the surface was regenerated by injecting a 6 M guanidine hydrochloride solution supplemented with 250 mM sodium hydroxide for 60 s at 5 µL/min. This regeneration step allows to fully remove MBP-16E6, biotinylated peptide and CAPture reagent from the surface by dehybridizing the surface oligonucleotide from the CAPture reagent oligonucleotide.

Kinetic measurements were achieved by injecting a series of two-fold dilutions of MBP-16E6, starting from 5 µM. The assays involving the peptides E6AP$_{LXXLL}$ wt and E15R were performed in triplicates, those with E10A in duplicates and Q9A in singlicate. The uncertainties were estimated according to standard deviations.

2.7.2. Data processing

Data were interpreted as previously described [17], using first Biacore T200 Evaluation 1.0 software (GE Healthcare, Biacore Life Science, Uppsala, Sweden) and an in-house Python script for steady-state analysis assuming a simple 1:1 interaction binding isotherm model. The uncertainties correspond to the standard deviations between the replicates.

2.8. Estimation of K_D and ΔG from holdup binding intensities

Although the *BI* values contain information about the strength of the binding, a conversion needs to be undertaken in order to estimate the dissociation constant, K_D. This conversion was achieved using **Equation 3** that was previously described [18]:

$$K_D = \frac{[E6_{free}][LXXLL_{free}]}{[E6-LXXLL_{complex}]}$$

$$K_D = \frac{([E6_{total}] - BI \times [E6_{total}]) \times ([LXXLL_{total}] - BI \times [E6_{total}])}{BI \times [E6_{total}]} \quad \text{[Equation 3]}$$

This equation shows that an estimation of the total peptide concentration is required even when the MBP-16E6 concentration was set up at a fixed value of 4 µM. Assuming that the binding strengths are the same for a given interaction observed by SPR or holdup assays, the peptide concentration was fitted with **Equation 3** using experimental binding constants independently obtained by SPR for the interaction of MBP-16E6 with E6AP$_{LXXLL}$ wild-type and the mutants F8 and R15 (5 independent measurements). The estimated mean peptide concentration was 11.4 ± 2.4 µM. We finally assumed that all biotinylated single-point mutant E6AP$_{LXXLL}$ peptides were at the same concentration on the streptavidin resin, and used a peptide concentration of 11.4 µM for K_D extrapolation from holdup data. As reported previously [7], the repetition of holdup experiments obtained for an irrelevant "neutral" peptide having no specific interaction with the analyte had led to BI values that are almost all below 0.10 (98% of all wells) and with a standard deviation of less than 0.10 (considering 95% of the data). According to this, we applied a conservative safety factor of 2 that leads to BI = 0.20, which represents a very stringent threshold to retain only high-confidence binding event [7,18,19]. In practice, we converted all BI < 0.20 to the fixed value of 0.00 prior to K_D extrapolation. ΔG, and ΔΔG for relative analysis of mutant E6AP$_{LXXLL}$ as compared to wild-type (wt), were further calculated from K_D according to **Equation 4** and **Equation 5** below. Since the SPR measurements were performed at 25 °C and the holdup assay at room temperature, we applied a temperature of T = 298 K in the following equations.

$$\Delta G = RT \times \ln(K_D) \quad \text{[Equation 4]}$$

$$\Delta\Delta G = \Delta G_{mutant} - \Delta G_{wt} \quad \text{[Equation 5]}$$

The uncertainties were obtained by quadratic uncertainty

propagation assuming independent variables and using variance formula [20].

2.9. SPOT peptide array

2.9.1. SPOT synthesis

Peptide synthesis on nitrocellulose membrane was performed following a standard protocol previously described [21,22].

2.9.2. Interaction assay with SPOT peptide array

Nitrocellulose membranes with synthesized peptides were first activated in ethanol 100% for 10 min at room temperature under agitation. They were subsequently washed three times with TBS (Tris Buffered Saline: Tris HCl pH 8.00 50 mM, NaCl 150 mM) with 10 min incubation at room temperature under agitation for each washing step. The membranes were then saturated by casein blocking buffer (Sigma, B6429) diluted in TBS to the final recommended concentration according to the manufacturer. For saturation, membranes were incubated 3 h at 4 °C under agitation. They were then washed three times with TBS (5 min per washing step). A solution of purified 6His-MBP-16E6 was prepared as follows: 6His-MBP-16E6 20–40 μg/mL, maltose 10 mM, DTT 5 mM, TBS and casein following the manufacturer's recommendations. As for purification buffers, the buffer was equilibrated with argon to avoid intermolecular disulfide bridges. Maltose was added to the solution to allow specific interaction of MBP-16E6 with the spotted peptides by binding to MBP. The SPOT membrane was incubated with MBP-16E6 solution for 18 h at 4 °C under agitation. The membrane was then washed with TBS (3 times 4 min). For the detection of 6His-MBP-16E6, we used as a primary antibody, a monoclonal *anti*-polyhistidine antibody produced in mouse (Sigma, H1029) diluted in casein-TBS buffer at 1 μg/mL and incubated for 90 min at 4 °C under agitation. The membrane was washed with TBS (3 times 4 min). The secondary antibody used was an Anti-Mouse IgG with peroxidase activity (Sigma, A5906) diluted in casein-TBS buffer at 1 μg/mL and incubated for 1 h at 4 °C under agitation. The membrane was washed with TBS (3 times 4 min). A chemiluminescent substrate (Thermo Fisher Scientific, 34080) with 3 min reaction time at room temperature under agitation was used for revelation. The detection was performed with ImageQuant LAS 4000 (GE Healthcare) with exposure time between 1 and 20 s.

2.9.3. In-house python script for extraction of SPOT intensities

Image files previously saved into PNG format are read using the *imread* function of Python. In-house scripts are used to convert the image in grey scale and to integrate the SPOT intensity over a circle for each spot and for the background. The intensities were further normalized to values between 0 and 1 to facilitate the comparison with holdup BI. SPOT intensities equal to the background were set to zero while the highest value of the membrane was set to the highest BI for the same peptide array for easier comparison.

2.10. Structure visualization

The crystal structure of 16E6 in complex with E6AP-LXXLL wild-type (sequence ELTLQELLGEER) was previously reported by our team [12] and deposited on PDB with ID 4GIZ. The surface representation in the present study was generated using the UCSF Chimera package from the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH P41 RR-01081) [23,24]. Other representations were generated using the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

3. Results

3.1. The streamlined holdup assay

3.1.1. General principle

The holdup assay consists of probing the interaction between an analyte in solution (protein) and a ligand immobilized on resin (biotinylated peptide), as shown on Fig. 1. On a filter plate, the analyte is incubated in parallel with two distinct resin batches, one bearing the putative ligand and the other one a negative control molecule. The concentration of resin-bound ligand should largely exceed that of analyte (typically, at least 20 μM of ligand peptide for 4 μM of protein). The liquid phase containing the free analyte that did not interact with the ligand, is extracted by centrifugation while the analyte-ligand complex is trapped with the resin on the filter. The protein in the liquid fraction is quantified. We distinguish the amount of analyte after incubation with negative control, from the remaining amount of analyte after incubation with the ligand. The amount of analyte-ligand complex can be deducted by comparing these two measures. A value called binding intensity (thereafter called BI) can then be determined. In principle, BI values range from 0.00 (no detected interaction) to 1.00 (100% of analyte bound to ligand). Therefore, the BI value reflects the strength of the interaction and is correlated with the affinity.

The forthcoming sections describe the different aspects of the holdup assay, including the ligand, the analyte, the quantification of unbound analyte and the BI calculation.

3.1.2. The ligand. Design of biotinylated peptides corresponding to single-point mutants of $E6AP_{LXXLL}$

The immobilization system should i) allow oriented immobilization for homogeneous presentation of the ligand and ii) prevent the release of the ligand in solution for proper affinity estimation. Any leakage of the immobilized molecule from the resin to the final liquid fraction would distort affinity estimation, since a fraction of the bound analyte would not be retained on resin either. To fulfill these requirements, we used biotinylated peptides immobilized on streptavidin resin as the probed ligands while biotin-saturated resin was used as negative control. For a 96-well plate assay, we included 6 to 8 negative control wells placed at random positions on the plate, as shown on Appendices (Fig. A1). A polyethylene glycol derivative TTDS (1,13-diamino-4,7,10-trioxatridecan-succinicacid) serves as a linker between the biotin and the peptide for better accessibility of the ligand.

In the present study, we investigated the amino acid sequence determinants of the interaction between 16E6 protein and E6AP-LXXLL motif. The published structure of 16E6 in complex with E6AP-LXXLL motif [12] comprised a 12-residue long peptide (sequence: ELTLQELLGEER). The structure showed some contacts between 16E6 and the alanine linker preceding the LXXLL motif from E6AP, suggesting that amino acids located on the N-term of E6AP motif may participate in the interaction with 16E6. We based our present study on a 16-meric E6AP-LXXLL motif encompassing residues 402 to 417 in E6AP isoform II (sequence: $P^1E^2S^3S^4E^5L^6T^7L^8Q^9E^{10}L^{11}L^{12}G^{13}E^{14}E^{15}R^{16}$), thereafter called $E6AP_{LXXLL}$. Every designed mutant was bearing one single substitution as compared to the wild-type motif. The key residue E5 and residues T7 to E15 were replaced by alanine or amino acids with similar biochemical properties, allowing us to investigate the impact of subtle changes on the interaction and to estimate the contribution of each residue.

For one holdup assay, the required amount of each peptide is 6 nmol, corresponding for example to 14 μg for the biotin-TTDS linker-$E6AP_{LXXLL}$ (molecular weight = 2,360 Da).

3.1.3. The analyte. Purification of biologically active HPV16 E6

In the holdup approach originally reported, soluble crude extracts of MBP-fused PDZ (MBP-PDZ) domains were used as analytes [7]. In this former study, the same binding activity was observed for purified MBP-PDZ and bacterial extracts of overexpressed MBP-PDZ.

With difficult proteins prone to aggregation or other phenomena altering their activity in solution, the use of a purified analyte allows a better control of activity and reproducibility. Here, experiments have been performed using purified MBP-fused HPV 16E6 oncoprotein (MBP-16E6). HPV16 E6 protein is composed of two zinc-binding domains, named E6N and E6C. In addition to the 8 highly conserved cysteine residues involved in zinc ion coordination, this protein contains 6

non-conserved cysteine amino acids. When overexpressed in bacteria and released in oxidative conditions for cell lysis, the major fraction of the produced cysteine-rich MBP-16E6 accumulates as inactive soluble oligomers due to artifactual intermolecular disulfide bridges [25]. Using a non-purified cleared lysate of MBP-16E6 overexpression would have strongly altered the holdup results since up to 90% of the MBP-16E6 sample would be inactive in the assay. Thus, holdup assay was performed using isolated monomeric MBP-16E6 protein ensuring optimal analyte quality.

For this purpose, we used the solubility-enhanced mutant HPV16 E6 F47R 4C/4S for which the crystal structure HPV16 E6/E6AP$_{LXXLL}$ has been published [12]. In this construct, four non-conserved cysteines have been mutated to serine to prevent intermolecular disulfide bridges, while the phenylalanine 47 has been mutated to arginine to prevent dimerization of the E6N domain.

In addition, a particular purification protocol involving the elimination of inactive species for the isolation of active MBP-16E6 monomers was applied [12]. This purification strategy enables the elimination of residual oligomers by affinity chromatography on amylose resin, overnight ultracentrifugation and size-exclusion chromatography (SEC). In order to limit protein re-association over time, holdup assay was performed immediately after purification.

3.1.4. The quantification of the unbound analyte. Two possible protein quantification methods: capillary electrophoresis or tryptophan fluorescence

3.1.4.1. Size-specific protein quantification by capillary electrophoresis. In the former publication reporting the holdup assay [7], the final holdup samples containing the unbound analyte to quantify were soluble crude extracts of MBP-fused PDZ domains overexpressed in bacteria. Such extracts contained, in addition to the overexpressed domain, many bacterial proteins. Therefore, capillary electrophoresis was used to separate proteins according to their migration time in a high-throughput, accurate and reproducible fashion. The readout is based on fluorescence emission of labels that have been chemically added to the denatured proteins prior to migration, indicating that peak intensities are proportional to protein concentrations. The time-to-mass conversion results in plots of protein molecular weight versus fluorescence intensity named electropherograms. Within the electropherograms, we had to identify and to quantify the particular peak corresponding to the overexpressed domain. Depending on the overexpression level, on the quality of the capillary electrophoresis runs and on the amount of data to treat, this task can become arduous and computationally challenging [Jané et al., *Meth Mol Biol, 2020, in press*].

This difficulty to process electropherograms is bypassed when using a purified sample since it does not contain any bacterial protein background. Fig. 2 shows three typical examples of electropherograms recorded with purified protein samples. On each graph, one can observe the peaks of the three proteins present in the holdup sample: MBP-16E6 and the two internal standard proteins, lysozyme and BSA. MBP-16E6, lysozyme and BSA migrate at 64.0 ± 1.3 kDa (theoretical molecular weight MW$_{theo}$: 62.5 kDa), 15.7 ± 0.3 kDa (MW$_{theo}$: 14.3 kDa) and 77.9 ± 1.5 kDa (MW$_{theo}$: 66.3 kDa), respectively. The intense peak on the left of lysozyme is a lower molecular weight marker from the manufacturer's kit, used to detect the beginning of the sample migration. The three panels in Fig. 2 display holdup sample electropherograms after incubation with an E6AP$_{LXXLL}$ binding ligand, plotted with the biotin negative control for better visualization of MBP-16E6 depletion. On the right panel (Fig. 2A), MBP-16E6 was incubated with E6AP$_{LXXLL}$ wild-type. The strong depletion of MBP-16E6 in the "ligand" sample as compared to the "negative control" sample indicates that the majority of MBP-16E6 was retained on the filter with E6AP$_{LXXLL}$ wild-type. On the opposite, only a very small fraction of MBP-16E6 was depleted after incubation with E6AP$_{LXXLL}$ V8 mutant (Fig. 2B), indicating that there was no interaction between MBP-16E6 and E6AP$_{LXXLL}$ V8 peptide. For each of the two panels, the rather good superimposition of the two electropherograms reflects the absence of migration issues and in particular that equal volumes of the two samples were loaded on the capillary electrophoresis instrument. In addition, the fact that the intensities of the lysozyme or BSA standard protein peaks are similar for "ligand" and "negative control" samples rules out any specific interaction between lysozyme or BSA with E6AP$_{LXXLL}$ as compared to resin saturated with biotin. On Fig. 2C, the superimposition of biotin and L11A samples is not as clear as in the previous examples due to shifted profiles and possibly volume variation. In this case, the internal standard proteins are determinant for proper analyte quantification. Indeed, after normalization, the resulting BI is 0.05 ± 0.03. This result indicates that there was no interaction between MBP-16E6 and the L11A mutant of E6AP$_{LXXLL}$ although this was not clearly visible on the electropherograms.

3.1.4.2. Total protein quantification by intrinsic tryptophan fluorescence. While microfluidic capillary electrophoresis is a powerful and reproducible analytical approach, it requires certain preparation time and maintenance. For measuring 96 samples, the

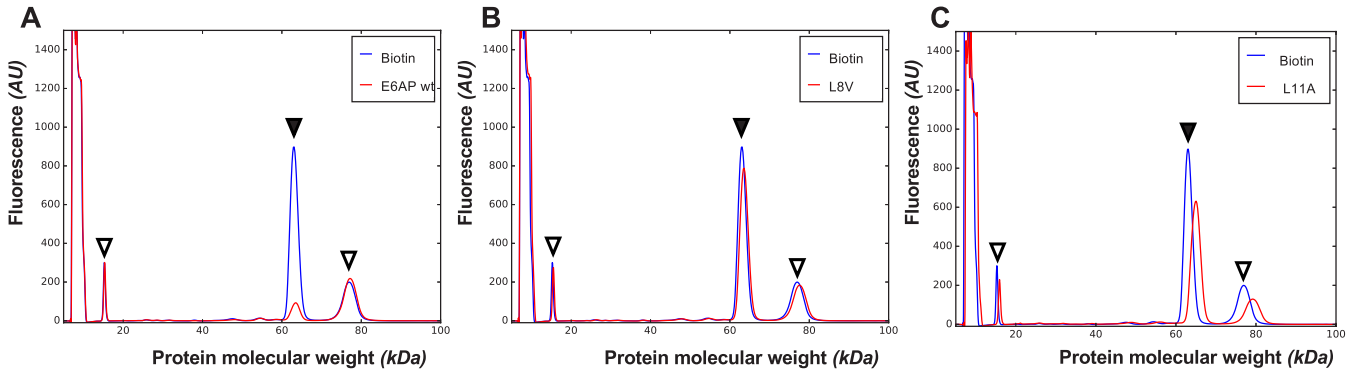


Fig. 2. Electropherograms of holdup samples of MBP-16E6 after incubation with E6AP$_{LXXLL}$. The final samples from holdup were analyzed by capillary electrophoresis for quantification of the bound fraction of analyte (MBP-16E6). The proteins were labelled with fluorescent dye prior to electrophoresis using manufacturer's kit. The fluorescence intensity shown in the *y*-axis is proportional to the protein amount. During electrophoresis, the proteins are separated according to their molecular weight, shown on the *x*-axis. Each holdup sample is a mixture of analyte (MBP-16E6, indicated by a black arrow) and standards (lysozyme and BSA, indicated by white arrows). The analyte-standard mixture was incubated either with the negative control (biotin, represented with blue line) or with biotinylated E6AP$_{LXXLL}$ peptide ligand (represented by the red line). **A**. The tested ligand is E6AP$_{LXXLL}$ wild-type (sequence $P^1E^2S^3S^4E^5L^6T^7L^8Q^9E^{10}L^{11}L^{12}G^{13}E^{14}E^{15}R^{16}$). Note the strong depletion of MBP-16E6, indicating the high affinity of the interaction. **B**. The tested ligand is the L8V mutant of E6AP$_{LXXLL}$. MBP-16E6 is almost not depleted with E6AP V8 as compared with biotin, reflecting the absence of detectable interaction. **C**. The tested ligand is the L11A mutant of E6AP$_{LXXLL}$. As for L8V mutant, there was not interaction with MBP-16E6. Note the migration shift and the variations of protein peak heights between the tested ligand and the negative control, showing the importance of internal controls for estimating analyte depletion.

time needed is 30 min for preparing the chip and reagents, and 1 h 30 for the measurement. In addition, the capillary often gets clogged or may contain bubbles impairing migration, thus requesting the user to prime or prepare fresh reagents several times per day and carefully maintain the chip and the device. This generates expensive maintenance or repair costs which add up to the price of the chips and reagents cited before. Therefore, if the purified protein construct contains tryptophan residues (which is the case for any MBP-fused construct), another strategy for quantifying the unbound analyte is to use intrinsic tryptophan fluorescence measurement. 16E6 contains only one tryptophan residue, however MBP carrier protein contains 8 tryptophan residues which significantly increase the extinction coefficient of the MBP-16E6 fusion protein. Fluorescence spectroscopy of tryptophan is a fast and sensitive method for quantifying a purified protein analyte in holdup samples. It is less time-consuming and about 60-fold cheaper than capillary electrophoresis. In addition, the fluorescence approach has the obvious advantage of delivering, for each sample, two values (the fluorescence intensities of the protein tryptophan and of the internal controls at distinct wavelengths) while the capillary electrophoresis delivers a full electropherogram including numerous peaks that have to be analyzed. However, since the fluorescence method allows the quantification of all proteins present in a sample, great care must be taken when purifying the protein analyte to minimize the presence of protein contaminants.

We tested two internal markers for the correction of volume variation: fluorescein and mCherry. Their fluorescence absorption and emission wavelengths (Fluorescein: $\lambda_{absorption}$ = 485 nm; $\lambda_{emission}$ = 520 nm and mCherry: $\lambda_{absorption}$ = 575 nm; $\lambda_{emission}$ = 620 nm) do not overlap with those of tryptophan ($\lambda_{absorption}$ = 280 nm; $\lambda_{emission}$ = 350 nm). For the readout by fluorescence, we systematically include control wells for estimating the background signal. Working buffer is incubated in the filter plate with biotin-saturated resin and the flow-through is measured at the wavelengths of tryptophan and internal standards. The resulting background is subtracted to all fluorescence values of the corresponding wavelength prior to further processing.

3.1.4.3. The internal markers for accurate analyte quantification. During the liquid phase extraction, some volume variations between the different wells can be observed. In order to normalize the amount of analyte, we systematically included one, if not two, internal standard molecules that did not interact with the peptide nor with the negative control. Thus, the concentrations of the standards are supposed to be constant regardless of the presence of the ligand or analyte. The chosen standards were adapted to the quantification method of the analyte. For MBP-E6 quantification by capillary electrophoresis or SDS-PAGE, we recommend lysozyme and BSA as internal standard proteins, because they have very distinct molecular weights from MBP-16E6 and their peaks flank the MBP-16E6 peak on the electropherogram. They allow double normalization in *x* and *y* axes, which is more reliable in case of important loaded volume variation and/or time-to-molecular weight conversion. For quantitative analysis, the heights of both standards and analyte were systematically extracted and processed for binding intensity calculation, as further detailed in section 3.1.5. When MBP-16E6 was quantified by measuring its tryptophan fluorescence, fluorescein and mCherry were used as standards. The amount of internal standard is expected to be independent from the amount of MBP-16E6. In our setup, lysozyme, BSA, fluorescein and mCherry fluorescence signals were in a constant range for different MBP-E6 intensities, with a mean fluorescence intensity of 241 ± 39 AU, 141 ± 25 AU, 39 500 ± 1400 AU and 5900 ± 280 AU, respectively (Fig. 3). The amounts of lysozyme, BSA and fluorescein in the flow-through seemed completely independent of the amount of analyte (Fig. 3A–C). Their slight variabilities can be attributed to variations of volume and do not affect the quality of the normalization. On the opposite, the amount of mCherry seems to slightly increase with the

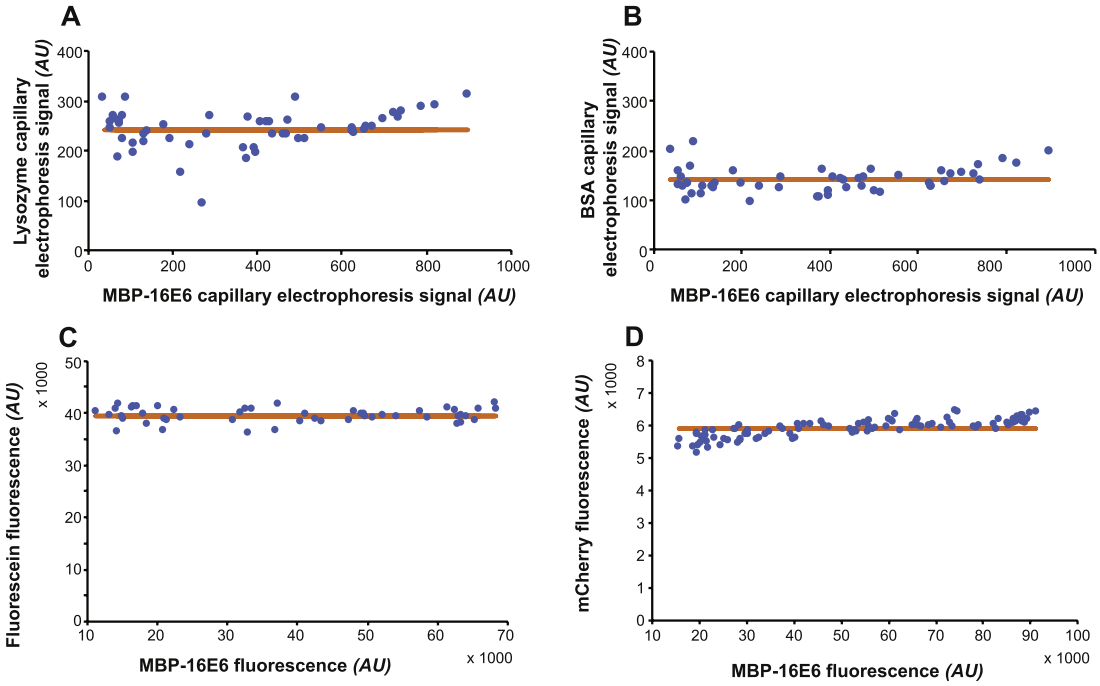


Fig. 3. Fluorescence intensity of the analyte (MBP-16E6) versus the internal markers (lysozyme, BSA, fluorescein, mCherry). The choice of an internal marker for the holdup assay is based on its ability to flow through the resin independently from the analyte. Thus, the concentration of internal marker is expected to be constant at any analyte concentration. On this figure, MBP-16E6 intensity was plotted versus the intensities of lysozyme (**A**), BSA (**B**), fluorescein (**C**) and mCherry (**D**). Lysozyme and BSA are the internal markers used for quantification by capillary electrophoresis while fluorescein and mCherry were tested as internal markers for quantification by fluorescence. The blue dots are the experimental data points from holdup assays. On each plot, the mean intensity of the internal control is indicated by a horizontal orange line. These graphs show that the intensities of the four internal markers are in a constant range regardless of MBP-16E6 intensity. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

amount of MBP-16E6 (Fig. 3D), suggesting that fluorescein is more reliable than mCherry as an internal standard for the holdup assay using fluorescence detection.

The second criterion for selecting an appropriate internal standard is that it should have limited interaction with the filter, the streptavidin resin or the plastic of the well. Fig. 4 shows a retention test of different mixtures of analyte and internal standards in filter plate containing biotin-saturated streptavidin resin. The test was performed with lysozyme and BSA quantified by capillary electrophoresis (Fig. 4A), or with fluorescein and mCherry quantified by fluorescence (Fig. 4B). To facilitate data visualization, the amount of each molecule is plotted as normalized intensity based on its highest signal in the retention test. Thus, a decreased signal indicates partial retention as observed for lysozyme and BSA. In absence of MBP-16E6 and BSA, 70.6 ± 0.3% of lysozyme flowed through the resin while 82.7 ± 5.2% of BSA flowed through the resin in absence of MBP-16E6 and lysozyme. Since the holdup is a comparative assay, this partial retention does not impact the quality of the holdup data if it remains in a constant range for all tested ligands and negative controls. This is the case for lysozyme and BSA since their levels remain constant during a holdup assay for different MBP-E6 concentration (Fig. 3). Regarding mCherry, the retention test showed complete retention in absence of MBP-16E6, indicating non-specific interaction with the streptavidin resin, the biotin or the filter plate. Thus, fluorescein seems more reliable than mCherry as an internal marker for the holdup assay with fluorescence readout.

Finally, another important parameter determining the quality of an internal standard is its stability over time. However, in the holdup assay depicted herein, the time elapsed from the preparation of a fresh standard solution to the measurement is of maximum 1 h. Thus, we did not assess the stability of the internal standard for such a short handling time.

3.1.5. Calculation of the binding intensity

The holdup assay allows relative affinity estimation by quantifying the analyte depletion at equilibrium after incubation with an immobilized ligand [7]. This analysis is enabled by the calculation of binding intensity (BI), which is the ratio of bound analyte (obtained by calculating the difference between the total and the remaining amount of protein) as compared with a negative control. The analyte and internal standard molecules in the final fraction are quantified either by capillary electrophoresis or fluorescence spectroscopy. First, the amount of analyte is normalized by the amount of one of the internal standard molecules. Second, the BI reports on the fraction of depleted analyte after incubation with the ligand molecule as compared to the amount of analyte after incubation with a negative control molecule. In principle, a BI of 0.00 indicates that the concentration of analyte remaining in the filtrate is the same after incubation with the putative ligand or incubation with the negative control. On the contrary, a BI equal to 1.00 means that the analyte cannot be detected anymore in the filtrate after incubation with the ligand.

Fig. 5 shows the correlation of the BI values from the two methods used for quantifying the analyte: capillary electrophoresis and fluorescence spectroscopy. For the readout by capillary electrophoresis, Fig. 5A shows the correlation between lysozyme and BSA normalizations. The correlation between mCherry and fluorescein normalizations is plotted in Fig. 5B for the fluorescence readout. In both cases, the slope of the linear fit is very close to 1 (0.971 for capillary electrophoresis and 1.067 for fluorescence), showing high correlation between the internal standards. The error bars showing the standard deviation between two independent experiments are higher for capillary electrophoresis than for fluorescence, suggesting a better robustness of the latter. Finally, the comparison between capillary electrophoresis and fluorescence spectroscopy as holdup readouts is plotted on Fig. 5C. The two datasets are correlated as they could be fitted by linear regression. However, the BI are significantly lower for fluorescence than for capillary electrophoresis. For the same interaction between $E6AP_{LXXLL}$ wild-type and MBP-16E6, the BI estimated by the capillary electrophoresis setup is 0.922 ± 0.043 and 0.778 ± 0.033 for the fluorescence readout. This difference can be explained by the fact that capillary electrophoresis enables specific quantification of one protein based on its molecular weight. On the opposite, fluorescence spectroscopy only allows total protein quantification without any selection. Thus, any protein present in solution with the analyte (bacterial contaminant, keratin …) is included in the quantification by tryptophan fluorescence. All added together, these contaminant proteins form a background of non-interacting proteins that flow through the resin regardless of the immobilized ligand. In theory, a BI equal to 1.00 corresponds to 100% of analyte bound to the ligand immobilized on resin and thus, no analyte detected in the flow-through. If some non-interacting proteins are in solution with the analyte after the holdup, they are present in the flow-through even if the analyte is entirely retained on the resin by a high-affinity ligand. Since fluorescence spectroscopy does not allow differentiation between different proteins, the presence of non-interacting protein contaminants would give the same result as if a fraction of the analyte does not bind the ligand. Due to this background of non-interacting protein, the BI from the fluorescence

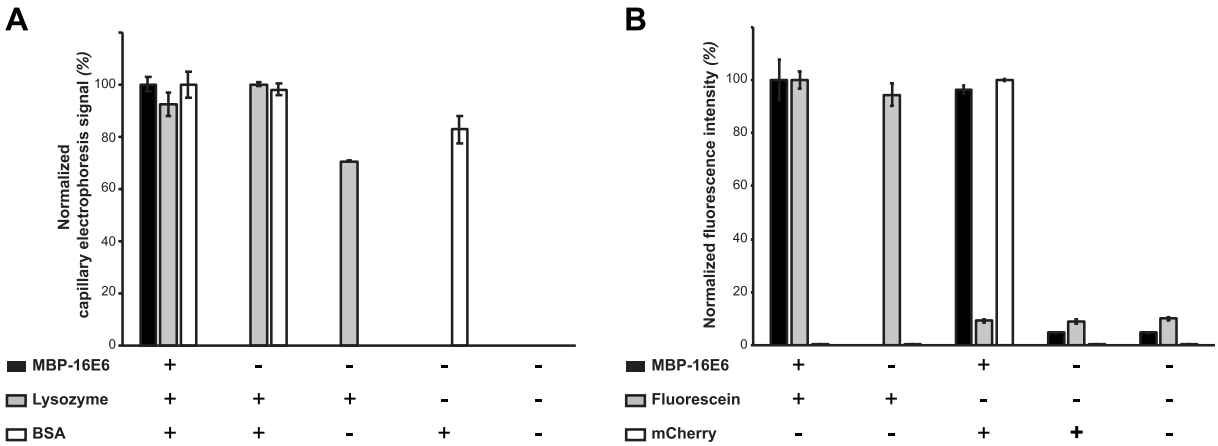


Fig. 4. Retention test of internal controls mixed with MBP-16E6 on streptavidin resin saturated with biotin. The normalized capillary electrophoresis signal (**A**) or fluorescence intensity (**B**) are plotted for each tested mixture of MBP-16E6 and standards. 100% indicates that the molecule flowed through the resin without any loss whereas 0% indicates that the molecule was completely retained on the filter with the resin. The assay was performed in duplicate and the experimental standard deviations are shown as error bars. Note that mCherry was entirely retained while lysozyme was only partially retained on the filter with biotin-saturated resin [color online only].

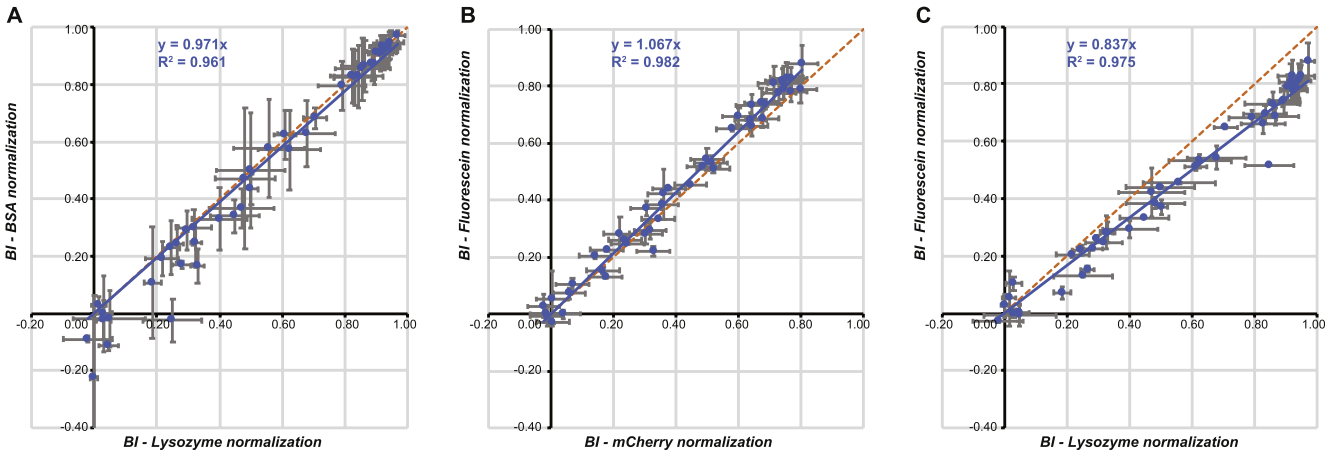
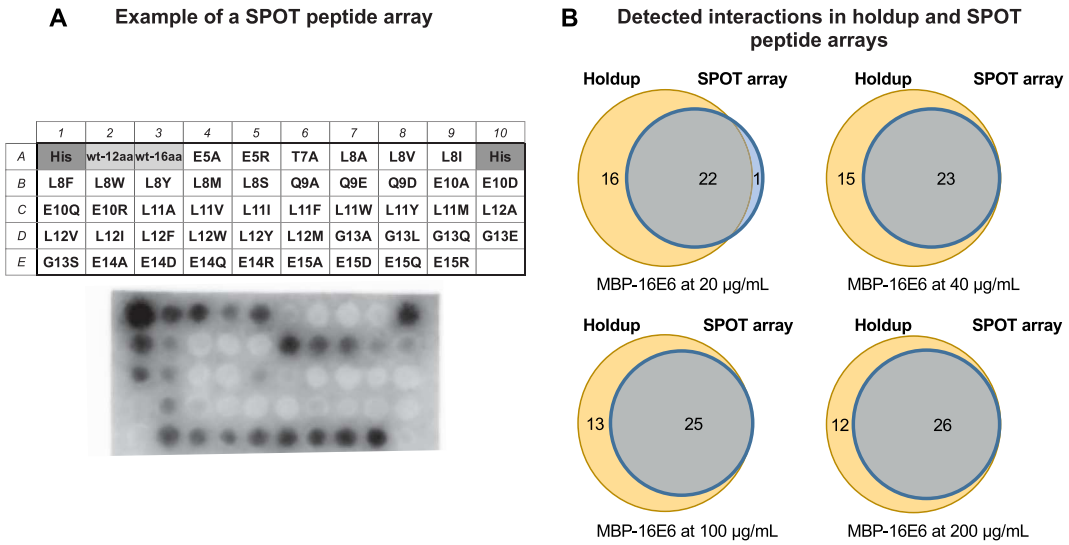


Fig. 5. Comparison of holdup binding intensities (BI) normalized by four internal markers. The holdup assay can be performed using either capillary electrophoresis or fluorescence spectroscopy for quantification of unbound analyte. For each approach, two alternative internal markers were tested. When the analyte was quantified by capillary electrophoresis, lysozyme and BSA were used for normalization (**A**). Note the high correlation between BI normalized by lysozyme and by BSA. When the analyte was quantified by tryptophan fluorescence, the internal markers were mCherry and fluorescein (**B**). Finally, it is apparent from the scatter plot **C** that the BI(lysozyme) from capillary electrophoresis are higher than the BI(fluorescein) that were obtained by measuring fluorescence. The plotted BI values are the mean of duplicated experiments with standard deviation shown as error bars. The linear regression line is plotted on each graph and its equation with coefficient of determination R^2 are indicated in the top left corner. The function $y = x$ is plotted in dotted orange line on each graph. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

approach are lower than the BI from the capillary electrophoresis approach.

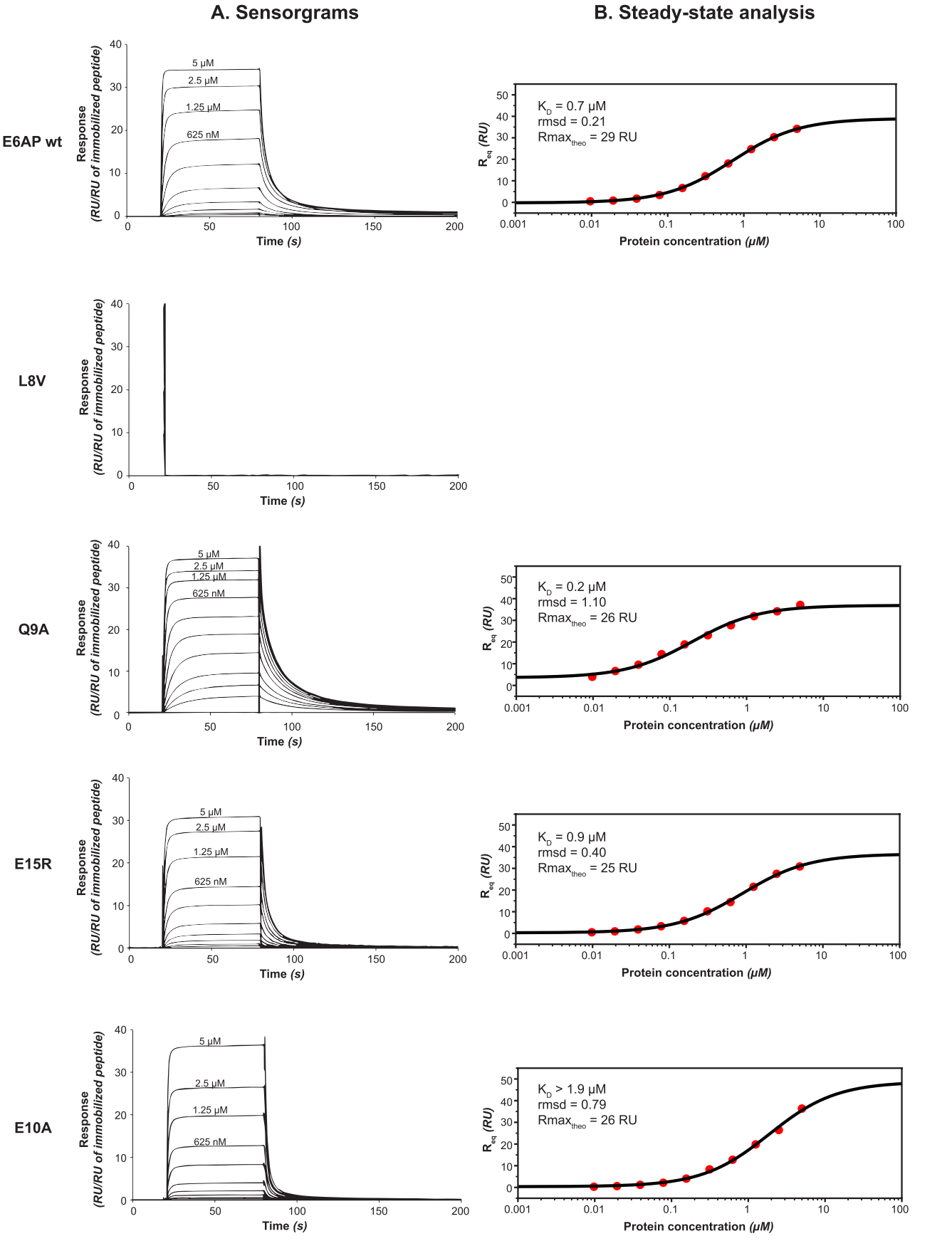
These results validate the two internal standards tested for each approach, as they give consistent results. Lysozyme and BSA can be used independently or together in a holdup assay. We recommend using both in each assay, so that they complement and possibly replace each other in case of low-quality measurement. The fact that the BI normalized by mCherry are consistent with those normalized by fluorescein suggests that mCherry did not interfere with the tested 16E6-E6AP$_{LXXLL}$ interactions. Nevertheless, since we have previously noted that mCherry strongly interacts with biotin-streptavidin resin in absence of MBP-16E6, we recommend using fluorescein which does not interact with the immobilization support.

The holdup assay with fluorescence readout resulted in reproducible data that were consistent with the results from capillary electrophoresis. The major weakness of this approach is the background due to total protein quantification. To limit this background, it is strongly recommended to purify protein samples carefully in order to avoid high amounts of contaminants in solution with the analyte. Even though the BI values are lower with total protein quantification, the ranking of binding ligands by affinity remains reliable. The conversion of BI into binding free energy ΔG is a robust solution to convert these relative affinity estimations into some comparable values, as further detailed in section 3.3.



	1	2	3	4	5	6	7	8	9	10
A	His	wt-12aa	wt-16aa	E5A	E5R	T7A	L8A	L8V	L8I	His
B	L8F	L8W	L8Y	L8M	L8S	Q9A	Q9E	Q9D	E10A	E10D
C	E10Q	E10R	L11A	L11V	L11I	L11F	L11W	L11Y	L11M	L12A
D	L12V	L12I	L12F	L12W	L12Y	L12M	G13A	G13L	G13Q	G13E
E	G13S	E14A	E14D	E14Q	E14R	E15A	E15D	E15Q	E15R	

Fig. 6. Interaction assay by SPOT peptide array. A. Example of a SPOT peptide array with mutant E6AP$_{LXXLL}$ after testing the interaction with MBP-16E6. The plan of spotted peptides is displayed above the membrane: the wild-type motif with 12 amino acid sequence ELTLQELLGEER is labelled wt-12aa and the wild-type motif with 16 amino acid sequence $P^1E^2S^3S^4E^5L^6T^7L^8Q^9E^{10}L^{11}L^{12}G^{13}E^{14}E^{15}R^{16}$ is labelled wt-16aa. The single-point mutants are based on wt-16aa sequence: the initial residue is indicated with its position in the wt-16aa motif, followed by the replacing residue. The membrane was incubated with 40 µg/mL 6His-tagged MBP-16E6, which was further detected by an anti-His antibody. The SPOT membrane entails His spots as positive controls for detection by anti-His antibody. **B**. Detected interactions in holdup and SPOT arrays tested for different MBP-16E6 concentrations. The selection threshold for a significant interaction based on binding intensity is 0.20. Holdup conditions were constant (capillary electrophoresis setup with lysozyme normalization) whereas SPOT arrays were tested with MBP-16E6 concentrations ranging between 20 and 200 µg/mL.



(caption on next page)

Fig. 7. Representative SPR data for estimating the interaction properties of MBP-16E6 with several E6AP$_{LXXLL}$: E6AP wt and the mutants L8V, Q9A, E15R and E10A. A. Sensorgrams for the interaction of MBP-16E6 with the different E6AP$_{LXXLL}$, plotting response normalized by the peptide immobilization level versus time. The flat response for the L8V mutant indicates that the interaction is abolished. **B.** Steady-state analysis for each interaction plotting the equilibrium response obtained in a 5-s window versus protein concentration. rmsd stands for root mean square deviation between experimental data (red dots) and fitted curve (black line). Rmax$_{theo}$ corresponds to the maximal expected signal assuming that each molecule of the immobilized ligand is involved in a 1:1 interaction with the analyte. This value is normalized by the immobilization level of the ligand, therefore Rmax$_{theo}$ = MW$_{analyte}$/MW$_{ligand}$. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.2. Validation by SPOT and SPR

3.2.1. SPOT peptide array

We aimed at comparing the sensitivity of the holdup assay to quantify affinities, with SPOT peptide arrays, which are often used for positional scanning. We tested the interactions between 16E6 protein in solution and SPOT peptide arrays entailing our peptide library of mutant E6AP$_{LXXLL}$ on membrane. Fig. 6A shows a SPOT array on membrane after incubation with 6His-tagged MBP-16E6 and detection of bound 16E6 by an anti-His antibody. Using an in-house python script, we quantified the intensity of each spot and normalized it. Since the quantity of analyte retained on the membrane should also be related to affinity, we therefore directly compared the normalized SPOT intensities with BI from the holdup assay. For a qualitative analysis, we translated these BI values into binary results (binds/does not bind) (Appendices,Table A1). We defined 0.20 as the BI threshold for a significant interaction and compared the number of binders detected by holdup (constant conditions) with those obtained by SPOT array (incubated with increased concentrations of MBP-16E6). Fig. 6B shows that the overlap between the two approaches increased with the concentration of MBP-16E6 incubated with the SPOT membrane. At the highest tested MBP-16E6 concentration, we note that 12 presumably weak interactions undetected by SPOT assay are still detected by holdup.

In conclusion, while both SPOT and holdup assay allowed efficient and accurate detection of E6-peptide interactions with affinity in the low micromolar range, the holdup assay turned out to be more sensitive than SPOT array for E6-peptide interactions in the mid-micromolar range.

3.2.2. SPR assay

We used SPR to measure the MBP-16E6 binding affinity constants of a selection of LXXLL biotinylated peptides displaying the highest affinities according to holdup data. The first aim was to validate the ranking obtained by holdup data using an orthogonal method. The second aim was to independently estimate K_D values allowing us to evaluate the peptide concentration in the holdup assay and subsequently to convert all BI values into K_D, and ΔG values, as discussed in the next paragraph.

We assayed the interactions of MBP-16E6 with the E6AP$_{LXXLL}$ wild-type and four mutants (Fig. 7). As shown by the absence of response on the sensorgram, the mutation L8V abolishes the interaction whereas the affinity is modulated by the mutations Q9A, E15R and E10A. The dissociation constants (K_D) were determined by steady-state analysis. Considering the biological replicates, the K_D for the interaction between MBP-16E6 and E6AP$_{LXXLL}$ wild-type is 0.84 ± 0.12 μM. The mutation Q9A enhances the affinity by around 4-fold, which reaches 0.2 μM (singlicate measurement). The mutation E15R has moderate impact on the affinity since the corresponding K_D is 1.01 ± 0.15 μM and the mutation E10A seems to decrease the affinity by at least 2-fold. The above K_D values are consistent with the BI measured by holdup assay: the BI of MBP-16E6 interaction with E6AP wt is 0.92 ± 0.04 when the BI of the affinity-enhancing mutation Q9A is 0.97 ± 0.02. The BI of E15R peptide is 0.94 ± 0.03. This value is close to the BI of E6AP wt, which confirms the K_D estimated by SPR.

For most of the tested interactions, we note that the experimental Rmax (highest experimental response at equilibrium) is close to the fitted Rmax: on Fig. 7B, the highest red dot is close to the plateau of the fitted curve for the ligands E6AP wt, Q9A and E15R. However, this is not the case for the peptide E10A, indicating that the tested conditions did not allow the analyte to fully saturate the surface. Thus, we only report a rough estimation for the corresponding dissociation constant. In addition, we observe that the fitted Rmax (value of the plateau) is rather consistent with the theoretical Rmax$_{theo}$ indicated on Fig. 7B. The Rmax$_{theo}$ value corresponds to the expected maximal signal inferred from the molecular weights of the analyte and ligand and assuming a 1:1 binding model. The consistency between the fitted and theoretical Rmax confirms i) that both analyte and ligand were biologically active for specific binding, ii) that the interaction occurs with a 1:1 stoichiometry. Lastly, the rmsd (root mean square deviation) never exceeds 5% of the fitted Rmax value, which denotes the good quality of the fit and the reliability of the reported dissociation constants.

Our SPR results confirm the slight affinity modulation of the single-point mutations as first observed with holdup binding intensities. The dissociation constants of five E6AP$_{LXXLL}$ peptides were accurately estimated by SPR, allowing us to convert binding intensities into dissociation constants for the entire peptide array as detailed in the next section.

3.3. Estimation of the free energy (ΔG) of 16E6-E6AP$_{LXXLL}$ interactions

BI values can be used to estimate binding affinity constants provided that the concentration of peptide attached on the resin is determined. This can be achieved directly by quantifying the peptide before and after immobilization on the resin, or indirectly by using the dissociation constant of one or several analyte-ligand couples measured independently [7,18]. In the present study, we estimated by SPR the dissociation constants of MBP-16E6 for three variants of the E6AP$_{LXXLL}$ motif. We then used these dissociation constants to estimate the peptide concentration in the holdup experiments, which was 14.8 ± 5.2 μM. The peptide concentration determined in that way was subsequently used to convert all the experimentally determined BI values into K_D and ΔG (Appendices, Table A2), assuming the peptide concentration to be always the same within the holdup experiments. For proper comparison of the E6AP$_{LXXLL}$ mutants with the wild-type motif, we calculated for each mutant the ΔΔG = ΔG$_{MUTANT}$ - ΔG$_{WT}$ (Fig. 8). Some mutations display ΔΔG values close to 0, suggesting a neglectable impact on the affinity with 16E6 (Q9D, E14D, E14A, E15Q and E15R). On the contrary, we identified mutations enhancing the affinity (Q9A, Q9E and E15D) and several other mutants abrogating completely the interaction, most of them targeting L8, L11 and L12. This result is consistent with the well-defined consensus motif LXXLL. The holdup assay thus allowed us to establish a quantitative interaction map of 45 single point mutants of the E6AP$_{LXXLL}$ binding motif.

4. Discussion

4.1. Advantages of the holdup assay for quantifying affinities in the micromolar range

The methods for studying protein-protein interactions are either highly accurate with quantitative biophysical methods requiring specific equipment and expertise (for instance ITC, SPR), or less accurate with qualitative biochemical assays easier to set-up in a biochemistry laboratory (mainly GST-pulldown assay, co-immunoprecipitation). In addition, very few approaches allow the quantitative assessment of affinities in the mid-micromolar range. The holdup assay with the adapted protocol presented here allows quantitative determination of affinity constants of weak protein-peptide interactions by using an easy setup and standard lab equipment.

The major difference between the traditional pulldown assay and the holdup is the quantification of the unbound fraction in the latter, instead of eluting the complex formed on the resin and directly quantifying it [26].

The advantage of quantifying the unbound analyte is that it allows subtraction of non-specific binding which may arise from the resin or the plastic plate. Since the holdup is a comparative assay, each sample incubated with a ligand is compared with a negative control in the closest possible conditions. The non-specific binding to any other component than the ligand is expected to be the same in both conditions, therefore the analyte depletion can only be due to specific binding of the ligand. In addition, the washing steps included in pulldown assay prevent the identification of low-affinity fast-dissociating complexes which can be more easily detected in the holdup equilibrium assay.

Quantitative estimations are essential in perturbation experiments of protein-protein binding. Very recently, a chromatographic assay enabling such detailed analysis was published [27]. The analyte in complex with the ligand immobilized on resin packed in a column was eluted by gradient of pH and salt. Since the holdup assay does not require any chromatographic system, the gradient elution cannot be performed but could be replaced by comparative incubation in serial buffers at various pH and ionic strengths for the analyte solution.

4.2. Prerequisites and advice for setting up the holdup assay

There are several important prerequisites for setting up the holdup assay, which are discussed below and summarized in Fig. 9 as well as the different possible experimental setups. As for most *in vitro* protein-protein interaction assays, the main requirement for holdup in terms of sample quality is that the two interacting proteins should be biochemically active and stable over the time of the experiment (no degradation nor oligomerization preventing specific interaction). If a holdup assay is performed with an analyte protein that is partly or fully inactive (due to misfolding, aggregation, presence of an inhibitory molecule in the binding site …), the results may falsely indicate an absence of binding (false negative) or a weaker affinity than the actual one. It is therefore crucial to assess the optimal quality of the analyte proteins before starting any interaction assay. The presence of contaminants or cleavage products can be visualized on SDS-PAGE gel and the oligomerization state can be checked by gel-filtration or dynamic light scattering. If the protein is capable of interacting specifically with its ligand even in the presence of contaminants, the holdup assay can be performed with overexpressed analyte in total extracts, as previously described [7].

We report two possible techniques for quantifying the free analyte in the final holdup fractions: capillary electrophoresis and fluorescence spectroscopy, the latter being equally accurate for protein quantification in holdup samples and a better choice for the user in terms of budget and time. However, when no fluorimeter is available in a laboratory, it remains possible to use SDS-PAGE as a readout to identify and quantify the protein bands as previously described [16]. It should be mentioned that various setups for high-throughput and reproducible SDS-PAGE, that may represent competitive and cost-effective approaches to microfluidic capillary electrophoresis, are commercially available.

4.3. Preparative SPOT synthesis and SPOT peptide arrays for interaction assays

Using the benchtop setup presented herein, the manual holdup assay can in principle reach a throughput of up to a few hundred protein-peptide interaction points measured per day. Based on the setup described in the present study, each single interaction point requires 6 nmol of peptide presenting the proper sequence. For higher throughput and lower amount of peptide per well, it is possible to upscale the holdup assay from 96-well plate to 384-well plates. Obtaining all the required peptides in sufficient amounts and purity may represent practical limitations (high costs) if the peptides have to be synthesized by a company. To circumvent that problem, it should be possible to use biotinylated peptides synthetized in large numbers by preparative SPOT synthesis. This approach can yield up to 1000 biotinylated peptides in a single synthesis run. A “chemical purification step”, consisting of an acetylating procedure prior to biotinylating, can be applied for optimal purity (> 90%) of biotinylated peptides. The peptides are then detached from the nitrocellulose membranes and re-solubilized prior to the assay. The amount of peptide per spot ranges from 50 to 100 nmol, which allows to perform up to 8–16 individual binding assays. This approach was previously reported for the preparation of a single replacement scan peptide library used for the development of an antimicrobial peptide [28].

Besides, SPOT arrays are frequently used for screening a library of mutant peptides. They are well adapted if the purpose of the study is to perform a semi-quantitative screen and to select high-affinity peptides [29]. However, many protein-motif interactions are transient and can dissociate during the washing steps. In order to accurately detect, compare and quantify the affinity of such interactions, we demonstrated in the present study that the manual holdup is a suitable approach with limited requirements in terms of equipment, that allows to quantify even low affinity interactions not detected by SPOT assays.

4.4. Converting holdup binding intensities into equilibrium dissociation constants

For each tested ligand/analyte pair, the holdup experiment provides a binding intensity that can be in principle converted into a steady-state

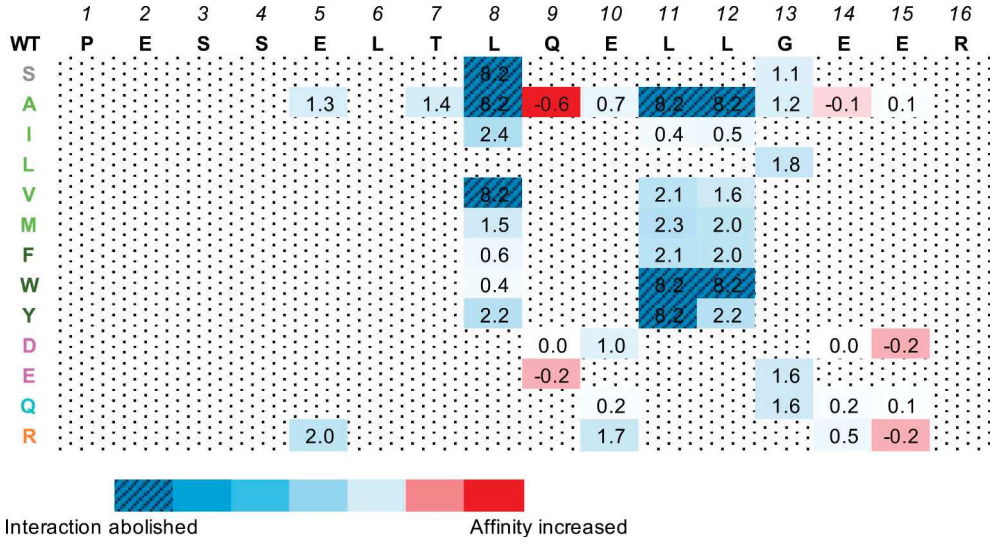


Fig. 8. Interaction of 16E6 with single-point mutant peptides of E6AP$_{LXXLL}$ binding motif. ΔG (in kcal.mol^{-1}) was calculated for wild-type and mutant E6AP$_{LXXLL}$ peptides according to the SPR normalization of the holdup data (capillary electrophoresis, lysozyme normalization). In this figure we show the values of ΔΔG = Δ G$_{MUTANT}$ - ΔG$_{WT}$. The mean uncertainty was estimated at 0.5 kcal mol^{-1}. Positive values (in blue) correspond to a loss of affinity as compared to the interaction with wild-type motif, whereas a gain of affinity is shown by negative value (red). A value of ΔΔG = 8.43 kcal mol^{-1} shows a complete loss of interaction, displayed in hatched blue. The color code on the left column indicates the biochemical properties of the side chain of the mutated amino acids: small (grey), aliphatic (light green), aromatic (dark green), negative (magenta), positive (turquoise), charged (orange).

