Characterization and interaction of the HPV16 major capsid protein L1 and minor capsid protein L2

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Nora Isabel Roos

aus Stuttgart

Tübingen

2022

Table of Content

I. ZUSAMMENFASSUNG

Das Kapsid des Humanen Papillomavirus (HPV) Typ 16 besteht aus 360 Kopien des monomeren Kapsidproteins L1 welches in 72 Pentameren in einer ikosaedrischen Symmetrie von T = 7 angeordnet ist. Die L1-Pentamere können sich selbst zu virusähnlichen Partikeln assemblieren, die erfolgreich als prophylaktischer Impfstoff gegen HPV-Infektionen eingesetzt werden. Das Kapsidprotein L2 befindet sich im inneren Teil des HPV16 Kapsids und die Anzahl variiert zwischen 12-72 L2 Monomeren. Diese Anordnung von L1 und L2 lässt eine Interaktion der beiden Kapsidproteine annehmen. Diese wurde bereits in zellbasierten pull-down Experimenten und zusätzlich qualitativ auf Strukturebene gezeigt. Allerdings ist bis zum jetzigen Zeitpunkt nicht bekannt mit welcher Bindungsaffinität diese Interaktion stattfindet und inwiefern verschiedene Einflüsse auf zellulärer Ebene eine Rolle spielen. Die Charakterisierung der L1 und L2 Interaktion ist wichtig für ein besseres Verständnis des viralen Eintrittes von HPV16 in die Zelle. Um die Interaktion von L1 und L2 zu charakterisieren und die beiden Kapsidproteine auf biochemischer Ebene zu analysieren, optimierten wir die rekombinante Proteinexpression in *E. coli* sowie die Aufreinigung der viralen Strukturproteine. Für Interaktionsstudien wurde eine L1-non-assembly-Variante (L1nav) verwendet da diese Variante nur L1-Pentamere ausbildet und keine Kapside. Anschließend erfolgte die Charakterisierung des L1nav und L2 Proteins in Bezug auf: Stabilität, sekundäre Protein-Struktur, Oligomerisierungszustand und Homogenität unter verschiedenen Pufferbedingungen. Nachfolgend kann die Bindungsaffinität (K_d) der gereinigten Kapsidproteine mittels verschiedener Methoden (EL(2)ISA, SPR oder Anisotropie Messung) und unter verschiedenen Bedingungen qualitativ und quantitativ ermittelt werden. Die Ergebnisse zeigen eine Interaktion von L1 und L2 unter pH 7.4 mit einem K_d von 620 nM ± 296 nM im Gegensatz dazu zeigt sich bei pH 5.5 ein K_d von 36 nM ± 69 nM. Bereits publizierte Ergebnisse zeigen, dass der pH-Wert in infiziertem Gebärmutterhalsgewebe mit der HPV-Infektion ansteigt und dies vermutlich die Stabilität von L1 verringern kann. Diese Stabilitätsverminderung kann dazu führen, dass eine Konformationsänderung begünstigt wird und der N-Terminus von L2 exponiert wird. Diese Konformationsänderung spielt eine wichtige Rolle bei dem viralen HPV Eintritt in die Zelle. Zusätzlich zeigen unsere Ergebnisse einen ähnlichen K_d für eine L2 Mutante. Dieser L2 Mutante (L2ΔL1bs) fehlt die L1 Bindungsdomäne und aus diesem Grund sollte L2ΔL1bs nicht mehr an L1 binden. Allerdings bindet L2ΔL1bs mit 392 nM ± 229 nM im Vergleich zum L2 volllängen Protein bei pH 7.4. Dies lässt darauf

schließen, dass eine zweite L1 Bindungsdomäne auf dem L2 Protein vorhanden sein könnte. Diese Ergebnisse sind für das tiefere und biochemische Verständnis des viralen Eintritts in die Zelle wichtig. Auch konnten wir zusätzliche Interaktionspartner (L1nav Monomer und CypA) von L2 durch das EL(2)ISA „Screening-Instrument" ausfindig machen. Bis zu diesem Zeitpunkt wird davon ausgegangen, dass nicht alle Interaktionspartner von L2 während des viralen Eintritts in die Zelle bekannt sind. Nach einem ersten „Screening" kann eine weitere Charakterisierung der Interaktion, in Bezug auf Bindungsaffinität oder Komplexeigenschaften, erfolgen. Das vollständige Verständnis des intrazellulären Transportes von HPV16 in den Zellkern ist wichtig, da höchstwahrscheinlich nicht nur ein Virustyp diesen Transportweg kapert. Aus diesem Grund tragen Studien über virale Proteine seit langem zur Entdeckung und Charakterisierung von zellulären Proteinen und Signalwegen bei.

II.   ABSTRACT

The Human Papillomavirus (HPV) Type 16 capsid consists of 360 monomer copies of the major capsid protein L1 and is arranged in 72 pentamers to a T = 7 icosahedral symmetry. L1 pentamers itself can assemble to virus-like-particles, which are successfully used as a prophylactic vaccine against HPV infections.

The minor capsid protein L2 is located in the inner part of the HPV 16 capsid and the amount differs from 12-72 L2 monomer copies. This arrangement of L1 and L2 suggests an interaction of these two capsid proteins. This interaction has already been shown in cell-based experiments but also partially qualitatively at the structural level. However, until now it is not known with which binding affinity this interaction takes place and to what extent different influences on the cellular level play a role. Characterization of the L1 and L2 interaction is important for a better understanding of the viral entry of HPV16 into the cell.

To characterize the interaction of L1 and L2 and to analyze the two capsid proteins at the biochemical level, we optimized recombinant protein expression in *E. coli* and purification of the viral structural proteins. For interaction studies, a L1-non-assembly variant (L1nav) was used since this variant forms only L1-pentamers and no capsids. Subsequently, characterization of the L1nav and L2 protein was performed in terms of stability, secondary protein structure, oligomerization state and homogeneity under different buffer conditions. Subsequently, the binding affinity (Kd) of the purified capsid proteins can be determined qualitatively and quantitatively by different methods (EL(2)ISA, SPR or anisotropy measurement) and under different conditions.

The results demonstrate an interaction of L1 and L2 under pH 7.4 with a binding affinity of 620 nM ± 296 nM in contrast, pH 5.5 shows a binding affinity of 36 nM ± 69 nM. Previously published results indicated that pH in infected cervical tissue increases with HPV infection and this is thought to reduce the stability of L1. This decrease in stability may lead to promote a conformational change and expose the N-terminus of L2. This conformational change plays an important role in viral HPV entry into the cell. In addition, our results show a similar K_d for an L2 mutant. This L2 mutant (L2ΔL1bs) lacks the L1 binding domain and for this reason L2ΔL1bs should no longer bind to L1. However, L2ΔL1bs binds at 39 nM ± 229 nM compared to the L2 full-length protein at pH 7.4. These results suggest that a second L1 binding domain may be present on the L2 protein. These results are important for deeper and biochemical understanding of viral entry into the cell. We were also able to locate additional interaction

partners (L1nav monomer and CypA) of L2 by the EL(2)ISA "screening tool". Up to this point, it is assumed that not all interaction partners of L2 are known during viral entry into the cell. After initial "screening", further characterization of the interaction, in terms of binding affinity or complex properties, can be performed. Fully understanding the intracellular transport of HPV16 into the nucleus is important, as most likely not only one type of virus hijacks this transport pathway. For this reason, viral protein studies have long contributed to the discovery and characterization of cellular proteins and signaling pathways.

III.   LIST OF ABBREVIATIONS

Amp	Ampicillin
AB	Antibody
APS	Ammonium Peroxodisulfate
ATP	Adenosine triphosphate
AUC	Analytical Ultracentrifugation
Biotin-L2	Biotin labelled L2
BSA	Bovine Serum Albumin
Cam	Chloramphenicol
CMC	Critical Micelle Concentration
CV	Column volume
DoB	Degree of Biotinylation
DoL	Degree of Labeling
FL-L2	Fluorescein labelled L2
G-factor	Device factor
GMP	Good manufacturing practise
GroEL/ES	Chaperon complex consisting of larger protein GroEL and smaller Protein GroES
GST	Glutathione S-transferase
His$_6$	Hexa-Histidin-Tag
HPV	Human papilloma virus
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl-β-D-thiogalactopyranosid
Kan	Kanamycin
K$_d$	Dissociation constant
LB	Luria-Bertani Medium
L1nav	L1 non-assembly variant
L1/L2	L1/L2 complex
L2/vDNA	Complex of L2 and viral DNA
L2ΔL1bs	L2 mutant
min	Minutes
MWCo	Molecular weight cut off
Ni-NTA	Nickel-Nitrilotriacetic acid
o/n	over night
OD$_{600}$	Optical density at 600 nm
RLU	Relative Light Unit
rpm	Revolutions per minute
RT	room temperature
SDS PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
s	Seconds
SEC	Size exclusion chromatography
SOC	Super Optimal broth with Catabolite repression
SPR	Surface Plasmon Resonance
Srt A	Sortase A
SUMO	Small Ubiquitin-Related Modifier
TB	Terrific-Broth-Medium
TEM	Transmission electron microscopy
TGN	Tans-Golgi Network
VLP	Virus-like particle
v/v	Volume per volume
w/v	Weight per volume
UC	Ultracentrifugation

1 INTRODUCTION

1.1 Human Papillomavirus: general information & gene regulation

The human papillomavirus (HPV) belongs to the family of *Papillomaviridae* and causes diseases that vary from warts to cancer [8]. HPV is a non-enveloped, double stranded DNA virus with a genome of approximately 8 kbp and is encapsulated by two capsid proteins [6]. The genome is divided into the early and late genes. HPV is classified, based on the L1 protein sequence, into two main phylogenetic genera called α- and β-HPV [8]. In addition, there are three further genera called *γ-, μ-, ν*-HPV [9]. At present, more than 200 HPV types from different genera are known (https://pave.niaid.nih.gov/). The largest group is the alpha genera, which can infect both mucosal as well as cutaneous tissues. The alpha HPV types that infect the mucosal tissue are best studied and can be divided based on their carcinogenic potential into high-risk and low-risk types [10, 11]. Thereby, one of the best described diseases caused by HPV alpha type is cervical cancer, the fourth common cancer in women worldwide [12, 13] and the second most common cancer in less developed region [12]. Worldwide, 570.000 cases in women can be attributed to HPV each year [14]. The numbers are different in less developed countries and developed countries [15].

The viral genome is divided in three different regions: the early genes (E1-E7) which encode non-structural regulatory proteins. These play an important role in the viral replication and manifestation of the virus in the cell. The late genes (L1 and L2) encode the structural capsid proteins and play an important role in viral release [16] and the long control region (LCR) which includes promotor, enhancer and origin of replication as shown in figure 1.

