Structural insights into the interaction of a high-risk human papillomavirus wild-type oncoprotein E6

Dissertation
zur Erlangung des akademischen Grades doctor rerum naturalium
(Dr. rer. nat.)

vorgelegt dem Rat der Biologisch-Pharmazeutischen Fakultät
der Friedrich-Schiller-Universität Jena

von
Dipl. Biotechnologe André Mischo
geboren am 23.02.1982 in Bottrop (D)
„Dass ich erkenne, was die Welt im Innersten zusammenhält”

Goethes Faust
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>51Z2</td>
<td>C-terminal zinc binding domain of the oncoprotein E6 derived from HPV 51</td>
</tr>
<tr>
<td>AP</td>
<td>Alkaline phosphatase</td>
</tr>
<tr>
<td>APC</td>
<td>Adenomatous polyposis coli</td>
</tr>
<tr>
<td>BPV</td>
<td>Bovine papillomavirus</td>
</tr>
<tr>
<td>CD</td>
<td>Circular dichroism spectroscopy</td>
</tr>
<tr>
<td>CRPV</td>
<td>Cottontail rabbit papillomavirus</td>
</tr>
<tr>
<td>E6AP</td>
<td>Human papillomavirus E6-associated protein / Ubiquitin-protein ligase E3A</td>
</tr>
<tr>
<td>hDlg</td>
<td>Human Dlg-1</td>
</tr>
<tr>
<td>hDlgPDZ2</td>
<td>PDZ domain 2 of hDlg</td>
</tr>
<tr>
<td>HPV</td>
<td>Human papillomavirus</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>ISG</td>
<td>Interferon stimulated gene</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>NOE</td>
<td>Nuclear Overhauser Effect</td>
</tr>
<tr>
<td>PC</td>
<td>Proproteinconvertase</td>
</tr>
<tr>
<td>PDZ-BM</td>
<td>PDZ-binding motif</td>
</tr>
<tr>
<td>PDZ (domain)</td>
<td>Post synaptic density-95, Discs large, Zona occludens-1 domain</td>
</tr>
<tr>
<td>pGlu</td>
<td>Pyroglutamate (5-oxo-proline)</td>
</tr>
<tr>
<td>PV</td>
<td>Papillomavirus</td>
</tr>
<tr>
<td>SPR</td>
<td>Surface plasmon resonance</td>
</tr>
<tr>
<td>ZBD</td>
<td>Zinc binding domain</td>
</tr>
</tbody>
</table>

Standard abbreviations, e.g. for chemical names, follow recommendations of the *International Union of Pure and Applied Chemistry* and are not listed here.
# Table of contents

1 Introduction  
   *Opening remarks*  
   1  
1.1 Papillomaviruses  
   *A brief historical account*  
   1  
   *Papillomavirus taxonomy and phylogeny*  
   2  
1.2 Human papillomaviruses  
   *Epidemiology*  
   3  
   *Infection and vaccines*  
   4  
   *Genome organization*  
   6  
   *Productive life cycle and epithelial responses*  
   8  
   *The oncoproteins E6 and E7*  
   10  
1.3 The oncoprotein E6  
   *Regulation of E6 during viral life cycle*  
   11  
   *E6 interactions and consequences thereof*  
   11  
   *Biophysical properties of E6*  
   16  
1.4 Scope of work and publication summary  
   17  

2 Publications  
2.1 Publication 1  
   *Recombinant production of isotope-labeled peptides and spontaneous cyclization of amino-terminal glutamine into pyroglutamic acid*  
   19  
2.2 Publication 2  
   *NMR assignment of a PDZ domain in complex with a HPV51 E6 derived N-terminally pyroglutamic acid modified peptide*  
   27  
2.3 Publication 3  
   *Structural insights into a wildtype domain of the oncoprotein E6 and its interaction with a PDZ domain*  
   31  

3 Discussion  
3.1 Results summary  
3.2 Technical aspects  
   *Secondary structure of E6 protein constructs*  
   82  
   *Recombinant peptide generation and N-terminal pyroglutaminylation*  
   83
3.3 Solution structure of a wild-type E6 domain
   \( E6 \) proteins share a globally conserved structure \( 86 \)
   Local structural E6 differences do exist \( 87 \)
   Conformational plasticity of E6 proteins \( 88 \)
3.4 Comments on the E6-hDlg complex
   \( 51Z2 \) interacts with hDlgPDZ2, increasing affinity via an extended PDZ binding motif \( 90 \)
   Structural basis of the extended E6 PDZ-BM hDlgPDZ2 interaction \( 91 \)
   Functional implications of the E6-hDlg interaction \( 93 \)
   Mechanisms of high affinity PDZ-BM PDZ domain interactions \( 94 \)
3.5 More E6 interactions?
   Additional PDZ domains possibly targeted by E6 \( 98 \)
   E6 and different interaction motifs \( 99 \)

4 Summary \( 101 \)

5 Zusammenfassung \( 102 \)

6 References \( 103 \)

7 Acknowledgements – Danksagung \( 115 \)

8 Curriculum vitae \( 116 \)

9 Ehrenwörtliche Erklärung \( 119 \)
1 Introduction

Opening remarks

In the dissertation at hand, the human papillomavirus oncoprotein E6 is investigated with a focus on E6 structure and interactions. The introduction (sections 1.1 to 1.3) should provide a detailed picture of this virus family and recapitulate facts that are not or only in passing covered by the publications included (sections 2.1 to 2.3). These publications represent the main focus of my research at the Fritz Lipmann Institute, Jena. One additional publication in Angewandte Chemie (Tietze, 2012) to which I contributed addresses a different topic and was therefore not included here. In the final discussion, I briefly summarize the content of each of the included publications, discuss their combined implications and draw conclusions on the structure-function relationship of the oncoprotein E6.

1.1 Papillomaviruses

A brief historical account

Already in 1842, it was postulated that cervical cancer is a sexually transmitted disease (Rigoni-Stern, 1842). In 1933 Richard Shope demonstrated that an isolated animal virus was responsible for the formation of warts in cotton tail and domestic rabbits (Shope, 1933). This virus was subsequently termed Cottontail Rabbit PapillomaVirus (CRPV). Two years later, CRPV was identified to be causative for skin cancer in infected domestic rabbits by Peyton Rous (Rous, 1935). This was the first evidence that a virus causes cancer and consequently Peyton Rous was awarded the nobel prize “for his discovery of tumor-inducing viruses” in 1966 (www.nobelprize.org). In 1965 it was revealed that human papillomaviruses (HPVs), frequently encountered in human warts, are double stranded DNA viruses (Crawford, 1965; Klug 1965) and in subsequent years, it was speculated that HPVs might be causative for cervical cancer (e.g. zur Hausen, 1977). Only in the early 1980’s, HPVs were detected in cervical cancer biopsies (Gissmann, 1983; Dürst, 1983; Boshart, 1984), a finding supporting the oncovirus theory. Suffice it to say that in 2008 Harald zur Hausen was awarded the nobel prize “for his discovery of human papillomaviruses causing cervical cancer” (www.nobelprize.org).

The following sections present recent research on HPV with particular focus on its carcinogenicity.
1.1 Papillomaviruses

_Papillomavirus taxonomy and phylogeny_

In order to classify the large and increasing number of papillomavirus strains (PVs), including HPVs, the taxonomic system presented below is widely accepted. A complete PV genome is considered to be a new type when the DNA sequence of the L1-encoding region (the major PV capsid protein) differs by more than 10% from the closest known type (de Villiers, 2004, Bernard, 2010). A phylogenetic tree based on the L1 sequence is constructed and the new type is placed within a species (the “lower-order phylogenetic assembly”) and within a genus (the “higher-order phylogenetic assembly”) (de Villiers, 2004). For example, HPV 51 is classified as α5, implying that it resides in species 5 of the genus alpha-PV in the _papillomaviridae_ family (Figure 1; Bernard, 2010). All viruses with higher L1 sequence conservation are referred to as subtypes (90-98% identity) or variants (98-99% identity) of existing types (Hazard, 2007).

The _papillomaviridae_ family comprises at least 241 virus types according to the PV database “Papillomavirus Episteme” (http://pave.niaid.nih.gov/#home). The majority of these entries are HPVs with, as of February 2013, 144 reference genomes listed in the PV database.

*Figure 1:* Phylogenetic tree of 189 papillomavirus types (Bernard, 2010). PV genera, species and types are indicated.
1.2 Human papillomaviruses

Epidemiology

Certain types grouped in the alpha PV genus infect keratinocytes of the genital tract. Virtually all cervical cancer cases are attributed to the presence of alpha HPVs (Parkin, 2006). 529,000 new cases of cervical cancer and 274,000 cervical cancer-related deaths have been reported for 2008 (Ferlay, 2010). It should be noted however, that the vast majority of alpha HPV infections does not result in cancer (Gravitt, 2011). Therefore, these HPV infections favor malignant transformation, but supplementing factors appear necessary for cancer development (Münger, 2004). This is in line with the conceptual view of cancerogenesis as a multistep process (Hanahan, 2000) in which some but not all of its “features” are provided by HPV oncogenic molecules. HPV 16 and 18 are responsible for approximately 55% and 16%, respectively, of all cervical cancers (Smith, 2007). These and several other types frequently detected within cervical cancer biopsies are, hence, termed high-risk types. In recent years further types were considered high-risk, since the database on cervical cancer increases steadily and thus the occurrence of rare types can be more reliably evaluated. The larger data set also allows for assessment whether rare HPV types per se do contribute to cancer formation or whether co-infection of these types e.g. with HPV 16 and 18 feigned oncogenicity (IARC, 2012). Currently, HPV types from the genus alpha papillomavirus species 5, 6, 7, 9 and 11 can be considered as high-risk (IARC, 2012). Within a given species the riskiness of the individual types varies (reviewed in Bernard, 2010) and the IARC classifies each HPV type to be carcinogenic (Group 1), probably carcinogenic (Group 2A), possibly carcinogenic (Group 2B) or not carcinogenic (Group 3) to humans (IARC, 2012). Alpha-PVs infecting the genital tract but not or rarely associated with cancer formation (Group 3) are termed low-risk, e.g. types of alpha-PV species 1, 3 and 10 (including the HPV types 6 and 11) (Doorbar, 2006). Among the alpha-PVs, there are types that do not infect mucosa at all but instead cutaneous epithelium, e.g. alpha-PV species 2, 4 and 8 (Doorbar, 2006), which is common for HPVs of other PV genera (de Villiers, 2004).

In addition to cervical cancer, HPV also causes a subset of penile, vulval and vaginal, anal, mouth and oropharyngeal cancers, estimated to account for 68,400 additional cancer incidents in 2002 (Parkin, 2006). Furthermore, skin infecting HPV types may contribute to formation of non-melanoma skin cancer (zur Hausen, 2009). An HPV
etiology for a subset of cancers of the oesophagus, prostate, bladder, mamma, lung and other organs is discussed but not widely accepted (reviewed in Petersen, 2008). Besides cancer caused by HPV, cutaneous HPVs cause a variety of cutaneous manifestations that in some instances may go along with an increased risk for malignant transformation (reviewed in Cardoso, 2011). Common warts occurring in 3.5% of adults (and more in children) represents the most common cutaneous HPV-related burden (Cardoso, 2011). Genital warts caused by low-risk (alpha) HPVs are reported to occur in 0.2 to 5.6% of men in various European countries (Hartwig, 2012). Though incidence and prevalence may vary worldwide and gender specifically, interpolating estimated prevalence for genital warts and common warts in the current world population of slightly above 7,000,000,000 (United Nations, 2011) approximately results in 70,000,000 (1%) and 245,000,000 (3.5%) people with genital and common warts, respectively. Collectively, the hundred thousands of cancer-related deaths, together with several hundred millions of more ‘benign’ manifestations attributable to HPV worldwide, underscore the tremendous impact of this pathogen on mankind.

**Infection and vaccines**

All HPVs infect and replicate in human keratinocytes (de Villiers, 2004). This tissue and host tropism may be at least partially due to the fact that the early viral promoter (at least of HPV 18) was shown to be active in keratinocytes and cervical carcinoma cells, whereas it was inactive in non-epithelial cell-lines (reviewed in Thierry, 2009). Moreover, HPV pseudovirions consisting of the HPV capsid proteins L1 and L2 were shown to preferentially infect basal keratinocytes (Kines, 2009).

Cervical cancers caused by HPV arise predominantly in the transformation zone located between ecto- and endocervix. It has been speculated that tumors arise there as a consequence of predominant infections at these sites or, alternatively, due to a particularly suited microenvironment (Bodily, 2011).

Using pseudovirions and an animal model mimicking microlesions *in vivo*, it was demonstrated that HPV particles bind to certain heparan sulfate proteoglycans (Selinka, 2007; Kines, 2009), as components of the basement membrane which is normally occluded by an intact epithelium (Figure 2). Thus, microlesions that could originate from sexual activity are necessary for the docking of virus capsids to these structures (reviewed in Assmann, 2011). Upon binding to the heparan sulfate
proteoglycan of the basement membrane, the HPV capsid undergoes conformational rearrangements and exposes the N-termini of the capsid L2 protein which is in turn cleaved by the proprotein convertases (PCs) furin and/or PC5/6 (Kines, 2009). Only these proteolytically processed viral particles infect the basal keratinocytes adjacent to the microlesion, targeting an L1-specific, yet still unidentified receptor (Kines, 2009).

Figure 2: HPV virions first bind to heparan sulfate proteoglycans on the basement membrane separating the epidermis from the dermis (i). The induced conformational change (ii) exposes a furin or PC5/6 cleavage site on L2. After L2 cleavage by one of these proprotein convertases (PC; iii) an L2 neutralizing epitope is exposed and a previously unexposed L1 region binds to an unidentified secondary receptor on the invading edge of the epithelial cells (iv; Schiller, 2010). Figure adapted and modified from (Kines, 2009).

Upon receptor binding HPV is internalized and relocated to endosomes (Day, 2003). Since L2 has membrane penetrating properties, the L2-viral genome complex escapes from the endosome (Kamper, 2006). It is currently debated whether the viral genome is transported to the nucleus via microtubules (Sapp, 2009). The viral genome entry into the nucleus is dependent upon the usual dissociation of the nuclear membrane during mitosis (Sapp, 2009; Assmann, 2011). Transcription of the viral genome is facilitated by association of the L2-genome complex with nuclear bodies (reviewed in Schiller, 2010). Following infection, the viral genome is maintained as a stable episome (reviewed in Doorbar, 2006). Initial binding of the viral capsid to the basement membrane and processing prior to binding to cell surfaces predominantly leads to viral infection of the basal keratinocytes (Kines, 2009). There is a remarkable long delay between initial capsid binding and viral genome transcription (Schiller, 2010) and a concomitant long
exposition time of viral antigens on the cell surface. The host immune system will be focused at the microlesions during wound healing; therefore the delayed HPV gene expression causes a delay of intracellular responses to HPV gene expression which could be beneficial for immune evasion (Schiller, 2010). The strategy to target basal keratinocytes guarantees the infection of a replication-competent cell in which HPV can easily maintain an infectious state and reprogram the hijacked cell for virus production (Assmann, 2011; Kines, 2009).

As the viral particles reside on the basement membrane of microlesions for hours, and because the microlesions are sites of cell-based and humoral immune response, the high efficiency of current HPV vaccines can be explained (Schiller, 2010). The current vaccines Cervarix and Gardasil target the viral L1 protein of HPV types 16 and 18 or 6, 11, 16 and 18, respectively. These vaccines also confer some degree of cross-protection against the HPV types 31, 33 and 45 (Malagón, 2012). Antibodies directed against the epitope exposed after furin or PC5/6 cleavage of the viral L2 protein have cross-neutralizing potential even extending beyond PV genus boundaries that could increase the protective effects of next generation HPV vaccines (Gambhira, 2007, Schiller, 2010). Vaccinations are expensive and require a certain infrastructure, similar to the pap-smear testing regularly performed by gynecologists to early detect neoplasias within the cervix. Indeed, the relative 5-year survival rate after cancer diagnosis in a country is positively correlated with the respective gross domestic product implying that prognosis is especially poor for women in developing countries (Ferlay, 2010). Vaccination of all women against HPV seems unlikely in the near future. Moreover, if men are not also vaccinated, the HPV malignancies in men (Hartwig, 2012) would persist and men would serve as a reservoir for HPVs. Thus, a detailed mechanistic understanding of how HPV causes malignancies is equally needed in order to develop countermeasures against the malicious activities of HPV.

**Genome organization**

In order to understand which tools are utilized by HPV, a brief overview of the proteins frequently encoded by HPV genomes is given before dealing with the viral life cycle per se. HPVs are double-stranded DNA viruses that usually encode six early (E1, E2, E4, E5, E6 and E7) and the two already mentioned late viral capsid proteins (L1 and L2; reviewed in Doorbar, 2006; Doorbar, 2012; Figure 3). Upon successful infection, the viral genome is maintained within the host cell as episome, requiring the
viral replication proteins E1 and E2 (reviewed in Doorbar, 2006). The early proteins E4 and E5 are thought to mainly contribute to virus assembly and release (Doorbar, 2006, Bodily, 2011). HPV faces the paradoxical situation to be dependent upon cellular DNA replication and to infect keratinocytes, a cell type that is differentiating and will exit the cell cycle (Stanley, 2012). The proteins E6 and E7 are thought to create a cellular environment permissive for viral DNA replication, thereby overcoming this challenge (Stanley, 2012). Distantly related PVs do not always contain E6 and E7 open reading frames (ORFs). For example, the gamma-HPV types 101, 103 and 108, the avian Fringilla coelebs and Psittacus erithacus timneh PVs do not possess an E6 ORF (Terai, 2002; Chen, 2007; Nobre, 2009), while the cetacean Phocoena spinipinnis PV lacks E7 (Van Bressem, 2007). It appears that E6 and E7 functions may not be required by all PVs, their functionality may be swapped or may be performed by other viral proteins (Chen, 2007; Nobre, 2009; Tomaić, 2009a). The icosahedral capsid with the encapsidated genome is assembled from 72 L1 pentamers and up to one L2 protein per L1 pentamer (Modis, 2002; Buck, 2008; Wolf, 2010; Figure 3). A detailed picture of the productive viral life cycle is drawn in the next section.

**Figure 3:** HPV 16 genome organization (modified from Doorbar, 2006) superimposed onto an atomic model of the viral capsid solely based on L1 (modified from Modis, 2002). The locus control region (LCR) contains regulatory elements for gene expression and genome replication (reviewed in Doorbar, 2006). The L2 protein is thought to reside within the center of the L1 pentamers (Buck, 2008) which can be recognized as a cleft within the yellow surface regions of each L1 pentamer in the capsid model.
Productive life cycle and epithelial cell responses

Following cellular infection, the HPV life cycle is closely linked to epithelial differentiation (reviewed in Moody, 2010; Figure 4). The HPV genomes are replicated and maintained at about 100 copies per cell (Moody, 2010). Maintenance in the infected basal keratinocytes requires E1 and E2 to initiate viral DNA replication and E6 and E7 to modulate cell cycle regulators to maintain long-term replication competence (reviewed in Bodily, 2011). During the maintenance phase, the viral proteins involved are expressed at extremely low levels, thereby preventing the induction of cellular signals to trigger an inflammatory response probably contributing to escape from immune surveillance (Bodily, 2011). The basal HPV-infected keratinocytes replicate and the additional daughter cells will also be HPV positive, probably mediated by the tethering of the viral episomes to mitotic chromosomes via E2 (McBride, 2006). Daughter cells usually leave the basal epithelial layer while simultaneously differentiating in order to constitute the productive, epidermal barrier (Doorbar, 2006). The starting differentiation induces high levels of viral proteins in these cells that are not easily accessible for the immune system (Frazer, 2009). This immune-evasion strategy has the drawback that HPV is dependent upon cellular replication proteins in these differentiating keratinocytes, which would normally exit from the cell-cycle and cease replication (Stanley, 2012, Bodily, 2011). E7 and E6 maintain replication competence and override cellular responses to this unscheduled replication by, among many others, interfering with pRb and p53 functions, respectively (Doorbar, 2006; Doorbar, 2012). Upon differentiation, the transcript and protein levels of E1, E2, E4 and E5 increase dramatically due to the induction of the viral differentiation-dependent late promoter (Bodily, 2011, Doorbar, 2006), leading to an increased viral copy number of up to several thousands per cell (Bodily, 2011). Following genome amplification, capsid proteins L1 and L2 are induced and virions are assembled and shed from the epithelial surface (Bodily, 2011).
1.2 Human papillomaviruses

Figure 4: Productive life cycle of the mucosal HPV 16 (modified from Doorbar, 2006). HPV 16 exploits the early p97 promoter for E6/E7 expression and the late p670 promoter for the remaining proteins. The p670 transcript is alternatively spliced resulting in usage of early or late polyadenylation sites, governing whether exclusively E1, E2, E4 and E5 or additionally L1 and L2 are expressed, respectively. E2 expression leads to induction of late gene expression by inhibiting the polyadenylation of early genes, thereby exerting a functional switch from genome amplification to the production and release of the fully assembled virus (reviewed in Thierry, 2009). Cells with active E6/E7 are depicted with red nuclei, while cells with additionally active E1, E2, E4 and E5 needed for genome amplification are shown in green. Cells expressing E4, L2 and L1 for viral assembly and release are depicted in green with yellow nuclei. Desquamating cells containing complete virions are depicted in yellow.

In conclusion, viral replication and assembly occur in the differentiating keratinocyte, a cell destined for death and desquamation far from the sites of immune response. Thus, there is no virus-induced cell death and therefore no inflammation, rendering HPV almost invisible for host defenses such as e.g. epidermal Langerhans cells and dermal, stromal dendritic cells. These immune cells remain ignorant of the pathogen for long periods of time (Stanley, 2012).

Since natural HPV infections do not lead to viremia, and because released virus particles are shed from the epithelial surfaces, there is poor access to vascular and lymphatic channels and lymph nodes, often preventing seroconversion (Gravitt, 2011). Moreover proinflammatory cytokines as well as antigen-processing and -presenting molecules are actively down-regulated by HPV (Karim, 2011, Reiser, 2011). The innate host defense is also compromised e.g. by reducing the cellular interferon kappa transcription which in turn negatively affects the transcription of interferon-stimulated genes (Reiser, 2011; Stanley, 2012). Although it is clear that the
1.2 Human papillomaviruses

HPV vaccines elicit a humoral immune response against capsid proteins of certain types, the understanding of natural immunity is limited (Gravitt, 2011). While impaired cellular immunity reduces clearance of HPV infection, a few high-risk HPV infections (up to 20%) do persist even in immunocompetent individuals (Frazer, 2009, Stanley, 2012). The majority of genital infections usually resolves within 10-18 months (Richardson, 2003). Very often, lesions caused by HPV regress as a result of a successful T-cell mediated immune response directed specifically against the E2 and E6 proteins (Woo, 2010). It is predominantly the persisting infections that progress to neoplasias that either regress and clear or persist or progress to invasive cancer (Gravitt, 2011).

The oncoproteins E6 and E7

In up to 90% of cervical cancers, HPV sequences are found integrated into the host genome and often the viral episome is no longer present (Pett, 2007). The minimal consensus integrant sequence encompasses the E6/E7 open reading frame (ORF), thereby implying that these proteins seem to be sufficient for malignancy (Pett, 2007). Disruption of host gene functions by these integration events can further perturb host cell metabolism if the other gene allele is inactivated (Schmitz, 2012).

E6 and E7 of high-risk HPVs are necessary and sufficient for the immortalization of primary, cultured human keratinocytes and to prevent their normal terminal differentiation (Dürst, 1987, Münger, 1989, Hudson, 1990). Abrogation of E6/E7 expression levels by various means in cervical cancer derived cell lines (such as e.g. HeLa) leads to apoptosis or senescence (summarized in Magaldi, 2012). Moreover, it was recently shown that primary human cervical carcinoma cells require continuous E6 and E7 expression for their ongoing proliferation (Magaldi, 2012). The oncoproteins E6 and E7 are always encoded in the alphapapillomaviridae genomes (see Papillomavirus Episteme database).

Within the high-risk alpha HPVs, the E6 and E7 oncoproteins are necessary for the formation and maintenance of cervical cancer. In the following section the E6 protein is discussed in detail, since this oncoprotein was the subject of my research.
1.3 The oncoprotein E6

Regulation of E6 during viral life cycle

E6 has no known enzymatic activity and exerts its functions by interacting with numerous cellular targets (Tungteakkhun, 2008). Before dealing with selected E6 interactions, it should be noted that the E6 expression is tightly regulated during the viral life cycle. E2 down-regulates viral transcription of E6/E7 and induces transcription of L1/L2 (Thierry, 2009; Johansson, 2012), thereby potentially exerting a functional switch from maintaining a cellular environment suited for viral replication to the production and release of fully assembled viruses. This E2 function can be counteracted by the E1 protein (Johansson, 2012). High-risk but not low-risk HPV types encode for a combined E6/E7 ORF (reviewed in Stacey, 2000), and spliced E6 isoforms (the transcripts are often referred to as E6*) exist for most high-risk HPV types (Sedman, 1991; Mesplède, 2012). HPV 18 E6* interacts with full-length E6 (Pim, 1999), possibly by competing for E6-liganded zinc (Barbosa, 1989; Heer, 2011). The interaction of E6* to full-length E6 is thought to counteract certain E6 functions (Pim, 1997; Pim, 1999, Filippova, 2007) while maintaining others (Pim, 2009). Conceptually, E6* may be viewed as a possibility to fine-tune the activity of E6 compared to the E7 activity by maintaining certain E6 functions while counteracting others. In the next sections, the focus lies on selected full-length E6 interactions.

E6 interactions and consequences thereof

The oncoprotein E6 exerts its malignant functions in cooperation with E7. In a three-stage model of carcinogenesis (initiation, promotion, progression), E7 mainly acts during earlier stages, while E6 mainly acts during the later stages (Song, 2000). The HPV phylogeny is reflected by the E6 interactome (White, 2012). Due to the multitude of E6 interactions (Tungteakkhun, 2008), only selected interactors are commented upon. Recent large scale E6 interaction studies (White, 2012, Neveu, 2012) add new targets and suggest that some E6 interactors are HPV type specific, whereas others seem to be targeted by E6 proteins of a larger subset or even all HPV types. Selected E6 interactions involved in the multistep process of cancer formation according to the model of Hanahan and Weinberg (Hanahan, 2000; Figure 5) are discussed below.
1.3 The oncoprotein E6

Figure 5: Conceptual hallmarks of cancer formation (adapted from Hanahan, 2000). These hallmarks can be acquired in various orders or in parallel. Cancer formation is often accompanied by genomic instability that favors the establishment of the hallmarks (Hanahan, 2000). Recently, cancer-related inflammation was shown to also contribute to cancer, particularly by establishing an inflammatory microenvironment in which DNA damage repair is impaired thereby contributing to genomic instability (reviewed in Colotta, 2009). Proteins for which their E6 targeting contributes to the cancer hallmarks as discussed in the main text are indicated in red.