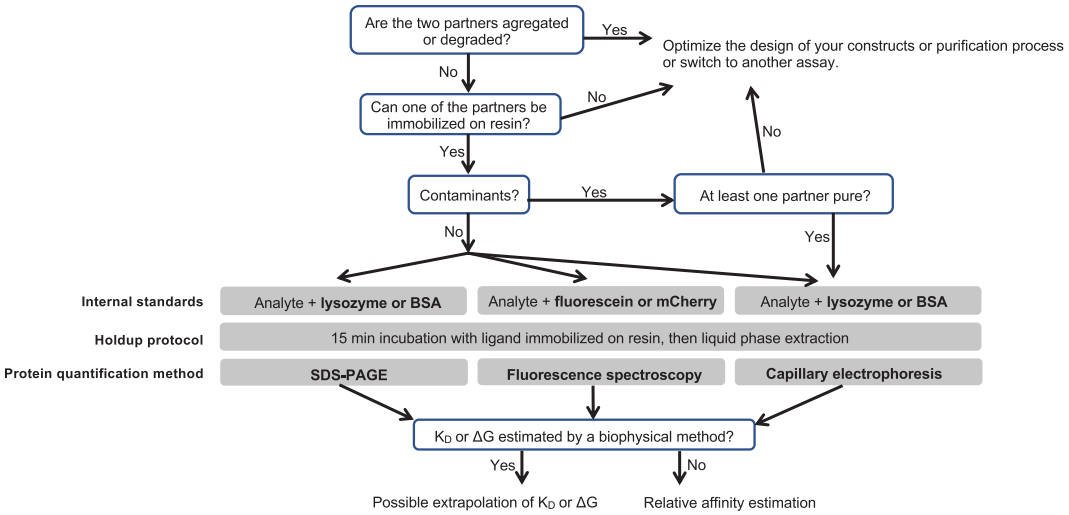


Fig. 9. Adaptation process of holdup assay to the user's subject and needs. This scheme describes the main steps by which the holdup assay can be adapted to different protein-protein or protein-peptide interaction systems and how the final output can be interpreted according to the setup chosen. The choice of the resin for running reliable holdup assays should allow stable immobilization without interfering with the tested interaction. The experimental setup is described in the grey squares, the parameters to be adapted in bold [color online only].

dissociation constant and then free energy. Such extrapolation requires accurate estimation of the concentrations of free analyte and complex, but also of the concentration of ligand immobilized on the resin. This can be achieved either by direct peptide quantification after immobilization or by indirect estimation using the dissociation constants of several interaction pairs measured by other means. Accurate quantification of peptide having no aromatic residue can be challenging since it requires specific equipment (spectrophotometer capable to record accurate data at 214 nm) [30]. In addition, direct quantification of immobilized peptide does not consider the percentage of inactive peptide that may be present on the resin (for instance insoluble peptide unable to interact). Those quantification issues can be circumvented by calibrating the holdup results with several known values of dissociation constant, which allows to estimate the concentration of active peptide present on the resin. The dissociation constants used for this purpose can be measured with other orthogonal methods for cross-validation, or reported in published data bearing in mind that such data should have been obtained in conditions as close as possible to the ones used for holdup in order to allow reasonably reliable extrapolation (buffer composition, same ligand and analyte). In the present study, we investigated the recognition specificities of 16E6 for an array of 45 single-point mutation E6AP$_{LXXLL}$ peptides and calibrated our holdup data by SPR measurements on a subset of ligands in order to extrapolate ΔΔG values of each tested interaction.

4.5. Structural aspects revealed by this study

We interpreted our mutagenesis interaction data with the crystal structure of 16E6 in complex with wild-type E6AP$_{LXXLL}$ that was previously published by our team [12]. As shown on Fig. 10A, the LXXLL motif folds onto an α-helix that binds to the basic-hydrophobic pocket between the two zinc-binding domains of 16E6. In addition, the residues of 16E6 forming the interface with the peptide are mainly positive, aliphatic and aromatic (Fig. 10A and B). According to the binding data summarized in Fig. 8, we observed that the ΔΔG values are mainly positive, indicating that most of the tested mutations decreased the affinity or even abolished the interaction with 16E6 as compared to the wild-type motif. In particular, our data confirmed that the three leucine residues L8, L11 and L12 that determine the LXXLL consensus are amino acid sequence determinants since none of the tested

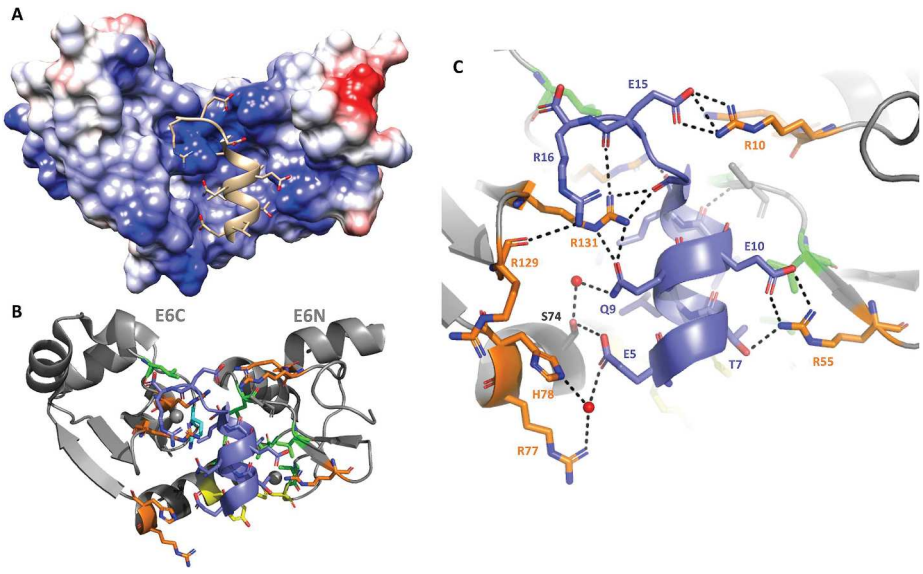


Fig. 10. The previously solved structure of 16E6 in complex with E6APLXXLL wild-type. A. Charge distribution at the surface of 16E6. The surface is colored in blue and red for positive and negative potentials, respectively. E6AP$_{LXXLL}$ is shown in beige. **B.** Overview of the 16E6-E6AP$_{LXXLL}$ complex. The two zinc-binding domains E6N and E6C are indicated in grey. The secondary structure of 16E6 is shown with ribbon representation. E6AP$_{LXXLL}$ wild-type (sequence $E^5L^6T^7L^8Q^9E^{10}L^{11}L^{12}G^{13}E^{14}E^{15}R^{16}$) is shown in blue and the residues from 16E6 at the peptide interface are colored as follows: S: grey; A, I, L, V, M: green; F, W, Y: yellow; Q: turquoise; R, H: orange. **C.** Detailed view of 16E6-E6AP$_{LXXLL}$ interactions using the same color code as in B. Hydrogen bonds are indicated by black dashed lines and water molecules by red spheres. PDB ID: 4GIZ [12].

mutations increased the affinity for 16E6. It is interesting to note that the leucine residues L8 slightly tolerates aromatic residues such as tryptophan or tyrosine while the two other leucine residues L11 and L12 do not tolerate any other hydrophobic residue. On the opposite, we identified 3 mutations that improved the affinity of 16E6 for $E6AP_{LXXLL}$: Q9A, Q9E and E15D. In the native motif, the Q9 residue from $E6AP_{LXXLL}$ stabilizes the structure by making two hydrogen bonds with R131 in 16E6 (Fig. 10C). The mutation of this polar residue into the aliphatic alanine residue might enhance the contacts within the hydrophobic interface, compensating the loss of hydrogen bonds. In addition, our results indicated the substantial participation of the flanking residues E5 and E15. E5 makes a direct hydrogen bond with S74 and a water-mediated interaction with H78 and R77 in 16E6. E15 forms three hydrogen bonds with R10 in 16E6. The E15D mutation slightly increases the affinity, which can be due to the conservation of charges on the lateral chain. These results give hints for the development of affinity-enhanced mutants blocking LXXLL motifs capture by 16E6 for therapeutic purpose. They could also allow the identification of additional protein targets of 16E6 by extending the definition of the recognized motif. It will be interesting to apply the same approach to probe the LXXLL binding preferences of E6 proteins from other papillomaviruses, compare them to those of E6 from HPV16 and interpret preference variations in regard of their biological and pathological properties.

5. Conclusions

Streamlined holdup is a fast, easy and versatile assay for accurate affinity estimation without the need for any specific lab equipment. The results are consistent with high-precision biophysical methods like SPR and show higher sensitivity for low affinity interactions. For optimal results, the user can choose to work either with crude extracts or purified proteins, depending on the solubility of the analyte and the ease of handling non-purified protein samples. Streamlined holdup can be readily adapted for pairwise affinity estimation of any protein-protein interaction system, including protein complexes. We hope that the biochemistry community will take advantage of this method by using and adapting it to their subject of interest. By giving hints on the affinity parameters, this approach allows a deeper analysis than binary assays (bind/does not bind), thus a better understanding of protein-protein interactions. In summary, with the same samples and materials as those required for a pull-down with purified proteins, the holdup is more sensitive to low and medium affinities and allows an accurate ranking of the interaction partners based on their affinity with the target protein. This versatile approach can in principle be applied to the interaction of two full-length folded proteins. In addition, the interaction preferences of 16E6 with some single-point mutants of $E6AP_{LXXLL}$ allowed us to identify the residues that determine the interaction. In the present case, these key positions are not limited to the three leucine residues from the consensus sequence but also include acidic residues upstream and downstream the LXXLL motif.

Subject category

Chromatographic techniques.

Funding

This work received institutional support from le Centre National de la Recherche Scientifique (CNRS), Université de Strasbourg, Institut National de la Santé et de la Recherche Médicale (INSERM) and Région Alsace. The work was supported in part by grants from Ligue contre le Cancer (équipe labellisée 2015 and fellowship to A.B.), Fondation pour la Recherche Médicale (fellowship to A.B.), National Institutes of Health [Grant R01CA134737], Instruct (ESFRI), and the French Infrastructure for Integrated Structural Biology (FRISBI). The authors declare that the content is solely their responsibility and does not represent the official views of the National Institutes of Health.

Author contributions

Anna Bonhoure: Conceptualization, methodology, investigation, data analysis, visualization, writing — original draft, **Anne Forster**: Conceptualization, investigation, writing — editing, **Khaled Ould Babah**: Conceptualization, investigation, writing — editing, **Gergő Gógl**: investigation, data analysis, writing — editing, **Pascal Eberling**: Resources, writing — editing, **Camille Kostmann**: investigation, writing — editing, **Rudolf Volkmer**: Resources, writing — editing, **Victor Tapia Mancilla**: Resources, writing — editing, **Gilles Travé**: Conceptualization, supervision, project management, funding acquisition, formal analysis, visualization, writing — review and editing; **Yves Nominé**: methodology, software, formal analysis, visualization, writing — review and editing.

Declaration of competing interest

The authors declare that they have no competing interest that could have influenced the work reported in the present paper.

Acknowledgements

The authors thank B. Kieffer for the Python script used for intensity quantification on SPOT arrays and colleagues of the Travé team for helpful discussion and support. The authors also thank the reviewers for their careful reading of the manuscript.

Appendices

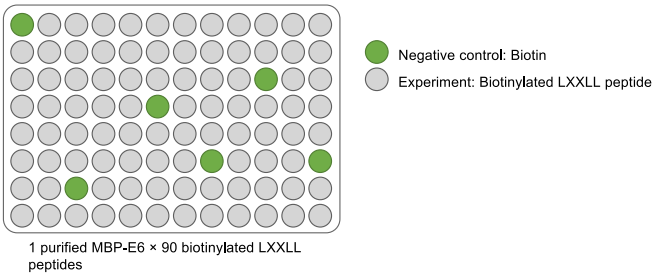


Fig. A1. Plate organization for holdup assay. In a single assay on 96-well plate, we can test the interaction between 1 MBP-E6 against up to 90 biotinylated LXXLL peptides immobilized on streptavidin resin. The plate typically contains 6 to 8 negative control wells containing biotin, randomly placed [color online only].

Table A1

Comparison between holdup binding intensities and normalized intensities of SPOT peptide arrays at increasing concentrations of MBP-16E6. From duplicated SPOT assay at 40 μg/mL MBP-16E6, we observed a mean standard deviation of ± 0.1. The holdup binding intensities were obtained with the protocol including quantitative capillary electrophoresis and lysozyme normalization. The mean standard deviation observed from duplicated assays is ± 0.05. The above standard deviations are indicated in the same unit as the normalized values in the table [color online only].

		Holdup	SPOT with increased concentration of MBP-HPV16 E6 F47R 4C/4S (in *μg/mL*)			
		BI	*20*	*40*	*100*	*200*
wt	PESSELTLQELLGEER	0.92	0.6	0.8	0.8	0.8
E5A	PESSALTLQELLGEER	0.61	0.3	0.5	0.5	0.4
E5R	PESSRLTLQELLGEER	0.32	0.3	0.6	0.7	0.8
T7A	PESSELALQELLGEER	0.56	0.1	0.2	0.2	0.3
L8A	PESSELTAQELLGEER	0.00	0.2	0.1	0.2	0.0
L8V	PESSELTVQELLGEER	0.01	0.1	0.1	0.1	0.0
L8I	PESSELTIQELLGEER	0.22	0.5	0.2	0.3	0.1
L8F	PESSELTFQELLGEER	0.83	0.5	0.8	0.5	0.7
L8W	PESSELTWQELLGEER	0.86	0.2	0.4	0.4	0.4
L8Y	PESSELTYQELLGEER	0.27	0.0	0.1	0.1	0.1
L8M	PESSELTMQELLGEER	0.50	0.0	0.0	0.0	0.0
L8S	PESSELTSQELLGEER	0.05	0.1	0.1	0.1	0.1
Q9A	PESSELTLAELLGEER	0.97	0.9	1.0	1.0	1.0
Q9E	PESSELTLEELLGEER	0.95	0.7	0.6	0.9	0.9
Q9D	PESSELTLDELLGEER	0.92	0.5	0.7	1.0	0.7
E10A	PESSELTLQALLGEER	0.79	0.3	0.4	0.6	0.4
E10D	PESSELTLQDLLGEER	0.71	0.5	0.4	0.5	0.4
E10Q	PESSELTLQQLLGEER	0.89	0.4	0.6	0.6	0.5
E10R	PESSELTLQRLLGEER	0.45	0.3	0.3	0.4	0.4
L11A	PESSELTLQEALGEER	0.05	0.1	0.0	0.1	0.1
L11V	PESSELTLQEVLGEER	0.28	0.0	0.0	0.1	0.1
L11I	PESSELTLQEILGEER	0.86	0.1	0.2	0.3	0.4
L11F	PESSELTLQEFLGEER	0.30	0.2	0.2	0.1	0.2
L11W	PESSELTLQEWLGEER	0.03	0.1	0.1	0.0	0.0
L11Y	PESSELTLQEYLGEER	-0.02	0.0	0.0	0.0	0.0
L11M	PESSELTLQEMLGEER	0.24	0.0	0.0	0.0	0.0
L12A	PESSELTLQELAGEER	0.03	0.1	0.1	0.1	0.0
L12V	PESSELTLQELVGEER	0.50	0.2	0.1	0.1	0.2
L12I	PESSELTLQELIGEER	0.83	0.4	0.3	0.3	0.4
L12F	PESSELTLQELFGEER	0.33	0.1	0.1	0.2	0.3
L12W	PESSELTLQELWGEER	0.19	0.0	0.0	0.1	0.1
L12Y	PESSELTLQELYGEER	0.25	0.0	0.0	0.1	0.0
L12M	PESSELTLQELMGEER	0.33	0.1	0.0	0.0	0.0
G13A	PESSELTLQELLAEER	0.62	0.2	0.2	0.1	0.1
G13L	PESSELTLQELLLEER	0.40	0.1	0.1	0.1	0.1
G13Q	PESSELTLQELLQEER	0.48	0.1	0.1	0.1	0.2
G13E	PESSELTLQELLEEER	0.47	0.1	0.1	0.1	0.1
G13S	PESSELTLQELLSEER	0.68	0.1	0.0	0.1	0.0
E14A	PESSELTLQELLGAER	0.93	0.7	0.5	0.6	0.4
E14D	PESSELTLQELLGDER	0.92	0.7	0.5	0.5	0.7
E14Q	PESSELTLQELLGQER	0.89	0.6	0.5	0.8	0.6
E14R	PESSELTLQELLGRER	0.84	0.6	0.5	0.7	0.5
E15A	PESSELTLQELLGEAR	0.90	0.8	0.7	0.8	0.8
E15D	PESSELTLQELLGEDR	0.94	0.9	0.7	0.6	0.6
E15Q	PESSELTLQELLGEQR	0.92	0.8	0.8	0.7	0.8
E15R	PESSELTLQELLGERR	0.94	1.0	0.9	0.8	0.9

Table A2

Conversion of holdup binding intensities into dissociation constants (K_D) and free energy (ΔG). The binding intensities in the present table were obtained with the holdup protocol including quantitative capillary electrophoresis and normalization with lysozyme. The raw BI from duplicate experiments are indicated with the corresponding standard deviations. The first step was to apply the threshold of 0.20 to BI values, meaning converting any BI < 0.20 into 0.00. Then, the peptide concentration attached on resin was back calculated from the affinities determined by SPR measurements (11.4 μM) and was subsequently used to extrapolate the K_D and the ΔG. ΔΔG was calculated by comparison with E6AP$_{LXXLL}$ wt ($\Delta\Delta G = \Delta G_{MUTANT} - \Delta G_{WT}$) [color online only].

		BI	BI<0.2 converted to 0.0	K_D (μM)	ΔG (kcal.mol^{-1})	ΔΔG (kcal.mol^{-1})
wt	PESSELTLQELLGEER	0.92 ± 0.04	0.92	0.93 ± 0.62	-8.22 ± 0.40	
E5A	PESSALTLQELLGEER	0.61 ± 0.03	0.61	7.86 ± 3.37	-6.96 ± 0.25	1.26 ± 0.47
E5R	PESSRLTLQELLGEER	0.32 ± 0.06	0.32	28.89 ± 11.26	-6.19 ± 0.23	2.03 ± 0.46
T7A	PESSELALQELLGEER	0.56 ± 0.11	0.56	9.91 ± 4.17	-6.82 ± 0.25	1.40 ± 0.47
L8A	PESSELTAQELLGEER	0.00 ± 0.01	0.00	ND	ND	ND
L8V	PESSELTVQELLGEER	0.01 ± 0.02	0.00	ND	ND	ND
L8I	PESSELTIQELLGEER	0.22 ± 0.05	0.22	49.97 ± 19.07	-5.86 ± 0.23	2.36 ± 0.46
L8F	PESSELTFQELLGEER	0.83 ± 0.07	0.83	2.42 ± 1.29	-7.65 ± 0.32	0.57 ± 0.51
L8W	PESSELTWQELLGEER	0.86 ± 0.07	0.86	1.77 ± 1.10	-7.84 ± 0.37	0.38 ± 0.54
L8Y	PESSELTYQELLGEER	0.27 ± 0.02	0.27	37.89 ± 14.42	-6.03 ± 0.23	2.19 ± 0.46
L8M	PESSELTMQELLGEER	0.50 ± 0.02	0.50	12.67 ± 5.22	-6.67 ± 0.24	1.55 ± 0.47
L8S	PESSELTSQELLGEER	0.05 ± 0.12	0.00	ND	ND	ND
Q9A	PESSELTLAELLGEER	0.97 ± 0.02	0.97	0.34 ± 0.28	-8.82 ± 0.48	-0.60 ± 0.63
Q9E	PESSELTLEELLGEER	0.95 ± 0.04	0.95	0.62 ± 0.54	-8.46 ± 0.52	-0.24 ± 0.65
Q9D	PESSELTLDELLGEER	0.92 ± 0.05	0.92	0.95 ± 0.67	-8.21 ± 0.42	0.01 ± 0.58
E10A	PESSELTLQALLGEER	0.79 ± 0.08	0.79	3.06 ± 1.54	-7.52 ± 0.30	0.70 ± 0.50
E10D	PESSELTLQDLLGEER	0.71 ± 0.04	0.71	4.97 ± 2.22	-7.23 ± 0.26	0.99 ± 0.48
E10Q	PESSELTLQQLLGEER	0.89 ± 0.05	0.89	1.41 ± 0.82	-7.97 ± 0.34	0.25 ± 0.53
E10R	PESSELTLQRLLGEER	0.45 ± 0.08	0.45	15.99 ± 6.47	-6.54 ± 0.24	1.68 ± 0.46
L11A	PESSELTLQEALGEER	0.05 ± 0.03	0.00	ND	ND	ND
L11V	PESSELTLQEVLGEER	0.28 ± 0.03	0.28	34.78 ± 13.33	-6.08 ± 0.23	2.14 ± 0.46
L11I	PESSELTLQEILGEER	0.86 ± 0.09	0.86	1.89 ± 1.21	-7.80 ± 0.38	0.42 ± 0.55
L11F	PESSELTLQEFLGEER	0.30 ± 0.00	0.30	32.28 ± 12.39	-6.12 ± 0.23	2.10 ± 0.46
L11W	PESSELTLQEWLGEER	0.03 ± 0.02	0.00	ND	ND	ND
L11Y	PESSELTLQEYLGEER	-0.02 ± 0.08	0.00	ND	ND	ND
L11M	PESSELTLQEMLGEER	0.24 ± 0.03	0.24	42.50 ± 16.11	-5.96 ± 0.22	2.26 ± 0.46
L12A	PESSELTLQELAGEER	0.03 ± 0.03	0.00	ND	ND	ND
L12V	PESSELTLQELVGEER	0.50 ± 0.11	0.50	12.82 ± 5.27	-6.67 ± 0.24	1.55 ± 0.47
L12I	PESSELTLQELIGEER	0.83 ± 0.07	0.83	2.29 ± 1.23	-7.69 ± 0.32	0.53 ± 0.51
L12F	PESSELTLQELFGEER	0.33 ± 0.02	0.33	27.91 ± 10.82	-6.21 ± 0.23	2.01 ± 0.46
L12W	PESSELTLQELWGEER	0.19 ± 0.03	0.00	ND	ND	ND
L12Y	PESSELTLQELYGEER	0.25 ± 0.09	0.25	41.08 ± 16.21	-5.98 ± 0.23	2.24 ± 0.46
L12M	PESSELTLQELMGEER	0.33 ± 0.02	0.33	27.30 ± 10.60	-6.22 ± 0.23	2.00 ± 0.46
G13A	PESSELTLQELLAEER	0.62 ± 0.10	0.62	7.40 ± 3.22	-6.99 ± 0.26	1.23 ± 0.47
G13L	PESSELTLQELLLEER	0.40 ± 0.09	0.40	19.69 ± 7.87	-6.41 ± 0.24	1.81 ± 0.46
G13Q	PESSELTLQELLQEER	0.48 ± 0.09	0.48	13.81 ± 5.65	-6.62 ± 0.24	1.60 ± 0.47
G13E	PESSELTLQELLEEER	0.47 ± 0.10	0.47	14.48 ± 5.90	-6.60 ± 0.24	1.63 ± 0.46
G13S	PESSELTLQELLSEER	0.68 ± 0.09	0.68	5.73 ± 2.58	-7.14 ± 0.27	1.08 ± 0.48
E14A	PESSELTLQELLGAER	0.93 ± 0.03	0.93	0.81 ± 0.52	-8.30 ± 0.38	-0.08 ± 0.55
E14D	PESSELTLQELLGDER	0.92 ± 0.04	0.92	0.92 ± 0.62	-8.23 ± 0.40	-0.01 ± 0.56
E14Q	PESSELTLQELLGQER	0.89 ± 0.05	0.89	1.32 ± 0.79	-8.01 ± 0.35	0.21 ± 0.53
E14R	PESSELTLQELLGRER	0.84 ± 0.08	0.84	2.10 ± 1.22	-7.74 ± 0.34	0.48 ± 0.53
E15A	PESSELTLQELLGEAR	0.90 ± 0.01	0.90	1.18 ± 0.57	-8.08 ± 0.29	0.14 ± 0.49
E15D	PESSELTLQELLGEDR	0.94 ± 0.04	0.94	0.70 ± 0.51	-8.39 ± 0.44	-0.17 ± 0.59
E15Q	PESSELTLQELLGEQR	0.92 ± 0.04	0.92	1.03 ± 0.63	-8.16 ± 0.36	0.06 ± 0.54
E15R	PESSELTLQELLGERR	0.94 ± 0.03	0.94	0.66 ± 0.45	-8.42 ± 0.40	-0.20 ± 0.56

References

[1] M. Kumar, M. Gouw, S. Michael, H. Sámano-Sánchez, R. Pancsa, J. Glavina, A. Diakogianni, J.A. Valverde, D. Bukirova, J. Čalyševa, N. Palopoli, N.E. Davey, L.B. Chemes, T.J. Gibson, ELM—the eukaryotic linear motif resource in 2020, Nucleic Acids Res. 48 (2020) D296–D306, https://doi.org/10.1093/nar/gkz1030.

[2] P. Tompa, N.E. Davey, T.J. Gibson, M.M. Babu, A million peptide motifs for the molecular biologist, Mol. Cell 55 (2014) 161–169, https://doi.org/10.1016/j.molcel.2014.05.032.

[3] N.E. Davey, M.-H. Seo, V.K. Yadav, J. Jeon, S. Nim, I. Krystkowiak, C. Blikstad, D. Dong, N. Markova, P.M. Kim, Y. Ivarsson, Discovery of short linear motif-mediated interactions through phage display of intrinsically disordered regions of the human proteome, FEBS J. 284 (2017) 485–498, https://doi.org/10.1111/febs.13995.

[4] S.G. Choi, J. Olivet, P. Cassonnet, P.-O. Vidalain, K. Luck, L. Lambourne, K. Spirohn, I. Lemmens, M. Dos Santos, C. Demeret, L. Jones, S. Rangarajan, W. Bian, E.P. Coutant, Y.L. Janin, S. van der Werf, P. Trepte, E.E. Wanker, J. De Las Rivas, J. Tavernier, J.-C. Twizere, T. Hao, D.E. Hill, M. Vidal, M.A. Calderwood, Y. Jacob, Maximizing binary interactome mapping with a minimal number of assays, Nat. Commun. 10 (2019) 3907, https://doi.org/10.1038/s41467-019-11809-2.

[5] D.M. Ichikawa, C. Corbi-Verge, M.J. Shen, J. Snider, V. Wong, I. Stagljar, P.M. Kim, M.B. Noyes, A multireporter bacterial 2-hybrid assay for the high-throughput and dynamic assay of PDZ domain–peptide interactions, ACS Synth. Biol. 8 (2019) 918–928, https://doi.org/10.1021/acssynbio.8b00499.

[6] H.Q. Nguyen, J. Roy, B. Harink, N.P. Damle, N.R. Latorraca, B.C. Baxter, K. Brower, S.A. Longwell, T. Kortemme, K.S. Thorn, M.S. Cyert, P.M. Fordyce, Quantitative mapping of protein-peptide affinity landscapes using spectrally encoded beads, ELife 8 (2019), https://doi.org/10.7554/eLife.40499.

[7] R. Vincentelli, K. Luck, J. Poirson, J. Polanowska, J. Abdat, M. Blémont, J. Turchetto, F. Iv, K. Ricquier, M.-L. Straub, A. Forster, P. Cassonnet, J.-P. Borg, Y. Jacob, M. Masson, Y. Nominé, J. Reboul, N. Wolff, S. Charbonnier, G. Travé, Quantifying domain-ligand affinities and specificities by high-throughput holdup assay, Nat. Methods 12 (2015) 787–793, https://doi.org/10.1038/nmeth.3438.

[8] S. de Sanjose, W.G. Quint, L. Alemany, D.T. Geraets, J.E. Klaustermeier, B. Lloveras, S. Tous, A. Felix, L.E. Bravo, H.-R. Shin, Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study, Lancet Oncol. 11 (2010) 1048–1056.

[9] M.L. Gillison, L. Alemany, P.J.F. Snijders, A. Chaturvedi, B.M. Steinberg, S. Schwartz, X. Castellsagué, Human papillomavirus and diseases of the upper airway: head and neck cancer and respiratory papillomatosis, Vaccine 30 (2012) F34–F54, https://doi.org/10.1016/j.vaccine.2012.05.070.

[10] J.M. Huibregtse, M. Scheffner, P.M. Howley, Localization of the E6-AP regions that direct human papillomavirus E6 binding, association with p53, and ubiquitination of associated proteins, Mol. Cell Biol. 13 (1993) 4918–4927, https://doi.org/10.1128/MCB.13.8.4918.

[11] L.V. Ronco, A.Y. Karpova, M. Vidal, P.M. Howley, Human papillomavirus 16 E6 oncoprotein binds to interferon regulatory factor-3 and inhibits its transcriptional activity, Gene Dev. 12 (1998) 2061–2072.

[12] K. Zanier, S. Charbonnier, A.O.M.O. Sidi, A.G. McEwen, M.G. Ferrario, P. Poussin-Courmontagne, V. Cura, N. Brimer, K.O. Babah, T. Ansari, I. Muller, R.H. Stote, J. Cavarelli, S. Vande Pol, G. Travé, Structural basis for hijacking of cellular LxxLL motifs by papillomavirus E6 oncoproteins, Science 339 (2013) 694–698.

[13] J.M. Huibregtse, M. Scheffner, P.M. Howley, A cellular protein mediates association of p53 with the E6 oncoprotein of human papillomavirus types 16 or 18, EMBO J. 10 (1991) 4129–4135, https://doi.org/10.1002/j.1460-2075.1991.tb04990.x.

[14] M. Scheffner, B.A. Werness, J.M. Huibregtse, A.J. Levine, P.M. Howley, The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53, Cell 63 (1990) 1129–1136, https://doi.org/10.1016/0092-8674(90)90409-8.

[15] D. Martinez-Zapien, F.X. Ruiz, J. Poirson, A. Mitschler, J. Ramirez, A. Forster, A. Cousido-Siah, M. Masson, S.V. Pol, A. Podjarny, G. Travé, K. Zanier, Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53, Nature 529 (2016) 541–545, https://doi.org/10.1038/nature16481.

[16] S. Charbonnier, K. Zanier, M. Masson, G. Travé, Capturing protein–protein complexes at equilibrium: the holdup comparative chromatographic retention assay, Protein Expr. Purif. 50 (2006) 89–101, https://doi.org/10.1016/j.pep.2006.06.010.

[17] A. Bonhoure, A. Demenge, C. Kostmann, L. San José, E. De la Cal, P. Armisen, Y. Nominé, G. Travé, One-step affinity purification of fusion proteins with optimal monodispersity and biological activity: application to aggregation-prone HPV E6 proteins, Microb. Cell Factories 17 (2018), https://doi.org/10.1186/s12934-018-1039-z.

[18] G. Gógl, B. Biri-Kovács, F. Durbesson, P. Jané, Y. Nominé, C. Kostmann, V. Bilics, M. Simon, A. Reményi, R. Vincentelli, G. Travé, L. Nyitray, Rewiring of RSK–PDZ interactome by linear motif phosphorylation, J. Mol. Biol. (2019), https://doi.org/10.1016/j.jmb.2019.01.038.

[19] G. Gógl, P. Jané, C. Caillet-Saguy, C. Kostmann, G. Bich, A. Cousido-Siah, L. Nyitray, R. Vincentelli, N. Wolff, Y. Nominé, N.N. Sluchanko, G. Travé, Dual Specificity PDZ- and 14-3-3-Binding Motifs: A Structural and Interactomics Study, Structure, (2020), https://doi.org/10.1016/j.str.2020.03.010.

[20] H.H. Ku, Notes on the use of propagation of error formulas, J. Res. Natl. Bur. Stand. 70C (1966) 263–273.

[21] H. Wenschuh, R. Volkmer-Engert, M. Schmidt, M. Schulz, J. Schneider-Mergener, U. Reineke, Coherent membrane supports for parallel microsynthesis and screening of bioactive peptides, Peptide Sci. 55 (2000) 188–206, https://doi.org/10.1002/1097-0282(2000)55:3<188::AID-BIP20>3.0.CO;2-T.

[22] M. Bleeke, S. Fest, N. Huebener, C. Landgraf, B. Schraven, G. Gaedicke, R. Volkmer, H.N. Lode, Systematic amino acid substitutions improved efficiency of GD2-peptide mimotope vaccination against neuroblastoma, Eur. J. Canc. 45 (2009) 2915–2921, https://doi.org/10.1016/j.ejca.2009.07.026.

[23] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera - a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612, https://doi.org/10.1002/jcc.20084.

[24] M.F. Sanner, A.J. Olson, J.-C. Spehner, Reduced surface: an efficient way to compute molecular surfaces, Biopolymers 38 (1996) 305–320, https://doi.org/10.1002/(SICI)1097-0282(199603)38:3<305::AID-BIP4>3.0.CO;2-Y.

[25] Y. Nominé, T. Ristriani, C. Laurent, J.-F. Lefèvre, É. Weiss, G. Travé, Formation of soluble inclusion bodies by HPV E6 oncoprotein fused to maltose-binding protein, Protein Expr. Purif. 23 (2001) 22–32, https://doi.org/10.1006/prep.2001.1451.

[26] A. Louche, S.P. Salcedo, S. Bigot, Protein–protein interactions: pull-down assays, in: L. Journet, E. Cascales (Eds.), Bacterial Protein Secretion Systems, Springer New York, New York, NY, 2017, pp. 247–255, , https://doi.org/10.1007/978-1-4939-7033-9_20.

[27] H.R. Reese, C.C. Shanahan, J. Lembo, L. Tsonev, A. Hirsh, S. Menegatti, Chromatographic assay to probe the binding energy and mechanisms of homologous proteins to surface-bound ligands, J. Chromatogr. B 1136 (2020) 121927, https://doi.org/10.1016/j.jchromb.2019.121927.

[28] D. Pulido, G. Prats-Ejarque, C. Villalba, M. Albacar, M. Moussaoui, D. Andreu, R. Volkmer, M. Torrent, E. Boix, Positional scanning library applied to the human eosinophil cationic protein/RNase 3 N-terminus reveals novel and potent anti-biofilm peptides, Eur. J. Med. Chem. 152 (2018) 590–599, https://doi.org/10.1016/j.ejmech.2018.05.012.

[29] V.E. Tapia Mancilla, V.T. Mancilla, R. Volkmer, Peptide arrays on planar supports, Methods Mol. Biol. 1352 (2016) 3–17, https://doi.org/10.1007/978-1-4939-3037-1_1.

[30] B.J.H. Kuipers, H. Gruppen, Prediction of molar extinction coefficients of proteins and peptides using UV absorption of the constituent amino acids at 214 nm to enable quantitative reverse phase high-performance liquid Chromatography – Mass spectrometry analysis, J. Agric. Food Chem. 55 (2007) 5445–5451, https://doi.org/10.1021/jf070337l.

2 Interaction preferences of multiple E6 oncoproteins for LXXLL motifs

2.1 E6 oncoproteins from different HPV types have distinct LXXLL interaction preferences

2.1.1 Introduction

Papillomaviruses are small non-enveloped viruses containing compact, circular double-stranded DNA genome of about 8,000 bp. They infect squamous epithelial cells from many different hosts, including mammals, birds, reptiles. The Human Papillomaviruses (HPV) infect human epithelia. They vary in their tissue tropism, by targeting mucosal or cutaneous epithelia. HPVs also induce a high diversity of phenotypes in the infected host, from benign proliferation (warts, condylomas) to different cancers. Thus, they can be classified according to their oncogenic risk (low-risk or high-risk). HPVs are the main etiological agents of cervix cancer. In particular, HPV 16 and 18 are responsible for 61 and 10 % of worldwide cervix cancer cases, respectively (Serrano *et al.*, 2015). They also induce various anogenital cancers (vulva, vagina, anus, penile). Some head and neck cancers are HPV-positive, with HPV 16 causing 71 % of the cases (Castellsagué *et al.*, 2016). Some HPVs can cause squamous cell carcinoma of the skin in patients with epidermodysplasia verruciformis, a rare genetic disease affecting primary immunity (Accardi and Gheit, 2014). The so-called low risk HPV 6 and 11 cause genital warts or condylomas (Yanofsky *et al.*, 2012).

In addition to the phenotypic classification based on tissue tropism and oncogenic risk, the phylogenetic tree of papillomaviruses is based on sequence homology of L1 gene. This late gene encodes for the major capsid protein and is the most conserved sequence in the viral genome. The nucleotide sequence identity of L1 gene defines the different categories of HPV phylogenetic classification. Genera are designated by Greek letters (α, β, γ, μ, ν) and gather HPVs with at least 60 % sequence identity, species have 70 to 80 % L1 sequence identity, types have a minimum of 90 % identity and variants have less than 10 % sequence variation from any other known HPV genome (de Villiers, 2013).

The HPV genome is extremely compact: this circular DNA contains overlapping open reading frames and produces polycistronic RNA that undergo alternative splicing. The genome encodes for early (E) and late (L) genes. Among the early proteins, the oncoproteins E6, E7 and E5 trigger cell proliferation and promote viral replication. E6 and E7 disrupt cell functions by interacting with many protein targets. E6 and E7 are small proteins, of about 150 and 100 amino acids, respectively. Thus, they take advantage of protein motif hijacking strategies to target a high number of cellular proteins by interactions between motifs and folded domains. The viral protein can mimic an interaction motif existing in the cell and interact with host folded domains. As an example of the mimicking strategy, the E7 oncoprotein contains LXCXE motif which is present is various cellular proteins and is recognized by the "pocket proteins" of retinoblastoma protein family, including the tumor suppressor Rb and the proteins p107 and p130 (Harden and Munger, 2017). By interacting with Rb, the E7 oncoprotein dissociates Rb from the transcriptional activator E2F, which induces S phase entry (Münger and Howley, 2002). The second strategy, which is the main topic of the present work, is the capture of a host protein motif by a viral folded domain, as LXXLL motifs captured by the hydrophobic pocket formed between the two zinc-binding domains of E6 oncoprotein (Zanier *et al.*, 2013). In addition to the motif hijacking strategy, E6 and E7 oncoproteins are also involved in domain-domain interactions. For instance, the third conserved region of HPV18 E7 (CR3) folds into a zinc-binding domain and interacts with the catalytic domain of the tyrosine phosphatase PTPN14 (Yun *et al.*, 2019). E7 recruits the ubiquitin ligase UBR4 to form a tripartite complex with PTPN14, which leads PTPN14 to its proteasomal degradation (White *et al.*, 2016). As a result, the keratinocyte differentiation is inhibited (Hatterschide *et al.*, 2019). Once bound to the LXXLL motif of E6AP ubiquitin ligase, HPV16 E6 binds the core domain of the tumor suppressor p53 (Martinez-Zapien *et al.*, 2016). Lastly, the β-1 HPV 5 and 8 E6 oncoproteins bind the histone acetyltransferase CBP/p300 (Howie *et al.*, 2011). One should note that the three domain-domain interactions cited lead to the proteasomal degradation of the bound host protein.

So far, no consensus cellular protein target is known to be shared by all HPV E6 oncoproteins. However, all E6 proteins expressed by mammalian papillomaviruses share a common fold, consisting of two zinc-binding domains connected by a linker helix. The hydrophobic pocket between the zinc-binding domains and the strong

conservation of an arginine residue (R102 in HPV16 E6) allow the recognition of helical LXXLL motifs (Zanier *et al.*, 2013). However, while all mammalian papillomavirus E6 proteins appear to bind LXXLL motifs, they can target cellular proteins by other domain-motif interactions than the LXXLL motif capture. High-risk mucosal E6 proteins harbor a C-terminal PDZ-binding motif (PBM), which is recognized by PDZ domain-containing proteins, such as MAGI1 (Vincentelli *et al.*, 2015).

The E3 ubiquitin ligase E6AP was the first cellular protein reported as a target of HPV16 E6 oncoprotein (Huibregtse *et al.*, 1991). The E6 protein captures an acidic LXXLL motif present in E6AP (Huibregtse *et al.*, 1993) and the resulting heterodimer recruits the tumor suppressor p53 by interacting with its core domain, which leads to the proteasomal degradation of p53 (Scheffner *et al.*, 1993). Another protein targeted by its LXXLL motif is the transcriptional co-activator MAML1. E6 oncoproteins from cutaneous β-HPVs capture an LXXLL motif located at the C-terminus of MAML1. This interaction prevents MAML1 transactivation and represses Notch signaling pathway, which is responsible for keratinocyte differentiation (Brimer *et al.*, 2012). Recent works suggested a tendency of LXXLL motif recognition preferences among the different HPV types: α-HPVs E6 proteins bind E6AP motif whereas β-HPVs E6 oncoproteins rather interact with MAML1 motif (Brimer *et al.*, 2017; White *et al.*, 2012a). Additional E6-binding human proteins containing LXXLL motifs were reported. The bovine papillomavirus type 1 (BPV1) E6 interacts with the focal adhesion protein paxillin, leading to cell transformation (Tong and Howley, 1997; Wade *et al.*, 2008). BPV1 E6 binds to several paxillin LXXLL motifs (Brimer *et al.*, 2017) and the structure of BPV1 E6 in complex with one of these motifs was solved by X-ray crystallography (Zanier *et al.*, 2013). The interaction between HPV16 E6 and the interferon-regulatory factor 3 (IRF3) inhibits its transcriptional activity (Ronco *et al.*, 1998) and occurs through the recognition of an LXXLL motif (PDB:6SJA; Poirson *et al.*, in preparation).

2.1.1.1 Objective

The present study aimed at deciphering the LXXLL interaction preferences of a set of diverse E6 oncoproteins by quantitative biophysical approach. We used independent complementary methods to accurately measure the affinities of E6 proteins from different HPVs representative of the phylogenetic and phenotypic diversity for an array of LXXLL peptide motifs originating from various host proteins. The selected E6

oncoproteins are expressed by HPV types belonging to the most represented genera (α, β and γ) and to the δ-genus BPV1. This set includes mucosal, cutaneous and high-risk HPV types.

2.1.1.2 My contribution

I attempted to isolated active E6 proteins from HPV 1, 3, 5, 8, 11 by the purification strategy that I previously developped (see Results section 1.1). I produced and purified the recombinant MBP-fused proteins 16E6 F47R 4C/4S, 18E6 F49R, 23E6 Whim variant, 38E6 C118A and 49E6 C8A following protocols previously established. I optimized the purification of 197E6 C79Y C89Q. I cloned triple fusion proteins, entailing the solubility-enhancing MBP carrier, E6 protein and LXXLL peptide. I produced the construct 18E6 F49R fused to the mutated E6AP LXXLL motif L8F E15R and additional constructs which could not be used for biophysical assay nor crystallography due to auto-association and solubility issues. I performed SPR and holdup assay on the E6 proteins that could be isolated under active form able to specifically interact with LXXLL peptides. I processed the data according to the methodology that I applied for the project detailed in Results section 1.2.

2.1.2 Material and Methods

2.1.2.1 Protein expression and purification

In order to limit the formation of intermolecular disulfide bridges, non-conserved cysteine residues exposed to the solvent were mutated. The solubility-enhanced mutant of HPV16 E6 (uniprot ID: P03126) includes 4 mutations of exposed cysteine residues (C80S; C97S; C111S and C140S). In addition, the mutation F47R prevents dimerization of the protein. Similarly, HPV18 E6 (uniprot ID: P06463) was mutated to prevent its dimerization (F49R). Solvent-exposed cysteines of the three other E6 presented in this study were mutated as follows: HPV38 E6 (uniprot ID: Q80907) (C118A); HPV49 E6 (uniprot ID: P36813) (C8A) and HPV197 E6 (C79Y; C89Q). A variant of HPV23 E6 was selected for the present study: with respect to the reference sequence (uniprot ID: P50776), it contains the substitutions H5R, T13N, S19L, T20P, H62N. In addition, a solvent-exposed cysteine residue was mutated for solubility-enhancement (C114V). BPV1 E6 was cloned without any mutation nor substitution (uniprot ID: P06931). All reference sequences of wild-type E6 proteins are available on PaVE database (pave.niaid.nih.gov).

E6 proteins from HPV 16, 18, 23, 38, 49 and BPV1 were fused to a mutant MBP (vector pETXM1) as previously described for crystallographic purpose (Zanier *et al.*, 2013). Sequence encoding for 197E6 was cloned in the NcoI Acc65I sites of pETM-41 vector. Triple fusion constructs (MBP-E6-LXXLL) were obtained by ligating the PCR fragment encoding for E6 protein with hybridized oligonucleotides encoding for the LXXLL peptides within the NcoI Acc65I sites of the vector pETXM1 encoding for N-terminal fusion to MBP.

Proteins were recombinantly expressed in *Escherichia coli* BL21 (DE3) grown in LB medium supplemented with glucose 0,2 % (w/v) and kanamycine 50 µg/mL. Bacteria were grown at 37°C until $OD_{600} \approx 0.7$, induced by adding 0.5 mM IPTG (Isopropyl β-D-1-thiogalactopyranoside) and $ZnSO_4$ 100 µM, incubated overnight at 16°C and harvested by centrifugation.

In order to avoid the formation of intermolecular disulfide bridges, all purification buffers were extensively degassed and equilibrated with argon before adding reducing agent.

MBP-E6 proteins from HPV 16, 18 (with and without fusion to E6AP LXXLL peptide mutant L8F E15R), 23, 38, 49 and BPV1 were purified as previously described (Zanier *et al.*, 2013). Briefly, bacterial pellets were resuspended in Buffer A (Tris HCl pH 8 50 mM, NaCl 400 mM, DTT or betamercaptoethanol 2 mM for 38E6) supplemented with glycerol 5 % (w/v), RNAse 0.25 µg/mL, DNAse 0.25 µg/mL, lysozyme at approximately 1 µg/mL and EDTA-free protease inhibitor cocktail (Roche) at the concentration recommended by the manufacturer. Cells were disrupted by three Microfluidizer cycles and centrifuged 45 min at 150,000x*g* 4°C. For MBP-mediated purification, supernatant was loaded on packed amylose resin (New England BioLabs), washed with buffer A and eluted with buffer B (Tris HCl pH 8 50 mM, NaCl 400 mM, DTT or betamercaptoethanol 2 mM, maltose 15 mM, protease inhibitor cocktail at 20 % of the recommended concentration). In order to eliminate MBP-E6 aggregates, the sample was then ultracentrifuged 15 h at 200,000x*g* 4°C in swing rotor. As a final step to isolate monomeric protein species, the supernatant was purified by size-exclusion chromatography in Buffer A on HiLoad 16/600 or 26/600 Superdex 200 columns (GE Healthcare).

E6 from HPV197 was purified by IMAC on nickel resin through its N-terminal 6His tag. Bacterial cells were disrupted according to the same protocol as described above (with 5 mM DTT and 10 mM imidazole added in the lysis buffer). After centrifugation 45 min at 150,000x*g* 4°C, the cleared lysate was loaded on HisTrap FF column (GE Healthcare). After washing at high salt (Buffer A2: Tris HCl pH 8 50 mM, NaCl 1 M, imidazole 10 mM, DTT 5 mM, protease inhibitor cocktail at 20 % of the recommended concentration), protein was gradient eluted from 10 to 300 mM imidazole (5 CV). Finally, protein eluted at 54 mM imidazole was injected in HiLoad 16/600 Superdex 200 column (GE Healthcare) for size-exclusion chromatography in Buffer A.

Small-scale batch purification on nickel resin was performed according to a protocol previously described (Bonhoure *et al.*, 2018).

2.1.2.2 Peptide synthesis

Biotinylated peptides were synthesized by either JPT Innovative Peptide Solutions with >70% purity or by peptide synthesis platform from IGBMC with > 80 % purity. They are all N-terminally conjugated to biotin via a TTDS spacer (N-(13-amino-4,7,10-trioxa-tridecayl)-succinamic acid) (Iris Biotech GMBH). The lyophilized peptides were resuspended in water at 5 mM and stored at -20°C.

2.1.2.3 *In vitro* protein-peptide interaction assays: holdup assay, surface-plasmon resonance and SPOT peptide arrays

The holdup chromatographic retention assay was performed as previously described, using microfluidics capillary electrophoresis for the quantification of unbound MBP-E6 (Bonhoure *et al.*, 2020). Similarly, the interaction assays by surface plasmon resonance and SPOT peptide arrays were performed based on previously published protocols (Bonhoure *et al.*, 2020).

2.1.3 Results

2.1.3.1 Selection and production of a pool of E6 proteins

Performing accurate quantitative binding assays requires protein samples with optimal quality, which may be challenging for the aggregation-prone E6 oncoproteins (Nominé *et al.*, 2001). E6 proteins contain at least 8 conserved cysteine residues required for the coordination of two zinc ions and several additional non-conserved cysteine residues that form intermolecular disulfide bridges in oxidative conditions. The resulting

oligomeric species can no longer interact specifically with LXXLL peptides since the hydrophobic binding pocket is not accessible. In order to maximize their solubility of recombinant E6 proteins, we used the same strategy as applied for structure resolution by X-ray crystallography (Zanier *et al.*, 2013). First, all E6 were expressed in fusion with the solubility-enhancer Maltose-Binding Protein (MBP). Second, we mutated the solvent-exposed cysteine residues with a rationale aiming at preserving the intrinsic protein fold while preventing artifactual disulfide bridges. Lastly, we explored different purification approaches in order to maximize the fraction of active protein. We tried to isolate active MBP-fused E6 protein from HPV 1, 3, 5, 8, 11, 16, 18, 23, 31, 38, 49 and 197. **Table 3** shows a summary of our attempts with different mutations and purification procedures. We tested wild-type as well as solubility-enhanced mutants. The starting point for the purification approach was based on a strategy used for crystallographic purpose, which consisted of an amylose affinity chromatography followed by an overnight ultracentrifugation eliminating a certain fraction of MBP-E6 oligomers. Size-exclusion chromatography was used as a final step for isolating active monomers. In addition, we made purification screenings by a small-scale purification approach, consisting of a nickel affinity performed in batch according to a previously published protocol (Bonhoure *et al.*, 2018). The criteria used for assessing the quality of the resulting protein sample were the monomeric state of the protein and its ability to bind LXXLL peptides. The optimization of the purification procedure was a determining factor for isolating active MBP-fused HPV197 E6 protein. We failed at purifying active MBP-E6 with the strategy optimized for crystallography. However, we obtained biologically active samples after small-scale nickel batch purification, which prompted us to attempt a scaled-up Nickel affinity chromatography followed by SEC. This last purification trial allowed us to obtain enough protein for probing the interaction preferences against the full array of LXXLL peptides. Most MBP-E6 proteins isolated as monomers could bind LXXLL motifs, except MBP-fused HPV11 E6 which did not bind any of the tested peptides. This observation may indicate that HPV11 E6 does not interact with isolated LXXLL motifs but binds other regions in the target protein, as demonstrated for the ubiquitin ligase E6AP (Drews *et al.*, 2020). Another explanation is that our peptide library lacked the preferential targets of HPV11 E6. We can hypothesize that we were able to purify active HPV11 E6 even though it did not interact with our LXXLL motifs.

HPV type	Genus	Species	Solubility-enhancing mutations	Expression vector	Expression in bacteria	Purification procedure	Recovery ratio (%) after overnight ultracentracentrifugation	Presence of monomeric protein		Binding to an LXXLL motif	
								y/n	By which assay ?	y/n	By which assay ?
3	α	2	C146S	pETM41	Expressed	BI-D	NA	n	DLS	n	SPR
11	α	10	wt	pETXM1	Expressed	A-U-S	27	y	SEC	n	holdup
11	α	10	C112S C144S	pETM41	Expressed	BI-D	NA	y	DLS	n	SPR
16	α	9	F47R C80S C97S C111S C140S	pETXM1	Expressed	A-U-S	62	y	SEC	y	holdup, SPOT, SPR
18	α	7	F49R	pETXM1	Expressed	A-U-S	ND	y	SEC	y	holdup, SPOT, SPR
31	α	9	C97S C111S	pETXM1	Expressed	A-U-S	ND	y	SEC	y	holdup, FP
5	β	1	C53/69V	pETM41	Not expressed						
8	β	1	wt	pETXM1	Expressed	A-U-S	23	n	SEC		
8	β	1	C52/74V	pETM41	Expressed	BI-D	NA	y	SEC	n	SPR
23	β	2	H5R T13N S19L T20P H62N C114V	pETXM1	Expressed	A-U-S	ND	y	SEC	y	SPR
38	β	2	C118A	pETXM1	Expressed	A-U-S	57	y	SEC	y	holdup, SPR
49	β	3	C8A	pETXM1	Expressed	A-U-S	93	y	SEC	y	holdup, SPR
197	γ	24	C79Y C89Q	pETXM1	Expressed	A-U-S	88	n	SEC		
197	γ	24	C79Y C89Q	pETM41	Expressed	BI-D	NA	y	SEC	y	SPR
197	γ	24	C79Y C89Q	pETM41	Expressed	I-S	NA	y	SEC	y	holdup, SPR
BPV 1	δ	4	wt	pETXM1	Expressed	A-U-S	52	y	SEC	y	holdup, SPOT
1	μ	1	C16N	pETM41	Expressed	BI-D	NA	n	SEC		
1	μ	1	C16N	pETM41	Expressed	BI-D	NA	NA	NA	n	SPR
1	μ	1	wt	pETXM1	Expressed	A-U-S	1	n	SEC	n	holdup

Table 3 : Summary of HPV E6 purification attempts. The HPV type, genus and species are indicated according to phylogenetic classification. We used two expression vectors: one encoding for N-terminal fusion to an MBP mutant optimized for crystallization (pETXM1), the second one encoding for N-terminal fusion to 6His-MBP-TEVsite which enables Nickel affinity chromatography and cleavage of the purification tags by TEV protease. All recombinant proteins were produced in *E.coli* BL21(DE3) following the same expression protocol. We tried three purification strategies: the A-U-S: Amylose affinity - Ultracentrifugation - SEC that was previously described for crystallographic purpose (Zanier *et al.*, 2013), the BI-D: Batch IMAC nickel affinity - Desalting, which is a small-scale purification used for screening (Bonhoure *et al.*, 2018) and the I-S: IMAC Nickel affinity - SEC was used for isolating active samples of HPV197 E6 in sufficient amount for holdup assay. The recovery ratio is indicated in percentage, 0 % meaning that the protein entirely aggregated during centrifugation. The monodispersity of the purified protein was assessed by either Dynamic Light Scattering (DLS) or Size-Exclusion Chromatography (SEC). The binding to LXXLL motifs was probed by holdup chromatographic retention assay, SPOT peptide array or Surface-Plasmon Resonance (SPR). NA stands for Not Applicable, ND stands for Not Determined and dashed lines indicate that the final sample was not suitable for interaction assay, due to the absence of monomeric protein, the absence of binding to LXXLL motifs or the impossibility to purify the protein due to an absence of expression in *E.coli*.