Viral Protein	Functions and Features
E1	It forms a heterodimer complex with E2 and controls viral replication
E2	It regulates early gene promoter and together with E1 viral DNA replication
E4	It may mediate the viral particle realease by destabilizing cytokeratin network
E5	It stimulates mitogenic signals of growth factors
E6	It inactivates many cellular proteins and is one of the major viral oncoproteins
E7	It inactivates many cellular proteins and is one of the major viral oncoproteins
L1	It is the major capsid protein and is the component of the HPV prophylactic vaccine
L2	It is the minor capsid protein

Figure 1: Organization and explanation of the HPV16 genome. Left: Location of the different genes in the genome. Right: function of the encoded protein (Tommasino 2014 [7])

1.2 Major capsid protein HPV16 L1

The outer shell of the HPV virion consists of 72 pentamers of the major capsid protein L1 which build up the 55 nm large capsid predicted in gray in figure 2 A and in blue in figure 2C [6]. These pentamers are arranged in a T = 7d icosahedral [17, 18] symmetry. The L1 protein intrinsically can self-assemble to virus-like particles (VLPs) which are used as prophylactic vaccines [19]. These VLPs contain no viral genetic material and no minor capsid protein L2 and are therefore non-infectious.

The L1 monomer (58 kDa) as shown in figure 3 A is a classical "jelly roll" ß-sandwich which is demonstrated by eight ß-strands arranged in two four-stranded sheets (turquoise: B-I strands figure 3 A) and connected via loops (purple loops figure 3 A) [4, 20]. The alpha helices are
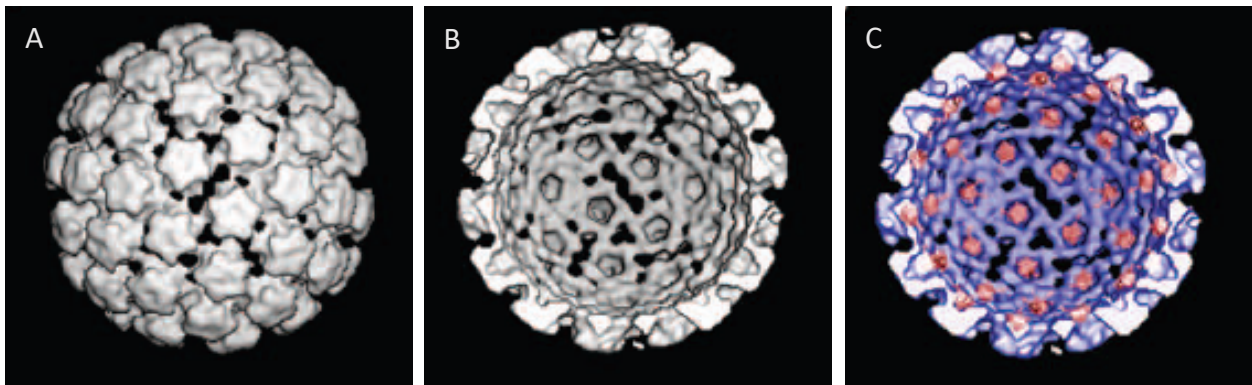
Figure 2: Capsid structure of HPV16 with HPV16L2. L1 protein in shown in grey/blue and L2 in red. Computerized reconstruction of L1 only capsid in A and B shown and L1+L2 capsid in C. (modified from Buck 2008 [6])
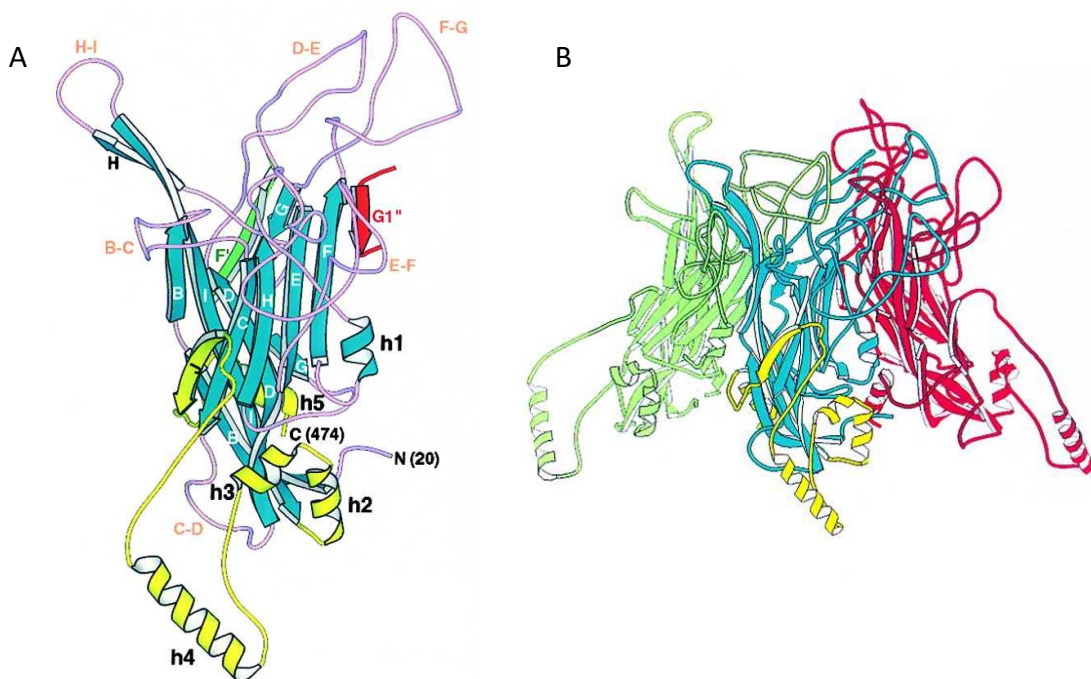
Figure 3: Tertiary structure of HPV16 L1. (A) L1 monomer from amino acid 20-474. The helices as well as the loops are labeled (h1-h5; B-I). **(B)** L1 pentamer shown on the 5-fold axis, for a better overview the two monomer subunits in the background were deleted (modified from Chen 2000 [4])

2

represented in yellow and labeled as h1-h5 (figure 3 A). The C - termini (residue 383-474 are significantly alpha-helical, and the helices h2, h3 and h4 are located on the pentamer surface to contact neighboring pentamers (figure 3). Additionally, the last 31 amino acids at the C-terminus are disordered and it is shown that this part is important for VLP assembly [21]. In addition, the N-terminus and the C-terminus serve as connectors between the L1 pentamers as well [4]. For interaction studies in this work, the N-terminus, C-terminus and the alpha helical h4 part of L1 were removed to prevent the assembly of VLPs and urge the production of L1 pentamers only [4]. Therefore, the variant is called L1 non-assembly variant (L1nav). Another component of HPV is the minor capsid protein L2 which is located in the inner area of the virion presumably close to the L1 major capsid protein as shown in figure 2 C (blue HPV16L1, red HPV16L2) [6]. So far it is not clear how many L2 proteins are present in one capsid. The range differs in between 12-72 L2 monomers [6, 22].

1.3 Minor capsid protein HPV16 L2

The minor capsid protein L2 is a key player during the viral infection. It plays a role in viral entry, uncoating, internalization and in trafficking the virion from the endosome to the nucleus [23]. It is highly conserved among pseudo virions and therefore enables a cross recognition of neutralizing antibodies between different HPV types [24].

The protein has a size of 55 kDa and consists of 473 amino acids [25, 26]. Furthermore, predicted structure analysis of L2 shows it is basically an unstructured

Figure 4: predicted structure of HPV16L2 protein based on the amino acid sequence by AlphaFold (alphafold.ebi.ac.uk)

disordered protein (figure 4). The predicted structural elements in figure 4 are shown in matching colours to figure 5. However, the calculations and further analyses indicate an alpha helical N-terminus. A furin cleavage site is located on aa 1-12 (figure 5) [27]. The amino acid sequence 45-67 predicts a membrane spanning region (transmembrane domain) which allows interaction and insertion into membranes (figure 5). The synthesized peptide (aa 45-67) shows

in vivo an alpha helical structure by a micellar state [28]. The predicted structure in figure 4 also shows an alpha helical region for aa 45-67 (shown in turquoise).
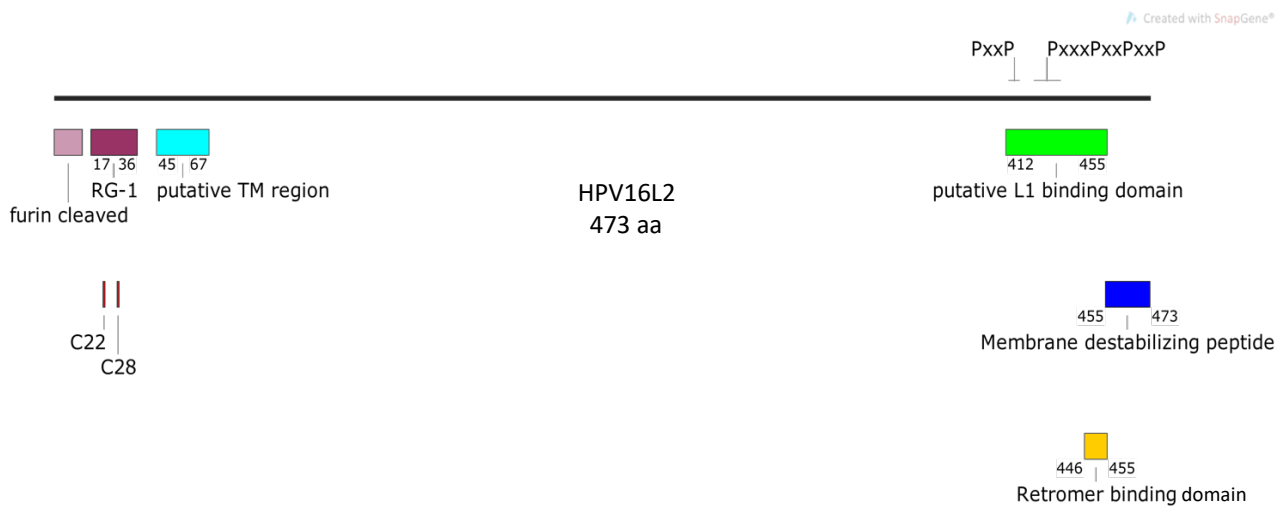
Figure 5: HPV16L2 gene sequence (created with Snap Gene). Important and described parts of the L2 genome; N-terminal and C-terminal

The putative L1-binding region is located from aa 412-455 [29] at the predominant unstructured C-terminus of L2 (figure 5). This L1-binding region is characterized by several PxxP motifs. Similar PxxP motifs have been associated with protein-protein interaction in several systems [23]. Furthermore, these PxxP motifs were observed in the similar C-terminal region of L2 from different HPV types which suggests a highly conserved region. At the N-terminus two highly conserved cysteines (C22 and C28) are located (figure 5). Further studies indicate that these cysteines might play a key role in late-stage stabilization of the HPV capsid. This was found out in studies modulating the accessibility of L2 on the surface of raft-derived virions [23, 30]. In addition, further studies show that individual regions (aa 32-81, 212-231, 272-291 and 347-381) of L2 are presented at the surface of the L1/L2 virion [31]. As already described the RG- 1 epitope, located at the L2 N-terminus (aa 17-36) can be recognized by a monoclonal antibody and is accessible on immobilized virions but not on virions in solution [32]. This led to the assumption that the RG-1 epitope is exposed after binding to the extracellular matrix and furin cleavage of L2 [33]. On the last aa 455-473 of the C-terminus, the so-called membrane destabilisation peptide (MDP) is located (figure 5). This part plays probably a key role in endosomal escape [34]. The main domain sequences of L2 have been mapped by mutation and deletion of the different parts and very little is known about the higher structure of L2 due to the largely unstructured protein structure [23].