The probably most characterized action of E6 is to induce degradation of p53 (Scheffner, 1990) by interacting with both p53 (Werness, 1990) and the E3 ubiquitin ligase E6AP (Huibregtse, 1991) with a concomitant alteration of the E6AP substrate specificity. This leads to p53 ubiquitylation and subsequent degradation (Scheffner, 1993). The E6-mediated degradation of p53 (Scheffner, 1990) facilitates evasion from apoptosis (Howie, 2009), might enhance tumor angiogenesis (Ravi, 2000), contributes to insensitivity to anti-growth signals (el-Deiry, 1993) and contributes to genomic instability, often detected in cells expressing high-risk E6 and E7 proteins (Rihet, 1996; Moody, 2010; Figure 5).

p53 degradation seems to be an exclusive trait of high-risk HPVs as a study that monitored p53 degradation by 29 E6 proteins came to the conclusion that all high-risk
or probable high-risk E6 induce degradation of p53, whereas the tested low-risk, cutaneous and one E6 protein of unknown status do not (Mesplède, 2012). Instead of mediating degradation of p53, cutaneous beta HPVs repress Notch signaling by binding to Mastermind-like coactivators, a property not shared by alpha HPVs (Tan, 2012). p53 degradation is also induced by E6 in the absence of E6AP (Massimi, 2008), implying other E3 ligases involved or other mechanisms of p53 degradation. One study showed that E6 also causes proteasome independent p53 degradation (Camus, 2007), possibly explaining the E6AP-independent mechanism, though the mechanistic details remain unclear. Several other E6 interactors are also targeted for degradation, and in some cases E6AP is involved (reviewed in Beaudenon, 2008). Moreover, the E6 interaction with E6AP was shown to stabilize E6 in vivo (Tomaić, 2009b) and E6AP was responsible for tethering E6 to components of the proteasome (White, 2012).

E6 induces via E6AP, the degradation of NFX1-91 (Gewin, 2001; Katzenellenbogen, 2009). Degradation of this transcriptional repressor results in increased hTERT transcription, the catalytic subunit of human telomerase (Gewin, 2001; Howie, 2009). An alternatively spliced version of NFX1, NFX1-123, which augments hTERT expression, is stabilized by E6 in vivo (Katzenellenbogen, 2009). Moreover, E6 directly interacts with and stabilizes hTERT (Liu, 2009a). The resulting increased telomerase activity prevents telomere shortening and strongly contributes to increased replicative potential (Figure 5).

E6 has also an influence on the host immune system. Recently, interferon kappa was identified as the interferon predominantly expressed in keratinocytes and full-length E6 proteins of HPV 16, 18 and 31 were shown to down-regulate kappa interferon and kappa interferon stimulated gene (ISG) transcription (Reiser, 2011). By this mechanism, ISGs with antiviral activity, genes involved in interferon signaling and pro-apoptotic genes plus pathogen recognition receptor genes are down-regulated (Reiser, 2011). Moreover, combined HPV 16 E6 and E7 expression leads to down-regulation of toll-like receptor 9 transcription, which in turn reduces ISG expression levels (Hasan, 2007). However, the E6 protein is also targeted by the immune system, since E6 specific T-cell responses have been reported (Woo, 2010).

Among the mucosal alpha HPVs, the E6 proteins from high-risk types possess a C-terminal PDZ-binding motif (PDZ-BM), whereas the E6 from low-risk or cutaneous types do not (Thomas, 2008b). PDZ-BMs are usually located at C-termini of proteins.
and they typically consist of four residues (Stiffler, 2007). A simplified classification system of PDZ-BMs exists (Stiffler, 2007) and class I PDZ-BMs, to which the C-termini of E6 match, have the sequence X-S/T-X-φ, where φ denotes a hydrophobic residue (Thomas, 2008b). The high-risk E6 proteins interact via their PDZ-BM with a multitude of PDZ domain-containing proteins (reviewed in Thomas, 2008b). Phosphorylation of the S/T PDZ-BM residue allows for specific binding of at least one protein (14-3-3ζ that does not contain PDZ domains), to this phosphorylated E6 region (Boon, 2013). The S/T PDZ-BM phosphorylation abolishes E6 PDZ interactions (Liu, 2007; Thomas, 2008b).

PDZ is an acronym referring to Post synaptic density-95, drosophila Discs large and Zona occludens-1, the first proteins in which these domains have been identified (e.g. reviewed in Nagano, 1998). PDZ domain-containing proteins, (in the following abbreviated PDZ-proteins), are multi-domain scaffold proteins, often with more than one PDZ domain (reviewed in Feng, 2009; Javier, 2011, Figure 6), usually involved in regulating cell adhesion and cell polarity (reviewed in Humbert, 2006; Bryant, 2008; Javier, 2011; Figure 6). In several diseases, PDZ-protein levels are altered (Facciuto, 2012). Interaction with PDZ-proteins seems to be a common strategy for viruses (reviewed in Javier, 2008; Javier, 2011; Figure 6). In one documented case one single mutation in the PDZ-BM of a viral protein resulted in altered PDZ interactions culminating in a switch from viral virulence to an attenuated state (Préhaud, 2010).
1.3 The oncoprotein E6

Figure 6: Certain proteins - often containing PDZ-domains - involved in cell adhesion and cell polarity are targeted by various viral proteins, among them HPV E6 (boxed in red; Figure modified from Javier, 2011). The complexes regulating polarity are Crumbs, Par and Scribble, mainly regulating apical, tight junction and basolateral polarity, respectively. For each protein, the total number of present PDZ domains (obtained from Javier, 2011 or from www.expasy.org) is indicated in parentheses. The domain organization of hDlg is presented as an example (modified from Feng, 2009; L27: L27 domain; PDZ: PDZ domain; SH3: Src homology 3 domain; GK: guanylate kinase-like domain).

Among the PDZ-proteins, which interact with E6 are hDlg (also known as SAP-97 or Dlg1), Scribble, MAGI-1 to -3, PATJ, MUPP1 and PSD-95 (reviewed in Thomas, 2008b; Pim, 2012). The presence of Scribble or hDlg stabilizes E6 against proteasomal degradation in vivo (Nicolaides, 2011), similar to the E6AP-mediated E6 stabilization (Tomaić, 2009b) and similar to an E6 stabilizing interaction observed for USP15 (Vos, 2009). Subsequent to E6 binding, the PDZ-proteins, including hDlg, are usually degraded with potential involvement of E6AP (Thomas, 2008b). Certain PDZ-proteins are also degraded in presence of E6 but in the absence of E6AP (Massimi, 2008). The E6 interaction with and degradation of hDlg (Kiyono, 1997, Gardiol,
1.3 The oncprotein E6

1999) disrupts the hDlg function as a negative regulator of G0/G1 to S phase transition (Ishidate, 2000) and confers insensitivity to anti-growth signals upon the host cell (Figure 5). As hDlg depleted cells show resistance to anoikis, a specialized form of apoptosis (Massimi, 2012), hDlg targeting also contributes to evading apoptosis. Since hDlg is involved in regulating epithelial-polarity (reviewed in Bryant, 2008; Javier, 2011), the E6-mediated degradation might promote epithelial-to-mesenchymal transition, thought to contribute to invasion and metastasis (Thompson, 2005). It seems that the multitude of E6-disrupted hDlg (Roberts, 2012) and other PDZ-protein functions are together very important for tumorigenesis. Consistent with this notion, transgenic mice with E6 proteins lacking the PDZ-BM develop smaller and less invasive tumors than mice with the full-length E6 protein (Shai, 2007).

It was demonstrated that PDZ-domains arranged in tandem with short linker sequences often serve as a structural and functional unit called supramolecule or supramodule (see e.g. Feng, 2009; McCann, 2011). Among the PDZ-proteins targeted by E6 are at least two proteins with known tandem PDZ domains that arrange as a supramodule, hDlg and PSD-95 (Feng, 2009; encircled region in the scheme of hDlg; Figure 6). The structural basis for the hDlg-related PSD-95 PDZ1-2 tandem is available (McCann, 2011).

In summary, E6 exerts its function by multiple interactions contributing to cancer formation, though knowledge on the mechanistic, structural details potentially conducive to defining drug targets is still fragmentary.

**Biophysical properties of E6**

In order to understand how E6 exerts and to develop countermeasures against its malignant functions, namely its numerous interactions, the biophysical and biochemical properties of E6 alone and in complexes with its cellular interactors should be precisely known. Moreover, as previously discussed, not all E6 proteins interact with all known E6 interactors. Therefore it will be also of importance to investigate E6 proteins from different HPV types.

Full-length E6 consists of approximately 150 residues and includes two zinc binding domains each coordinating one zinc ion via cysteines embedded in a motif of the form CXXC-X29-CXXC (Barbosa, 1989; Figure 7). E6 tends to aggregation and polydisperity in vitro (Nomine, 2001) and in vivo (García-Alai, 2007). The aggregation propensity of full-length E6 was shown to reside mainly in the amino-
1.3 The oncoprotein E6

terminal zinc binding domain of HPV 16 E6 (Liu, 2009b). The NMR solution structures of mutated N-terminal and of the refined mutated C-terminal zinc binding domain of HPV 16 E6 have been solved (Zanier, 2012; Nomine, 2006). The utilized mutations result in an incapability of the mutated E6 to bind to p53 and to cause p53 degradation (Nomine, 2006, Ristriani, 2009). The structure of mutated, full-length HPV 16 E6 in complex with a peptide derived from E6AP was solved recently as well as a structure of bovine PV-1 E6 in complex with a paxillin-derived peptide (Zanier, 2013).

Figure 7: Organization of the E6 proteins (upper panel) and available structures for the high-risk HPV 16 derived mutants (lower panel; PDB entries 2LJX and 2LJZ). The zinc ions of zinc binding domains (ZBD1 and ZBD2) are depicted as spheres (diameter 1.4 Å) in the lower panel. The PDZ-BM at the C-terminus is exclusively present on high-risk HPV E6 proteins. Several residues on the mutated HPV 16 E6 structures are labeled according to the position on full-length E6. In the full-length protein, the E6 domains adopt the same fold (Zanier, 2012; Zanier, 2013). NMR signals of residues preceding S80 undergo line broadening on full-length E6 likely due to dynamic processes (Zanier, 2012). Therefore the domains may not adopt a defined orientation relative to each other in the free full-length E6. Mutated full-length HPV 16 E6 in complex with a peptide derived from E6AP contains a long linker helix between the E6 ZBDs that is formed by the residues preceding S80 which leads to a defined domain orientation of E6AP-complexed E6 (Zanier, 2013).

1.4 Scope of work and publication summary

In order to elucidate how E6 exerts its malignant functions, i.e. its numerous interactions, the biophysical and biochemical properties of E6 alone and in functional complexes have to be investigated.

Up to now, no structural data on a wild-type HPV E6 are available. Screening of wild-type E6 proteins of types other than HPV 16 for intrinsic solubility could on the one
hand lead to a structure of a non-mutated E6 and on the other hand shed light on structural similarities and differences among the E6 proteins.

To determine the NMR solution structure of a wild-type, full-length E6 protein, various E6 constructs from different HPV types were screened for solubility and NMR-amenability (publication 3). Yet only the C-terminal zinc binding domains of screened high-risk E6 types were soluble and among these, the C-terminal zinc domain of HPV 51 E6 (termed 51Z2) was amenable for NMR analysis and the 51Z2 solution structure was subsequently solved (publication 3). Since 51Z2 contains a PDZ-BM in the less-ordered C-terminal 11 residues, the interaction with a PDZ domain, PDZ2 of hDlg was investigated (publication 3). Several studies of HPV 18 E6-derived peptides of up to seven residues complexed to PDZ domains of hDlg and MAGI-1 are available (Liu, 2007, Zhang, 2007) as well as one structure where it was demonstrated that E6 binds not only with the canonical PDZ-BM to MAGI-1 (Charbonnier, 2008, Charbonnier, 2011). As PDZ domains of murine Dlg (with at least 98% sequence identity to hDlg PDZ domains) show specificity for 9-residue peptides, we hypothesized that E6 could interact via more residues in addition to the canonical PDZ-BM with hDlg. Interestingly, the 51Z2-hDlgPDZ2 interaction involved the nine C-terminal E6 residues, more than found previously. To structurally characterize this complex, a $^{13}$C and $^{15}$N-labeled peptide covering the less-ordered E6 residues was generated (publication 1). This peptide was produced by exploiting the intein system. Surprisingly the peptide spontaneously auto-converted its amino-terminus from glutamine to pyroglutamate and this could be biotechnologically exploited for the production of stabilized proteins or peptides as discussed in publication 1. As the pyroglutaminylation on the peptide did not perturb the complex formation with hDlgPDZ2, the resonances on the hDlgPDZ2-complexed doubly-labeled peptide and the resonances of the peptide-complexed doubly-labeled hDlgPDZ2 were assigned (publication 2). Finally, the NMR solution structure of the E6-derived 11mer peptide in complex with hDlgPDZ2 was solved (publication 3). The implications of biophysical E6 properties, as deduced from the wild-type E6 domain, and of the E6 and complex structures are discussed in publication 3. Collectively, these results contribute mechanistic insight on atomic scale level into the malignancy promoting actions of the HPV oncoprotein E6.
2 Publications

Each of the subsequent publications contributes either directly or indirectly to the characterization of the oncoprotein E6 and its association with hDlg. The respective summary highlights the contribution of each paper towards the dissertation at hand.

2.1 Publication 1

*Recombinant production of isotope-labeled peptides and spontaneous cyclization of amino-terminal glutamine into pyroglutamic acid*

Authors: A. Mischo, O. Ohlenschläger, K. H. Gührs, M. Görlach.
Contributions: AM and MG conceived experiments. AM did wet-lab work. AM and OO performed NMR experiments. AM evaluated kinetic data. KHG performed mass spectroscopy. AM, OO and MG wrote the manuscript.

Summary: This manuscript describes how a stable-isotope labeled peptide was generated *via* recombinant expression as intein-fusion. The purified peptide contains the region of HPV 51 E6 that interacts with hDlg PDZ domain 2 (hDlgPDZ2; see publication 3). The freshly generated peptide (\(^{13}\)C and \(^{15}\)N labeled) in complex with unlabeled hDlgPDZ2 was utilized for NMR spectroscopy in order to elucidate the solution structure of this complex. Unexpectedly, NMR spectral changes over time as well as mass spectroscopy collectively demonstrated auto-cyclization of the amino-terminal peptide residue Gln into pyroglutamate (pGlu). The pGlu conversion kinetics was evaluated by NMR spectroscopy. Though accidental in our system, it might be possible to exploit this spontaneous conversion to generate and subsequently characterize physiological and disease-relevant pGlu-containing peptides such as truncated amyloid β species or to produce N-terminally pGlu-modified and exoprotease-resistant proteins for biotechnological applications.
2.1 Recombinant Production of Isotope-Labeled Peptides and Pyroglutamylation

NMR is a powerful tool for the elucidation of biomolecular structures and their dynamics. Labeling with the stable isotopes $^{15}$N and $^{13}$C paved the way to determine the structures of larger systems. In addition, examination of peptide–protein complexes is facilitated through the use of isotope-labeled peptides. However, the chemical synthesis of isotope-labeled peptides is expensive and recombinant production has yielded a number of peptide-specific protocols in the past.

For structural analysis of the complex between the hIg POZ domain 2 (POZ2) and the HPV oncoprotein E6, we generated a doubly labeled ($^{15}$C and $^{13}$N) 11-residue target peptide (QRTROQNETQV) derived from the C-terminus of HPV E6 by recombinant expression in an insect cell line. Typically, in silico methods are employed to mediate ligation of protein segments. In our approach, the intein just liberates the isotope-labeled peptide through amide bond cleavage at the fusion site in the presence of thiols.

Purification of the soluble intein–peptide fusion protein was carried out by affinity chromatography based on the chitin-binding domain of the intein carrier. On-column cleavage was performed in the presence of thiols (Figure S1 in the Supporting Information), followed by reversed-phase HPLC in order to purify the liberated peptide (Figure S2). All major HPLC peaks were collected and subsequently analyzed by $^{[1,15]}$HN HSQC NMR spectroscopy and MALDI-TOF MS. The material eluting at 14 min yielded a $^{[1,15]}$HN HSQC spectrum corresponding to the composition of the target peptide (Figure S3A). MALDI-TOF MS (Figure S3B, Table 1A) detected a mass corresponding to the calculated mass of the $^{15}$C and $^{13}$N-labeled target peptide (1493.8 Da). MS analysis also indicated the presence of two minor species that probably represent a sodium–peptide adduct and a species with a mass of 1475.9 Da (i.e., $18$ Da, detected as [M–$18$ Da]+ of 1476.9 Da). The target peptide was lyophilized and stored at $-80^\circ$C until further use. The average yield of the target peptide was 10.6 nmol g$^{-1}$ ($1.58 \mu$g g$^{-1}$) of Escherichia coli wet cell mass.

During the resonance assignment of the target peptide bound to the unlabeled POZ2, we observed unexpected NMR signals in a number of through-bond correlation experiments linking the amide proton of the N-terminal residue (Gln) to its own side-chain nuclei (Figure 1), which suggested a cyclization to pyroglutamine. Such a cyclization involves liberation of the Gln1 side chain $^{15}$NH$^3$ and of one of its $\alpha$-amino protons to release $^{15}$NH$_3$ (Scheme 1). This explains the observation of an amide proton signal for the N-terminal residue and is consistent with a mass difference of $-18$ Da between the native and the pygu-containing peptide (Table 1A). The observed spectral "peculiarities" and the mass difference of $18$ Da thus clearly indicate spontaneous pyroglutamine formation in the N-terminal glutamine. We monitored this Gln-to-pyGlu conversion by recording a series of $^{[1,15]}$HN HSQC spectra and observed decreasing signal intensities for the Gln1 side chain NH$_3^+$ protons with a concurrent increase in the pyGlu amide proton signal and two altered backbone amide proton signals for Arg2 and Thr3 (Figure 1E).

Customarily, kinetics of pyGlu formation can conveniently be followed by MS. Here, though, the kinetics were evaluated by

Table 1. A) Masses (MALDI-TOF) observed in the HPLC fraction containing the target peptide. B) Peptide masses from ESI-MS of the peptide in combination with PDZ2. Note that in B) no residual mass with 1493.8 Da (m/z=2H$^+$: expected as 747.9 Da) was observable.

<table>
<thead>
<tr>
<th>Peptide Mass (Da)</th>
<th>Mass (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1476.9 Da (M+H–18 Da)$^{[1,5]}$</td>
<td>1493.8 Da</td>
</tr>
<tr>
<td>1475.9 Da (M+H–18 Da)$^{[1,5]}$</td>
<td>1493.8 Da</td>
</tr>
<tr>
<td>1476.9 Da (M+H–18 Da)$^{[1,5]}$</td>
<td>1493.8 Da</td>
</tr>
</tbody>
</table>

Scheme 1. Pyroglutamic acid formation by $\gamma$-glutamyl acid (top) or $\gamma$-glutamylamine (bottom). In pyGu-containing proteins the pyGu hydroxamate moiety (boxed) constituting part of the carboxyl group is not present because of peptide bond formation.
2.1 Recombinant Production of Isotope-Labeled Peptides and Pyroglutaminylation

NMR spectroscopy with the labeled peptide at hand and MS served as tool to characterize the endpoint of this reaction. The increase in the well-resolved NMR signal of the pGlu amide proton (relative to a reference signal that did not change by more than 10% for the spectra recorded with equal acquisition parameters) was monitored. The downstream time for peptide analysis after HPLC separation was estimated as 10 h. Nonlinear regression was employed to fit data according to a first-order reaction (Figure 2). At the endpoint of pGlu formation (ca. 500 h), a small aliquot of the NMR sample was removed and subjected to ESI MS (Figure S5, Table 1B). Only pGlu-modified peptide—and no residual unmodified peptide—was detected, confirming that the N terminus of the peptide had been virtually completely converted into pGlu during this time. Accordingly, no further NMR spectral changes were observed. Spontaneous pGlu formation, also reported to occur from amino-terminal glutamic acid, is dependent upon buffer conditions, temperature, pressure, and possibly on adjacent residues, as well as on local structure. The spontaneous cyclization for the target peptide here was shown to progress to 50% within 61.8 h (Figure 2), implying degrees of pGlu modification of the N terminus of approximately 85 or 98% after 7 and 14 days, respectively. Spontaneous cyclization thus takes place within days, but probably has to be evaluated separately for each system studied. The kinetics of the cyclization observed here are very close to what has been observed for free glutamine and faster than for glutamate. In vivo conversions of amino-terminal glutamine or glutamate residues of proteins/peptides to pGlu are catalyzed by glutaminyl cyclases (EC 2.3.2.5). Although a comprehensive assessment of the biological significance of pGlu formation has, to the best of our knowledge, yet to be undertaken, the physiological relevance of amino-terminal pGlu was initially demonstrated by the finding that thyrotropin-releasing hormone (TRH; pGlu-His-Pro-NH₂) binds to its receptor with at least 216-fold increased affinity relative to the noncyclic derivative Gin-His-Pro-NH₂. Importantly, pGlu formation at the N terminus alters the solubility and increases both the amyloidogenicity and the toxicity of amino-terminally truncated Aβ peptides that start with Glu or Glu1. Moreover, pGlu-modified Aβ is reported to aggregate behavioral defects in transgenic mice. pGlu-modified amyloid peptides represent approximately 50% of the total amyloid in senile plaques. For Aβ peptides it was shown that glutaminyl cyclases contribute to pGlu formation. Hence, it might be interesting to assess the toxicities.
and formation of higher-order aggregates of such Aβ degradation products by use of Aβ fragments derived from intact
fusions. This would have the advantage of avoiding significantly altered peptide properties and toxicities as reported for chemically synthesized Aβ peptides.10,21

In summary, we have demonstrated that the intact system can be utilized to prepare preparative amounts of doubly la-beled peptides. The advantages of this system include avoidance of hazardous cleaving chemicals and of (costly) proteases, the latter of which can in turn require, and leave behind, non-
native residues in the peptide. Our data clearly demonstrate the formation of pGlu by spontaneous self-cyclization of the N-terminal glutamine. This self-cyclization could provide for an N-terminus that is less prone to exopeptidolytic attack and, hence, stabilize the peptide in biochemical assays. Because the self-cyclization observed here occurs after the purification of the peptide, any specific artifact of the intact system used for recombinant expression and purification of the peptide can safely be excluded. Because the pGlu protects against Edman degradation or other processes that require a free N-terminus, it might have to be removed by use of pyroglutamate exopeptidase (EC 3.4.19.3).22 We note that the formation of a protective pGlu N-terminus or the removal of spontaneously formed pGlu can be achieved with commercially available enzymes.23

In conclusion, spontaneous pGlu formation from amino-termi-nal glutamine (or glutamic acid) residues might be a complicating factor that deserves consideration when generating pro-teins or peptides from fusions by any cleavage method. However, it constitutes an added protective benefit for the produc-tion of isotope-labeled peptides.

Experimental Section

For recombinant expression, the coding sequence for the peptide of interest was ligated into the intein-coding vector pYB21 (New England Biolabs). Expression in E.coli BL21 (DE3) was achieved in minimal media, containing 10 g/l glucose and 10 g/l ammonium chloride, at 20°C upon induction with isopropyl-β-D-thiogalactopyranos-ide (IPTG; 0.4 mM). The soluble fusion protein was immobilized on a chitin-bead column. After washing of the column, the peptide was liberated by addition of dithiothreitol (DTT) and 2-mercapto-
ethanesulfonic acid (MESNa; 50 mM each). Liberated peptide was collected, dialyzed, concentrated, and further purified by HPLC on a C8 reversed-phase column, with application of a 0–60% acetonitrile gradient in water supplemented with TFA (0.1%).

Purified peptide was dialyzed against water and analyzed by NMR spectroscopy and MALDI-TOF MS. [H,15N]HSQC's were used to follow the conversion of the N-terminal Gin to pGlu of the peptide in complexation with the P2D2. As soon as no further spectral changes were observable, we re-examined the complexed peptide by ESI MS. The experimental details are given in the Supporting Information.

Abbreviations

hDlgl: human disks large homologue 1, HPV: human papillomavi-rus, pGlu: pyroglutamate (5-oxo-proline), P2D2 domain: PSD95 Dlg1 Ze-1 domain.

Acknowledgements

We acknowledge the funding of the Fritz Lipmann Institute by the State of Thuringia and the Federal Republic of Germany. A.M. was funded by the Leibniz Graduate School on Ageing (LSGA). We thank Dr. Jörg Leppert and Dr. Ramadurai Ramachandran (both at the Fritz Lipmann Institute, Jena) for technical support and Prof. Dr. Jens Wirths (Goethe University, Frankfurt am Main) for helpful comments on our manuscript.

Keywords: inteins • NMR spectroscopy • peptides • pyroglutamic acid • recombinant production

184.  
202.  
175.  
2376.  
991.  
277.  
3090.  
7978.  

Received: March 13, 2012  
Published online on May 22, 2012
Supporting Information

(C) Copyright Wiley-VCH Verlag GmbH & Co. KGaA, 69451 Weinheim, 2012

Recombinant Production of Isotope-Labeled Peptides and Spontaneous Cyclization of Amino-Terminal Glutamine into Pyroglutamic Acid

André Mischc, Oliver Ohlenschläger, Karl-Heinz Gührs, and Matthias Görlach

cbic_201200178_sm_misellaneous_information.pdf
2.1 Recombinant Production of Isotope-Labeled Peptides and Pyroglutaminylated Peptides

Supporting Information

DNA and cloning

cDNA of the target peptide was generated by annealing two complementary primers (forward sequence 5'-AGGCGCCGCCGCTACGTCGAAGCACGACGGTGCCT3' and reverse sequence 5'-TCGAGATATGGCCTTCTTGAGTGGTGTTGTGATG-3'), by digesting the annealed DNA with BspQI and Boll and ligating it into BspQI and BamHI cleaved pYB21 plasmid according to manufacturer's instructions (New England Biolabs, Frankfurt, Germany). E. coli DH5α clones harboring the plasmid with the target peptide fused to the C-terminus of the Sco VMA intein/chitin binding domain were identified by sequencing.