Despite all our efforts, we were able to produce suitable samples for only half of the HPV E6 proteins that we attempted to isolate, namely HPV 16, 18, 23, 31, 38, 49, 197. In addition to this set of E6 proteins from various HPVs, we tried to isolate E6 from papillomaviruses used as models, such as cottontail rabbit papillomavirus (CRPV) and bovine papillomavirus type 1. We successfully isolated active samples of the latter, which we included in our collection of E6 proteins.

2.1.3.2 Selection and synthesis of a pool of LXXLL motifs

Before the beginning of my thesis, an array of 46 LXXLL motifs from cellular proteins was designed by my supervisor Gilles Travé and Khaled Ould Babah, a former PhD student. This library was developped in order to decipher E6 interaction preferences for diverse peptide sequences (**Table 4**). The rationale was to investigate whether the interaction between E6 and some of its published or putative protein targets was mediated by LXXLL motifs. To this aim, the first selected targets were host proteins hijacked by E6 oncoproteins that were reported in the literature. PDZ domain-containing proteins were excluded, in order to rule out any PBM-mediated interaction and better focus on interactions exclusively driven by LXXLL motifs. Some putative targets were added to the list of cellular proteins: they were mainly chosen because of their similarities with published targets. For instance, one homolog of the tumor suppressor p53 (*Bos taurus*) was included to complement the human motif that was identified in pulldown experiments using E6 oncoproteins as bait. In addition, cellular proteins that were not published as E6 targets but that were good candidates (regarding their interesting peptide motifs) were included. The second step consisted in identifying potential E6 binding motifs in the sequences of the selected proteins. The motif search was inclusive: the putative interaction motifs should correspond to the consensus LXXLL expanded to ΦXXΦΦ, with the residues V, I, F, W, Y or M replacing whole or part of the three L residues from the consensus. Neighbouring residues with an acidic side chain (D, E) were considered as a favorable context for E6 binding. The manual identification of binding motifs corresponding to this consensus sequence was not systematic. For better accessibility of the binding sequence, most motifs were identified in disordered regions predicted using IUPred. The majority of the peptide motifs are 12-mers. Some motifs were extended up to 17 residues. In the case of p53bos_1 and p53bos_2, designing these longer peptides allows one to evaluate the contribution of the neighbouring residues, as they might increase the affinity of an E6

protein for the peptide. In the case of E6AP peptide, the published structure of HPV16 E6 / $E6AP_{LXXLL}$ complex showed some contacts with the N-terminus of the peptide (Zanier *et al.*, 2013), suggesting that the 4 N-terminal residues might participate to the interaction. For better identification of E6AP-binding E6 proteins, a 16-mer peptide motif was designed as it entails most of the peptide residues that are potentially recognized by an E6 oncoprotein.

Some peptides of our library did not detectably interact with the purified E6 proteins. Hence, while **Table 4** presents the list of all tested peptides, the next figures only entail the peptides that bound at least one of the tested E6 above the significance threshold (BI>0.20).

Peptide name	Peptide sequence	Peptide length	Position of the motif	Disorder score	E6 binding? *y/n*	Protein name	Uniprot ID	Reference
ADA3_1	QLELETLLSSAS	12	47-58	0.124	n	Transcriptional adapter 3	O75528	(Kumar *et al.*, 2002; Zeng *et al.*, 2002)
ADA3_2	VRTLEELLKPPE	12	170-181	0.489	n			
AP1B1	GGGLDSLMGDEP	12	660-671	0.478	y	AP-1 complex subunit beta-1	Q10567	(Tong *et al.*, 1998)
AP1G1_1	GGELLDLLGDIN	12	653-664	0.429	y	AP-1 complex subunit gamma-1	O43747	(Tong *et al.*, 1998)
AP1G1_2	ANDLLDLLGGND	12	625-636	0.672	y			
BRCA1_1	IQKVNEWFSRSD	12	379-390	0.350	n	Breast cancer type 1 susceptibility protein	P38398	(Zhang *et al.*, 2005)
BRCA1_2	FSRSDELLGSDD	12	386-397	0.467	n			
BRCA1_3	ERTLKYFLGIAG	12	1698-1709	0.018	n			
CBP	HKQLSELLRGGS	12	67-78	0.527	n	CREB-binding protein	Q92793	(Patel *et al.*, 1999; Zimmermann *et al.*, 1999)
DBC1	KSQLQRLLQELR	12	882-893	0.308	n	Deleted in breast cancer gene 1 protein	Q8N163	
DNJA1	LSLLEKLLPERK	12	339-350	0.312	y	DnaJ homolog subfamily A member 1	P31689	(Rozenblatt-Rosen *et al.*, 2012)
E6AP	PESSELTLQELLGEER	16	402-417	0.670	y	E6-associated protein	Q05086	(Brimer *et al.*, 2017; Grace and Münger, 2017; Rozenblatt-Rosen *et al.*, 2012; White *et al.*, 2012a)
E6BP	GFVSLEEFLGDY	12	204-215	0.306	y	Reticulocalbin-2	Q14257	(Chen *et al.*, 1995; Grace and Münger, 2017)
E6TP1	PTKLSDFLITGG	12	240-251	0.473	n	E6-targeted protein 1	O43166	(Gao *et al.*, 1999)
EBP1	DAELKALLQSSA	12	351-362	0.473	n	ErbB3-binding protein 1	Q9UQ80	
Emerin	DTELTTLLRRYN	12	9-20	0.270	n	Emerin	P50402	
Gps2	FLQLKKVLHEEE	12	87-98	0.300	n	G protein pathway suppressor 2	Q13227	(Degenhardt and Silverstein, 2001)
IRF3	EDILDELLGNMV	12	137-148	0.631	y	Interferon regulatory factor 3	Q14653	(Grace and Münger, 2017; Ronco *et al.*, 1998)
MAML1	MSDLDDLLGSQ	11	1006-1016	0.601	y	Mastermind-like protein 1	Q92585	(Brimer *et al.*, 2017; Grace and Münger, 2017; Tan *et al.*, 2012; White *et al.*, 2012a)
MED13	YTDLDNLFNSDE	12	780-791	0.430	y	Mediator complex subunit 13	Q9UHV7	
p300	QQALQNLLRTLR	12	2048-2059	0.451	n	Histone acetyltransferase p300	Q09472	(Muench *et al.*, 2010; White *et al.*, 2012a)
p53	FSDLWKLLPENN	12	19-30	0.412	y	Cellular tumor antigen p53	P04637	(Grace and Münger, 2017; White *et al.*, 2012a, 2014)
p53bos_1	FSDLWNLLPENN	12	19-30	0.307	y	Cellular tumor antigen p53 *Bos taurus*	P67939	
p53bos_2	SQETFSDLWNLLPENN	16	15-30	0.319	y			

PAB1	DNSELLHMLESP	12	589-600	0.286	n	Polyadenylate-binding protein 1	P11940	
Piasy_1	VSDLQMLLGFVG	12	17-28	0.054	y	Protein inhibitor of activated STAT protein gamma	Q8N2W9	(Bischof *et al.*, 2006)
Piasy_2	FNMLDELLKPTE	12	139-150	0.153	y			
PXN_1	MDDLDALLADLE	12	1-12	0.568	y	Paxillin	P49023	(Grace and Münger, 2017; Vande Pol *et al.*, 1998)
PXN_2	LSELDRLLLELN	12	142-153	0.423	y			
PXN_4	TRELDELMASLS	12	263-274	0.413	y			
SGT1	PDQLDQLLQEAV	12	420-431	0.569	y	Human suppressor of GCR two	O95905	
TGFB1_1	MEDLDALLSDLE	12	1-12	0.814	y	Transforming growth factor beta-1-induced transcript 1 protein	O43294	(Grace and Münger, 2017)
TGFB1_2	LCELDRLLQELN	12	90-101	0.256	y			
TGFB1_3	TLELDRLMASLS	12	155-166	0.439	y			
TRIP-Br1_1	SASMASLLEDLS	12	112-123	0.310	n	Transcriptional regulator interacting with the PHD-bromodomain 1	Q9UHV2	(Gupta *et al.*, 2003)
TRIP-Br1_2	DDGLEGLFEDID	12	167-178	0.298	n			
TRIP-Br1_3	EAELDYLMDVLV	12	212-223	0.575	n			
TRIP-Br1_4	LDYLMDVLVGTQ	12	215-226	0.570	y			
TRIP-Br1_5	EAELDYLMDVLVGTQ	15	212-226	0.581	y			
TSC2_1	LQTLQDILGDPG	12	1391-1402	0.668	y	Tuberous sclerosis 2 protein	P49815	(Lu *et al.*, 2004)
TSC2_2	ELQTLQDILGDP	12	1390-1401	0.670	y			
WIZ	INILQELLATSA	12	807-818	0.338	y	Widely-interspaced zinc finger-containing protein	O95785	
XRCC1	PEELGKILQGVV	12	313-324	0.394	n	X-ray repair cross-complementing protein 1	P18887	(Iftner *et al.*, 2002)
Zyxin	IDSLSSLLDDMT	12	148-159	0.495	y	Zyxin	Q15942	(Rozenblatt-Rosen *et al.*, 2012)

Table 4 : LXXLL motifs of HPV E6 putative cellular targets. The abbreviated names of the cellular target proteins are followed by a number when several motifs were identified in the same protein. In the peptide sequence, the residues corresponding to the consensus ΦXXΦΦ are indicated in red. The position of the motif is indicated based on the canonical protein sequence, as defined on UniProt database. The disorder score is the mean IUPred value for the motif sequence: it ranges from 0 to 1 and a score above 0.5 indicates disorder. Some of these motifs never detectably bound any of the E6 oncoproteins that we tested by holdup, as indicated on the table by the column "E6 binding?".

2.1.3.3 Affinity profiling of seven papillomavirus E6 oncoproteins for LXXLL peptide motifs

2.1.3.3.1 Cellular LXXLL motifs peptide array

We deciphered the interactions preferences of seven E6 oncoproteins for the peptide library presented in **Table 4**. We tested the LXXLL interaction preferences of the MBP-fused solubility-enhanced mutants 16E6 F47R 4C/4S, 31E6 C97S C111S, 18E6 F49R, 38E6 C118A, 49E6 C8A, 197E6 C79Y C89Q and BPV1 E6. For sake of clarity, the MBP-fused E6 proteins will hereafter be designated without mentioning the solubility-enhancing mutations, that is to say MBP-16E6, MBP-31E6, MBP-18E6, MBP38E6, MBP-49E6, MBP-197E6 and MBP-BPV1E6.

Figure 27 displays the so-called binding profiles resulting from the holdup assay. The plots present the values resulting from the holdup assay, which are called binding intensities (BI). They are calculated from analyte depletion with the tested peptide as compared to a negative control. The BI reflects the strength of each E6/LXXLL interaction, as they range from 0 (no interaction) to 1 (strong interaction). On the left side of the figure, a distance tree shows the sequence similarities between the tested E6 proteins: as it coincides with the phylogenetic classification of the concerned HPV types, the genus and species of each HPV E6 protein are indicated. Note that the set of tested E6 oncoproteins is phylogenetically diverse as the three most common HPV genera are represented together with the δ-genus BPV1. The peptides were ranked by decreasing affinity with MBP-16E6. The overview of the seven interaction profiles first indicates that all tested MBP-E6 had very distinct interaction preferences within the peptide array. They differ in several aspects: i) the subset of peptide motifs which they recognize and their ranking (preference), ii) the number of peptides for which significant binding is reached (specificity) and iii) the strength of their interactions (affinity). Regarding the selectivity of the MBP-E6 for LXXLL motifs, it is important to stress that our data give an indication of E6 interaction preferences within a limited peptide library that is not representative of all available motifs from the proteome. The peptide library that we constituted for this study entails the main E6 protein targets reported in the literature and some putative target motifs. It is thus informative for the most significant targets identified to date. Nonetheless, the E6 proteins tested in our study may have additional protein targets that were not yet reported nor included in the peptide array.

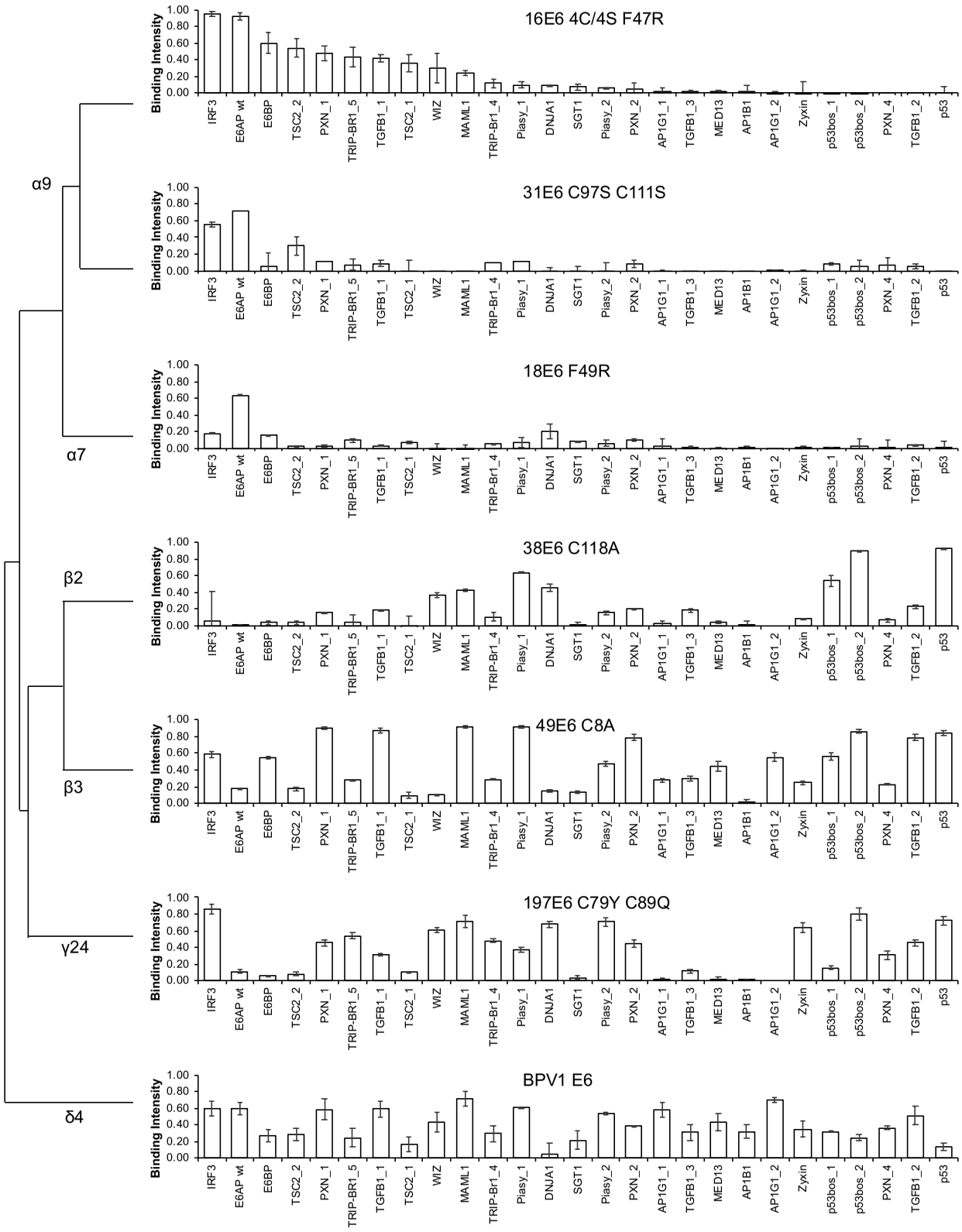
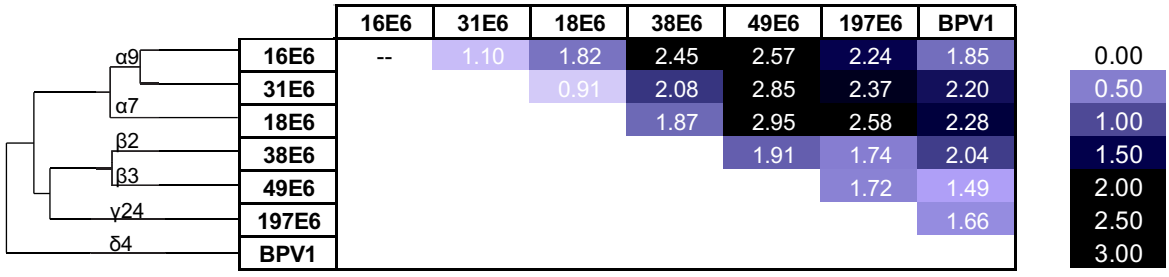


Figure 27 : Holdup binding profiles of seven MBP-fused E6 proteins for an array of LXXLL motifs from putative cellular targets. On the right side, an average distance tree was generated based on the protein sequences of the solubility-enhanced mutants of E6 oncoproteins. The genera and species of each type of papillomavirus are indicated on the distance tree. The binding intensities are proportional to the affinity, ranging from 0 (no interaction) to 1 (highest quantifiable affinity). For all interaction profiles, the LXXLL peptides were ranked by decreasing affinity with 16E6 4C/4S F47R.

One could expect the binding preferences of MBP-31E6 and MBP-18E6 to be similar to MBP-16E6. It is the case for E6AP and IRF3 LXXLL motifs: they have the highest affinities for MBP-16E6, they both interact with MBP-31E6 and E6AP LXXLL motif is recognized by MBP-18E6 (**Figure 27**). As MBP-31E6 and MBP-18E6 bind very few LXXLL peptides of the tested library, one could argue that their biochemical quality does not allow them to interact with LXXLL motifs, that is to say that the two proteins are present under the form of inactive oligomers. Another explanation is that the tested peptide library entails only few peptides that match their binding preferences. We tested a second peptide library made of single-point mutants of E6AP LXXLL motif that will be further detailed in section 2.1.3.3.2. When tested with this second library, MBP-31E6 and MBP-18E6 reached significant interaction with many LXXLL peptides (**Figure 29**). This observation supports the hypothesis that the purified MBP-31E6 and MBP-18E6 were active and able to interact with LXXLL motifs as long as the peptide library entails their preferred peptide motifs. Our results suggest that MBP-31E6 and MBP-18E6 bind specifically E6AP and IRF3 LXXLL motif. However, one cannot rule out that they may bind other motifs that are not represented within our peptide library, as stated in the previous paragraph.



	16E6	31E6	18E6	38E6	49E6	197E6	BPV1
16E6	--	1.10	1.82	2.45	2.57	2.24	1.85
31E6			0.91	2.08	2.85	2.37	2.20
18E6				1.87	2.95	2.58	2.28
38E6					1.91	1.74	2.04
49E6						1.72	1.49
197E6							1.66
BPV1							

Table 5: Euclidean distances between the seven binding profiles of MBP-fused E6 proteins determined by the holdup assay (presented in **Figure 27**). At the left of the table, a phylogenetic tree entailing the genus and species of each tested HPV E6 oncoprotein is displayed. The color code used to highlight the Euclidean distances values is presented on the right of the table. Note the low values for MBP-31E6/MBP-18E6 comparison, which indicates that both proteins have similar binding profils.

Table 5 presents the Euclidean distance between each pair of binding profiles, based on the holdup data presented in **Figure 27**. This pairwise comparison emphasizes that E6 proteins belonging to the same genus have related binding preferences, for instance the α-genus MBP-16E6, MBP-18E6 and MBP-31E6. On the contrary, the highest Euclidean distance is reached for MBP-18E6/ MBP-49E6 comparison, which is consistent as both proteins do not share any common target peptide. Interestingly,

E6 proteins could be split into two groups based on the proximity of their binding profiles: on the one hand, the α-genus high-risk mucosal MBP-16E6, MBP-18E6 and MBP-31E6 and on the other hand, the β and ɣ-genus cutaneous MBP-38E6, MBP-49E6 and MBP-197E6. The lower distance between MBP-BPV1E6/MBP-16E6 may suggest that MBP-BPV1E6 is intermediate between those two groups. Indeed, MBP-BPV1E6 bound E6AP LXXLL motif at detectable level (BI = 0.60 ± 0.09), unlike MBP-38E6, MBP-49E6 and MBP-197E6.

As noted in the previous paragraph, the comparison of the binding profile of the α-genus MBP-16E6 versus the β-genus MBP-38E6 and MBP-49E6 reveals a clear reshuffling of their binding preferences. While E6AP LXXLL motif appears as the preferential target of α-genus HPV E6 proteins (Brimer *et al.*, 2017), it does not interact with MBP-38E6 and its interaction with MBP-49E6 is below the significance threshold fixed at 0.2. MBP-38E6 and MBP-49E6 share common peptide targets, for instance the LXXLL motifs of human and bovine p53 (p53 and p53_bos2), MAML1 and piasy_1. Beside their similar preferences, one can note than MBP-49E6 binds more LXXLL motifs within the peptide library than MBP-38E6. In addition, MBP-49E6 bound piasy_1 and MAML1 LXXLL motifs with higher affinity than MBP-38E6: BI(MBP-38E6/MAML1) = 0.42 ± 0.02 versus BI (MBP-49E6/MAML1) = 0.72 ± 0.07.

The ɣ-genus MBP-197E6 also target some of the β-genus preferential targets, such as p53, p53_bos2 and MAML1. In addition, MBP-49E6 and most particularly MBP-197E6 target IRF3 LXXLL motif.

Concerning the interaction profile of MBP-BPV1E6, one can note that its overall affinity for the tested LXXLL motifs is lower than the β and ɣ-genus MBP-E6. Its highest BI was reached for MAML1: BI(MBP-BPV1E6/MAML1) = 0.72 ± 0.06. In addition, **Figure 27** shows that MBP-BPV1E6 is the protein that bound most peptides in the library as compared to all seven tested MBP-E6. As expected based on the literature (Vande Pol *et al.*, 1998), MBP-BPV1E6 interacts with PXN_1, PXN_2 and PXN_4, which originate from its prototypical target, paxillin. The BI between MBP-BPV1E6 and the three paxillin motifs range from 0.4 to 0.6. The conversion of these values into ΔG free energy revealed that at T= 298K they correspond to ΔG(MBP-BPV1E6/PXN_1) = -7.2 ± 0.5 kcal.mol^{-1} and ΔG(MBP-BPV1E6/PXN_2) = ΔG(MBP-BPV1E6/PXN_4) = -6.6 ± 0.5 kcal.mol^{-1} (**Figure 28**). The associated dissociation constant for PXN_1 is K_D(MBP-

BPV1E6/PXN1) = 5.7 µM. This value constrats with the previously published affinity for BPV1 E6 and the first LXXLL motif of paxillin, which was determined by SPR at 35.1 nM (Ould M'hamed Ould Sidi *et al.*, 2011). The difference between these two values can be explained by two factors. First, the peptide sequence of the paxillin motif slightly differs between the two studies: while our peptide sequence was MDDLDALLAD**L**E, the previous study was performed using a peptide with the sequence DDLDALLAD**K**E. The presence of a lysine residue instead of a leucine may enhance the affinity by adding a hydrogen bond with the positively charged side chain of lysine. Unfortunately, the crystal structure of BPV1 E6 in complex with paxillin LXXLL motif was solved using a third peptide sequence MDDLDALLAD that does not entail the substituted residue (Zanier *et al.*, 2013), impeding the verification of this first hypothesis. The second factor that could explain the higher affinity measured in Ould M'hamed Ould Sidi *et al.* is the salt concentration that was set at 200 mM in the former SPR study versus 300 mM in the present holdup assay. As assayed in (Zanier *et al.*, 2005), the affinity of an E6/LXXLL interaction is dependent on ionic strength. Hence, the variation of peptide sequence coupled to difference buffer composition may explain the higher K_D value estimated in our study.

These rather low affinities for each isolated peptide can be compensated when BPV1 E6 interacts with full-length paxillin. The presence of several binding motifs within the same protein target may increase the overall affinity by an avidity effect. Interestingly, MBP-49E6 binds PXN_1 and PXN_2 at higher affinity than MBP-BPV1E6. MBP-49E6 also bound the corresponding motifs within HIC5, the homolog of paxillin (TGFB1_1 and TGFB1_2). As displayed on **Figure 28**, the sequences of PXN_1 and PXN2 only differ from TGFB1_1 and TGFB1_2 by 2 residues, which explains that the homologous motifs are recognized by the same MBP-E6 protein at comparable affinities.

In our study, we included the LXXLL motif of human p53 (p53) and two peptides from the same motif in bovine p53 (p53_bos1 and p53_bos2). While p53 and p53_bos1 only differ by a single residue, this substitution is sufficient to impact drastically its recognition by E6 oncoproteins. MBP-BPV1E6 does not interact with human p53 peptide but binds p53_bos1 and p53_bos2 with similar affinities (-6.4 and -6.2 ± 0.5 $kcal.mol^{-1}$, respectively). On the contrary, the K/N substitution from p53 to p53_bos1 abolished the interaction with MBP-197E6 and lowered the affinity of MBP-38E6 and MBP-49E6. Interestingly, adding four residues SQET in N-terminal (p53_bos2) could

compensate the amino acid substitution and restore the interaction with MBP-38E6, MBP-49E6 and MBP-197E6.

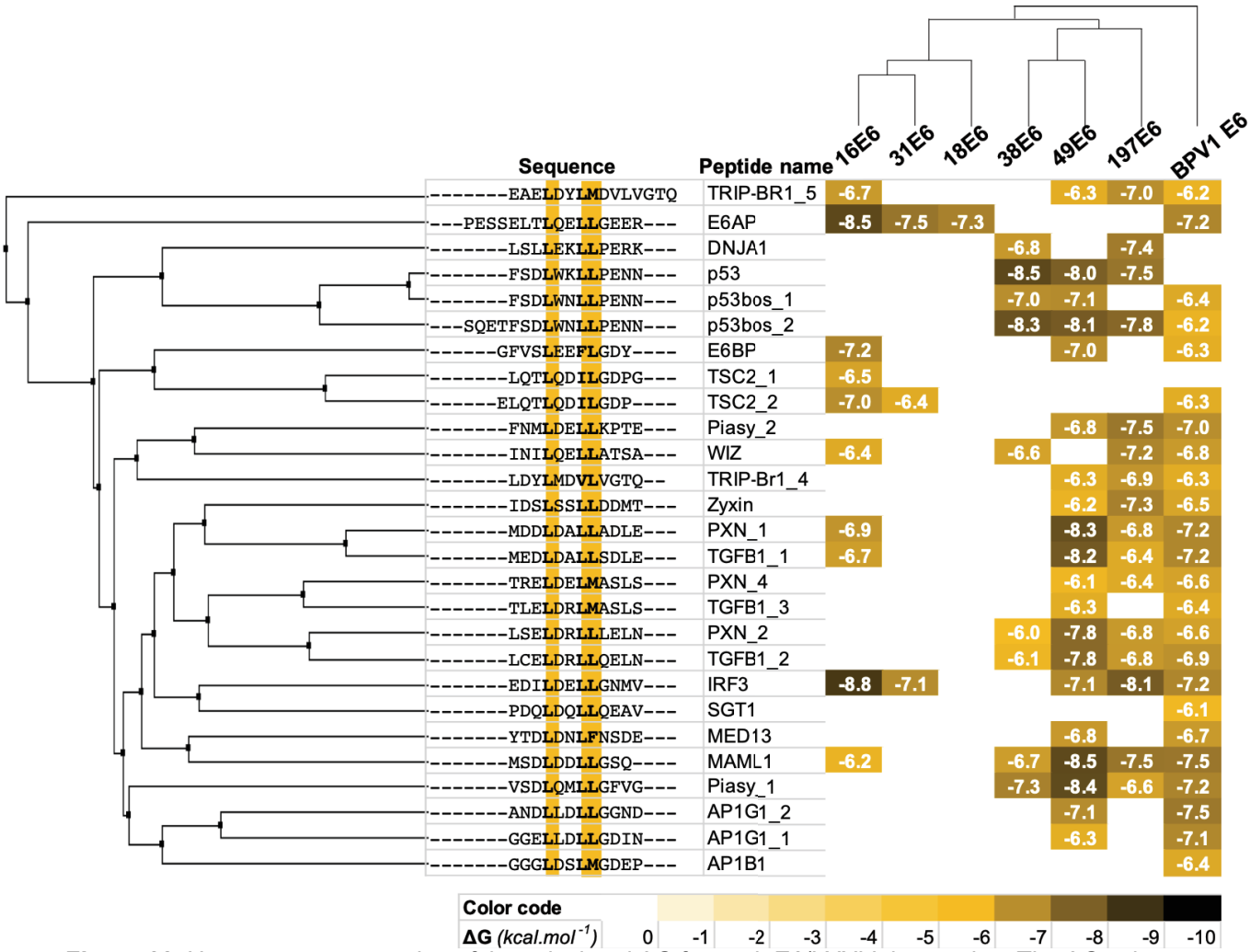


Figure 28: Heatmap representation of the calculated ΔG for each E6/LXXLL interaction. The ΔG values were estimated by calibrating holdup BI with three interactions quantified by SPR, as previously described (Bonhoure *et al.*, 2020). The color code of the heatmap for the corresponding ΔG values (in kcal.mol^{-1}) is at the bottom. Both MBP-fused E6 proteins and LXXLL peptides were aligned according to their amino acid sequence and an average distance tree is displayed on the figure. For sake of clarity, the solubility-enhancing mutations of each MBP-fused E6 protein are not indicated on the figure: 16E6 F47R 4C/4S; 31E6 C97S C111S; 18E6 F49R; 38E6 C118A; 49E6 C8A; 197E6 C79Y C89Q. The mean uncertainty was estimated at 0.5 kcal.mol^{-1}.

In order to estimate the significance of our *in vitro* results in a cellular context with full-length proteins, we compared our fragmental protein-motif interaction data with published AP-MS studies. **Table 6** presents a summary of the E6 interaction partners reported in (Eckhardt *et al.*, 2018; Grace and Münger, 2017; Rozenblatt-Rosen *et al.*, 2012; Wang *et al.*, 2016; White *et al.*, 2012a) compared with the ΔG that we estimated. To date, the human proteins targeted by 49E6 have not been investigated by proteomics. HPV76 belongs to the same β3 species as HPV49. The protein 76E6 was used as a bait in a study published in 2012 (White *et al.*, 2012a). Since 76E6 and 49E6

share 80 % identity in their protein sequence, we assume their binding preferences to be similar.

While p53 was co-purified with the α-genus 16E6, 31E6 and 18E6, we did not detect any direct interaction between these E6 proteins and the LXXLL motif within p53. This observation confirms that high-risk mucosal HPV E6 proteins do not interact with p53 by capturing its LXXLL motif but recruit E6AP ubiquitin ligase to form a tripartite complex in which E6 binds p53 by an interface located in its core domain (Martinez-Zapien *et al.*, 2016). The co-purification of both E6AP and p53 is consistent with this mode of interaction.

E6-binding protein (E6AP) is also called Reticulocalbin-2 (RCN2) or Calcium-binding protein ERC-55. This protein was reported to interact with high-risk HPV E6 and BPV1 E6 according to a pull-down assay published in 1995 (Chen *et al.*, 1995). Our results show a binding of E6BP LXXLL motif to MBP-16E6, MBP-BPV1E6, MBP-49E6 but not with MBP-31E6.

Paxillin and its homolog HIC5 (also named TGFB1) were co-purified with 197E6 but not with 76E6, which is similar to 49E6. Given that 49E6 bound LXXLL motifs within paxillin and HIC5 at higher affinity than 197E6, one could expect these protein targets to be co-purified with 76E6. The slight differences in the amino acid sequence of 76E6 and 49E6 may partly explain this result. In addition, the study reporting 197E6 target proteins (Grace and Münger, 2017) was performed in distinct experimental conditions than the publication reporting 76E6 target proteins (White *et al.*, 2012a). In particular, the washing steps can be critical since harsh washing conditions can dissociate transient complexes and prevent the identification of a certain number of interacting proteins.

16E6		31E6		18E6		38E6		49E6/76E6		197E6		
Protein target	ΔG (kcal.mol^{-1})	Protein target	ΔG (kcal.mol^{-1})	Protein target	ΔG (kcal.mol^{-1})	Protein target	ΔG (kcal.mol^{-1})	Protein target	ΔG (kcal.mol^{-1})	Protein target	ΔG (kcal.mol^{-1})	
E6AP	**-8.5**	E6AP	**-7.5**	E6AP	**-7.3**					E6AP	0.0	
p53	0.0	p53	0.0	p53	0.0	p53	**-8.5**			p53	**-7.5**	
E6BP	**-7.2**	E6BP	0.0			E6BP	0.0	E6BP	**-7.0**			
						MAML1	**-6.7**	MAML1	**-8.5**	MAML1	**-7.5**	
										IRF3	**-8.1**	
										Paxillin	PXN_1	**-6.8**
											PXN_2	**-6.8**
											PXN_4	**-6.4**
										TGFB1	TGFB1_1	**-6.4**
											TGFB1_2	**-6.8**
											TGFB1_3	0.0

Color code

ΔG (kcal.mol^{-1}) 0 -1 -2 -3 -4 -5 -6 -7 -8 -9 -10

Table 6: Comparison with published protein targets from AP-MS studies. References: (White *et al.*, 2012a) for 16E6, 18E6, 38E6, 76E6 which is similar to 49E6; (Rozenblatt-Rosen *et al.*, 2012) for 16E6 and 18E6; (Wang *et al.*, 2016); (Eckhardt *et al.*, 2018) for 31E6 and (Grace and Münger, 2017) for 197E6 .

As previously reported, β and ɣ-genus 38E6, 76E6 and 197E6 interact with MAML1 (Brimer *et al.*, 2017). E6AP was co-purified with 197E6, however our results indicate no detectable interaction between 197E6 and the LXXLL motif of E6AP. Our results tend to confirm that the LXXLL motifs of E6AP and MAML1 are preferentially targeted by α-genus mucosal HPV E6 and β-genus cutaneous HPV E6, respectively. In our set of purified MBP-E6, MBP-BPV1E6 was the only oncoprotein to bind both E6AP and MAML1 at equivalent affinity.

Interestingly, IRF3 was co-purified with 197E6 but not with 16E6. The first study reporting the interaction between 16E6 and IRF3 and its effect of inhibiting IRF3 transcriptional activity was published in 1998 (Ronco *et al.*, 1998). Our results confirm the interaction between 16E6 and IRF3 LXXLL motif and our team participated to solve the crystal structure of 16E6/IRF3 LXXLL motif complex, which reinforces the validity and the significance of IRF3 as a target of HPV E6 proteins.

Our results provide reliable quantitative information that support and complement the previously published interaction data. Using a fragmental approach, we confirmed that LXXLL motifs drive the recognition of several host target proteins by HPV E6 oncoproteins.

2.1.3.3.2 Single-point mutants of E6AP LXXLL motif

The second peptide library that we used for assessing the interaction preferences of our set of MBP-E6 proteins entails single-point mutants of the LXXLL motif of E6AP ubiquitin ligase. Like the cellular motifs peptide array, the array of E6AP mutants was designed by Gilles Travé and Khaled Ould Babah (**Figure 29**). It previously allowed the identification of the sequence determinants ruling the interaction of MBP-16E6 with its prototypical target (Bonhoure *et al.*, 2020). Comparing several mutants of the motif can lead to the identification of affinity-increasing mutations. In the present study, we tested the interactions of 31E6 and 18E6 proteins that were previously reported to bind the wild-type LXXLL motif of E6AP (hereafter named E6APwt) (Eckhardt *et al.*, 2018; White *et al.*, 2012a). We also included E6 from BPV1 and 49E6. The interaction profiles of the three α-HPV E6 proteins (HPV16, 18, 31) present some similarities. 16E6 has the highest affinity for E6APwt, while 31E6 and 18E6 are in a lower range. This decrease of affinity for E6APwt is correlated with an increased sensitivity to the tested single-point mutations. Over a total of 45 tested mutants, 8 abolished the interaction

with 16E6, 17 with 31E6 and 28 with 18E6 (**Figure 29**). As a comparison, only 6 mutants abolished completely the interaction of BPVE6, which is phylogenetically very distant from high-risk HPVs. Hence, this second peptide array confirms that each tested E6 has its own interaction preferences.

Including 49E6 to the study is particularly informative since it does not detectably bind E6APwt. The holdup binding intensity was 0.19 ± 0.03, which is below the threshold of interaction significance that is fixed at 0.2. Several mutations increased the affinity of 49E6 for E6AP LXXLL motif. The highest affinity was reached with L8W mutant, at a $\Delta\Delta G$ value of -8.3 ± 0.5 kcal.mol^{-1} and the interaction was confirmed by SPR (data not shown). At a lower extent, the substitution of the first leucine residue of the consensus LXXLL (L8) by other hydrophobic residues still increases the affinity for 49E6, for instance with isoleucine (L8I), phenylalanine (L8F) and methionine (L8M). Similarly, the mutation of the last leucine of the consensus into a phenylalanine residue increases the affinity of 49E6 (L12F). By comparing all tested E6 proteins, we note that some mutations on the consensus leucine residues are tolerated with a limited impact on the affinity, for instance L8F, L8W and L11I. This observation may lead to a certain reconsideration of the consensus, which appears more flexible than three leucine residues interspaced by random amino acids. Recently, a publication reported that BPV1 E6 was able to bind peptides entailing LXXLF consensus motif instead of LXXLL, based on phage display selections (Brimer *et al.*, 2017). Hence, our observation that some E6 proteins can select other hydrophobic residues than leucine in the consensus LXXLL is consistent with recently published data.

		16E6	31E6	18E6	49E6	BPV1 E6
PESSELTLQELLGEER	wt	-8.5	-7.5	-7.3	0.0	-7.2
PESSALTLQELLGEER	E5A	1.3	1.3	7.3	-7.1	-0.1
PESSRLTLQELLGEER	E5R	2.1	7.5	7.3	-7.4	-0.1
PESSELALQELLGEER	T7A	1.4	1.3	1.2	-7.3	-0.4
PESSELTAQELLGEER	L8A	8.5	7.5	7.3	0.0	7.2
PESSELTVQELLGEER	L8V	8.5	7.5	7.3	0.0	0.8
PESSELTIQELLGEER	L8I	2.4	7.5	7.3	-6.3	0.5
PESSELTFQELLGEER	L8F	0.6	0.7	-0.7	-6.5	0.6
PESSELTWQELLGEER	L8W	0.4	0.4	0.6	-8.3	0.2
PESSELTYQELLGEER	L8Y	2.3	7.5	7.3	0.0	0.8
PESSELTMQELLGEER	L8M	1.6	7.5	-0.3	-6.7	0.4
PESSELTSQELLGEER	L8S	8.5	7.5	7.3	0.0	0.8
PESSELTLAELLGEER	Q9A	-0.6	0.2	0.1	-7.2	-0.2
PESSELTLEELLGEER	Q9E	-0.2	0.6	1.1	-6.7	0.0
PESSELTLDELLGEER	Q9D	0.0	0.8	1.0	-7.4	0.0
PESSELTLQALLGEER	E10A	0.7	0.4	0.3	-7.1	0.0
PESSELTLQDLLGEER	E10D	1.0	0.9	1.2	-6.8	0.2
PESSELTLQQLLGEER	E10Q	0.3	0.4	-0.3	-7.2	0.0
PESSELTLQRLLGEER	E10R	1.7	0.9	0.7	-6.9	0.3
PESSELTLQEALGEER	L11A	8.5	7.5	7.3	0.0	7.2
PESSELTLQEVLGEER	L11V	2.2	7.5	7.3	0.0	0.8
PESSELTLQEILGEER	L11I	0.4	0.8	0.1	0.0	0.5
PESSELTLQEFLGEER	L11F	2.2	7.5	7.3	0.0	0.8
PESSELTLQEWLGEER	L11W	8.5	7.5	7.3	0.0	7.2
PESSELTLQEYLGEER	L11Y	8.5	7.5	7.3	0.0	7.2
PESSELTLQEMLGEER	L11M	2.3	7.5	7.3	0.0	1.0
PESSELTLQELAGEER	L12A	8.5	7.5	7.3	0.0	7.2
PESSELTLQELVGEER	L12V	1.6	7.5	7.3	0.0	7.2
PESSELTLQELIGEER	L12I	0.5	0.8	7.3	0.0	0.9
PESSELTLQELFGEER	L12F	2.1	1.0	7.3	-6.8	-0.5
PESSELTLQELWGEER	L12W	8.5	7.5	7.3	0.0	0.5
PESSELTLQELYGEER	L12Y	2.3	7.5	7.3	0.0	0.2
PESSELTLQELMGEER	L12M	2.1	1.0	7.3	0.0	0.0
PESSELTLQELLAEER	G13A	1.3	0.3	7.3	-6.1	0.2
PESSELTLQELLLEER	G13L	1.9	0.7	7.3	0.0	0.3
PESSELTLQELLQEER	G13Q	1.6	0.6	7.3	0.0	0.0
PESSELTLQELLEEER	G13E	1.7	0.9	7.3	0.0	0.5
PESSELTLQELLSEER	G13S	1.1	0.2	0.7	0.0	0.2
PESSELTLQELLGAER	E14A	-0.1	-0.2	7.3	-6.5	0.5
PESSELTLQELLGDER	E14D	0.0	0.1	7.3	-6.7	0.0
PESSELTLQELLGQER	E14Q	0.2	0.0	7.3	-6.2	0.5
PESSELTLQELLGRER	E14R	0.5	-0.2	7.3	0.0	0.3
PESSELTLQELLGEDR	E15D	-0.2	-0.3	-0.7	-7.1	-0.3
PESSELTLQELLGEQR	E15Q	0.1	0.0	-0.5	-7.0	-0.2
PESSELTLQELLGERR	E15R	-0.2	-0.3	-0.9	-7.6	-0.3

Color code										
ΔΔG *(kcal.mol^{-1})*	10	8	6	4	2	0	-2	-4	-6	-8	-10

Figure 29: Calculated ΔΔG for five MBP-fused E6 oncoproteins interacting with a peptide array of E6AP LXXLL motif single-point mutants. The ΔG of the E6APwt is indicated at the top of the table. The variation of free energy was calculated according to $\Delta\Delta G = \Delta G_{MUTANT} - \Delta G_{WT}$: a negative ΔΔG is correlated with an increased affinity (shown in red) while a positive ΔΔG indicates a loss of affinity (in blue) as compared to E6APwt. Hatched cells indicate an abolished interaction. The interaction data of MBP-fused 16E6 were published in a former paper (Bonhoure *et al.*, 2020). For sake of clarity, the solubility-enhancing mutations of each MBP-fused E6 protein are not indicated on the figure: 16E6 F47R 4C/4S; 31E6 C97S C111S; 18E6 F49R; 38E6 C118A; 49E6 C8A; 197E6 C79Y C89Q. The mean uncertainty was estimated at 0.5 kcal.mol^{-1} according to the standard deviation of duplicated data.

For 18E6, our holdup data indicated a slight increase of the affinity with the mutants L8F and E15R as compared to E6APwt (**Figure 29**). We characterized by SPR the interaction of 18E6 with each of these mutants and with a mutant combining the two affinity-enhancing mutations (**Figure 31**). From a K_D higher than 4.8 µM with E6APwt, the affinity increases up to K_D = 2.0 ± 0.4 µM with the double mutant L8F E15R. Our concentration range did not allow us to reach saturation with the wild-type motif and the single-point mutants L8F and E15R, which is why our K_D estimation is less accurate with these peptides than with the double mutant. We compared the affinity of 18E6 with 16E6 for the same peptides: 16E6 interacts with E6APwt at higher affinity than 18E6, which is in agreement with previously published data (Zanier *et al.*, 2005). We designed a "triple fusion" construct of 18E6, which is composed of the solubility-enhancing MBP carrier (Raran-Kurussi *et al.*, 2015), the protein 18E6 with the mutation F49R preventing its dimerization (Zanier *et al.*, 2012) and the double mutant of E6AP LXXLL motif L8F E15R as a C-terminal fusion (**Figure 30**).

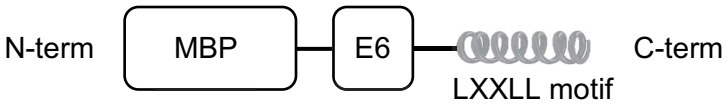


Figure 30: Triple fusion design, which entails the solubility-enhancing carrier MBP, HPV E6 protein and its preferential LXXLL target motif. The N-terminus (N-term) and C-terminus (C-term) of the protein are indicated.

As assessed by SPR, the affinity-enhanced LXXLL motif fused to 18E6 blocks its hydrophobic pocket and prevents 18E6 from interacting with E6AP LXXLL motif peptides immobilized on the surface of the sensor chip. The size-exclusion chromatography of the triple fusion revealed that the protein is predominantly present as monomers (data not shown). This observation indicates that for each molecule, 18E6 binds its own fused LXXLL peptide. The local concentration of the fused peptide is increased in the environment surrounding the hydrophobic pocket, which further increases the occupation of the pocket. The fusion of interacting proteins has long been described for facilitating the crystallization of a protein-peptide complex, as it stabilizes the complex while maintaining a 1:1 stoichiometry (Kobe *et al.*, 2015). In our case, the triple fusion combines the advantages of using MBP as a "crystallization chaperone" allowing molecular replacement (Waugh, 2016) and the LXXLL peptide fusion stabilizing the protein/peptide complex. The team took advantage of the

interaction results and of the triple fusion construct to solve the structure of 18E6 F49R in complex with the double mutant L8F E15R E6AP LXXLL motif (PDB: 6SJV).

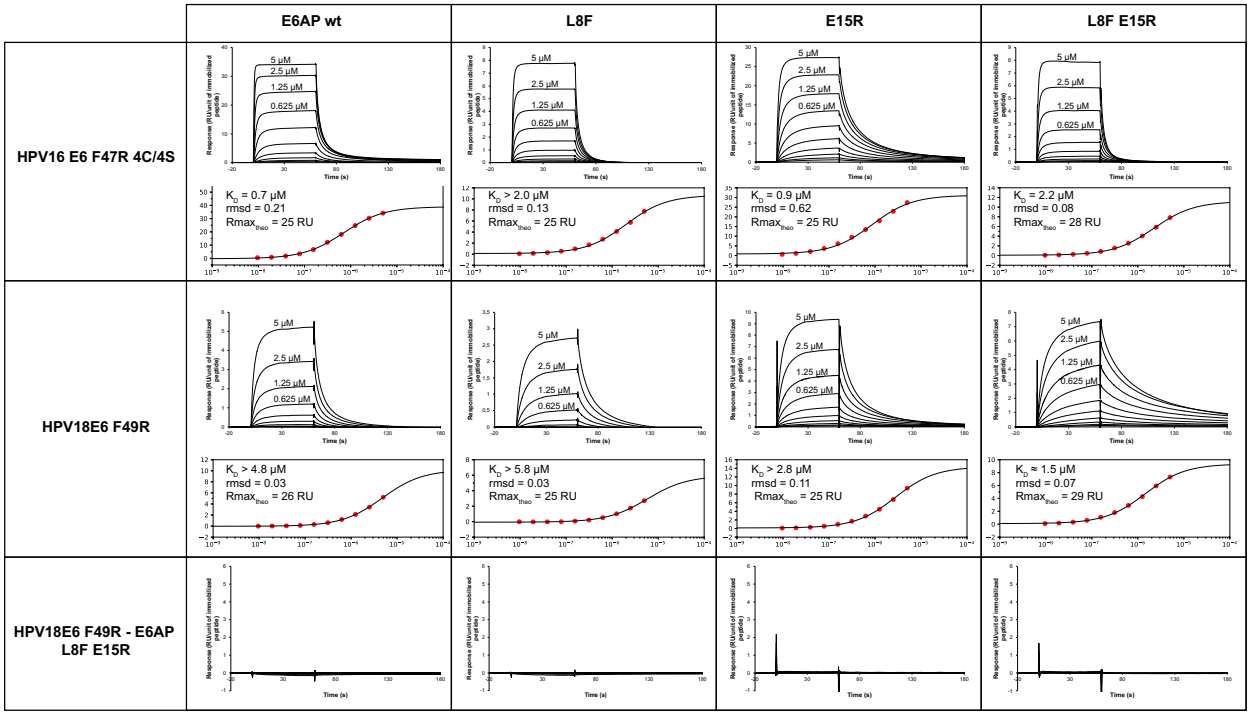


Figure 31: SPR measurements illustrating the affinity enhancement of two combined mutations of E6AP LXXLL motif for MBP-fused HPV18 E6 F49R. The MBP-E6 proteins that were used as analyte in SPR setup are indicated on the left for each row, namely the MBP-fused solubility-enhanced mutants HPV16 E6 F47R 4C/4S and HPV18 E6 F49R, and a fusion of MBP-HPV18E6 with the affinity-enhanced E6AP LXXLL motif. The corresponding E6AP LXXLL peptides (wild-type and mutants L8F, E15R and L8R E15R) are indicated on top of the figure for each column. For each pair, the sensorgrams and the steady-state analysis that enabled affinity quantification are displayed. $Rmax_{theo}$ corresponds to the highest expected response for a 1:1 interaction knowing the amount of immobilized peptide and the molecular weights of the peptide and the protein. rmsd stands for root mean square deviation. The K_D(MBP-18E6/E6AP L8F E15R) is an estimation that can possibly be biased by the fact that the curves did not reach equilibrium before the dissociation phase. Note that the fusion of MBP-18 E6 F49R with L8R E15R E6AP LXXLL motif prevents the interaction with any of the tested peptides, since the hydrophobic pocket of HPV18 E6 is blocked by the fused peptide.

2.1.3.3.3 Crystal structures of E6/LXXLL complexes

The holdup and SPR interaction data that were generated in the present study served as a basis for the identification of stable E6/LXXLL complexes, suitable for structure determination by X-ray crystallography. E6 proteins are stabilized upon binding to a high-affinity LXXLL peptide motif. The first structures of 16E6 and BPV1 E6 could be solved by using their prototypical LXXLL target motifs, E6AP and paxillin, respectively. The same strategy was applied by the team to solve the structures of 18E6 and 49E6. The first peptide library entailing LXXLL motifs from various cellular targets allowed us to identify MAML1 LXXLL motif as a high-affinity binder of 49E6 (see section 2.1.3.3.1). For 18E6, we combined two affinity-enhancing mutations that were identified using the peptide library of single-point mutants of E6AP LXXLL motif (detailed in section 2.1.3.3.2). These two E6/LXXLL complexes were further stabilized for crystallization by designing "triple fusion" proteins: MBP-E6-LXXLL, as detailed in section 2.1.3.3.2.

This strategy allowed the team to solve the structures of 18E6/E6AP LXXLL motif L8F E15R and 49E6/MAML1 LXXLL motif, at resolutions of 1.75 Å and 1.90 Å, respectively (**Figure 32**).

The fold of 18E6 and 49E6 is similar to 16E6: the protein is made of two zinc-binding domains E6N and E6C linked by a hydrophobic helix. The LXXLL peptide is bound by a hydrophobic pocket formed between the zinc-binding domains. Similar to the wings of a butterfly, E6N and E6C are flexible and able to adjust their opening to make contacts with the LXXLL motifs. Indeed, the superimposition of the two E6/LXXLL structures reveals that E6N and the LXXLL peptide are extremely well aligned while E6C can undergo slight changes in conformation and in its relative position as compared E6N and LXXLL peptide. Due to the flexibility of E6C, it is challenging to predict E6 interaction preferences based solely on their structure. Together with some residues from E6N and E6C, the residues from the linker helix that are in contact with the LXXLL peptide participate to the interaction. Their involvement in determining the preferred motifs is easier to anticipate than E6N and E6C interfaces since they belong to a more rigid moiety within E6 protein. The combination of linker helix interface modelling and interaction data enables better prediction E6 LXXLL recognition features.

Hence, the interaction data did not only allow the identification of high-affinity target peptides but also nicely complement the structural data for a better understanding of E6 interaction preferences in terms of LXXLL motif.

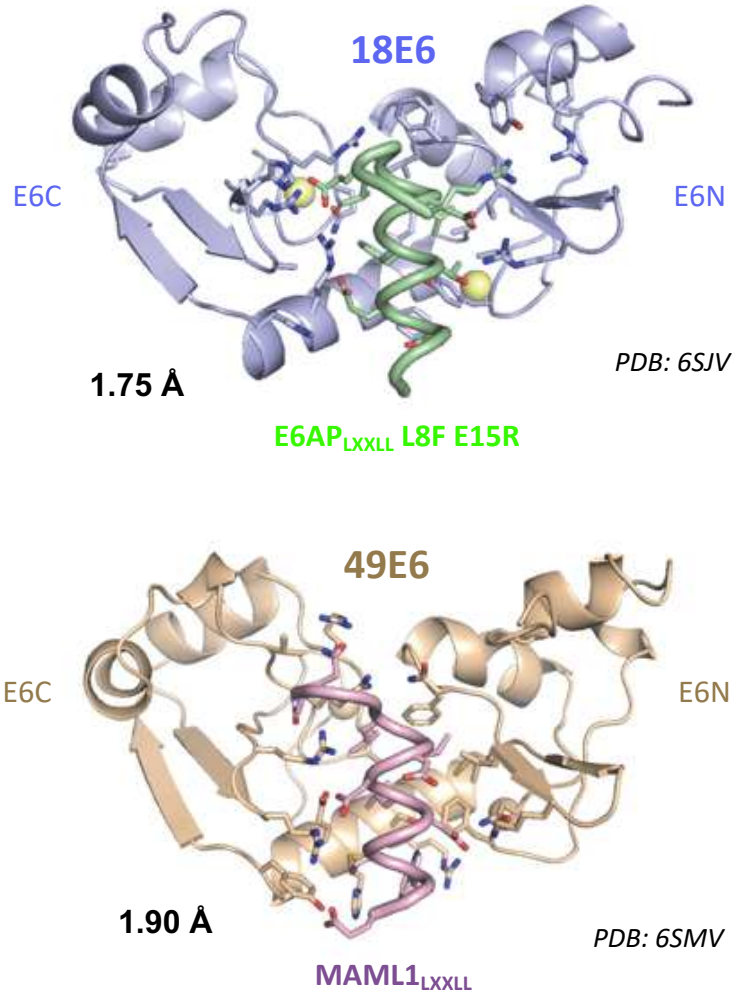


Figure 32: Crystal structures of E6/LXXLL complexes. The interaction data from the present study allowed the identification of stable E6/LXXLL complexes whose structures could be solved by X-ray crystallography. 18E6 F49R forms a stable complex with the affinity-enhanced mutant of E6AP LXXLL motif (L8F E15R) while 49E6 C8A stably interacts with MAML1 LXXLL motif. Adapted from (Suarez, 2017).

2.1.3.4 Further investigation of the p53 LXXLL motif binding by β-genus cutaneous HPV E6 proteins

The interaction of β-genus HPV E6 proteins with p53 and the consequent effects on p53 transactivation have been carefully examined (White *et al.*, 2014). However, the residues within p53 that are necessary for interaction with β-genus HPV E6 have not yet been identified. When the cellular LXXLL motifs peptide array was designed, a peptide bearing the sequence of an LXXLL motif located within p53 was included in

order to test whether it could mediate an interaction with some E6 proteins. The localization of this motif within p53 coincides with the binding sites of the ubiquitin ligase MDM2 and the coactivator p300, as indicated on **Figure 33**.

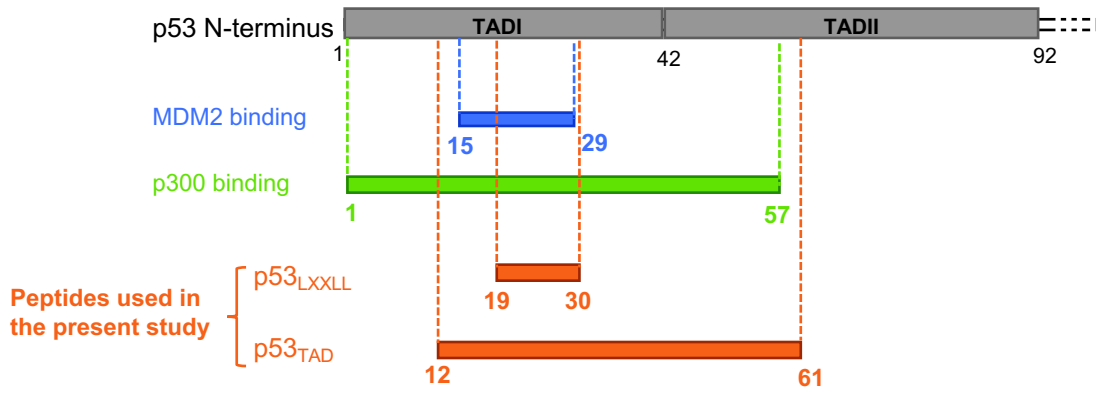


Figure 33: Domain organization of p53 N-terminus and binding sites of MDM2 and p300. The borders of transactivation domains I and II (TADI and TADII) are indicated in grey (Harms and Chen, 2006). The binding sites of MDM2 N-terminus and p300 are is indicated in blue and in green, respectively (Fernandez-Fernandez and Sot, 2011). The peptides $p53_{LXXLL}$ and $p53_{TAD}$ correspond to p53 regions, which are indicated in orange.