1.4   Medical application of capsid proteins - HPV prophylactic vaccine

Nowadays, there are different prophylactic vaccines against HPV infection available. These vaccines are based on virus-like particles of different HPV types and can efficiently prevent an HPV infection of the respective included HPV type [35]. The various L1 proteins of different HPV-types are expressed for this purpose either in yeast, insect cells or *E. coli* [36-39]. At this point in time available vaccines are: Cervarix® (GSK, HPV 16 and 18) produced in insect cells, Gardasil® (Merck& Co., HPV 6, 11, 16, 18) and Gardasil-9® (Merck&Co., HPV 6, 11, 16, 18, 31, 33, 45, 52, 58) produced in yeast and Cecolin® (Innovax, HPV 16 and 18) produced in *E. coli* [39-43]. Another vaccine that is currently in clinical stage I trials [44] is developed by inserting the so-called RG-1 peptide (aa 17-36) of HPV16L2 into the DE-surface loop of HPV16L1 [45, 46]. This novel vaccine results in a chimeric HPV16L1-HPV16L2 virus-like particle (RG1-VLP) [45]. It had been demonstrated that the RG-1 peptide of L2 is conserved in several HPV types [46] and can mediate cross-neutralization via low-titer antibodies and a B-cell epitope that is recognized by mAb RG1. This was observed in *in vivo* studies in animal models [46]. Due to its high conservation within many HPV types, the RG-1 epitope may represent an attractive target for the development of a wide-spectrum HPV vaccine [47, 48].

Even though there are already good prophylactic vaccines available, and the cost of vaccination is covered by health insurance in developed countries, the latest numbers from the RKI from 2015 show that only 31.1% of people in Germany over 15 years are fully vaccinated and the numbers in the different federal states differ extremely in Germany [49]. Because of this low vaccination rate and resulting further spread of HPV infections, continuous research on HPV is important.

1.5 Viral entry of HPV16

The viral entry of HPV16 is not completely understood but the currently prevailing model suggests, that the capsid proteins L1 and L2 play a key role in the process of viral entry (figure 6) [50, 51].
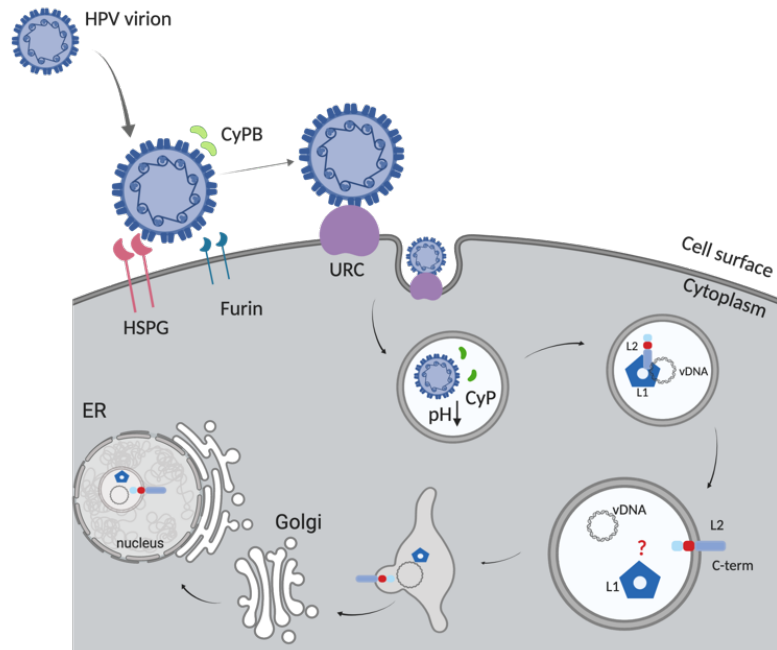
Figure 6: Model for viral entry of HPV16 (created with BioRender). HSPG receptor binding of viral capsid followed by furin-dependent cleavage and binding to a secondary receptor complex with subsequent receptor mediated endocytosis. The endosome migrates through the TNG/ER to the nucleus.

1.5.1 Receptor binding of HPV capsid

The current published data does not provide a full explanation on viral entry and all models shown are up for debate [33]. However, the general opinion is that the first step of viral entry includes the interaction of the viral capsid protein L1 with the heparin sulfate proteoglycans (HSPG) receptor [52, 53] either on the surface or on the extracellular matrix [33, 54, 55]. It has been shown that the interaction on the HSPG receptor with the virion is taking place at the amino acid 278 and 361 of HPV16L1 capsid protein [56, 57]. In addition, it was shown that modifications e.g., sulfation of the HSPG receptor have a crucial effect on HPV binding. In particular, this applies to HPV16 and HPV33 [58]. The receptor binding of L1 to HSPG leads to a conformational change of the L1 protein and this event is thought to take part in the transfer of virions to a secondary receptor [59].

1.5.2 Exposure of L2 N-terminus, furin cleavage and interaction with secondary receptor complex

After receptor attachment and the additional interaction with cyclophilin B, the N-terminus of L2 is exposed and afterwards cleaved independently by furin (aa 1-12) [60-62]. Cyclophilin B is a peptidyl-prolyl cis-/trans-isomerase and associated with syndecan-1 in the extracellular part [63, 64]. This step is necessary for cellular entry. However, it is shown that the entry can take place without furin cleaved L2 but the viral capsid proteins are then trapped in the endosome [32, 60]. In addition, kallikrein-8 (KLK8), a secreted trypsin-like serine protease, has been shown to cleave the L1 capsid independently of cyclophilins and furin [65]. This cleavage facilitates conformational changes of the minor capsid protein L2 and which subsequently exposes an important epitope of L2. Due to that, L1 cleavage facilitates access to L2 [66]. Published data hypothesizes that particle binding causes the L2 aa 13-31 motif to be displayed on the capsid surface. Afterwards the capsid is interacting with a second receptor complex (assumed: CD151 tetraspanin, α-6-β-1 integrin, annexin A2 heterotetramer (A2t), obscurin-like 1 (OBSL1) [67-71]. Binding to the second receptor complex results in endosomal uptake of the virus capsid; however, it is unclear whether L1 or L2 binds to the secondary receptor complex. Moreover, except for furin cleavage, the role of L2 for internalization is unclear; nevertheless, *in vivo* studies show an interaction of L2 (108-120 aa) with the cellular membrane. This leads to the assumption that the N-terminus of L2 is at least exposed up to aa 120 and could interact with the secondary receptor complex [72].

1.5.3 Receptor-mediated endosomal uptake and endosomal escape

The endosomal uptake is the next important step and independent of caveolin and dynamin [73] but dependent on clathrin [34]. Clathrin-dependent endocytosis is characterized as endocytosis which, subsidiary on the transport signal, is additionally associated with another specific type of endocytosis, e.g. lysosomal pathway [74]. Thereby the virion overcome the cellular membrane and is located afterwards in the endosome, where the pH drops to 5.5. This acidification together with cyclophilin interaction is important in virus uncoating [62, 75]. Most of the L1 protein segregates from the L2/vDNA complex and remains in endocytic compartments, where it is degraded by lysosomes, while the L2/vDNA complex escapes from the endosome [76]. Further studies suggested, that the L2/vDNA complex escape by penetrating the endosomal membrane by the so-called membrane destabilization peptide on the C-terminus of L2 (aa 455-473) [34]. Another important region was lately described by

Bronnimann et. al. [28]: the so-called transmembrane domain (TMD) that is located at the N-terminus of L2 (aa 45-67) and includes several highly conserved GxxxG motifs. Mutated GxxxG motifs show storage of L2 in the endosome and leads to the conclusion that these motifs play a key role in endosomal escape [28]. Further studies showed a membrane interaction of L2 with the Transmission Electron Microscopy (TEM) *in vitro* [77]. Regardless, it is not completely understood how L1, L2 and vDNA are interacting in the endosome and how the L2/vDNA complex can overcome the vesicular membrane in detail.

1.5.4 Transport of L2/vDNA complex

After endosomal escape the viral trafficking continues with the transport of L2/vDNA to the Trans Golgi Network via various cellular factors including Sorting Nexin 17 (SNX17), Sorting Nexin 27 (SNX27), retromer complex components, dynein, and gamma-secretase [78]. The retromer complex serves as a cargo recognition core and consists of three subunits: Vps26, Vps29 and Vps35. All subunits are required for retromer-mediated endosomal sorting and HPV infection [79, 80]. Because in cells depleted of retromer, L2/vDNA complex fail to arrive the Trans Golgi Network TGN [81]. However, the mechanism of how retromer supports L2/vDNA trafficking remains unknown. Another important region on the C-terminus of L2 is the dynein binding region (aa 456-461) [78]. Dynein is known as a microtubule motor protein which may also be important in transporting the L2/vDNA complex to the nucleus [78]. Furthermore, the N-terminus of L2 at amino acid 41-44 interacts with the ER receptor syntaxin-18. This mechanism also could be responsible for the nuclear entry of the HPV genome [82]. It is suggested, that after endosomal escape, the L2/vDNA complex can interact and enter a syntax-18-postive vesicle to travel towards the nucleus via the ER/TGN [74, 82, 83]. After the onset of mitosis, the L2/vDNA complex buds out of the TGN/ER in a transport vesicle, which is then transported along the microtubule to the condensed chromosomes [78, 84].

1.5.5 Nuclear entry of L2/vDNA complex

During nuclear envelope breakdown the cell gets structurally reorganized and viral DNA can enter the cell nucleus [85]. In addition to nuclear entry during nuclear breakdown, a nuclear retention signal at amino acid 296-316 of L2 might play a role in viral infection. Here, it is suggested that during the metaphase of nuclear breakdown, this region is responsible for the association of the L2/vDNA complex with the nuclear matrix [86]. After successful cell division and a reformed nuclear envelope, the viral DNA is released from the transport vesicle in the

nucleus. So far, the model of viral entry of HPV virions contains many unanswered questions, but there is a clear indication that L2 as well as L1 plays a key role.