Protein expression

The target intein-peptide fusion was recombinantly expressed in E. coli BL21 (DE3) cells with ampicillin (100 mg/L) for selection. Bacteria from a single colony were grown at 37°C to an OD600 of 1.5 in 1 L LB culture, harvested, washed with PBS and added to 250 mL modified M9 medium (6 g/L Na2HPO4, 4 g/L uniformly labeled 13C-D-glucose, 3 g/L KH2PO4, 1 g/L NH4Cl, 0.5 g/L NaCl, 2 mL MgSO4, 0.1 mL CaCl2, 5 mg/L thiamine, 1 mL MgCl2, 1 mL Na2EDTA, 1 mL folic acid, 1 mL niacinamide, 1 mL D-pantothenate, 1 mL pyridoxal, 0.1 mL riboflavin). After 30 minutes at 37°C, cells were cooled on ice and protein expression was induced by IPTG addition (0.4 mM). Subsequent overnight expression (16 h) was performed at 20°C. Harvested cells were stored at -80°C until further use.

Peptide purification

All buffers were filtered and degasified with N2 prior to use. 3-4 g wet biomass were resuspended in 30 mL buffer A (50 mM NaCl, 20 mM Tris, 20 μM PMSE, traces of DNBs and RNase A, pH 8.5) and homogenized four times using a French Press (cell pressure: 18,000 psi). The supernatant obtained by centrifugation (30 min, 10,000 x g, 4°C) of the lysate and containing the soluble target fusion protein was applied to a manually-packed, gravity-flow operated chitin-bead (New England Biolabs) column (1 mL / g biomass). The column was pre-equilibrated with 20 column volumes (CV) buffer B (same as A without PMSE, DNBs or RNase A) and the supernatant was applied four times to the column. Subsequently, the column was washed with 20 CV of buffer C (same as B with 1 mM NaCl) followed by 20 CV of buffer B. Target peptide was liberated from the intein by subsequent on-column cleavage by flushing the column with 3 CV of buffer D (same as B with 50 mM DTT and 50 mM MESNA). The column was sealed and the reaction was allowed to proceed at 20°C for three days (Sm Figure S1). Subsequently, the peptide was eluted by four CV of buffer B. Eluent fractions were extensively dialyzed against water at 4°C overnight, pooled and concentrated to a volume of less than 2 mL. The pooled material was subjected to two aliquots to C18 reverse-phase HPLC (Vydac 218TP C18 5 μm column (Grace Deerfield, United States), Sm Figure S2) operated at 1 mL/min. HPLC buffer conditions were 3 min 100% B to 5% B (1% trifluoroacetic acid (TFA) in water, 30 min linear gradient to 100% B (80% acetonitrile (v/v) in B), 5 min 100% B, 2 min linear gradient to 100% B, 2 min 100% B). The target peptide, as detected at 205 nm, eluted with a retention time of 14 min. Fractions of interest were dialyzed extensively at 4°C against water, lyophilized and stored at -80°C until further use. In addition to several HPLC peaks containing unlabeled and/or low molecular weight species (data not shown), one peptide-component eluted around 21 min. This component is a peptide (MKH45KVLIVGSLG) originating from the extreme N-terminus of the intein-peptide fusion that is also liberated from the Sco VMA moiety during thiold-induced cleavage. This occurence of the peptide thus corroborates the successful thiol-induced cleavage of the intein fusion.

The identity of both peptides was validated via 1H, 13C; HSQC spectra and MALDI-TOF mass spectrometry (for the target peptide Sm Figure S3, Table 1a, for the other peptide Sm Figure S4 and Sm Table S1). Prior to NMR spectroscopy and MS, samples were resuspended in 200 μL distilled, autoclaved water including 10% D2O and target peptide concentration was determined from the absorption at 205 nm.

Mass spectrometry

HPLC-purified and water-dissolved HPLC fractions were serially diluted with TFA (0.1%) containing acetonitrile (30%). For MALDI-MS a dried droplet preparation was performed using 1 μL of appropriately diluted sample and 1 μL of a saturated solution of octa-cyano-4-hydroxyphenoic acid (in 95% acetonitrile, 5% TFA). MALDI mass spectra were acquired with an Ultraflex mass spectrometer (Bruker, Bremen, Germany) and evaluated by the flexAnalysis and BioTools software packages (Bruker, Bremen, Germany). Samples containing protein and peptide mixtures were analyzed after NMR experiments by ESI MS (Sm Figure S5). Here, the sample was diluted to a final peptide concentration of 5 μM/mL with formic acid (0.1%) containing acetonitrile (5%) in water. An aliquot of 10 μL of the diluted sample was subjected to a nanoHPLC at a flow rate of 250 nL/min using an Easy nanoLC (Proxeon, Odense, Denmark) equipped with a (5 x 0.3 mm) Zorbax trap column (Agilent, Waldbronn, Germany) and a 10 cm x 75 μm analytical column (Nanocaparations, Nieuwkoop, Netherlands). Separation of the components was achieved by application of a gradient from 5% to 80% acetonitrile in formic acid (0.1%). The efflux was directly sprayed into the orifice of a LTQ Orbitrap XL mass spectrometer (Thermo Scientific, Dreieich, Germany) and mass spectra were recorded. The evaluation of the spectra was performed by Xcalibur (Thermo Scientific, Dreieich, Germany) and deconvolution of the protein mass spectra was achieved by Xtract (Thermo Scientific, Dreieich, Germany).

NMR spectroscopy

All NMR spectra as implemented in Topspin 2.1 (Bruker Biospin, Rheinsetten, Germany) were acquired at 293 K on a Bruker 600 MHz Avance III spectrometer equipped with a cryo-probe. Lyophilized, water-resuspended HPLC fractions were evaluated in 3 mm NMR tubes and characterized by 1H and 13C-NMR spectroscopy, respectively. The purified, 15N and 13C labelled target peptide (1.25 mM) was analyzed in Na-phosphate (20 mM), TCEP (4 mM), D2O (10%), NaCl (0.05%), pH 6.5 in presence of unlabeled hDIIg (SAP-97) PDZ domain (2 mM). Pyroglutaminyl acid formation of the first peptide residue became apparent during the resonance assignment of the complexed peptide using HNCACB, HN(CAC)2, CCC(O)NH, HCC(O)NH and HCC(O)NH experiments, respectively. All spectra were processed with Topspin 2.1 and analyzed with CARA.
2.1 Recombinant Production of Isotope-Labeled Peptides and Pyroglutaminylation

![Image of gel electrophoresis](image1)

Figure S1. Coomassie-stained 10% nonreducing SDS-polyacrylamide gel of chitin-bead attached target intein-peptide fusion of four representative chitin-bead columns (lanes 1 and 2, 3 and 4, 5 and 6, 7 and 8, respectively). Odd lanes: column material prior to flushing with thiols. Even lanes: column material post 3 days incubation in presence of thiols. The expected positions on the gel for fusion protein (28.0 kDa) and liberated intein (30.0 kDa) are indicated by an asterisk (*) and by a minus sign (−), respectively. Note that the liberated peptides GPRTRGRQNETEQ (1.5 kDa, theoretical pIca = 11.7) and MKIEEQKLVGISEG (1.7 kDa, theoretical pIca = 4.8) were not detectable even by Schöpper gels for small peptides, even in combination with highly sensitive silver staining combined with a protein fixation protocol enhancing the staining of small or highly charged peptides (data not shown).[17-19]

![Image of HPLC](image2)

Figure S2. Representative C18 reverse phase HPLC run (see peptide purification section for details). The target peptide elutes at 14 min (peak highlighted in red) and the MKIEEQKLVGISEG peptide elutes at 21.2 min (blue) indicating the success of thiol-induced intein-fusion cleavage.

![Image of mass spectrometry](image3)

Figure S3. Characterization of the water-dissolved peptide eluting at 14 min HPLC retention time (the freshly prepared target peptide). A) An [14N, 15N] HSOC showed 10 intense NH signals, 4 NH doublets (linked by horizontal lines), and one spectrally folded resonance (starred, probably an arginine side-chain). This matches to the target peptide harboring three Gln and one Asn with at least one of three sidechain Arg signals observable. B) MALDI-TOF identified molecular masses of 1476.9, 1494.9 and 1516.5 Da corresponding to [M+H]+, [M+Na]+ and [M+K]+ target peptide forms, respectively. These results are in accordance with a calculated mass of 1493.6 Da for the 15N labeled target peptide. For implications of the "M - 18Da" species see text.
2.1 Recombinant Production of Isotope-Labeled Peptides and Pyroglutaminylation

Figure S4. Identification of the water-soluble peptide MKIEEGLVLVGSEG. A) A [1H, 15N] HSQC showed 13 NH signals, no NH, or folded arginine side-chains and matches to a peptide with approximately 15 residues without any Gin, Asn and Arg residues. B) MALDI-TOF identified molecular masses of 1690.1, 1712.2 and 1728.1 Da corresponding to the [M + H]+, [M + Na]+ and [M + Na + Methionine oxidized] peptide, respectively. These results are in accordance with a calculated mass of 1690.0 Da for that peptide.

Figure S5. ESI/MS of the 15N and 13C-labeled target peptide. The exclusive presence of a 1475.8 Da mass (detected as [m+2H]+, 738.0 Da) and absence of 1493.6 Da species ([m+2H]+ expected to appear as 747.3 Da; position marked by an arrow) corroborate the complete conversion of amino-terminal glutamine to pGlu after completion of NMR experiments (post 500 h).

Table S51. MALDI-MS derived masses of the HPLC peak eluting at 21.2 min

<table>
<thead>
<tr>
<th>Mass (Da)</th>
<th>Charge</th>
</tr>
</thead>
<tbody>
<tr>
<td>1690.1</td>
<td>+2</td>
</tr>
<tr>
<td>1712.2</td>
<td>+2</td>
</tr>
<tr>
<td>1728.1</td>
<td>+2</td>
</tr>
</tbody>
</table>

[9] m_{theo} calculated mass for 100% 15N and 13C labeling of the MKIEEGLVLVGSEG peptide.

2.2 NMR assignment of a PDZ domain in complex with a HPV 51 E6 derived peptide

2.2 Publication 2

*NMR assignment of a PDZ domain in complex with a HPV51 E6 derived N-terminally pyroglutamic acid modified peptide*

Authors: A. Mischo, O. Ohlenschläger, R. Ramachandran and M. Görlach.

Contributions: AM and MG conceived experiments. AM did wet-lab work. AM, RR and OO performed NMR experiments. AM evaluated NMR spectra. AM, OO and MG wrote the manuscript.

Status: Published in *Biomolecular NMR assignments*, 2012, DOI: 10.1007/s12104-012-9374-1.

Summary: Having generated the pGlu-modified, $^{13}$C- and $^{15}$N-labeled E6-derived peptide (publication 1), it was demonstrated that utilizing the pyroglutaminylated peptide compared to an acetylated one did not perturb the hDlgPDZ2-E6 interaction. Thus, the $^1$H, $^{13}$C and $^{15}$N resonances of the pGlu-modified, $^{13}$C- and $^{15}$N-labeled peptide in complex with unlabeled hDlgPDZ2 and the $^1$H, $^{13}$C and $^{15}$N resonances of $^{13}$C- and $^{15}$N-labeled hDlgPDZ2 in complex with unlabeled, acetylated peptide were assigned and deposited in the BioMagResBank. Resonance assignment is a *conditio sine qua non* for determining biomolecular structures by NMR spectroscopy and, here, to elucidate the structural basis of the hDlgPDZ2-E6 interaction (publication 3).
2.2 NMR assignment of a PDZ domain in complex with a HPV51 E6 derived N-terminally pyroglutamic acid modified peptide

André Mischo · Oliver Ohlenschläger · Ramadurai Ramachandran · Matthias Görlich

Received: 6 October 2011 / Accepted: 25 February 2012 © Springer Science+Business Media B.V. 2012

Abstract The resonance assignment of an amino-terminal pyroglutamic acid containing peptide derived from the E6 protein of human papillomavirus (HPV) type 51 in complex with PDZ domain 2 of hDdg/SAP-97 is reported. The assignments include $^1$H, $^{13}$C and $^{15}$N resonances for the protein and peptide in the complex and all of the peptide’s pyroglutamic acid nuclei.

Keywords HPV · E6 · hDdg · SAP-97 · PDZ · Pyroglutamic acid

Biological context

Human Papilloma Virus (HPV) infections by high-risk HPV types are intimately linked to the development of cervical cancer (reviewed in zur Hausen 2009), accounting for 274,000 cancer-related deaths among women in 2008 worldwide (Ferlay et al. 2010). The HPV oncoproteins E6 and E7 are responsible for cellular transformation and maintenance of the malignant phenotype (Narisawa-Saito and Kiyono 2007). E6 binds to a multitude of cellular proteins contributing to reprogramming of HPV infected, basal keratinocytes. This in turn maintains cellular DNA replication competence, upon which viral replication depends (Yugawa and Kiyono 2009). E6-interaction mediated degradation of cellular targets through involvement of the E3 ubiquitin ligase E6AP (and possibly other E3s) include e.g. p53 and hDdg/SAP-97 (human Ddg/synapse-associated-protein-97; Narisawa-Saito and Kiyono 2007).

In order to elucidate the binding of a peptide representing exactly the unstructured, flexible C-terminus of the E6 protein from high-risk HPV type 51 (data not shown) harbouring a PDZ domain binding-motif which binds to PDZ domain 2 of hDdg/SAP-97, we chose to determine their complex structure by NMR spectroscopy. This is also consistent with a recent report (Charbonnier et al. 2011), that residues upstream of a mutated HPV16 E6 PDZ-binding motif contribute to such interaction with a different PDZ-domain. Due to the repetitive nature of the peptide sequence (QRTRQRNETQV) which includes three arginines we obtained (a) the peptide resonance assignment utilising $^{13}$C and $^{15}$N labelled peptide in complex with the unlabelled PDZ domain and (b) the protein resonance assignment employing the $^{13}$C and $^{15}$N labelled PDZ domain and unlabelled peptide.

Methods and experiments

The in vitro-based preparation of the labelled peptide will be published elsewhere. The N-terminal Gln (Q) residue of the E6 derived peptide (QRTRQRNETQV; HPV51 E6 residues 141-151) converted spontaneously into pyroglutamic acid (pGlu) upon release from the intact carrier protein. The plasmid encoding the C-terminally His6-tagged hDdg PDZ domain 2 (hDdg residues 318-406; UniProtKB (http://www.uniprot.org) entry Q12959) was obtained from ADDGENE (http://www.addgene.org) and expression and purification of the PDZ domain was performed as described by Liu et al. (2007).

The purified $^{13}$C/$^{15}$N labelled peptide containing the N-terminal pGlu was examined in 20 mM Na-Phosphate, 4
2.2 NMR assignment of a PDZ domain in complex with a HPV 51 E6 derived peptide

mM TCEP, 10% D2O, 0.05% NaN3, pH 6.5 at a concentration of 1.25 mM in a 3 mM NMR tube in presence of 3 mM unlabelled PDZ domain. Purified 800 μM 13C/15N labelled hDlg PDZ domain 2 was examined in presence of 2 mM unlabelled, amino-terminally acetylated, chemically synthesised QRTRQRNETQV peptide in a 5 mM NMR tube under the same buffer conditions. 100 μM of 13C/15N labelled PDZ domain in presence of 2.5-fold excess of unlabelled pGlu modified peptide was assessed using [1H,15N]-HSQC and [1H,13C]-HSQC experiments.

All spectra were acquired at 293 K on a Bruker 600 MHz Avance III spectrometer equipped with a cryo-probe, processed with Topspin (Bruker BioSpin) and analysed with CARA (Keller 2004). For experiments to be carried out in D2O as sole solvent, samples were freeze-dried and resuspended in D2O in their respective original volume.

The 1H, 13C and 15N resonances of the labelled biomolecules in presence of a 2.5 fold excess of the respective unlabelled interaction partner were assigned using [1H,15N]-HSQC, HNCACB, HN(CA)CO, HNCO, HNHA experiments (mainly for backbone atoms) and [1H,13C] aliphatic/HSQC, C(C)(C)/NH, H(CC)CONH, H(CC)CH-COSY, H(CC)CH-TOCSY (side-chains) as well as [1H,13C] aromatic/HSQC, (HB)CB(CGCD)HD, (HB)CB(CGDCDE) HE, aromatic H(CC)CH-COSY, [1H,13C] aromatic]-NOESY-HSQC (hDlig aromatic side chains), respectively.

Assignments and data deposition

The well dispersed [1H,15N]-HSQC of hDlig PDZ domain 2 in complex with the N-terminally acetylated QRTRQR NETQV peptide is shown in Fig. 1a and indicates a well-folded PDZ domain. Only very minor differences for the chemical shift of terminal PDZ domain residues (E319 and the His-tag residues H409-411) in complex with unlabelled pGlu-modified peptide were observed in [1H,15N]-HSQC (Fig. 1a, red contours) and [1H,13C]-HSQC (not shown)

![Fig. 1](image)

Fig. 1 [1H,15N]-HSQC spectra of a hDlig PDZ domain 2 in presence of unlabelled N-acetylated peptide (blue contours) and pGlu-modified peptide (red contours) and b the pGlu-TRTRQRNETQV peptide each in complex with its unlabelled interaction partner in 2.5-fold excess. Backbone resonances numbered according to position in the corresponding full-length protein. In a assigned resonances are numbered 318–406 for the PDZ domain and 407–411 for the GSHH residues of the C-terminal histidine tag. Star Folded and probable Arg side chain resonances. Asn and Gln side-chain NH2 resonances: connected by horizontal lines and labelled with SC. N-terminal pGlu resonance: boxed in b

© Springer
NMR assignment of a PDZ domain in complex with a HPV 51 E6 derived peptide

The assignment for the PDZ domain has a completeness of 99.6, 99.7 and 98.0% for $^1$H, $^{13}$C and $^{15}$N resonances, respectively, excluding hydroxyls, the Arg guanidino groups, Lys NH$^\text{F}$, His H$^{\text{H}}\text{N}^\text{Q}$ and H$^{\text{C}}\text{N}^\text{Q}$ as well as the Phe and Tyr C$^\text{Q}$ and the Tyr C$^\text{F}$. A total of 1040 resonances of the PDZ domain were assigned. This adds 119 mainly side chain assignments for the PDZ domain as compared to an earlier assignment of this PDZ domain in complex with a different HPV18 E6 derived peptide comprising only 6 residues (Liu et al. 2007). We also note a significant difference in the $^{15}$N chemical shift for residues K327 and T394 which are located outside of the peptide binding region of the PDZ domain.

The [H, $^{13}$N]-HSQC of pGlu-RTRQRNETQV in complex with the unlabelled PDZ domain is presented in Fig. 1b. The peptide assignment is complete for the customarily observable $^1$H, $^{13}$C and $^{15}$N resonances and includes the complete assignment for the pGlu residue of the peptide.

Assignments for the PDZ domain (89 residues) with attached linker and hexahistidine tag (8 extra residues) and the PDZ bound peptide (11 residues) were deposited in the BioMagResBank under the accession number 17942.

Acknowledgements We acknowledge the funding of the Fritz Lipmann Institute by the State of Thuringia and the Federal Government of Germany. AM was funded by the Leibniz Graduate School on Aging (LGSa). We thank S. Rothenmund (IZKF Leipzig, Germany) for the synthesis of unlabelled peptides.

Conflict of Interest The authors declare that they have no conflict of interest.

References


2.3 Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction

2.3 Publication 3

*Structural insights into a wildtype domain of the oncoprotein E6 and its interaction with a PDZ domain*

Authors: A. Mischo, O. Ohlenschläger, P. Hortschansky, R. Ramachandran and M. Görlach.

Contributions: AM and MG conceived experiments. AM purified and biophysically characterized proteins. PH measured surface plasmon resonance-based interaction data. AM, RR and OO recorded NMR spectra. AM evaluated NMR spectra. AM and OO performed structure calculations, AM, OO, PH and MG wrote the manuscript.

Status: Submitted to PLOSOne on February 08, 2013.

Summary: In this manuscript, the first structure of a wild-type HPV E6 domain is presented, and differences and similarities to other available mutated HPV 16 or BPV 1 structures are discussed. Our E6 domain experiences conformational plasticity at elevated temperatures. The interaction of the wild-type E6 domain with a PDZ domain, hDlgPDZ2 was investigated. Surface plasmon resonance and NMR data revealed that E6 interacts with hDlgPDZ2 by more residues than anticipated and these additional residues significantly contribute to affinity. Having generated a $^{13}$C- and $^{15}$N-labeled E6-derived peptide containing all E6 residues involved in binding (publication 1) and having demonstrated that pyroglutaminylation of that peptide did not perturb the complex (publication 2), the structure of the hDlgPDZ2 in complex with that peptide was solved. The complex structure reveals that E6 interacts with hDlgPDZ2 by additional contacts including charge complementarity. General implications of this mechanism for PDZ-interactions are discussed.
Structural insights into a wildtype domain of the oncoprotein E6 and its interaction with a PDZ domain

Authors and affiliations
Andre Mischo¹, Oliver Ohlenschläger¹, Peter Hortschansky², Ramadurai Ramachandran¹ and Matthias Görlach¹*.

¹ Biomolecular NMR Spectroscopy, Leibniz Institute for Age Research - Fritz Lipmann Institute, 07745 Jena, Germany
² Molecular and Applied Microbiology, Leibniz Institute for Natural Product Research and Infection Biology - Hans-Knöll-Institute, 07745 Jena, Germany

* E-mail: mago@fli-leibniz.de

Abstract
The high-risk human papilloma virus (HPV) oncoproteins E6 and E7 interact with key cellular regulators and are etiological agents for tumorigenesis and tumor maintenance in cervical cancer and other malignant conditions. E6 induces degradation of the tumor suppressor p53, activates telomerase and deregulates cell polarity. Analysis of E6 derived from a number of high risk HPV finally yielded the first structure of a wild-type HPV E6 domain (PDB 2M3L) representing the second zinc-binding domain of HPV 51 E6 (termed 51Z2) determined by NMR spectroscopy. The 51Z2 structure allows to pinpoint HPV-type specific structural differences between E6 proteins. The observed temperature sensitivity of the well-folded wild-type E6 domain implies a significant malleability of the oncoprotein in vivo. Hence, the structural differences between individual E6 and their malleability appear, together with HPV type-specific surface exposed side-chains, to provide the structural basis for the different interaction networks reported for individual E6 proteins. Furthermore, the interaction of 51Z2 with a PDZ domain of hDlg was analyzed. Human Dlg constitutes a prototypic representative of the large family of PDZ proteins regulating cell polarity, which are common targets of high-risk HPV E6. Nine C-terminal residues of 51Z2 interact with the second PDZ domain of hDlg2. Surface plasmon resonance in conjunction with the NMR spectroscopy derived complex
2.3 Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction

structure (PDB 2M3M) indicate that E6 residues N-terminal to the canonical PDZ-BM of E6 significantly contribute to this interaction and increase affinity. This complex structure allows to deduce how residues outside of the classical PDZ-BM of E6 and probably of PDZ-binding proteins of other viruses enhance their affinity towards PDZ domains in order to successfully compete with cellular proteins for their respective PDZ targets.

Introduction

Infections are responsible for 17.8% of all cancers [1]. Several human papilloma virus (HPV) types are the most frequent viral cancer-causing agents and are linked to well over half a million incidences of cervical cancer each year [1,2,3]. Though vaccines against HPV types 6, 11, 16 and 18 are available, these HPV types cover only ~ 70% of cervical cancer cases and the vaccines do not cure already existing infections [3,4]. Development of therapies against HPV-caused malignancies thus requires further mechanistic insight into how oncogenic HPV drive tumor development and maintenance.

The HPV oncoproteins E6 and E7 of tumor-associated so-called “high-risk” HPVs prevent the differentiation of HPV-infected keratinocytes and immortalize primary, cultured human keratinocytes [5,6,7]. Moreover, primary human cervical carcinoma cells require E6 and E7 expression for proliferation [8]. Upon depletion of E6 and/or E7, HPV-transformed cervical cancer derived cell lines undergo apoptosis or senescence [9,10,11,12,13,14]. Apoptotic HPV positive cancer cells may transform human primary fibroblasts by horizontal gene transfer of the E6/E7 ORF [15]. The double-stranded DNA genome of HPV is replicated as episome in infected cells, but almost 90% of cervical carcinomas contain genome-integrated HPV sequences, encompassing at least the E6/E7 open reading frame (ORF) [16]. The integration and subsequent loss of episomal HPV sequences leads to absence of E2 which usually negatively regulates the transcription of E6/E7 [16]. Collectively these findings suggest that the “malicious couple” E6 and E7 is necessary for formation and maintenance of cervical cancer.

The HPV oncoprotein E6 has no known enzymatic activity. It interacts with numerous cellular proteins and these interactions contribute to reprogramming keratinocytes so as to prevent normal differentiation and to maintain cellular replication competence in order to amplify the HPV genome [17,18].
The inactivation of p53 is a general hallmark of tumorigenesis and tumors frequently show mutations and/or reduced p53 levels [19]. E6 interacts with p53 [20] and the E3 ubiquitin ligase E6AP [21]. The E6 interaction leads to an alteration of the substrate specificity of E6AP ultimately resulting in proteasomal p53 degradation [22,23]. E6 also induces p53 degradation independent of E6AP [24] and also independent of ubiquitin-mediated proteasomal degradation [25]. Interestingly, p53 degradation is a common mode of action of the high-risk HPVs and is rarely found for other HPV types [26,27].

A second common feature of tumors constitutes an increase in telomerase activity [19]. This increase results both from direct interaction of E6 with the enzyme and via induction of telomerase reverse transcriptase at the transcriptional level [28,29]. Again, this property appears to be shared primarily among the oncogenic HPV types such as e.g. HPV 16, 18 and 51 [30].