The holdup assay revealed that MBP-49E6, MBP-38E6 and MBP-197E6 recognize $p53_{LXXLL}$ peptide motif. The interaction of this region with β-genus HPV E6 proteins might compete with other interaction partners of p53 whose binding site overlaps $p53_{LXXLL}$. The N-terminus of MDM2 binds p53 15-29 region (Yu *et al.*, 2006). Several domains within p300 also interact with the TAD domains of p53: they are indicated on **Figure 34**.

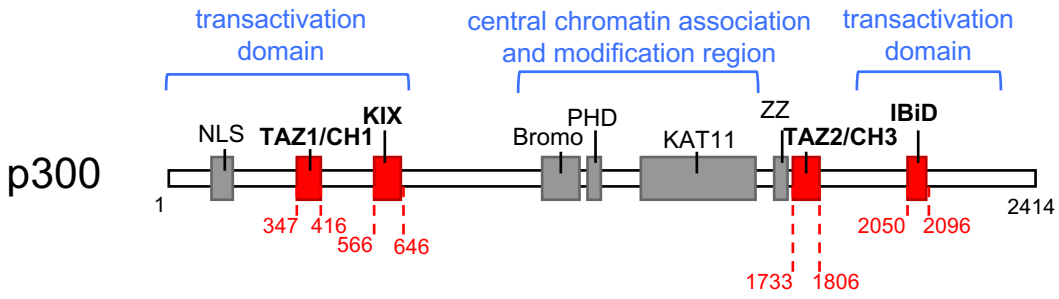


Figure 34: Domain organization of p300. The p53-interacting domains are indicated in red and the functions of the three main regions of p300 are indicated in blue. NLS: Nuclear Localization Signal; TAZ: transcriptional-adaptor zinc-finger domain; CH: cysteine/histidine-rich region; KIX: kinase inducible domain of CREB interacting domain; Bromo: bromodomain; PHD: plant homeodomain finger; KAT11: lysine acetyltransferase domain; ZZ: ZZ-type zinc finger domain; IBiD: IRF3-interacting domain (Chan and Thangue, 2001).

The affinities of MDM2 and the p300 domains TAZ1, KIX, TAZ2 and IBiD for p53 TAD domain (region 1-57) have been determined (**Table 7**). They range from 18 ± 1.8 nM to 4.40 ± 0.18 µM.

		p53 TAD (1-57)
		K_D *(nM)*
p300 domains	TAZ2	18 ± 1.8
	TAZ1	770 ± 50
	KIX	3700 ± 400
	IBiD	4400 ± 180
MDM2		200 ± 20

Table 7: Dissociation constants of p300 domains and MDM2 for p53 (1-57) determined by fluorescence anisotropy. Adapted from (Teufel *et al.*, 2009).

Since we detected an interaction between $p53_{LXXLL}$ and the E6 proteins 38E6 and 49E6, we further characterized their interaction by SPR. We quantified the affinity of β-genus HPV E6 proteins for p53 in order to determine whether their interaction may compete with p300 or MDM2 binding. To this aim, we performed SPR assays on MBP-38E6, MBP-49E6 and the MBP-fused E6 protein from HPV23 WHIM variant (MBP-$23E6_{WHIM}$). Unfortunately, we were unable to generate exploitable data for MBP-$23E6_{WHIM}$ due to aspecific effects observed for the highest concentrations (30 µM; 15 µM and 7.5 µM). The remaining concentration points do no provide sufficient information for K_D determination. These effects can be attributed to the partial oligomerization of MBP-23E6WHIM: further optimization of the purification protocol can help to solve this issue (choice of the reducing agent).

Figure 35 presents the SPR data that could be obtained for MBP-38E6 and MBP-49E6. MBP-49E6 interacts with $p53_{LXXLL}$ and $p53_{TAD}$ at lower affinity than 38E6. The concentration range did not allow an accurate estimation of MBP-49E6 affinity for p53 peptides, however the data indicate that K_D(MBP-49E6/$p53_{LXXLL}$) > 5 µM and K_D(MBP-49E6/$p53_{TAD}$) > 7 µM. 38E6 displayed higher affinity for both peptides, with K_D(MBP-38E6/$p53_{LXXLL}$) = 1.7 µM and K_D(MBP-38E6/$p53_{TAD}$) = 1.4 µM. In addition, the sensorgrams indicate a longer dissociation time for MBP-38E6 as compared to MBP-49E6. We also note that 38E6 binds $p53_{LXXLL}$ and $p53_{TAD}$ at equivalent affinities. This observation suggests that the interaction between 38E6 and p53 is mostly driven by the LXXLL motif and that the close neighboring regions within p53 TAD domain have a limited impact on the interaction.

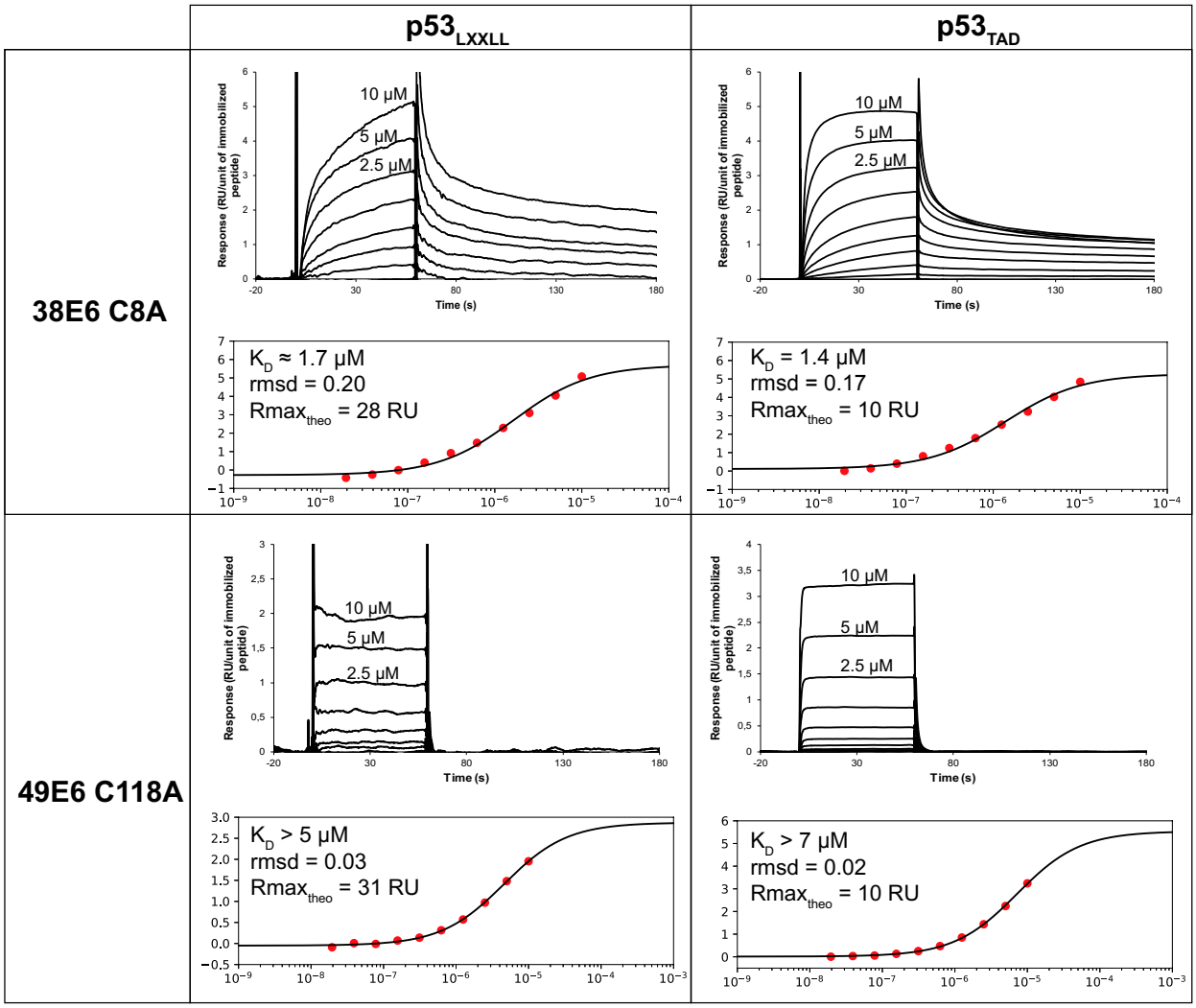


Figure 35: SPR analysis of the interaction between β-genus E6 proteins and fragments of p53. We compared the interaction features of MBP-fused 38E6 and 49E6 for two biotinylated peptides: $p53_{LXXLL}$, which corresponds to the 19-30 region within p53 (sequence: FSDLWKLLPENN) and $p53_{TAD}$, which corresponds to the region 12-61 of p53. $Rmax_{theo}$ corresponds to the highest expected response for a 1:1 interaction and rmsd stands for root mean square deviation. The experiment was performed in singlicate. For the interaction MBP-38E6/$p53_{LXXLL}$, the equilibrium could not be reached in the chosen experimental conditions: the corresponding K_D is a first estimation.

38E6 interacts with p53 at higher affinity than the KIX and IBiD domains from p300. Hence, the binding of 38E6 might compete with these domains for interacting with p53, which may explain the effects of 38E6/p53 interaction on p53 transactivation.

The *in vitro* data presented herein report the interaction of β-genus 38E6 and 49E6 to p53 via an LXXLL motif that is not bound by α-genus 16E6 and 18E6. To complement our data by interaction assays involving full-length proteins, GPCA experiments were performed by the team of our collaborator Murielle Masson (BSC, Illkirch-Graffenstaden). Since GPCA consists in testing the interaction between two full-length

proteins overexpressed in mammalian cells, it is a very potent assay for the detection of direct binary interactions.

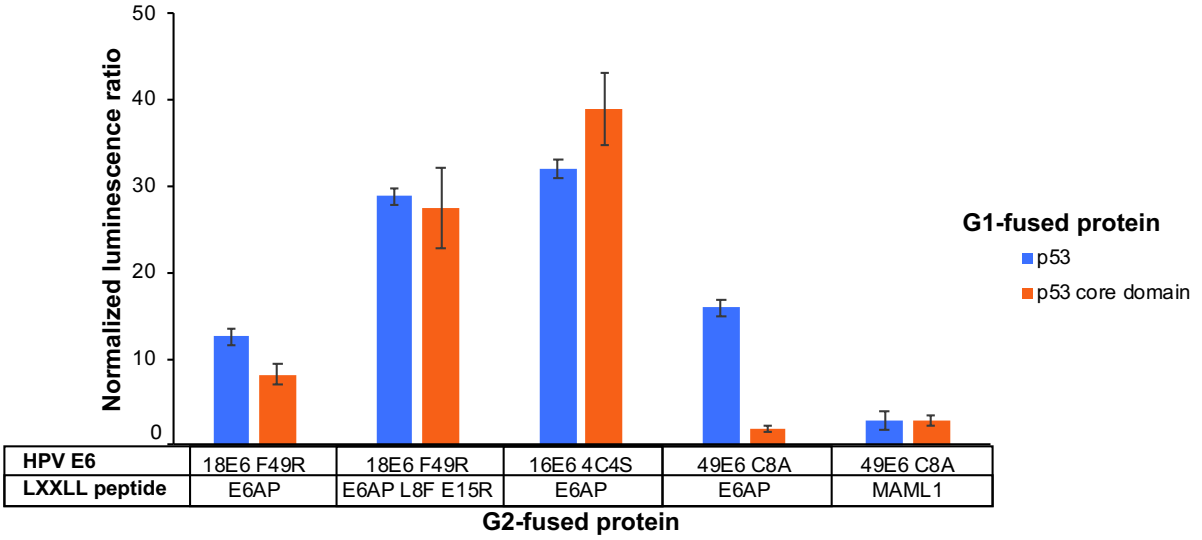


Figure 36: GPCA assay for deciphering the interaction of p53 with 18E6, 16E6 and 49E6. In the present experiment, the complementation of *Gaussia princeps* luciferase is achieved when the two fragments G1 and G2 are gathered by a direct interaction between G1-p53 and a G2-E6-LXXLL. The normalized luminescence ratio is proportional to the strength of the interaction. The tested proteins are fused to LXXLL motifs in order to assess whether LXXLL binding prevents or is required for E6 / p53 interaction.

Figure 36 presents the results obtained by GPCA for comparing the interaction features of 18E6, 16E6 and 49E6 with p53. As previously reported, the high-risk mucosal 16E6 and 18E6 require the formation of a stable complex with E6AP LXXLL motif for interacting with the core domain of p53 (Martinez-Zapien *et al.*, 2016). 16E6 fused to E6AP LXXLL motif produces a high luminescence signal when interacting with either p53 full-length or with only p53 core domain. Similarly, 18E6-E6AP$_{LXXLL}$ fusion appear to interact with both p53 and p53 core domain. The interaction of 18E6 with p53 is improved upon fusion to the affinity-enhanced mutant of E6AP LXXLL motif, E6AP L8F E15R. On the contrary, 49E6 fused to E6AP LXXLL motif is able to bind p53 full-length but not the isolated core domain. Fusing 49E6 to MAML1 LXXLL motif abolishes the interaction with p53. According to the holdup data, 49E6 interacts with E6AP LXXLL motif at low affinity (below the significance threshold) while $\Delta G(49E6/p53_{LXXLL})$ = -8.0 kcal.mol^{-1} and $\Delta G(49E6/MAML1_{LXXLL})$ = -8.5 kcal.mol^{-1} at T = 298 K. Since 49E6 interacts at higher affinity for MAML1$_{LXXLL}$ than for p53$_{LXXLL}$, the fused MAML1 LXXLL motif prevented 49E6 from interacting with p53 LXXLL motif by competitive binding. On the contrary, the affinity of 49E6 for E6AP$_{LXXLL}$ was not high

enough to induce a competition preventing 49E6-E6AP$_{LXXLL}$ fusion from interacting with p53.

Taken together, these results indicate that the interaction of β-genus HPV E6 proteins with p53 is mediated by the recognition of an LXXLL motif within p53. This mode of interaction differs from the tripartite complex involving α-genus E6 proteins, the LXXLL motif of E6AP ubiquitin ligase and the core domain of p53. The interaction of β-genus HPV E6 proteins with p53 does not results in p53 proteasomal degradation but impacts p53 transactivation. Our results suggest that this interaction may compete with p300 domains for binding to p53, which may provide a piece of explanation for the disruption of p53 transactivation.

2.1.4 Discussion

E6 proteins are challenging to use for biophysical studies, due to their propensity for forming inactive oligomers. Coupling a given E6 with a binding LXXLL peptide is an effective strategy for optimizing their purification (Ould M'hamed Ould Sidi *et al.*, 2011) or for structural studies (Zanier *et al.*, 2013).

In the present study, we investigated the binding preferences of seven different E6 proteins for two LXXLL peptide library entailing the main E6 targets reported in the literature. The resulting data enabled the identification of two E6/LXXLL motif complexes whose structure could be solved by X-ray crystallography. Hence, the first structures of the high-risk mucosal 18E6 and the cutaneous 49E6 could first be determined. The structures allow a better comprehension of the molecular basis of E6/LXXLL interaction. However, due to the high flexibility of E6 proteins and in particular the E6C domain, the structural information does not allow a complete understanding of the LXXLL recognition preferences of each E6 protein.

Using a robust strategy combining the holdup assay and SPR, we were able to establish E6 interaction profiles. The peptide library consisting of single-point mutants of E6AP LXXLL motif gave hints about the residues that determine the affinity of the interaction. The results indicate that the sequence recognized by E6 proteins might entail phenylalanine or tryptophane instead of the first leucine of the consensus LXXLL. A double mutant of E6AP LXXLL motif was selected for its enhanced affinity with 18E6.

This double mutant was used for determining the structure of 18E6, as it allowed the formation of a stable complex.

The cellular motifs peptide library allowed the comparison of E6 interaction preferences for diverse peptide sequences. Thus, some similarities between E6 belonging to the same phylogenetic group were observed. In addition to the previously reported targets, our data suggest a novel mode of interaction between p53 and β-genus 38E6 and 49E6. Unlike 16E6, 38E6 and 49E6 capture an LXXLL motif within p53 sequence without any interaction with p53 core domain. The p53 LXXLL motif overlaps p300-binding regions. Our affinity measurements suggest that 38E6 might prevent the binding of p300 domains by competitive binding. This observation might explain that β-genus E6 disrupt p53 transactivation without inducing its proteasomal degradation (Cornet *et al.*, 2012; White *et al.*, 2012a, 2014). Interestingly, p53 retains its transactivating activities in high-risk mucosal HPV-positive cells despite lowered protein levels (Bañuelos *et al.*, 2003; Butz *et al.*, 1995).

2.2 Study of the interaction preferences of ancestral E6s of the alpha genus

2.2.1 State of the art

Papillomaviruses co-evolved with their hosts by acquiring abilities that allow them to maximize their replication while adapting themselves to the host evolution. Deciphering PV evolution provides hints to understand why certain PV became oncogenic and at which stage of their evolution they gained carcinogenic capacities (Willemsen and Bravo, 2019). To this aim, the α-genus is particularly interesting to study since it entails high-risk mucosal HPVs (such as the α9-species HPV16 and 31 and the α7-species HPV18), low-risk HPVs causing genital warts (e.g. α10 HPV11 and 6) and low-risk cutaneous HPVs (e.g. α2 HPV3).

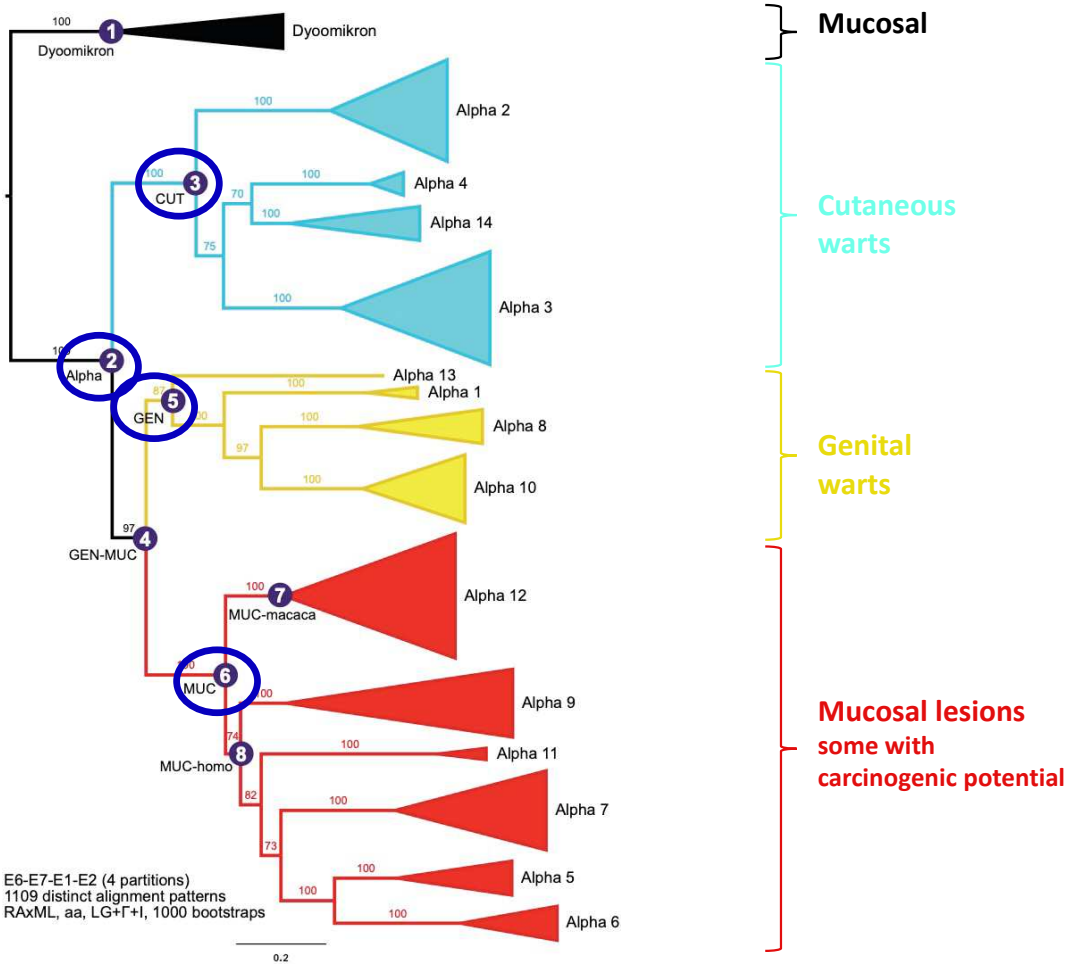


Figure 37: Phylogenetic tree of α-genus HPVs generated by Anouk Willemsen and Ignacio Bravo. The phylogenetic tree was generated using a maximum likelihood algorithm (RAxML). The common ancestors on which the study was focused are indicated by purple circles: 2Alpha, 3CUT, 5GEN and 6MUC. The phenotypes induced by each group of PVs are indicated on the right side of the figure. The triangles indicate the number of HPV types within each species.

This project was done in collaboration with Anouk Willemsen and Ignacio Bravo (MiVEGEC, IRD Montpellier). They traced the evolution of α-genus PVs and generated the phylogenetic tree presented in **Figure 37**. Hence, the difference between the phenotypes induced by α-PVs finds an explanation in the evolutionary divergences that splitted the PVs in the species and types as we know them nowadays. Using a bayesian algorithm, Anouk WIllemsen and Ignacio Bravo were able to reconstitute the probable E6 protein sequences of four common ancestors: 2alpha, 3CUT, 5GEN and 6MUC. As their names indicate, they are the ancestors of all **α**-genus, low-risk PVs causing **cut**aneous warts, low-risk PVs causing **gen**ital warts and **muc**osal PVs.

2.2.2 Objective

We analyzed the interaction preferences of four ancestral E6 proteins for a set of LXXLL peptide motifs in order to investigate how the LXXLL interaction features of α-genus E6 proteins evolved and how they affected the phenotypes caused by each PV type.

2.2.3 My contribution

I produced and purified the MBP-fused ancestral E6 proteins. I checked the quality of the purified samples by Dynamic Light Scattering (DLS) and performed interaction assays by SPR and holdup assay.

2.2.4 Material and Methods

The sequences encoding the ancestral E6 proteins were cloned in Acc65I/NcoI restriction sites of pETM41 vector by the company Genscript. Small-scale nickel purification and SPR was performed as previously described (Bonhoure *et al.*, 2018). The holdup assay was performed as a scale-up 384-well plate version of the protocol detailed in (Bonhoure *et al.*, 2020). The experiment was executed in the platform PCBIS with the kind help of Sophie Gioria who programmed the pipetting robots.

2.2.5 Results

We selected some α-genus HPV types in order to have a point of comparison with the ancestral E6 proteins. High-risk mucosal α9-species 16E6 was already available in our lab. As a complement and to cover the major groups depicted in **Figure 37** (cutaneous, genital, mucosal), we attempted to purify α10-species HPV11 (genital) and α2-species

HPV3 (cutaneous). Unfortunately, we were unable to isolate active samples of 11E6 and 3E6, despite the mutation of solvent-exposed cysteines (see **Table 3**).

Thus, we conducted the study using five MBP-fused proteins: 16E6 and the four ancestral E6 proteins 2Alpha, 3CUT, 5GEN and 6MUC. In order to perform interaction assays in parallell on freshly purified samples, we took advantage of the nickel purification protocol previously developed for this purpose (Bonhoure *et al.*, 2018). The oligomeric state of the purified samples was assessed by Dynamic Light Scattering: all samples were monomeric, which is a positive indication regarding their interaction abilities.

	$E6AP_{LXXLL}$				**$IRF3_{LXXLL}$**			
	K_D (µM)	**rmsd**	**$Rmax_{exp}$ *(RU)***	**$Rmax_{exp/theo}$**	**K_D (µM)**	**rmsd**	**$Rmax_{exp}$ *(RU)***	**$Rmax_{exp/theo}$**
16E6	1.2	0.9	150	0.7	1.0	9.6	328	1.1
2Alpha	0.2	1.0	178	0.8	3.0	8.4	267	1.0
3CUT	0.3	4.2	192	0.8	3.0	6.8	301	1.0
5GEN	> 9.9	2.6	112	0.5	> 19.8	5.5	279	0.9
6MUC	13.6	1.6	139	0.6	7.2	1.9	319	2.2

Table 8: Preliminary dissociation constants of four ancestral E6 proteins for E6AP and IRF3 LXXLL peptide, estimated by SPR. rmsd: root mean square deviation. $Rmax_{theo}$: highest signal expected for a 1:1 interaction (theoretical). $Rmax_{exp}$: highest signal according to the fit.

We first assessed the interaction features of the MBP-fused ancestral E6 proteins by SPR, using two LXXLL motif targets typical from α-genus HPV E6: E6AP and IRF3 LXXLL motifs (**Table 8**). Interestingly, 2Alpha and 3CUT bound E6AP LXXLL motif at more than 10-fold higher affinity than 6MUC. 6MUC is evolutionary closer to 16E6 than 2Alpha or 3CUT and still, its affinity of E6AP LXXLL motif is significantly lower than 16E6. Our conditions did not allow us to reach the saturation plateau for 5GEN, indicating that its K_D for E6AP LXXLL motif is above 9.9 µM. Not all α-genus E6 proteins can bind IRF3 LXXLL motif, since α7-species 18E6 does not bind IRF3 LXXLL. Surprisingly, 2Alpha and 3CUT interacted with IRF3 LXXLL motif in the same affinity range as 16E6. 6MUC bound stronger to IRF3 LXXLL motif than to E6AP LXXLL motif. The interaction with 5GEN was ocurring at low affinity, which could not be precisely determined in our conditions.

We complemented these data by a holdup assay performed in parallell on the five freshly purified MBP-E6 proteins (**Figure 38**). Since the resulting data were generated within a single experiment and have not yet been validated by an independent replicate, they should be considered as preliminary data awaiting confirmation.

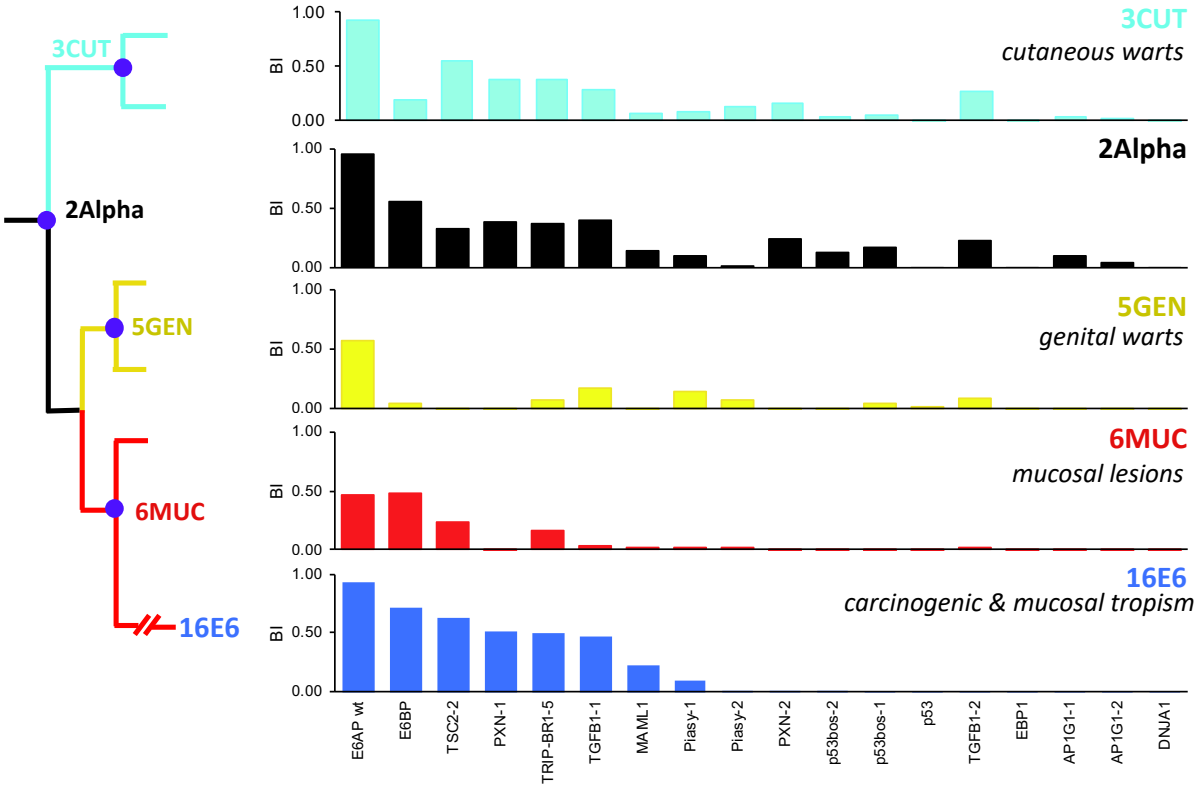


Figure 38: Preliminary interaction profiles of 16E6 and ancestral α-genus E6 proteins assessed by holdup assay. The MBP-fused E6 proteins were tested for their interaction with a part of the library of LXXLL motifs from cellular targets (see **Table 4**). The present data were generated within a single experiment and have not yet been reproduced.

Concerning the interaction between E6AP LXXLL motif, the results of the holdup assay tend to confirm the SPR results. Similar to 16E6, 2Alpha and 3CUT bind E6AP$_{LXXLL}$ at high affinity while 6MUC and 5GEN interact at much lower affinity with this same motif. Most of the peptides bound by 16E6 are also recognized by 2Alpha and 3CUT. The ancestral proteins 2Alpha and 3CUT also bind additional targets such as TGFB1-2 and PXN-2. Our previous results did not indicate any interaction of an α-genus E6 with these two motifs, whose sequences are very similar. However, we note that they are recognized by β-genus 38E6 and 49E6 and ɣ-genus 197E6 (**Figure 28**). This unexpected result might indicate that the interaction preferences of cutaneous α-genus E6 proteins are similar to those of cutaneous β-genus E6 proteins, despite their phylogenetic distance. This evolutionary convergence would explain their similarities

in term of tropism and induced phenotype. Hence, the binding to paxillin/TGFB1 might be a marker of cutaneous tropism in E6 proteins.

Strikingly, 5GEN protein appears to bind only E6AP LXXLL motif at BI (MBP-5GEN/E6AP$_{LXXLL}$) = 0.6. The overall low affinity and the few binding peptides may indicate that 5GEN is partially inactive. The same observation can also be applied to 6MUC. Despite the quality check and the purification protocol maximizing monomeric fraction, the best option ensuring optimal protein quality is the large scale purification that was reported for MBP-16E6 (Zanier *et al.*, 2013). The small-scale purification is a very good tool for screening the interaction preferences of a set of protein, however the final results to be published should ideally be generated with protein purified at large scale. Another possibility is that the tested set of peptides does not entail the preferential targets of 5GEN and 6MUC or that they have limited LXXLL motif recognition abilities (like 11E6 which might be unable to bind LXXLL motif as suggested in the section 2.1.3.2).

2.2.6 Discussion

The present project led us to approach the question of E6/LXXLL interactions from another angle: that of evolution. The extraordinary broad diversity of PV types identified to date results from a long co-evolution allowing the virus to adapt to a certain niche. Hence, some PVs have evolved toward a cutaneous tropism and a low oncogenic risk while other PVs belonging to the same genus have a mucosal tropism and a high oncogenic risk.

In this section, we presented preliminary results deciphering the LXXLL interaction preferences of several ancestral E6 proteins from the genus α. Our data indicate that all tested ancestral E6 proteins share the ability to interact with E6AP LXXLL motif, which is the prototypical target of 16E6. Except 5GEN that interacts at much lower affinity, they also share the ability to interact with IRF3 LXXLL motif. Surprisingly, the common ancestor of high-risk mucosal HPVs 6MUC displayed rather low affinity for these two motifs. As detailed in the results section, this lower affinity can be due to the intrinsic interaction preferences of the protein. However, one cannot rule out that the proteins 6MUC and 5GEN were partly inactive during the experiments, which would have resulted in an apparent higher K_D. Our promising yet preliminary results must be reinforced by reproducing the interaction assays with protein purified using a large-

scale purification procedure. Even though the parallel batch purification was optimized for maximizing the fraction of monomeric protein, its efficiency in isolating monomeric does not compete with a purification protocol entailing an overnight ultracentrifugation and a SEC. Nevertheless, the nickel batch purification has been put to advantage in such project implying the comparison of four proteins which had never been purified until then.

Our holdup results suggest a possible convergent evolution between cutaneous PVs from genus α and those from genera β and ɣ. The LXXLL motifs of paxillin and TGFB1 appeared as potential markers of cutaneous PV E6 proteins. This exciting hypothesis should be further confirmed and further examined. Regarding the good biochemical properties of the ancestral E6 proteins 2Alpha and 3CUT, we could attempt to cystallize them in complex with either E6AP or IRF3 LXXLL motif. The determination of their structure would allow a comparison with the structure of the β-genus 49E6. In addition, since we could not isolate 11E6 nor 3E6 for structure determination, the structure of these ancestral proteins might be an interesting alternative for the structural study of α-genus PV E6 proteins other than 16E6 and 18E6.

3 Adjunct project: Study of the interaction between ubiquitin ligases E6AP and HERC2

3.1 State of the art

Apart from being the prototypical target of E6 from high-risk mucosal HPVs, defects in the expression E6AP ubiquitin ligase are related with neurodevelopmental disorders. Mutations in *UBE3A* (the gene encoding E6AP) are linked with Angelman Syndrome (AS), a genetic disorder that impairs the development of the nervous system (Kishino *et al.*, 1997). The patients suffering from AS are characterized in particular by mental retardation, easily excitable behavior, gait ataxia, frequent smiling and laughter (Kalsner and Chamberlain, 2015). Also named "puppet syndrom", AS is named after the English pediatrician Harry Angelman, who first described AS in 1965 (Angelman, 1965). The global prevalence of AS is estimated between 1/10,000 and 1/20,000 (angelman-afsa.org, consulted on September, 7th 2020). *UBE3A* is located in the 15q11-q13 locus. The involvement of other genes located in this same locus has been highlighted, in particular *HERC2* (Harlalka *et al.*, 2013). *HERC2* encodes an E3 ubiquitin ligase that physically interacts with E6AP at the protein level (Kühnle *et al.*, 2011). Together with a third protein named NEURL4, HERC2 and E6AP form a high molecular weight complex: the HUN complex. To date, the function of the HUN complex remains poorly understood, but a recent study reported its role in regulating the actin cytoskeleton as well as its association with mTORC1 has been recently reported (Martínez-Noël *et al.*, 2018). mTORC1 is a regulator of metabolism whose dysregulation is associated with neurodevelopmental disorders.

Hence, E6AP and HERC2 are linked both by the proximity of their genes and by their physical interaction at the protein level. The regions involved in their interaction were characterized by a study published in 2011 (Kühnle *et al.*, 2011). As displayed on **Figure 39**, the second RCC1-like domain of HERC2 (RDL2) mediates the interaction with a 50-residue region located within E6AP (150-200 in Isoform I). The crystal structures of RLD1 and RLD3 were determined by X-ray crystallography: they revealed a beta-propeller fold. Since RLD2 shares 45 % and 57 % identity with RLD1 and RLD3, respectively, one can expect that RLD2 domain adopts a similar beta-propeller fold. Regarding E6AP, there is to date no available structure of the full-length protein. In addition to the LXXLL motif bound by 16E6, the structures of the N-terminal zinc

binding domain as well as the C-terminal catalytic domain (HECT) have been determined. However, the arrangement of the different domains and how they may impact the protein interaction features of E6AP remains unknown. Most particularly for the RLD2-binding site, the region 150-200 is predicted as disordered.

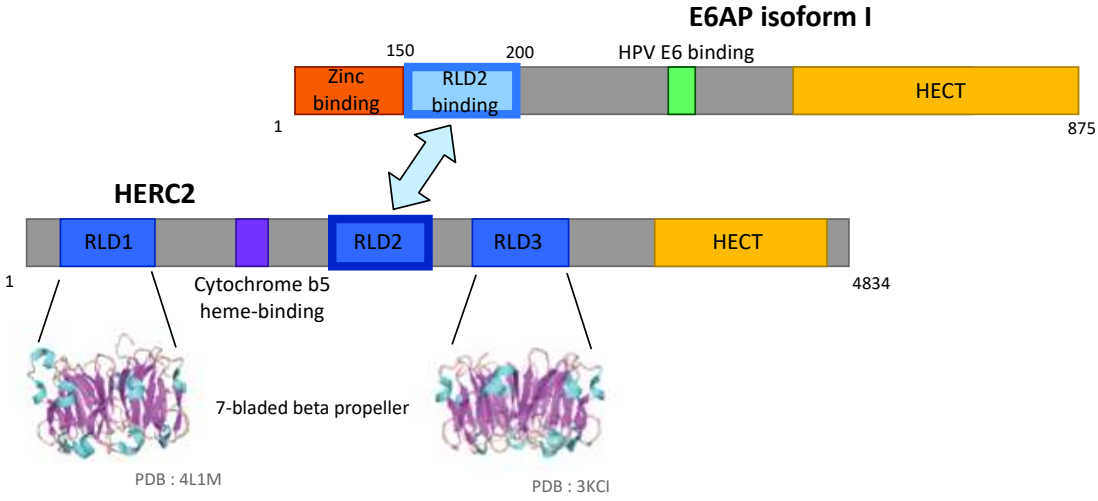


Figure 39: Domain organization of HERC2 and E6AP. RLD: RCC1-like domain; HECT: Homologous to E6AP C-terminus. E6AP/HERC2 interaction is mediated by the 150-200 region within E6AP and HERC2 RLD2 domain (Kühnle *et al.*, 2011). Unlike RLD1 and RLD3 domains, the structure of RLD2 domain has not been determined to date.

3.2 Objective

As specialists of protein-motif interactions, we aimed at reducing RLD2-binding region within E6AP to the minimal interaction unit. Identifying shorter binding motif would facilitate the structure determination of RLD2 complex by crystallography. Indeed, it would reduce the number of flexible loops that are not driving the interaction and make the formation of diffracting crystals more difficult. In addition, the identification of a short binding motif recognized by RLD2 domain paves the way to the identification of a novel domain-motif interaction network.

3.3 My contribution

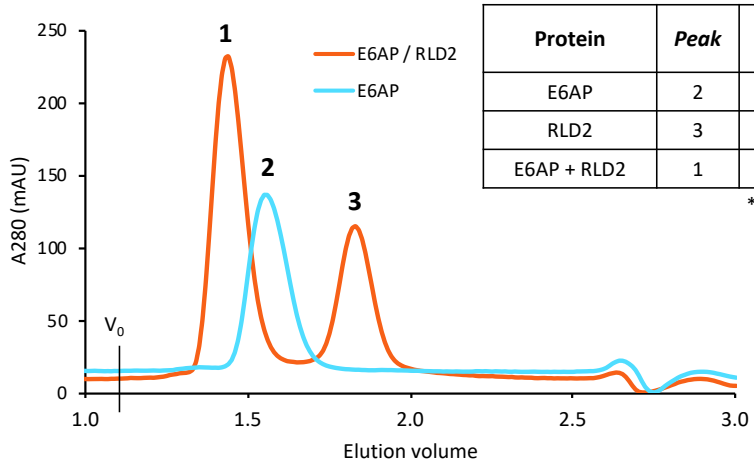
I designed E6AP overlapping peptides which were synthesized by the Peptide synthesis service from the IGBMC. I produced and purified recombinant MBP-fused RDL2 domain and performed analytical SEC, holdup and SPR interaction assays. I also performed preliminary crystallization screenings on MBP-RLD2 in absence of any ligand (apo).

3.4 Material and Methods

The RLD2 domain from HERC2 (sequence: 2,941-3,342 from HERC2, uniprot ID: O95714) was cloned in pETM41 and pETXM1 vectors using the restriction sites NcoI Acc65I. MBP-RLD2 was purified according to the purification protocol published for 16E6 F47R 4C/4S (Zanier *et al.*, 2013). MBP-E6AP was purified by Eduardo Howard and Irina Suarez. SPR and holdup assay were performed as previously described (Bonhoure *et al.*, 2020). Analytical SEC was performed as detailed in (Bonhoure *et al.*, 2018).

3.5 Results

Our first step for studying the interaction between HERC2 RLD2 domain and E6AP was to produce active samples of both proteins. To do so, we applied the same strategy as developped for E6 oncoproteins. RLD2 domain and E6AP full-length were expressed as MBP fusions for improving their solubility and careful purification entailing overnight ultracentrifugation for the elimination of inactive oligomers was performed. Despite those precautions, the purification of full-length E6AP is challenging and has been the subject of numerous optimizations by Eduardo Howard and Irina Suarez, researchers in the team. Once we could obtain samples suitable for biophysical assays, we first checked whether full-length E6AP and HERC2 RLD2 domain were able to interact by size-exlusion chromatography (SEC) (**Figure 40**).



Protein	*Peak*	V_e exp (mL)	MW_{exp}* (kDa)	MW_{theo} (kDa)	Rh_{exp}* (Å)
E6AP	2	1.54	215	141	52
RLD2	3	1.83	75	83	38
E6AP + RLD2	1	1.42	334	224	60

* Calculated from column calibration

Figure 40: Analytical SEC of two MBP-fused proteins: HERC2 RLD2 domain and full-length E6AP. A table indicates the experimental and theoretical molecular weights of each protein and of the complex (MW_{exp} and MW_{theo}), as well as the elution volume (V_e) and the experimentally determined hydrodynamic radius (Rh_{exp}). The void volume (V_0) is indicated on the chromatogram.

As illustrated by the chromatogram on **Figure 40**, we made two injections of our MBP-fused proteins: E6AP/RLD2 with an excess of RLD2 and E6AP alone. Peak 3 was attributed to the excess of RLD2, as its molecular weight calculated from the elution volume was very close to the theoretical molecular weight: MW_{exp}(MBP-RLD2) = 75 kDa and MW_{exp}(MBP-RLD2) = 83 kDa. The conversion of elution volume into molecular weight is based on the calibration of the column by globular proteins. However, depending on the shape of the proteins in solution, one can observe some variations between MW_{theo} and MW_{exp}. For instance, MW_{exp}(MBP-E6AP) is 1.5 fold higher than the expected value MW_{theo}(MBP-E6AP). This difference can be explained because MBP-E6AP is probably not globular nor spherical in solution, thus its migration in the column is slowed down by its shape. Finally, the peak of MBP-E6AP/MBP-RLD2 complex is visible on the chromatogram (peak **1**). Its MW_{exp} is close to the sum of MW_{exp}(MBP-RLD2) and MW_{exp}(MBP-E6AP). The shift between peak 1 (RLD2/E6AP) and peak 2 (E6AP alone) confirms that the peak 1 is due to the formation of a complex. The elution volume of the peak 1 is distinct from the void volume V_0, which further confirms the validity of our MW estimation. Indeed, a protein peak eluted in the void volume cannot undergo size estimation since it is out of the size range that the column can separate at adequate resolution.

In order to reduce RLD2-binding region within E6AP, we designed overlapping peptides and compared their affinity for HERC2 domain (**Figure 41**).

E6AP 150-200 Isoform I

Polypeptide	HTKEELKSLQAKDEDKDEDEKEKAACSAAAMEEDSEASSSRIGDSSQGDNN
1/8	HTKEELKSLQAKDED
2/8	LKSLQAKDEDKDEDE
3/8	AKDEDKDEDEKEKAA
4/8	KDEDEKEKAACSAAA
5/8	KEKAACSAAAMEEDS
6/8	CSAAAMEEDSEASSS
7/8	MEEDSEASSSRIGDS
8/8	EASSSRIGDSSQGDNN

Figure 41: Design of overlapping peptides within E6AP sequence for mapping the minimal interaction unit with the RLD2 domain of HERC2. Based on the polypeptide reported in (Kühnle *et al.*, 2011), eight 15-mer overlapping peptides were designed for interaction assays with MBP-fused RLD2 domain.

We included the full 150-200 region in a polypeptide and designed eight 15-mer overlapping peptides, with an overlap of five residues. We then attempted to compare

the binding of MBP-RLD2 to these peptides by SPR. We estimated K_D(MBP-RLD2/E6AP polypeptide) = 0.6 µM with rmsd = 3.5 RU, $Rmax_{exp}$ = 123 RU and $Rmax_{exp}/Rmax_{theo}$ = 0.34. The high difference between rmsd and $Rmax_{exp}$ indicates that the good quality of the fit used to estimate the K_D. However, the low ratio $Rmax_{exp}/Rmax_{theo}$ might indicate a partial aggregation of the polypeptide immobilized on the surface. This hypothesis seems probable regarding the difficulties of synthesis and resuspension assciated with the polypeptide. On the contrary, MBP-RLD2 is monomeric in solution and does not significantly forms oligomers one isolated, as illustrated on **Figure 40**.

Unfortunately, we were only able to estimate the dissociation constant of the polypeptide and did not detect any interaction with the fragments. It is very likely that reducing the interface implies a drastic decrease of affinity, which is out of the detection range of SPR. Hopefully, the holdup assay is more sensitive to low affinities than SPR since is does not entail any fluidic system likely to wash away the transient complexes. Hence, we compared the interaction preferences of MBP-RLD2 for our set of overlapping peptides by the holdup assay (**Figure 42**).

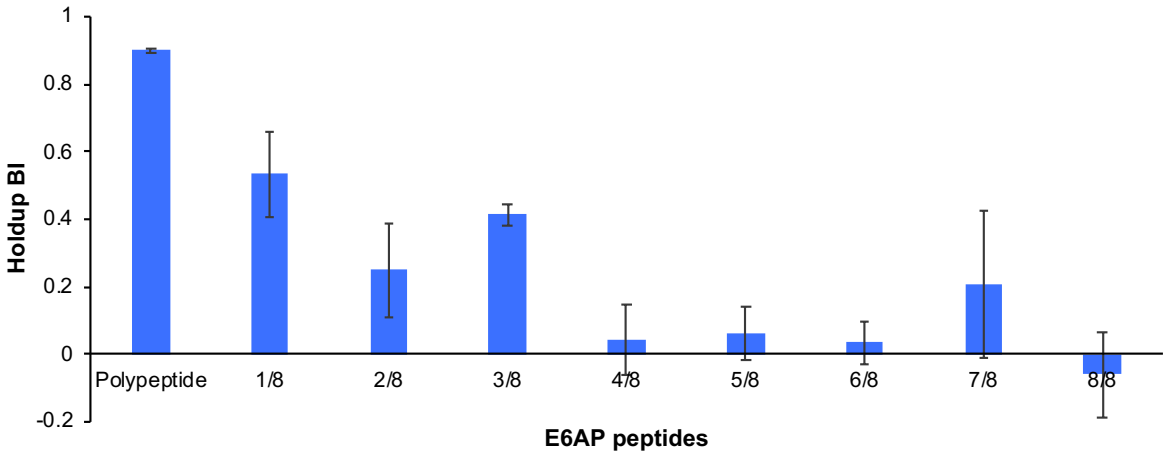


Figure 42: Holdup assay for mapping E6AP-HERC2 interaction. The interaction between MBP-fused RLD2 domain of HERC2 and biotinylated E6AP peptides was quantified by holdup assay. The nomenclature used for E6AP peptides is the same as presented in **Figure 41**. The standard deviations according to duplicated data are shown as error bars.

First, the holdup assay confirmed the lower affinity of MBP-RLD2 for the 15-mer peptides as compared to the 50-mer polypeptide. It also revealed higher affinities for the fragments 1/8 and 3/8 as compared to the other fragments.

While performing interaction assays, I conducted first crystallization trials with MBP-RLD2 in apo from, that is to say in absence of any binding peptide. Unfortunately, the protein did not crystallize in absence of its ligand.

This project was then continued by Auguste Demenge, a PhD student in the team who started in 2019 a thesis in structural biology focused on E6AP and its interaction with HERC2 interactions. He took advantage of the "triple fusion" strategy that previously proved to be effective for determining the structures of E6/LXXLL complexes (**Figure 30**) and used MBP-RLD2-E6AP peptide fusions. He carried out several crystallization screenings on several complexes involving MBP-RLD2 and E6AP peptides (polypeptide, 1/8 and 3/8). He could successfully determine the structure of HERC2 RLD2 domain in complex with E6AP 3/8 peptide (**Figure 43**).

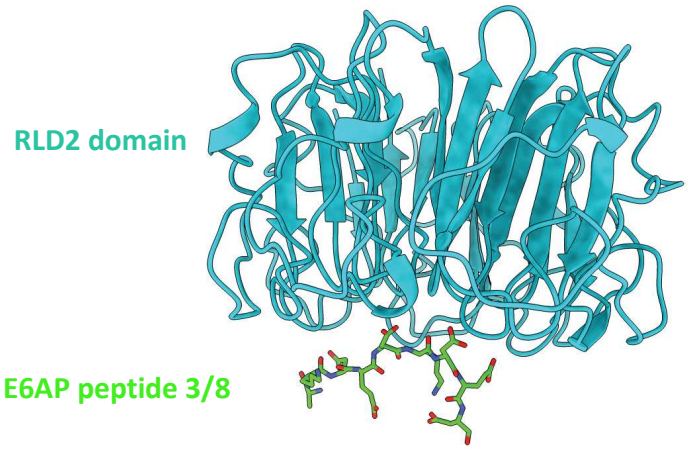


Figure 43: Crystal structure of HERC2 RLD2 domain in complex with E6AP 3/8 peptide, determined by Auguste Demenge. The structure was solved at a resolution of 3.0 Å. The 3/8 fragment of E6AP (in green) binds the bottom of RLD2 7-bladed beta propeller (in blue).

As postulated based on the homology between RLD domains, RLD2 adopts a 7-bladed beta-propeller fold. The 3/8 fragment of E6AP peptide binds the bottom of the beta-propeller. Interestingly, the 3/8 peptide contained a repeated motif within its sequence: AKDED**KDED**EKEKAA. Since this motif could possibly be part of a consensus recognized by different proteins, I attempted to identify novel RLD2-binding peptides by proteomic peptide phage display. I tested different ionic strength conditions, in order to promote interactions involving the charged side chains of the motif. Unfortunately, I was unable to select any binding peptides by this approach. Regarding the very low affinity of 15-mer peptides for RLD2 whose interaction would not even be detected by SPR, it is likely that the washing steps of the phage display procedure removed the

transient interacting peptides, despite the adaptations of the protocol for selecting domain-motif interactions.

3.6 Discussion

In this project, we took advantage of the robust methodology that we developped initially for the study of E6/LXXLL interactions. This approach has now proven to be effective for other domain-motif interactions. Regarding the purification procedure, consequent amounts of active MBP-RLD2 could be isolated using the protocol that is routinely used for E6 proteins. The combination of SPR and holdup assay was particularly advantageous in this project since the affinity of RLD2 for overlapping 15-mer peptides is out of SPR detection range. SPR is a potent tool for the accurate quantification of high-affinity binders whereas holdup is more sensitive to medium-low affinities (in the micromolar range).

By restricting the interface to the minimal interaction unit within E6AP, the first structure of RLD2/E6AP complex could be determined. Both E6AP and HERC2 are involved in AS, but the exact mechanism by which the dysfunction of these proteins leads to a neurodevelopmental disorder remains yet poorly understood. There is a real need for biophysical and structural characterization of these two ubiquitin ligases, as they are at the origin of such disabling syndrome for which no cure is available to date. The interest in understanding AS and the promising preliminary results that could be obtained justify the beginning of a thesis deciphering E6AP/HERC2 interaction.

The identification of the E6AP fragments that drive the interaction with HERC2 highlighted the presence of a repeated KDED motif, which has not been reported in the Eukaryotic Linear Motif datablase to date (elm.eu.org, consulted on September, 7th 2020). Other proteins bearing this motif might interact with RLD2 domain, which could possibly be a novel protein-interaction platform. Phage display trials were unsuccessful but the results presented herein indicate that the holdup assay is a suitable approach for identifying novel RLD2-binding peptides. As part of Auguste Demenge's thesis, a peptide library entailing hundreds of peptides deriving from E6AP binding motif has been generated. The next steps consist in testing the interaction of not only RLD2 but also RLD1 and RLD3 with this library and to investigate whether additional binding motifs can be identified using the holdup assay.

At last, determining the structure of E6AP full-length would be the milestone of this project, since it is of prime interest for both HPV-related cancers and AS. The identification of HERC2 RLD2 domain as a stable interaction partner might stabilize E6AP and facilitate its structure determination either by X-ray crystallography or by cryo-electron microscopy (Cryo-EM). The optimization of E6AP purification led to significant progress and trials are still ongoing for determining the appropriate conditions for Cryo-EM.

Discussion

The E6 oncoproteins are an attractive subject of study: how can such small viral proteins target so many host cellular proteins and induce carcinogenesis? They are at the origin of a high number of cancers worldwide, however their biochemical properties make them difficult to isolate *in vitro*, which delays the analysis of their interaction features.

Designing customized biochemical approaches for the study of E6 oncoproteins

The first part of my thesis work consisted in developing strategies for isolating active E6 samples and quantifying their affinity for an array of LXXLL peptide motifs. When overexpressed in bacteria and isolated *in vitro*, most E6 proteins tend to form inactive oligomers. However, proper biophysical studies require active, monomeric E6 samples. Designing MBP-fused constructs in which the solvent-exposed cysteine residues are mutated is an effective way to reduce the propensity of a given E6 to oligomerize. However, it is a long trial and error process, since some MBP-E6 proteins still aggregate despite the mutations aiming at enhancing their solubility. Hence, the design of a fast, parallelizable purification protocol maximizing the proportion of active MBP-E6 proteins (Bonhoure *et al.*, 2018) was a great advantage for the screening of numerous mutated E6 from different HPV types (**Table 3**).

Using this methodology, we directly checked the LXXLL motif binding properties of the purified MBP-E6 samples by SPR interaction assay. This screening workflow allowed us to identify which E6 proteins were suitable for detailed interaction analysis. The next step consisted in quantifying the affinity of E6/LXXLL interactions by a sensitive, high-throughput assay. The holdup chromatographic retention assays was previously published by the team for establishing the quantitative interactome of all existing PDZ domains for the PBM of HPV16 and 18 (Vincentelli *et al.*, 2015). This approach, which has proven to be reliable for the study of protein-motif interactions, required some adjustments before being applied to E6/LXXLL interactions. First, handling 266 PDZ domains overexpressed in bacterial extracts was achieved using pipetting robots. Since we used purified MBP-E6 samples for deciphering E6/LXXLL interactions, the holdup assay could be achieved on the bench, using multichannel pipettes. Second, while capillary electrophoresis was well suited for quantifying the analyte depletion in bacterial extracts, it could readily be replaced by intrinsic tryptophan fluorescence for

quantifying the depletion of purified protein. Thus, we published a methodological article in which we precisely described our standardized protocol for performing the benchtop holdup interaction assay and processing the data (Bonhoure *et al.*, 2020). We applied these methodological developments as a basis to select a set of HPV E6 proteins, quantify and compare their interaction preferences in terms of LXXLL peptides by a rigorous, standardized approach.