1.6 Interaction of HPV16 L1 and HPV16 L2

Currently there are several publications describing the interaction of the HPV 16 major capsid protein L1 with the minor capsid protein L2. Most of these results were provided by experiments in cell culture. Only a few publications investigated the interaction on biochemical level. Therefore, HPV11 L1 and GST tagged L2 were co-expressed and co-purified. Afterwards the complex was analyzed due to stoichiometry via size exclusion chromatography and FPLC. Additionally, the carboxy-terminal L1 binding domain was investigated with truncated HPV11 L2 variants by immunoblotting analysis. Afterwards, the ability of L1/L2 binding of eight different papillomavirus serotypes was compared. These results show a L1 binding domain at aa 412-455 located in the C-terminus of L2 for HPV16 [29]. Another publication investigated the L1/L2 interaction of HPV16 based on the compiled protein sequence via MUSCLE (by Log-Expectation, MUltiple Sequence Comparision). Therefore, the L1 protein structure (from RCSB PDB – 2R5H) and the predicted L2 protein structure (from 3D-Jigsaw and swissmodel.expasy.org and SAM-T09) is used. It is suggested that L2 bind the L1 capsomers but not the VLP, this indicates that L2 co-assembled with L1 rather than being inserted into the already formed capsid [87]. To determine the interaction of L1/L2 in the capsid Cryo-EM analysis was performed with VLP´s (with and without L2). However, the capsids and VLP´s were produced in 293T cells. The results suggest a maximum amount of 72
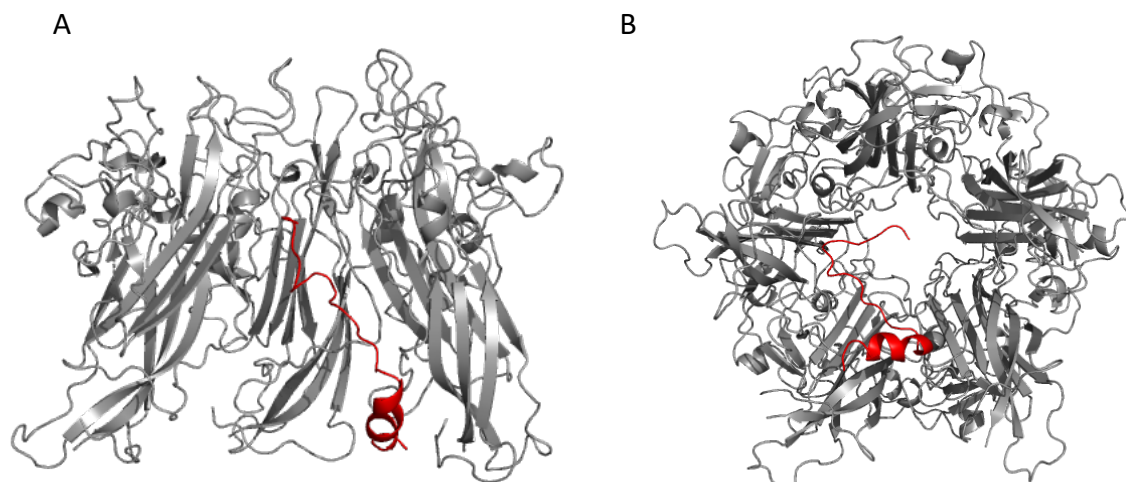
Figure 7: VP1-VP2 complex (modified from PDB 1CN3). VP1 pentamer is shown in gray and the VP2 peptide (aa 269-296) in red. For comparison the model of VP2 [3] has been transformed into the superposition of VP1 [5]. **(A)** View on the 5-fold axis. **(B)** View from the inner part of the "capsid".

L2 molecules per capsid (figure 2) [6]. Another indication for the mechanism of L1/L2 interaction comes from polyomavirus research. The polyomavirus is a dsDNA virus with the capsid proteins VP1, VP2 and VP3. In literature interesting parallels due to the location of the putative L1 binding site on L2 as well as hydrophobic nature of L1/L2 interaction and stoichiometry compared to the mouse polyomavirus VP1/VP2 interaction are shown [3, 29, 88]. Furthermore, the X-ray crystallography analysis demonstrates a close resemblance of VP1 with L1 and additionally the results from Buck et. al. suggest that L2 is eventually radically permuted relative to the L1 capsomer axis. This model for L2 organization in HPV is remarkably similar to the arrangement of VP1 and VP2 [3, 4, 6, 20, 89]. These results provide an indication that the L1/L2 interaction may resemble the VP1/VP2 interaction. For VP1 and VP2, an interaction is already described [3]. Here, the VP2 peptide interacts with three monomers of VP1 (figure 7). When comparing the peptide sequence used from VP2 to L2, it becomes apparent that this sequence is located near the putative L1 binding domain on the L2 genome (figure 5). In addition, the secondary structure prediction of L2 (MPI tool kit [90]) for aa 450-461 shows an alpha helical structure. This would also fit to the structure of the VP2 peptide as shown. These analogies propose a similar binding of L1 to L2.

Recent results show an interaction of L1 and L2 in near atomic resolution [91]. On this occasion a protein-like density which corresponds to L2 were found and a clear separation between the L1 chain and putative L2 was shown. Taken these *in vitro* and *in vivo* results together, an interaction between HPV16 L1 and L2 is verified. Nevertheless, this interaction is poorly understood and needs further biochemical investigations.

1.7 Objective

The aim of this thesis was to investigate the interaction of the purified HPV16 capsid proteins L1 and L2 *in vitro* due to a second L1 binding site located on L2, different pH and other interaction partners. Therefore, the proteins were produced, purified and afterwards characterized on a biochemical level. As mentioned in chapter 1.5, L1 and L2 play a major role in receptor-mediated endocytosis, endosomal escape, and transport of viral DNA into the nucleus. However, it is not yet known exactly what role the L1/L2 complex plays during this process. For this reason, a pronounced understanding of the viral entry is important to identify possible drug targets to treat an HPV infection and help the development of antiviral measures. Although a prophylactic vaccine is available, only 31 % in Germany are vaccinated [49]. It is already known that L1 and L2 interact with each other in the capsid and during viral entry. This interaction must change during viral entry from the cell to the nucleus [3, 6, 29, 87]. This means the interaction of L1 and L2 in the capsid probably differs from the interaction of L1 and L2 in the endosome. To characterize the change of the interaction we need to address the binding affinity (K_d) of L1 and L2 in different environments to compare the changes in interaction. Thereby the focus of this thesis refers to the change of interaction in the endosome (acidification and CypA and B) [62] and the interaction of L1 and L2 mimicking capsid conditions (pH 7.4). Additionally, we want to confirm the L1 binding domain at the C-terminus of L2 and we want to investigate an eventual second L1 binding domain at the N-terminus of L2 [29]. The first step of the project is to characterize the recombinant and purified capsid protein L1 and L2 regarding their purity, homogeneity, secondary structure and functionality to verify suitable material for interaction studies. To determine binding between HPV16L2 and potential binding partners we established a modified ELISA (EL(2)ISA) as a "screening-tool". Subsequently to quantify the binding affinity an anisotropy measurement and a surface plasmon resonance (SPR) measurement are performed. For the anisotropy the L2 is fluorescently labeled whereas for SPR measurement L2 is biotinylated via *Sortase A* [92] reaction. For all experiments, we need purified L1 and L2 protein as well as SUMO-Protease, TEV-Protease and *Sortase A*. The expression and purification of recombinant proteins needed to be established first to ensure a sufficient yield of proteins in reasonable purity and stability for biochemical characterization. The final goal, characterization of the L1 and L2 proteins and their interaction allows insights into the molecular basis of viral entry, thereby providing a better understanding.

2 MATERIALS AND METHODS

2.1 Materials

2.1.1 Equipment

Table 1: used equipment and manufacturer

Usage	Equipment	Manufacturer
Protein Expression	HT Infors Minitron	INFORS AG, Bottmingen
	Certomat®	B. BRAUN BIOTECH INTERNATIONAL, Melsungen
	Baffled flasks 5L	GLASGERÄTEBAU OCHS, Bovenden
	Bioreactor Labfors	INFORS AG, Bottmingen
	BioPhotometer	EPPENDORF AG, Hamburg
Protein Purification	Äkta® Pure	GE HEALTHCARE, Munich
	Äkta® Purifier	GE HEALTHCARE, Munich
	Superloop 10 mL	GE HEALTHCARE, Munich
	Superloop 50 mL	GE HEALTHCARE, Munich
	Superloop 150 mL	GE HEALTHCARE, Munich
	HisTrap® FF 1 mL	GE HEALTHCARE, Munich
	HisTrap® FF 5 mL/ HisTrap® crude FF 5 mL	GE HEALTHCARE, Munich
	HiTrap® SP HP 5 mL	GE HEALTHCARE, Munich
	HiTrap® Capto Q 5 mL	GE HEALTHCARE, Munich
	HiTrap® Heparin high performance 5 mL	GE HEALTHCARE, Munich
	HiLoad 16/600 Superdex 200 pg	GE HEALTHCARE, Munich
	HiLoad 16/600 Superdex 75 pg	GE HEALTHCARE, Munich
	Superdex 200 increase 10/300	GE HEALTHCARE, Munich
	Superose 6 10/300	GE HEALTHCARE, Munich
	Syringe Filter Millex HA MCE/PES	MERCK MILLIPORE, Darmstadt
	Membrane 0.45 mm/ 0.22 mm	MERCK MILLIPORE, Darmstadt
	Vivaspin 6, 3 kDa MWCo, PES	SARTORIUS, Goettingen
	Vivaspin 20, 10 kDa MWCo, PES	SARTORIUS, Goettingen
	Amicon® 15, 30 kDa MWco, Cellulose	MERCK MILLIPORE, Darmstadt
	Duropores® 0.45 mm PVDF	MERCK MILLIPORE, Darmstadt
	Syringes 1 mL /5 mL / 10 mL/ 20 mL /50 mL	BD PLASTIPAK, Heidelberg
	Thermo spectronic french press FA078	THERMO FISHER, Waltham (USA)
	Table centrifuge 5417R	Eppendorf AG, Hamburg
Centrifuges	Rotina 420R	HETTICH, Beverly (USA)
	Sorvall Lynx 6000	THERMO FISHER, Waltham (USA)
	Optima^{TM} LE-80K Ultracentrifuge	BECKMAN COULTER, Krefeld
	Electrophoresis Chamber Mini-PROTEAN II	BIO-RAD, Munich
	analytical ultracentrifuge XL-I, An-50 Ti rotor, double sector cells	Beckman Coulter, Palo Alto, CA, USA

Protein Analysis	SDS PAGE PowerPac200	BIO-RAD, Munich
	PROTEAN II TransBlot® Cell	BIO-RAD, Munich
	Licor Odyssey Fc	LICOR BIOTECHNOLOGY, Bad Homburg
	Spectrophotometer Specord 200	ANALYTIK JENA, Jena
	TRISTAR^2 S LB 942 Multimode reader	BERTHOLD TECHNOLOGIES, Bad Wildbad
	Monolith NT.115	NANOTEMPER TECHNOLOGIES GmbH, Munich
	Thermomixer F1.5	EPPENDORF AG, Hamburg
	FS-5, Edinburgh Instruments, temperature controller SC-25	Edinburgh, UK
	Biacore, T200	Cytiva, Münch
	Dynamic light scattering, Zetasizer Nano ZS	Malvern Panalytical, GB Malvern
	Transmission Electron Microscope, Tecnai Spirit	Thermo Fisher Scientific, Waltham (USA)
	JASCO J-815 Circular Dichroism Spectrometer	JASCO, Pfungstadt, Germany
Cuvettes	Fluorescence Ultra-Micro-Cuvette 100 µl, 10 mm	Hellma, Müllheim Germany
	Macro-Cuvette 101, Quartzglass Suprasil, 3 ml, 10 mm	Hellma, Müllheim Germany
	Demontierbare Küvette 106-QS SD d=0.01mm / d=0.2mm / d=0.5mm	Hellma, Müllheim Germany
	Macro-Cuvette 110-QS, 2mm, plug	Hellma, Müllheim Germany
	Küvettenhalter 013.000, Küvetten-Typ 106	Hellma, Müllheim Germany
Thermomixers	Magnetic Mixer MR3001K	HEIDOLPH INSTRUMENTS LABORTECHNIK, Schwabach
Others	Nanodrop 1000	THERMO FISHER, Waltham (USA)
	Nanophotometer	IMPLEN GmbH, Munich
	pH Detector pH526	WTW, Weilheim
	Scanner HP Scanjet 3800	HP, Paolo Alto, USA
	Sterile bank HeraSafe	HERAEUS HOLDING GMBH, Hanau
	Vortexer relaxTop	HEIDOLPH INSTRUMENTS LABORTECHNIK, Schwabach
	Autoclave HRM-242 II	HIRAYAMA, Tokyo
	Pierce^{TM} Streptavidin Coated High Capacity Plates, 96-Well	THERMO FISHER, Waltham (USA)