Furthermore, E6 of high-risk types interacts with a number of PDZ-domain containing proteins (PDZ proteins) via its C-terminal PDZ-binding motif (PDZ-BM) [31,32]. These PDZ proteins, usually containing multiple PDZ domains, act frequently as hub proteins in signaling and/or regulation of cell polarity and are typically degraded upon E6 interaction [31]. Deletion of the PDZ-BM from the high-risk type HPV 31 E6 results in a retarded growth-rate of infected cells and a reduced copy number of the viral episome [33]. The absence of the PDZ-BM of HPV 16 E6 in an in vivo mouse model resulted in smaller cervical tumors [34]. This strongly suggests a tumor promoting mechanism based upon E6’s targeting of PDZ proteins. Interestingly, PDZ proteins are common targets for human tumorigenic and nontumorigenic viruses [35,36]. The proteins targeted by these viruses control formation of tight junctions, cell adhesion and apoptosis [36]. This targeting appears to be a common strategy to support viral replication and transmission to new hosts. A prominent case here constitutes the observation that one single mutation in the PDZ-BM of the envelope protein of rabies virus dramatically changed its PDZ protein target spectrum and resulted in a switch from the virulent to an attenuated state [37].

One PDZ domain containing protein targeted by E6 for degradation in vivo is the multi-domain protein hDlg (human Dlg/hDlg1/SAP-97) [38]. The E6 dependent reduction of hDlg levels has been demonstrated by an in vitro degradation assay for a number of high-risk HPVs, among them HPV 16, 18 and 51 [32]. Human Dlg is part of the Scribble polarity complex that controls basolateral polarity [39] and it is
required for adherens junction formation and differentiation of epithelial cells [40]. Human Dlg is expressed in human keratinocytes and localizes to nuclear, cytoplasmic, membrane-associated and cytoskeletal pools that are thought to exert different functions [41]. Different isoforms of hDlg with different cellular localization and translation efficiency by alternative splicing [41,42]. The hDlg-APC (Adenomatous Polyposis Coli) complex negatively regulates cell cycle progression from G0/G1 to S phase [43] and hDlg-depleted keratinocytes show an increased resistance to anoikis [44]. Human Dlg recruits the Src homology 3 domain-containing (RhoG-specific) guanine nucleotide exchange factor (SGEF) to the cytoskeleton and induces SGEF activity [45]. Interestingly, this cytoskeletal pool of hDlg is not degraded by E6 [45] Moreover, E6 interacts with SGEF in an hDlg-dependent manner and maintains high RhoG activity, thereby increasing invasive capacity [45]. Thus, the role of hDlg in tumor formation seems to be ambivalent and E6 apparently specifically abrogates certain tumor-suppressive hDlg activities [41].

Full-length E6 consists of approximately 150 residues and includes two zinc-binding domains (ZBDs) each coordinating one zinc ion via cysteines arranged in a motif of the form CXXC-X_{29}-CXXC [46]. Recombinant full-length E6 tends to precipitate upon tag-removal [47] and the aggregation propensity mainly resides in the amino-terminal zinc-binding domain of E6 [48]. Notably, continued efforts have recently born out fruits and resulted in NMR spectroscopy derived solution structures of the N-terminal and of the C-terminal zinc-binding domain of HPV 16 E6 [49,50]. However, all of these E6 constructs contained mutations in order to ensure sample monodispersity and to prevent aggregation [50]. Unfortunately, these mutations abolish the hallmark property of full-length E6 to bind and degrade p53 [49,51]. In particular, the F47R mutation prevents dimerization of E6 [50]. Dimerization (and possibly further aggregation), however, may be required for full E6 activity [50,51]. The interaction of E6-derived short 6 or 7mer peptides, respectively, with PDZ domains of hDlg was previously elucidated by X-ray crystallography (PDZ2, PDZ3) [52] and NMR spectroscopy (PDZ2) [53].

Up to now, no structural data on a wild-type HPV E6 are available. So we reasoned that addressing wild-type E6 proteins of high-risk types other than HPV 16 could on the one hand lead to a structure of a wild–type E6 and on the other hand shed light on structural similarities and differences among the E6 proteins including their interaction with target proteins. Here we determined the solution structure of the wild-
2.3 Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction

type C-terminal zinc-binding domain of E6 derived from the high-risk HPV 51 and unraveled the structural basis of its interaction with the PDZ domain 2 of hDlg.

Results
Assessment of solubility of amino-terminally His$_6$-tagged, recombinant E6 constructs derived from the high-risk types 16, 18, 26, 31, 45, 51, 59 and 97 as well as for the non-tumorigenic cutaneous type 1a was performed after bacterial expression in presence of 10 µM of Zn$^{2+}$ as solubility of E6 constructs is dependent on low µM concentrations of zinc [54]. The results are summarized in Table S1. None of the wild-type sequence full-length E6 constructs nor their respective N-terminal zinc-binding domain (ZBD) turned out to be soluble (Table S1). However, all C-terminal ZBDs of high-risk E6 proteins were at least partially soluble except for HPV 59 which could not be expressed at all although codon-optimized DNA sequences for expression in $E$. coli had been employed (Table S1). The soluble constructs are denoted as E6Z2 (e.g. 51Z2 stands for the second, i.e. C-terminal, ZBD of HPV 51 E6).

The soluble E6Z2s (Table S2) were purified and tested for monodispersity as detailed in the SI. For constructs that appeared to be monodisperse, [¹H,¹⁵N]-HSQC spectra were recorded (see SI for details). If such a spectrum contained an acceptable signal distribution, homogeneous peak intensities and a signal number consistent with the expected number of observable amide groups, the respective protein was considered amenable to further NMR spectroscopic analysis. In the end, only 26Z2 and 51Z2 fulfilled these criteria (Table S2). Since 26Z2 exhibited signs of unfolding as indicated by spectral changes after a few days (Figure S1, Table S2), we finally concentrated our efforts on 51Z2. Results of the initial characterization of this E6 domain are presented in Figure S2. [¹H,¹⁵N]- HSQC spectra of 51Z2 recorded at increasing temperatures from 4 °C to 45 °C resulted in reversible signal disappearance above 20 °C (Figure 1). Only side chain signals, a few signals from backbone amides of the flexible C-terminal E6 tail and of the cysteines involved in zinc coordination were detected at elevated temperatures (Figure 1).

Resonance assignment (Figure 2) and Nuclear Overhauser Effect (NOE) NMR-spectroscopy was carried out at 10 °C (Table S3). The resonance assignment for customarily NMR observable nuclei is virtually complete for the E6 residues (see SI for details). Subsequently, 1501 NOEs, torsion angle constraints derived from 63
experimentally determined HNHA $^3$J-couplings as well as TALOS+ derived constraints (Table 1, SI) were utilized for distance geometry-based structure calculation with CYANA [55]. The calculated final structures were subjected to energy minimization in water employing CNS [56]. Resonance assignment and the structure of 51Z2 have been deposited in the BMRB and PDB databases (entries 18967 and 2M3L, respectively). The structural statistics are given in Table 1.

The calculated structural ensemble (Figure 3a) exhibits a backbone r.m.s.d. of 0.62 Å for the structured 51Z2 region (residues 80 to 140 of the full length E6) while the backbone r.m.s.d. drops to 2.37 Å when including the less ordered E6 C-terminus (residues 141-151, numbering according to full-length sequence). This C-terminus gave rise to only few NOEs consistent with a less ordered organization. Backbone torsion angles are located exclusively in most favored (85.3%) and additionally allowed (14.7%) regions of the Ramachandran plot (Table 1). The methyl groups of Ile88 show significantly upfield-shifted resonances (H$^\delta_1$: -0.628 ppm; H$^\delta_2$: -1.296 ppm). In the calculated structure, Ile88 H$^\delta_2$ is located directly above the aromatic ring of Trp132, while Ile88 H$^\delta_1$ is proximal to the aromatic ring of Phe125. Thus, ring current effects [57] explain the observed upfield-shift of these resonances and are consistent with the presented 51Z2 structure.

51Z2 exhibits a $\beta_1\alpha_1\beta_2\alpha_3\beta_4\beta_5\alpha_3$ topology (Figure 3a-c). All $\alpha$-helices are located on one side of 51Z2, while all $\beta$-sheets are juxtaposed on the opposite hemisphere of the domain (Figures 3b and 3c). The closest-to-mean structure of the well-folded 51Z2 portion is presented in Figures 3b and 3c. The N-terminal part of $\alpha_3$ carries regular helical geometry, while its C-terminal residues arrange as 3$\_10$ helix. One anti-parallel $\beta$-sheet on 51Z2 is composed of the strands $\beta_1$, $\beta_4$ and $\beta_5$, stabilized by charge-charge interactions and side chain hydrogen bonds, e.g. from R81 H$^\eta_1$ to E127 O$^{\varepsilon_1}$. A second short anti-parallel $\beta$-sheet consisting of $\beta_2$ and $\beta_3$ is arranged almost perpendicular to the first sheet. This allows for the formation of hydrogen bonds between R102 H$^N$ and W132 O, H104 H$^N$ to G134 O and G134 H$^N$ to R102 O which stabilize the arrangement of both $\beta$-strands (Figure 3d). A turn-like structural element, located between $\alpha_1$ and $\beta_2$, is stabilized by ionic interactions and the correlated formation of a hydrogen bond between the side chains of residues K94 H$^{\varepsilon_2}$ and D98 O$^{\delta_2}$ and additionally fixed by a hydrogen bond between L96 H$^N$ to E89 O$^{\varepsilon_1}$. A bi-dentate hydrogen bond between Q135 O$^{\varepsilon_1}$ and R105 H$^{\eta_1}$ as well as R105 H$^{\eta_{21}}$
2.3 Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction

to T143 stabilizes the spatial arrangement of the turn of the second β-sheet and the third α-helix (Figure 3e).

51Z2 contains a carboxy-terminal PDZ-BM and the full-length protein causes the degradation of hDlg in vitro [32]. Since HPV 16 and 18 E6 interact with hDlgPDZ2 and because the available structural hDlg-E6 interaction data contained only short E6 peptides of 6 or 7 residues [52,53,58] we investigated how 51Z2, representing a complete and wild-type domain of a high-risk HPV E6, interacts with hDlgPDZ2. An interaction is observed, as several chemical shifts of amide groups arising from the C-terminus of 51Z2 were perturbed in presence of hDlgPDZ2 (Figure 4a). Surprisingly, the perturbation affected the C-terminal nine E6 residues (143 to 151), more than anticipated from either the canonical PDZ-BM or from the already published hDlgPDZ2-E6 peptide complex structures [52,53]. From the gradual disappearance of resonances of the unbound 51Z2 and reappearance of the same number of resonances in the bound state of this domain we concluded that this interaction occurs within the slow exchange regime on the NMR time scale. Importantly, no further E6 resonances experienced chemical shift perturbation, clearly indicating that the interaction with the hDlgPDZ2 is confined to the disordered C-terminal region harboring the PDZ-BM of HPV51 E6. When titrated with an 11mer peptide representing the complete E6 disordered C-terminus (Ac-QRTRQRNETQV, corresponding to HPV 51 E6 residues 141 to 151, further referred to as E6CT11), the hDlgPDZ2 also showed significant spectral changes (Figure 4b). Moreover, hDlgPDZ2 bound to E6CT11 exhibited additional chemical shift perturbations when compared to hDlgPDZ2 in complex with an N-terminally truncated 6mer peptide (Ac-RNETQV, further referred to as E6CT6). In order to assess the affinity of the two different representative peptides of the E6 C-terminus, surface plasmon resonance (SPR) with hDlgPDZ2 and the E6CT11 and E6CT6 peptides, respectively, was performed. The SPR data indicate a contribution to binding of the additional residues as the E6CT11 binds with higher affinity to hDlgPDZ2 than the E6CT6 (Kd 9.6 µM versus 28.3 µM, respectively; Figure 5). Owing to fast association and dissociation (Figure 5 left panels), a kinetic analysis of the SPR data is more error-prone. The kinetically derived Kd values, however, are virtually identical to the values of the steady state analyses (Figure 5 right panels). The kinetic analysis reveals that the association is approximately 2-fold faster and dissociation 1/3rd slower for the E6CT11 as compared to the E6CT6, resulting in a total difference in affinity by a factor of 3.
In order to elucidate this difference in structural terms, the hDlgPDZ2-E6CT11 complex structure was determined exploiting stable isotope labeled E6CT11 (experimental details see Table S4). During the production of $^{13}\text{C}$ and $^{15}\text{N}$ labeled E6CT11 using the intein system, spontaneous cyclization of its amino-terminal glutamine to pyroglutamate was observed [59]. This covalent automodification, however, has no bearing on complex formation as shown earlier [60]. Completion of the resonance assignment and structure determination of hDlgPDZ2 complexed with the E6CT11 (for details see SI, BMRB entry 17942, [60] and PDB ID: 2M3M) allowed for the identification of hDlg residues additionally affected by the extended peptide (Figure 6a). In the final complex structure (Figures 6b and 6c), the most perturbed residues (backbone amide groups of A334, G 335, G336, H341, Y349, E385, T389 and side chain amide group of N339) are situated in a region of the PDZ domain that is close to the peptide binding region (Figure 6c). Only one additionally perturbed backbone amide of I353 is located far away from the E6CT11 binding region of hDlgPDZ2. Of particular note are the perturbed residues E385 and T389 of hDlgPDZ2, which are contacted by the amino-terminal peptide residues T143 and R144 that were lacking in the previously characterized hDlgPDZ2-E6 complexes [52,53].

The solution structure of the E6CT11 complexed hDlgPDZ2 (Table 1; Figures 6b and 6c) forms a compact domain with a $\beta_1\beta_2\beta_3\alpha_1\beta_4\beta_5\alpha_2\beta_6$ topology, in line with the topology of hDlgPDZ2 complexed with shorter peptides [52,53] and consistent with the common fold of PDZ domains [61]. The core of hDlgPDZ2 is formed by the side chains of a number of hydrophobic amino acids of the $\beta$-strands (I320, L322, L329, F331, I333, V350, L365, L371, L400, V402) and L391 emanating from the central $\alpha$-helix. In complex with the E6CT11 peptide, this $\alpha$-helix can be regarded as an anchoring element by presenting its charged amino acids E385 and K392, located at the $\alpha_2$ N- and C-terminus, respectively, into the solvent. Of the two peptide residues attaching to this anchor, residue Q150 is located in the C-terminal part of the peptide (E148-V151) essentially forming an additional b-strand ($\beta^*$) anti-parallel to strand $\beta_2$ of hDlgPDZ2 upon complexation. The second residue of the peptide, R142, could contribute to short-lived side chain charge-charge interactions with E385 of hDlgPDZ2. However, the R142 of E6 is located in the very N-terminus of the peptide. As deduced from the order parameters (Figure 6, inset), this entity experiences a higher motional freedom than the C-terminal part of the E6CT11 peptide which is
2.3 Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction

‘immobilized’ as integral part of the hDlgPDZ2 β-sheet architecture. For residues T143 and R144 of E6, however, well-defined side chain interactions to T389 and E385 of hDlgPDZ2 are present, consistent with the observed chemical shift perturbation for these residues (Figures 6a and 6c). The R146 side chain also contacts the PDZ domain at residues G338 and N339, the amide resonances of which are perturbed as well in the complex.

In conclusion, the main contribution of binding of the E6CT11 peptide to the hDlgPDZ2 is the formation of a stable and rigid additional β-strand. Overall, the peptide forms a kinked shape-complementary structure allowing its C-terminus to align anti-parallel to the β-strand 2 of the PDZ domain while the very N-terminus of the peptide is separated by a turn-like structure and thus a clash with the turn of the hDlgPDZ2 involving residues V337-N339 is prevented. Additional contacts of residues that are located N-terminal to the canonical PDZ-BM which is located on the β* strand establish further contacts between E6 and the hDlgPDZ2, involving charge complementarity: Arg 144 on E6 interacts with E 385 on hDlgPDZ2 (Figure 6c). These additional contacts are consistent with the increased affinity as seen in the SPR experiments (Figure 5).

Discussion

We set out to characterize a full length, wildtype E6 structure. The result of these efforts is the structure of the C-terminal zinc-binding domain (ZBD) of HPV 51 E6 (residues 80 to 151; 51Z2) and represents the first structural characterization of a non-mutated HPV E6 domain. The International Agency for Research on Cancer classifies HPV 51 as high-risk [62] and in HPV 51 infected cells, p53 and hDlg are degraded [32,63]. All wild-type, high-risk E6 constructs containing the amino-terminal ZBD tested here were prone to aggregation and insolubility. Dimerization and aggregation propensity of HPV 16 E6 resides mainly in the amino-terminal ZBD [48,50]. In combination with our solubility screening this strongly suggests that dimerization and/or aggregation mediated by the E6 amino-terminal domain may be a property shared among high-risk E6 proteins.

The hydrodynamic radius and secondary structure content estimated from circular dichroism spectroscopy of the wild-type 51Z2 are consistent with the final, monomeric 51Z2 structure. The reversibly disappearing 51Z2 backbone signals above 20 °C in vitro suggest protein motion at a timescale that causes line broadening at
higher temperatures, whereas the zinc coordinating residues remain rigid. These findings derived from a wild-type E6 domain allow us to suggest that in cervical epithelial cells, where the E6 protein is expressed at 37 °C, this particular domain and potentially the entire E6 protein may be more malleable *in vivo* than their published structures suggest at first glance. As the fold is re-adopted upon cooling down (Figure 1), it could be argued that the present low temperature conformation represents a low energy state and that this might be the conformation the C-terminal E6 domain adopts upon complex formation. The close structural match of 51Z2 to the corresponding domain of ligand-bound BPV E6 (Figure 7) supports this idea. The presence of E6 interactors such as hScrib and hDlg [64] or E6AP [65] stabilizes E6 *in vivo*, supporting our hypothesis that E6 interactors may stabilize an energetically favored E6 fold that might be less prone to proteolytic turnover *in vivo*. Consistent with this HPV 16 E6 interacts with p53 and experiences an increased stability *in vivo* in presence of commonly E6 binding LXXLL-containing peptides [66]. The malleability of E6 is reminiscent of intrinsically disordered proteins that undergo a disorder-to-order transition upon productive complex formation with specific ligands [67]. In the case of E6 this might be functional in the context of binding to the multitude of cellular E6 interaction partners [18] and further studies are needed to address the dynamic aspects of E6 plasticity (ZBD2; this paper) and dimerization (ZBD1; [50,51]) of wild-type E6.

A structural comparison of the unbound, wild-type 51Z2 to the corresponding unbound, four-fold mutated ZBD2 of HPV 16 E6 and to the evolutionarily distant bovine papilloma virus 1 (BPV) E6 in complex with the LD1 motif of paxillin reveals an identical general topology for E6 (Figure 7). Thus, it is reasonable to assume a similar fold for the corresponding domains of at least other high-risk or even all E6 proteins. To analyze this similarity in more detail, sequences of E6 shared by the HPV types for which there is reasonable evidence for their carcinogenicity [62] were aligned and conserved residues were identified (highlighted in Figure S3). In the following, residues are numbered according to their corresponding position in HPV 51 E6. Among the conserved residues, cysteines 103, 106, 136 and 139 coordinate the Zn\(^{2+}\) ion. Residues V83, L88, L96, L99, I101, L110, W132 and G134 form the E6 core. G85 constitutes the begin of the first α-helix of ZBD2. Residues S82, Y84, T87, R102, P109, P112, E114, K115, R124, H126, I128, T149 and V 151 are solvent exposed and prone to contribute to binding of cellular targets of E6. We
2.3 Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction

also note a hitherto unrecognized E6 sequence element involving the conserved residues P109 and P112 located on the E6 surface. This PXXP motif is present in all oncogenic E6 types (Figure S3) but rarely found in nononcogenic types (representative types see Figure S4; PXXP is present in HPV 7, 32, 40, 43, 91 E6). PXXP could be recognized by protein domains targeting proline-rich sequences (such as SH3; [68,69]). The conserved, charged E6 residues E114 and K115 in spatial proximity to PXXP could further enhance binding affinity and specificity of this PXXP motif as observed for other SH3 interactions [68].

Interestingly, a direct correlation between E6 phylogeny and their protein-protein interaction networks exists [70], i.e. the target spectrum of closely related E6 is more similar than that of evolutionary more distantly related E6. Thus, although the E6 proteins may share a generally identical structural topology, subtle structural variations could explain the altered target spectrum of different E6 proteins [70,71], strongly arguing in favor of structural characterization of further E6 proteins embedded in different interaction networks.

Our analysis allows the identification of significant local structural differences between individual E6 proteins. Particularly, β4 and β5 of 51Z2 position differently as compared to the HPV 16 E6Z2 domain, but similar to the corresponding region of the only available full-length E6 structure in a complex, the BPV E6 bound to LD1 of paxillin (Figure 7). The conserved I128 (Figure S3) is a key residue for the E6-E6AP interaction [71,72,73] and a single I128T E6 mutation in the genome of the high-risk HPV 31 causes viral episome loss after a few cell passages [74]. In the unbound 51Z2 I128 resides on β4 that orients different with respect to the HPV 16 E6 structure (vide supra). In 51Z2 I128 is proximal to the conserved residues H126 and S82 on β4 and β1, respectively. HPV 51 is, as HPV 16, frequently detected in premalignant neoplasias, but proportionally significantly less present in cervical cancers [4,75,76](Figure 8). In a three-stage model of carcinogenesis (initiation, promotion, progression), the later stages of cervical cancer progression are governed by E6 functions [77]. Therefore, the different E6 conformation in the region around I128, subtle as it is, could entail altered interaction properties and/or a change in relative positioning between E6 and LXXLL-containing proteins such as e.g. E6AP or paxillin. Since the E3 ubiquitin ligase E6AP is often involved in E6-mediated target degradation [22,23,78], an alteration of this particular interaction might have implications for the proteasomal degradation of other E6 interactors as well and may
influence the fate of cells infected by different HPV types. This invites the speculation that the subtle difference in the structures of HPV 16 and 51 E6 contribute to the different prevalence in high-grade cervical lesions of these two HPV types. Of course, other regions or properties of E6 or even differences between the respective E7 proteins are likely to be also relevant to the difference in oncogenicity of the two high-risk HPV types and to clarify this will require substantial but potentially fruitful future work.

51Z2 in the unbound state adopts a globular fold with a solvent exposed and flexible C-terminus (Figures 3 and S3) comprising the PDZ-BM of high-risk HPV, including the conserved T149 and V151 PDZ-BM key residues [31]. This structure provides for an E6 C-terminus accessible for binding to target PDZ domains and the HPV 51 E6 PDZ-BM binds to hDlgPDZ2 (Figures 4, 5 and 6). The structure of the E6CT11-bound hDlgPDZ2 domain is similar to the available structures of hDlgPDZ2 in complex with shorter peptides [52,53]. However, as a single mutation on a PDZ-BM dramatically alters viral virulence in a different context [37], a full analysis of all E6 residues involved in this PDZ binding appeared indicated. The previously available structures of hDlg PDZ domains complexed with E6-derived peptides contained the 4 most C-terminal residues forming the canonical E6 PDZ-BM and up to three additional residues [52,53]. Our analysis of the 51Z2 interaction with hDlgPDZ2 by way of SPR and subsequent structure determination reveals that the C-terminal 9 E6 residues contribute to the interaction. Importantly, presence of residues N-terminal to the canonical PDZ-BM significantly increase PDZ-binding affinity (Figure 5). Using a library of synthetic peptides it was shown that optimal substrate specificity and affinity of several PDZ-domains requires 9 residue peptides [79]. This study included all three PDZ domains of murine Dlg that carry a greater than 98% sequence identity with the hDlg PDZ domains. Contribution of E6 residues upstream of the canonical PDZ-BM was also observed for the MAGI1 PDZ1 interaction [80]. Yet in that system, 2 of 8 MAGI1 PDZ1 contacting E6 peptide residues interact with a region outside the canonical PDZ-domain. In our system, however, at least nine E6CT11 residues contact canonical PDZ-domain residues of hDlgPDZ2. Hence, ‘supernumeral’ residues outside the canonical E6 PDZ-BM establish specific contacts with PDZ domains in different ways be it via contacts to bona fide PDZ domain residues [this study] or to residues outside the classical PDZ domain [80]. Evidently, such interactions may fine-tune E6-PDZ interactions. Here, the E6CT11 and E6CT6
peptides bind with affinities of slightly below 10 µM and slightly above 28 µM to hDlgPDZ2. An affinity to PDZ-domains below 10 µM is exceptional; more often the affinity lies in the 10 to 100 µM range [81]. Thus, the charge complementarity (and additional contacts) of the E6CT11 guarantees for high affinity, which could allow E6 to successfully compete with cellular molecules for hDlg binding. As basic residues upstream to the canonical PDZ-BM are present on all high-risk E6 proteins (though they are not conserved position-specifically; Figure S3), the charge complementarity could be a general property of high-risk E6 proteins for binding to certain PDZ domains. Moreover, since viruses often target PDZ-domains [35,36], the ‘affinity boost’ by charge complementarity could also be one mechanism of other viral PDZ-targeting proteins to be ‘better binders’ than their competing cellular proteins.

A crystal structure of the C-terminal 11 residues of APC (APC_CT11) in complex with hDlgPDZ2 reveals only one peptide per asymmetric unit containing in turn five PDZ domains. Here, only the C-terminal six APC peptide residues gave rise to interpretable electron density [82]. The non-visible peptide residues, however, apparently contributed to affinity in that complex. The hDlgPDZ2-E6CT11 structure might explain the findings of the hDlgPDZ2-APC_CT11 complex of [82]: the increased flexibility observed for the N-terminus of E6CT11 in the hDlgPDZ2 complex might also apply to the APC_CT11-hDlgPDZ2 complex, which in turn may have compromised a complete X-ray structural analysis of the APC peptide residues.