Summary of the scientific input provided by this thesis

We could isolate seven E6 oncoproteins (16E6, 31E6, 18E6, 38E6, 49E6, 197E6, BPV1 E6) and analyze their interaction preferences for two LXXLL peptide libraries: i) one with single-point mutations within E6AP, the prototypical LXXLL target of high-risk HPV E6 proteins, ii) the second peptide library entailed LXXLL motifs from distinct host proteins, putative or published E6 targets. Our interaction results allowed the identification of high-affinity LXXLL motifs that could be crystallized in complex with E6 proteins. In addition, these data complement structural information for a better understanding of E6 recognition features in terms of LXXLL motif. We confirmed that E6AP LXXLL motif is bound by α-genus HPV E6 proteins while MAML1 LXXLL is recognized by cutaneous β and ɣ-genus E6 proteins (Brimer *et al.*, 2017). In addition, our results indicate that β-genus HPV E6 proteins interact with an LXXLL motif located within p53 TAD domain. This interaction might prevent p300/p53 interaction by competitive binding. This finding could explain the disruption of p53 transactivation induced by β-genus HPV E6 (White *et al.*, 2014) and not by α-genus HPV E6 proteins (Bañuelos *et al.*, 2003), since the latter do not bind p53 LXXLL motif.

The project in collaboration with Anouk Willemsen and Ignacio Bravo allowed us to study the LXXLL interaction preferences of α-genus ancestral E6 proteins. We examined four E6 proteins from common ancestors of: i) the whole α-genus, ii) low-risk cutaneous HPVs, iii) HPV types causing genital warts and iv) mucosal HPV types, including both high and low oncogenic risk. Our preliminary results suggest that the binding to E6AP LXXLL motif is a feature shared by all α-genus regardless of their tropism. We also identified that the ancestral cutaneous α-genus E6 protein can interact with LXXLL motifs from paxillin and HIC5, similarly to the cutaneous β-genus 38E6 and 49E6 and ɣ-genus 197E6. This result, which remain to be confirmed by further assays, suggests a convergent evolution for cutaneous E6 proteins.

We took advantage of our expertise in quantifying protein-motif interactions to decipher the interaction of E6AP with a cellular E3 ubiquitin ligase named HERC2. The minimal interaction unit of E6AP required for interacting with HERC2 RLD2 domain was mapped using overlapping peptides. The 15-mer interacting peptide that was identified in this study could be crystallized in complex with HERC2 RLD2 domain, providing for the first time the structural basis of E6AP/HERC2 interaction.

The CCR4-Not complex as a target of β2 E6 oncoproteins

Our peptide library entailing LXXLL motifs from various published targets of E6 oncoproteins provides insight on the minimal interaction motifs mediating the interaction. However, our peptide array did not include any motif from a subunit of the CCR4-Not complex, which was published as a target of E6 proteins from β2-HPV types 17a and 38 and HPV197 (Grace and Münger, 2017; White *et al.*, 2012a). We identified good candidates as LXXLL motifs recognized by HPV E6 proteins, in particular in the subunits NOT11 and TAB182. The interactions between these peptides and our set of MBP-E6 proteins will be tested to complement our publication reporting E6/LXXLL interaction preferences.

Perspectives for extending the number of identified LXXLL motifs targeted by E6 proteins

Our workflow combining holdup assay and SPR provides a robust tool for the characterization of LXXLL motif-containing proteins targeted by E6 oncoproteins. The first results generated allow the affinity-based ranking of LXXLL motifs for each tested E6 protein. However, even though our peptide library entails the main E6 targets reported in the literature, we did not perform an unbiased screening including all accessible protein targets in the human proteome. During the third year of my thesis, I attempted to select E6-binding peptide motifs by proteomic peptide phage display. The procedure that I applied in Uppsala University was optimized for the selection of transient, low affinity binding peptides (Davey *et al.*, 2017). Unfortunately, no relevant binding motif was enriched for any of the five tested E6 proteins. It is probably due to the biochemical instability of purified MBP-E6, which are likely to oligomerize during the panning cycles despite the adapted protocol. In order to confirm our results and possibly identify novel E6 targets within the whole human proteome, we consider using

an AP-MS protocol adapted for low-interaction binders. We can take advantage of our expertise in isolating biologically active MBP-E6 proteins overexpressed in bacteria to prepare biotinylated E6 baits to be immobilized on streptavidin resin. Non-transfected mammalian cell extracts can be incubated with resin saturated with MBP-E6 proteins and be affinity-purified, with few gentle washing steps in order to preserve transient interactions. Another interesting possibility is to perform a holdup assay and measure the depletion of host proteins after incubation of the cell extract with E6 proteins immobilized on beads. This can be achieved provided that the target human proteins are sufficiently abundant in the cell extract. This strategy has already been successfully applied to another project studied in the team, involving PDZ-PBM interactions (data not shown). To cross-validate this approach and test whether the interaction with an E6 oncoprotein is mediated by an LXXLL motif, we plan to use GPCA as an orthogonal method. The advantage of the GPCA approach for cross-validating holdup data is that the interaction is probed in mammalian cells between two full-length proteins instead of peptide motifs. The two interacting proteins overexpressed in mammalian cells can be mutated in order to map the residues driving the interaction, as previously performed on a recent paper deciphering the interactions of E6 and E7 oncoproteins with the ubiquitin-proteasome system (Poirson *et al.*, 2017).

Hypothesis: LXXLL motif-containing E6 targets belong to a cellular interaction network whose central interactant is CBP/p300

Viral proteins evolve to highjack an existing interaction network in the host cell, by capturing or mimicking conserved binding motifs. It is this likely that HPV E6 oncoproteins evolved to target proteins belonging to a common interaction network, based on the recognition of their acidic LXXLL motif. For a long time, helical LXXLL motifs have been reported as mediating interactions involved in transcriptional regulation (Plevin *et al.*, 2005), which includes but is not limited to co-activators/nuclear receptors interactions (Heery *et al.*, 2001).

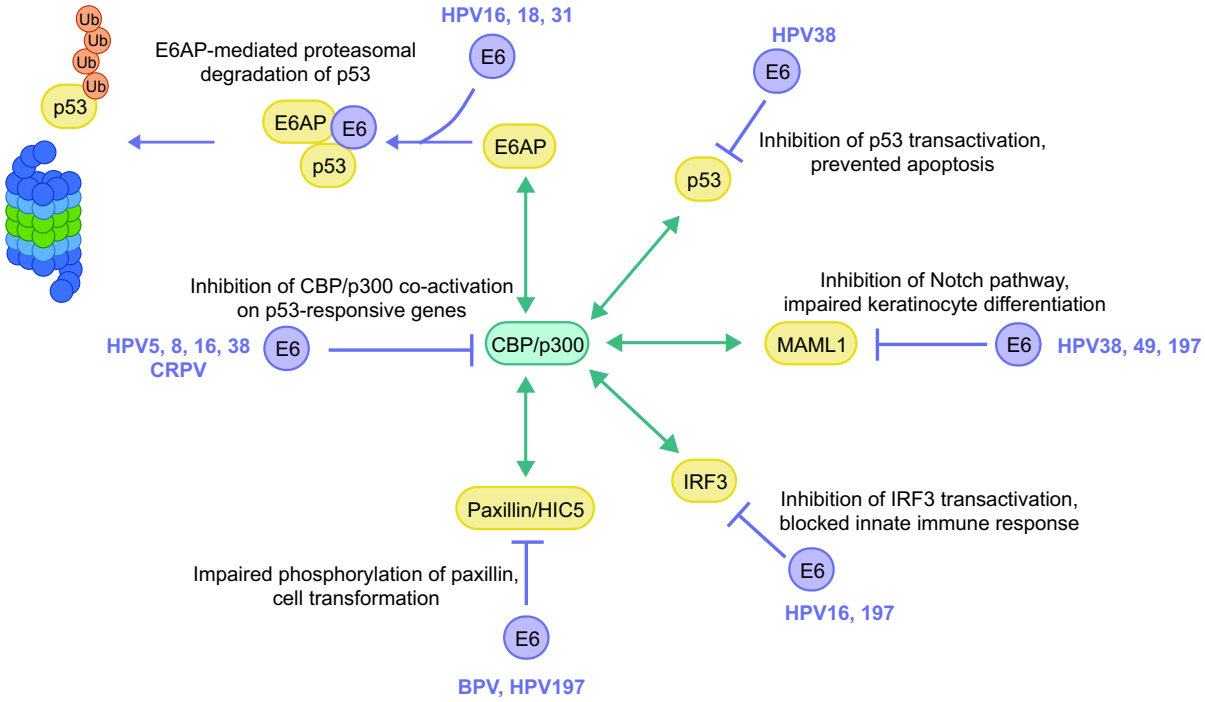


Figure 44: Schematic overview of host proteins targeted by HPV E6 oncoproteins. The LXXLL motif-containing proteins (indicated in yellow boxes) all interact with CBP/p300 co-activators (indicated in blue). E6 proteins from various HPV types (indicated in purple) interact either with CBP/p300 interaction partners through LXXLL recognition (for instance with E6AP, MAML1, IRF3) or with CBP/p300 itself. Each interaction has a biological consequence: deregulation of the cell cycle, impaired differentiation or prevented antiviral response.

The main LXXLL motif-containing targets of E6 oncoproteins are involved in transcriptional regulation and share a common interaction partner: the histone acetyltransferases CBP and its close homolog CBP. The ubiquitin-ligase E6AP (Catoe and Nawaz, 2011), the tumor suppressor p53 (Krois et al., 2016), the transcription factors MAML1 (Wallberg et al., 2002) and IRF3 (Qin et al., 2005) and the transcriptional adapter ADA3 (Germaniuk-Kurowska et al., 2007) were all reported to interact with CPB/p300. The CCR4-Not complex, which is also targeted by some E6 proteins, is an interaction partner of CBP/p300 (Sharma et al., 2016).

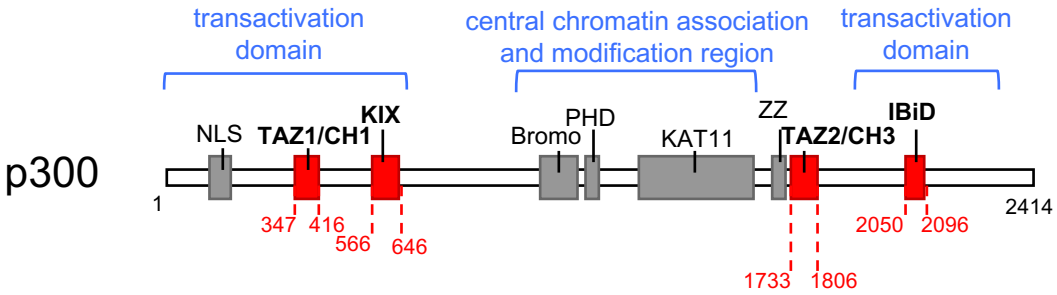


Figure 34: Domain organization of p300. The p53-interacting domains are indicated in red and the function of the three main regions of p300 is indicated in blue. NLS: Nuclear Localization Signal; TAZ: transcriptional-adaptor zinc-finger domain; CH: cysteine/histidine-rich region; KIX: kinase inducible domain of CREB interacting domain; Bromo: bromodomain; PHD: plant homeodomain finger; KAT11: lysine acetyltransferase domain; ZZ: ZZ-type zinc finger domain; IBiD: IRF3-interacting domain (Chan and Thangue, 2001).

The KIX domain of CBP (see **Figure 34**) was reported to interact with LXXLL motifs, in particular with an LXXLL motif located in the transactivation domain of c-Myb (Zor *et al.*, 2004). During my thesis, I produced and purified the KIX domain from CBP and tested whether it was able to bind LXXLL motifs from E6 target proteins. I could not detect any significant interaction with any of the peptides from the library of cellular targets, neither by holdup assay nor by SPR (data not shown). Some E6 proteins directly target p300, such as α-genus HPV 16 E6 and β-genus HPV 38, 5 and 8. E6 oncoproteins from HPV 5 and 8 appear to associate stronger with p300 and induce its proteasomal degradation without recruiting E6AP. The degradation of p300 protein is regulated by phosphorylation by AKT, a RAC-Alpha Serine/Threonine protein Kinase. HPV8 E6 was reported to bind to AKT phosphorylation site within p300 C-terminus, thus competing with AKT kinase for p300 binding. By preventing p300 phosphorylation by AKT, HPV8 E6 destabilizes p300 protein level, which leads to the inhibition of keratinocyte differentiation (Howie *et al.*, 2011). The interaction of HPV38 E6 with p300 was reported to be necessary for immortalization of human foreskin keratinocytes. Similarly to CRPV E6 (from Cottontail Rabbit Papillomavirus), HPV38 E6 binds to p300 without inducing its degradation. However, the interaction of p300 with CRPV E6 prevents p53 acetylation, which is essential for stabilizing p53 binding to the p21 promoter. Thus, by inhibiting p21 induction, CRPV E6 prevents p53-induced apoptosis (Muench *et al.*, 2010). HPV16 E6 was reported to bind three regions within CBP and p300: CH1 and CH3 domains and C-terminus. This interaction resulted in the inhibition of CBP/p300 co-activation, which includes p53- and NF-κB-responsive elements (Patel

et al., 1999; Zimmermann *et al.*, 1999). Hence, the fact that E6 proteins target diverse p300-interacting proteins and p300 itself suggests that the conserved interaction network inhibited by E6 proteins is driven by CBP/p300. Blocking different factors from this network is advantageous, since it leads to the inhibition of apoptosis (p53), innate immune response (IRF3) and keratinocyte differentiation (MAML1). This hypothesis, proposed in (Suarez and Travé, 2018), should be explored for future studies aiming at characterizing the interactome of E6 oncoproteins.

Posters

Hijacking of LXXLL motifs in human proteins by HPV E6 oncoprotein

- *Quantitative & Structural analysis* -

Bonhoure A., Forster A.[1], Ould Babah K.[1], Altschuh D.[1], Kostmann C., Suarez I. P., Cousido-Siah A., Mitschler A., Podjarny A., Nominé Y. and Travé G.

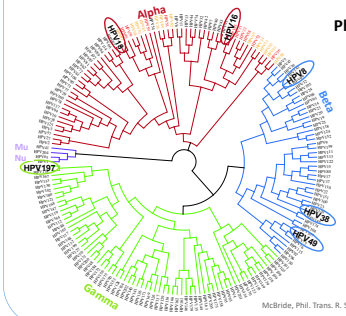
1: Former team member

Institut de **G**énétique et de **B**iologie **M**oléculaire et **C**ellulaire – 1 rue Laurent Fries – BP 10142 – 67404 Illkirch CEDEX
anna.bonhoure@igbmc.fr

E6 oncoprotein from Human Papillomaviruses

Human PapillomaViruses (HPV)

- Small DNA viruses infecting epithelia
- Pathogenicity-based categories:
  - Low-risk: warts and condylomas
  - High-risk: Cancer (cervical, head & neck, skin, anal)



McBride, Phil. Trans. R. Soc. B 2017

Phylogenic classification

- 225 HPV types, organized in 5 genera

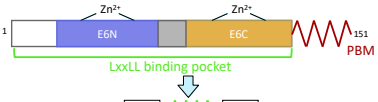
α : mucosal tropism
HPV16 → 61 % of cervical cancer
HPV18 → 10 % of cervical cancer
+ head and neck cancers…
de Sanjose *et al.*, Lancet Oncol. 2010

β : cutaneous tropism (in general)
HPV5 and HPV8 → provoke squamous carcinoma in patients with epidermodysplasia verruciformis
HPV49 ,HPV38

γ: cutaneous tropism
HPV197 → skin tumors
Arroyo Mühr *et al.*, Int. J. Cancer 2015

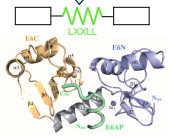
HPV oncoproteins

- 2 oncoproteins: E6 and E7
- E6: wide interaction network, in particular with cellular proteins through domain-motifs interactions



LXXLL-containing proteins

	Affects	
p53 degradation	←	E6AP
Innate immune response	←	IRF3
Notch signalling	←	MAML1
Focal adhesion	←	Paxillin
Transcriptional co-activators	←	P300/CBP

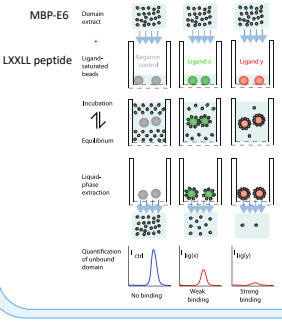


Interaction of E6 hydrophobic pocket with **LXXLL motifs**
Zanier *et al.* 2013 / PDB: 4GIZ

OBJECTIVE
Build a "quantitative" database of the E6 interactome:
- To characterize the binding preferences of different HPV types
- To provide tools for the design of specific inhibitors

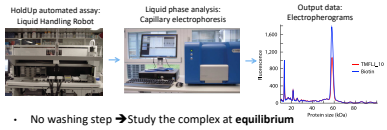
Methods

◆ HoldUp: Comparative chromatographic retention assay



Vincentelli *et al.* Nat Methods 2015

Experimental Setup



- No washing step ➔Study the complex at **equilibrium**
- **High-throughput** approach with quantitative information
- Detection of affinities in the **µM** range

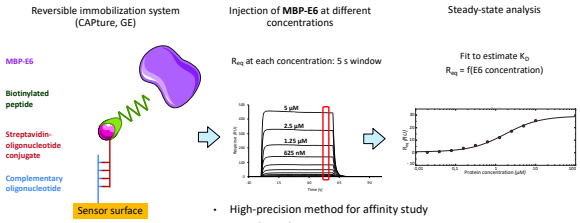
$$BI = \frac{I_{cnt} - I_{lig}}{I_{cnt}}$$

K_D evaluation

$$K_D \approx \frac{I_{lig} - I_{bg}}{I_{ctrl} - I_{bg}} \times C_{pept}$$

⇒ **Step 1**: Quantitative screening

◆ Surface Plasmon Resonance

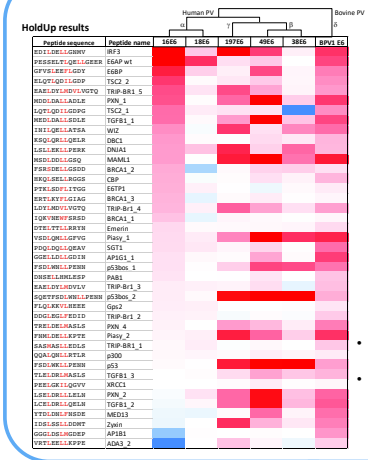


- High-precision method for affinity study
- **Low-throughput**

⇒ **Step 2**: Affinity estimation of strongest binders

Interaction data from 2 peptide arrays

Putative cellular targets



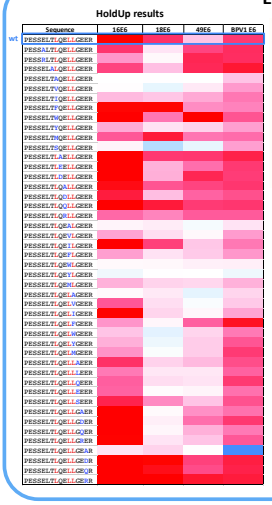
SPR data

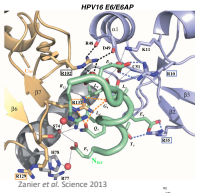
LXXL Peptide	MBP-E6	estimated K_D (µM)	Rmax (RU)	χ^2 (RU)	Rmax exp/theo
E6AP wt	16E6	0.71	371.4	2.8	1.57
IRF3	16E6	0.23	1221.0	18.1	1.48
E6AP wt	18E6	4.73	245.9	0.6	0.35
IRF3	197E6	0.78	1152.0	7.2	0.27
MAML1	197E6	1.42	656.2	3.2	0.18
PXN_1	49E6	2.50	660.1	5.6	0.74
p53_hum	49E6	5.87	226.3	0.9	0.26
MAML1	49E6	4.68	157.4	0.3	1.08
p53_hum	38E6	11.28	63.0	1.0	0.37
Piasy_1	38E6	30.93	51.7	0.2	0.12

- Reshuffling of E6 interaction preferences between the different HPV types
- Our *in vitro* protein-motif interaction data are in adequation with proteomic assays

White & Howley Virology 2013

E6AP single-point mutants



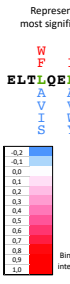


Zanier *et al.* Science 2013

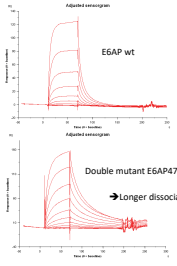
18E6 has high affinity for a **double mutant** of E6AP LXXLL motif

E6AP wt PESSELTLQELLGEER
➔ **E6AP47** and **E6AP9** bind stronger to 18E6 as compared to WT LxxLL
E6AP47 PESSELTLQELLGERR
E6AP9 PESSELTFQELLGEER

Representation of the most significant mutations

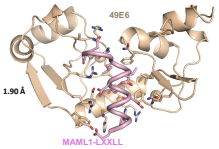


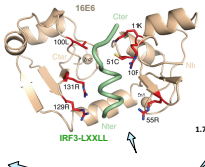
SPR data

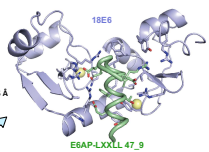


E6AP47_9 PESSELTFQELLGERR

X-ray structures







HoldUp / SPR data

197E6 / MAML1-LXXLL
Crystallization screenings ongoing
to be tested: IRF3-LXXLL

38E6 / MAML1-LXXLL or p53-LXXLL
Purification trials: protein oxidation & aggregation
to be tested: MBP-E6 / SUMO-LXXLL

16E6 recruits E6AP ubiquitin ligase via its LXXLL motif and then interacts with p53 central core domain
➔ Proteasomal degradation of p53 → p53 cell level

38E6 interacts with p53 LXXLL motif
➔ Stabilization of p53 → p53 cell level

References

Zanier, K., Charbonnier, S., Sidi, A.O.M.O., McEwen, A.G., Ferrario, M.G., Poussin-Courmontagne, P., Cura, V., Brimer, N., Babah, K.O., Ansari, T., et al. (2013). Structural basis for hijacking of cellular LxxLL motifs by papillomavirus E6 oncoproteins. Science *339*, 694–698.

Vincentelli, R., Luck, K., Poirson, J., Polanowska, J., Abdat, J., Blémont, M., Turchetto, J., Iv, F., Ricquier, K., Straub, M.-L., et al. (2015). Quantifying domain-ligand affinities and specificities by high-throughput holdup assay. Nature Methods *12*, 787–793.

Charbonnier, S., Zanier, K., Masson, M., and Travé, G. (2006). Capturing protein–protein complexes at equilibrium: The holdup comparative chromatographic retention assay. Protein Expression and Purification *50*, 89–101.

Structural and biophysical study of E6AP– HERC2 ubiquitin ligases interaction

Demenge A., Bonhoure A., Kostmann C., Roth A., Cousido-Siah A., Mitschler A., Podarny A. and Travé G.

Institut de Génétique et de Biologie Moléculaire et Cellulaire – 1 rue Laurent Fries – BP 10142 – 67404 Illkirch CEDEX
anna.bonhoure@igbmc.fr

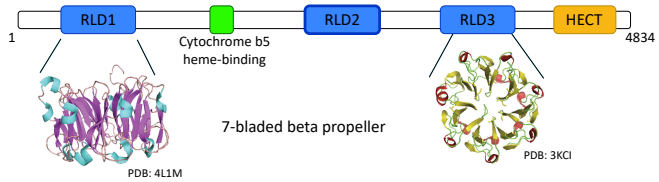
INTRODUCTION

UBE3A (coding for E6AP protein) and *HERC2:* genes located on chromosome 15q11-13, both E3 ubiquitin ligases

Coordinated gene expression: their loss of function by mutation results in Angelman syndrome, a neurodevelopmental disorder

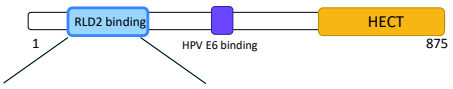
HERC2

RLD: RCC1-like domain
RLD2 shares 45 % and 57 % identity with RLD1 and RLD3 respectively



E6AP

Founding member of HECT family of E3 ubiquitin ligases
Interacts with E6 oncoprotein from HPV and degrades p53 tumor suppressor



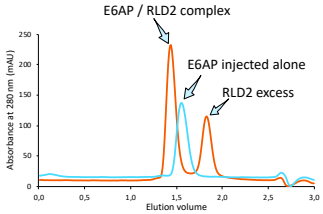
Kühnle *et al.* reported interaction between RLD2 domain (HERC2) and residues 150-200 (E6AP)

Physical interaction: HERC2 binds UBE3A and acts as an allosteric regulator

What is the structural basis of E6AP– HERC2 interaction?

BIOPHYSICAL STUDY

E6AP full-length and RLD2 domain from HERC2 were produced in *E. coli* and purified as MBP fusions



Interaction confirmed with full-length E6AP by SEC

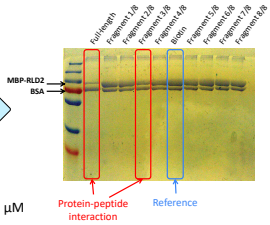
We designed overlapping peptides to reduce the region of E6AP responsible for RLD2 interaction

E6AP 150 - 200
HTKEELKSLQAKDEDKDEDEKEKAACSAAAMEEDSEASSSRIGDSSQGDNN

HTKEELKSLQAKDED
LKSLQAKDEDKDEDE
AKDEDKDEDEKEKAA
KDEDEKEKAACSAAA
KEKAACSAAAMEEDS
CSAAAMEEDSEASSS
MEEDSEASSSRIGDS
EASSSRIGDSSQGDNN
,,

We analyzed by SPR the interaction RLD2 / E6AP 150-200: $K_D \approx 0{,}6$ µM

HoldUp interaction assay



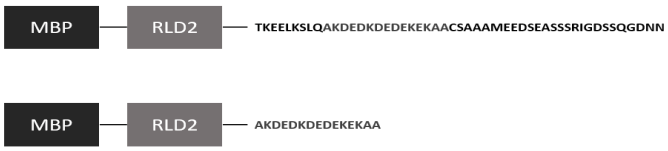
Sample	Binding intensity
Biotin	0,00
Full-length	0,87
Fragment 1/8	0,03
Fragment 2/8	0,00
Fragment 3/8	0,42
Fragment 4/8	0,12
Fragment 5/8	0,07
Fragment 6/8	-0,13
Fragment 7/8	0,02
Fragment 8/8	-0,04

Fragment 3/8:
AKDEDKDEDEKEKAA

We identified 15 residues in E6AP that seem particularly involved in the interaction with RLD2

X-RAY CRISTALLOGRAPHY

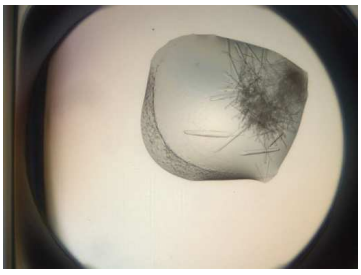
We produced and purified two different constructs. The first one is a fusion of MBP, RLD2 and peptide E6AP 150-200. The second is a fusion of MBP, RLD2 and fragment 3/8.



Purification of the protein

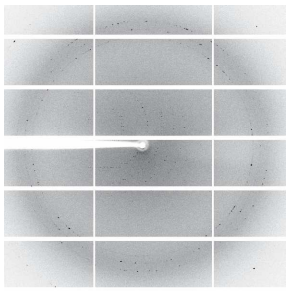
1) Affinity column amylose to purify MBP-protein fusion
2) Ultracentrifugation to eliminate aggregated protein
3) gel filtration chromatography to improve the pureness of the protein
4) Concentration of the protein
5) Crystallization assay
6) Diffraction of X ray by the crystal
7) Determination of the structure by molecular replacement

crystals and structure obtained with the construct MBP-RLD2-E6AP(peptide 3/8)

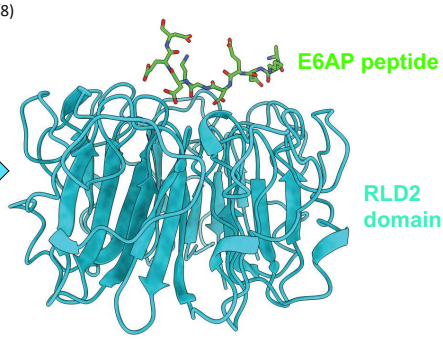


Crystals: 40µm 40µm 300µm

Crystallization condition:
Protein: 17,5 mg/mL
0.1M Tris HCl pH7.5
0.2M lithium sulphate
0.2M sodium chloride
5% Polyethylene glycol
4° C



Diffraction pattern register at SLS synchrotron source at Paul Scherrer Institut



RCC1 like domain (RLD) family
7 bladed beta propeller
Binding of the peptide on the "top" side
Analysis of domain/peptide interface (ongoing) will guide mutagenesis experiments to probe function of the complex in neuronal development

References

Kalsner, L., and Chamberlain, S.J. (2015). Prader-Willi, Angelman, and 15q11-q13 Duplication Syndromes. Pediatric Clinics of North America *62*, 587–606.

Kühnle, S., Kogel, U., Glockzin, S., Marquardt, A., Ciechanover, A., Matentzoglu, K., and Scheffner, M. (2011). Physical and Functional Interaction of the HECT Ubiquitin-protein Ligases E6AP and HERC2. Journal of Biological Chemistry *286*, 19410–19416.

Charbonnier, S., Zanier, K., Masson, M., and Travé, G. (2006). Capturing protein–protein complexes at equilibrium: The holdup comparative chromatographic retention assay. Protein Expression and Purification *50*, 89–101.

Vincentelli, R., Luck, K., Poirson, J., Polanowska, J., Abdat, J., Blémont, M., Turchetto, J., Iv, F., Ricquier, K., Straub, M.-L., et al. (2015). Quantifying domain-ligand affinities and specificities by high-throughput holdup assay. Nature Methods *12*, 787–793.

One-step purification of monomeric E6 oncoproteins from Human Papillomaviruses

Bonhoure A., Demenge A. and Travé G.

Institut de Génétique et de Biologie Moléculaire et Cellulaire – 1 rue Laurent Fries – BP 10142 – 67404 Illkirch CEDEX
anna.bonhoure@igbmc.fr

Human PapillomaViruses (HPV)

- ✓ Small DNA viruses infecting epithelia
- ✓ Pathogenicity-based categories:
  - Low-risk: warts and condylomas
  - High-risk: malignant proliferation (cervix cancer, "head and neck" cancer)

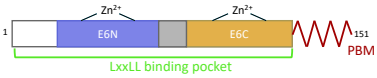
Examples of HPV types

High-risk mucosal	High-risk cutaneous
HPV16, 18	**HPV5, 8, 38**
Low-risk mucosal	**Low-risk cutaneous**
HPV6, 11, 32	**HPV1, 3, 9**

High-risk mucosal HPVs

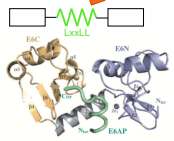
- ✓ 2 oncoproteins: E6 and E7
- ✓ E6: wide interaction network, in particular with cellular proteins through domain-motifs interactions

Two binding motifs: a **pocket** that captures short sequence motifs with the consensus LxxLL / a conserved C-terminal PDZ binding motif (**PBM**)

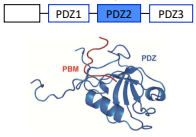


LxxLL-containing proteins

	Affects	
p53 degradation	←	E6AP
Innate immune response	←	IRF3
Notch signalling	←	MAML1
Focal adhesion	←	Paxillin
Transcriptional co-activators	←	P300/CBP



Interaction of E6 hydrophobic pocket with **LxxLL motifs**
Zanier *et al.* 2013 / 4GIZ.pdb



Interaction of E6 C-terminus motif (**PBM**) with **PDZ domains**
Ramirez *et al.* 2015

PDZ domain-containing proteins

~270 PDZ domains distributed over ~140 proteins in the human proteome

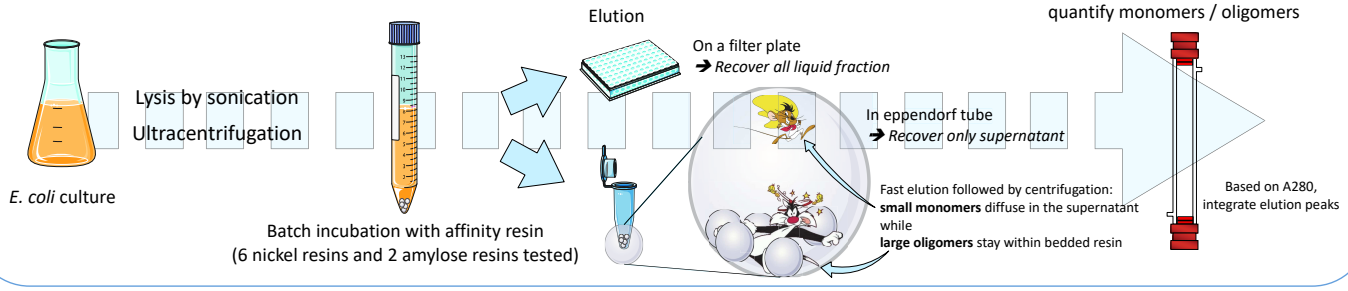
	Affects	
MAGI proteins	→	Cell junction scaffold
SCRIB	→	Cell polarity
DLG proteins	→	

PROBLEM: Even solubilized by MBP (Maltose-Binding Protein), E6 proteins can form soluble aggregates unable to interact with LxxLL motifs for protein-protein assays

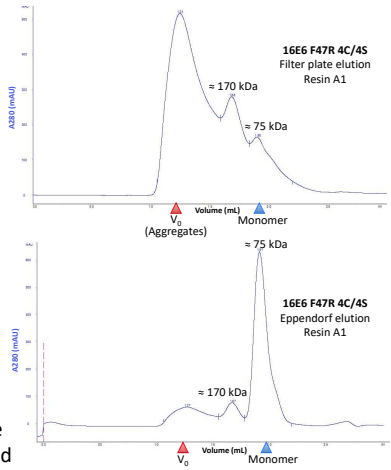
OBJECTIVE: Purify fully active monomers

➡ Develop a fast and simple protocol to purify **monomeric** E6 from different HPV types in parallel

PURIFICATION PIPELINE



Gel-filtration chromatograms
Calibrated S200 5/150 GL



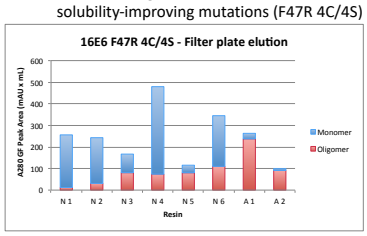
Step 1
Changing the elution method

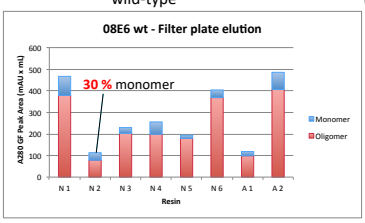
Step 2
Optimize incubation time between cleared lysate and affinity resin

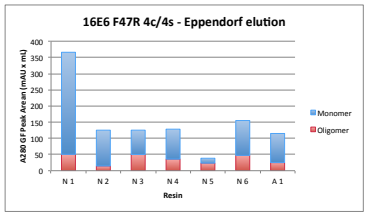
Tests performed on two E6 proteins

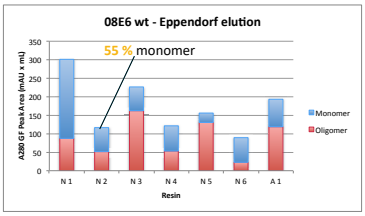
E6 from high-risk mucosal **HPV16**, with solubility-improving mutations (F47R 4C/4S)

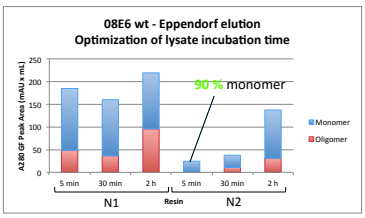
E6 from high-risk cutaneous **HPV08**, wild-type











	Affinity	Nickel ligand	Features	Metal Ion Capacity / Binding Capacity
N 1	Nickel	NTA	6 % cross-linked agarose	approx 15 µmol Ni^{2+}/mL gel
N 2	Nickel	IDA	10 % cross-linked agarose, small pores excluding particles > 500 kDa	20-40 µmol Ni^{2+}/mL gel
N 3	Nickel	IDA	6 % cross-linked agarose	20-40 µmol Ni^{2+}/mL gel
N 4	Nickel	NTA	6 % cross-linked agarose	5-19 µmol Ni^{2+}/mL gel
N 5	Nickel	IDA	6 % agarose	5-19 µmol Ni^{2+}/mL gel
N 6	Nickel	NTA	6 % agarose	5-19 µmol Ni^{2+}/mL gel
A 1	Maltose		Cross-linked agarose	7 mg MBP-protein/mL gel
A 2	Maltose		6 % cross-linked agarose	10 mg MBP-protein/mL gel

The purified protein is significantly **enriched in monomer**, after one single purification step
But the total protein amount is **decreased**

➡ Less protein but higher **quality**

Are these monomeric protein able to interact with their known LxxLL peptides?

UPCOMING: Perform interaction assays on these purified proteins to test whether they are biologically active

References

Zanier, K., Charbonnier, S., Sidi, A.O.M.O., McEwen, A.G., Ferrario, M.G., Poussin-Courmontagne, P., Cura, V., Brimer, N., Babah, K.O., Ansari, T., et al. (2013). Structural basis for hijacking of cellular LxxLL motifs by papillomavirus E6 oncoproteins. Science *339*, 694–698.

Ramirez, J., Poirson, J., Foltz, C., Chebaro, Y., Schrapp, M., Meyer, A., Bonetta, A., Forster, A., Jacob, Y., Masson, M., et al. (2015). Targeting the Two Oncogenic Functional Sites of the HPV E6 Oncoprotein with a High-Affinity Bivalent Ligand. Angewandte Chemie International Edition *54*, 7958–7962.

Vincentelli, R., Luck, K., Poirson, J., Polanowska, J., Abdat, J., Blémont, M., Turchetto, J., Iv, F., Ricquier, K., Straub, M.-L., et al. (2015). Quantifying domain-ligand affinities and specificities by high-throughput holdup assay. Nature Methods *12*, 787–793.

Charbonnier, S., Zanier, K., Masson, M., and Travé, G. (2006). Capturing protein–protein complexes at equilibrium: The holdup comparative chromatographic retention assay. Protein Expression and Purification *50*, 89–101.

Appendix: Introduction en français

1 Biologie des papillomavirus

1.1 Généralités

Les Papillomavirus sont de petits virus non enveloppés à ADN circulaire double-brin. Ils infectent les épithéliums des mammifères, mais également des oiseaux, poissons et reptiles. Certains d'entre eux ne provoquent que des verrues, condylomes et proliférations bénignes appelées papillomes. D'autres peuvent provoquer différents types de cancer (col de l'utérus, et autres cancers ano-génitaux, ORL, peau).

1.1.1 Historique

Les verrues génitales chez l'humain ont fait l'objet d'écrits dont les plus anciens datent de l'Antiquité (Oriel, 1971). Cependant, le lien de causalité entre les papillomavirus et les verrues n'a pu être fait qu'à partir du vingtième siècle, grâce aux progrès de la méthode scientifique. Le papillomavirus a dans un premier temps été mis en évidence chez les animaux, notamment via les travaux de Richard Shope sur les lapins (Shope and Hurst, 1933). Chez le lapin, l'implication du papillomavirus dans les cancers de la peau a été rapidement identifiée (Rous and Beard, 1935), alors qu'on considérait que le papillomavirus humain (HPV) ne causait que des verrues et proliférations bénignes (papillomes). En 1950, des particules virales de HPV sont visualisées pour la première fois par microscopie électronique (Strauss *et al.*, 1950). C'est dans les années 1970 que les études sur les virus oncogènes ont connu un essor, avec de premières tentatives d'isolation d'ADN viral à partir d'échantillons tumoraux (Wolf *et al.*, 1975). Par ses études sur des verrues et lésions cutanées, le virologue français Gérard Orth identifie les premiers types de HPV (Orth *et al.*, 1977, 1978). Les travaux menés par le médecin allemand Harald zur Hausen ont marqué un tournant dans la caractérisation des papillomavirus humains : après avoir identifié un nouveau type de HPV présent dans des condylomes (Gissmann and zur Hausen, 1980), il a été le premier à isoler les génome de HPV16 et HPV18 à partir de biopsies de cancer du col de l'utérus (Boshart *et al.*, 1984; Durst *et al.*, 1983). Ces découvertes majeures ainsi que ses articles mettant en évidence l'implication du HPV dans divers cancers humains (zur Hausen, 1991, 2002) lui ont valu le prix Nobel de Médecine en 2008.

En parallèle, des études sont menées pour caractériser les propriétés d'un papillomavirus infectant le bétail: le papillomavirus bovin (BPV) (Cheville and Olson, 1964; Pamukcu *et al.*, 1959), dont les génomes des deux premiers types sont isolés en 1978 (Lancaster and Olson, 1978). Par la suite, il s'est avéré que certaines des caractéristiques du BPV s'appliquent également au HPV (Munday, 2014). Des études ont montré que le BPV est capable d'infecter non seulement les bovins mais également les hamsters, les ânes ainsi que les chevaux (Koller and Olson, 1972). Le BPV a en particulier une forte prévalence chez les chevaux : il provoque des tumeurs cutanées invasives appelées sarcoïdes qui peuvent causer une ulcération et gêner la marche selon leur localisation (Bogaert *et al.*, 2008; Taylor and Haldorson, 2013). A ce jour, aucune infection de HPV chez d'autres animaux n'a été rapportée à l'exception d'une étude rapportant l'amplification d'ADN issu de HPV9 dans des échantillons de papillome félin (Munday *et al.*, 2007). Ce résultat n'ayant pas été reproduit depuis sa publication, il est aujourd'hui probablement attribué à une contamination des échantillons (Munday *et al.*, 2019). Les similarités observées entre HPV et BPV ont fait de ce dernier un organisme modèle pour l'étude du papillomavirus, tant pour la compréhension des mécanismes infectieux et carcinogènes que pour le développement de vaccins (Campo, 2006).

1.1.2 Classification

Il existe une grande diversité de papillomavirus, infectant non seulement les êtres humains mais les Vertébrés au sens large (mammifères, oiseaux, serpents, tortues, poissons). Les papillomavirus sont fréquemment désignés par le nom de leur hôte (*Felis domesticus* PV ; Papillomavirus Humain ; Papillomavirus Bovin...).

Les papillomavirus ont dans un premier temps été classés dans la famille des *Papovaviridae*, qui regroupait les Papillomavirus, les Polyomavirus dont le virus simien 40 (SV-40). Les caractéristiques à l'origine de la création de ce taxon sont l'absence d'enveloppe ainsi que le génome se présentant sous forme d'ADN circulaire double-brin. Cependant, ces virus présentent des organisations génomiques bien distinctes. En 2002, le Comité International de Taxonomie des Virus (ICTV) prend la décision de scinder le groupe des *Papovaviridae* en deux groupes : les *Papillomaviridae* et les *Polyomaviridae* (de Villiers *et al.*, 2004). Il est toutefois à noter que les Papillomavirus et les Polyomavirus présentent des homologies de séquence protéique significatives

et l'identification d'hybrides papilloma-polyomavirus BPCV1 et 2 appuie l'hypothèse d'une origine commune (Rector and Van Ranst, 2013).

Cet exemple illustre l'impact positif qu'ont eu les développements technologiques sur l'étude des papillomavirus depuis leur identification, permettant une meilleure compréhension de leurs propriétés. La mise au point de lignées cellulaires permettant la culture de HPV *in vitro* a également facilité l'identification des nouveaux types viraux (de Villiers *et al.*, 2004). Enfin, les premières méthodes de séquençage ADN ont permis de cloner puis de séquencer les génomes complets de HPV, en commençant par HPV1 au début des années 1980 (Danos *et al.*, 1980, 1982). Depuis les années 2010, le séquençage nouvelle génération à haut débit a permis de décupler le nombre de types de HPV identifiés (Arroyo *et al.*, 2013), en particulier les HPV ayant un tropisme cutané (Ekström *et al.*, 2011).

La classification actuellement en vigueur est basée sur l'homologie de séquence du gène codant pour la protéine de capside L1, étant donné qu'il s'agit du gène le mieux conservé du génome viral. Les papillomavirus sont regroupés dans la famille des *Papillomaviridae*, divisée en genres désignés par des lettres grecques (alpha, beta, gamma, mu, nu...) eux-mêmes sous divisés en espèces puis en types. Les HPV appartenant au même genre ont au minimum 60 % d'identité de séquence nucléotidique du cadre de lecture ouvert de L1, contre 70 à 80 % pour les HPV de la même espèce. Un nouveau type de HPV est défini par au minimum 10 % de variation de séquence avec tout autre type connu. Enfin, les génomes ayant moins de 10 % de différence de séquence L1 sont appelés des variants (de Villiers, 2013).

Tout isolat d'un nouveau type potentiel de HPV est envoyé au Centre de Référence International HPV, actuellement hébergé à l'Institut Karolinska situé à Stockholm, en Suède. Ce centre a pour rôle de confirmer, d'enregistrer et de référencer dans l'arbre phylogénétique tous les nouveaux types de HPV ainsi que de conserver des échantillons ADN de tous les génomes connus de HPV (Bzhalava *et al.*, 2015). Récemment, un centre de référence pour les Papillomavirus infectant les animaux a été implanté dans l'Université d'Arizona, Tucson, États-Unis (Van Doorslaer and Dillner, 2019). De plus, le Papillomavirus Épisteme (PaVE) est une base de données qui a été créé afin de compiler l'ensemble des génomes annotés et de fournir des outils d'alignement de séquence et de construction d'arbres phylogénétiques spécialement

dédiés aux papillomavirus (Van Doorslaer *et al.*, 2017). A ce jour, 227 types de référence de HPV ont été identifiés et 215 papillomavirus ciblant les animaux. La **Figure 45** représente l'arbre phylogénétique des HPV identifiés à ce jour (Papillomavirus Épistème, pave.niaid.nih.gov, consulté le 25 mars 2020).

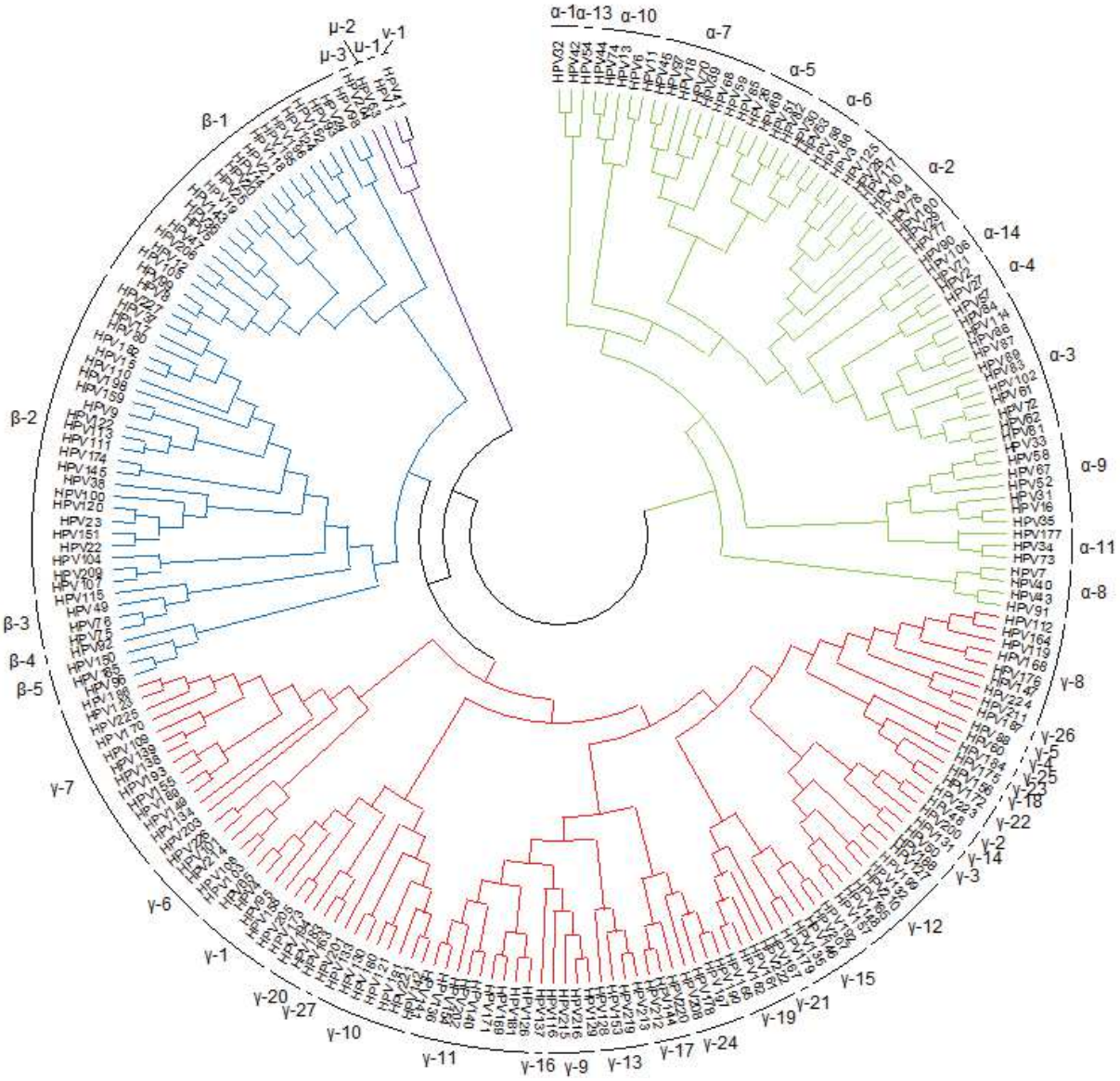


Figure 45: Arbre phylogénétique des HPV, d'après le Centre International de Référence HPV (hpvcenter.se, consulté le 25 mars 2020). Les genres alpha, beta, gamma, mu et nu sont représentés respectivement en vert, bleu, rouge et violet. Les espèces sont indiquées par des nombres.

Étant donné la grande diversité d'effets pathologiques provoqués par les HPV, d'autres classifications basées sur leur risque oncogène ainsi que sur leur tropisme existent en marge de la classification taxonomique.

Les HPV dits à "haut risque" peuvent provoquer des cancers tandis que les HPV à "bas risque" ne causent que des proliférations bénignes (condylomes, verrues). Le **Tableau 1** résume l'ensemble des HPV considérés comme carcinogènes et notamment les types reconnus en 2012 par l'Organisation Mondiale de la Santé comme substances cancérigènes.

Risque oncogène	Genre	Espèce	Types de HPV
Élevé	Alpha	5	51
		6	56
		7	18 ; 39 ; 45 ; 59
		9	16 ; 31 ; 33 ; 35 ; 52 ; 58
Probablement élevé		7	68
Possiblement élevé		5	26 ; 69 ; 82
		6	30 ; 53 ; 66
		7	70 ; 85 ; 97
		9	67
		11	34; 73
	Beta	1	5*; 8*
Bas	Alpha	1	54
		8	40;43
		1	42
		13	54

Tableau 1: Classification des HPV par risque oncogénique. D'après (Muñoz *et al.*, 2006) et d'après les données de l'Agence Internationale de Recherche contre le Cancer (monographs.iarc.fr, consulté le 25 mars 2020). * Les HPV5 et 8 ne sont reconnus cancérigènes que chez les patients atteints d'Épidermodysplasie verruciforme.

D'autre part, on distingue les HPV selon leur tropisme tissulaire. Les HPV ayant un tropisme cutané infectent les épithéliums kératinisés, comme la peau. Les HPV à tropisme muqueux ciblent les épithéliums non-kératinisés, comme la muqueuse utérine ou l'oropharynx par exemple. On note que la plupart des HPV muqueux appartient au genre phylogénétique alpha, tandis que les genres beta, gamma, mu et nu regroupent une majorité de HPV cutanés. Cependant, ceci n'est pas une règle absolue. Certains HPV alpha ont un tropisme cutané (HPV2, 3, 7, 10, 27, 28, 57) et certains types comme ceux du genre beta 3 semblent capables de se localiser dans les deux types d'épithéliums (Gheit, 2019; Hampras *et al.*, 2017).

1.1.3 Capside, génome et cycle infectieux

1.1.3.1 Capside

Les papillomavirus sont des virus non-enveloppés, dotés d'une capside ayant un diamètre de 60 nm environ. La structure atomique des capsides de HPV1 et BPV1 a été déterminée pour la première fois en 1991, par cryo-microscopie électronique (Baker *et al.*, 1991). La capside est icosaédrique et a une symétrie T = 7. Elle se compose de 360 copies de la protéine de capside L1 et jusqu'à 72 copies de la protéine L2, organisées en 72 pentamères (Li *et al.*, 2016).

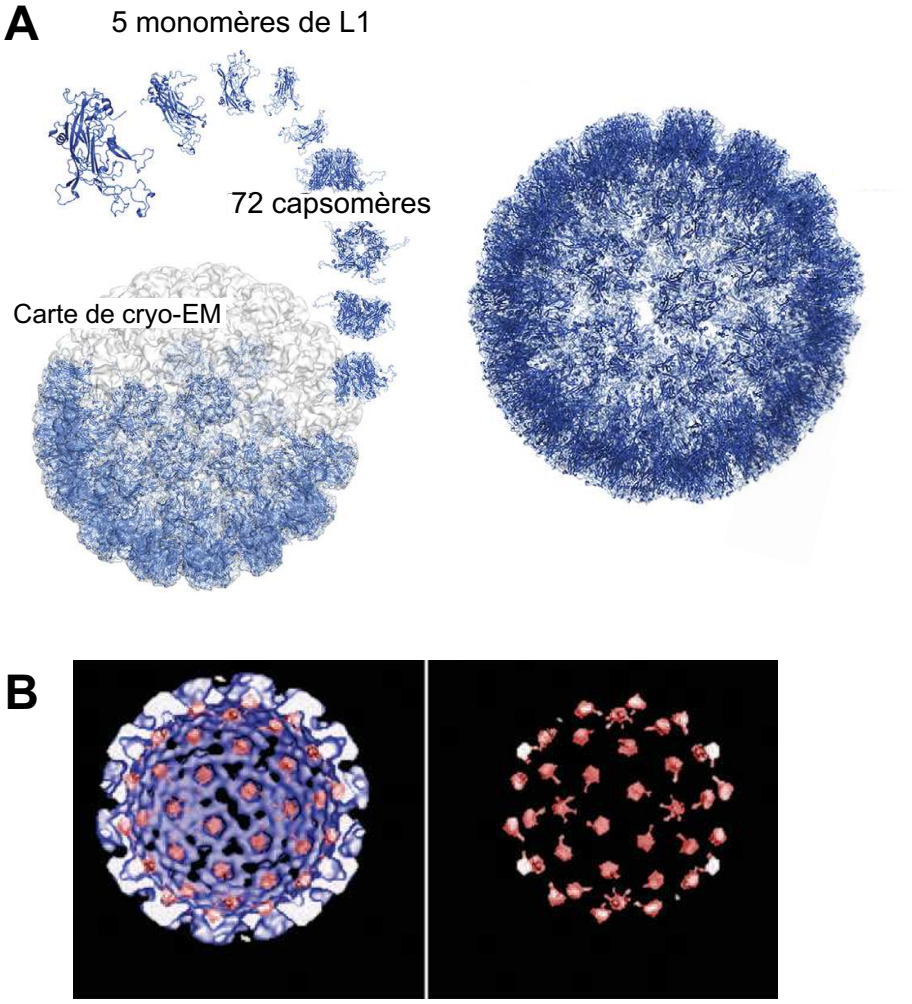


Figure 46 : Structure de la capside de HPV et positionnement des protéines L1 et L2. **A**. Structure de capside de HPV59 reconstituée avec la protéine majoritaire L1 uniquement. La capside est composée de 72 pentamères de L1. La structure a été obtenue en combinant des données de cristallographie avec une carte de cryo-microscopie électronique. Adapté de (Li *et al.*, 2016). **B**. Position de la protéine de capside minoritaire L2 dans une capside de HPV16. Les unités de protéine L2 sont représentées en rouge, superposée avec l'intérieur d'une capside composée uniquement de protéine L1 (gauche) ou seules (droite). Adapté de (Buck *et al.*, 2008)

La capside peut être reconstituée *in vitro* en exprimant uniquement la protéine majoritaire L1 ou en co-exprimant les protéines L1 et L2 (Hagensee *et al.*, 1993). La protéine L2 est en grande partie enfouie dans la capside, à l'exception de son extrémité N-terminale (Buck *et al.*, 2008; Guan *et al.*, 2017).