2.1.2 Chemicals

Table 2: used chemicals and manufacturer

Chemical	Manufacturer
2-Mercaptoethanol	SIGMA ALDRICH, Taufkirchen
Acetic Acid 96 %	CARL ROTH, Karlsruhe
Acetone p.A.	BIOFROXX GmbH, Einhausen
Albumin (BSA)	SERVA Electrophoresis GmbH, Heidelberg
Ammonium Peroxodisulfate (APS)	CARL ROTH, Karlsruhe
Ampicillin > 97 %	CARL ROTH, Karlsruhe
Amylose Resin	NEW ENGLAND BioLabs Inc., Ipswich (USA)
Anti Foam DF204	SIGMA ALDRICH, Steinheim
L-arginine monohydrochloride PharmaGrade	(SAFC), MERCK, Darmstadt
D/L-arginine hydrochloride > 98% (TLC) #MKBZ1464V	SIGMA ALDRICH, Taufkirchen
Benzonase® > 90 %	MERCK, Darmstadt
Biotin > 98.5 %	CARL ROTH, Karlsruhe
Brij® 58	SIGMA ALDRICH, Taufkirchen
Bromophenol Blue	SIGMA ALDRICH, Taufkirchen
$CaCl_2 \cdot 2\ H_2O$ > 98 %	CARL ROTH, Karlsruhe
$CaCl_2 \cdot 7\ H_2O$ > 98 %	CARL ROTH, Karlsruhe
$CoCl_2 \cdot 6\ H_2O$	CARL ROTH, Karlsruhe
Coomassie® Brilliant blue R250	MERCK, Darmstadt
$CuSO_4 \cdot 5\ H_2O$ > 99.5 %	CARL ROTH, Karlsruhe
Cyclohexylamino propanesulphonic acid CAPS > 99 %	SIGMA ALDRICH, Steinheim
D-Glucose > 99.5 % CELLPURE®, water-free	CARL ROTH, Karlsruhe
Ethanol absolute	APPLICHEM, Darmstadt
$FeCl_3 \cdot 6\ H_2O$ > 99 %	CARL ROTH, Karlsruhe
Glycerol, ROTIPURAN® > 99.5 %, water-free	CARL ROTH, Karlsruhe
Glycin PUFFERAN® > 99 %	MP Biomedicals, Illkirch (FR)
HEPES > 99.5 %	CARL ROTH, Karlsruhe
Imidazole PUFFERAN® > 99 %	CARL ROTH, Karlsruhe
Isopropanol	SAV-LP GmbH, Flintsburg
Isopropyl-β-D-thiogalactopyranosid (IPTG) Dioxan-free	THERMO FISHER SCIENTIFIC, Waltham (USA)
K_2HPO_4 > 99 %, H_2O free	CARL ROTH, Karlsruhe
Kanamycin sulphate	CARL ROTH, Karlsruhe
KH_2PO_4 > 99 %, H_2O free	CARL ROTH, Karlsruhe
Lysozyme, 200.000 U/mg, lyophilized	CARL ROTH, Karlsruhe
Maltose	CARL ROTH, Karlsruhe
Methanol	Honeywell, Seelze
$MgCl_2 \cdot 7\ H_2O$	MERCK, Darmstadt
NH_4Cl	SIGMA ALDRICH, Steinheim
PAGE – Ruler® Prestained Protein Ladder	THERMO FISHER SCIENTIFIC, Waltham (USA)
PAGE – Ruler® Unstained Protein Ladder	THERMO FISHER SCIENTIFIC, Waltham (USA)

Peptone ex. Casein, tryptic digestion (Tryptone)	CARL ROTH, Karlsruhe
Protease Inhibitor cocktail (EDTA-free)	ROCHE, Grenzach-Wyhlen
Rotiphorese® Gel 30 (37.5:1)	CARL ROTH, Karlsruhe
SDS ultra-pure > 99.5 %	CARL ROTH, Karlsruhe
Sodium Chloride (NaCl) > 99 %	CARL ROTH, Karlsruhe
Tris(2-carboxyethyl)phosphine hydrochloride (TCEP)	ALFA AESAR (THERMO FISHER SCIENTIFIC), Kandel
Tris(hydroxymethyl)aminomethan (TRIS)	MP Biomedicals, Illkirich (FR)
Tween®20	SIGMA-ALDERICH, Steinheim
Yeast extract	CARL ROTH, Karlsruhe
$ZnCl_2$, H_2O free	CARL ROTH, Karlsruhe
$ZnSO_4 \cdot 7 H_2O$	SIGMA ALDRICH, St. Louis (USA)

2.1.3 Peptides

Table 3: used peptides

Name	Sequence
FL-GGG	5(6)-Carboxyfluorescein-LPETGGRR
Biotin-GGG	Biotin-LPETGGRR

2.1.4 Media

All used media were sterilized via autoclaving at 121°C for 30min.

Table 4: used media and recipe

Medium	Components
Feed (for fermentation)	30 % (w/v) Glucose, 25 mL/L $MgSO_4$ (1 M), 20 mL/L trace metal stock, 20 mL/L Thiamin, 0.1 % (v/v) Antibiotics (Kan), 80 g/L NH_4Cl Adjust to pH 8, sterile filtered
LB-agar	10 g/L Peptone, 15 g/L Agarose, 5 g/L Yeast extract, 5 g/L NaCl
LB-Medium	10 g/L Peptone, 5 g/L Yeast extract, 5 g/L NaCl
M9-Medium (for fermentation)	6 g/L Na_2HPO_4 anhydride, 2.86 g/L KH_2PO_4, 0.5 g/L NaCl, 0.5 g/L NH_4Cl, 8 g/L Glucose, 1 mL/L Trace metal stock (1000 x), 1 mM $MgSO_4$, 1 mg/mL Biotin, 10 mg/mL Thiamin, 0.1 % (v/v) Antibiotics (Kan), 0.3 mM $CaCl_2$
SOC-Medium	20 g/L Peptone, 5 g/L Yeast extract, 0.5 g/L NaCl, 10 mM MgCl2, 10 mM MgSO4, 20 mM Glucose
TB-Medium	24 g/L Yeast extract, 12 g/L Peptone, 5 g/L NaCl, 0.4 % (v/v) Glycerol, 5 % (v/v) 20 x TB Buffer

2.1.5 Stock solutions

Table 5: used stock solutions with recipe

Stock	Components
10 x M9 Buffer (autoclaved)	60 g/L Na_2HPO_4 anhydride, 28.6 g/L KH_2PO_4, 5 g/L NaCl
10 x PBS	80 g/L NaCl, 11.5 g/L $Na_2HPO_4 \cdot 7\ H_2O$, 2 g/L KCl, 2 g/L KH_2PO_4, pH 7.2
20 x TB Buffer (autoclaved)	46.3 g/L KH_2PO_4, 250 g/L K_2HPO_4, 5 g/L $MgSO_4 \cdot 7\ H_2O$, pH 7.2
5 x SDS running buffer	15.1 g/L TRIS, 72 g/L Glycine, 5 g/L SDS
Ampicillin	100 mg/mL solved in deionized H_2O
APS	10 % (w/v) solved in deionized H_2O
Biotin	
$CaCl_2$	1 M in deionized H_2O
Chloramphenicol	20 mg/ml in 100 % EtOH
IPTG	1 M solved in deionized H_2O
Kanamycin	30 mg/mL solved in deionized H_2O
$MgCl_2$	1 M in deionized H_2O
$MgSO_4$	1 M in deionized H_2O
NaCl	5 M in deionized H_2O
SDS	20 % (w/v) dissolved in deionized H_2O
Thiamin	
Trace metal stock (1000 x) (autoclaved)	2 g/L $CaCl_2 \cdot 2\ H_2O$, 0.72 g/L $ZnSO_4 \cdot 7\ H_2O$, 0.4 g/L $MnSO_4 \cdot H_2O$, 40.2 g/L Na_2-EDTA $\cdot 2\ H_2O$, 33.4 g/L $FeCl_3 \cdot 6\ H_2O$, 0.64 g/L $CuSO_4 \cdot 5\ H_2O$, 0.72 g/L $CoCl_2 \cdot 6\ H_2O$
Tris-HCl pH 6.8 @ RT	1 M
Tris-HCl pH 7.4 @ 8°C	1 M
Tris-HCl pH 8 @ 8°C	1 M
Tris-HCl pH 8.8 @ RT	1 M
$ZnCl_2$	0.5 M in deionized H_2O

2.1.6 Buffers

All buffers were filtered and degassed for 30 min before usage. This was done to avoid larger aggregates, which can lead to unspecific light scattering during measurements and to fulfill the requirements to be used during protein purification with the ÄKTA® System.

Table 6: used buffers with recipe

Usage	Buffer	Components
SDS-PAGE	5 x SDS sample buffer	0.25 M Tris • HCl, 5 % (w/v) SDS, 5 % (v/v) 2-Mercaptoethanol, 50 % (w/v) Glycerol, 0.005 % (w/v) Bromphenol Blue, pH 8 at RT
	Coomassie Staining Solution	0.05 % (w/v) CoomassieR Brilliant blue R250, 10 % (v/v) Acetic acid, 25 % (v/v) Propan-2-ol
	Destaining solution	10 % Acetic acid, 10 % Ethanol, 80 % H_2O
Western Blot	Western Blot Buffer	2.44 g/L Tris • HCl, 11.26 g/L Glycin, 0.02 g/L SDS, 10 % (v/v) Methanol, pH 8.2
Test expression	Test expression cell lysis buffer	200 mM NaCl, 50 mM Tris • HCl pH 8 @ 8 °C, 5 % (w/v) Glycerol, 10 mM EDTA, 3.5 mg/ml lysozyme
Purification soluble	Lysis Buffer	50 mM Tris • HCl pH 8 at 8 °C, 200 mM NaCl, 5 % (w/v) Glycerol, 1 mM TCEP
	Ni-NTA A1	50 mM Tris • HCl pH 8 at 8 °C, 200 mM NaCl, 5 % (w/v) Glycerol, 1 mM TCEP
	Ni-NTA A2	50 mM Tris • HCl pH 8 at 8°C, 1 M NaCl, 5 % (w/v) Glycerol, 1 mM TCEP
	Ni-NTA B	50 mM Tris • HCl pH 8 at 8 °C, 200 mM NaCl, 5 % (w/v) Glycerol, 1 mM TCEP, 1 M Imidazol
	SEC	50 mM Tris • HCl pH 8 at 8 °C, 200 mM NaCl, 5 % (w/v) Glycerol
	CD Buffer L1nav	10 mM Tris • HCl pH 8 at 8 °C, 200 mM NaCl, 5 % (w/v) Glycerol
Purification insoluble	Isolation Buffer	10 mM Tris • HCl pH 7.4 at 8 °C, 5 mM EDTA
	solubilization buffer	6 M Guanidine Hydrochloride, 20 mM Tris • HCl pH 7.4 @ 8 °C
	Ni-NTA SB A	6 M Guanidine Hydrochloride, 20 mM Tris • HCl pH 7.4 @ 8 °C
	Ni-NTA SB B	6 M Guanidine Hydrochloride, 20 mM Tris • HCl pH 7.4 @ 8 °C, 1 M Imidazol
	Refolding Buffer	50 mM Tris • HCl pH 7.4 at 8 °C, 200 mM NaCl, 5 % (w/v) Glycerol, 0.5 M L-arginine
	Refolding Buffer CD	10 mM Tris • HCl pH 7.4 at 8 °C, 200 mM NaCl, 5 % Glycerol, 0.5 M DL-arginine
Interaction studies	CD Buffer pH 7.4	18 mM Na_2HPO_4, 1 mM $C_6H_8O_7$, 200 mM NaCl, 5 % (w/v) Glycerol, 1.6% (w/v) 58 Brij
	SPR Buffer pH 7.4	50 mM Tris • HCl pH 7.4 at 8 °C, 200 mM NaCl, 5 % (w/v) Glycerol, 0.16% (w/v) 58 Brij, ± 1 mM TCEP
	SPR Buffer pH 5.5	11 mM Na_2HPO_4, 4.4 mM $C_6H_8O_7$, 200 mM NaCl, 5 % (w/v) Glycerol, 0.16% (w/v) 58 Brij