In conclusion, here we provide a virtually complete structural rational for extended PDZ-BM PDZ interactions for the first time. Based on this, it is tempting to speculate that the mechanism of PDZ-domain binding with residues upstream of the canonical PDZ-BM may be a common strategy of the PDZ-BM harboring high-risk E6 proteins to increase substrate affinity and specificity. Moreover, interaction with an extended PDZ-BM may be a common mechanism of at least a subset of PDZ-domains [61]. For example, PDZ domains apparently employ long-range networks for fine-tuning their binding selectivity [83] and two mutations, one of which is located far away from the peptide interface, in a PDZ domain confer altered binding specificities to this domain [84]. The single hDlgPDZ2 residue I353 affected differentially after binding of E6CT11 versus E6CT6 and located far away from the peptide interface might have been perturbed due to the presence of such a long range network. This raises the question if the HPV 51 E6 extended PDZ-BM may even induce changes in places remote from the PDZ 2 of hDlg as the PDZ domains 1 and 2 of the hDlg-related PSD-
are organized as a supramolecule and retain their relative orientation in the context of the full-length protein [85]. Since PDZ1 and 2 of hDlg are, as the domains of PSD-95, connected by a short linker, such a supramolecular organization can also be envisioned for hDlg PDZ1 and 2 as there is evidence that this PDZ1-2 region forms a single conformational and functional unit [86]. The additional E6 residues that contribute to PDZ2 binding do not reach into the hypothetical hDlg PDZ1-PDZ2 interface and would thus not interfere directly with formation of a hDlg PDZ1-2 supramolecule as deduced from applying the PSD-95 PDZ1-2 structural model [85] to hDlg here (not shown). Hence it should be interesting to further investigate the E6–hDlg interaction by using extended hDlg constructs to cover aspects of long range networks and of the PDZ supramolecular organization.

In summary, the first structure of a wild-type E6 domain and of the complex between its complete PDZ-BM and the PDZ domain 2 of hDlg revealed a number of E6 properties common for high-risk HPV E6 but also important differences between individual E6. The fully characterized interaction of E6’s extended PDZ-BM to one PDZ domain suggests future routes towards the elucidation of PDZ-supramolecule E6 interaction. The implications of the E6 structural variations and the impact of the probable malleability of E6 in vivo with regard to interactions, such as e.g. with E6AP, may be addressed based on our findings in order to solve the puzzle of how the E6 proteins exert their malignant potential via interaction with a multitude of cellular targets.

Materials and Methods

A detailed description of methods is provided in Supporting Text 1. Briefly, recombinant proteins were generated in E. coli BL21 (DE3). Cells were disrupted and protein solubility in cell-lysates was assessed. Soluble proteins were FPLC-purified and initially characterized by CD spectroscopy, dynamic light scattering, gel filtration chromatography and NMR spectroscopy. The affinity of E6-derived peptides to hDlgPDZ2 was elucidated utilizing Surface Plasmon Resonance. hDlgPDZ2 was generated by bacterial expression and FPLC purification. The generation of the labeled E6CT11 peptide was performed as previously described [59]. The purified, monomeric 51Z2 and the purified, monomeric hDlgPDZ2 in complex with purified E6CT11 were subsequently structurally characterized by standard solution state NMR techniques.
Acknowledgements

We thank Dr. J. Leppert and A. Heller (both at the Fritz Lipmann Institute) for technical support, Prof. Dr. M. Dürst (Jena University Hospital) for the provision of several E6 plasmids and Dr. S. Rothemund (IZKF Leipzig) for the synthesis of unlabelled peptides. A. M. was funded by the Leibniz Graduate School on Ageing (LGSA). The Fritz Lipmann Institute (FLI) and the Hans Knöll Institute (HKI) are members of the Science Association ‘Gottfried Wilhelm Leibniz’ (WGL) and are financially supported by the Federal Government of Germany and the State of Thuringia.
2.3 Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction

References

2.3 Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction

2.3 Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction


Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction

2.3 Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction

J Biomol NMR 31: 129-141.


2.3 Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction

Figures (including figure legends)

Figure 1. Temperature sensitivity of the HPV 51 C-terminal zinc-binding domain 51Z2.

Nine $[^1H-^{15}N]$-HSQC spectra of an 51Z2 sample were recorded with identical spectral parameters and only the temperature was increased as indicated in the spectral plots. Residues that were still observable at 45 °C are indicated in the 45 °C spectrum. The assignment for the 45 °C signals was based on the assignment for 10 °C sample temperature (Figure 2) by tracking peak positions with increasing temperatures. Tentative assignments which were not unambiguous are given in parentheses. One control shows the 51Z2 spectrum after returning to the start temperature.
2.3 Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction

Figure 2. Resonance assignment of 51Z2.

$[^1\text{H},^{15}\text{N}]$-HSQC of 1.1 mM 51Z2 with indicated assignments and residue numbering according to full-length E6 sequence. Side chain signals are labeled with SC, and pairs of SC signals are linked by horizontal lines. Two starred resonances: folded signals probably representing arginine-side chains that could not unambiguously be assigned. Experimental details for this spectrum can be found in Table S3. The assignment has been deposited in the BMRB (Accession number: 18967).
2.3 Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction

Figure 3. Solution structure of the C-terminal zinc-binding domain of HPV 51 E6 (51Z2).

The bound zinc ion is represented as grey sphere. A Stereo view of the 51Z2 bundle of the 20 structures with the lowest energy after CNS refinement. B Ribbon view of the structural ensemble as in A with labeled secondary structure elements. The less ordered C-terminal residues 141-151 are omitted for clarity. C represents the rotated ensemble of B. D Backbone hydrogen bonds between residues on β2 and β5 strands stabilize the arrangement of both 51Z2 β-sheets. Side chains omitted for clarity. E The bidentate H-bond involving Arg105, Gln135 and Thr143 stabilizes the C-terminal α3 helix. The coordinates of 51Z2 have been deposited (PDB: 2M3L).
2.3 Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction

Figure 4. Interaction of 51Z2 with hDlgPDZ2.

100 µM of $^{15}$N labeled sample without (blue) or with (red) 3 fold excess of unlabeled interaction partner. A Labeled 51Z2 and unlabeled hDlgPDZ2. Only the spectral region with perturbed resonances is shown for clarity. The signal intensity for the nine C-terminal 51Z2 residues as well as resonances of side chains N147 and Q150 is diminished and a corresponding number of new signals is observed in presence of hDlgPDZ2 (arrowheads). The assignments (BMRB entry 18967) of perturbed residues on free 51Z2 are indicated. B labeled hDlgPDZ2 and unlabeled E6CT11, derived from the flexible 51Z2 C-terminus including the E6 residues perturbed upon binding. Resonance assignments of perturbed residues of hDlgPDZ2 are not indicated for clarity, since virtually all resonances show chemical shift differences. For resonance assignment of the hDlgPDZ2 complexed with the E6CT11, see ([60]; BMRB entry 17942).
2.3 Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction

Figure 5. 51Z2 derived E6CT6 peptide versus E6CT11 peptide binding to hDlgPDZ2.

SPR data. A E6CT6, B E6CT11. Sensorgrams of E6CT6 and E6CT11 binding injected in triplicate (black lines) are shown overlaid with the best fit derived from a 1:1 interaction model including a mass transport term (orange lines). Peptide concentrations of 3.125, 6.25, 12.5, 25, 50, 100 and 200 µM are shown. The binding parameters were obtained by kinetic analysis of association and dissociation phases (left panels) or by steady state analysis (right panels) utilizing signals of plateaus depicted in the corresponding left panel.
2.3 Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction

Figure 6. Interaction of E6CT11 with hDlgPDZ2.

A Combined $^1$H and $^{15}$N chemical shift perturbation (as detailed in SI) of 100 µM hDlgPDZ2 in complex with 300 µM E6CT11 peptide versus 300 µM E6CT6 peptide. Residues without observable amide shifts are denoted with X. The inset of a region of the corresponding HSQC spectra show unperturbed as well as perturbed signals. Red contours: hDlgPDZ2 complexed with E6CT11, blue contours: hDlgPDZ2 complexed with E6CT6. Note that the side chain amide signals of Asn339 were also perturbed by more than 2x the average value.

B Structure of the hDlgPDZ2-E6CT11 complex. The bundle of 20 best E6CT11 structures (residues 141 to 151, dark grey) is shown together with a ribbon of the closest-to-mean hDlgPDZ2 structure (hDlg residues 318-406). Peptide structures were
fitted to residues 143 to 151 and the termini are indicated. Secondary structure elements are labeled. C Details of the hDlgPDZ2-E6CT11 complex. hDlgPDZ2 backbone trace depicted in light gray. PDZ side chains (heavy atoms) of residues showing most perturbed combined amide group chemical shifts (backbone and Asn339 side chain; Figure 6a) are depicted in green and labeled, while the closest-to-mean E6CT11 peptide structure (heavy atoms, residues 143-151) is presented in dark gray. The boxed inset depicts per-residue backbone order parameters of the complexed E6CT11 peptide. D Schematic depiction of intermolecular hydrogen bonds and salt bridges in the closest-to-mean complex structure. Indicated side-chains start at the Cβ atom. Hydrogen bonds are indicated as dashed lines. Secondary structure elements β* and β2 are emphasized by arrows; hDlgPDZ2 residues appear gray, while peptide residues are depicted in black.
Superimposition (details: see SI) of the 51Z2 closest-to mean structure (folded part, residues 80-140, blue) onto the corresponding regions of HPV 16 E6 (red, PDB ID 2LJZ, r.m.s.d. 2.27 Å) and BPV E6 in complex with the LD1 motif of paxillin (gray, PDB ID 3PY7, paxillin omitted for clarity, r.m.s.d. 1.61 Å). The overall topology is conserved. Notably, the b4 and b5 strands and their connecting loop of 51Z2 and BPV position similar, while for HPV 16 E6, the corresponding region orients differently (upper right corner; encircled and highlighted).
Figure 8. Comparison of HPV 51 and HPV 16 prevalence.

Meta-analysis of prevalence of HPV 51 (blue) and HPV 16 (red) in asymptomatic epithelia (AE; [75]) and in (pre-)cancerous stages low-grade squamous intraepithelial lesions (LSIL; [76]), high-grade squamous intraepithelial lesions (HSIL; [4]) and squamous cell carcinoma (SCC; [4]). While the fraction of HPV 16 increases with severity of neoplasia, the HPV 51 fraction decreases after the LSIL stage.
2.3 Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction

Tables

**Table 1. Statistics for the 20 best 51Z2 and hDlgPDZ2-E6CT11 peptide complex structures.a**

<table>
<thead>
<tr>
<th>Structure</th>
<th>51Z2</th>
<th>hDlgPDZ2-E6CT11</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOE-based distance restraints</td>
<td>1501</td>
<td>2760</td>
</tr>
<tr>
<td>intra-residual (</td>
<td>i-j</td>
<td>=0)</td>
</tr>
<tr>
<td>sequential (</td>
<td>i-j</td>
<td>=1)</td>
</tr>
<tr>
<td>medium-range (2≤</td>
<td>i-j</td>
<td>≤4)</td>
</tr>
<tr>
<td>long-range (</td>
<td>i-j</td>
<td>≥5)</td>
</tr>
<tr>
<td>inter-chain</td>
<td>not applicable</td>
<td>162</td>
</tr>
<tr>
<td>Zinc geometry constraints (upper and lower)</td>
<td>2 x 9</td>
<td>not applicable</td>
</tr>
<tr>
<td>Restrained H-bonds</td>
<td>15</td>
<td>32</td>
</tr>
<tr>
<td>Dihedral angle restraints</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Φ/Ψ</td>
<td>71/73</td>
<td>100/99</td>
</tr>
<tr>
<td>3 J HNHA coupling derived</td>
<td>63</td>
<td>97</td>
</tr>
<tr>
<td>χ1</td>
<td>60</td>
<td>78</td>
</tr>
<tr>
<td>Leu χ2</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Ile χ21</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>CYANA target function</td>
<td>0.26 - 1.39</td>
<td>0.81 - 1.58</td>
</tr>
<tr>
<td>R.m.s.d. after water refinement [Å]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totalc</td>
<td>2.991</td>
<td>1.433</td>
</tr>
<tr>
<td>Ordered (backbone)c</td>
<td>0.623</td>
<td>0.496</td>
</tr>
<tr>
<td>Ordered (heavy atoms)c</td>
<td>1.255</td>
<td>0.964</td>
</tr>
<tr>
<td>Ramachandran plot [%]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>most-favored</td>
<td>85.3</td>
<td>82.6</td>
</tr>
<tr>
<td>additionally allowed</td>
<td>14.7</td>
<td>16.3</td>
</tr>
<tr>
<td>generously allowed</td>
<td>0</td>
<td>1.2</td>
</tr>
<tr>
<td>disallowed</td>
<td>0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

a: Methods detailed in the Supplementary Information.

b: including the intermolecular (inter-chain) distance restraints.
c: Residue numbers 80 to 151 refer to position of the full-length HPV 51 E6 protein (UniProtKB entry: P26554), while residues 318 to 406 refer to positions of the full length hDlg protein (UniProtKB entry: Q12959). Flexible non-native residues (amino-terminal GSHM of 51Z2 and carboxy-terminal His_6-Tag of hDlgPDZ2) were not included in this structural statistics. For 51Z2 r.m.s.d. calculations include residues 80-151 or 80-140, respectively (see text). For the hDlgPDZ2-E6CT11 complex, r.m.s.d. calculations include residues 318-406 (hDlg) and 141-151 (E6) or 318-406 (hDlg) and 143-151 (E6), respectively.
2.3 Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction

Supporting Information - Figures

Figure S1. Spectral changes of 26Z2 over time.

Freshly prepared $^{15}$N-labeled 26Z2 (250 µM) was subjected to $[^1H,^{15}N]$-HSQC NMR spectroscopy (blue contours). After 10 days at 4 °C, the spectra (red contours) showed significant differences. Low peak-dispersion suggests an increased proportion of unfolded protein. Both spectra were recorded with 16 scans at a Bruker Avance III 750 MHz NMR spectrometer. Sample conditions were 135 mM NaCl, 45 mM L-Arg, 45 mM L-Glu, 9 mM DTT, pH 7.4, 4 °C.
Figure S2. Biophysical characterization of 51Z2.

For experimental details, see supplementary text 1. 

A Analytical gel-filtration. The chromatogram of 51Z2 run on a TSK gel G3000SWxl column is presented in blue, the column calibration is shown in orange with molecular weights of reference proteins indicated. 51Z2 (calculated MW 8.9 kDa) eluted as a 9 to 10 kDa sized protein indicating a monomeric state. 

B Dynamic light scattering. 51Z2 exhibits a hydrodynamic radius of 1.58 nm, which corresponds to an approx. 10 kDa sized protein assuming a globular shape. 

C Circular dichroism spectrum of purified 51Z2. The secondary structure content of 51Z2 was estimated from the CD spectrum using CDNN [87].
ClustalW2 [88] was utilized for alignment of the E6 proteins from oncogenic/possibly oncogenic HPV types (according to IARC, [62]). The figure was prepared using jalview [89]. Residues with a jalview-implemented conservation score [90] of 9 or higher were colored in green. The oncogenic HPV types [62] phylogenetically belong to the genus alpha-papillomaviridae and to the species indicated on the left table-side.
### Figure S4. Alignment of the high-risk HPV 51 E6 to representative low-risk and cutaneous E6 proteins.

The residues corresponding to conserved positions among high-risk types (Figure S3) are bracketed. The bracket color is green, if any residue present among the high-risk E6 proteins at that position is encountered (Figure S3) or red, if the residue is never observed at the respective position of any high-risk E6 protein.
### Supporting Information - Tables

#### Table S1. E6 domain architecture and solubility of recombinant E6 constructs.

<table>
<thead>
<tr>
<th>HPV type</th>
<th>E6FL</th>
<th>E6Z1</th>
<th>E6Z2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a (co)</td>
<td>No expression</td>
<td>No expression</td>
<td>Insoluble</td>
</tr>
<tr>
<td>16 (co)</td>
<td>Insoluble</td>
<td>Insoluble</td>
<td>Partially soluble</td>
</tr>
<tr>
<td>18</td>
<td>Insoluble</td>
<td>Insoluble</td>
<td>Partially soluble</td>
</tr>
<tr>
<td>26</td>
<td>Insoluble</td>
<td>Insoluble</td>
<td>Soluble</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Residue</th>
<th>1</th>
<th>79</th>
<th>80</th>
<th>151</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ZBD1</td>
<td>ZBD2</td>
<td>PDZ-BM</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E6FL</th>
<th>E6Z1</th>
<th>E6Z2</th>
</tr>
</thead>
<tbody>
<tr>
<td>His&lt;sub&gt;6&lt;/sub&gt;</td>
<td>His&lt;sub&gt;6&lt;/sub&gt;</td>
<td>His&lt;sub&gt;6&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

**HPV type**
- 1a (co)
- 16 (co)
- 18
- 26
### 2.3 Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction

<table>
<thead>
<tr>
<th>Residue</th>
<th>Solubility</th>
<th>Residue</th>
<th>Solubility</th>
<th>Residue</th>
<th>Solubility</th>
</tr>
</thead>
<tbody>
<tr>
<td>31 (co)</td>
<td>Insoluble</td>
<td>Insoluble</td>
<td>Partially soluble</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>Insoluble</td>
<td>Insoluble</td>
<td>Partially soluble</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51(co)</td>
<td>Insoluble</td>
<td>Insoluble</td>
<td>Soluble</td>
<td></td>
<td></td>
</tr>
<tr>
<td>59 (co)</td>
<td>Insoluble</td>
<td>Insoluble</td>
<td>No expression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>97 (co)</td>
<td>Insoluble</td>
<td>Insoluble</td>
<td>Partially soluble</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Top: E6 domain architecture. Residue numbers correspond to full-length HPV 51 E6 and for each construct zinc-binding domain (ZBD) borders were chosen by alignment to HPV 16 E6 according to [49]. Recombinant constructs were expressed in *E. coli* BL21(DE3) with an N-terminal His<sub>6</sub>-tag linked to full-length (E6FL) or individual zinc-binding domains (E6Z1 or E6Z2), respectively. PDZ-BM denotes the PDZ-binding motif present in high-risk HPV derived E6 and missing in HPV 1a E6. Bottom: Solubility of recombinant E6 constructs in *E. coli* cell extract. Coomassie stained SDS-PAGE slices of equivalent amounts of (T) total, (S) soluble and (I) insoluble E6 containing fractions. E6 bands are indicated by (>). Codon-optimized genes for expression in *E. coli* are denoted with (co).
2.3 Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction

Table S2. Final screening results of E6 constructs for NMR spectroscopy.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Monodispersity</th>
<th>HSQC</th>
<th>Long-term stability</th>
</tr>
</thead>
<tbody>
<tr>
<td>51Z2</td>
<td>Passed</td>
<td>Passed</td>
<td>Passed</td>
</tr>
<tr>
<td>26Z2</td>
<td>Passed</td>
<td>Passed</td>
<td>Failed</td>
</tr>
<tr>
<td>18Z2</td>
<td>Passed</td>
<td>Failed</td>
<td></td>
</tr>
<tr>
<td>45Z2</td>
<td>Passed</td>
<td>Failed</td>
<td></td>
</tr>
<tr>
<td>16Z2</td>
<td>Failed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31Z2</td>
<td>Failed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>97Z2</td>
<td>Failed</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Purified, soluble E6 constructs (compare Table S1) were screened for monodispersity by gel-filtration and dynamic light scattering. Monodisperse constructs always behaved as monomers. For those constructs, $[^{1}H,^{15}N]$-HSQC were recorded and assessed. This test was considered as passed when the number of amino acids in the construct was matching with the number of observable amide resonances and the resonances in turn exhibited similar peak intensities. Long-term stability was considered as passed when no spectral changes occurred after two weeks post acquisition of the first $[^{1}H,^{15}N]$-HSQC.

Table S3. Details of NMR experiments and samples for the 51Z2 structure determination.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Data points, NS (tmix)</th>
<th>Sample / Solvent</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[^{1}H,^{15}N]$-HSQC</td>
<td>2048/256, 4</td>
<td>1.13 mM in H$_2$O</td>
</tr>
<tr>
<td>HNCACB</td>
<td>2048/64/160, 32</td>
<td>1.13 mM in H$_2$O</td>
</tr>
<tr>
<td>CC(CO)NH</td>
<td>2048/96/168, 16</td>
<td>1.13 mM in H$_2$O</td>
</tr>
<tr>
<td>H(CCCO)NH</td>
<td>2048/64/128, 32</td>
<td>1.13 mM in H$_2$O</td>
</tr>
<tr>
<td>HNCO</td>
<td>2048/64/128, 8</td>
<td>1.13 mM in H$_2$O</td>
</tr>
<tr>
<td>HNFA</td>
<td>2048/128/96, 16</td>
<td>1.13 mM in H$_2$O</td>
</tr>
<tr>
<td>HNCA CO</td>
<td>2048/64/128, 32</td>
<td>1.13 mM in H$_2$O</td>
</tr>
<tr>
<td>$[^{1}H,^{13}C]$-HSQC (ali)</td>
<td>2048/96/280, 16 (180 ms)</td>
<td>1.13 mM in H$_2$O</td>
</tr>
<tr>
<td>$[^{1}H,^{13}C]$-HSQC (ali)</td>
<td>1024/128, 16</td>
<td>1.13 mM in H$_2$O</td>
</tr>
<tr>
<td>H(C)CH-TOCSY</td>
<td>2048/128/144, 8 (12 ms)</td>
<td>0.90 mM in D$_2$O</td>
</tr>
<tr>
<td>H(C)CH-COSY</td>
<td>2048/128/128, 16</td>
<td>0.90 mM in D$_2$O</td>
</tr>
<tr>
<td>$[^{1}H,^{13}C]$-NOESY- HSQC (ali)</td>
<td>2048/192/256, 8 (120 ms)</td>
<td>0.90 mM in D$_2$O</td>
</tr>
<tr>
<td>$[^{1}H,^{13}C]$-HSQC (aro)</td>
<td>1024/256, 4</td>
<td>0.90 mM in D$_2$O</td>
</tr>
<tr>
<td>HDBC(CGC)H HD</td>
<td>1024/128, 448</td>
<td>0.90 mM in D$_2$O</td>
</tr>
<tr>
<td>(HDBC)(GCDCE)HE</td>
<td>1024/128, 512</td>
<td>0.90 mM in D$_2$O</td>
</tr>
<tr>
<td>H(C)CH-COSY aro.</td>
<td>2048/60, 64</td>
<td>1.00 mM in D$_2$O</td>
</tr>
<tr>
<td>$[^{1}H,^{13}C]$-NOESY- HSQC (aro)</td>
<td>2048/80/1024, 8 (120 ms)</td>
<td>1.00 mM in D$_2$O</td>
</tr>
</tbody>
</table>

Bruker AvanceIII NMR spectrometers with $^{1}H$ resonance frequencies of 750 MHz or 600 MHz were utilized. All experiments were performed with $^{13}$C and $^{15}$N-labeled 51Z2 and the respective implementations of pulse programs in the TOPSPIN v. 2.1 software bundle. Resolution, number of scans (NS) and, where appropriate, mixing times ($t_{mix}$) are given. Protein concentration was determined spectrophotometrically using a molar extinction coefficient of 13980 M$^{-1}$ cm$^{-1}$ as calculated from the amino acid sequence by the Protparam tool [91] at www.expasy.org/protparam/. Spectra were recorded at 283 K. 51Z2 buffer conditions were 90 mM NaCl, 45 mM L-arginine, 45 mM L-glutamate, 9 mM DTT, 0.05 % (w/v) NaN$_3$, pH 7.4 in 90% H$_2$O/10% D$_2$O or in 100% D$_2$O.
2.3 Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction

Table S4. Details of NMR experiments and samples for the hDlgPDZ2-E6CT11 complex structure determination.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Data points, NS (t&lt;sub&gt;mix&lt;/sub&gt;)</th>
<th>Solvent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong> 800 µM labeled hDlgPDZ2 + 2000 µM unlabeled E6CT11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[H-15N]-HSQC</td>
<td>2048/512, 32</td>
<td>H2O</td>
</tr>
<tr>
<td>HNCACB</td>
<td>2048/64/160, 32</td>
<td>H2O</td>
</tr>
<tr>
<td>CC(CO)NH</td>
<td>2048/64/160, 32</td>
<td>H2O</td>
</tr>
<tr>
<td>H(CCCO)NH</td>
<td>2048/64/160, 32</td>
<td>H2O</td>
</tr>
<tr>
<td>HNCO</td>
<td>2048/64/128, 16</td>
<td>H2O</td>
</tr>
<tr>
<td>HNHA</td>
<td>2048/128/48, 16</td>
<td>H2O</td>
</tr>
<tr>
<td>H(CCCO)CO</td>
<td>2048/64/128, 32</td>
<td>H2O</td>
</tr>
<tr>
<td>[H-15N]-NOESY-HSQC</td>
<td>2048/112/280, 8 (120 ms)</td>
<td>H2O</td>
</tr>
<tr>
<td>[H-15N]-NOESY-HSQC, f1 filtered (intermol. NOEs)</td>
<td>2048/64/128, 32 (120 ms)</td>
<td>H2O</td>
</tr>
<tr>
<td>[H-13C]-HSQC (ali)</td>
<td>1024/512, 32</td>
<td>D2O</td>
</tr>
<tr>
<td>[H-13C]-HSQC (ali)</td>
<td>2048/512, 32</td>
<td>D2O</td>
</tr>
<tr>
<td>H(C)CH(TOCSY)</td>
<td>2048/128/160, 8 (12 ms)</td>
<td>D2O</td>
</tr>
<tr>
<td>H(C)CH(COSY)</td>
<td>2048/128/160, 8</td>
<td>D2O</td>
</tr>
<tr>
<td>[H-13C]-NOESY-HSQC</td>
<td>2048/192/208, 8 (120 ms)</td>
<td>D2O</td>
</tr>
<tr>
<td>[H-13C]-NOESY-HSQC, f1 filtered (intermol. NOEs)</td>
<td>2048/64/128, 32 (120 ms)</td>
<td>D2O</td>
</tr>
<tr>
<td>[H-13C]-HSQC (aro)</td>
<td>2048/96, 128</td>
<td>D2O</td>
</tr>
<tr>
<td>H(B)CH(CGD)HD</td>
<td>1024/72, 512</td>
<td>D2O</td>
</tr>
<tr>
<td>[H-13C]-NOESY-HSQC (aro)</td>
<td>2048/44/288 (120 ms)</td>
<td>D2O</td>
</tr>
</tbody>
</table>