1.1.3.2 Génome

Les papillomavirus ont un génome à ADN double-brin circulaire, d'environ 7-8 kb. On distingue trois zones distinctes dans ce génome : la région régulatrice URR (Upstream Regulatory Region), la région précoce (Early) et la région tardive (Late)(**Figure 47**).

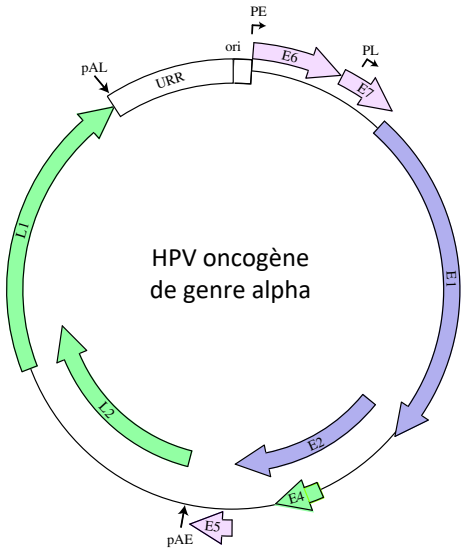


Figure 47: Organisation du génome de HPV oncogène de genre alpha. La région régulatrice URR est représentée en blanc, la région précoce en violet et la région tardive en vert. Les signaux de polyadénylation des régions précoce et tardive sont notés respectivement pAE et pAL. De même, les promoteurs qui contrôlent l'expression de la région précoce (PE) et de la région tardive (PL) sont indiqués. L'origine de réplication (ori) est incluse dans la région régulatrice. L'organisation du génome varie en fonction du genre phylogénétique de HPV(Doorbar, 2018). Adapté de (McBride, 2017).

Les trois régions sont séparées par des signaux de polyadénylation : l'un à la fin de la région précoce (pAE), le second à la fin de la région tardive (pAL). Le génome est très compact : plusieurs cadres de lecture ouverts se recoupent et le génome viral produit des transcrits polycistroniques qui subissent des épissages alternatifs (Johansson and Schwartz, 2013). Les différentes protéines virales encodées par les régions précoce et tardive sont indiquées sur le **Tableau 2**.

Protéine	Activité la mieux connue
E1	Hélicase se liant à l'origine de réplication du génome.
E2	Protéine de liaison à l'ADN, permet de charger l'hélicase E1, attache le génome viral aux chromosomes de la cellule hôte, régulateur transcriptionnel
E8^E2	Protéine de liaison à l'ADN, répresseur qui limite la réplication productive.
E1^E4 ou E4	Protéine tardive, réorganise le réseau de kératines dans les cellules différenciées afin de faciliter la transmission virale.
E5	Protéine encodée uniquement par les HPV du genre alpha, favorise la prolifération cellulaire en activant les récepteurs aux facteurs de croissance et permet au virus d'échapper au système immunitaire.
E6	Oncoprotéine qui perturbe les fonctions cellulaires par des interactions protéine-motif, soit en capturant les motifs LXXLL des protéines hôtes, soit en interagissant avec les protéines à domaine PDZ dans le cas des HPV à haut-risque muqueux. Peut mener certaines protéines à leur dégradation, comme p53 dans le cas de HPV16.
E7	Oncoprotéine qui dégrade les protéines de la famille Rb, modulateur épigénétique.
L1	Protéine de capside majoritaire, 360 copies par virion.
L2	Protéine de capside minoritaire, jusqu'à 72 copies par virion. L2 se lie à l'ADN et participe à transporter le pseudogénome jusqu'au noyau ainsi qu'à compacter le génome dans la capside virale.

Tableau 2: Présentation des protéines encodées par le génome de HPV. D'après (McBride, 2017)

Le génome comporte également deux promoteurs, dont l'un situé avant la séquence codant pour E6 et contrôlant l'expression de la quasi-totalité des gènes de la région précoce (PE) (Zheng and Baker, 2006). Le second promoteur est situé dans le cadre de lecture ouvert de E7 et contrôle l'expression de la région tardive (PL). Ce promoteur PL n'est activé que dans les kératinocytes différenciés tandis que le promoteur PE est activé dans les cellules épithéliales basales, ce qui correspond à l'expression virale au cours du cycle infectieux représentée sur la **Figure 48**.

1.1.3.3 Cycle infectieux

1.1.3.3.1 Entrée dans la cellule hôte

Les papillomavirus infectent les épithéliums pluristratifiés, muqueux ou cutanés selon leur tropisme. Le virus infecte les cellules de la couche basale en s'infiltrant par des micro-abrasions (Roberts *et al.*, 2007). Grâce au processus de cicatrisation et de renouvellement tissulaire, ces cellules basales se divisent avant de migrer vers la surface de l'épithélium et de se différencier. Le cycle infectieux du papillomavirus suit la progression de ces cellules : leur différenciation induite une forte réplication virale ainsi que l'expression des gènes viraux. Les virions sont finalement assemblés dans les couches supérieures avant d'être relargués dans l'environnement par des squames (Graham, 2017). Dans le cas des infections menant au cancer du col de l'utérus, les papillomavirus dits à haut-risque muqueux semblent cibler une population cellulaire très particulière : les cellules malpighiennes de la jonction entre l'endocol et l'ectocol (Herfs *et al.*, 2012).

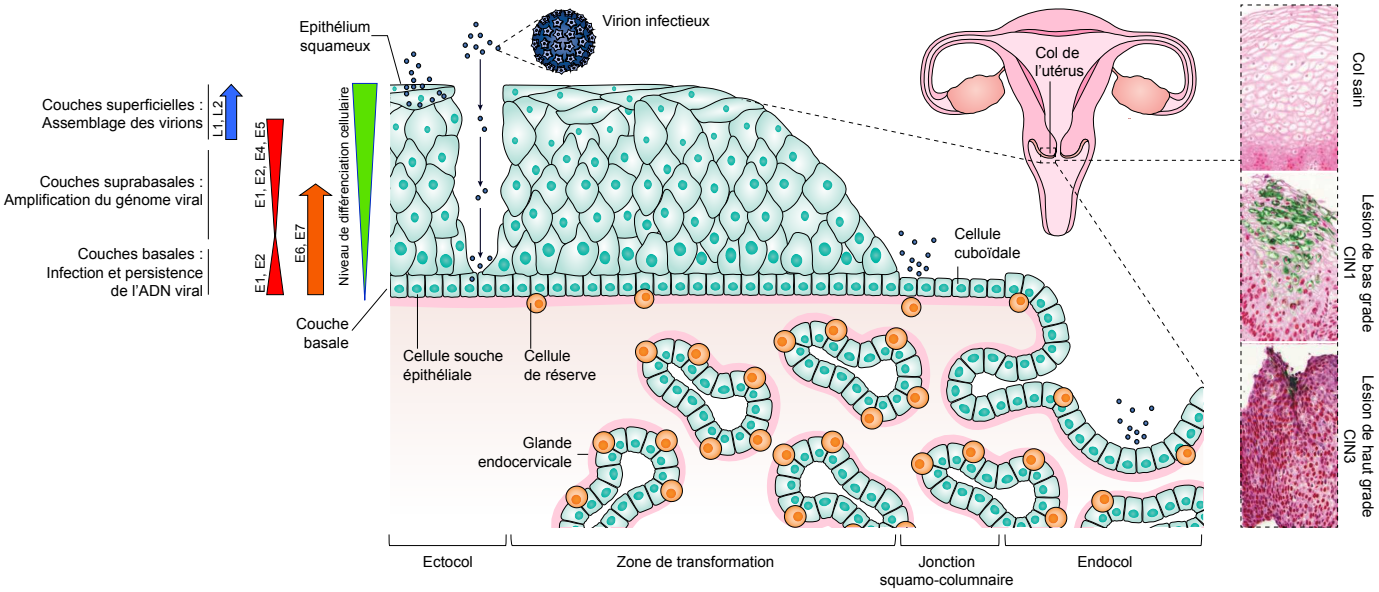


Figure 48: Cycle infectieux du HPV au niveau du col de l'utérus. Figure modifiée de (Graham, 2017) et (Schiffman *et al.*, 2016). Le HPV infecte les épithéliums au niveau de micro-abrasions ou, dans le cas du col de l'utérus, en ciblant les cellules cuboïdales et les cellules de réserve (jonction squamo-cellulaire) ou les cellules de l'endocol. En infectant ces cellules peu différenciées et en active division cellulaire afin d'assurer le renouvellement de l'épithélium, le virus assure une persistance de l'ADN viral en maintenant un faible nombre de copies de son génome dans un ensemble de cellules proches de la membrane basale. Grâce à l'action des oncoprotéines E6 et E7, l'amplification virale est activée dans les cellules qui ont migré au niveau de la couche suprabasale. Enfin, l'expression des protéines de capside L1 et L2 dans les couches superficielles de l'épithélium permet l'assemblage des virions qui sont ensuite libérés dans l'environnement par desquamation. Les images d'immunohistochimie sur la droite de l'image présentent un col sain, ainsi que des néoplasies cervicales intraépithéliales (CIN) de grade 1 et 3. La protéine E4 de HPV est colorée en vert et la coloration rouge indique un marqueur du cycle cellulaire (complexe de maintenance des minichromosomes).

A la surface de la capside virale, la protéine L1 est exposée au solvant et interagit avec des protéoglycanes à héparane sulfate (HSPG) présents au niveau de la membrane basale. Les résidus lysine 278 et 361 de la protéine L1 sont les points d'attache aux HSPG (Raff *et al.*, 2013). La cyclophiline B est une isomérase liée aux HSPG, en particulier le syndecan-1. Elle induit un changement conformationnel de la protéine de capside virale L2, qui facilite le clivage de l'extrémité N-terminale de L2 par une endoprotéase cellulaire, la furine et/ou PC5/6 (Bienkowska-Haba *et al.*, 2009; Kines *et al.*, 2009; Richards *et al.*, 2006). A ce jour, aucun modèle décrivant les voies de signalisation menant à l'internalisation du HPV n'a été établi à l'unanimité mais plusieurs récepteurs potentiellement impliqués ont été mis en évidence : intégrine α_6, récepteurs de facteurs de croissance épidermique et de kératinocytes, tétraspanines, sous-unité S100A10 de l'hétérotétramère Annexine A2 (Raff *et al.*, 2013; Woodham *et al.*, 2012). La voie d'internalisation semble impliquer la caveoline 1 et la dynamine 2, comme cela a été démontré pour HPV31 (Smith *et al.*, 2007). La particule virale est ensuite transportée par la voie endosomale. Pendant ce transport, la capside virale se dissocie sous l'effet du pH acide et des protéases et chaperonnes cellulaires. La protéine L2 ayant la capacité de se lier à une molécule d'ADN sans reconnaître une séquence spécifique (Zhou *et al.*, 1994), elle forme un complexe avec le pseudogénome qui est ensuite transporté jusqu'au noyau (Aksoy *et al.*, 2017). Le complexe L2-pseudogénome entre dans le noyau pendant que la cellule hôte est en mitose, alors que l'enveloppe nucléaire est rompue (Aydin *et al.*, 2014).

1.1.3.3.2 Maintien du génome

Le génome viral est maintenu dans le noyau des cellules basales épithéliales sous forme d'épisome, défini comme un ADN circulaire extrachromosomique capable de se répliquer de manière autonome et de s'intégrer à l'ADN chromosomique (Mellin *et al.*, 2002).

Le papillomavirus synchronise la réplication de son génome à celle de la cellule infectée, qui dans le cas d'une cellule basale est peu différenciée et se divise activement. Les protéines virales E1 et E2 initient la réplication virale : la protéine précoce E2 recrute et positionne l'hélicase E1 sur l'origine de réplication. La protéine E1 désenroule l'ADN et recrute la machinerie de réplication cellulaire (topoisomérase I, ADN polymérase α primase, ADN polymérase δ...). La protéine E2 arrime le génome

viral aux chromosomes des deux cellules filles, ce qui permet de répartir les copies du génome viral dans les cellules filles, en donnant ainsi lieu à une infection persistance dans un nombre croissant de cellules (McBride, 2008). Dans un premier temps, l'amplification est limitée à 10 à 50 copies par cellule afin d'éviter d'activer la réponse immunitaire dans les cellules basales. La protéine E2 joue un rôle de répresseur sur le promoteur de la région précoce PE en bloquant l'accès aux facteurs transcriptionnels et en altérant la conformation de la chromatine (Graham, 2017).

1.1.3.3.3 Amplification du génome

Le papillomavirus active la multiplication de la cellule hôte grâce à l'action des oncoprotéines E6 et E7. Leur expression est activée dans les couches basale et suprabasale de l'épithélium, puis elle diminue dans les couches superficielles (**Figure 48**). Ce mécanisme permet de compenser la diminution progressive de la multiplication cellulaire au fur et à mesure que celles-ci se différencient.

D'une part, l'oncoprotéine E6 capture la protéine hôte E6AP (E6-Associated Protein) en liant son motif LXXLL (Huibregtse *et al.*, 1991, 1993). La protéine E6AP est une E3 ubiquitine ligase, elle intervient à la dernière étape de la réaction d'ubiquitination et est responsable de la reconnaissance du substrat protéique sur lequel une molécule d'ubiquitine sera fixée. Son interaction avec la protéine virale E6 modifie ses spécificités d'interaction. Ainsi, l'hétérodimère E6/E6AP lie le suppresseur de tumeurs p53 qui est alors ubiquitiné, ce qui provoque sa dégradation par le protéasome (Scheffner *et al.*, 1990, 1993).

Chez les HPV à haut risque de genre alpha, la protéine E7 se lie au suppresseur de tumeurs Rb (protéine du Rétinoblastome) et perturbe son interaction avec les facteurs de transcription E2F (Liu *et al.*, 2006; Slebos *et al.*, 1994). Dans les kératinocytes en cours de différenciation, la division cellulaire est réprimée par l'action inhibitrice de Rb hypophosphorylé qui s'associe avec E2F. L'interaction de la protéine E7 lève l'inhibition exercée par Rb et permet à la cellule de passer de la phase G1 à la phase S du cycle mitotique (Giarrè *et al.*, 2001).

En maintenant les cellules infectées en phase active de division cellulaire, le HPV amplifie de façon drastique son génome viral, atteignant plusieurs milliers de copies par cellule (Doorbar *et al.*, 2015). Afin de mener la réplication virale à terme, les

protéines E1 et E2 sont fortement exprimées, comme indiqué sur la **Figure 48** (Graham, 2017).

1.1.3.3.4 Assemblage et libération des particules virales

Lorsque les cellules infectées migrent vers les couches superficielles de l'épithélium, la phase de réplication virale productive est initiée. Par des changements de sites d'épissage, les transcrits produits à partir du promoteur PL au signal de polyadénylation pAL donnent lieu à l'expression des protéines de capside L1 et L2 (**Figure 47**). La protéine L2 est recrutée au niveau des régions de réplication par la protéine E2, ce qui permet l'encapsidation du génome. La maturation virale intervient dans les kératinocytes sénescents les plus proches de la surface, dans lesquels l'environnement est fortement oxydant. Dans ce contexte, de nombreux ponts disulfures se forment entre les protéines L1 qui s'assemblent afin de former des capsides extrêmement stables. La protéine E4 facilite le relargage des virions infectieux en perturbant le réseau de kératines (Doorbar *et al.*, 2012).

1.1.3.3.5 Intégration dans le génome

L'intégration de l'ADN viral dans le génome de l'hôte n'est pas un événement faisant normalement partie du cycle de vie du HPV. Au contraire, un tel événement marque la fin du cycle infectieux, étant donné que le génome viral linéarisé et intégré dans l'ADN chromosomique n'est plus suffisamment compact pour être encapsidé et transmis à un nouvel hôte. Il semble que l'intégration dans l'ADN chromosomique soit un événement rare et qui se produit de façon aléatoire, au niveau de sites dits fragiles présentant une certaine homologie avec des séquences présentes dans le génome viral (Schmitz *et al.*, 2012; Wentzensen *et al.*, 2004).

Dans la plupart des cancers invasifs liés au HPV, on note que l'ADN viral est intégré au génome humain. Cet événement provoque généralement une expression accrue des oncoprotéines E6 et E7, qui stimulent alors la prolifération cellulaire sur le long terme et mènent à la cancérogenèse. Le mécanisme le plus couramment proposé pour expliquer cette dérégulation est l'altération de la fonction inhibitrice de E2 sur les régions codantes de E6 et E7. Un défaut dans la séquence du gène E2 ou l'altération des sites de liaison reconnus par E2 peuvent permettre de lever l'inhibition transcriptionnelle que cette protéine exerce (**Figure 49**).

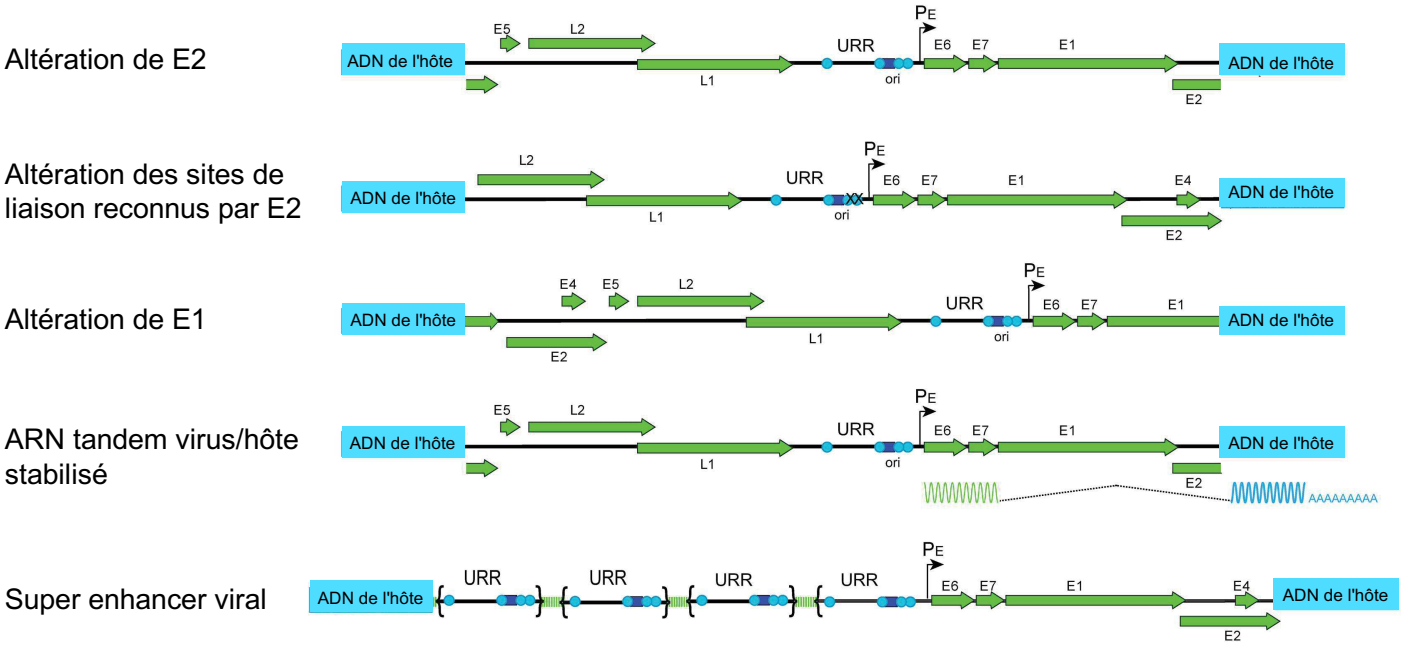


Figure 49: Modèles d'intégration dans le génome hôte menant à la cancérogenèse. La région régulatrice (URR) contient les sites de liaison de E2 (indiqués par des cercles bleu clair) et l'origine de réplication du génome viral (signalée par un carré bleu foncé). Le promoteur qui régule la région précoce est noté PE. Adapté de (McBride and Warburton, 2017).

Ce n'est cependant pas le seul mécanisme qui peut entraîner la cancérogenèse par une surexpression des oncoprotéines. L'altération du gène E1 peut entraîner des dommages de l'ADN et augmenter l'instabilité génomique au niveau du locus d'intégration. La formation d'un transcrit qui fusionne les séquences E6 et E7 avec des séquences provenant de l'hôte a souvent pour effet de stabiliser l'ARN et de favoriser sa traduction. La duplication de la région régulatrice URR en amont des gènes E6 et E7 peut former un enhancer qui amplifie l'expression des deux oncogènes (McBride and Warburton, 2017).

On note par ailleurs que tous les cancers provoqués par les HPV et notamment par les oncogènes E6 et E7 ne sont pas systématiquement associés à des événements d'intégration de l'ADN viral dans le génome de l'hôte. Une étude récente menée sur

des modèles murins a démontré la possibilité d'un mécanisme de "hit-and-run" des oncoprotéines E6 et E7 de HPV38 du genre beta: après une expression transitoire des oncogènes au niveau de la couche basale de l'épithélium cutané couplée à une exposition aux ultra-violets, on observe la mise en place d'un phénotype cancéreux qui n'est pas atténué par la délétion des gènes E6 et E7 (Viarisio *et al.*, 2018).

1.2 Pathologies HPV et traitement

1.2.1 Épidémiologie

Les HPV à haut risque muqueux sont responsables de respectivement 8.6 % et 0.8 % des cancers diagnostiqués chez les femmes et chez les hommes à l'échelle mondiale. Les zones les plus touchées dans le monde sont les pays en développement où on dénombre 85 % des décès liés aux cancers HPV (**Figure 50**). Cette différence s'explique en partie par l'accès limité aux mesures de prévention telles que les vaccins prophylactiques et les frottis de dépistage. Dans les pays développés, ces mesures permettent de prévenir jusqu'à 80 % des cancers du col. De même, l'accès aux traitements qui s'appliquent lorsque le cancer est à un stade très avancé est limité dans les pays en développement où le taux de décès est très élevé (Organisation Mondiale de la Santé).

Incidence des cancers attribués aux HPV à l'échelle mondiale
Taux normalisés selon l'âge pour 100 000 individus par pays

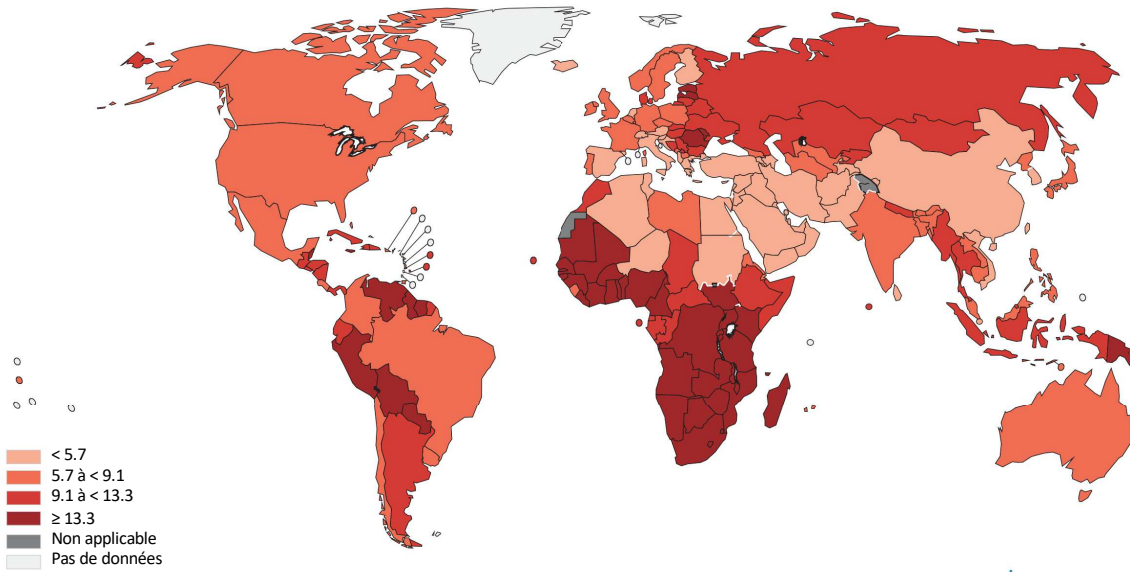


Figure 50 : Incidence des cancers attribués aux HPV à l'échelle mondiale. Données de (de Martel *et al.*, 2020), mise en forme de la carte sur le site de l'observatoire mondial du cancer géré par le centre international de recherche sur le cancer (gco.iarc.fr, consulté le 9 avril 2020).

Les HPV à haut risque muqueux sont les principaux agents étiologiques du cancer du col de l'utérus, qui est le quatrième cancer le plus fréquent chez les femmes. Ils sont également à l'origine de 88 % des cancers de l'anus, de 25 % des cancers de la vulve, 78 % des cancers du vagin, de 50 % des cancers du pénis ainsi que 35% des cancers de la sphère ORL (oropharynx, cavité orale, larynx) (de Martel *et al.*, 2017). Pour ces différents cancers, on note que certains types de HPV ont une plus forte prévalence et ce même au sein des HPV classifiés comme à haut risque muqueux. HPV16 et HPV18 sont responsables de respectivement 61 % et 11 % des cas de cancer de l'utérus (**Figure 51**). HPV16 est le type de HPV ayant la plus forte prévalence pour les cancers anogénitaux (vulve, vagin, anus, pénis). Il est également à l'origine de 71 % des cancers de la sphère ORL, qui englobe la cavité buccale, le pharynx et le larynx.

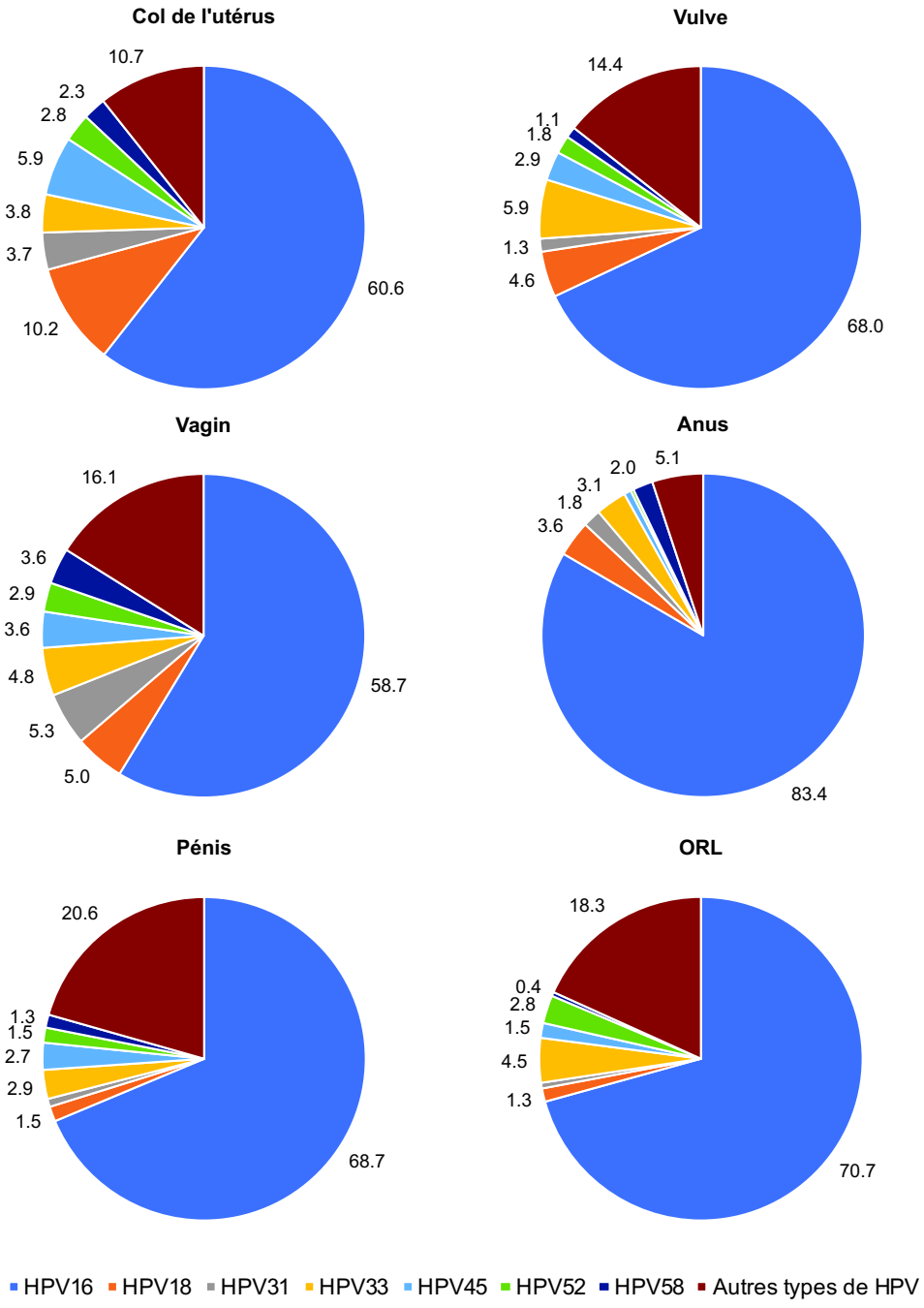


Figure 51 : Prévalence de différents types de HPV à haut risque muqueux dans les cancers du col de l'utérus, du vagin, de la vulve, du pénis, de l'anus et de la sphère ORL, exprimée en pourcentages. Données à l'échelle mondiale issues de (Alemany *et al.*, 2016; Castellsagué *et al.*, 2016; Serrano *et al.*, 2015).

D'autre part, les HPV à tropisme cutané sont à l'origine des verrues, qui peuvent être situées sur le visage, les mains, pieds, coudes, genoux et parties génitales. HPV2 est à l'origine de la plupart des verrues vulgaires : situées sur les mains, elles disparaissent spontanément dans 65 % des cas. Les verrues plantaires sont causées soit par HPV1 (myrmécie) soit par HPV2 (verrue mosaïque). Les verrues des bouchers

sont un type de verrue vulgaire lié à HPV7 qui touche particulièrement les mains des personnes manipulant de la viande ou du poisson. Pour ces dernières, le traitement s'avère souvent difficile en raison du taux de récidive qui s'élève à 50 % (Aubin and Guerrini, 2007). Concernant les verrues génitales ou condylomes, 75 à 90 % sont causés par les HPV à bas risque muqueux HPV6 et HPV11. Avec 1 % de la population américaine sexuellement active touchée, ces verrues représentent un véritable problème de santé publique. En l'absence de traitement, l'infection HPV peut être liée à l'apparition d'un carcinome verruqueux, un cancer de bas grade bien différencié et qui métastase rarement. La tumeur de Buschke and Löwenstein ou condylome acuminé géant est un type de carcinome verruqueux localisé dans la région anogénitale qui est souvent associé à une infection par HPV6 et HPV11 (Yanofsky *et al.*, 2012). De plus, les HPV de types 6 et 11 sont associés à la papillomatose respiratoire récurrente : cette maladie respiratoire rare se caractérise par la présence de papillomes (petites tumeurs bénignes) au niveau des muqueuses des voies aéro-digestives supérieures, en particulier le larynx. Généralement bénigne, elle peut chez l'enfant atteindre la trachée et les bronches, ce qui peut s'avérer fatal en cas d'occlusion des voies aériennes. Dans sa forme infantile, l'infection par le HPV à l'origine de la papillomatose a lieu par transmission placentaire de la mère au fœtus : le risque de développer la maladie quelques mois ou années après la naissance est 231 fois plus élevé si la mère a eu des verrues anogénitales pendant la grossesse (Fusconi *et al.*, 2014).

L'épidermodysplasie verruciforme (EV), aussi appelée syndrome de Lutz-Lewandowsky, est une maladie de la peau rare qui se manifeste par un défaut de l'immunité primaire et une sensibilité accrue aux infections par des HPV de genre beta. Il s'agit d'une maladie génétique à transmission autosomique récessive. Dans 75 % des cas, elle est causée par des mutations des gènes EVER1 et EVER2 situés sur le chromosome 17 : ces gènes codent pour des protéines membranaires ayant un rôle dans la régulation du taux de zinc intracellulaire et dont la perte de fonction augmente drastiquement la sensibilité du revêtement cutané aux HPV (Cardoso and Calonje, 2011). Les types de HPV mis en cause sont principalement des HPV de l'espèce 1 du genre beta : les HPV5 et HPV8 sont le plus souvent impliqués, mais on compte également les HPV14, 20, 47... De même, certains HPV de l'espèce 2 du genre beta sont impliqués, par exemple HPV38, ainsi que des HPV de l'espèce 3 du genre beta,

comme HPV49. Enfin, certains HPV du genre alpha peuvent également être impliqués, comme HPV3 et HPV10. La maladie se manifeste généralement avant l'âge de dix ans par l'apparition de verrues plates et macules squameuses qui persistent durant toute la vie du patient. L'aspect de ces lésions squameuses en forme d'écailles rappelle l'écorce d'un arbre, à l'origine d'un surnom donné à l'EV : "la maladie de l'homme-arbre".

Figure 52 : Abul Bajandar, patient atteint d'EV et surnommé "l'homme-arbre". Originaire du Bangladesh et alors âgé de 26 ans, Abul Bajandar a subi en 2015 un total de 16 opérations chirurgicales afin de retirer environ 5 kg d'excroissances cutanées de ses mains. Malheureusement, la maladie a récidivé dans les années qui ont suivi, en dépit des multiples opérations pratiquées. Source de l'image maxisciences.com, site consulté le 13 avril 2020.

Les personnes atteintes d'EV ont un risque accru de développer un carcinome épidermoïde, en particulier au niveau des zones de peau exposées au soleil. Chaque cellule tumorale renferme de multiples copies du génome viral sous forme d'épisomes. Chez 90 % des patients ayant un carcinome épidermoïde et atteints d'EV, les génomes des HPV5 et 8 ont été identifiés (Accardi and Gheit, 2014). Ces deux types de HPV ont été classifiés comme agents possiblement cancérogènes chez les patients atteints d'EV (**Tableau 1**). On note que les patients atteints d'EV ne sont sensibles qu'aux infections par des HPV, ils ne sont pas particulièrement vulnérables à d'autres agents pathogènes. L'immunité humorale est préservée : des anticorps dirigés contre les protéines E6 et E7 sont détectés chez 70 % des patients, ce qui indique également que l'expression de ces oncoprotéines est requise pour l'évolution cancéreuse de la maladie (Orth, 2010).

1.2.2 Cancer du col de l'utérus : infection, persistance et progression cancéreuse

Les HPV infectent les muqueuses et la peau par contact direct. L'infection par un HPV est relativement fréquente : on estime que plus de 70 % des hommes et femmes sexuellement actives seront en contact avec un HPV au moins une fois dans leur vie. Dans la plupart des cas de HPV à tropisme muqueux, l'infection a lieu au cours des premiers rapports sexuels et reste asymptomatique. Par la suite, la charge virale diminue jusqu'à ce que le virus devienne indétectable 5 ans après l'infection. Dans 90 % des cas, cette diminution est un signe de clairance virale, c'est-à-dire d'élimination totale des particules virales dans le tissu. Pour les 10 % restants, le virus est toujours présent à un niveau faible en phase latente (papillomavirus.fr, consulté le 13 avril 2020). Les probabilités de persistance du virus et de progression précancéreuse dépendent en partie du type de HPV : le risque de cancer associé au HPV16 est bien plus élevé que dans le cas d'autres HPV considérés comme étant à haut-risque (**Figure 51**). La réponse immunitaire de l'hôte est également un facteur déterminant: les patients infectés par le virus de l'immunodéficience humain (VIH) ont une plus forte incidence des cancers provoqués par les HPV (de Martel *et al.*, 2015). Enfin, certains facteurs comportementaux comme le tabac, la multiparité et la prise à long terme de contraceptifs hormonaux peuvent avoir une influence modérée sur le risque de cancer (Schiffman *et al.*, 2016).

La progression depuis un épithélium cervical sain à un cancer à cellules squameuses passe par des stades intermédiaires de lésions précancéreuses appelées néoplasies cervicales intraépithéliales (CIN). Ces lésions sont caractérisées par un nombre croissant de cellules présentant des anomalies morphologiques comme des défauts de stratification de l'épithélium, des anomalies nucléaires ou des défauts de différenciation. Les différents grades de CIN peuvent être définis par la proportion de cellules dysplasiques dans l'épithélium cervical. Dans le cas d'une lésion CIN1 de bas grade, les cellules dysplasiques sont présentes dans le tiers inférieur de l'épithélium, tandis qu'elles occupent deux tiers et jusqu'à l'épithélium entier pour les lésions de haut grade CIN2 et CIN3 (**Figure 53**).

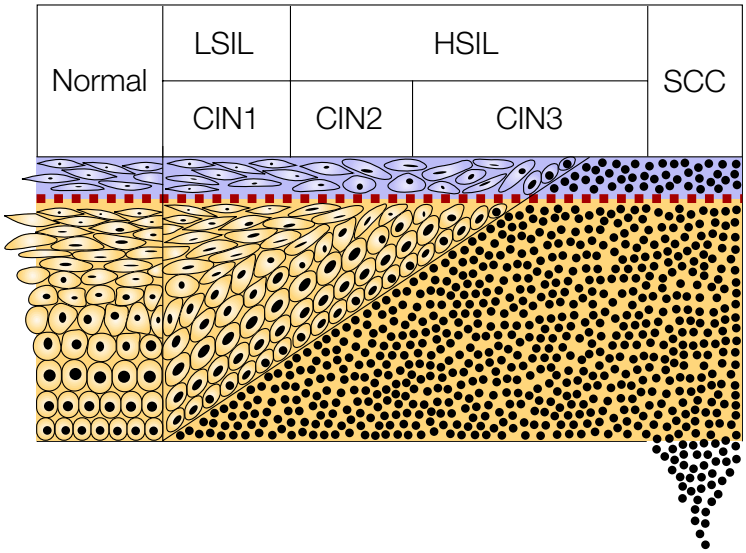


Figure 53: Lésions néoplasiques intraépithéliales du col de l'utérus : classification par système de tiers. La progression depuis l'épithélium du col de l'utérus normal vers le cancer à cellules squameuses (SCC) est représentée par le nombre croissant de cellules dysplasiques. Pour les lésions squameuses intraépithéliales de bas grade (LSIL) ou néoplasie intraépithéliale de grade 1 (CIN1), les cellules dysplasiques sont situées dans le tiers inférieur de l'épithélium. Les lésions de haut grade (HSIL) sont caractérisées par des cellules dysplasiques occupant deux tiers (CIN2) ou plus (CIN3). Adapté de (Baldwin *et al.*, 2003).

Les lésions de bas grade CIN1 sont associées à une régression spontanée qui ne requiert aucun traitement particulier. De 70 à 80 % pour l'ensemble des femmes, le taux de régression des CIN1 augmente jusqu'à 90 % pour les patientes de moins de 25 ans. En revanche, 0.2 à 4 % des CIN3 progressent vers un cancer invasif dans un intervalle de 12 mois. Il est important de noter qu'une période de 25 à 30 ans en moyenne sépare l'infection par un HPV du développement du cancer invasif (Martin and O'Leary, 2011).

1.2.3 Mesures préventives et curatives

Différentes mesures permettent de limiter le risque de cancer du col de l'utérus chez la femme. En premier lieu, des mesures préventives permettent de limiter l'infection par les HPV afin de réduire le risque de développer un cancer. On compte notamment le port du préservatif, qui semble réduire le risque d'infection HPV au niveau du col de l'utérus et des parties génitales chez la femme (Winer *et al.*, 2006) et chez l'homme (Pierce Campbell *et al.*, 2013). D'autre part, des vaccins prophylactiques ont été développés afin de prévenir l'infection de différents types de HPV. Le vaccin bivalent Cervarix a été commercialisé en 2008 afin de prévenir les infections par HPV16 et 18. Le groupe Merck a développé en 2007 un vaccin tétravalent nommé Gardasil,

contenant des particules virales (VLP) L1 de HPV6, 11, 16, 18. A ces 4 types de HPV s'ajoutent les HPV31, 33, 45, 52 et 58 inclus dans le Gardasil 9 nonavalent en circulation depuis 2018. Ce dernier est destiné à remplacer ses prédécesseurs en garantissant une meilleure couverture contre les principaux HPV à haut risque muqueux (HPV16, 18, 31, 33, 45, 52, 58) et les deux principaux HPV à bas risque responsables de verrues génitales (HPV6 et 11). Étant donné que l'infection par des HPV intervient généralement au début de la vie sexuelle, la vaccination est recommandée pendant la préadolescence (11 à 14 ans) et jusqu'à 19 ans pour les jeunes filles n'ayant pas encore eu de relations sexuelles. La vaccination contre les HPV a démarré depuis environ 10 ans et la prévalence des HPV à haut risque muqueux a significativement baissé dans les pays où le taux de vaccination est le plus élevé, notamment en Australie (Machalek *et al.*, 2018). Malgré ces résultats positifs, le taux de vaccination reste bas et les campagnes de vaccination sont peu suivies aux États-Unis (Walling *et al.*, 2016) et en France, où une partie de la population reste méfiante (Lefèvre *et al.*, 2018).

En complément de ces mesures préventives, des dépistages réguliers sont effectués afin de pouvoir détecter et traiter toute lésion précancéreuse avant son évolution en cancer. Le frottis cervico-utérin ou test de Papanicolaou est un prélèvement cytologique permettant de vérifier la présence de cellules dysplasiques au niveau de l'épithélium cervical. Le résultat du frottis est classifié d'après le système Bethesda. Le test HPV est un examen qui complète le frottis et fournit des résultats plus fiables : il permet de détecter la présence de HPV avec une sensibilité de plus de 95 %, et d'identifier les types de HPV avec une spécificité de plus de 90 %. A terme, le test HPV pourraient remplacer le frottis comme approche de dépistage du cancer du col de l'utérus (Hill *et al.*, 2017; Ronco *et al.*, 2010). Si ces premiers examens indiquent la présence de dysplasies et de HPV à haut risque muqueux, on réalise une colposcopie afin d'observer au microscope le col de l'utérus, le vagin et la vulve pour y détecter d'éventuelles lésions précancéreuses. En cas d'anomalie, les traitements mis en œuvre sont dans un premier temps la cryothérapie ou le laser pour les lésions précancéreuses. Selon la gravité de la lésion, la conisation peut être envisagée : il s'agit d'une opération chirurgicale qui consiste à enlever tout ou partie du col utérin. Ces traitements visant à exciser les CIN sont efficaces à 90 - 95 %. En cas de cancer du col de l'utérus, les traitements appliqués sont la chimiothérapie, la radiothérapie et

la chirurgie. Aucun traitement spécifique du HPV n'est disponible à ce jour, bien que des essais cliniques soient en cours pour des molécules inhibant la liaison de E1 et E2 à l'ADN (Bosch *et al.*, 2013).

2 L'oncoprotéine E6

2.1 Structure et biochimie

L'oncoprotéine E6 est en partie responsable de la cancérogenèse induite par les infections HPV : cette petite protéine d'environ 151 résidus est capable d'interagir avec un grand nombre de protéines hôtes, perturbant ainsi les réseaux d'interactions préexistants dans la cellule et compromettant les fonctions de régulation du cycle cellulaire. De fait, la caractérisation structurale des protéines E6 est indispensable à l'étude de leurs mécanismes moléculaires d'interaction et permet le développement d'inhibiteurs bloquant spécifiquement les interactions des E6. Mais alors que la protéine E6 a été identifiée comme oncogène dès la fin des années 1980 (Androphy *et al.*, 1987), il a fallu attendre 2006 pour que la structure du domaine C-terminal de liaison au zinc soit résolue par Résonance Magnétique Nucléaire (Nominé *et al.*, 2006) et 2013 pour la structure cristallographique de la protéine E6 entière (Zanier *et al.*, 2013). Ce délai est en grande partie dû à la difficulté d'exprimer la protéine E6 recombinante et de l'isoler sous sa forme active. Produite en bactérie, la protéine E6 se retrouve dans les corps d'inclusion lorsqu'elle n'est pas fusionnée à une protéine qui augmente sa solubilité (Nominé *et al.*, 2001). La Maltose-Binding Protein (MBP) améliore grandement la solubilité de l'oncoprotéine E6. Cependant, la fusion MBP-E6 est en partie présente sous forme d'oligomères micellaires comprenant le polypeptide E6 mal replié enfoui et les unités de MBP solubles et exposées au solvant (**Figure 54**).

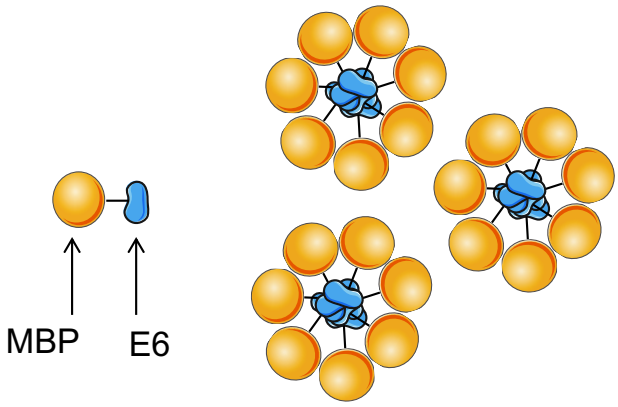


Figure 54 : Oligomères micellaires de l'oncoprotéine E6 fusionnée à la MBP. La protéine MBP est fusionnée au N-terminus de l'oncoprotéine E6. Une partie des fusions MBP-E6 tend à former des oligomères où les polypeptides E6 mal repliés sont regroupés au centre de la micelle tandis que les unités de MBP demeurent solubles et sont exposées au solvant.

Ces difficultés d'isoler l'oncoprotéine E6 sous forme soluble et repliée ont conduit certains spécialistes de HPV à la considérer comme une protéine nativement non repliée et dont le désordre intrinsèque participe à l'activité oncogène des HPV à haut risque (Uversky *et al.*, 2006).

Pour comprendre pourquoi la protéine E6 présente une si forte tendance à l'auto-association, il nous faut revenir à son organisation structurale et à sa composition en acides aminés (**Figure 55**).

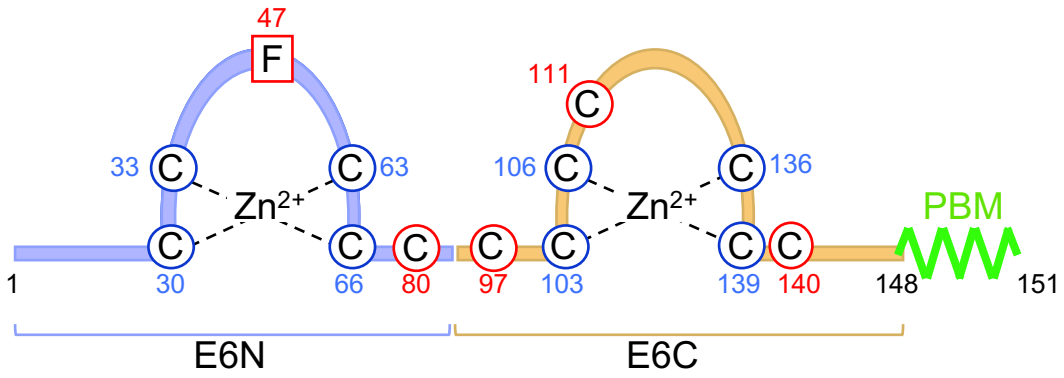


Figure 55 : Organisation de l'oncoprotéine E6 de HPV16. Les deux domaines de liaison au zinc sont notés E6N et E6C et sont représentés respectivement en violet et en orange. Chaque ion zinc est coordonné par 4 cystéines, représentées dans des cercles bleus. Les cystéines non conservées qui ont été mutées en sérine afin de maximiser la solubilité de la protéine sont annotées dans des cercles rouges. Le résidu phénylalanine en position 47 a été muté en arginine afin d'éviter la dimérisation de la protéine E6. Le motif de liaison aux domaines PDZ (PBM) est présent à l'extrémité C-terminale. Figure adaptée de (Poirson, 2016).

Les E6 de HPV à haut-risque muqueux comportent un motif de liaison aux domaines PDZ (PBM). Les protéines E6 de HPV infectant les mammifères se composent de deux domaines de liaison au zinc, l'un situé dans la partie N-terminale (E6N) et le second en partie C-terminale (E6C) (Suarez and Travé, 2018). Chaque ion zinc requiert quatre cystéines pour sa coordination : ces résidus sont fortement conservés car ils sont indispensables au bon repliement de la protéine. A ces huit cystéines s'ajoutent des cystéines non conservées dont certaines sont exposées à la surface de la protéine : en conditions oxydantes, ces résidus sont susceptibles de former des ponts disulfures intramoléculaires qui participent fortement au phénomène d'auto-association de E6. L'oncoprotéine E6 de HPV16 comporte au total 14 résidus cystéine. Parmi les six résidus non impliqués dans la coordination des ions zinc, quatre sont exposés au solvant : ces derniers ont été mutés en sérine afin de limiter la formation de ponts disulfure artefactuels entre plusieurs molécules. D'autre part, le résidu phénylalanine

en position 47 favorise la dimérisation du domaine E6N (Zanier *et al.*, 2012). Afin d'éviter les phénomènes d'avidité lors des tests d'interaction, ce résidu a été muté en arginine pour empêcher la dimérisation de la protéine.

Le mutant de HPV16 E6 combinant les mutations F47R et les quatre cystéines mutées en sérine est celui dont la structure a pu être résolue par cristallographie aux rayons X à une résolution de 2.6 Å (Zanier *et al.*, 2013). La structure présente le mutant de HPV16 E6 en complexe avec sa cible prototypique, le motif LXXLL de l'ubiquitine ligase E6AP de séquence $E^{1}L^{2}T^{3}L^{4}Q^{5}E^{6}L^{7}L^{8}G^{9}E^{10}E^{11}R^{12}$. La visualisation de ce complexe à une résolution atomique permet de voir que le motif se replie en hélice alpha et se loge dans une poche hydrophobe formée par les deux domaines de liaison au zinc. La structure révèle également que les résidus glutamate en amont et en aval du consensus LXXLL semblent participer à l'interaction par des liaisons hydrogène, ce qui suggère que le consensus reconnu par E6 de HPV16 ne se limite pas aux trois résidus leucine mais inclut un environnement acide de part et d'autre du motif (**Figure 56**).

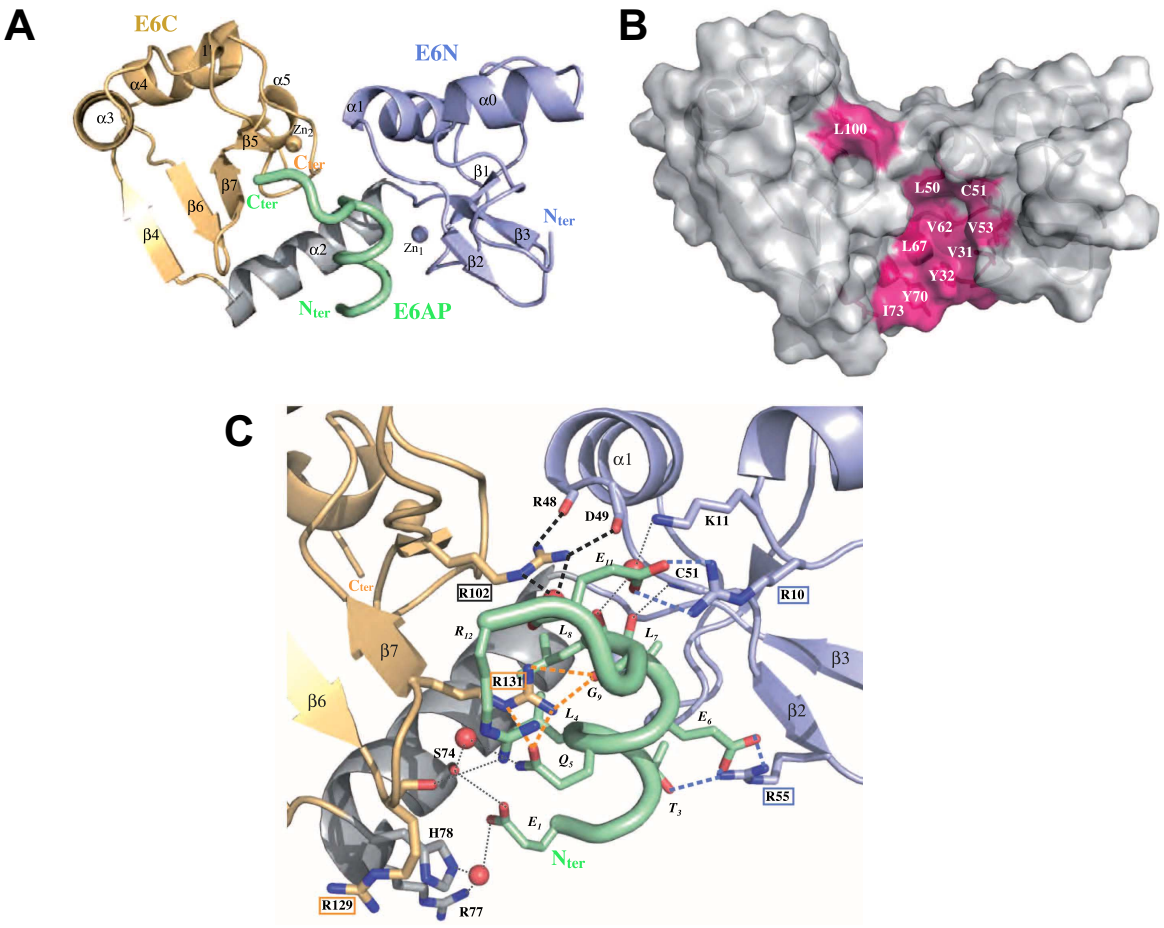


Figure 56 : Structure de E6 de HPV16 en complexe avec le motif LXXLL de l'ubiquitine ligase E6AP, de séquence peptidique $E^1L^2T^3L^4Q^5E^6L^7L^8G^9E^{10}E^{11}R^{12}$. **A**. Structure du complexe E6 HPV16 / E6AP. Les domaines E6N et E6C et les deux ions zinc correspondant sont représentés respectivement en violet et orange. Le linker, replié sous forme d'hélice alpha, est représenté en gris. Le motif LXXLL de E6AP (représenté en vert) adopte une conformation d'hélice alpha lorsqu'il est en complexe avec la protéine E6. **B**. Le site de liaison du motif LXXLL est une poche hydrophobe formée entre les deux domaines de liaison au zinc, dont les résidus sont indiqués en rose. **C.** Liaisons hydrogène entre le motif LXXLL de E6AP et la protéine E6. Les interactions sont représentées par des tirets et les molécules d'eau qui participent à ces interactions sont représentées par des sphères rouges. Adapté de (Zanier *et al.*, 2013)

3 Approches pour l'étude des interactions protéine-protéine

3.1 Méthodes quantitatives à faible débit

3.1.1 Principes de base pour l'étude quantitative des interactions protéine-protéine

Les interactions protéine-protéine peuvent être décrites par différentes constantes, qui reflètent la force d'interaction ou encore les cinétiques d'association et de dissociation d'un complexe. Ces constantes et les stratégies permettant de les déterminer sont communes à la majorité des méthodes biophysiques de quantification d'affinité.

La constante de dissociation, notée K_D, est la valeur la plus souvent utilisée pour quantifier l'affinité d'une interaction. Dans le cas de l'interaction de deux molécules A et B formant un complexe AB, la réaction est décrite d'après l'**Équation 1**.

$$A + B \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} AB$$

Équation 1

Dans cet exemple, le K_D est la constante de réaction associée à la dissociation du complexe AB à l'équilibre. Elle dépend des concentrations des composés A et B libres en solution ainsi que de la concentration de complexe AB à l'équilibre (**Équation 2**).

$$K_D = \frac{[A] \times [B]}{[AB]} = \frac{k_{off}}{k_{on}}$$

Équation 2

En pratique, la stratégie la plus courante permettant de déterminer le K_D d'une interaction consiste à incuber les deux partenaires à différentes concentrations : la concentration du partenaire A reste fixe tandis que le partenaire B est ajouté à des concentrations croissantes. Le mélange A + B est incubé jusqu'à atteindre l'équilibre puis le complexe AB est quantifié par différentes approches biophysiques. On obtient un signal proportionnel à la concentration de complexe formé en fonction de la concentration de l'un des deux partenaires. La **Figure 57** présente un exemple de données de résonance plasmonique de surface (SPR) permettant la détermination du

K_D entre E6 HPV16 fusionnée à la MBP (notée MBP-16E6) et le motif LXXLL de E6AP (noté $E6AP_{LXXLL}$).