2.1.7 Antibodies

Table 7: used antibodies, dilution and manufacturer

Antibody	Species	Used dilution	Manufacturer
α-HPV16L1 sc-57834	mouse	1:200	Santa Cruz Biotechnology, USA
α-HPV16L2 sc-65709	mouse	1:200	Santa Cruz Biotechnology, USA
α-His-Tag D3I10	mouse	1:1000	Cell Signaling Technology, USA
α-His-Tag sc-8026 Lot#D0618	mouse	1:500	Santa Cruz Biotechnology, USA
α-GroEL	mouse	1:200	Enzo Life Sciences, USA
α-CypA sc-134310	mouse	1:200	Santa Cruz Biotechnology, USA
α-CypB sc-130626	mouse	1:200	Santa Cruz Biotechnology, USA
α-mouse IgG IRDye® 680RD Goat	goat	1:10000	LI-COR Biotechnology - GmbH, USA
α-mouse IgG IRDye® 800RD Goat	goat	1:10000	LI-COR Biotechnology - GmbH, USA

2.1.8 *Escherichia coli* strains

Table 8: used *Escherichia Coli* strains

E. coli strain	Application
Top10	Plasmid amplification
BL21(DE3)	Protein expression
BL21(DE3)pGro7	Protein expression

2.1.9 Software

Table 9: used software

Software	Application
GraphPad Prism 9	Plotting Data
ImageJ 1.52a	Densiometric analysis
Microsoft Office 2016	Data analysis, Presentation, Protocol, writing
Endnote x9	Literature management
PyMOL 2.5.2	Protein structure analysis
SnapGene 5.2	Plasmid design and management

2.2 Methods

2.2.1 Cloning

All used constructs (figure 8) were either synthesized at GenScript, cloned via Gibson Cloning [93] or ligated via site directed mutagenesis [94, 95] (table S1). For Gibson Cloning and Ligation, the primers were designed via SnapGene 5.2 and ordered at Metabion. The amplification of the plasmid DNA of the selected vector and insert was carried out via a polymerase chain reaction (PCR) with the proofreading Q5® High-Fidelity DNA Polymerase. Afterwards, the amplified products were either assembled via Gibson assembly [93] or re-ligated with DNA ligase (section 2.2.3)[94, 95].

	plasmid	variant	protease
His$_6$ / SUMO / L1 (1, 6, 114, 619)	pETSUMOadapt	His$_6$-SUMO-L1	SUMO-Protease
His$_{10}$ / L1 (1, 10, 515)	pET19b	His$_{10}$-L1	enterokinase
L1 / His$_6$ (1, 505, 511)	pET22b(+)	L1-His$_6$	-
His$_6$ / SUMO / L1nav (1, 6, 114, 573)	pETSUMOadapt	His$_6$-SUMO-L1nav	SUMO-Protease
His$_{10}$ / L1nav (1, 10, 499)	pET19b	His$_{10}$-L1nav	enterokinase
GST / L1nav (1, 218, 653)	pGEX-4T1	GST-L1nav	thrombin
His$_6$ / TEV / GGGL2 (1, 6, 16, 489)	pET28a(+)	His$_6$-TEV-GGGL2	TEV-Protease
His$_6$ / TEV / GGGL2ΔL1bs (1, 6, 16, 446)	pET28a(+)	His$_6$-TEV-GGGL2ΔL1bs	TEV-Protease

Figure 8: Overview of the used L1 and L2 constructs. The first three constructs show the L1 constructs with different tags. The subsequent three constructs describe the L1nav constructs with different tags and the last two constructs present the different L2 constructs. Additionally, the used proteases for tag cleavage and the utilized plasmids are described.

2.2.2 Gibson cloning

For Gibson Cloning the Gibson Assembly® Master Mix (NEB #E2611) (New England Biolabs, USA) was used. The Gibson reaction was performed as described in the protocol of the kit. Subsequent transformation of Gibson assembled plasmid DNA in competent Top10 *E. coli* bacteria cells (NEB #C3019H) was performed as described in section (2.2.4).

2.2.3 Ligation

For deletion or insertion from an existing insert the Q5® Site-Directed Mutagenesis Kit (E0552S) (New England Biolabs, USA) was used. The reaction was performed as described in the protocol of the kit.

2.2.4 Transformation

The chemical competent *E. coli* expression strains BL21(DE3) or BL21(DE3)pGro7 were used for protein expression. The *E. coli* strain BL21pGro7 (Takara Biotechnology, China) contains additionally the genes from the GroEL and GroES chaperon system [96, 97]. These genes were expressed under the control of an *araB* promotor for a supplementary overexpression of the chaperon system GroEL/ES. To generate the competent BL21(DE3)pGro7 bacteria cells the plasmid was isolated from the BL21pGro7 strain (Takara Biotechnology) and transformed afterwards in BL21(DE3). Therefore, competent *E. coli* bacteria were thawed on ice and 1 µl of plasmid which harbor the encoding expression cassette of recombinant target protein was added. Afterwards, cells were incubated for 15 min on ice and heat-shocked at 42 °C for 45 s. Subsequently, the heat-shocked cells were incubated on ice for 15 min. Then, 1 ml of the SOC media was added and incubated for 1 h at 37 °C whilst gentle shaking at 300 rpm. The suspension was centrifuged at 5.000 g for 1 min and the supernatant was discarded. The pellet was subsequently resuspended in 100 µl SOC media and plated on an agar plate with respective antibiotics selecting for bacteria cells transformed with the desired plasmid. The plate was then incubated over night at 37 °C.

2.2.5 Sodium Dodecylsulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

A reducing Sodium Dodecylsulfate Polyacrylamide Gel Electrophoresis (reducing SDS-PAGE) was used for separation of proteins according to their molecular weight. This method also gives information about the purity of produced and purified proteins. The negative charged SDS binds to the protein and leads to a similar charge-to-mass ratio. Due to the negative charge of SDS, the proteins migrate through the gel to the positively charged anode. In addition, larger proteins move slower than smaller proteins because of their weaker electrophoretic mobility, thus create a separation. An SDS-PAGE gel consists of two layers, a stacking gel (top) with lower % of acrylamide and pH where the loaded proteins are concentrated and a separating gel (bottom) with higher % of acrylamide and pH where the proteins are separated according to size.

For this purpose, separating gel with 12% and 15% were used (table 10) to produce the gels, depending on the size of the proteins to be separated.

Table 10: used ingredients for SDS PAGE

Gel	Components	Volume (for 1 gel)
6 % (stacking gel)	1 M Tris • HCl pH 6.8	0.9 mL
	30 % Acrylamide	0.6 mL
	H_2O	2 mL
	20 % SDS	18 µL
	10 % APS	25 µL
	TEMED	2.5 µL
12 % (separating gel)	1 M Tris • HCl pH 8.8	2 mL
	30 % Acrylamide	2.9 mL
	H_2O	3.1 mL
	20 % SDS	40.5 µL
	10 % APS	50 µL
	TEMED	5 µL
15 % (separating gel)	1 M Tris • HCl pH 8.8	2 mL
	30 % Acrylamide	3.6 mL
	H_2O	2.3 mL
	20 % SDS	40.5 µL
	10 % APS	50 µL
	TEMED	5 µL

Samples were reduced by adding 1:5 (v/v) reducing 5 x SDS sample buffer and then incubated at 95 °C for at least 5 min. This resulted in denaturation of the proteins, which were then separated on a reducing SDS-PAGE using 200 V for 40 – 60 min at RT. After separation, protein bands were visualized by Coomassie staining. For this, the polyacrylamide gel was heated in the Coomassie staining solution for 30 sec and incubated on a shaker at RT for 15 min. After this incubation, the Coomassie staining solution was removed and the gel was incubated in decolorizing solution (30 % (v/v) EtOH, 10 % (v/v) acetic acid) until the protein bands are clearly visible and the background was decolorized.

2.2.6 Western Blot

To identify and verify the different purified proteins a Western Blot was used. Therefore, an SDS-PAGE (section 2.2.5) was transferred onto a nitrocellulose membrane via a TransBlot® chamber filled with 2.5 L Western Blot buffer, applying constant voltage of 90 V for 90 min. Any free protein binding sites were then blocked by incubating the membrane in 5 % (w/v) Albumin-PBS solution for 1 h at room temperature. The nitrocellulose membrane was then washed with PBS-T (0.05 % (v/v) Tween) and afterwards incubated o/n at 4 °C with the primary antibody against the protein of interest. The next day the membrane was washed with PBS-T (0.05 % (v/v) Tween) and incubated with a fluorescently labelled secondary antibody

(1:10.000) (table 7) for 1 h at RT. After repeating the washing steps three times the protein was detected by the fluorescence of the secondary antibody with the imaging system Li-Core Odyssey Fc.

2.2.7 Quantification of soluble protein (Western Blot)

To quantify the solubility of the expressed proteins, a densiometric quantification of proteins in Western Blots was performed. Therefore, the software from Li-Core Odyssey Fc was used. The protein expressed under different conditions (expression host, tag and temperature) and the solubility of the different L1 constructs were tested by loading the same amount of biomass on a reducing SDS-PAGE. After testexpression and cell lysis the lysates were centrifuged at 30.000 g, 8 °C for 30 min to separate the soluble proteins (supernatant, SN) from the insoluble proteins (pellet, P). The supernatant and pellet were loaded equally on a reducing SDS-PAGE and the signal of supernatant and pellet was summed up and used as 100 % produced protein. The signal from pellet and supernatant was calculated and compared to the total protein amount.

$$protein\ amount\ (soluble\ or\ insoluble)\% = \frac{specific\ antibody\ signal\ (soluble\ or\ insoluble\ fraction)}{specific\ antibody\ signal\ (soluble\ +\ insoluble\ fraction)} * 100\%$$

Additionally, GroEL was used as an indicator to quantify the efficiency of cell lysis because GroEL/ES is a highly soluble protein [98] and in case of 100 % cell lysis GroEL should be located only in the soluble fraction (supernatant) only after centrifugation of the sample. GroEL signal in pellet fraction represents an incomplete cell lysis. Exemplary calculation is shown in the supplementary part figure S3.