<p>| <strong>B</strong> 1250 µM labeled E6CT11 + 3122 µM unlabeled hDlgPDZ2 | | |</p>
<table>
<thead>
<tr>
<th>Experiment</th>
<th>Data points, NS (t&lt;sub&gt;mix&lt;/sub&gt;)</th>
<th>Solvent</th>
</tr>
</thead>
<tbody>
<tr>
<td>[H-15N]-HSQC</td>
<td>2048/128, 8</td>
<td>H2O</td>
</tr>
<tr>
<td>HNCACB</td>
<td>2048/48/118, 32</td>
<td>H2O</td>
</tr>
<tr>
<td>CC(CO)NH</td>
<td>2048/36/152, 16</td>
<td>H2O</td>
</tr>
<tr>
<td>H(CCCO)NH</td>
<td>2048/36/112, 16</td>
<td>H2O</td>
</tr>
<tr>
<td>HNCO</td>
<td>2048/48/128, 16</td>
<td>H2O</td>
</tr>
<tr>
<td>HNHA</td>
<td>2048/80/32, 32</td>
<td>H2O</td>
</tr>
<tr>
<td>H(CCCO)CO</td>
<td>2048/48/128, 32</td>
<td>H2O</td>
</tr>
<tr>
<td>[H-15N]-NOESY-HSQC</td>
<td>2048/24/160, 48 (120 ms)</td>
<td>H2O</td>
</tr>
<tr>
<td>[H-15N]-NOESY-HSQC, f1 filtered (intramol. NOEs)</td>
<td>2048/24/96, 72 (120 ms)</td>
<td>H2O</td>
</tr>
<tr>
<td>[H-13C]-HSQC (ali)</td>
<td>1024/448, 32</td>
<td>H2O</td>
</tr>
<tr>
<td>[H-13C]-HSQC (ali)</td>
<td>1024/384, 64</td>
<td>D2O</td>
</tr>
<tr>
<td>H(C)CH(TOCSY)</td>
<td>2048/128/160, 8 (12 ms)</td>
<td>D2O</td>
</tr>
<tr>
<td>H(C)CH(COSY)</td>
<td>2048/128/128, 8</td>
<td>D2O</td>
</tr>
<tr>
<td>[H-13C]-NOESY-HSQC</td>
<td>2048/128/128, 16 (120 ms)</td>
<td>D2O</td>
</tr>
<tr>
<td>[H-13C]-NOESY-HSQC, f1 filtered (intermol. NOEs)</td>
<td>2048/96/96, 40 (120 ms)</td>
<td>D2O</td>
</tr>
<tr>
<td>[H-13C]-NOESY-HSQC, f1 filtered (intramol. NOEs)</td>
<td>2048/96/96, 32 (120 ms)</td>
<td>D2O</td>
</tr>
</tbody>
</table>

Bruker AvanceIII NMR spectrometers with 1H resonance frequencies of 750 MHz or 600 MHz were utilized. All experiments were performed with 13C and 15N-labeled 51Z2 and the respective implementations of pulse programs in the TOPSPIN v. 2.1 software bundle. The experiments were performed with one interaction partner in 13C and 15N-labeled form and the other, unlabeled interaction partner in 2.5-fold excess. A the PDZ domain was labeled B the peptide was labeled. Resolution, number of scans (NS) and, where appropriate, mixing times (t<sub>mix</sub>) of experiments are given. hDlgPDZ2 concentration was determined spectrophotometrically using a molar extinction coefficient of 2980 M<sup>-1</sup> cm<sup>-1</sup> as calculated from the amino acid sequence by the ProtParam tool [91] at www.expasy.org/protparam/. Peptide concentration was determined utilizing the peptide bond absorbance [92]. Spectra were recorded at 293 K. Buffer conditions for all experiments were 20 mM sodium phosphate, 4 mM TCEP, 0.05 % (w/v) NaN<sub>3</sub>, pH 6.5 in 90% H<sub>2</sub>O/10% D<sub>2</sub>O or in 100% D<sub>2</sub>O.
Supplementary text 1: Methods

DNA constructs for recombinant HPV E6 expression
Vectors encoding full-length E6 proteins of HPV 18, 26 and 45 were kindly provided by Matthias Dürst (Jena University Hospital). Vectors encoding codon-optimized full-length E6 reading frames flanked by NdeI/BclI restriction sites for cloning purposes and for recombinant expression in *E. coli* were obtained from DNA2.0. E6 constructs were generated as detailed in Table S1, spanning the full-length E6 protein or the amino- or carboxy-terminal part of E6, respectively.

Vectors encoding full-length codon-optimized E6 genes were transformed into chemically competent *E. coli* GM2163 (dam’/dcm’) and after amplification, DNA was isolated using the QIAGEN Plasmid Midi Kit. This DNA was digested with NdeI/BclI (New England Biolabs, NEB) following manufacturer’s instructions.

All other inserts were generated by amplification of desired E6 regions by PCR. A typical PCR protocol consisted of 2 min initial denaturation, 35 cycles of 30 s denaturation (94 °C), 45 s annealing (57 °C) and 70 s elongation (72 °C) and a final elongation step for 5 min at 72 °C. 100 µl PCR reactions were carried out and contained 2 µM of each primer, 400 µM dNTPs each and 2 U *Pwo* polymerase with the corresponding polymerase buffer (PEQLAB). Forward primers for the desired protein fragment contained 5’-extensions with the NdeI cleavage site followed by E6 sequences. Reverse primers were designed to match the desired E6 region and extended to provide a stop codon followed by a BclI or BamHI cleavage site. The generated PCR fragments were run on agarose gels, excised and purified using the QIAquick Gel Extraction Kit (QIAGEN). These DNAs were then digested with NdeI and BamHI or with NdeI and BclI to obtain the desired sticky ends. The inserts were ligated with T4 DNA ligase (Life Technologies) into the pET15b vector (50 ng) predigested with NdeI and BamHI. 1 to 10 µL of each ligation mixture was transformed into *E. coli* DH5α and DNA from several colonies was isolated by QIAprep Spin Miniprep Kit (QIAGEN). Sequence identity of His<sub>6</sub>-tagged E6 constructs was verified by sequencing (Eurofins MWG).

Expression and solubility screen of E6 constructs
Cells transformed with a vector encoding the E6 construct in question were grown to 0.7 - 0.8 OD<sub>600</sub> at 37 °C with LB medium, transferred to modified M9 medium (see
below) containing 10 µM Zn\(^{2+}\) and were induced after 30 min with 0.4 mM ITPG. Proteins were expressed overnight at 20 °C (~16h). Solubility was assessed after disruption of cells in suspension in lysis buffer (see below) by FRENCH® Press (for details see next section), subsequent centrifugation (10,000 x g, 4 °C, 30 min) and resuspending the pellet via ultra-sonication in a volume equal to the volume of the corresponding supernatant (see Table S1 for results).

**Expression and purification of soluble E6 proteins**

All E6 protein constructs cloned into pET15b were expressed in *E. coli* BL21 (DE3) using 100 mg/L ampicillin for selection. Freshly transfected cells were grown to an OD\(_{600}\) of 0.7 to 0.8 in 1L LB medium at 37 °C. Cells were harvested, washed and transferred to 250 mL modified M9 medium (6 g/L Na\(_2\)HPO\(_4\), 4 g/L D-glucose, 3 g/L KH\(_2\)PO\(_4\), 1 g/L \(^{15}\)NH\(_4\)Cl, 0.5 g/L NaCl, 2 mM MgSO\(_4\), 0.1 mM CaCl\(_2\), 10 µM ZnSO\(_4\), 5 mg/L thiamine, 1 mg/L biotine, 1 mg/L choline chloride, 1 mg/L folic acid, 1 mg/L niacinamide, 1 mg/L D-pantothenate, 1 mg/L pyridoxal, 0.1 mg/L riboflavin). After 30 min at 37 °C, the temperature was lowered to 20 °C and protein expression was induced with 0.4 mM IPTG. Cells were harvested by centrifugation after 16h and stored at -80 °C until further use. For purification, four to five grams of harvested bacteria were resuspended in 10 mL of lysis buffer (300 mM NaCl, 50 mM sodium phosphate, 10mM imidazole, pH 7.3, filtered, degassed) per gram (wet weight) of cells. A pinch of lyophilized RNAse A and DNAse I and one “Complete EDTA-free Protease Inhibitor” tablet (Roche) per 50 mL buffer were added. Cells were homogenized four times using a Standard FRENCH® Pressure Cell (THERMO Scientific; cell pressure: 18,000 psig). The supernatant obtained by centrifugation (30 min, 10,000 x g, 4 °C) of the lysate and containing the soluble target His\(_6\)-E6 construct was loaded two times onto a manually-packed, gravity-flow operated column with 0.5 g NiNTA agarose (QIAGEN) per gram (wet weight) bacteria pre-equilibrated with 10 column volumes (CV) of lysis buffer. Subsequently, each column was washed with lysis buffer, high-salt buffer (lysis buffer with 500 mM NaCl), and wash buffer (lysis buffer with 70 mM imidazole) until the absorbance at 280 nm reached baseline. For eluting the E6 protein constructs 0.75 CV of elution buffer (lysis buffer with 250 mM imidazole) were applied stepwise until no further protein eluted as assessed by absorbance at 280 nm. For 51Z2, typically 10 CV lysis buffer, 6
CV high-salt buffer and 16 CV wash buffer were employed before elution with a total of 6.75 CV elution buffer.

Subsequently, the pool containing the respective E6 construct was dialyzed against NMR buffer (150 mM NaCl, 50 mM L-arginine, 50 mM L-glutamine, 10 mM DTT, pH 7.4 filtered, degassed) and supplemented by thrombin (10 U per mg of E6 protein) in order to liberate E6 from the amino-terminal His₆-tag during overnight dialysis at 4 °C. The resulting E6 construct always contained the amino-terminal GSHM sequence not part of the E6 moiety but originating from the thrombin cleavage site.

Following concentration to 2 mL via Vivaspin 20 (3 kDa cutoff, GE healthcare), the dialyzed pool was subjected to gelfiltration using a HiLoad 16/600 Superdex 75 pg column (GE healthcare) operated at 1 mL / min NMR buffer on an ÄKTA Avant or ÄKTA Explorer FPLC system (GE Healthcare). The fractions containing the E6 protein were pooled and concentrated as above. The E6 concentration was assessed using the absorbance at 280 nm of a fresh sample with extinction coefficients calculated by the ProtParam tool [1] available at www.expasy.org/protparam/. These samples were used for biophysical assays.

**Biophysical characterization of soluble E6 proteins**

Analytical gelfiltration at 20 °C of E6 constructs was performed by injecting 50 µL of approx. 100 µM concentrated protein onto a TSKgelG3000SWxl column at a flow rate of 0.75 mL/min of NMR buffer. The column was calibrated at the same flow rate (see Figure S2) with thyroglobulin (669 kDa), alcohol dehydrogenase (150 kDa), carbonic anhydrase (29 kDa), cytochrome c (12.4 kDa) and aprotinin (6.5 kDa). Proteins were detected at 280 nm.

Dynamic light scattering of soluble E6 proteins (at least 100 µM in NMR buffer) was carried out on a Viscotec 802 DLS instrument. Light scattering was monitored at 90° relative to the incident laser beam. Thirty transients of 3 sec each were recorded at 20°C in a 50 µL volume. Processing of the recorded scattering intensities and the autocorrelation functions, evaluation of the mass weighted distribution of hydrodynamic radii as well as radii conversion to molecular weights utilizing a mass model of globular proteins was carried out with the Viscotec OMNISIZE software v3.0.

[¹H-¹⁵N]-HSQC spectra were recorded for each ¹⁵N labeled E6 construct that was monodisperse as assessed by gel-filtration and dynamic light scattering. These
Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction

constructs were 18Z2, 26Z2, 45Z2 and 51Z2 (Table S2). Sodium azide and D$_2$O was added to the E6 constructs in NMR buffer, effectively resulting in at least 100 µM E6 construct in 135 mM NaCl, 45 mM L-arginine, 45 mM L-glutamate, 9 mM DTT, 0.05% (w/v) NaN$_3$, pH 7.4 in 90% H$_2$O/10% D$_2$O. Spectra were recorded at 20 °C. Results of the biophysical characterization of soluble, purified E6 constructs are summarized in Table S2. The un-stability of 26Z2 over time is illustrated in Figure S1. The biophysical characterization of 51Z2 is given in Figure S2.

Expression and purification of 51Z2

The expression was performed as described for the other soluble E6 proteins except that $^{13}$C and $^{15}$N labeled 51Z2 was generated for NMR spectroscopy. To that end uniformly $^{13}$C-labeled D-glucose and $^{15}$N-labeled NH$_4$Cl were used as the sole carbon and nitrogen source, respectively, in modified M9 medium. It turned out, that reducing expression time to 6 h increased yield per time unit. The purification was performed as described above except that the NaCl concentration in the SEC buffer was reduced from 150 mM to 100 mM.

Circular dichroism of 51Z2

Circular dichroism (CD) was employed to estimate the secondary structure content of 51Z2 (Figure S2c, S2d). The protein buffer was exchanged to into CD buffer (10 mM sodium phosphate, pH 7.4, filtered, degassed) via NAP 5 columns (GE healthcare). For 14.8 µM 51Z2 in a 1 mm pathlength cuvette, the CD spectrum (190-260 nm) was recorded on a JASCO spectropolarimeter J-710 using averaging of 8 scans at 20 °C. Following blank subtraction and unit conversion to molar ellipticity the resulting CD data were deconvoluted using CDNN [2].

Peptide generation

All unlabeled peptides (E6CT6: Ac-RNETQV and E6CT11: Ac-QRTRQRNETQV or pGlu-RTRQRNETQV) were chemically synthesized and kindly provided by S. Rothemund (IZKF Leipzig, Germany) in HPLC-purified form. For NMR spectroscopy of the hDlgPDZ2-E6CT11 peptide complex, $^{13}$C and $^{15}$N labeled E6CT11 (pGlu-RTRQRNETQV) was generated applying the intein system as previously described [3]. The concentration of peptides in solution was evaluated spectrophotometrically utilizing absorbance of peptide bonds at 205 nm [4].
Expression and purification of hDlgPDZ2

The plasmid encoding hDlgPDZ2 was obtained from addgene (www.addgene.org). hDlgPDZ2 was expressed and purified as described previously [5] utilizing the C-terminal His₆-tag for affinity chromatography followed by gelfiltration.

Surface plasmon resonance (SPR) of hDlgPDZ2

Real-time analyses were performed on a Biacore 2000 system (GE Healthcare) at 25 °C and data were processed with Scrubber 2.0c (BioLogic Software). His₆-tagged hDlgPDZ2 was immobilized on flow cell 2 of a NTA sensor chip (GE Healthcare) by using a capture coupling method that results in the capture of hDlgPDZ2 in a non-random orientation by the His₆-tag after EDC/NHS activation [6]. 2780 RU were found to be covalently bound after injection of 1 µM hDlg-PDZ2 (50 µL) in running buffer (150 mM NaCl, 10 mM HEPES, 3 mM EDTA, 0.005% (v/v) surfactant polysorbate 20, pH 7.4) at a flow rate of 5 µL/min. E6CT6 and E6CT11 peptides were injected in running buffer at concentrations between 3 and 200 µM. Association and dissociation times were set to 1 minute at a flow rate of 30 µL/min. Each injection was performed at least 3 times. Regeneration was achieved with 10 mM glycine/HCl, pH 2.0 for 40 seconds at a flow rate of 30 µL/min. Refractive index errors due to bulk solvent effects were corrected with responses from non-coated flow cell 1 (activated with ECD/NHS and subsequently deactivated by ethanolamine/HCl, pH 8.5) as well as subtracting blank injections. Kᵤ values were calculated from the kinetic rate constants for peptide-hDlgPDZ2 complex formation and dissociation derived from a 1:1 interaction model including a mass transport term as well as by steady-state affinity analysis.

Comparison of hDlgPDZ2 binding to E6CT6 and E6CT11 peptides via NMR spectroscopy

[^1H-^15N]-HSQC-spectra of ^15N (and ^13C) labeled hDlgPDZ2 at 100 µM in presence of increasing concentrations (0, 20, 40, 70, 100, 200, 300 µM) of E6CT6 or E6CT11, respectively were recorded. From 200 to 300 µM peptide concentration, the signals did not show any significant further changes indicating the endpoint of titration, i.e. saturation of hDlgPDZ2 with peptide. In order to deduce differences of the complex
2.3 Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction

of the E6CT11 versus the E6CT6 with hDlgPDZ2, the combined $^1$H and $^{15}$N chemical shifts ([7]; scaling factor $^1$H / $^{15}$N: approx. 1/5.5) obtained at the endpoints of titration (in presence of 300 µM E6CT11 or E6CT6 peptide, respectively) were compared.

NMR spectroscopy and structure calculation

NMR spectroscopy of 51Z2 in 90 mM NaCl, 45 mM L-arginine, 45 mM L-glutamate, 9 mM DTT, 0.05 % (w/v) NaN$_3$, pH 7.4 in 90% H$_2$O/10% D$_2$O was performed at 10 °C. For acquisition of 51Z2 NMR spectra in 100 % D$_2$O, an 51Z2 NMR sample (90 % H$_2$O, 10 % D$_2$O) was lyophilized and resuspended in an equal volume of 100 % D$_2$O. [$^1$H-$^{13}$C]-HSQC spectra were recorded to assess stability of 51Z2 during this procedure (data not shown).

For NMR spectroscopy of hDlgPDZ2 at 20 °C the same buffer as previously described [5] was utilized (20 mM sodium phosphate, pH 6.5, filtered, degassed, 4 mM TCEP in 90% H$_2$O/10% D$_2$O or in 100% D$_2$O) with the modification that 0.05% (w/v) sodium azide was included to avoid microbial growth. For NMR spectra of 51Z2 in presence of hDlgPDZ2 the 51Z2 conditions were utilized. Spectra of hDlgPDZ2 in complex with either E6CT11 or E6CT6 (51Z2-derived) peptides were recorded at the hDlgPDZ2 conditions.

$^1$H, $^{13}$C and $^{15}$N resonances were assigned based upon the experiments listed in Tables S3 and S4. The 51Z2 assignment, based on experiments listed in Table S3, has a total completeness of 85.1, 88.4 and 64.0 % for $^1$H, $^{13}$C and $^{15}$N atoms, respectively, with missing resonance assignments located within the first four residues (GSHM) that are vector-encoded and not part of the E6 moiety and atoms that are usually not detected by standard NMR techniques such as e.g. atoms within arginine guanidinium-groups or hydroxyl-protons. The assignment of 51Z2 was deposited in the BMRB, entry 18967. The assignment of the hDlgPDZ2-E6CT11 complex has been deposited in the BMRB, entry 17942 [8].

Assignments and NOEs were evaluated using CARA [9]. For structure calculation, CYANA 3.9 was employed [10,11] and in general, the best 20 of 100 structures were analyzed further. Upper limit distance constraints were derived by calibration of NOE peak intensities. For E6, calibration classes of 2.9, 3.7, 5.4 and 5.9 Å for non-exchangeable and 4.4 and 5.7 Å for exchangeable hydrogens were used, respectively. For the complexed hDlgPDZ2 calibration classes of 2.8, 3.5, 4.0, 5.0, 5.7 and 6.1 Å were employed. For the 51Z2 structure, constraints for maintenance of tetrahedral
zinc coordination were included in analogy to the experimentally determined E7 zinc coordination that also carries four zinc-coordinating cysteines [12] which is consistent with the zinc coordination of HPV 16 E6 [13,14]. Titrations of 51Z2 with excess of EDTA led to severe perturbations of 51Z2 $[^1H,^{15}N]$-HSQC spectra confirming the importance of zinc for the 51Z2 structure (data not shown) and the chemical shifts of the Ca and Cβ nuclei of the 51Z2 cysteines 103, 106, 136 and 139 are consistent with their role in zinc-coordination [15]. Structural calculations were augmented by torsion angle constraints derived from TALOS+ [16] and torsion angles constraints derived from HNHA experiments [17] in combination with the FOUND module [18]. For hydrogen bonds that were consistently formed during initial rounds of structure calculation, hydrogen bond constraints were also employed in final rounds of structure calculation. Even though intermolecular hydrogen bonds were not utilized during the complex calculations, the C-terminal residues of the E6CT11 arrange as additional β-strand on the hDlgPDZ2-E6CT11 complex with the consistent formation of hydrogen bonds for these residues. Presence of cis prolines (ω torsion angles = 0°) was assessed by evaluating proline Cβ and Cγ chemical shifts [19]. All prolines present in 51Z2 and in the hDlgPDZ2-E6CT11 complex are in trans configuration relative to the preceding residue (ω = 180°). For hDlgPDZ2 that contains eight Leu and Ile residues each, the conformation of Leu and Ile side chains was additionally evaluated: the chemical shift difference between Cδ1 and Cδ2 of Leu side chains was evaluated and the trans probability was calculated [20] for each methyl group. When the trans probability for one methyl group was above 0.75, and if a single conformation of the side chain in at least 15 out of 20 structures was observed, the χ2 angle was constrained to the conformation mostly present in the structures (trans: 180°+/-30°, gauche 60°+/-30°; according to [20], such residues have a low probability of rotamer interconversions). For Ile, Cδ1 chemical shifts were evaluated to calculate the probabilities of gauche- and trans conformation [21]. When one of these probabilities was above 0.75 and at least 15/20 structures were consistent with this particular side chain conformation, the angle constraint for the corresponding χ21 angle was included (gauche-:300°+/-30°, trans:170°+/-30°).

After the final round of refinement with CYANA, the resulting best 40 out of 200 CYANA E6 and hDlgPDZ2-peptide complex structures were water-refined using RECOORD scripts [22] for CNS [23] and the 20 structures with the lowest total energy were used for further evaluation. These have been deposited in the PDB,
2.3 Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction

entries 2M3L (E6) and 2M3M (complex). Non-native residues (amino-terminal GSHM of 51Z2 and the carboxy-terminal GSHHHHHH tag of hDlgPDZ2) were not depicted in the figures of this paper and were not included in the r.m.s.d. calculations. Figures were prepared using MOLMOL [24]. PDBsum [25] providing for PROCHECK functionality was employed to evaluate torsion angle distribution of non-proline, non-glycine residues.

Superimposition of 51Z2 to other available E6 structures

The 51Z2 closest-to-mean structure of the folded core (residues 80-140) was superimposed onto the corresponding regions of HPV 16 E6 (residues 80-140; PDB ID 2LJZ) and BPV E6 in complex with the LD1 motif of paxillin (PDB ID 3PY7). When a standard sequence alignment of these proteins is generated, the BPV E6 sequence contains a supernumerary (loop) residue at position 507. Thus, for the structural alignment with identical number of atoms, for BPV E6 only residues 453-506 and 508-514 were utilized. Residue 507 is nevertheless shown in the corresponding Figure 7.

References for Supplementary Text 1

2.3 Structural insights into a wildtype domain of the oncoprotein E6 and its PDZ interaction


3 Discussion

3.1 Results summary

Infections with high-risk types of human papillomaviruses are the cause of virtually all cervical cancers (Ferlay, 2010) and of a subset of penile, vulvar and vaginal, anal, mouth and oropharyngeal cancers (Parkin, 2006). HPV exerts its malignant properties by the combined action of the oncoproteins E6 and E7 which interact with numerous cellular proteins (Song, 2000; Doorbar, 2006; Tungteakkhun, 2008; Doorbar, 2012; Stanley, 2012). The initial aim of this work was to elucidate the structure of a wild-type high-risk, full-length E6 and to address how E6 interacts with cellular targets.

The propensity of recombinantly expressed HPV 16 E6 to form insoluble aggregates is well-documented (Nomine, 2001; Ristriani, 2009). This property hampers structural characterization in solution and has been tackled by mutagenesis (Nomine, 2006; Zanier, 2012; Zanier, 2013). Here, aggregation and insolubility were addressed by screening twenty-seven wild-type E6 protein constructs of nine HPV types (of high-risk HPV 16, 18, 26, 31, 45, 51, 59, 97 and of the more distant, cutaneous HPV 1a) for intrinsic solubility after expression in E. coli. Extensive testing of expression conditions was performed, i.e. varying expression temperature, medium composition, codon-optimization for bacterial expression and co-expression with chaperones. Optimized conditions, i.e. low expression temperature (20°C) combined with defined zinc content in the expression medium (10 µM), only yielded soluble C-terminal Zinc Binding Domains (ZBD2) of several HPV types (E6 ZBD2 domains of HPV 16, 18, 26, 31, 45, 51, 97), but no soluble full-length E6 or amino-terminal zinc binding domain (ZBD1) of E6. These findings agree with a model in which the E6 aggregation propensity mainly resides in the ZBD1 (Liu, 2009b; Zanier, 2012). The soluble E6 ZBD2 domains were affinity purified, exploiting a hexahistidine tag fusion. Subsequent to tag removal by thrombin protease, these proteins were further purified using gel filtration chromatography. Circular dichroism spectroscopy, analytical gel filtration, dynamic light scattering and [1H,15N]-HSQC ‘fingerprint’ NMR spectra were utilized for the analysis of the soluble ZDB2 domains. Of the soluble, purified proteins, only the ZBD2 of HPV 51 E6 (termed 51Z2) was amenable to further NMR analysis, as it was the only monodisperse, monomeric protein, exhibited an [1H,15N]-HSQC spectrum indicating a folded state and was long-term stable. The technical results and implications are discussed in section 3.2.
3.1 Results summary

The solution structure of 51Z2 was solved by NMR spectroscopy and structural similarities and differences to other E6 proteins were revealed. This will be discussed in section 3.3.

NMR spectra recorded at different temperatures indicate conformational plasticity of 51Z2 at temperatures above 20°C in vitro, a phenomenon that might have physiological relevance and is discussed in section 3.3.