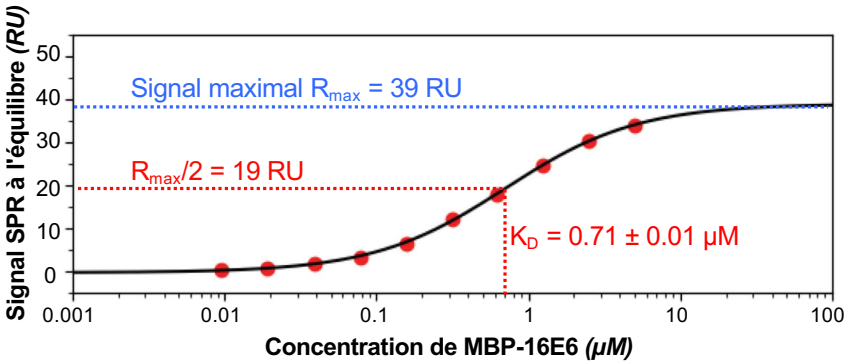


Figure 57 : Exemple de graphe pour la détermination de K_D. Les données ci-dessus sont issues d'un test réalisé par Résonance Plasmonique de Surface (SPR) pour l'interaction entre la protéine 16E6 F47R 4C/4S fusionnée à la MBP (notée MBP-16E6) et le peptide biotinylé du motif LXXLL de E6AP (séquence PESSELTLQELLGEER). Le signal SPR à l'équilibre est normalisé par le signal de peptide E6AP immobilisé. La protéine MBP-16E6 a été incubée à des concentrations comprises entre 9.6 nM et 5 µM. Les données expérimentales sont représentées par des points rouges et le fit quadratique par une ligne noire. Le K_D de l'interaction est égal à la concentration de MBP-16E6 pour laquelle le signal atteint la moitié du signal maximal.

En immobilisant une quantité constante de peptide $E6AP_{LXXLL}$, la protéine MBP-16E6 a été injectée successivement à des concentrations croissantes. Pour chaque injection, on extrait le signal à l'équilibre, qui est représentatif de la quantité de complexe MBP-16E6/$E6AP_{LXXLL}$. Sur un graphe représentant ce signal à l'équilibre en fonction de la concentration de MBP-16E6, on effectue une régression quadratique, qui permettra de connaître le K_D. Celui peut également être lu sur le graphique : il s'agit de la concentration de protéine pour laquelle le signal est égal à la moitié du signal maximal (**Figure 57**). Pour pouvoir effectuer cette interprétation de données, il est important d'adapter les gammes de concentrations des deux partenaires afin de bien encadrer le point d'inflexion où se situe le K_D et de voir un plateau aux concentrations les plus élevées. De plus, le temps nécessaire pour atteindre l'équilibre est propre à chaque interaction, il fait partie des paramètres à optimiser pour pouvoir quantifier le complexe à l'équilibre.

La constante de dissociation est également liée aux constantes cinétiques d'association (k_{on}) et de dissociation (k_{off}). Au-delà de l'affinité, les cinétiques de formation d'un complexe sont très informatives pour décrire l'aspect dynamique d'une interaction dans un système vivant qui est rarement à l'équilibre. En pharmacocinétiques, ces données sont primordiales pour déterminer le temps d'action

d'une molécule et donc la dose et la fréquence d'administration d'un médicament. La détermination de ces constantes cinétiques requiert une méthode permettant de mesurer en continu la formation de complexe depuis la mise en présence des deux partenaires jusqu'à atteindre l'équilibre puis la phase de dissociation. La SPR est la technique la plus couramment utilisée pour étudier les cinétiques d'interaction de biomolécules.

Enfin, les paramètres thermodynamiques décrivent l'énergie de la réaction de formation du complexe. La variation d'énergie libre est notée ΔG, elle est liée aux variations d'enthalpie (ΔH), d'entropie (ΔS) ainsi qu'à la température T (**Équation 3**).

$$\Delta G = \Delta H - T\Delta S$$

Équation 3

L'énergie libre traduit la stabilité d'un complexe : si la variation d'énergie libre est négative, le système est favorable à la formation du complexe AB qui est plus stable que les partenaires A et B non liés. A ce jour, la seule approche permettant de déterminer la contribution des paramètres thermodynamiques d'entropie et d'enthalpie d'une interaction en une seule mesure est la titration calorimétrique isotherme (ITC). Les autres approches telles que la résonance plasmonique de surface (SPR), la thermophorèse à micro-échelle (MST) et la résonnance magnétique nucléaire (RMN) permettent d'estimer les paramètres thermodynamiques en effectuant plusieurs mesures à différentes températures.

Enfin, l'énergie libre associée à l'association de deux partenaires est reliée au K_D, d'après l'**Équation 4**.

$$\Delta G = RT \times ln(K_D)$$

Équation 4

3.1.2 Résonance plasmonique de surface

La résonance plasmonique de surface (SPR) est une technique optique de détection des interactions entre deux partenaires. Elle s'applique aux protéines comme aux acides nucléiques, lipides, glucides, petites molécules, etc. Depuis la publication du phénomène d'excitation de plasmons de surface par réflection d'un faisceau de

lumière en 1968 (Kretschmann and Raether, 1968), la SPR a évolué et trouvé de nombreuses applications, notamment dans l'étude des biomolécules.

L'un des partenaires d'interaction est appelé l'analyte : il est injecté par un système de flux au niveau d'une puce avec une surface en or sur laquelle est immobilisé le second partenaire d'interaction, appelé ligand. Plusieurs canaux sont disponibles sur la surface de la puce. L'un des canaux sert de référence tandis que le ou les ligands sont immobilisés sur les canaux restants. L'analyte est injecté sur l'ensemble des canaux de la puce : un signal est émis en cas de liaison de l'analyte. Le signal du canal de référence résulte exclusivement de la liaison éventuelle de l'analyte sur la surface tandis que le signal des canaux actif est la somme de la liaison à la surface et au ligand. Ainsi, le signal dû à la reconnaissance spécifique de l'analyte et du ligand peut être déduit par soustraction.

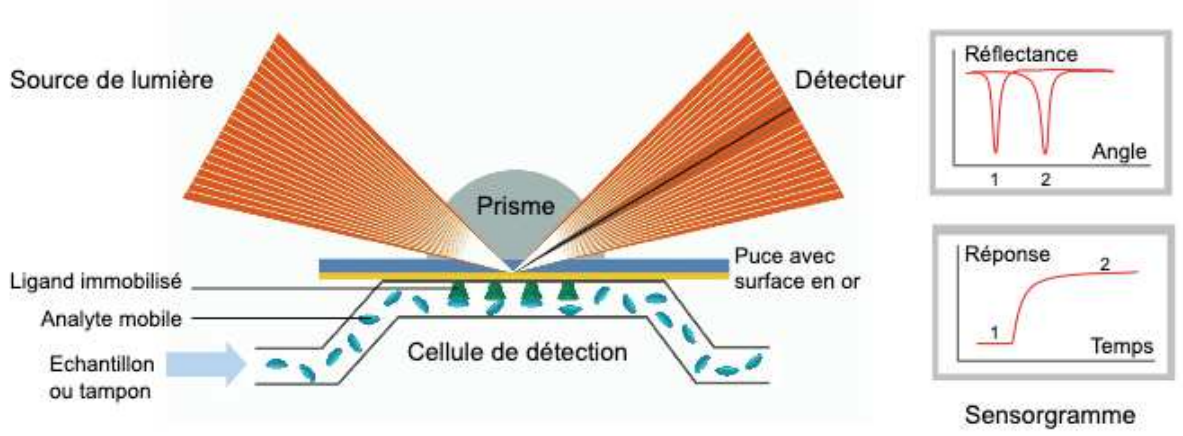


Figure 58 : Principe de la résonance plasmonique de surface. Le ligand est immobilisé sur la surface en or d'une puce de détection et l'analyte est injecté par un système de flux. Un faisceau de lumière polarisée passe à travers un prisme et atteint la surface en or d'une puce de détection : le rayon de lumière incident a un angle permettant la réflexion interne totale. Dans ces conditions, l'angle de réflectance est différent selon que le ligand est libre (1) ou lié à l'analyte (2). Le sensorgramme est le résultat qui présente l'association et la dissociation du complexe en fonction du temps. Adapté de (Gueneau and Dufour, 2019).

Le principe de la SPR est illustré sur la **Figure 58**. Un faisceau de lumière polarisé traverse un prisme avec un angle permettant la réflexion interne totale du faisceau à travers le prisme. Dans ces conditions, le champ électrique des photons s'étend jusqu'à la surface d'or et interagit avec les électrons libres qui s'y trouvent. Les photons sont alors convertis en plasmons, c'est-à-dire en densité électronique de surface, et la lumière n'est plus réfléchie (voir la zone sombre du faisceau réfracté). Lorsque le ligand immobilisé sur la surface d'or est lié par un analyte, cela provoque un

changement de l'indice de réfraction à la surface, qui déplace la zone d'ombre du faisceau réfléchi. Ce décalage de l'intensité de réflectance de l'angle 1 à l'angle 2 est montrée sur la partie supérieure droite de la figure. L'intensité de la réponse en fonction du temps est représentée sur le sensorgramme, qui sert de base à l'interprétation des résultats. Le tracé faisant la transition entre les états 1 et 2 correspond à une cinétique d'association bimoléculaire. Ainsi, la SPR permet de détecter à haute sensibilité et en temps réel l'association de deux molécules, ce qui permet de déterminer les paramètres cinétiques de l'interaction. D'autre part, cette approche est parfois utilisée pour estimer la concentration d'analyte biologiquement actif (Drescher *et al.*, 2018).

L'immobilisation du ligand peut se faire par de différentes approches : il s'agit d'un paramètre déterminant à optimiser pour le bon déroulement de la mesure d'interaction. La **Figure 59** présente différentes stratégies d'immobilisation de ligand couramment utilisées en SPR et qui peuvent également être appliquées à d'autres méthodes impliquant l'immobilisation d'une protéine en préservant son repliement et son activité.

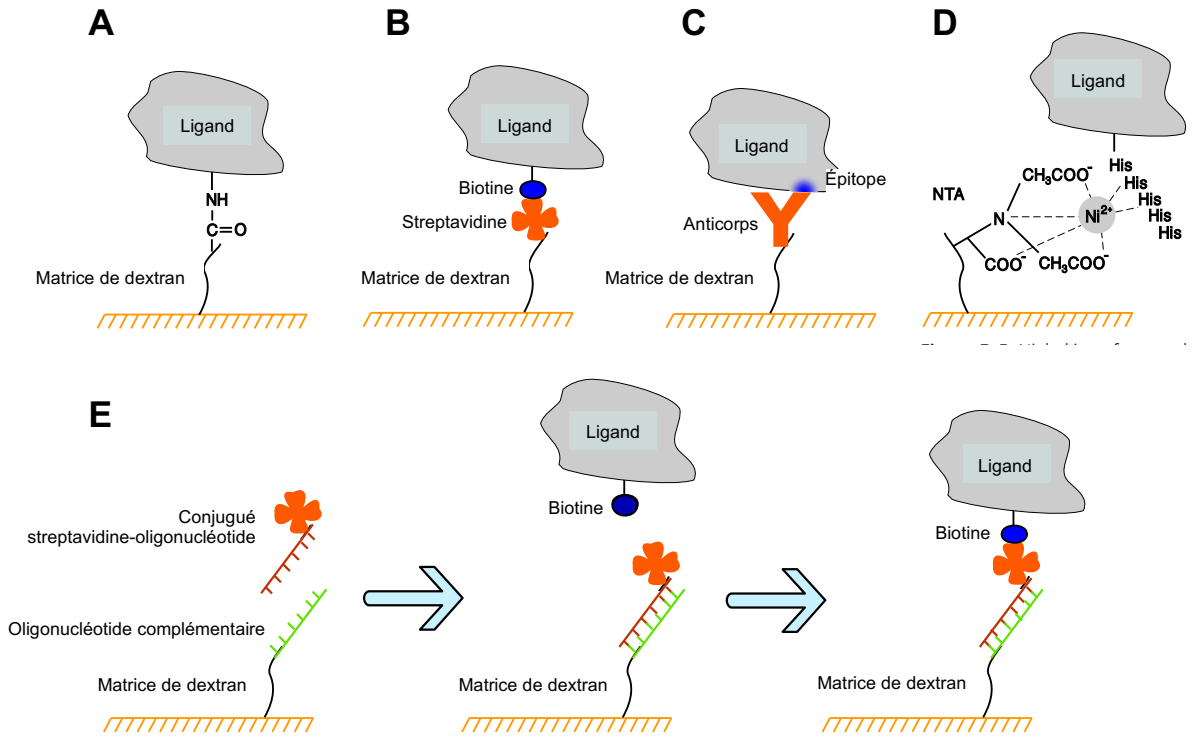


Figure 59 : Modes d'immobilisation de ligand sur la surface de la puce de détection SPR. Tous les types de surface incluent une matrice de dextran qui assure une meilleure accessibilité du ligand immobilisé. **A**. Immobilisation covalente par couplage des amines. **B**. Capture à haute affinité de ligand biotinylé par streptavidine immobilisée sur la surface. **C**. Capture de ligand par un anticorps spécifique. **D**. Capture de ligand porteur d'étiquette poly-histidine par Ni-NTA (ion Nickel immobilisé sur surface par de l'acide nitrilotriacétique). **E**. Système CAP pour l'immobilisation réversible de ligand biotinylé. Un oligonucléotide est immobilisé sur la surface et s'hybride avec un conjugué streptavidine-oligonucléotide complémentaire. Le ligand biotinylé se lie à la streptavidine et peut être lavé de la surface par déshybridation des oligonucléotides. Figure adaptée du manuel d'utilisation Biacore, GE Healthcare.

Le ligand peut être immobilisé par couplage covalent ou par capture. Le couplage covalent est effectué par liaison de groupements fonctionnels réactifs : dans le cas des protéines, il s'agit le plus souvent de groupements amine, mais il existe également des réactions de couplage impliquant les groupements thiol ou aldéhyde. L'immobilisation covalente par les amines ne requiert aucune séquence spécifique et permet de lier l'extrémité N-terminale ainsi que les résidus porteurs d'un groupement amine sur leur chaîne latérale. En revanche, le ligand peut être immobilisé dans une orientation qui rend son site d'interaction inaccessible. De plus, la réaction de couplage implique une étape de pré-concentration du ligand sur la surface par attraction électrostatique. La surface est porteuse de charges négatives, les conditions de pH sont donc choisies pour permettre au ligand de porter des charges positives : le pH est généralement compris entre la constante logarithmique d'acidité de la surface (notée pka et égale à 3,5) et le point isoélectrique de la protéine. Cette étape peut être limitante pour les

protéines sujettes à l'agrégation en conditions acide et exclut les protéines dont le point isoélectrique est inférieur à 3,5 (Gueneau and Dufour, 2019).

La capture se fait par le biais d'une molécule intermédiaire immobilisée sur la surface. Elle permet une immobilisation orientée du ligand pour une accessibilité optimale du site de liaison à l'analyte. En revanche, l'affinité de l'interaction entre le ligand et la molécule de capture doit être suffisamment élevée pour limiter la fuite de ligand pendant le test d'interaction avec l'analyte. Les différents modes de capture incluent l'utilisation d'un couple biotine-streptavidine, d'un anticorps reconnaissant spécifiquement un épitope du ligand ou encore l'immobilisation d'ions Nickel sur la surface permettant de capturer les protéines dotées d'une étiquette poly-histidine. Enfin, un système de capture combinant l'hybridation de deux oligonucléotides complémentaires et l'interaction biotine-streptavidine permet d'immobiliser de façon réversible un ligand biotinylé. Cette stratégie est avantageuse lorsque de nombreux ligands doivent être testés en utilisant une seule puce de détection. Elle permet également d'éviter la perte d'activité d'un ligand immobilisé sur une surface pendant une longue période et permet ainsi de rallonger considérablement la durée d'utilisation d'une même puce de détection.

La SPR étant une technique d'interaction entre un ligand immobilisé et un analyte en solution, le bon déroulement d'une expérience dépend en partie de l'efficacité du transport de masse à la surface de la puce. Dans la configuration utilisée en SPR, l'analyte en solution est mis en contact avec une surface d'or par un flux laminaire. En l'absence de toute interaction entre l'analyte et la surface, la concentration d'analyte est homogène. Lorsque l'analyte est recruté pour interagir avec le ligand, cela crée un gradient de concentration entre la surface et le reste du volume. Pendant l'injection de la solution d'analyte, celui-ci peut être localement déplété dans la zone proche de la surface. A l'opposé, pendant l'injection de tampon, l'analyte peut être retenu sur la surface. La diffusion à travers ce gradient permet de renouveler la quantité d'analyte présente à la surface de la puce : ce phénomène est indispensable aux phases d'accrochage et de décrochage d'analyte sur les molécules de ligand. Dans le cas où le transport de masse est plus lent que les cinétiques d'association et de dissociation du complexe analyte-ligand, la mesure enregistrée en SPR serait représentative du transport de masse et non plus de l'interaction. La phase d'association serait plus lente car l'analyte à la surface ne serait pas renouvelé assez rapidement et la phase de

dissociation serait également ralentie car l'analyte pourrait se lier à nouveau sur la surface (Schuck and Zhao, 2010). En pratique, plus la surface comporte de sites d'interactions disponibles, plus le transport de masse doit être efficace pour assurer le renouvellement d'analyte. Immobiliser le ligand en petite quantité et à un flux élevé garantit une répartition homogène des sites de liaison sur l'ensemble de la surface et limite la ré-association de l'analyte pendant la phase de dissociation. Le choix d'un flux élevé pour l'ensemble de l'expérience permet également de se placer dans des conditions optimales de transport de masse (Gueneau and Dufour, 2019).

Enfin, la méthode dite d'interférométrie de bio-couche présente certaines similarités avec la SPR : il s'agit également d'une technique optique de détection d'interactions bimoléculaires, permettant de mesurer les cinétiques d'association et de dissociation de deux molécules. Le profil d'interférence de la lumière blanche est mesuré dans du tampon et lorsque les deux partenaires sont mis en présence. Le changement entre ces deux états produit un signal qui permet de quantifier en temps réel l'association et la dissociation des deux partenaires.

3.1.3 Titration calorimétrique isotherme

La titration calorimétrique isotherme est à ce jour l'approche la plus couramment utilisée pour déterminer les paramètres thermodynamiques d'une interaction. Elle s'applique aussi bien aux protéines qu'aux acides nucléiques, lipides et petites molécules (Falconer, 2016). Cette technique ne requiert aucune sonde fluorescente ni immobilisation de l'un des deux partenaires et permet l'étude thermodynamique de l'interaction entre deux molécules en solution. En une seule expérience, elle permet de déterminer l'affinité (K_D), la variation d'enthalpie (ΔH), la variation d'entropie (ΔS), la variation d'énergie libre (ΔG) et la stœchiométrie d'une interaction.

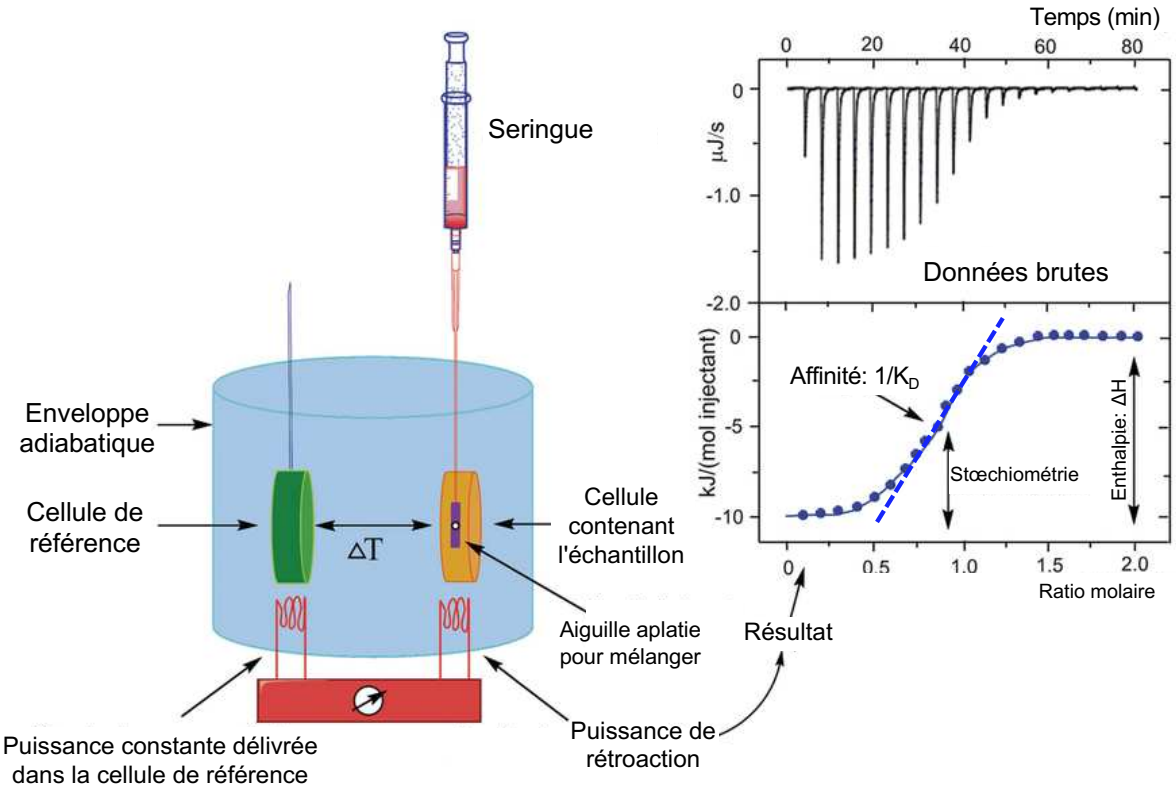


Figure 60 : Principe de la titration calorimétrique isotherme. L'appareil contient deux cellules entourées une enveloppe adiabatique qui les isole des changements de température extérieurs. L'une des cellules contient du tampon de travail ou de l'eau (cellule de référence) et les deux partenaires d'interaction sont mis en présence dans la cellule contenant l'échantillon, grâce à une seringue qui injecte le ligand. Une température constante est maintenue dans les deux cellules grâce à un dispositif qui compense les dégagements de chaleur dus à la formation du complexe dans la cellule contenant l'échantillon. La puissance de rétroaction nécessaire pour chaque injection de ligand est enregistrée. Le résultat final est appelé l'isotherme : ce graphe permet de déterminer la variation d'enthalpie, la constante de dissociation ainsi que la stœchiométrie de la réaction. Adapté de (Song *et al.*, 2015).

Le principe a été publié pour la première fois en 1990 (Freire *et al.*, 1990). L'appareil comporte deux cellules identiques entourées d'une enveloppe adiabatique (**Figure 60**). L'une des cellules contient le tampon de travail seul ou de l'eau (cellule de référence) et la seconde cellule contient l'échantillon : les deux partenaires d'interaction y sont mis en présence à l'aide d'une seringue. La concentration du partenaire d'interaction présent dans la cellule demeure constante tandis que le second partenaire (ligand) est injecté progressivement par la seringue. Une puissance constante est appliquée sur la cellule de référence. Afin de maintenir les deux cellules à une température égale, un courant de rétroaction appliqué sur la cellule contenant l'échantillon compense les variations de chaleur provoquées par la réaction de formation du complexe. Cette réaction peut être exothermique ou endothermique, selon qu'elle dégage ou qu'elle consomme de la chaleur. Dans le cas d'une réaction exothermique, le dégagement de chaleur dû à la formation de complexe est capté par

l'appareil et le courant de rétroaction est réduit afin que la température demeure constante (Ramesh, 2019). L'appareil enregistre les variations d'intensité du courant de rétroaction, qui sont représentés par des pics de valeur négative dans le cas d'une réaction exothermique (voir données brutes sur la **Figure 60**). Chaque pic correspond à un point de la gamme de concentration. La quantité de ligand augmente graduellement jusqu'à ce que la réaction atteigne un plateau de saturation lorsque la cellule contient un excès de ligand. Les pics sont ensuite intégrés et tracés en fonction de la quantité de ligand injecté : le graphe qui en résulte est appelé un isotherme. La stœchiométrie de la réaction ainsi la variation d'enthalpie (ΔH) peuvent être directement lus sur l'isotherme et la constante de dissociation (K_D) peut être déterminée par un modèle de liaison.

L'ITC est une méthode puissante pour l'étude des interactions car elle permet de déterminer les contributions énergétiques et d'avoir accès aux paramètres intrinsèques de l'interaction. Ainsi, cette approche permet d'identifier tout changement conformationnel ou protonation de l'un des deux partenaires requis pour l'interaction. De tels éléments améliorent grandement la compréhension des mécanismes par lesquels deux molécules sont amenées à interagir. D'autre part, l'ITC permet de quantifier à haute précision des affinités extrêmement élevées, comme dans le cas du couple biotine-streptavidine.

L'une des limitations liées à l'ITC est qu'elle requiert d'importantes quantités de chaque partenaire d'interaction, qui doivent tous deux être solubles et stables à forte concentration (de l'ordre de 10 à 100 µM). De plus, aucune réaction autre que la formation du complexe ne doit avoir lieu pendant la mesure afin de ne pas interférer sur les dégagements de chaleur mesurés (ex: réaction d'oxydo-réduction entre l'oxygène et des agents réducteurs présents dans le tampon) (Baranauskiene *et al.*, 2019).

3.1.4 Thermophorèse à micro-échelle

La thermophorèse à micro-échelle (MST) est une approche pour l'étude d'interactions bimoléculaires qui combine détection de fluorescence et thermophorèse. Elle peut s'utiliser sur des molécules porteuses d'une sonde fluorescente ou en utilisant uniquement la fluorescence intrinsèque du tryptophane.

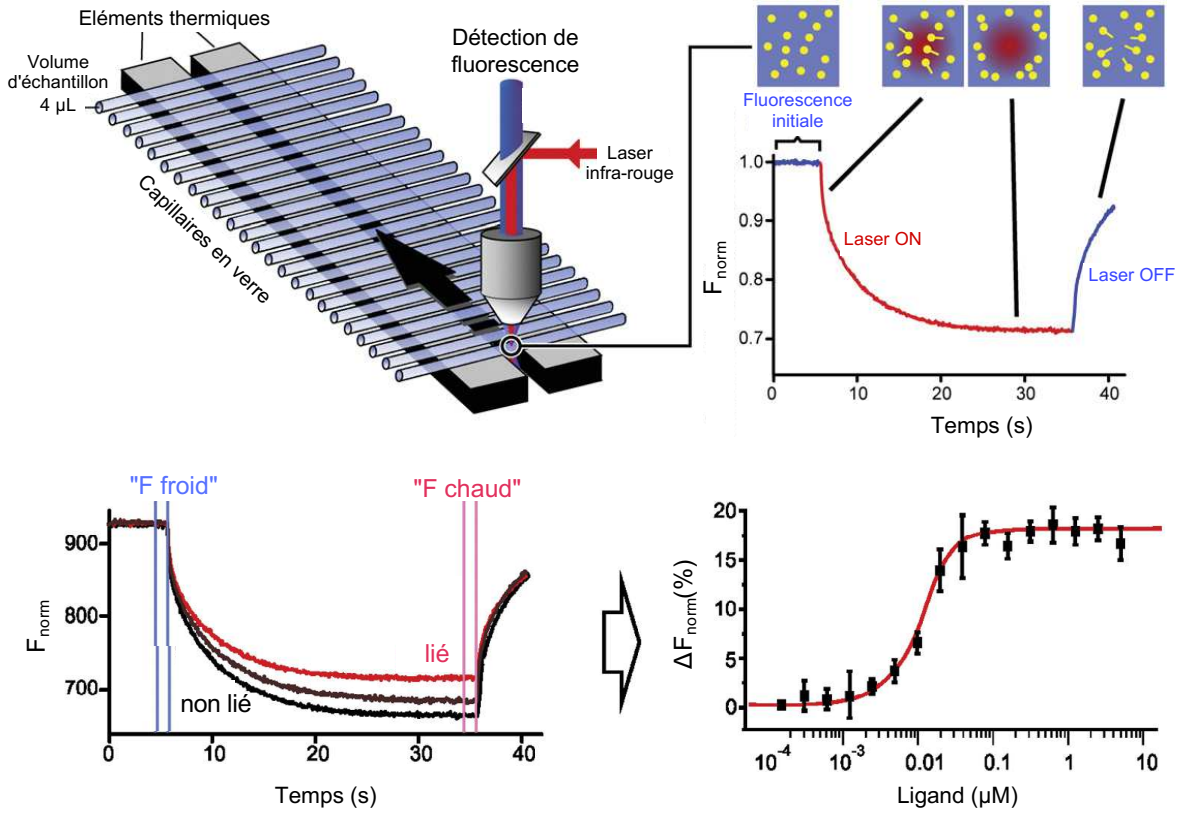


Figure 61 : Principe de la thermophorèse à micro-échelle. L'expérience nécessite différents mélanges des deux partenaires d'interaction, avec une gamme de concentration du ligand et une concentration constante du partenaire d'interaction émettant une fluorescence. Les échantillons contenant les deux partenaires d'interaction se trouvent dans des capillaires en verre. Un laser IR induit un gradient de température dans l'échantillon, qui active le mouvement thermophorétique et l'excitation du fluorophore. La fluorescence initiale est appelée "F_{froid}" et la fluorescence après excitation du fluorophore est appelée "F_{chaud}". La variation de la fluorescence normalisée (ΔF_{norm}) pour chaque concentration de ligand permet de déterminer la constante de dissociation. Adapté de (Alexander *et al.*, 2014) et (Jerabek-Willemsen *et al.*, 2014)

Le principe de la thermophorèse a été publié pour la première fois en 1856 (Ludwig *et al.*, 1856). L'échantillon est présent dans des capillaires de verre d'une capacité de 4 µL environ (**Figure 61**). Un laser infra-rouge induit un gradient de température microscopique au niveau de l'échantillon (de l'ordre de 2 à 6 K). Ce gradient active le mouvement thermophorétique moléculaire, qui provoque l'excitation des fluorophores. La fluorescence émise est captée par le même objectif que le laser infra-rouge. Ainsi, l'appareil enregistre la déplétion ou l'accumulation des fluorophores dans le gradient de température induit par le laser. La constante de dissociation est déterminée par une gamme de concentration de ligand incubée avec une quantité constante de partenaire émettant un signal fluorescent. Ces solutions sont placées dans différents capillaires qui peuvent être scannés en une seule expérience. La variation de fluorescence

mesurée pour chaque point de concentration permet de déterminer la constante de dissociation du complexe, selon un modèle similaire à celui présenté sur la **Figure 57**.

La MST peut s'appliquer à l'hybridation de deux molécules d'ADN, aux interactions ADN-protéine ainsi qu'aux interactions protéine-protéine. Les constantes de dissociation peuvent être estimées pour des fortes affinités, de l'ordre du nM et pM). Cette approche présente l'avantage de consommer de faibles quantités d'échantillon et de délivrer un résultat en peu de temps de mesure. Les protéines étudiées peuvent être utilisées sous forme purifiée ou exprimées en fusion avec une protéine fluorescente dans des lysats cellulaires. De plus, elle peut être utilisée pour étudier les étapes de dénaturation des protéines en présence de concentrations croissantes d'agent chaotropique (ex: urée, guanidine). La limite principale de cette approche est que bien que la mesure reste possible en l'absence de fluorophores, c'est-à-dire en se basant uniquement sur la fluorescence intrinsèque des résidus tryptophane, une majorité de protéines contient un grand nombre de résidus aromatiques susceptibles de provoquer des interférences (Sparks and Fratti, 2019). Il est donc souvent préférable d'utiliser des protéines fusionnées fluorescentes ou couplées à un fluorophore, ce qui ne permet pas toujours de préserver le repliement des partenaires et peut générer des interactions non spécifiques dues à la présence du fluorophore.

References

Accardi, R., and Gheit, T. (2014). Cutaneous HPV and skin cancer. La Presse Médicale *43*, e435–e443.

Aksoy, P., Gottschalk, E.Y., and Meneses, P.I. (2017). HPV entry into cells. Mutation Research/Reviews in Mutation Research *772*, 13–22.

Albert, T.K., Hanzawa, H., Legtenberg, Y.I.A., de Ruwe, M.J., van den Heuvel, F.A.J., Collart, M.A., Boelens, R., and Timmers, H.Th.M. (2002). Identification of a ubiquitin–protein ligase subunit within the CCR4–NOT transcription repressor complex. The EMBO Journal *21*, 355–364.

Alemany, L., Cubilla, A., Halec, G., Kasamatsu, E., Quirós, B., Masferrer, E., Tous, S., Lloveras, B., Hernández-Suarez, G., Lonsdale, R., Tinoco, L., Alejo, M., Alvarado-Cabrero, I., Laco, J., Guimerà, N., Poblet, E., Lombardi, L.E., Bergeron, C., Clavero, O., Shin, H.-R., Ferrera, A., Felix, A., Germar, J., Mandys, V., Clavel, C., Tzardi, M., Pons, L.E., Wain, V., Cruz, E., Molina, C., Mota, J.D., Jach, R., Velasco, J., Carrilho, C., López-Revilla, R., Goodman, M.T., Quint, W.G., Castellsagué, X., Bravo, I., Pawlita, M., Muñoz, N., Bosch, F.X., and de Sanjosé, S. (2016). Role of Human Papillomavirus in Penile Carcinomas Worldwide. European Urology *69*, 953–961.

Alexander, C.G., Wanner, R., Johnson, C.M., Breitsprecher, D., Winter, G., Duhr, S., Baaske, P., and Ferguson, N. (2014). Novel microscale approaches for easy, rapid determination of protein stability in academic and commercial settings. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics *1844*, 2241–2250.

Amacher, J.F., Cushing, P.R., Brooks, L., Boisguerin, P., and Madden, D.R. (2014). Stereochemical Preferences Modulate Affinity and Selectivity among Five PDZ Domains that Bind CFTR: Comparative Structural and Sequence Analyses. Structure *22*, 82–93.

Androphy, E.J., Hubbert, N.L., Schiller, J.T., and Lowy, D.R. (1987). Identification of the HPV-16 E6 protein from transformed mouse cells and human cervical carcinoma cell lines. The EMBO Journal *6*, 989–992.

Angelman, H. (1965). 'Puppet' Children A Report on Three Cases. Developmental Medicine & Child Neurology *7*, 681–688.

Arroyo, L.S., Smelov, V., Bzhalava, D., Eklund, C., Hultin, E., and Dillner, J. (2013). Next generation sequencing for human papillomavirus genotyping. Journal of Clinical Virology *58*, 437–442.

Au, W.C., Moore, P.A., Lowther, W., Juang, Y.T., and Pitha, P.M. (1995). Identification of a member of the interferon regulatory factor family that binds to the interferon-stimulated response element and activates expression of interferon-induced genes. PNAS *92*, 11657–11661.

Aubin, F., and Guerrini, J.-S. (2007). Aspects cliniques bénins de l'infection à papillomavirus humains. Bulletin de l'Académie Nationale de Médecine *191*, 585–599.

Aydin, I., Weber, S., Snijder, B., Samperio Ventayol, P., Kühbacher, A., Becker, M., Day, P.M., Schiller, J.T., Kann, M., Pelkmans, L., Helenius, A., and Schelhaas, M. (2014). Large Scale RNAi Reveals the Requirement of Nuclear Envelope Breakdown for Nuclear Import of Human Papillomaviruses. PLoS Pathog *10*.

Bachelerie, F. (2010). CXCL12/CXCR4-axis dysfunctions: Markers of the rare immunodeficiency disorder WHIM syndrome. Disease Markers *29*, 189–198.

Baker, T.S., Newcomb, W.W., Olson, N.H., Cowsert, L.M., Olson, C., and Brown, J.C. (1991). Structures of bovine and human papillomaviruses. Analysis by cryoelectron microscopy and three-dimensional image reconstruction. Biophysical Journal *60*, 1445–1456.

Balabanian, K., Lagane, B., Pablos, J.L., Laurent, L., Planchenault, T., Verola, O., Lebbe, C., Kerob, D., Dupuy, A., Hermine, O., Nicolas, J.-F., Latger-Cannard, V., Bensoussan, D., Bordigoni, P., Baleux, F., Le Deist, F., Virelizier, J.-L., Arenzana-Seisdedos, F., and Bachelerie, F. (2005). WHIM syndromes with different genetic anomalies are accounted for by impaired CXCR4 desensitization to CXCL12. Blood *105*, 2449–2457.

Baldwin, P., Laskey, R., and Coleman, N. (2003). Translational approaches to improving cervical screening. Nat Rev Cancer *3*, 217–226.

Baleja, J.D., Cherry, J.J., Liu, Z., Gao, H., Nicklaus, M.C., Voigt, J.H., Chen, J.J., and Androphy, E.J. (2006). Identification of inhibitors to papillomavirus type 16 E6

protein based on three-dimensional structures of interacting proteins. Antiviral Research *72*, 49–59.

Banks, L., Pim, D., and Thomas, M. (2012). Human tumour viruses and the deregulation of cell polarity in cancer. Nature Reviews Cancer *12*, 877–886.

Bañuelos, A., Reyes, E., Ocadiz, R., Alvarez, E., Moreno, M., Monroy, A., and Gariglio, P. (2003). Neocarzinostatin Induces an Effective p53-Dependent Response in Human Papillomavirus-Positive Cervical Cancer Cells. J Pharmacol Exp Ther *306*, 671–680.

Baranauskiene, L., Kuo, T.-C., Chen, W.-Y., and Matulis, D. (2019). Isothermal titration calorimetry for characterization of recombinant proteins. Current Opinion in Biotechnology *55*, 9–15.

Barbosa, M. s., Edmonds, C., Fisher, C., Schiller, J. t., Lowy, D. r., and Vousden, K. h. (1990). The region of the HPV E7 oncoprotein homologous to adenovirus E1a and Sv40 large T antigen contains separate domains for Rb binding and casein kinase II phosphorylation. The EMBO Journal *9*, 153–160.

Barrow-Laing, L., Chen, W., and Roman, A. (2010). Low- and high-risk human papillomavirus E7 proteins regulate p130 differently. Virology *400*, 233–239.

Basukala, O., Sarabia-Vega, V., and Banks, L. (2020). Human papillomavirus oncoproteins and post-translational modifications: generating multifunctional hubs for overriding cellular homeostasis. Biological Chemistry *401*, 585–599.

Bedell, M.A., Jones, K.H., Grossman, S.R., and Laimins, L.A. (1989). Identification of human papillomavirus type 18 transforming genes in immortalized and primary cells. Journal of Virology *63*, 1247–1255.

Bienkowska-Haba, M., Patel, H.D., and Sapp, M. (2009). Target Cell Cyclophilins Facilitate Human Papillomavirus Type 16 Infection. PLoS Pathog *5*.

Bischof, O., Schwamborn, K., Martin, N., Werner, A., Sustmann, C., Grosschedl, R., and Dejean, A. (2006). The E3 SUMO Ligase PIASy Is a Regulator of Cellular Senescence and Apoptosis. Molecular Cell *22*, 783–794.

Bogaert, L., Martens, A., Van Poucke, M., Ducatelle, R., De Cock, H., Dewulf, J., De Baere, C., Peelman, L., and Gasthuys, F. (2008). High prevalence of bovine papillomaviral DNA in the normal skin of equine sarcoid-affected and healthy horses. Veterinary Microbiology *129*, 58–68.

Bonhoure, A., Demenge, A., Kostmann, C., San José, L., De la Cal, E., Armisen, P., Nominé, Y., and Travé, G. (2018). One-step affinity purification of fusion proteins with optimal monodispersity and biological activity: application to aggregation-prone HPV E6 proteins. Microbial Cell Factories *17*.

Bonhoure, A., Forster, A., Babah, K.O., Gógl, G., Eberling, P., Kostmann, C., Volkmer, R., Tapia Mancilla, V., Travé, G., and Nominé, Y. (2020). Benchtop holdup assay for quantitative affinity-based analysis of sequence determinants of protein-motif interactions. Analytical Biochemistry *603*, 113772.

Bonilla, X., Parmentier, L., King, B., Bezrukov, F., Kaya, G., Zoete, V., Seplyarskiy, V.B., Sharpe, H.J., McKee, T., Letourneau, A., Ribaux, P.G., Popadin, K., Basset-Seguin, N., Chaabene, R.B., Santoni, F.A., Andrianova, M.A., Guipponi, M., Garieri, M., Verdan, C., Grosdemange, K., Sumara, O., Eilers, M., Aifantis, I., Michielin, O., de Sauvage, F.J., Antonarakis, S.E., and Nikolaev, S.I. (2016). Genomic analysis identifies new drivers and progression pathways in skin basal cell carcinoma. Nature Genetics *48*, 398–406.

Bosch, F.X., Broker, T.R., Forman, D., Moscicki, A.-B., Gillison, M.L., Doorbar, J., Stern, P.L., Stanley, M., Arbyn, M., Poljak, M., Cuzick, J., Castle, P.E., Schiller, J.T., Markowitz, L.E., Fisher, W.A., Canfell, K., Denny, L.A., Franco, E.L., Steben, M., Kane, M.A., Schiffman, M., Meijer, C.J.L.M., Sankaranarayanan, R., Castellsagué, X., Kim, J.J., Brotons, M., Alemany, L., Albero, G., Diaz, M., and Sanjosé, S. de (2013). Comprehensive Control of Human Papillomavirus Infections and Related Diseases. Vaccine *31*, H1–H31.

Boshart, M., Gissmann, L., Ikenberg, H., Kleinheinz, A., Scheurlen, W., and zur Hausen, H. (1984). A new type of papillomavirus DNA, its presence in genital cancer biopsies and in cell lines derived from cervical cancer. The EMBO Journal *3*, 1151–1157.

Boyer, S.N., Wazer, D.E., and Band, V. (1996). E7 Protein of Human Papilloma Virus-16 Induces Degradation of Retinoblastoma Protein through the Ubiquitin-Proteasome Pathway. Cancer Res *56*, 4620–4624.

Braun, J.E., Huntzinger, E., Fauser, M., and Izaurralde, E. (2011). GW182 Proteins Directly Recruit Cytoplasmic Deadenylase Complexes to miRNA Targets. Molecular Cell *44*, 120–133.

Brimer, N., and Vande Pol, S.B. (2014). Papillomavirus E6 PDZ interactions can be replaced by repression of p53 to promote episomal human papillomavirus genome maintenance. J. Virol. *88*, 3027–3030.

Brimer, N., Lyons, C., Wallberg, A.E., and Pol, S.V. (2012). Cutaneous papillomavirus E6 oncoproteins associate with MAML1 to repress transactivation and NOTCH signaling. Oncogene *31*, 4639–4646.

Brimer, N., Drews, C.M., and Vande Pol, S.B. (2017). Association of papillomavirus E6 proteins with either MAML1 or E6AP clusters E6 proteins by structure, function, and evolutionary relatedness. PLOS Pathogens *13*, e1006781.

Buck, C.B., Cheng, N., Thompson, C.D., Lowy, D.R., Steven, A.C., Schiller, J.T., and Trus, B.L. (2008). Arrangement of L2 within the Papillomavirus Capsid. Journal of Virology *82*, 5190–5197.

Burnouf, D., Ennifar, E., Guedich, S., Puffer, B., Hoffmann, G., Bec, G., Disdier, F., Baltzinger, M., and Dumas, P. (2012). kinITC: A New Method for Obtaining Joint Thermodynamic and Kinetic Data by Isothermal Titration Calorimetry. J. Am. Chem. Soc. *134*, 559–565.

Butler, I.B., Schoonen, M.A.A., and Rickard, D.T. (1994). Removal of dissolved oxygen from water: A comparison of four common techniques. Talanta *41*, 211–215.

Butz, K., Shahabeddin, L., Geisen, C., Spitkovsky, D., Ullmann, A., and Hoppe-Seyler, F. (1995). Functional p53 protein in human papillomavirus-positive cancer cells. Oncogene *10*, 927–936.

Butz, K., Ristriani, T., Hengstermann, A., Denk, C., Scheffner, M., and Hoppe-

Seyler, F. (2003). siRNA targeting of the viral E6 oncogene efficiently kills human papillomavirus-positive cancer cells. Oncogene *22*, 5938–5945.

Bzhalava, D., Eklund, C., and Dillner, J. (2015). International standardization and classification of human papillomavirus types. Virology *476*, 341–344.

Campo, M.S. (2006). Bovine Papillomavirus: old system, new lessons? In Papillomavirus Research: From Natural History to Vaccine and Beyond, (Norfolk, England), p. 34.

Cardoso, J.C., and Calonje, E. (2011). Cutaneous manifestations of human papillomaviruses: a review. Acta Dermatovenerol Alp Pannonica Adriat *20*, 145–154.

Carpenter, R.C. (1895). Calorimeter.

Cassonnet, P., Rolloy, C., Neveu, G., Vidalain, P.-O., Chantier, T., Pellet, J., Jones, L., Muller, M., Demeret, C., Gaud, G., Vuillier, F., Lotteau, V., Tangy, F., Favre, M., and Jacob, Y. (2011). Benchmarking a luciferase complementation assay for detecting protein complexes. Nature Methods *8*, 990–992.

Castellsagué, X., Alemany, L., Quer, M., Halec, G., Quirós, B., Tous, S., Clavero, O., Alòs, L., Biegner, T., Szafarowski, T., Alejo, M., Holzinger, D., Cadena, E., Claros, E., Hall, G., Laco, J., Poljak, M., Benevolo, M., Kasamatsu, E., Mehanna, H., Ndiaye, C., Guimerà, N., Lloveras, B., León, X., Ruiz-Cabezas, J.C., Alvarado-Cabrero, I., Kang, C.-S., Oh, J.-K., Garcia-Rojo, M., Iljazovic, E., Ajayi, O.F., Duarte, F., Nessa, A., Tinoco, L., Duran-Padilla, M.A., Pirog, E.C., Viarheichyk, H., Morales, H., Costes, V., Félix, A., Germar, M.J.V., Mena, M., Ruacan, A., Jain, A., Mehrotra, R., Goodman, M.T., Lombardi, L.E., Ferrera, A., Malami, S., Albanesi, E.I., Dabed, P., Molina, C., López-Revilla, R., Mandys, V., González, M.E., Velasco, J., Bravo, I.G., Quint, W., Pawlita, M., Muñoz, N., Sanjosé, S. de, and Xavier Bosch, F. (2016). HPV Involvement in Head and Neck Cancers: Comprehensive Assessment of Biomarkers in 3680 Patients. J Natl Cancer Inst *108*.

Catoe, H.W., and Nawaz, Z. (2011). E6-AP facilitates efficient transcription at estrogen responsive promoters through recruitment of chromatin modifiers. Steroids *76*, 897–902.

Celegato, M., Messa, L., Goracci, L., Mercorelli, B., Bertagnin, C., Spyrakis, F., Suarez, I., Cousido-Siah, A., Travé, G., Banks, L., Cruciani, G., Palù, G., and Loregian, A. (2020). A novel small-molecule inhibitor of the human papillomavirus E6-p53 interaction that reactivates p53 function and blocks cancer cells growth. Cancer Letters *470*, 115–125.

Chan, H.M., and Thangue, N.B.L. (2001). p300/CBP proteins: HATs for transcriptional bridges and scaffolds. Journal of Cell Science *114*, 2363–2373.

Charbonnier, S., Zanier, K., Masson, M., and Travé, G. (2006). Capturing protein–protein complexes at equilibrium: The holdup comparative chromatographic retention assay. Protein Expression and Purification *50*, 89–101.

Charbonnier, S., Nominé, Y., Ramírez, J., Luck, K., Chapelle, A., Stote, R.H., Travé, G., Kieffer, B., and Atkinson, R.A. (2011). The Structural and Dynamic Response of MAGI-1 PDZ1 with Noncanonical Domain Boundaries to the Binding of Human Papillomavirus E6. Journal of Molecular Biology *406*, 745–763.

Chemes, L.B., Sánchez, I.E., Smal, C., and Prat-Gay, G. de (2010). Targeting mechanism of the retinoblastoma tumor suppressor by a prototypical viral oncoprotein. The FEBS Journal *277*, 973–988.

Chemes, L.B., Camporeale, G., Sánchez, I.E., de Prat-Gay, G., and Alonso, L.G. (2014). Cysteine-Rich Positions Outside the Structural Zinc Motif of Human Papillomavirus E7 Provide Conformational Modulation and Suggest Functional Redox Roles. Biochemistry *53*, 1680–1696.

Chen, J.J., Reid, C.E., Band, V., and Androphy, E.J. (1995). Interaction of papillomavirus E6 oncoproteins with a putative calcium-binding protein. Science *269*, 529–531.

Cherry, J.J., Rietz, A., Malinkevich, A., Liu, Y., Xie, M., Bartolowits, M., Davisson, V.J., Baleja, J.D., and Androphy, E.J. (2013). Structure Based Identification and Characterization of Flavonoids That Disrupt Human Papillomavirus-16 E6 Function. PLOS ONE *8*, e84506.

Cheville, N.F., and Olson, C. (1964). Epithelial and Fibroblastic Proliferation in

Bovine Cutaneous Papillomatosis. Pathologia Veterinaria *1*, 248–257.

Cole, S.T., and Danos, O. (1987). Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome: Phylogeny of papillomaviruses and repeated structure of the E6 and E7 gene products. Journal of Molecular Biology *193*, 599–608.

Collart, M.A. (2016). The Ccr4-Not complex is a key regulator of eukaryotic gene expression. WIREs RNA *7*, 438–454.

Cornet, I., Bouvard, V., Campo, M.S., Thomas, M., Banks, L., Gissmann, L., Lamartine, J., Sylla, B.S., Accardi, R., and Tommasino, M. (2012). Comparative Analysis of Transforming Properties of E6 and E7 from Different Beta Human Papillomavirus Types. Journal of Virology *86*, 2366–2370.

Danos, O., Katinka, M., and Yaniv, M. (1980). Molecular Cloning, Refined Physical Map and Heterogeneity of Methylation Sites of Papilloma Virus Type 1a DNA. European Journal of Biochemistry *109*, 457–461.

Danos, O., Katinka, M., and Yaniv, M. (1982). Human papillomavirus 1a complete DNA sequence: a novel type of genome organization among papovaviridae. The EMBO Journal *1*, 231–236.

Davey, N.E., Seo, M.-H., Yadav, V.K., Jeon, J., Nim, S., Krystkowiak, I., Blikstad, C., Dong, D., Markova, N., Kim, P.M., and Ivarsson, Y. (2017). Discovery of short linear motif-mediated interactions through phage display of intrinsically disordered regions of the human proteome. The FEBS Journal *284*, 485–498.

DeCaprio, J.A., Ludlow, J.W., Figge, J., Shew, J.-Y., Huang, C.-M., Lee, W.-H., Marsilio, E., Paucha, E., and Livingston, D.M. (1988). SV40 large tumor antigen forms a specific complex with the product of the retinoblastoma susceptibility gene. Cell *54*, 275–283.

Degenhardt, Y.Y., and Silverstein, S.J. (2001). Gps2, a Protein Partner for Human Papillomavirus E6 Proteins. Journal of Virology *75*, 151–160.

Delury, C.P., Marsh, E.K., James, C.D., Boon, S.S., Banks, L., Knight, G.L., and

Roberts, S. (2013). The Role of Protein Kinase A Regulation of the E6 PDZ-Binding Domain during the Differentiation-Dependent Life Cycle of Human Papillomavirus Type 18. Journal of Virology *87*, 9463–9472.

Doorbar, J. (2018). Host control of human papillomavirus infection and disease. Best Practice & Research Clinical Obstetrics & Gynaecology *47*, 27–41.

Doorbar, J., Quint, W., Banks, L., Bravo, I.G., Stoler, M., Broker, T.R., and Stanley, M.A. (2012). The Biology and Life-Cycle of Human Papillomaviruses. Vaccine *30*, F55–F70.

Doorbar, J., Egawa, N., Griffin, H., Kranjec, C., and Murakami, I. (2015). Human papillomavirus molecular biology and disease association. Reviews in Medical Virology *25*, 2–23.

Drescher, D.G., Selvakumar, D., and Drescher, M.J. (2018). Chapter One - Analysis of Protein Interactions by Surface Plasmon Resonance. In Advances in Protein Chemistry and Structural Biology, R. Donev, ed. (Academic Press), pp. 1–30.

Drews, C.M., Brimer, N., and Pol, S.B.V. (2020). Multiple regions of E6AP (UBE3A) contribute to interaction with papillomavirus E6 proteins and the activation of ubiquitin ligase activity. PLOS Pathogens *16*, e1008295.

Durst, M., Gissmann, L., Ikenberg, H., and zur Hausen, H. (1983). A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions. Proceedings of the National Academy of Sciences *80*, 3812–3815.

Dymalla, S., Scheffner, M., Weber, E., Sehr, P., Lohrey, C., Hoppe-Seyler, F., and Hoppe-Seyler, K. (2009). A novel peptide motif binding to and blocking the intracellular activity of the human papillomavirus E6 oncoprotein. J Mol Med *87*, 321–331.

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

Eckhardt, M., Zhang, W., Gross, A.M., Von Dollen, J., Johnson, J.R., Franks-Skiba, K.E., Swaney, D.L., Johnson, T.L., Jang, G.M., Shah, P.S., Brand, T.M., Archambault, J., Kreisberg, J.F., Grandis, J.R., Ideker, T., and Krogan, N.J. (2018). Multiple Routes to Oncogenesis Are Promoted by the Human Papillomavirus–Host Protein Network. Cancer Discovery *8*, 1474–1489.

Eichten, A., Westfall, M., Pietenpol, J.A., and Münger, K. (2002). Stabilization and Functional Impairment of the Tumor Suppressor p53 by the Human Papillomavirus Type 16 E7 Oncoprotein. Virology *295*, 74–85.

Ekström, J., Bzhalava, D., Svenback, D., Forslund, O., and Dillner, J. (2011). High throughput sequencing reveals diversity of Human Papillomaviruses in cutaneous lesions. International Journal of Cancer *129*, 2643–2650.

Falconer, R.J. (2016). Applications of isothermal titration calorimetry – the research and technical developments from 2011 to 2015. Journal of Molecular Recognition *29*, 504–515.

Fernandez-Fernandez, M.R., and Sot, B. (2011). The relevance of protein-protein interactions for p53 function: the CPE contribution. Protein Engineering Design and Selection *24*, 41–51.

Finnigan, G.C., Duvalyan, A., Liao, E.N., Sargsyan, A., and Thorner, J. (2016). Detection of protein–protein interactions at the septin collar in Saccharomyces cerevisiae using a tripartite split-GFP system. Molecular Biology of the Cell *27*, 2708–2725.

Freire, E., Mayorga, O.L., and Straume, M. (1990). Isothermal titration calorimetry. Anal. Chem. *62*, 950A-959A.

Fryer, C.J., Lamar, E., Turbachova, I., Kintner, C., and Jones, K.A. (2002). Mastermind mediates chromatin-specific transcription and turnover of the Notch enhancer complex. Genes Dev. *16*, 1397–1411.

Fusconi, M., Grasso, M., Graco, A., Gallo, A., Campo, F., Remacle, M., Turchetta, R., Pagliuca, G., and De Vicentiis, M. (2014). Recurrent respiratory papillomatosis by HPV: review of the literature and update on the use of cidofovir. Acta

Otorhinolaryngol Ital *34*, 375–381.

Ganti, K., Broniarczyk, J., Manoubi, W., Massimi, P., Mittal, S., Pim, D., Szalmas, A., Thatte, J., Thomas, M., Tomaić, V., and Banks, L. (2015). The Human Papillomavirus E6 PDZ Binding Motif: From Life Cycle to Malignancy. Viruses *7*, 3530–3551.

Gao, Q., Srinivasan, S., Boyer, S.N., Wazer, D.E., and Band, V. (1999). The E6 Oncoproteins of High-Risk Papillomaviruses Bind to a Novel Putative GAP Protein, E6TP1, and Target It for Degradation. Mol Cell Biol *19*, 733–744.

García-Alai, M.M., Alonso, L.G., and 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.