2.2.8 Testexpression *E. coli*

For a test expression the different constructs were expressed in a small-scale expression at different temperatures (16 °C, 25 °C or 37 °C) and with or without the chaperon GroEL/ES. Therefore, a bacteria tube with 5 ml TB medium, 0.05 % (w/v) Glucose, the respective antibiotics (different for every construct) and recombinant *E. coli* strains [99]. The o/n culture was incubated at 37 °C and 130 rpm. 20 ml (in a 100 ml flask) of fresh TB medium, 0.05 % (w/v) Glucose and antibiotics required were inoculated to an $OD_{600} = 0.1$ and incubated at 37 °C until OD_{600} reached 0.5. The temperature was set to 16 °C, 25 °C or 37 °C at an OD_{600} of 0.7. Expression for the different constructs was induced with 1 mM IPTG and for the GroEL/ES chaperon, 2 mg/mL L-arabinose was used. To monitor the expression of recombinant protein

and cell growth, 500 µl sample was taken every hour. These samples were centrifuged at 20.000 g for 5 min and the supernatant was completely discarded. The pellet was resuspended in 100 µl 1 x SDS Buffer and incubated for 5 min at 95 °C and loaded according to the OD_{600} on a reducing SDS-PAGE. To analyze the solubility of the expressed protein, 14 ml culture were centrifuged at 12.000 g for 30 min after 4 h and the supernatant was discarded. The pellet was resuspended in 2 ml cell lysis buffer and lysed via ultra-sonification (3 cycles, each 30 sec ultrasound, 25 % power, M_s 72). The solution was centrifuged at 20.000 g for 30 min. 20 µl 5x SDS buffer were added to 80 µl supernatant while the pellet was resuspended in 100 µl 1 x SDS buffer and analyzed on an SDS-PAGE and Western Blot. The samples collected at different time point samples were applied to an SDS-PAGE according to the OD_{600} ratio between the different time points.

$$\left(\frac{OD_{600} \; timepoint \; 0h}{OD_{600} \; timepoint \; 1h} \, x20\mu l = volume \; timepoint \; 1h\right)$$

To calculate the solubility of the expressed protein, the supernatant and the pellet after cell lysis were analyzed with a reducing SDS-PAGE (section 2.2.5) and Western Blot (section 2.2.6).

2.2.9 Protein expression

2.2.9.1 Flasks

The recombinant *E. coli* strains were cultured in a 1 L baffled flask containing 200 ml TB medium, 0.05 % (w/v) glucose and supplemented with required antibiotics, o/n at 37 °C and 130 rpm (pre-culture). 1 L TB medium with 0.05 % (w/v) glucose and respective antibiotics in 5 L baffled flask were inoculated with the pre-culture to an OD_{600} = 0.1. The culture was incubated at 37 °C and 100 rpm until the OD_{600} reached 0.5. The temperature was set to 16 °C, 25 °C or 37 °C, depending on optimal condition defined through test expression, until OD_{600} reached 0.7. 1 mM IPTG was added to induce expression of recombinant protein. In case were GroEL/ES was co-expressed, 2 mg/ml L-arabinose was also added. The expression and bacterial cell growth were monitored by measuring the OD_{600} every hour and 500 µl SDS samples were collected each time.

After 4 h the culture was harvested at 6.000 g (Thermo Scientific Sorvall RC6 plus centrifuge) for 20 min and the biomass was stored at -20 °C. The SDS PAGE samples were centrifuged at 20.000 g for 5 min, resuspended in 100 µl 1x SDS sample Buffer and incubated for 5 min at 95 °C.

2.2.9.2 Process controlled Fermentation

To produce the L1nav protein the construct His$_6$-SUMO-L1nav the BL21(DE3)pGro7 *E. coli* expression strain was used. Therefore, a fed-batch fermentation in the bioreactor Labfors 5 was performed. The 200 ml pre-culture in fermentation medium (M9-based) [100] in 1 L baffled flask were inoculated with a single colony of BL21(DE3)pGro7/His$_6$-SUMO-L1nav and incubated at 37 °C, 130 rpm o/n. Afterwards, 1 L fermentation medium (M9-based) in 5 L baffled flask were inoculate with an OD$_{600}$ of 0.2 of the pre-culture. 4 L of the culture were incubated at 37 °C and 100 rpm until the culture reached an OD$_{600}$ of 1. The cells were harvested by centrifugation at 4.000 g, 4 °C for 30 min and the pellet was gently resuspended in 40 ml sterile phosphate-buffered saline (PBS).

The fed-batch fermentation was started with 3 L defined medium (M9-based) (table 4) containing 0.8 % (w/v) glucose. At the beginning, the temperature was set to 37 °C ± 0.5 °C for the batch phase and afterwards decreased for the feed-phase to 20 °C ± 0.5 °C. The pH was regulated to 6.8 ± 0.2 with 10 % (v/v) NH$_3$ and 10 % (v/v) H$_3$PO$_4$ and the oxygen was controlled at 30 % ± 10 % saturation by a cascade of first airflow and pure oxygen flow. The stirrer was set to 1.000 rpm and foam formation was avoided by addition of 1 ml antifoam (DF204).

The 3 L defined medium (M9-based) were inoculated with ~1 g dry biomass of the pre-culture. After ~2.5 h growth period at 37 °C the glucose was consumed, and the bioreactor was cooled down to 20 °C. The exponential phase was started with a μ_{set} of 0.1 h^{-1}, Y_{xs}= 0.5 and m_s= 0.04 to grow the bacterial under substrate limitation. Therefore, the maximal growth rate (μ_{max} = 0.7 ± 0.1) was determined before, under non-limited condition in 5 L baffled flask with 1 L LB medium at 20 °C, 100 rpm for 4 h. During the fed-batch process the OD$_{600}$ and glucose concentration were measured to analyze the biomass increase and the substrate consumption. After 17 h the culture was harvest by centrifugation at 6.000 g, 4 °C for 20 min and were stored at -20 °C. To calculate the dry biomass, the conversation factor of 0.3 g/L ($\frac{cdw\left(\frac{g}{l}\right)}{OD600} = \frac{0.3g}{L}$) was used [1, 101].

2.2.10 Cell lysis

2.2.10.1 Cell lysis of soluble protein - HPV16 L1 constructs

The cell pellet (30 g) was resuspended in 40 ml lysis buffer. Additionally, 3.5 mg/ml lysozyme, a ground up protease inhibitor pill (EDTA free), 3 mM MgCl$_2$ and 7 µl Benzonase (500.000 U, activity >250 U/µl) (to decrease the amount of nucleic acid in the solution) were freshly added and incubated for 30 min at 4 °C whilst gentle stirring. The bacteria lysis was performed with

4 cycles in the French pressure cell with a pressure of 600-1100 bar. In between the cycles the protein suspension was incubated on ice for 5 min to avoid aggregation. Afterwards the pH was checked and adjust with 2 M Tris • HCL to pH 8 at 8 °C and the suspension was incubated 1 h at 4 °C for further nucleic acid digestion by benzonase and centrifuged at 30.000 g for 1 h at 4 °C (SS-34 rotor, Thermo Scientific). The supernatant was filtered through a 0.45 µm syringe filter and purified via affinity chromatography. To analyze the cell lysis, 80 µl of the supernatant and pellet were mixed with 20 µl 5 x SDS reducing buffer and loaded on a reducing SDS-PAGE (section 2.2.5).

2.2.10.2 Cell lysis for insoluble protein (inclusion bodies) – HPV16 L2 constructs

The cell pellet (20 g) was resuspended in 40 ml Isolation buffer. After resuspension, 3.5 mg/ml lysozyme was added to degrade the cell wall. In addition, 7 µl Benzonase (500.000 U, activity >250 U/µl), 8 mM $MgCl_2$ as well as half of a ground protease inhibitor pill (EDTA free) were added, and the pH of the suspension was set to 8 due to the benzonase activity optimum at pH 8. Afterwards, the suspension was incubated on a magnetic stirrer for 30 min at 8 °C whilst gentle stirring. The bacteria lyses was performed in a French pressure cell for 4 cycles between 600-1100 bar. Subsequent, the other half of the protease inhibitor pill (EDTA free) was added, and the solution was incubated for 45 min at 8 °C while mixing with 100 rpm on a magnetic stirrer. The bacteria suspension was then centrifuged at 30.000 g, 8 °C for 30 min. For purification, the pellet was resuspended in 200 ml Isolation Buffer with the Ultrathurrax and the suspension was then centrifuged at 30.000 g for 30 min at 8 °C. DNA contamination on the pellet was visible as black or dark brown, slimy deposit and was scratched away carefully with a small spatula. This washing process was repeated 4 times. The 5th time the final white protein pellet was resuspended in 40 ml solubilization buffer for 2 h at RT, mixed at 250 rpm on a magnetic stirrer. Afterwards, the solution was centrifuged at 30.000 g for 30 min and the supernatant was filtered through a 0.45 µm syringe filter. To analyze the isolation and purification process, 80µl of the supernatant (SN) as well as the resuspended pellet (P) after each centrifugation step were collected and mixed with 20 µl 5 x SDS reducing buffer. The samples were resolved on a reducing SDS-PAGE (section 2.2.5)

2.2.11 Protein purification

2.2.11.1 Ni-NTA soluble protein purification – L1 constructs

Purification was performed on an ÄKTA® pure 25 FPLC. Therefore, Ni-NTA Buffer A1, Ni-NTA Buffer A2 for DNA removal and Ni-NTA Buffer B as the elution buffer was used. The system was washed with the pump wash for Ni-NTA buffer A1, A2, B. Afterwards a 3x 5 ml HisTrap®FFcrude column was connected to the system under flow (0.5 ml/min) and was equilibrated with 5 column volume (CV) Ni-NTA buffer A. The 500 ml sample was loaded on HisTrap®FFcrude column with a flow rate of 5 ml/min via sample pump. Thereupon, the column was washed with A1 for 5 CV and with A2 for 5 (CV). Subsequently, a pre-elution with a 3 % step Buffer B was performed and afterwards a gradient to 60 % over 15 CV was executed to elute target protein. The elution was collected in 2 ml fractions. To remove any tightly bound substances, the column was washed with 5 CV of 100% Buffer B. Accordingly, the column was then re-equilibrated with 5 CV of Buffer A. The GST-L1nav construct was purified as already described [102-105].For the analysis of protein purification 80 µl samples were taken from chosen fractions for a reducing SDS-PAGE (section 2.2.5) and mixed with 20 µl 5 x SDS Buffer. Furthermore, the protein concentration was determined spectroscopically.

2.2.11.2 Ni-NTA insoluble protein (inclusion bodies) purification – L2 constructs

Purification was performed on an ÄKTA® purifier FPLC and Ni-NTA SB A Buffer and Ni-NTA SB B Buffer were used. First, the system was washed with buffers SB A and SB B by a pump wash. Afterwards, a 3 x 5 ml HisTrap®FFcrude column was connected to the system via drop-to-drop procedure with 0.5 ml/min flow and was equilibrated with buffer SB A for 5 CV at 5 ml/min flowrate. Subsequently, 40 ml sample was loaded on the column via 50 ml superloop with 4 ml/min and the column was than washed with 5 CV buffer SB A to remove any unbound protein. Elution of the target protein was then performed by a gradient of 0 %-40 % for 5 CV buffer SB B and the protein was collected in 2 ml fractions. Any other tightly bound protein was than eluted with a 100 % buffer SB B step afterwards. Accordingly, the column was then re-equilibrated with 5 CV of Buffer SB A. For the analysis of protein purification 80 µl samples were taken from chosen fractions for a reducing SDS-PAGE (section 2.2.5) and mixed with 20 µl 5 x SDS Buffer. The elution peak with the target protein was collected and dialyzed o/n in a 12-14 kDa dialyzing tube against 5 L H2O to precipitate the protein. Precipitated protein was than centrifuged for 1 h at 100.000 g at 8 °C and the protein pellet was afterwards

resuspended in 2 ml IB A Buffer and the concentration was determined spectroscopically. The solubilized L2 proteins were stored at 4 °C.