51Z2 contains a PDZ-Binding Motif (PDZ-BM) located at the extreme carboxy-terminus. In the unbound state, the C-terminal 51Z2 residues including the PDZ-BM are solvent-exposed and adopt multiple conformations. Previous studies (Liu, 2007; Zhang, 2007) had demonstrated that peptides harboring the PDZ-BM of HPV 18 E6 interact with hDlg PDZ domain 2 (hDlgPDZ2). Therefore, it was investigated if and how 51Z2 interacts with hDlgPDZ2. [1H,15N]-HSQC NMR spectra of 51Z2 were recorded with increasing concentrations of hDlgPDZ2. An interaction is observed that perturbs the chemical shifts of nine C-terminal E6 residues. Previous studies revealed the interaction of E6-derived peptides with only six or seven residues with hDlgPDZ2 (Liu, 2007; Zhang, 2007). Surface plasmon resonance was employed to assess hDlgPDZ2 affinity of a peptide containing the complete flexible E6 C-terminal region (Ac-QRTRQRNETQV; termed E6CT11) versus the affinity of a truncated peptide (Ac-RNETQV; termed E6CT6). The additional residues of E6CT11 contribute to hDlgPDZ2 affinity. To elucidate the underlying structural mechanism, a structure determination of hDlgPDZ2 in complex with E6CT11 was undertaken. The complex structure reveals how the nine 51Z2 residues that were affected in presence of hDlgPDZ2 contribute to the hDlgPDZ2 interaction. The physiological implications of this interaction are shortly summarized in section 3.4 and further mechanisms of how E6 may further increase the affinity to hDlg in particular and to PDZ-domains in general are presented. Finally hypotheses on further PDZ-domain containing proteins and other proteins that could be targeted by E6 are presented in section 3.5. These ideas are experimentally approachable in the future and may be of relevance to HPV research in the future.
3.2 Technical aspects

Secondary structure of E6 protein constructs

After realizing that wild-type, full-length E6 or ZBD1 of E6 could not be obtained in a soluble form, the soluble, purified ZBD2 constructs were biophysically analyzed and ultimately the structure of 51Z2 was solved (publication 3).

The secondary structure content of several soluble, purified ZBD2 of E6 (termed e.g. 16Z2 for HPV 16 E6 ZBD2) was estimated by circular dichroism spectroscopy (CD; Table 1). The good agreement in the comparison of the CD-data of 51Z2 and the secondary structure composition back-calculated from the final 51Z2 solution structure corroborates the NMR-derived 51Z2 structure.

Table 1: Secondary structure content estimation (%) from circular dichroism spectroscopy recorded with freshly purified E6 constructs (identical buffer conditions at protein concentrations of 5 to 15 μM at 20°C). The secondary structure content of our 51Z2 structure (PDB: 2M3L) as well as the available structure of mutation-stabilized 16Z2 (Zanier, 2012; PDB: 2LJZ) are provided in parentheses.

<table>
<thead>
<tr>
<th>Construct</th>
<th>51Z2</th>
<th>26Z2</th>
<th>18Z2</th>
<th>97Z2</th>
<th>16Z2</th>
<th>31Z2</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-helix</td>
<td>19.4 (22.3)</td>
<td>23.5</td>
<td>18.4</td>
<td>26.5</td>
<td>25.2 (21.0)</td>
<td>22.6</td>
</tr>
<tr>
<td>Antiparallel β-sheet</td>
<td>29.2 (26.3)</td>
<td>18.5</td>
<td>24.3</td>
<td>12</td>
<td>8.6 (26.3)</td>
<td>13.5</td>
</tr>
<tr>
<td>Parallel β-sheet</td>
<td>6.2  (6.5)</td>
<td>6.3</td>
<td>6.4</td>
<td>7</td>
<td>6.4 (6.5)</td>
<td>5.7</td>
</tr>
<tr>
<td>β-turn</td>
<td>18.0 (18.4)</td>
<td>21.2</td>
<td>17.7</td>
<td>20.4</td>
<td>19.9 (18.4)</td>
<td>22.1</td>
</tr>
<tr>
<td>Other incl. random coil</td>
<td>27.5 (26.3)</td>
<td>29.6</td>
<td>32.5</td>
<td>33.2</td>
<td>39.8 (27.6)</td>
<td>35.2</td>
</tr>
<tr>
<td>Status*</td>
<td>1 (1)</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4 (1)</td>
<td>4</td>
</tr>
<tr>
<td>Phylogenetic species</td>
<td>α5</td>
<td>α5</td>
<td>α7</td>
<td>α7</td>
<td>α9</td>
<td>α9</td>
</tr>
</tbody>
</table>

* Ranked according to Table S2 from publication 3:
1: Monodisperse, NMR amenable, long-term stable
2: Monodisperse, NMR amenable, not long-term stable
3: Monodisperse, not NMR amenable
4: Polydisperse

The CD investigation also included polydisperse (i.e. organized as soluble aggregates of varied size) protein constructs as deduced from dynamic light scattering and gel filtration chromatography. The CD spectrum of proteins might be influenced by light
3.2 Technical aspects

scattering of polydisperse samples, but correction procedures for these effects are, to
our knowledge, not available. Interestingly, the polydisperse wild-type 16Z2 showed
marked differences in the secondary structure content compared to the published
16Z2 solution structure (Zanier, 2012) that was mutated to avoid aggregation.
Neglecting potential errors by light scattering, the CD analysis reveals that all
polydisperse E6 proteins contain a decreased content of antiparallel β-sheets and a
concomitant increase in random coil content, arguing for (a) different conformation(s)
of aggregated E6 domains. Oxidation of non-zinc coordinating cysteines appears
unlikely to be the reason for polydispersity, as purification was performed under
reducing conditions (10 mM DTT) and subsequent biophysical analysis was
performed with freshly purified proteins in deoxygenized solutions. Moreover the
polydisperse 97Z2 does not contain any non-zinc coordinating cysteines. Although
the aggregation of E6 is not the scope of this work, the underlying mechanism appears
to be different from amyloid-type processes which are frequently accompanied by
elevated β-sheet contents (Eichner, 2011).

Recombinant peptide generation and N-terminal pyroglutaminylination

On the way to solve the hDlg-E6CT11 complex structure, the generation of 13C- and
15N-labeled E6CT11 appeared indicated for unambiguous assignment of NMR
peptide resonances and for the unambiguous identification of intermolecular Nuclear
Overhauser Effects to be utilized for structural calculations. Chemical synthesis of
isotopically labeled peptides is possible but exceedingly expensive. Therefore it was
decided to prepare the E6CT11 by bacterial expression.
The generation of the stable-isotope labeled E6CT11 was achieved by the intein-
system that is traditionally exploited for ligation of larger protein fragments
(Dongsheng, 2009). The bacterial expression as a protein-peptide fusion allows the
generation of stable isotope labeled peptides that are properly expressed and which
are protected from cellular proteases (Chong, 1998a). The utilized intein-peptide
fusion offers the additional advantages that the intein contains a chitin-binding
domain for affinity purification and that subsequent tag-removal does not require
costly proteases (Chong, 1998a). It allows for thiol-induced intein-mediated cleavage
at a defined position nearly irrespective of the newly generated terminal residues
thereby releasing the target peptide (Chong, 1998a; Chong 1998b). The successful
production of the E6CT11 peptide from the fusion by this approach was demonstrated
3.2 Technical aspects

by mass spectrometry (MS). The publication 1 proved that the intein-system can be exploited for the generation of stable-isotope labeled peptides for NMR spectroscopy studies without undesired terminal residue extensions. However, during production of E6CT11, the amino-terminal glutamine unexpectedly auto-cyclized spontaneously into pyroglutamate (pGlu; Scheme 1).

Pyroglutaminylation became apparent as a size-reduced species as detected by MS and by NMR spectral changes indicating magnetization transfer routes that can only be explained by a cyclization of the amino-terminal residue; the process was finished after ~500 hours. The pGlu formation of amino-terminal glutamines is known for quite a while (Melville, 1935). Here, the impact of pyroglutaminylation on NMR spectra has been demonstrated for the first time and the apparent NMR spectral properties even allowed for a kinetics description of the underlying reaction, with 50% turnover after 61 h, similar to pGlu formation rates of free Gln (Schneider, 2003). pGlu also forms from amino-terminal glutamates, however, with a reduced rate (50% turnover require at least 24 weeks; Chelius, 2006). Publication 1 also serves as a gentle reminder to check for this modification when dealing with protein constructs with unprotected amino-terminal glutamines or glutamates. Amino-terminal pGlu carries a physiological role, as initially shown for the thyrotropin-releasing-hormone (pGlu-His-Pro-NH₂) which binds at least 216-times stronger to its receptor than the noncylcized derivative Gln-His-Pro-NH₂ (Hinkle, 1974). The pGlu-modification conferred more toxicity to amino-terminally truncated, oligomeric Amyloid β (Aβ) peptides (Schlenzig, 2009; Wittnam, 2012). It was demonstrated recently that these pGlu-modified Aβ peptides serve as seed for non-pGlu Aβ peptides that in turn adopt a more toxic conformation (Nussbaum, 2012). Since recombinantly produced peptides are not perfectly mimicked by chemically synthesized peptides (Finder, 2010), the spontaneous formation of pGlu can be exploited, as demonstrated here, to generate pGlu-modified peptides by biotechnological means, preferably by using glutamines as amino-terminal residue (due to faster kinetics as compared to amino-terminal glutamates). Even though enzymes for the catalysis of such a reaction (glutaminyl

\[ \text{H}_2\text{N} \quad \text{O} \quad \text{O} \quad \text{RTRQRNETQV} \quad \rightarrow \quad \text{O} \quad \text{NH}_2 \quad \text{O} \quad \text{RTRQRNETQV} + \text{NH}_3 \]

**Scheme 1**: Cyclization of the amino-terminal glutamine of E6CT11 into pyroglutamate.
cyclases) and for the removal of pGlu (pyroglutamate exopeptidase) exist, the publication 1 is, to the best of my knowledge, the first paper in which it was demonstrated that pyroglutaminylated proteins and peptides can be generated by recombinant expression without subsequent use of (costly) enzymes. Besides the generation of physiologically relevant proteins or peptides, pyroglutaminylation may be exploited as amino-terminal protecting group interfering with the action of exoproteases.

The resonances of hDlg PDZ domain 2 (hDlgPDZ2) had already be assigned in presence of an unlabeled amino-terminally Gln-acetylated E6CT11. The peptide acetylation (H3C-CO-NH-R, where NH is the amine of the first residue and R denotes the peptide residues) masks the primary amine of the N-terminus and results in a peptide bonded first residue. This imitates the situation of E6CT11 residues in the full-length E6 (or in 51Z2) peptide bonded to preceding residues. The stable isotope labeled E6CT11 that was obtained by the intein system had the same sequence, but instead of the amino-terminal acetylated Gln, it harbored the pyroglutamate (Scheme 1). To rule out that the complex is significantly affected when switching from an amino-terminal acetylated Gln to pyroglutaminylation of the peptide, the [1H,15N]- and [1H,13C]-HSQC spectra serving as a fingerprint spectra of isotope labeled hDlgPDZ2 were compared when complexed either with acetylated or pyroglutaminylated unlabeled E6CT11, respectively. The result (publication 2) shows that the complex was virtually unaffected which allowed the combined usage of the assignment data sets of: a) 13C- and 15N-labeled hDlgPDZ2 in complex with an 2.5 fold excess of acetylated unlabeled E6CT11 and b) 13C- and 15N-labeled pGlu-modified E6CT11 in complex with an 2.5 fold excess of unlabeled hDlgPDZ2. Usage of the combined assignment data sets allowed the assignment and evaluation of in total 2760 Nuclear Overhauser Effects utilized for the structure determination of the hDlgPDZ2-E6CT11 complex (publication 3).

Having discussed technical aspects on the way to the structure determination of 51Z2 and of the hDlgPDZ2-E6CT11 complex, the next sections will focus on the implications deduced from the unbound, wild-type 51Z2 structure (section 3.3) and the hDlgPDZ2-E6CT11 complex structure (section 3.4), respectively.
3.3 Solution structure of a wild-type E6 domain

*E6 proteins share a globally conserved structure*

After extensive screening, we solved the first structure of a wild-type HPV E6 domain (publication 3) by solution NMR spectroscopy. This structure is validated by torsion angles residing in favorable and allowed regions of the Ramachandran plot and, locally, by ring-current effects (publication 3).

Earlier available structures of HPV 16 E6 were solved utilizing stabilizing mutations that functionally compromise the full-length E6 to bind and degrade p53 (*Nomine*, 2006; *Ristiani*, 2009; *Zanier*, 2012, *Zanier*, 2013). The wild-type 51Z2 adopts the global conformation of the corresponding domains of mutated HPV 16 E6 (*Zanier*, 2012), of E6AP-peptide complexed mutated HPV 16 E6 (*Zanier*, 2013) and of the paxillin-peptide complexed, evolutionary more distant bovine papillomavirus 1 (BPV; *Zanier*, 2013) E6 proteins (Figure 8).

![Figure 8](image)

*Figure 8:* Superimposition of the unbound 51Z2 closest-to-mean structure (blue, PDB ID 2M3L, folded part, residues 80-140) onto the corresponding regions of unbound HPV 16 E6 (red, PDB ID 2LJZ), of BPV E6 in complex with the LD1 motif of paxillin (gray, PDB ID 3PY7, LD1 omitted for clarity) and of HPV 16 E6 in complex with an E6AP-derived peptide (orange, PDB ID 4GIZ, E6AP-peptide omitted for clarity). The overall topology is conserved. Notably, the β4 and β5 strands and their connecting loop of unbound 51Z2 positions similar to the available E6 structures in a complex, while the corresponding region orients differently in the unbound HPV 16 E6 (upper right corner; encircled and highlighted).
The PVs for which E6 structures are available are phylogenetically dissimilar; HPV 51 is grouped into α5, HPV 16 into α9 and BPV-1 into δ4 species (see Figure 1 of the introduction). Although these PV types are phylogenetically distant (at least the HPV types versus the bovine PV type), their global structure of the ZBD2 E6 domain is conserved. This strongly argues for a similar structure for the ZBD2 domains of all E6 proteins. If the structure of E6 proteins is conserved, how then can the different set of E6 interactors of various HPV types (Neveu, 2012; White, 2012) be explained that even permits a phylogenetic classification of an HPV solely based on the respective E6 interactome (Neveu, 2012)? Besides differences in surface charges and surface hydrophobicity, local structural differences clearly could determine to which interactors a given E6 binds.

Local structural E6 differences do exist
While the E6AP-peptide complexed HPV 16 E6 adopts the conformation of unbound HPV 51 E6, local structural differences do exist for the unbound conformations of these closely related high-risk E6 proteins (Figure 8). Unbound HPV 51 E6 has a different conformation around the region of the β4- and β5-strands in comparison to unbound HPV 16 E6 (publication 3; Zanier, 2012). In this region resides I128, which is conserved among the high-risk E6 proteins (publication 3). A single I128T E6 mutation in the context of the HPV 31 genome resulted in viral genomes that were no longer stably maintained, because episomes were lost after a few cell passages (Lee, 2007). I128T mutation of HPV 16 E6 impairs the interactions with LXXLL containing proteins (where X denotes any residue) E6AP and E6BP (Liu, 1999). Similar to a mutation of the I128 residue, the structural difference of the region around I128 for unbound HPV 51 E6 in comparison to unbound HPV 16 E6 may have implications for the virus as well, e.g. a different E6AP binding affinity and specificity of HPV 16 and HPV 51 E6 or a changed affinity towards other LXXLL containing proteins.

The E6AP interaction, attenuated by a conformational difference, might in turn affect proteins that are targeted by E6 via E6AP for proteasomal degradation (Beaudenon, 2008). Since E6 protein levels of various HPV types are increased in vivo by the E6AP interaction (Tomaić, 2009b; Thomas, 2013), an altered E6AP interaction could affect other E6 interactors simply by changing the E6 protein levels. In summary, the structural difference of the I128 region may contribute to changed affinity and a
changed set of interactors, as demonstrated to exist for various E6 proteins by large-scale interaction studies (Neveu, 2012; White, 2012).

The different conformation of the I128 region leading to possibly influenced affinity for certain interactions may contribute to the lower persistence of HPV 51 as compared to HPV 16 (mean infection persistence: 10 and 18 months, respectively; Richardson, 2003), potentially by interfering with episomal maintenance (vide supra and Lee, 2007).

In publication 3, the prevalence of the high-risk types HPV 51 and HPV 16 in precancerous lesions (Bruni, 2010; Clifford, 2005; Smith, 2007) and in squamous cell cervical carcinoma (Smith, 2007) were compared. HPV 51 is, as HPV 16, frequently present in precancerous lesions of the cervix, but HPV 16 is by far the most prevalent HPV in cervical cancer. The E6 protein governs the later stages of carcinogenesis (Song, 2000) and thus the structural difference around I128 of E6 might have a bearing on the carcinogenicity of individual HPV types.

Whether the structural differences between the respective E6 proteins and the potentially changed interaction characteristics are causative for these findings, or whether other properties of E6, of E7, of both and/or other viral characteristics are responsible, has to be elucidated in the future.

Conformational plasticity of E6 proteins

Up to now, the similarities and differences of E6 structures have been discussed. The present work has elucidated one aspect that was previously not noted. At elevated temperatures above 20°C in vitro, line broadening of most 51Z2 backbone signals is observed (publication 3) indicative of ms-to-μs motion (Wüthrich, 1986) of the wild-type ZBD2 of HPV 51 E6. Only signals of freely-rotating side-chains and few backbone signals of residues involved in zinc coordination are detected at elevated temperatures (publication 3). The less severe line broadening of residues near the zinc ion implies that zinc is still coordinated at elevated temperatures and the region around the zinc ion is more rigid, while other regions of 51Z2 are moving. In vivo, at 37°C, E6 might thus be more malleable than previously thought. The [1H,15N]-HSQC fingerprint spectra of 51Z2 prior to heating and after cooling down are identical, implying that the solved, low-temperature 51Z2 structure represents a low energy state. Thus, this might be the conformation, which E6 adopts in vivo when bound to
interactors. In line with this hypothesis, the available structures of E6 bound to target peptides show the same overall E6 topology (Figure 8).

Conformational plasticity is a property of at least the ZBD2 of E6. Since the ZBD1 of high-risk E6 is, up to now, not available as wild-type sequence for structural characterization, the reported dimerization of the ZBD1 (Zanier, 2012) could be related to plasticity of this domain as well. Several structures with mutated HPV 16 E6 are available (Nomine, 2006; Zanier, 2012; Zanier, 2013). In the HPV 16 E6-E6AP-peptide complex (Zanier, 2013), a so-called linker helix is present at the carboxy-terminal end of the ZBD1 which links ZBD1 with ZBD2. However, the isolated ZBD1 of HPV 16 E6 lacks the corresponding α-helix (Zanier, 2012) and in the unbound full-length protein, the NMR signals of the linker helix region undergo line broadening (Zanier, 2012), indicative for flexibility of this region in the unbound state.

Collectively, our results on the conformational plasticity of a wild-type E6 ZBD2, together with the results of Zanier et al. on the linker helix, suggest that E6 proteins may undergo a (partial) disorder-to-order transition upon binding to interactors, which is reminiscent of intrinsically-disordered proteins that also undergo a disorder-to-order transition upon productive complex formation with specific ligands (Turoverov, 2010).

The conformational plasticity may have functional consequences, either for the E6 protein stability per se or for the mode of the numerous E6 interactions:

**Stability:** It seems possible that E6 may be, at least partially, intrinsically disordered or flexible in vivo and may undergo structural stabilization if bound to interaction partners. Such a structural stabilization upon binding would explain the reported resistance to proteasomal degradation of E6 in vivo in presence of the interactors hDlg, E6AP, Scribble and USP15 (Nicolaides, 2011; Tomaić, 2009b; Vos, 2009).

**Interactions:** Conformationally flexible proteins may sample already (the) bound conformation(s) in the free state allowing for interactions by the conformational selection model originated from the Monod-Wyman-Changeux model of allostery (Monod, 1965; Okazaki, 2008; Figure 9). However, whether E6 interacts via this mechanism to its interactors and/or by an induced-fit mechanism (Koshland, 1958; Figure 9) remains to be elucidated. NMR spectroscopy offers particular tools for the characterization of short-lived protein transition states (Baldwin, 2009) which could
3.3 Solution structure of a wild-type E6 domain

be applied for characterization of short-lived E6 conformations at elevated, physiologically relevant temperatures.

Figure 9: A protein can be reshaped by a ligand, which can also be a protein, to form the protein-ligand complex (Koshland, 1958). This induced-fit mechanism is illustrated on the left side of the figure. Alternatively (on the right side of the figure), a protein can be present in multiple conformations (here: open versus closed), one of which resembles the ligand-bound state that is stabilized upon interaction to a ligand (conformational selection/population-shift model; Monod, 1965). The conformational plasticity of E6 allows for the sampling of multiple conformations at physiological temperatures. However, whether E6 interacts via this mechanism to its interactors and/or by an induced-fit mechanism remains to be elucidated (Figure modified from http://neurobio.drexel.edu).

3.4 Comments on the E6-hDlg complex

51Z2 interacts with hDlgPDZ2, increasing affinity via an extended PDZ binding motif

An important class of E6 interactors are PDZ-domain containing proteins (further referred to as PDZ-proteins), to which E6 binds by its C-terminal residues harboring a canonical class I PDZ binding motif (PDZ-BM) of the form X-S/T-X-φ, where X denotes any residue and φ denotes a hydrophobic amino acid (Thomas, 2008b). Exclusively high-risk, but not low-risk or cutaneous, E6 proteins possess a PDZ-BM (Thomas, 2008b).

Peptides of six or seven residues containing the HPV 18 E6 PDZ-BM bind to hDlgPDZ2 (Liu, 2007; Zhang, 2007). With the 51Z2 at hand including the PDZ-BM,
it was investigated, if and how 51Z2 interacts with hDlgPDZ2 and if so, which residues of this E6 domain contribute to this interaction. Titrations in combination with NMR spectroscopy elucidated that 51Z2 interacts with hDlgPDZ2 and that no residues of E6 (at least of the ZBD2), which are conformationally defined in the unbound 51Z2 state, are involved in binding, but that nine C-terminal residues, which are flexible in the unbound 51Z2 structure, are involved in the interaction with hDlgPDZ2 (publication 3). Surface plasmon resonance was utilized to demonstrate that the additional residues contribute to binding to hDlgPDZ2. A peptide (E6CT11; Ac-QRTRQRNETQV) harboring all eleven flexible C-terminal E6 residues (including the nine C-terminal residues involved in the hDlgPDZ2 interaction) has an affinity of slightly below 10 µM to hDlgPDZ2, whereas a truncated peptide (E6CT6; Ac-RNETQV) consisting of only the six C-terminal E6 residues bound to hDlgPDZ2 with 3-fold reduced affinity (slightly above 28 µM). Mutation of one single residue in the PDZ-BM of the envelope protein of rabies virus dramatically changed its PDZ target spectrum and resulted in a switch from virulence to an attenuated state (Préhaud, 2010). Therefore each and every residue involved in a PDZ interaction appears important with respect to biological function and warrants structural analysis of its role in such an interaction. For HPV 51 E6, the structural basis for the increased affinity was ultimately unraveled by solving the solution structure of the hDlgPDZ2-E6CT11 complex.

*Structural basis of the extended E6 PDZ-BM hDlgPDZ2 interaction*

All backbone φ and ψ torsion angles of the hDlgPDZ2-E6CT11 complex structure lie in favored and allowed regions of the Ramachandran plot (publication 3). The complexed PDZ domain has a high degree of structural similarity in terms of overall-topology to related PDZ domains (Luck, 2012) or to hDlgPDZ2 in complex with shorter E6 peptides (Liu, 2007; Zhang, 2007; Figure 10). Thus, the general topology of hDlgPDZ2 is unchanged when in complex with a longer peptide.

As previously discussed, pyroglutaminylation compared to acetylation of the amino-terminal E6CT11 residue did not affect the hDlgPDZ2 in the complex. This finding is consistent with the complex structure revealing that the first two amino-terminal E6CT11 residues are mostly disordered and do not contact the PDZ domain (publication 3). These findings agree with the fact that the resonances of residues Gln141 (the pGlu forming residue on the E6CT11) and Arg142 on 51Z2 are not
perturbed in presence of hDlgPDZ2 (publication 3). Thus, exactly the nine C-terminal HPV 51 E6 residues TRQRNETQV, which are disordered and solvent-exposed in the unbound 51Z2, contact hDlgPDZ2. In summary, the structure of the hDlgPDZ2 in complex with E6CT11 revealed that the newly identified E6 residues contributing to hDlgPDZ2 binding bend around the α-helix 2 to form additional contacts (publication 3).

![Figure 10: Superimposition of residues 319-401 that are present in all three structural datasets of hDlgPDZ2 in complex with E6-derived peptides. Blue: our complex with the E6CT11, solved by NMR spectroscopy (PDB Code 2M3M). Red: complex with a 7mer HPV 18 E6-derived peptide, solved by X-ray crystallography (Zhang, 2007; PDB Code 2I0L). Gray: complex with a 6mer HPV 18 E6-derived peptide, solved by NMR spectroscopy (Liu, 2007; PDB Code 2OQS). The E6-derived peptides are omitted for better visibility of hDlg structural elements. Peptides reside in the groove between the long α-helix 2 (on the right) and β-strand 2, forming a β-strand parallel to β-strand 2. The more N-terminal residues of E6CT11 bend around the α-helix 2 (see publication 3 and Figure 11 below).](image)

Tools for de novo backbone structure prediction by exclusive use of sequence information and chemical shifts, CS-Rosetta and CHESHIRE (Shen, 2008; Cavalli, 2007), were employed to predict the conformation of the complexed E6CT11 without the distance information derived from Nuclear Overhauser Effect spectroscopy (that were utilized for the hDlgPDZ2-E6CT11 structure determination). CS-Rosetta did not converge, whereas CHESHIRE yielded predictions (Figure 11) in which the backbone
3.4 Comments on the E6-hDlg complex

of the E6CT11 assumes an U-shape and nicely superimposes with the peptide conformation within the experimentally solved hDlgPDZ2 complex. This is one case - besides several published examples (Shen; 2008; Cavalli, 2007) - demonstrating the feasibility to predict structures with reasonable accuracy solely based on chemical shifts.