Genera, M., Samson, D., Raynal, B., Haouz, A., Baron, B., Simenel, C., Guerois, R., Wolff, N., and Caillet-Saguy, C. (2019). Structural and functional characterization of the PDZ domain of the human phosphatase PTPN3 and its interaction with the human papillomavirus E6 oncoprotein. Sci Rep *9*, 1–12.

Germaniuk-Kurowska, A., Nag, A., Zhao, X., Dimri, M., Band, H., and Band, V. (2007). Ada3 Requirement for HAT Recruitment to Estrogen Receptors and Estrogen-Dependent Breast Cancer Cell Proliferation. Cancer Res *67*, 11789–11797.

Gheit, T. (2019). Mucosal and Cutaneous Human Papillomavirus Infections and Cancer Biology. Front. Oncol. *9*.

Giarrè, M., Caldeira, S., Malanchi, I., Ciccolini, F., Leão, M.J., and Tommasino, M. (2001). Induction of pRb Degradation by the Human Papillomavirus Type 16 E7 Protein Is Essential To Efficiently Overcome p16INK4a-Imposed G1 Cell Cycle Arrest. Journal of Virology *75*, 4705–4712.

Gissmann, L., and zur Hausen, H. (1980). Partial characterization of viral DNA from human genital warts (condylomata acuminata). International Journal of Cancer *25*, 605–609.

Gógl, G., Jane, P., Caillet-Saguy, C., Kostmann, C., Bich, G., Cousido-Siah, A., Nyitray, L., Vincentelli, R., Wolff, N., Nominé, Y., Sluchanko, N.N., and Travé, G. (2020a). Dual Specificity PDZ- and 14-3-3-Binding Motifs: A Structural and Interactomics Study. Structure.

Gógl, G., Tugaeva, K.V., Eberling, P., Kostmann, C., Travé, G., and Sluchanko, N.N. (2020b). Recognition of high-risk HPV E6 oncoproteins by 14-3-3 proteins studied by interactomics and crystallography. BioRxiv 2020.07.24.220376.

Grace, M., and Münger, K. (2017). Proteomic analysis of the gamma human papillomavirus type 197 E6 and E7 associated cellular proteins. Virology *500*, 71–81.

Graham, S.V. (2017). The human papillomavirus replication cycle, and its links to cancer progression: a comprehensive review. Clinical Science *131*, 2201–2221.

Guan, J., Bywaters, S.M., Brendle, S.A., Ashley, R.E., Makhov, A.M., Conway, J.F., Christensen, N.D., and Hafenstein, S. (2017). Cryoelectron Microscopy Maps of Human Papillomavirus 16 Reveal L2 Densities and Heparin Binding Site. Structure *25*, 253–263.

Gueneau, E., and Dufour, G. (2019). Exploring molecular interactions in Biacore systems (Boulogne-Billancourt).

Guiley, K.Z., Liban, T.J., Felthousen, J.G., Ramanan, P., Litovchick, L., and Rubin, S.M. (2015). Structural mechanisms of DREAM complex assembly and regulation. Genes Dev. *29*, 961–974.

Gupta, S., Takhar, P.P.S., Degenkolbe, R., Heng Koh, C., Zimmermann, H., Maolin Yang, C., Guan Sim, K., I-Hong Hsu, S., and Bernard, H.-U. (2003). The human papillomavirus type 11 and 16 E6 proteins modulate the cell-cycle regulator and transcription cofactor TRIP-Br1. Virology *317*, 155–164.

Hagensee, M.E., Yaegashi, N., and Galloway, D.A. (1993). Self-assembly of human papillomavirus type 1 capsids by expression of the L1 protein alone or by coexpression of the L1 and L2 capsid proteins. Journal of Virology *67*, 315–322.

Hampras, S.S., Rollison, D.E., Giuliano, A.R., McKay-Chopin, S., Minoni, L.,

Sereday, K., Gheit, T., and Tommasino, M. (2017). Prevalence and Concordance of Cutaneous Beta Human Papillomavirus Infection at Mucosal and Cutaneous Sites. J Infect Dis *216*, 92–96.

Harden, M.E., and Munger, K. (2017). Human papillomavirus molecular biology. Mutation Research/Reviews in Mutation Research *772*, 3–12.

Harlalka, G.V., Baple, E.L., Cross, H., Kühnle, S., Cubillos-Rojas, M., Matentzoglu, K., Patton, M.A., Wagner, K., Coblentz, R., Ford, D.L., Mackay, D.J.G., Chioza, B.A., Scheffner, M., Rosa, J.L., and Crosby, A.H. (2013). Mutation of HERC2 causes developmental delay with Angelman-like features. Journal of Medical Genetics *50*, 65–73.

Harms, K.L., and Chen, X. (2006). The functional domains in p53 family proteins exhibit both common and distinct properties. Cell Death Differ *13*, 890–897.

Hatterschide, J., Bohidar, A.E., Grace, M., Nulton, T.J., Kim, H.W., Windle, B., Morgan, I.M., Munger, K., and White, E.A. (2019). PTPN14 degradation by high-risk human papillomavirus E7 limits keratinocyte differentiation and contributes to HPV-mediated oncogenesis. PNAS *116*, 7033–7042.

zur Hausen, H. (1991). Viruses in human cancers. Science *254*, 1167–1173.

zur Hausen, H. (2000). Papillomaviruses Causing Cancer: Evasion From Host-Cell Control in Early Events in Carcinogenesis. J Natl Cancer Inst *92*, 690–698.

zur Hausen, H. (2002). Papillomaviruses and cancer: from basic studies to clinical application. Nat Rev Cancer *2*, 342–350.

Heery, D.M., Hoare, S., Hussain, S., Parker, M.G., and Sheppard, H. (2001). Core L *XX* LL Motif Sequences in CREB-binding Protein, SRC1, and RIP140 Define Affinity and Selectivity for Steroid and Retinoid Receptors. Journal of Biological Chemistry *276*, 6695–6702.

Herfs, M., Yamamoto, Y., Laury, A., Wang, X., Nucci, M.R., McLaughlin-Drubin, M.E., Münger, K., Feldman, S., McKeon, F.D., Xian, W., and Crum, C.P. (2012). A discrete population of squamocolumnar junction cells implicated in the pathogenesis

of cervical cancer. PNAS *109*, 10516–10521.

Hill, B., Lam, S.F., Lane, P., MacAulay, C., Fradkin, L., and Follen, M. (2017). Established and Emerging Optical Technologies for the Real-Time Detection of Cervical Neoplasia: A Review. Journal of Cancer Therapy *08*, 1241.

Howie, H.L., Katzenellenbogen, R.A., and Galloway, D.A. (2009). Papillomavirus E6 proteins. Virology *384*, 324–334.

Howie, H.L., Koop, J.I., Weese, J., Robinson, K., Wipf, G., Kim, L., and Galloway, D.A. (2011). Beta-HPV 5 and 8 E6 promote p300 degradation by blocking AKT/p300 association. PLoS Pathog. *7*, e1002211.

Huh, K., Zhou, X., Hayakawa, H., Cho, J.-Y., Libermann, T.A., Jin, J., Harper, J.W., and Munger, K. (2007). Human Papillomavirus Type 16 E7 Oncoprotein Associates with the Cullin 2 Ubiquitin Ligase Complex, Which Contributes to Degradation of the Retinoblastoma Tumor Suppressor. Journal of Virology *81*, 9737–9747.

Huibregtse, J.M., Scheffner, M., and Howley, P.M. (1991). A cellular protein mediates association of p53 with the E6 oncoprotein of human papillomavirus types 16 or 18. The EMBO Journal *10*, 4129–4135.

Huibregtse, J.M., Scheffner, M., and Howley, P.M. (1993). Localization of the E6-AP regions that direct human papillomavirus E6 binding, association with p53, and ubiquitination of associated proteins. Molecular and Cellular Biology *13*, 4918–4927.

Huttlin, E.L., Bruckner, R.J., Navarrete-Perea, J., Cannon, J.R., Baltier, K., Gebreab, F., Gygi, M.P., Thornock, A., Zarraga, G., Tam, S., Szpyt, J., Panov, A., Parzen, H., Fu, S., Golbazi, A., Maenpaa, E., Stricker, K., Thakurta, S.G., Rad, R., Pan, J., Nusinow, D.P., Paulo, J.A., Schweppe, D.K., Vaites, L.P., Harper, J.W., and Gygi, S.P. (2020). Dual Proteome-scale Networks Reveal Cell-specific Remodeling of the Human Interactome. BioRxiv 2020.01.19.905109.

Iftner, T., Elbel, M., Schopp, B., Hiller, T., Loizou, J.I., Caldecott, K.W., and Stubenrauch, F. (2002). Interference of papillomavirus E6 protein with single-strand break repair by interaction with XRCC1. The EMBO Journal *21*, 4741–4748.

Jagadeeswaran, R., Surawska, H., Krishnaswamy, S., Janamanchi, V., Mackinnon, A.C., Seiwert, T.Y., Loganathan, S., Kanteti, R., Reichman, T., Nallasura, V., Schwartz, S., Faoro, L., Wang, Y.-C., Girard, L., Tretiakova, M.S., Ahmed, S., Zumba, O., Soulii, L., Bindokas, V.P., Szeto, L.L., Gordon, G.J., Bueno, R., Sugarbaker, D., Lingen, M.W., Sattler, M., Krausz, T., Vigneswaran, W., Natarajan, V., Minna, J., Vokes, E.E., Ferguson, M.K., Husain, A.N., and Salgia, R. (2008). Paxillin Is a Target for Somatic Mutations in Lung Cancer: Implications for Cell Growth and Invasion. Cancer Res *68*, 132–142.

James, C., and Roberts, S. (2016). Viral Interactions with PDZ Domain-Containing Proteins—An Oncogenic Trait? Pathogens *5*, 8.

Jerabek-Willemsen, M., André, T., Wanner, R., Roth, H.M., Duhr, S., Baaske, P., and Breitsprecher, D. (2014). MicroScale Thermophoresis: Interaction analysis and beyond. Journal of Molecular Structure *1077*, 101–113.

Johansson, C., and Schwartz, S. (2013). Regulation of human papillomavirus gene expression by splicing and polyadenylation. Nature Reviews Microbiology *11*, 239–251.

Kaelin, W.G. (1999). Functions of the retinoblastoma protein. BioEssays *21*, 950–958.

Kalsner, L., and Chamberlain, S.J. (2015). Prader-Willi, Angelman, and 15q11-q13 Duplication Syndromes. Pediatric Clinics of North America *62*, 587–606.

Kawai, T., and Malech, H.L. (2009). WHIM Syndrome: Congenital Immune Deficiency Disease. Curr Opin Hematol *16*, 20–26.

Kim, E., Niethammer, M., Rothschild, A., Nung Jan, Y., and Sheng, M. (1995). Clustering of Shaker-type K + channels by interaction with a family of membrane-associated guanylate kinases. Nature *378*, 85–88.

Kines, R.C., Thompson, C.D., Lowy, D.R., Schiller, J.T., and Day, P.M. (2009). The initial steps leading to papillomavirus infection occur on the basement membrane prior to cell surface binding. PNAS *106*, 20458–20463.

Kishino, T., Lalande, M., and Wagstaff, J. (1997). UBE3A/E6-AP mutations cause Angelman syndrome. Nature Genetics *15*, 70–73.

Kobe, B., Ve, T., and Williams, S.J. (2015). Fusion-protein-assisted protein crystallization. Acta Crystallogr F Struct Biol Commun *71*, 861–869.

Koller, L.D., and Olson, C. (1972). Attempted Transmission of Warts From Man, Cattle, and Horses and Deer Fibroma to Selected Hosts. Journal of Investigative Dermatology *58*, 366–368.

Kretschmann, E., and Raether, H. (1968). Notizen: Radiative Decay of Non Radiative Surface Plasmons Excited by Light. Zeitschrift Für Naturforschung A *23*, 2135–2136.

Krois, A.S., Ferreon, J.C., Martinez-Yamout, M.A., Dyson, H.J., and Wright, P.E. (2016). Recognition of the disordered p53 transactivation domain by the transcriptional adapter zinc finger domains of CREB-binding protein. PNAS *113*, E1853–E1862.

Kruiswijk, F., Labuschagne, C.F., and Vousden, K.H. (2015). p53 in survival, death and metabolic health: a lifeguard with a licence to kill. Nature Reviews Molecular Cell Biology *16*, 393–405.

Kühne, C., Gardiol, D., Guarnaccia, C., Amenitsch, H., and Banks, L. (2000). Differential regulation of human papillomavirus E6 by protein kinase A: conditional degradation of human discs large protein by oncogenic E6. Oncogene *19*, 5884–5891.

Kühnle, S., Kogel, U., Glockzin, S., Marquardt, A., Ciechanover, A., Matentzoglu, K., and Scheffner, M. (2011). Physical and Functional Interaction of the HECT Ubiquitin-protein Ligases E6AP and HERC2. Journal of Biological Chemistry *286*, 19410–19416.

Kumar, A., Zhao, Y., Meng, G., Zeng, M., Srinivasan, S., Delmolino, L.M., Gao, Q., Dimri, G., Weber, G.F., Wazer, D.E., Band, H., and Band, V. (2002). Human Papillomavirus Oncoprotein E6 Inactivates the Transcriptional Coactivator Human ADA3. Molecular and Cellular Biology *22*, 5801–5812.

Lancaster, W.D., and Olson, C. (1978). Demonstration of two distinct classes of

bovine papilloma virus. Virology *89*, 372–379.

Lee, C., and Laimins, L.A. (2004). Role of the PDZ Domain-Binding Motif of the Oncoprotein E6 in the Pathogenesis of Human Papillomavirus Type 31. Journal of Virology *78*, 12366–12377.

Lee, J.-O., Russo, A.A., and Pavletich, N.P. (1998). Structure of the retinoblastoma tumour-suppressor pocket domain bound to a peptide from HPV E7. Nature *391*, 859–865.

Lefèvre, H., Schrimpf, C., Moro, M.R., and Lachal, J. (2018). HPV vaccination rate in French adolescent girls: an example of vaccine distrust. Archives of Disease in Childhood *103*, 740–746.

Li, Z., Yan, X., Yu, H., Wang, D., Song, S., Li, Y., He, M., Hong, Q., Zheng, Q., Zhao, Q., Gu, Y., Zhang, J., Janssen, M.E.W., Cardone, G., Olson, N.H., Baker, T.S., Li, S., and Xia, N. (2016). The C-Terminal Arm of the Human Papillomavirus Major Capsid Protein Is Immunogenic and Involved in Virus-Host Interaction. Structure *24*, 874–885.

Liang, Y.-J., Chang, H.-S., Wang, C.-Y., and Yu, W.C.Y. (2008). DYRK1A stabilizes HPV16E7 oncoprotein through phosphorylation of the threonine 5 and threonine 7 residues. The International Journal of Biochemistry & Cell Biology *40*, 2431–2441.

Liu, S., Cai, X., Wu, J., Cong, Q., Chen, X., Li, T., Du, F., Ren, J., Wu, Y.-T., Grishin, N.V., and Chen, Z.J. (2015). Phosphorylation of innate immune adaptor proteins MAVS, STING, and TRIF induces IRF3 activation. Science *347*.

Liu, X., Clements, A., Zhao, K., and 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.

Liu, X., Yang, N., Figel, S.A., Wilson, K.E., Morrison, C.D., Gelman, I.H., and Zhang, J. (2013). PTPN14 interacts with and negatively regulates the oncogenic function of YAP. Oncogene *32*, 1266–1273.

Liu, Y., Liu, Z., Androphy, E., Chen, J., and Baleja, J.D. (2004). Design and Characterization of Helical Peptides that Inhibit the E6 Protein of Papillomavirus,. Biochemistry *43*, 7421–7431.

Lorenz, L.D., Rivera Cardona, J., and Lambert, P.F. (2013). Inactivation of p53 rescues the maintenance of high risk HPV DNA genomes deficient in expression of E6. PLoS Pathog. *9*, e1003717.

Lu, Z., Hu, X., Li, Y., Zheng, L., Zhou, Y., Jiang, H., Ning, T., Basang, Z., Zhang, C., and Ke, Y. (2004). Human Papillomavirus 16 E6 Oncoprotein Interferences with Insulin Signaling Pathway by Binding to Tuberin. J. Biol. Chem. *279*, 35664–35670.

Luck, K., Kim, D.-K., Lambourne, L., Spirohn, K., Begg, B.E., Bian, W., Brignall, R., Cafarelli, T., Campos-Laborie, F.J., Charloteaux, B., Choi, D., Coté, A.G., Daley, M., Deimling, S., Desbuleux, A., Dricot, A., Gebbia, M., Hardy, M.F., Kishore, N., Knapp, J.J., Kovács, I.A., Lemmens, I., Mee, M.W., Mellor, J.C., Pollis, C., Pons, C., Richardson, A.D., Schlabach, S., Teeking, B., Yadav, A., Babor, M., Balcha, D., Basha, O., Bowman-Colin, C., Chin, S.-F., Choi, S.G., Colabella, C., Coppin, G., D'Amata, C., De Ridder, D., De Rouck, S., Duran-Frigola, M., Ennajdaoui, H., Goebels, F., Goehring, L., Gopal, A., Haddad, G., Hatchi, E., Helmy, M., Jacob, Y., Kassa, Y., Landini, S., Li, R., van Lieshout, N., MacWilliams, A., Markey, D., Paulson, J.N., Rangarajan, S., Rasla, J., Rayhan, A., Rolland, T., San-Miguel, A., Shen, Y., Sheykhkarimli, D., Sheynkman, G.M., Simonovsky, E., Taşan, M., Tejeda, A., Tropepe, V., Twizere, J.-C., Wang, Y., Weatheritt, R.J., Weile, J., Xia, Y., Yang, X., Yeger-Lotem, E., Zhong, Q., Aloy, P., Bader, G.D., De Las Rivas, J., Gaudet, S., Hao, T., Rak, J., Tavernier, J., Hill, D.E., Vidal, M., Roth, F.P., and Calderwood, M.A. (2020). A reference map of the human binary protein interactome. Nature *580*, 402–408.

Ludwig, C., K.K. Hof- und Staatsdruckerei (Austria), Kaiserl. Akademie der Wissenschaften in Wien, and Braumüller (Firm) (1856). Diffusion zwischen ungleich erwärmten Orten gleich zusammengesetzter Lösung (Wien: Aus der K.K. Hof- und Staatsdruckerei, in Commission bei W. Braumüller, Buchhändler des K.K. Hofes und der K. Akademie der Wissenschaften).

Machalek, D.A., Garland, S.M., Brotherton, J.M.L., Bateson, D., McNamee, K., Stewart, M., Rachel Skinner, S., Liu, B., Cornall, A.M., Kaldor, J.M., and Tabrizi, S.N.

(2018). Very Low Prevalence of Vaccine Human Papillomavirus Types Among 18- to 35-Year Old Australian Women 9 Years Following Implementation of Vaccination. J Infect Dis *217*, 1590–1600.

Malecka, K.A., Fera, D., Schultz, D.C., Hodawadekar, S., Reichman, M., Donover, P.S., Murphy, M.E., and Marmorstein, R. (2014). Identification and Characterization of Small Molecule Human Papillomavirus E6 Inhibitors. ACS Chem. Biol. *9*, 1603–1612.

Mancilla, V.T., and Volkmer, R. (2016). Peptide Arrays on Planar Supports. Methods Mol. Biol. *1352*, 3–17.

de Martel, C., Shiels, M.S., Franceschi, S., Simard, E.P., Vignat, J., Hall, H.I., Engels, E.A., and Plummer, M. (2015). Cancers attributable to infections among adults with HIV in the United States. AIDS *29*, 2173–2181.

de Martel, C., Plummer, M., Vignat, J., and Franceschi, S. (2017). Worldwide burden of cancer attributable to HPV by site, country and HPV type: Worldwide burden of cancer attributable to HPV. Int. J. Cancer *141*, 664–670.

de Martel, C., Georges, D., Bray, F., Ferlay, J., and Clifford, G.M. (2020). Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. The Lancet Global Health *8*, e180–e190.

Martin, C.M., and O'Leary, J.J. (2011). Histology of cervical intraepithelial neoplasia and the role of biomarkers. Best Practice & Research Clinical Obstetrics & Gynaecology *25*, 605–615.

Martin, K., Trouche, D., Hagemeier, C., Sorensen, T.S., La Thangue, N.B., and Kouzarides, T. (1995). Stimulation of E2F1/DP1 transcriptional activity by MDM2 oncoprotein. Nature *375*, 691–694.

Martínez-Noël, G., Luck, K., Kühnle, S., Desbuleux, A., Szajner, P., Galligan, J.T., Rodriguez, D., Zheng, L., Boyland, K., Leclere, F., Zhong, Q., Hill, D.E., Vidal, M., and Howley, P.M. (2018). Network Analysis of UBE3A/E6AP-Associated Proteins Provides Connections to Several Distinct Cellular Processes. Journal of Molecular Biology *430*, 1024–1050.

Martinez-Zapien, D., Ruiz, F.X., Poirson, J., Mitschler, A., Ramirez, J., Forster, A., Cousido-Siah, A., Masson, M., Pol, S.V., Podjarny, A., Travé, G., and Zanier, K. (2016). Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53. Nature *529*, 541–545.

Maupin-Furlow, J. (2012). Proteasomes and protein conjugation across domains of life. Nat Rev Microbiol *10*, 100–111.

McBride, A.A. (2008). Chapter 4 Replication and Partitioning of Papillomavirus Genomes. In Advances in Virus Research, (Academic Press), pp. 155–205.

McBride, A.A. (2017). Oncogenic human papillomaviruses. Philosophical Transactions of the Royal Society B: Biological Sciences *372*, 20160273.

McBride, A.A., and Warburton, A. (2017). The role of integration in oncogenic progression of HPV-associated cancers. PLOS Pathogens *13*, e1006211.

Mellin, H., Dahlgren, L., Munck-Wikland, E., Lindholm, J., Rabbani, H., Kalantari, M., and Dalianis, T. (2002). Human papillomavirus type 16 is episomal and a high viral load may be correlated to better prognosis in tonsillar cancer. International Journal of Cancer *102*, 152–158.

Mesplède, T., Gagnon, D., Bergeron-Labrecque, F., Azar, I., Sénéchal, H., Coutlée, F., and Archambault, J. (2012). p53 Degradation Activity, Expression, and Subcellular Localization of E6 Proteins from 29 Human Papillomavirus Genotypes. Journal of Virology *86*, 94–107.

Molet, L. (2018). Génotypage des papillomavirus humains par séquençage haut-débit : conséquences dans le dépistage du cancer du col de l'utérus et apport conceptuel au virome cutané. phdthesis. Université Paris-Saclay.

Moody, C.A., and Laimins, L.A. (2010). Human papillomavirus oncoproteins: pathways to transformation. Nat Rev Cancer *10*, 550–560.

Muench, P., Probst, S., Schuetz, J., Leiprecht, N., Busch, M., Wesselborg, S., Stubenrauch, F., and Iftner, T. (2010). Cutaneous Papillomavirus E6 Proteins Must Interact with p300 and Block p53-Mediated Apoptosis for Cellular Immortalization and

Tumorigenesis. Cancer Res *70*, 6913–6924.

Munday, J.S. (2014). Bovine and Human Papillomaviruses: A Comparative Review. Vet Pathol *51*, 1063–1075.

Munday, J.S., Hanlon, E.M., Howe, L., Squires, R.A., and French, A.F. (2007). Feline Cutaneous Viral Papilloma Associated with Human Papillomavirus Type 9. Vet Pathol *44*, 924–927.

Munday, J.S., Sharp, C.R., and Beatty, J.A. (2019). Novel viruses: Update on the significance of papillomavirus infections in cats. Journal of Feline Medicine and Surgery *21*, 409–418.

Münger, K., and Howley, P.M. (2002). Human papillomavirus immortalization and transformation functions. Virus Research *89*, 213–228.

Münger, K., Basile, J.R., Duensing, S., Eichten, A., Gonzalez, S.L., Grace, M., and Zacny, V.L. (2001). Biological activities and molecular targets of the human papillomavirus E7 oncoprotein. Oncogene *20*, 7888–7898.

Muñoz, N., Castellsagué, X., de González, A.B., and Gissmann, L. (2006). Chapter 1: HPV in the etiology of human cancer. Vaccine *24*, S1–S10.

Nakagawa, S., and Huibregtse, J.M. (2000). Human scribble (Vartul) is targeted for ubiquitin-mediated degradation by the high-risk papillomavirus E6 proteins and the E6AP ubiquitin-protein ligase. Mol. Cell. Biol. *20*, 8244–8253.

Nogueira, M.O., Hošek, T., Calçada, E.O., Castiglia, F., Massimi, P., Banks, L., Felli, I.C., and Pierattelli, R. (2017). Monitoring HPV-16 E7 phosphorylation events. Virology *503*, 70–75.

Nominé, Y., Ristriani, T., Laurent, C., Lefèvre, J.-F., Weiss, É., and Travé, G. (2001). Formation of Soluble Inclusion Bodies by HPV E6 Oncoprotein Fused to Maltose-Binding Protein. Protein Expression and Purification *23*, 22–32.

Nominé, Y., Masson, M., Charbonnier, S., Zanier, K., Ristriani, T., Deryckère, F., Sibler, A.-P., Desplancq, D., Atkinson, R.A., Weiss, E., Orfanoudakis, G., Kieffer, B., and Travé, G. (2006). Structural and Functional Analysis of E6 Oncoprotein:

Insights in the Molecular Pathways of Human Papillomavirus-Mediated Pathogenesis. Molecular Cell *21*, 665–678.

Ohlenschläger, O., Seiboth, T., Zengerling, H., Briese, L., Marchanka, A., Ramachandran, R., Baum, M., Korbas, M., Meyer-Klaucke, W., Dürst, M., and Görlach, M. (2006). Solution structure of the partially folded high-risk human papilloma virus 45 oncoprotein E7. Oncogene *25*, 5953–5959.

Oriel, J.D. (1971). Natural history of genital warts. Sexually Transmitted Infections *47*, 1–13.

Orth, G. (2010). Génétique et sensibilité aux papillomavirus: le modèle de l'épidermodysplasie verruciforme. Bulletin de l'Académie Nationale de Médecine *194*, 923–941.

Orth, G., Favre, M., and Croissant, O. (1977). Characterization of a New Type of Human Papillomavirus That Causes Skin Warts. Journal of Virology *24*, 108–120.

Orth, G., Jablonska, S., Favre, M., Croissant, O., Jarzabek-Chorzelska, M., and Rzesa, G. (1978). Characterization of two types of human papillomaviruses in lesions of epidermodysplasia verruciformis. Proceedings of the National Academy of Sciences *75*, 1537–1541.

Ould M'hamed Ould Sidi, A., Ould Babah, K., Brimer, N., Nominé, Y., Romier, C., Kieffer, B., Vande Pol, S., Travé, G., and Zanier, K. (2011). Strategies for bacterial expression of protein–peptide complexes: Application to solubilization of papillomavirus E6. Protein Expression and Purification *80*, 8–16.

Paggi, M.G., Baldi, A., Bonetto, F., and Giordano, A. (1996). Retinoblastoma protein family in cell cycle and cancer: A review. Journal of Cellular Biochemistry *62*, 418–430.

Pamukcu, A.M., Brobst, D.F., Satter, E.J., and PRICEr, J.M. (1959). A Urinary Bladder Tumor Induced by a Bovine Cutaneous Papilloma Agent. Cancer Research *19*, 6.

Patel, D., Huang, S.-M., Baglia, L.A., and McCance, D.J. (1999). The E6 protein

of human papillomavirus type 16 binds to and inhibits co-activation by CBP and p300. The EMBO Journal *18*, 5061–5072.

Phelps, W.C., Yee, C.L., Münger, K., and Howley, P.M. (1988). The human papillomavirus type 16 E7 gene encodes transactivation and transformation functions similar to those of adenovirus E1A. Cell *53*, 539–547.

Pierce Campbell, C.M., Lin, H.-Y., Fulp, W., Papenfuss, M.R., Salmerón, J.J., Quiterio, M.M., Lazcano-Ponce, E., Villa, L.L., and Giuliano, A.R. (2013). Consistent Condom Use Reduces the Genital Human Papillomavirus Burden Among High-Risk Men: The HPV Infection in Men Study. The Journal of Infectious Diseases *208*, 373–384.

Plevin, M.J., Mills, M.M., and Ikura, M. (2005). The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. Trends in Biochemical Sciences *30*, 66–69.

Poirson, J. (2016). Interactome des oncoprotéines E6 et E7 des HPV : du système ubiquitine-protéasome à la voie de signalisation Hippo. Thèse de doctorat. Université de Strasbourg.

Poirson, J., Biquand, E., Straub, M.-L., Cassonnet, P., Nominé, Y., Jones, L., van der Werf, S., Travé, G., Zanier, K., Jacob, Y., Demeret, C., and Masson, M. (2017). Mapping the interactome of HPV E6 and E7 oncoproteins with the ubiquitin-proteasome system. The FEBS Journal *284*, 3171–3201.

Qin, B.Y., Liu, C., Srinath, H., Lam, S.S., Correia, J.J., Derynck, R., and Lin, K. (2005). Crystal Structure of IRF-3 in Complex with CBP. Structure *13*, 1269–1277.

Raff, A.B., Woodham, A.W., Raff, L.M., Skeate, J.G., Yan, L., Silva, D.M.D., Schelhaas, M., and Kast, W.M. (2013). The Evolving Field of Human Papillomavirus Receptor Research: a Review of Binding and Entry. Journal of Virology *87*, 6062–6072.

Ramesh, V. (2019). Biomolecular and Bioanalytical Techniques: Theory, Methodology and Applications (John Wiley & Sons).

Ramirez, J., Poirson, J., Foltz, C., Chebaro, Y., Schrapp, M., Meyer, A., Bonetta, A., Forster, A., Jacob, Y., Masson, M., Deryckère, F., and Travé, G. (2015). Targeting the Two Oncogenic Functional Sites of the HPV E6 Oncoprotein with a High-Affinity Bivalent Ligand. Angewandte Chemie International Edition *54*, 7958–7962.

Raran-Kurussi, S., Keefe, K., and Waugh, D.S. (2015). Positional effects of fusion partners on the yield and solubility of MBP fusion proteins. Protein Expression and Purification *110*, 159–164.

Rector, A., and Van Ranst, M. (2013). Animal papillomaviruses. Virology *445*, 213–223.

Richards, R.M., Lowy, D.R., Schiller, J.T., and Day, P.M. (2006). Cleavage of the papillomavirus minor capsid protein, L2, at a furin consensus site is necessary for infection. PNAS *103*, 1522–1527.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Séraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. Nature Biotechnology *17*, 1030–1032.

Roberts, J.N., Buck, C.B., Thompson, C.D., Kines, R., Bernardo, M., Choyke, P.L., Lowy, D.R., and Schiller, J.T. (2007). Genital transmission of HPV in a mouse model is potentiated by nonoxynol-9 and inhibited by carrageenan. Nature Medicine *13*, 857–861.

Ronco, G., Giorgi-Rossi, P., Carozzi, F., Confortini, M., Palma, P.D., Del Mistro, A., Ghiringhello, B., Girlando, S., Gillio-Tos, A., De Marco, L., Naldoni, C., Pierotti, P., Rizzolo, R., Schincaglia, P., Zorzi, M., Zappa, M., Segnan, N., and Cuzick, J. (2010). Efficacy of human papillomavirus testing for the detection of invasive cervical cancers and cervical intraepithelial neoplasia: a randomised controlled trial. The Lancet Oncology *11*, 249–257.

Ronco, L.V., Karpova, A.Y., Vidal, M., and Howley, P.M. (1998). Human papillomavirus 16 E6 oncoprotein binds to interferon regulatory factor-3 and inhibits its transcriptional activity. Genes & Development *12*, 2061–2072.

Rous, P., and Beard, J.W. (1935). The Progression to Carcinoma of Virus-

Induced Rabbit Papillomas (Shope). The Journal of Experimental Medicine *62*, 523–548.

Rozenblatt-Rosen, O., Deo, R.C., Padi, M., Adelmant, G., Calderwood, M.A., Rolland, T., Grace, M., Dricot, A., Askenazi, M., Tavares, M., Pevzner, S.J., Abderazzaq, F., Byrdsong, D., Carvunis, A.-R., Chen, A.A., Cheng, J., Correll, M., Duarte, M., Fan, C., Feltkamp, M.C., Ficarro, S.B., Franchi, R., Garg, B.K., Gulbahce, N., Hao, T., Holthaus, A.M., James, R., Korkhin, A., Litovchick, L., Mar, J.C., Pak, T.R., Rabello, S., Rubio, R., Shen, Y., Singh, S., Spangle, J.M., Tasan, M., Wanamaker, S., Webber, J.T., Roecklein-Canfield, J., Johannsen, E., Barabási, A.-L., Beroukhim, R., Kieff, E., Cusick, M.E., Hill, D.E., Münger, K., Marto, J.A., Quackenbush, J., Roth, F.P., DeCaprio, J.A., and Vidal, M. (2012). Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. Nature *487*, 491–495.

Saidu, N.E.B., Filić, V., Thomas, M., Sarabia-Vega, V., Đukić, A., Miljković, F., Banks, L., and Tomaić, V. (2019). PDZ Domain-Containing Protein NHERF-2 Is a Novel Target of Human Papillomavirus 16 (HPV-16) and HPV-18. Journal of Virology *94*.

Scheffner, M., Werness, B.A., Huibregtse, J.M., Levine, A.J., and Howley, P.M. (1990). The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. Cell *63*, 1129–1136.

Scheffner, M., Huibregtse, J.M., Vierstra, R.D., and Howley, P.M. (1993). The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. Cell *75*, 495–505.

Schiffman, M., Doorbar, J., Wentzensen, N., de Sanjosé, S., Fakhry, C., Monk, B.J., Stanley, M.A., and Franceschi, S. (2016). Carcinogenic human papillomavirus infection. Nat Rev Dis Primers *2*, 16086.

Schmitz, M., Driesch, C., Jansen, L., Runnebaum, I.B., and Dürst, M. (2012). Non-Random Integration of the HPV Genome in Cervical Cancer. PLoS One *7*.

Schuck, P., and Zhao, H. (2010). The Role of Mass Transport Limitation and Surface Heterogeneity in the Biophysical Characterization of Macromolecular Binding Processes by SPR Biosensing. Methods Mol Biol *627*, 15–54.

Serrano, B., de Sanjosé, S., Tous, S., Quiros, B., Muñoz, N., Bosch, X., and Alemany, L. (2015). Human papillomavirus genotype attribution for HPVs 6, 11, 16, 18, 31, 33, 45, 52 and 58 in female anogenital lesions. European Journal of Cancer *51*, 1732–1741.

Sharma, S., Poetz, F., Bruer, M., Ly-Hartig, T.B.N., Schott, J., Séraphin, B., and Stoecklin, G. (2016). Acetylation-Dependent Control of Global Poly(A) RNA Degradation by CBP/p300 and HDAC1/2. Molecular Cell *63*, 927–938.

Sheng, M., and Sala, C. (2001). PDZ Domains and the Organization of Supramolecular Complexes. Annual Review of Neuroscience *24*, 1–29.

Shope, R.E., and Hurst, E.W. (1933). Infectious papillomatosis of Rabbits with a note on the histopathology. J Exp Med. *58*, 607–624.

Slebos, R.J., Lee, M.H., Plunkett, B.S., Kessis, T.D., Williams, B.O., Jacks, T., Hedrick, L., Kastan, M.B., and Cho, K.R. (1994). p53-dependent G1 arrest involves pRB-related proteins and is disrupted by the human papillomavirus 16 E7 oncoprotein. PNAS *91*, 5320–5324.

Smith, J.L., Campos, S.K., and Ozbun, M.A. (2007). Human Papillomavirus Type 31 Uses a Caveolin 1- and Dynamin 2-Mediated Entry Pathway for Infection of Human Keratinocytes. Journal of Virology *81*, 9922–9931.

Solis, M., Goubau, D., Romieu-Mourez, R., Genin, P., Civas, A., and Hiscott, J. (2006). Distinct functions of IRF-3 and IRF-7 in IFN-alpha gene regulation and control of anti-tumor activity in primary macrophages. Biochemical Pharmacology *72*, 1469–1476.

Song, C., Zhang, S., and Huang, H. (2015). Choosing a suitable method for the identification of replication origins in microbial genomes. Front Microbiol *6*.

Songyang, Z., Fanning, A.S., Fu, C., Xu, J., Marfatia, S.M., Chishti, A.H., Crompton, A., Chan, A.C., Anderson, J.M., and Cantley, L.C. (1997). Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. Science *275*, 73–77.

Sparks, R.P., and Fratti, R. (2019). Use of Microscale Thermophoresis (MST)

to Measure Binding Affinities of Components of the Fusion Machinery. In SNAREs: Methods and Protocols, R. Fratti, ed. (New York, NY: Springer), pp. 191–198.

Strauss, M.J., Bunting, H., and Melnick, J.L. (1950). Virus-Like Particles and Inclusion Bodies in Skin Papillomas. Journal of Investigative Dermatology *15*, 433–444.

Suarez, I.P. (2017). Novel E6 structures from HPV18 and HPV49: understanding cellular target specificity (Lyon, France).

Suarez, I., and Travé, G. (2018). Structural Insights in Multifunctional Papillomavirus Oncoproteins. Viruses *10*, 37.

Tan, M.J.A., White, E.A., Sowa, M.E., Harper, J.W., Aster, J.C., and Howley, P.M. (2012). Cutaneous -human papillomavirus E6 proteins bind Mastermind-like coactivators and repress Notch signaling. Proceedings of the National Academy of Sciences *109*, E1473–E1480.

Tanaka, Y., and Chen, Z.J. (2012). STING Specifies IRF3 Phosphorylation by TBK1 in the Cytosolic DNA Signaling Pathway. Sci. Signal. *5*, ra20–ra20.

Taylor, S., and Haldorson, G. (2013). A review of equine sarcoid. Equine Veterinary Education *25*, 210–216.

Teufel, D.P., Bycroft, M., and Fersht, A.R. (2009). Regulation by phosphorylation of the relative affinities of the N-terminal transactivation domains of p53 for p300 domains and Mdm2. Oncogene *28*, 2112–2118.

Thomas, M., Myers, M.P., Massimi, P., Guarnaccia, C., and Banks, L. (2016). Analysis of multiple HPV E6 PDZ interactions defines type-specific PDZ fingerprints that predict oncogenic potential. PLoS Pathogens *12*, e1005766.

Tong, X., and Howley, P.M. (1997). The bovine papillomavirus E6 oncoprotein interacts with paxillin and disrupts the actin cytoskeleton. PNAS *94*, 4412–4417.

Tong, X., Boll, W., Kirchhausen, T., and Howley, P.M. (1998). Interaction of the Bovine Papillomavirus E6 Protein with the Clathrin Adaptor Complex AP-1. Journal of Virology *72*, 476–482.

Turner, C.E. (2000). Paxillin and focal adhesion signalling. Nature Cell Biology *2*, E231–E236.

Uversky, V.N., Roman, A., Oldfield, C.J., and Dunker, A.K. (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.

Van Doorslaer, K., and Dillner, J. (2019). The Launch of an International Animal Papillomavirus Reference Center. Viruses *11*, 55.

Vande Pol, S.B., Brown, M.C., and Turner, C.E. (1998). Association of Bovine Papillomavirus Type 1 E6 oncoprotein with the focal adhesion protein paxillin through a conserved protein interaction motif. Oncogene *16*, 43–52.

Van Doorslaer, K., Li, Z., Xirasagar, S., Maes, P., Kaminsky, D., Liou, D., Sun, Q., Kaur, R., Huyen, Y., and McBride, A.A. (2017). The Papillomavirus Episteme: a major update to the papillomavirus sequence database. Nucleic Acids Research *45*, D499–D506.

Venkatesan, K., Rual, J.-F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Zenkner, M., Xin, X., Goh, K.-I., Yildirim, M.A., Simonis, N., Heinzmann, K., Gebreab, F., Sahalie, J.M., Cevik, S., Simon, C., de Smet, A.-S., Dann, E., Smolyar, A., Vinayagam, A., Yu, H., Szeto, D., Borick, H., Dricot, A., Klitgord, N., Murray, R.R., Lin, C., Lalowski, M., Timm, J., Rau, K., Boone, C., Braun, P., Cusick, M.E., Roth, F.P., Hill, D.E., Tavernier, J., Wanker, E.E., Barabási, A.-L., and Vidal, M. (2009). An empirical framework for binary interactome mapping. Nature Methods *6*, 83–90.

Viarisio, D., Müller-Decker, K., Zanna, P., Kloz, U., Aengeneyndt, B., Accardi, R., Flechtenmacher, C., Gissmann, L., and Tommasino, M. (2016). Novel ß-HPV49 Transgenic Mouse Model of Upper Digestive Tract Cancer. Cancer Research *76*, 4216–4225.

Viarisio, D., Müller-Decker, K., Accardi, R., Robitaille, A., Dürst, M., Beer, K., Jansen, L., Flechtenmacher, C., Bozza, M., Harbottle, R., Voegele, C., Ardin, M., Zavadil, J., Caldeira, S., Gissmann, L., and Tommasino, M. (2018). Beta HPV38 oncoproteins act with a hit-and-run mechanism in ultraviolet radiation-induced skin

carcinogenesis in mice. PLOS Pathogens *14*, e1006783.

de Villiers, E.-M. (2013). Cross-roads in the classification of papillomaviruses. Virology *445*, 2–10.

de Villiers, E.-M., Fauquet, C., Broker, T.R., Bernard, H.-U., and zur Hausen, H. (2004). Classification of papillomaviruses. Virology *324*, 17–27.

Vincentelli, R., Luck, K., Poirson, J., Polanowska, J., Abdat, J., Blémont, M., Turchetto, J., Iv, F., Ricquier, K., Straub, M.-L., Forster, A., Cassonnet, P., Borg, J.-P., Jacob, Y., Masson, M., Nominé, Y., Reboul, J., Wolff, N., Charbonnier, S., and Travé, G. (2015). Quantifying domain-ligand affinities and specificities by high-throughput holdup assay. Nature Methods *12*, 787–793.

Wade, R., Brimer, N., and Pol, S.V. (2008). Transformation by Bovine Papillomavirus Type 1 E6 Requires Paxillin. Journal of Virology *82*, 5962–5966.

Wallberg, A.E., Pedersen, K., Lendahl, U., and Roeder, R.G. (2002). p300 and PCAF Act Cooperatively To Mediate Transcriptional Activation from Chromatin Templates by Notch Intracellular Domains In Vitro. Molecular and Cellular Biology *22*, 7812–7819.

Walling, E.B., Benzoni, N., Dornfeld, J., Bhandari, R., Sisk, B.A., Garbutt, J., and Colditz, G. (2016). Interventions to Improve HPV Vaccine Uptake: A Systematic Review. PEDIATRICS *138*, e20153863–e20153863.

Wang, J., Dupuis, C., Tyring, S.K., and Underbrink, M.P. (2016). Sterile α Motif Domain Containing 9 Is a Novel Cellular Interacting Partner to Low-Risk Type Human Papillomavirus E6 Proteins. PLOS ONE *11*, e0149859.

Wang, Z., Shen, D., Parsons, D.W., Bardelli, A., Sager, J., Szabo, S., Ptak, J., Silliman, N., Peters, B.A., Heijden, M.S. van der, Parmigiani, G., Yan, H., Wang, T.-L., Riggins, G., Powell, S.M., Willson, J.K.V., Markowitz, S., Kinzler, K.W., Vogelstein, B., and Velculescu, V.E. (2004). Mutational Analysis of the Tyrosine Phosphatome in Colorectal Cancers. Science *304*, 1164–1166.

Waugh, D.S. (2016). Crystal structures of MBP fusion proteins: Structures of

MBP Fusion Proteins. Protein Science *25*, 559–571.

Webb Strickland, S., Brimer, N., Lyons, C., and Vande Pol, S.B. (2018). Human Papillomavirus E6 interaction with cellular PDZ domain proteins modulates YAP nuclear localization. Virology *516*, 127–138.

Wentzensen, N., Vinokurova, S., and Doeberitz, M. von K. (2004). Systematic Review of Genomic Integration Sites of Human Papillomavirus Genomes in Epithelial Dysplasia and Invasive Cancer of the Female Lower Genital Tract. Cancer Res *64*, 3878–3884.

Werness, B.A., Levine, A.J., and Howley, P.M. (1990). Association of human papillomavirus types 16 and 18 E6 proteins with p53. Science *248*, 76–79.

White, E.A., Kramer, R.E., Tan, M.J.A., Hayes, S.D., Harper, J.W., and Howley, P.M. (2012a). Comprehensive Analysis of Host Cellular Interactions with Human Papillomavirus E6 Proteins Identifies New E6 Binding Partners and Reflects Viral Diversity. Journal of Virology *86*, 13174–13186.

White, E.A., Sowa, M.E., Tan, M.J.A., Jeudy, S., Hayes, S.D., Santha, S., Münger, K., Harper, J.W., and Howley, P.M. (2012b). Systematic identification of interactions between host cell proteins and E7 oncoproteins from diverse human papillomaviruses. PNAS *109*, E260–E267.

White, E.A., Walther, J., Javanbakht, H., and Howley, P.M. (2014). Genus Beta Human Papillomavirus E6 Proteins Vary in Their Effects on the Transactivation of p53 Target Genes. Journal of Virology *88*, 8201–8212.

White, E.A., Münger, K., and Howley, P.M. (2016). High-Risk Human Papillomavirus E7 Proteins Target PTPN14 for Degradation. MBio *7*.

Whyte, P., Williamson, N.M., and Harlow, E. (1989). Cellular targets for transformation by the adenovirus E1A proteins. Cell *56*, 67–75.

Willemsen, A., and Bravo, I.G. (2019). Origin and evolution of papillomavirus (onco)genes and genomes. Philos. Trans. R. Soc. Lond., B, Biol. Sci. *374*, 20180303.

Wilson, K.E., Li, Y.-W., Yang, N., Shen, H., Orillion, A.R., and Zhang, J. (2014).

PTPN14 Forms a Complex with Kibra and LATS1 Proteins and Negatively Regulates the YAP Oncogenic Function. J. Biol. Chem. *289*, 23693–23700.

Winer, R.L., Hughes, J.P., Feng, Q., O'Reilly, S., Kiviat, N.B., Holmes, K.K., and Koutsky, L.A. (2006). Condom Use and the Risk of Genital Human Papillomavirus Infection in Young Women. New England Journal of Medicine *354*, 2645–2654.

Winkler, G.S., Mulder, K.W., Bardwell, V.J., Kalkhoven, E., and Timmers, H.T.M. (2006). Human Ccr4-Not complex is a ligand-dependent repressor of nuclear receptor-mediated transcription. EMBO J *25*, 3089–3099.

Wolf, H., zur Hausen, H., Klein, G., Becker, V., Henle, G., and Henle, W. (1975). Attempts to detect virus-specific DNA sequences in human tumors: III. Epstein-Barr viral DNA in non-lymphoid nasopharyngeal carcinoma cells. Med Microbiol Immunol *161*, 15–21.

Woodham, A.W., Da Silva, D.M., Skeate, J.G., Raff, A.B., Ambroso, M.R., Brand, H.E., Isas, J.M., Langen, R., and Kast, W.M. (2012). The S100A10 Subunit of the Annexin A2 Heterotetramer Facilitates L2-Mediated Human Papillomavirus Infection. PLoS One *7*.

Wu, L., Aster, J.C., Blacklow, S.C., Lake, R., Artavanis-Tsakonas, S., and Griffin, J.D. (2000). MAML1, a human homologue of Drosophila mastermind, is a transcriptional co-activator for NOTCH receptors. Nature Genetics *26*, 484–489.

Yanofsky, V.R., Patel, R.V., and Goldenberg, G. (2012). Genital Warts. J Clin Aesthet Dermatol *5*, 25–36.

Yasumoto, S., Burkhardt, A.L., Doniger, J., and DiPaolo, J.A. (1986). Human papillomavirus type 16 DNA-induced malignant transformation of NIH 3T3 cells. Journal of Virology *57*, 572–577.

Yu, G.W., Rudiger, S., Veprintsev, D., Freund, S., Fernandez-Fernandez, M.R., and Fersht, A.R. (2006). The central region of HDM2 provides a second binding site for p53. Proceedings of the National Academy of Sciences *103*, 1227–1232.

Yun, H.-Y., Kim, M.W., Lee, H.S., Kim, W., Shin, J.H., Kim, H., Shin, H.-C., Park,

H., Oh, B.-H., Kim, W.K., Bae, K.-H., Lee, S.C., Lee, E.-W., Ku, B., and Kim, S.J. (2019). Structural basis for recognition of the tumor suppressor protein PTPN14 by the oncoprotein E7 of human papillomavirus. PLOS Biology *17*, e3000367.

Zanier, K., Charbonnier, S., Baltzinger, M., Nominé, Y., Altschuh, D., and Travé, G. (2005). Kinetic Analysis of the Interactions of Human Papillomavirus E6 Oncoproteins with the Ubiquitin Ligase E6AP Using Surface Plasmon Resonance. Journal of Molecular Biology *349*, 401–412.

Zanier, K., Ruhlmann, C., Melin, F., Masson, M., Ould M'hamed Ould Sidi, A., Bernard, X., Fischer, B., Brino, L., Ristriani, T., Rybin, V., Baltzinger, M., Vande Pol, S., Hellwig, P., Schultz, P., and Travé, G. (2010). E6 Proteins from Diverse Papillomaviruses Self-Associate Both In Vitro and In Vivo. Journal of Molecular Biology *396*, 90–104.

Zanier, K., ould M'hamed ould Sidi, A., Boulade-Ladame, C., Rybin, V., Chappelle, A., Atkinson, A., Kieffer, B., and Travé, G. (2012). Solution Structure Analysis of the HPV16 E6 Oncoprotein Reveals a Self-Association Mechanism Required for E6-Mediated Degradation of p53. Structure *20*, 604–617.

Zanier, K., Charbonnier, S., Sidi, A.O.M.O., McEwen, A.G., Ferrario, M.G., Poussin-Courmontagne, P., Cura, V., Brimer, N., Babah, K.O., Ansari, T., Muller, I., Stote, R.H., Cavarelli, J., Vande Pol, S., and Travé, G. (2013). Structural Basis for Hijacking of Cellular LxxLL Motifs by Papillomavirus E6 Oncoproteins. Science *339*, 694–698.

Zanier, K., Stutz, C., Kintscher, S., Reinz, E., Sehr, P., Bulkescher, J., Hoppe-Seyler, K., Travé, G., and Hoppe-Seyler, F. (2014). The E6AP Binding Pocket of the HPV16 E6 Oncoprotein Provides a Docking Site for a Small Inhibitory Peptide Unrelated to E6AP, Indicating Druggability of E6. PLoS ONE *9*, e112514.

Zeng, M., Kumar, A., Meng, G., Gao, Q., Dimri, G., Wazer, D., Band, H., and Band, V. (2002). Human Papilloma Virus 16 E6 Oncoprotein Inhibits Retinoic X Receptor-mediated Transactivation by Targeting Human ADA3 Coactivator. J. Biol. Chem. *277*, 45611–45618.

Zhang, Y., Fan, S., Meng, Q., Ma, Y., Katiyar, P., Schlegel, R., and Rosen, E.M.

(2005). BRCA1 Interaction with Human Papillomavirus Oncoproteins. J. Biol. Chem. *280*, 33165–33177.

Zhang, Y., Dasgupta, J., Ma, R.Z., Banks, L., Thomas, M., and Chen, X.S. (2007). Structures of a Human Papillomavirus (HPV) E6 Polypeptide Bound to MAGUK Proteins: Mechanisms of Targeting Tumor Suppressors by a High-Risk HPV Oncoprotein. Journal of Virology *81*, 3618–3626.

Zhao, C.Y., Szekely, L., Bao, W., and Selivanova, G. (2010). Rescue of p53 Function by Small-Molecule RITA in Cervical Carcinoma by Blocking E6-Mediated Degradation. Cancer Res *70*, 3372–3381.

Zhao, Y., Katzman, R.B., Delmolino, L.M., Bhat, I., Zhang, Y., Gurumurthy, C.B., Germaniuk-Kurowska, A., Reddi, H.V., Solomon, A., Zeng, M.-S., Kung, A., Ma, H., Gao, Q., Dimri, G., Stanculescu, A., Miele, L., Wu, L., Griffin, J.D., Wazer, D.E., Band, H., and Band, V. (2007). The Notch Regulator MAML1 Interacts with p53 and Functions as a Coactivator. J. Biol. Chem. *282*, 11969–11981.

Zheng, Z.-M., and Baker, C.C. (2006). Papillomavirus Genome Structure, Expression, And Post-Transcriptional Regulation. Front Biosci *11*, 2286–2302.

Zhou, J., Sun, X.Y., Louis, K., and Frazer, I.H. (1994). Interaction of human papillomavirus (HPV) type 16 capsid proteins with HPV DNA requires an intact L2 N-terminal sequence. Journal of Virology *68*, 619–625.

Zimmermann, H., Degenkolbe, R., Bernard, H.-U., and O'Connor, M.J. (1999). The Human Papillomavirus Type 16 E6 Oncoprotein Can Down-Regulate p53 Activity by Targeting the Transcriptional Coactivator CBP/p300. J. Virol. *73*, 6209–6219.

Zine El Abidine, A., Tomaić, V., Bel Haj Rhouma, R., Massimi, P., Guizani, I., Boubaker, S., Ennaifer, E., and Banks, L. (2017). A naturally occurring variant of HPV-16 E7 exerts increased transforming activity through acquisition of an additional phospho-acceptor site. Virology *500*, 218–225.

Zor, T., De Guzman, R.N., Dyson, H.J., and Wright, P.E. (2004). Solution Structure of the KIX Domain of CBP Bound to the Transactivation Domain of c-Myb. Journal of Molecular Biology *337*, 521–534.

Zuelzer, W.W., Evans, R.K., and Goodman, J. (1964). Myelokathexis — A New Form of Chronic Granulocytopenia. New England Journal of Medicine *270*, 699–704.

Université de Strasbourg

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

Anna BONHOURE

Interférence de l'oncoprotéine E6 de HPV avec l'interactome humain

Résumé

Les HPV sont les agents étiologiques du cancer du col de l'utérus, causant plus de 311 000 décès chaque année. Les HPV 16 et 18 sont à l'origine de 61 et 11 % des cancers du col. Les oncoprotéines E6 et E7 favorisent la réplication virale en stimulant la prolifération cellulaire. Chaque E6 produite par un HPV donné cible un certain ensemble de protéines hôtes qui détermine les effets pathologiques du virus. En particulier, E6 perturbe le fonctionnement de nombreuses protéines cellulaires en les capturant via des motifs de consensus "LXXLL". Le test chromatographique "holdup", développé par l'équipe et optimisé au cours de cette thèse, permet l'étude quantitative de nombreuses interactions protéines-motifs. Ainsi, nous avons déterminé les préférences d'interactions de protéines E6 de divers HPVs pour une banque de motifs LXXLL issus de protéines cellulaires. Les résultats obtenus ont permis de déterminer les structures de deux nouveaux complexes E6/LXXLL. Ces données permettent une meilleure compréhension de la carcinogenèse induite par certains HPV et ouvrent des opportunités pour le développement d'inhibiteurs spécifiques de E6.

Mots-clés: HPV, oncoprotéine E6, interaction domaine-motif, affinité, motif linéaire

Résumé en anglais

HPVs are the etiologic agents of cervical cancer, causing more than 311,000 deaths each year. HPV 16 and 18 cause 61 and 11% of cervical cancers. E6 and E7 oncoproteins promote viral replication by stimulating cell proliferation. Each E6 produced by a given HPV targets a certain set of host proteins that determines the pathological effects of the virus. In particular, E6 disrupts the function of many cellular proteins by capturing them via "LXXLL" consensus motifs. The "holdup" chromatographic test, developed by the team and optimized during this thesis, allows the quantitative study of numerous protein-motif interactions. Hence, we have determined the interaction preferences of a set of E6 proteins from various HPVs for a library of LXXLL motifs from target cell proteins. The results obtained allowed us to determine the structures of two new complexes for E6 with their prototypical LXXLL target motifs. These data allow a better understanding of the carcinogenesis induced by certain HPVs and pave the way for the development of specific E6 inhibitors.

Keywords: HPV, E6 oncoprotein, domain-motif interaction, affinity, linear motif