![Figure 11: Superimposition of representative E6CT11-structures derived from CHESHIRE (red) to the final E6CT11 in the complex (black). Backbones of most ordered peptide residues (143-151) in complex with hDlgPDZ2 from 20 structures were aligned to the respective atoms of 20 representative CHESHIRE structures (r.m.s.d.: 1.05 Å) and all peptide residues (141-151) are shown with the peptide termini indicated. For the figure, the hDlgPDZ2 was omitted (in the final complex structure, it is present in approximately the orientation depicted in figure 10). CHESHIRE was run by changing the amino-terminal peptide residue from the nonstandard pGlu to Gln.](image)

Functional implications of the E6-hDlg interaction

Having elucidated the complex structure of hDlgPDZ2-E6CT11, the physiological implications of the interaction are shortly summarized:

Since exclusively high-risk, but not low-risk or cutaneous E6 proteins contain a PDZ-BM (Thomas, 2008b), the E6 hDlg interaction is confined to high-risk E6 proteins. Except for cytoskeletal hDlg, which is stabilized by E6 and ensures, together with E6 and SGEF, high RhoG and invasive activities, the HPV E6 hDlg interaction results in hDlg degradation (Thomas, 2008b; Massimi, 2008; Subbaiah, 2012). The E6-mediated hDlg degradation has been confirmed for most high-risk HPV types, including HPV 16, 18 and 51 (Muench, 2009). As stated in the introduction, the E6 interaction with and degradation of hDlg (Kiyono, 1997; Gardiol, 1999) disrupts the hDlg function as a negative regulator of G0/G1 to S phase transition (Ishidate, 2000)
3.4 Comments on the E6-hDlg complex

and confers insensitivity to anti-growth signals to the host cell. The resistance to anoikis of hDlg depleted cells (Massimi, 2012), implies that these cells do not undergo apoptosis if detached from their cellular environment, which clearly favors metastasis (Liotta, 2004). The resistance to anoikis and the deregulation of the normal hDlg functions as regulator of epithelial-polarity may favor epithelial-to-mesenchymal transition, thought to contribute to invasion and metastasis (Thompson, 2005; Bryant, 2008; Javier, 2011). Taken together, deregulation of the cellular hDlg functions by E6 are very important for tumorigenesis (Shai, 2007; Roberts, 2012). The E6-hDlg interaction in addition leads to slower proteolytic turnover of E6 in vivo as well (Nicolaides, 2011), thereby increasing E6 steady-state levels.

Mechanisms of high affinity PDZ-BM PDZ domain interactions

Until very recently, the interaction of proteins/peptides to PDZ domains was thought to be mediated exclusively by the four C-terminal residues that form a β-strand upon interaction with PDZ-domains, and these four residues form the canonical PDZ-binding motif (PDZ-BM; reviewed in Luck, 2012). Our results extend the number of residues important for hDlgPDZ2 binding as compared to earlier structures (Liu, 2007, Zhang, 2007) to nine residues in total and the 3-fold increased affinity (of E6CT11 versus E6CT6) is both due to an increased on-rate as well as a decreased off-rate (publication 3). PDZ 1 of MAGI-1 is also targeted by E6 through more residues than the canonical PDZ-BM contains. In that case, however, six E6 peptide residues contact the PDZ domain and two E6 peptide residues contact residues outside the canonical PDZ domain boundary (Charbonnier, 2011), while our extended E6 PDZ-BM interacts only with bona-fide PDZ residues. Peptides with nine residues show the highest affinity for binding to the three PDZ domains of murine Dlg which are 98% to 100% identical to the PDZ domains of hDlg (Songyang, 1997). In that study, the optimal binding peptides contained amino-terminal basic residues (R/K), and of the nine HPV 51 E6 residues involved in PDZ binding (TRQRNETQV), the more amino-terminal Arg is involved in a salt-bridge to E385 conferring charge-complementarity to the E6CT11 peptide in addition to the shape complementarity provided by the β-strand formation of the four C-terminal peptide residues representing the canonical PDZ-BM (Figure 6 of publication 3). Thus, our investigation of the hDlgPDZ2-E6 complex provides a structural rationale for the findings of Songyang (Songyang, 1997) and addresses one possibility how extended peptides increase affinity and
specificity for PDZ-interactions: by additional contacts including charge complementarity. The high-risk E6 proteins all harbor a C-terminal, conserved class I PDZ-BM (Thomas, 2008b) extended by nonconserved, unstructured more aminoterminal residues, as deduced from the 51Z2 structure and sequence alignment (publication 3, Figure S3). Although these unstructured residues are not conserved, the high-risk E6 proteins all contain basic residues, mostly Arg. Thus, the charge complementarity of HPV 51 E6 to efficiently target hDlg appears conserved among the other high-risk E6 proteins as well. This hypothesis could be examined in the future by residue-specific mutagenesis of basic E6 residues preceding the canonical PDZ-BM and analyzing the resulting binding affinity. The E6CT11 interaction with hDlgPDZ22 has an affinity of slightly below 10 \( \mu \text{M} \) in contrast to the shortened E6CT6 peptide with an affinity of slightly above 28 \( \mu \text{M} \). An affinity of human peptides/proteins containing PDZ-BMs for PDZ-domain interactions below 10 \( \mu \text{M} \) is rare; more often the affinity lie in the 10 to 100 \( \mu \text{M} \) range (Wiedemann, 2004). If, \textit{in vivo} in a given compartment at a given time, the effective concentrations of the E6 and a cellular protein competing for hDlg is similar, the extended PDZ-BM of E6 guarantees for high hDlgPDZ2 affinity and contributes to replacement of already bound cellular hDlg interaction partners with lower affinity. This finding suggests how E6 manages, despite the presence of cellular competing proteins, to bind to hDlg and possibly other PDZ-domains. As viruses often interact with cellular PDZ domains (see introduction; Figure 6; Javier, 2008; Javier, 2011), it can be envisioned that other viruses may also exploit extended PDZ-BM to successfully compete with cellular interactors for their PDZ domain targets. It appears that not only viral proteins exploit extended PDZ-BMs, since the cellular protein Adenomatous Polyposis Coli (APC) interacts with hDlgPDZ22 by an extended PDZ-BM as well (Zhang, 2011). That study aimed at an X-ray structural characterization of the complex, but of the APC 11mer-peptide (RHSGSYLVTSV), only the six C-terminal residues gave rise to interpretable electron density. The increased flexibility observed for the amino-terminus of E6CT11 in the hDlgPDZ2 complex might also apply to the APC-hDlgPDZ2 complex, which seems to have compromised a complete X-ray structural analysis of the APC-peptide residues.

Within PDZ domains, long-range interaction networks exist that contribute to enhanced binding of interacting peptides/proteins (Gianni, 2011). For example, I341 of PSD-95 PDZ domain 3 is equivalent to the I353 of hDlgPDZ2. This residue is
3.4 Comments on the E6-hDlg complex

located far away from the peptide-binding site and, at least for PSD-95, nevertheless contributes to binding affinity and specificity (Gianni, 2011; McLaughlin, 2012). This could not only explain the different chemical shift perturbation characteristics for the I353 of hDlgPDZ2 upon binding of the short versus long E6-derived peptide (publication 3, E6CT6 versus E6CT11 binding, respectively) but also implies that long-range interactions within PDZ-domains also contribute to an enhanced binding of extended PDZ-BMs.

E6 dimers and aggregates harbor multiple C-termini each including one PDZ-BM. If such an E6 multimer simultaneously binds to two adjacent PDZ domains, the affinity might be even more increased. Indeed, there is evidence for dimers and aggregates of E6 of various HPV types in vitro and in vivo (e.g. García-Alai, 2007; Ristriani, 2009; Zanier, 2010; Zanier, 2012). Moreover, several PDZ-proteins contain supramodules, i.e. entities of two closely arranged PDZ domains (Feng, 2009). The PDZ domains 1 and 2 of the hDlg-related PSD-95 form a structural supramodule (McCann, 2011) and there is also evidence for hDlg PDZ1-2 to form such a supramodule (Lue, 1996; Feng, 2009). It was evaluated that E6 coimmuno-precipitates with hDlgPDZ1 to a similar extent as with hDlgPDZ2 (Gardiol, 1999). Derived from the PSD-95 supramodule structure (McCann, 2011), two E6 C-termini could simultaneously bind to hDlgPDZ1-2 and if that is the case, cooperative binding will enable dimeric or aggregated E6 to interact much stronger with the hDlg supramodule. Two affinities to each PDZ domain in the µM range (inferred from own results and from Gardiol, 1999) could result, if binding is cooperative, in a combined affinity in the nM or even pM range. Since most PDZ-proteins targeted by E6 do not harbor a PDZ supramodule (e.g. MAGI-1, MAGI-2 and MAGI-3, MUPP-1; inferred from long linkers between individual PDZ domains), the E6 dimerization/aggregation could serve as a structural and functional switch in modulation of PDZ targeting towards increased hDlg affinity and selectivity. Such phenomenon has already been described for the viral AdE4 protein (Chung, 2011), for which the aggregation state influences PDZ domain binding by specific interaction of an AdE4 trimer to one hDlg supramodule. In case of E6, this mechanism could allow for a successful hDlg targeting, even if the intracellular E6 levels are low, provided that these low E6 levels are sufficient for E6 dimerization/aggregation. Another protein targeted by E6 with a PDZ-supramodule is the abovementioned PSD-95 (Handa, 2007; Feng, 2009; McCann, 2011). However, up to now the precise regions on PSD-95 to which E6 binds, remain to be elucidated
3.4 Comments on the E6-hDlg complex

and thus it remains possible that the PSD-95 PDZ1-2 supramodule is targeted by E6. The analysis of E6 binding to any PDZ supramodule requires a defined multimeric E6 conformation that is still not available and not trivial to produce. For example, the C-terminal ZBDs of E6 may be N-terminally fused with alkaline phosphatase (AP). Since AP dimerizes such that both C-termini reside on the same side of a dimer (Kim, 1991), a dimerized AP-E6 fusion provides an E6-dimer-like molecule with two C-terminal extended PDZ-BMs. Such a pseudo-dimer could be investigated with regard to PDZ supramodule interactions (Scheme 2).

**Scheme 2**: S1Z2 or other ZBD2 of E6 (light blue circle) could N-terminally be fused with AP (blue ellipse) that forms dimers (Kim, 1991). The solvent-exposed C-terminal regions of the E6 domains, each with the extended PDZ-BM (light blue lines), could, similar to an E6 dimer, interact simultaneously with two adjacent PDZ-domains e.g. PDZ1 and PDZ2 of hDlg forming a supramodule (green rectangles linked by a green bar).

Besides the intramolecular PDZ domain interactions to form a supramodule, intermolecular interactions between PDZ domains of different proteins are also reported (Chang, 2011). hDlgPDZ2 is known to tightly bind to the single PDZ domain of nNOS (K_D 0.8 µM; Chang, 2011). Therefore it could be interesting to address a potential interplay of nNOS/hDlg/E6 on the cellular and also on the structural level.
3.5 More interactions?

Additional PDZ domains possibly targeted by E6

If one protein interacts with a specific PDZ domain in a PDZ-BM dependent manner, the same protein should target similar PDZ domains as well. For example, HTLV-1 (Human T-cell Lymphotropic Virus-1) tax is another viral protein known to bind hDlg via its PDZ-BM. Since the PDZ domains of hDlg and of pro-interleukin-16 share a significant sequence homology (Wilson, 2003), it was proposed and subsequently demonstrated that HTLV-1 tax also binds pro-interleukin-16 (pro-IL-16; Wilson, 2003). This raises the question whether high-risk E6 interacts with pro-IL-16 as well. (pro-)IL-16 is the only human interleukin containing PDZ domains, according to the UniProtKB database. So far, no reports addressed IL-16 in the context of an HPV infection (according to PubMed). HPV-infection of keratinocytes is known to affect other interleukins; these keratinocytes express reduced-levels of inflammatory interleukins IL-1 and IL-6 and elevated levels of the anti-inflammatory IL-10, which could positively contribute to HPV immune evasion (reviewed in Bodily, 2011). pro-IL-16 is cleaved by caspase 3 (Baier, 1997; Zhang, 1998) into native IL-16 and an N-terminal domain. The N-terminal domain translocates into the nucleus and induces cell cycle arrest in G0/G1 (Wilson, 2002). Native IL-16 is secreted and a potent chemoattractant for CD4-positive immune cells such as e.g. dendritic- or T-cells (reviewed in Cruikshank, 2000). In epithelial cells, pro-IL-16 expression and processing into active IL-16 is inducible, a process linked to inflammation (reviewed in Cruikshank, 2000). Interestingly, the HTLV-1 tax protein, interacting with both hDlg and pro-IL-16, has a PDZ-BM with the residues ETEV, identical to the PDZ-BM of HPV 35 E6 and quite similar to the E6 consensus PDZ-BM of ETQV (e.g. publication 3, figure S3). Similar to E6, the tax-protein does harbor basic residues upstream of the canonical PDZ-BM (C-terminal sequence of tax: EKHFRETEV) that could, as for E6 (publication 3), further contribute to binding affinity. However, the bulky aromatic residues of the tax C-terminus (His, Phe) are absent in E6 and if binding of E6 to pro-IL-16 does not occur, the reason may lie in the differences of the extended PDZ-BM. If an interaction does occur, it could provide an additional way of how E6 contributes to HPV immune evasion. In such a scenario the E6-pro-IL-16 interaction might lead to degradation of pro-IL-16 possibly by involvement of the E3 ubiquitin ligase E6AP thereby reducing the precursor levels for both native, chemoattractant IL-16 and for the amino-terminal pro-IL-16 cleavage product that
3.5 More interactions?

stalls the cell cycle. This could in turn reduce the cellular immune response as fewer T-cell migrate to and target HPV infected epithelia and could lead in addition to abrogation of induced G0/G1 cell cycle arrest by the amino-terminal processed portion of pro-IL-16.

**E6 and different interaction motifs**

E6 has no known enzymatic activity, but interacts with a multitude of cellular interactors (Tungteakkhun, 2008). Up to now, implications of E6’s extended PDZ-BM have been discussed, but there are more motifs contained within or recognized by E6. The high-risk E6 proteins contain a hitherto unnoticed, conserved PXXP motif and conserved residues in the vicinity that could allow for an interaction with SH3 domains (publication 3; Ball, 2005; Freund, 2008). Moreover, many proteins with which E6 interacts, contain an LXXLL motif, e.g. E6AP, E6BP and paxillin (Huibregtse, 1993; Heery, 1997; Torchia, 1997; Tungteakkun, 2008) and during the preparation of this thesis, the structural basis for this interaction was presented utilizing mutated, HPV 16 E6 and an E6AP-derived LXXLL peptide (Zanier, 2013). The E6-E6AP interaction requires E6 residues of both ZBDs (Liu, 1999; Zanier, 2013). The implications of a different local E6 structure around I128, a key residue for LXXLL binding have already been discussed. Because the binding of E6 to LXXLL motifs involves both ZBDs of E6 (Liu, 1999), an investigation of the affinity of purified E6 proteins to E6AP requires full-length E6 that is still not available as wild-type sequence. Zanier et al. generated their E6-E6AP(LXXLL-peptide) complex by expression and purification of an MBP-LXXLL fusion and mixing this fusion with mutation-stabilized HPV 16 E6 (Zanier, 2013) and subsequently solved the structure of the complex by X-ray crystallography. These E6 mutations however, render the E6 incapable of E6AP-mediated p53 degradation (Ristriani, 2009). It might still be possible in the future to solubilize wild-type full-length E6 by co-expression with extended regions of E6AP. Such an extended E6-LXXLL complex would contain a larger E6AP region and would be suited to unravel the structural basis for the E6-mediated target degradation. Vectors encoding the full-length E6AP protein (residues 1-865), the E6AP region (residues 280-865) sufficient for p53 degradation in presence of E6 and the E6AP region (residues 280-781) still binding but not polyubiquitinylating p53 in presence of E6 have been constructed according to (Huibregtse, 1993) during my work. As a first step, I have expressed the respective
proteins in *E. coli*. The 280-865 and 280-791 constructs appear particularly promising, since their expression levels are high and they are predominantly soluble. Interestingly, 13 out of 25 high-risk E6 proteins contain themselves an LXXLL motif (publication 3, Figure S3). 51Z2 has the sequence LYDLS (residues 96 to 100) and thus lacks an intact LXXLL motif. Moreover, L96 and L99 point ‘inward’ and contribute to the formation of the 51Z2 core. However, in the light of the conformational plasticity of 51Z2 and possibly of other E6 proteins at physiological temperatures, it seems reasonable to assume that an LXXLL motif of E6 transiently adopts an alternative conformation and becomes solvent-exposed. A transiently solvent-exposed LXXLL motif on E6 might be involved in the reported direct binding of HPV 18 E6 (containing LYNLL) to the nuclear receptors estrogen receptor, thyroid receptor and androgen receptor (*Wang*, 2003), as LXXLL is a common motif for binding of proteins to nuclear receptors (*Savkur*, 2004). Since E6 interacts with LXXLL containing proteins (*Huibregtse*, 1993; *Heery*, 1997; *Torchia*, 1997; *Liu*, 1999; *Tungteakkun*, 2008), a self-association mechanism of E6 involving this motif seems likely. HPV 16 E6 experiences an increased stability *in vivo* in presence of commonly E6 binding LXXLL-containing peptides (*Ansari*, 2012). Thus, E6 self-association may lead to increased stability of E6 *in vivo*.

Collectively, the presented data imply that many motifs contained within or recognized by E6 are short linear interaction motifs (SLIMs), *i.e.* short linear sequences that are usually involved in cellular protein-protein interactions. Viral proteins often mimic or target SLIMs in order to achieve interactions between viral and cellular proteins, thereby hijacking cellular interaction modules in order to reprogram infected cells for viral genome maintenance and production of infective virions (*Davey*, 2011). In case of HPV, the combined action of the oncoproteins E6 and E7 to reprogram infected keratinocytes can result in cancer (*Assmann*, 2011; *Stanley*, 2012; *Doorbar*, 2012). Further research on these oncoproteins will have to unravel the mechanism of how these oncoproteins exert their functions, *i.e.* their interactions. Since these viral proteins exploit cellular SLIMs (*Davey*, 2011), a detailed mechanistic understanding of the viral protein interactions will also deepen the insight into cellular protein-protein interactions.
4 Summary

The oncoproteins E6 and E7 of high-risk human papillomaviruses (HPVs) interact with cellular key regulators. The resulting reprogramming of HPV-infected keratinocytes facilitates viral replication and can lead to cervical cancer and other malignant conditions. E6 induces degradation of the tumor suppressor p53, activates telomerase and deregulates cell polarity. Here, the first structure of a wild-type HPV E6 domain (termed 51Z2) is presented (PDB 2M3L). Together with available structures of mutated HPV 16 E6 and of bovine PV 1 E6, the 51Z2 structure allows to pinpoint (H)PV-type specific structural differences between E6 proteins. The temperature sensitivity of the well-folded wild-type E6 domain implies conformational plasticity of E6 in vivo. Hence, the structural differences between individual E6 and their malleability appear, together with HPV type-specific solvent-exposed side-chains, to provide the structural basis for the different HPV type specific interaction networks reported for the respective E6 proteins. Surface plasmon resonance and NMR spectroscopy indicated that 51Z2 residues N-terminal to the canonical PDZ* binding motif (PDZ-BM) of E6 significantly contribute to the interaction with PDZ domain 2 of hDlg (hDlgPDZ2) and enhance affinity. hDlg constitutes a prototypic representative of the large family of PDZ domain containing proteins, regulates cell polarity and is usually degraded by an E6-mediated mechanism. In order to unravel the underlying structural basis for the extended interaction, a peptide (E6CT11) containing the hDlg-interacting E6 residues was generated by recombinant expression utilizing the intein system. The N-terminal Gln residue of E6CT11 underwent spontaneous cyclization into pyroglutamate (pGlu). This reaction is potentially useful for biotechnological applications. Being accidental in our system, the pGlu modification does not interfere with the E6CT11-hDlgPDZ2 interaction. The NMR solution structure of the hDlgPDZ2-E6CT11 complex (PDB 2M3M) reveals how an extended PDZ-BM forms additional contacts to a PDZ domain, thereby enhancing affinity. An extended PDZ-BM might also be utilized by other viral proteins in order to successfully compete with cellular proteins for their respective PDZ targets or by cellular proteins to increase affinity towards their targets. Dimeric or aggregated E6 could have an increased affinity towards a subset of ‘quasi-quaternary’ PDZ superstructures. Based on the above findings, additional proteins to which E6 binds or could bind are presented and the role of short linear interaction motifs contained within or recognized by the viral oncoprotein E6 is highlighted.

*PDZ domain: Acronym of the first proteins demonstrated to contain this domain: Post synaptic density-95, drosophila Discs large and Zona occludens-1.
5 Zusammenfassung


* PDZ-Domäne: Akronym der ersten Proteine, in denen diese Domäne nachgewiesen wurde: Post synaptic density-95, drosophila Discs large und Zona occludens-1.
6 References

References of this dissertation including citations of publications are listed below, sorted first by first author name and second by publication year.


References


Keller R, *The Computer Aided Resonance Assignment Tutorial*, CANTINA Verlag, **2004**.


6 References


References


6 References


Acknowledgements – Danksagung

I wish to express my sincere gratitude to my supervisor, Dr. Matthias Görlach, who not only offered me the opportunity to work on the structure and interactions of a challenging viral protein, but also offered invaluable advice, support and guidance and always lent me a sympathetic ear. Special thanks also go to the other members of my thesis committee, Prof. Dr. Frank Große and Prof. Dr. Matthias Dürst for fertile discussions and for supervision.

I am also extremely indebted to my colleagues of the Biomolecular NMR spectroscopy group at the FLI for introducing me to the world of structure biology, for providing a fruitful working atmosphere and for comments and discussions on manuscripts and my thesis. Especially I want to thank my colleagues Yvonne Ihle, Thomas Seiboth, Marina Baum, Sabine Häfner, Angelika Heller, Christiane Hirsch, Dr. Oliver Ohlenschläger, Dr. Jörg Leppert and Dr. Ramadurai Ramachandran.

My collaborators Dr. Bernhard Schlott, Dr. Karl-Heinz Gührs, Prof. Dr. Stephan Diekmann and Dr. Peter Hortschansky are also much appreciated. Finally I want to thank the FLI for providing such a stimulating, interdisciplinary atmosphere and for funding. I will now switch to German for private acknowledgements.

8 Curriculum vitae

André Mischo

geboren am 23.02.1982 in Bottrop

ledig

Berufserfahrung

2007 - heute  
Doktorand am Leibniz Institut für Altersforschung, Fritz-Lipmann-Institut e.V. (FLI), Jena

Dissertation im Rahmen der Leibniz Graduate School on Ageing and Age-Related Diseases

• Angestrebter Abschluss: Dr. rer. nat. (Friedrich-Schiller-Universität Jena)

• Thema: Structural insights into the interaction of a high-risk human papillomavirus wild-type oncoprotein E6

• Betreuer: Dr. Matthias Görlach

Studium

2002 - 2007  
Universität Bielefeld, Studiengang: molekulare Biotechnologie

Schwerpunkte inklusive Praktika: Bioanalytik, Biokatalyse, Molekulare Medizin, Prozessmesstechnik, Aufarbeitung biotechnologischer Produkte

Diplomarbeit: Production and characterization of proprotein convertase-inhibiting serpin variants

Abschluss: Diplom Biotechnologe, Note 1,1 (sehr gut)
Wehrdienst

2001 - 2002 Grundwehrdienst beim PzGren Btl. 192, Ahlen

Schulbildung

2001 Abitur am Helmholtzgymnasium Bielefeld

Bester des Abschlussjahrganges 2001: Note 1,2 (sehr gut)

Weiterbildungen

2012 Sachkunde gemäß §5 ChemVerbotsV

2011 Scientific Communication

2011 Technologietransfer-Kolloquium

2010 Isothermal Calorimetry Training

2009 International School of Biological Magnetic Resonance, Biophysics and Structure

2009 Medientraining

2008 Speed Reading

Präsentationen

2012 Structure and Interactions of the HPV Oncoprotein E6 (Vortrag & Poster*)

2011 Reprogramming ribosomes - a first step towards a new era of structural biology? (Vortrag)

Structure and Interactions of the HPV Oncoprotein E6 (Vortrag & Poster*)

From Proteins to Cancer: Structure and Interactions of the HPV Oncoprotein E6 (Poster*)

2010 Molecular Coin Slots (Vortrag)

Towards a Structure of the HPV Oncoprotein E6 (Vortrag*)
Structure and Interactions of the HPV Oncoprotein E6
(Vortrag*)

From Proteins to Cancer: the E6 Piece of the HPV Oncogenicity Puzzle (Poster*)

2009

Trapping Moving Targets with Small Molecules (Vortrag)

Structure Determination of the HPV Oncoprotein E6
(Vortrag*)

Strukturbestimmung des HPV Onkoproteins E6 (Vortrag*)

Towards a Structure of the HPV Oncoprotein E6 (Poster*)

2008

*p53 represses human papillomavirus type 16 DNA replication via the viral E2 protein (Vortrag)

Towards a structure of the HPV oncoprotein E6 (Vortrag*)

The HPV oncoproteins E6 and E7 (Poster*)

2007

Production and characterization of proprotein convertase-inhibiting serpin variants (Vortrag)

Mechanisms of Alternative Splicing - Generation of Protein Diversity (Vortrag)

Structure and interactions of the oncoprotein E6 (Vortrag*)

Novel Interaction between HPV E6 and BARD1 (BRCA1-Associated Ring Domain 1) and Its Biologic Roles (Vortrag)

Die mit * markierten Vorträge/Poster behandelten eigene Forschungsergebnisse.
9 Ehrenwörtliche Erklärung


Mir ist die geltende Promotionsordnung der biologisch-pharmazeutischen Fakultät der Universität Jena bekannt. Ich versichere demgemäß, dass ich keine Hilfe eines Promotionsberaters in Anspruch genommen habe und Dritte weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten, die im Zusammenhang mit dieser Dissertation stehen erhalten haben.

Die vorgelegte Dissertation wurde weder in gleicher noch ähnlicher Form bei einer anderen Fakultät oder Hochschule für eine andere staatliche oder andere wissenschaftliche Prüfung eingereicht.

Jena, den 14.04.13

André Mischo
1. Die vorgelegte Arbeit beschreibt die erste Struktur einer Wildtyp-Domäne des Onkoproteins E6 aus humanen Papillomaviren (HPV) sowie die Interaktion dieser Domäne mit einem prototypischen PDZ*-Protein.

2. Die hier aufgeklärte NMR-Struktur der Wildtyp-Domäne von E6 in Lösung erlaubt es, lokale strukturelle Unterschiede zwischen E6-Proteinen abzuleiten, die möglicherweise die unterschiedlichen Interaktionen von E6-Proteinen verschiedener HPV-Typen bedingen.


5. Das erweiterte, lineare PDZ-Bindemotiv von E6 stellt eine Möglichkeit dar, wie auch andere, virale oder zelluläre, Proteine die Affinität gegenüber ihren Ziel-PDZ-Domänen steigern könnten.


* PDZ-Domäne: Akronym der ersten Proteine, in denen diese Domäne nachgewiesen wurde: Post synaptic density-95, drosophila Discs large und zona occludens-1.