2.2.12 Refolding stability for L2 constructs

The L2 protein was purified under denaturing condition and needs to be refolded to be functional. To investigate the protein refolding stability under different buffer conditions, the light scattering of the sample was detected at 400 nm over time. Therefore, ~ 3 ml of chosen buffer was provided in a 1 x1 cm glass cuvette whilst stirring at 1200 rpm. The kinetic was measured at an excitation wavelength of 400 nm with a bandwidth (BW) of 0.5 and an emission of 400 nm with a BW of 1. Reference buffer was measured for 300 s and 1 µM pre-reduced (10 mM DTT, 30 min, 8 °C) protein was added and measured until 1800 s. The refolded sample was directly used to determine the temperature stability of the protein (section 2.2.14).

2.2.13 Preparative refolding of solubilized L2

For refolding, 10 mg/ml solubilized L2 was used and incubated for 30 min with 10 mM DTT at 8 °C to avoid unspecific disulphide bond formation. Before the refolding process could start, refolding buffer was incubated on ice in a 2 ml tube with a small stirrer. Pre-cooled pipettes, pipettes tips, tubes and buffers were used to avoid aggregation and the refolding was performed in the cold-room. To avoid aggregation a protein concentration of max. 10 mg/ml was used for refolding. After reduction with 10 mM DTT the pre-cooled 2 ml Eppendorf tube with a small stirrer and refolding buffer was put in a 30 ° angle on a magnetic mixer in the cold room. The magnetic mixer was set on the fastest level to prevent aggregation of the protein during refolding process. Through the constant mixing of the refolding buffer the solubilized protein sample was dropped into the buffer with a 5 s break in between the drops.

Depending on the application after the refolding, the sample was either used directly or dialyzed with a Dialysis Kit (Pur-A-Lyzer, Midi 3500, Sigma) in refolding buffer over night at 8 °C to remove any SB buffer leftovers. The sample was then centrifuged at 100.000 g for 1 h at 8 °C to get rid of protein aggregates and the supernatant was transferred into a fresh Eppendorf tube.

2.2.14 T_{trans} - temperature stability measurement for L1 and L2

The temperature-induced denaturation and aggregation was investigated by the tryptophan fluorescence (280 nm) for the L1 construct and for the L2 construct by the fluorescence of tyrosine (275 nm) due to the absence of tryptophanes. Therefore, a 1 x 1 cm glass cuvette containing 3 ml refolded protein with a concentration of 1 µM while stirring at 1200 rpm was used. For L2 constructs the sample was exited at 275 nm and the emission was measured from 275 nm-500 nm with a dwell time of 0.2 s while heated from 8 °C-60 °C with a heating rate of 0.2 °C/min, 1 °C stepwise with ± 0.3 K accuracy, with a 1 min equilibration at every step and a bandwidth of 1/5. The refolding buffer was also measured without protein to observe non-significant changed in the screened temperature range and to exclude changes in light scattering. The L1 constructs were excited at 280 nm and the emission was measured from 280 nm-360 nm.

2.2.15 Cleavage of fusion tag for L1 constructs

2.2.15.1 SUMO-Protease

To optimize the conditions for His_6-SUMO-tag cleavage, tests were performed with different concentration of SUMO-Protease, different incubation time and different temperatures (not all shown). Therefore, 1 mM TCEP (avoiding unwanted disulphide bonds) and His_6-SUMO-Protease [106] were added to the collected fractions, the samples were incubated and analysed via Western Blot (section 2.2.6). The final conditions were set to: (1) mass ratio 1:10 (protease:protein), (2) dialyzed for 48 h (tube 12-14 kDa) against 2 L A0 buffer without any reducing agent at 4 °C while gentle stirring. The dialyzed sample was then load on a HisTrap®FFcrude column 5 ml coupled online with a preparative HiLoad 16/600 Superdex 200 pg column to remove the His_6-SUMO-tag and His_6-SUMO-Protease followed by the separation of the L1nav/L1 Monomer and Pentamer. The concentration was measured spectroscopically, and the protein purity was checked via a 10 µg Coomassie stained reducing SDS-PAGE (section 2.2.5).

2.2.15.2 Thrombin

The fused GST-L1nav construct contains a Thrombin cleavage site. Therefore, the construct was incubated with different concentrations of Thrombin for different time periods and at different temperatures (done by Bastian Breiner). The final conditions, 1:8 mass ratio (protease:protein) (2.8 NIH Unit ~ 1 µg thrombin from human plasma (605195); Merck Millipore, Germany) with additional 10 mM $MgCl_2$ were used for GST-cleavage. The samples

were incubated on at RT or 8 °C. The cleavage efficiency from the different conditions was checked via reducing SDS-PAGE (section 2.2.5) and Western Blot analysis (section 2.2.6).

2.2.15.3 Enterokinase

The different His_6-L1 constructs contained a Enterokinase cleavage site and were cleaved thereby via enterokinase in a mass ratio 1:20 (protease:protein) (enterokinase from calf intestine, Roche) o/n at 30 °C and 8 °C. The cleavage efficiency from the different conditions and constructs was checked via reducing SDS-PAGE (section 2.2.5) and Western Blot analysis (section 2.2.6).

2.2.16 Analytical ultracentrifugation (AUC) for L1nav construct

The oligomeric state and homogeneity of L1nav protein was analysed via analytical ultracentrifugation (AUC). Therefore, 0.3-0.5 mg/ml L1nav protein under non-reducing, reducing or 1.6 % (w/v) Brij buffer conditions were analysed in a XL-I (Beckman Coulter) ultracentrifuge with an An-50 Ti rotor and double sector cells. For investigation of the sedimentation velocity the sedimentation of the sample was measured at 280 nm for 6 h and 8 °C at 40.000 rpm and scans were taken every 10 min. The L1nav construct was measured at 20 °C. To measure the molecular mass the samples were centrifuged at 8.000 rpm for 105 h and 8 °C. The data analysis for sedimentation velocity as well as the sedimentation equilibrium was performed with the program SEDFIT [107].

2.2.17 Transmission electron microscopy (TEM) for L1 constructs (L1nav/L1)

Transmission electron microscopy was used to analyse the pentameric structure of the L1 non-assembly variant and the His-tagged L1 VLPs. Therefore, a carbon-coated cupper grid was discharged and 100 µl/ml sample was loaded. Afterwards the sample was washed with water and stained with 1 % uranyl acetate for 5 min. The samples were than analysed with a Tecnai Spirit at 120 kV and images were taken with a F416 CMOS camera (TVIPS).

2.2.18 Cleavage of His$_6$-TEV-tag via TEV Protease

For the L2 constructs which contain a TEV-cleavage site, an in-house purified His$_6$-tagged TEV-Protease [108] (plasmid provided by Prof. Stehle) was used. Here, the TEV cleavage site was removed, and the subsequent triple-glycine-motif was exposed for *Sortase A* ligation (section 2.2.20). Therefore, 30 µM solubilized L2 was incubated with 10 mM DTT for 30 min at 8 °C to avoid any unspecific cysteine interaction during refolding L2. Thereafter, the reduced solubilized L2 was refolded via rapid dilution in refolding buffer with 0.5 M L-arginine (section 2.2.13). Previous experiments revealed a mass ratio
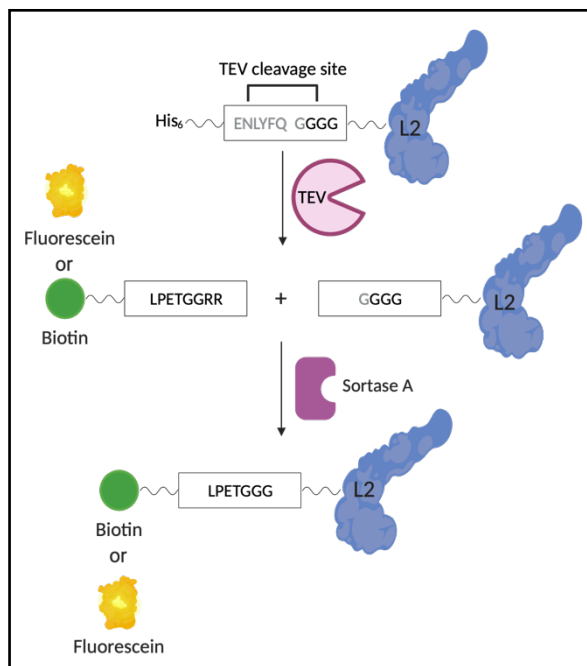
Figure 9: Scheme of TEV cleavage followed by *Sortase A* Ligation of L2 with fluorescein-LPETGGRR or biotin-LPETGGRR-peptide (modified from Desiree Frecot and created with BioRender).

1:5 (protease:protein) as a sufficient ratio for complete cleavage as well as an incubation time o/n at 8 °C while gentle stirring on a magnetic stirrer. The suspension was subsequently either loaded on a HisTrap®FFcrude 5 ml column to remove the His$_6$-TEV-Protease and the His$_6$-TEV-tag or directly labeled via *Sortase A* as described in section 2.2.20.

2.2.19 Circular Dichroism (CD)

The circular dichroism (CD) measurement was used to analyse the secondary structure of the refolded and cleaved L2 protein and the L1nav protein. Therefore, a far-UV CD spectroscopy in refolding CD buffer with low concentration of Tris • HCL and 1.6 % (w/v) Brij 58 was performed. The L2 samples were freshly prepared in refolding buffer and dialyzed against D/L-arginine over night at 8 °C with a concentration of 20 µM. The L1nav constructs were measured in A0 buffer with and without 0.16 % (w/v) Brij 58. Subsequently, the samples were centrifuged at 100.000 g 8 °C for 1 h to get rid of any aggregates. The samples were measured at 8 °C, wavelength 190-260 nm with a scan speed of 50 nm/min, 0.5 s response time, data pitch 0.5 nm and 64 accumulation per measurement. A cuvette with d = 0.01 mm for L2 and d = 10 mm for L1nav constructs were used and measured in a CD-spectrophotometer JASCO-815 equipped with a temperature controller.

2.2.20 *Sortase A* labeling

For the complex analysis of capsid protein L1 and L2, L2 had to be fluorescently labeled via *Sortase A*. *Sortase A* is used as a powerful tool to produce semi-synthetic proteins. Such semi-synthetic proteins are useful for various analytical methods such as Anisotropy. *Sortase A* is a transpeptidase of *Staphylococcus aureus* and responsible for anchoring district proteins to the bacterial cell wall. *Sortase A* recognizes a specific LPxTG motif [109] in order to be able to select the proteins, which are then anchored by *Sortase A* to the cell wall [110]. Afterwards, the threonine residue in the sequence it is cleaved by the formation of an intermediate thioester and then coupled to an N-terminal glycine residue which is anchored to the cell wall [111]. The TEV cleaved L2 constructs (section 2.2.18), including the cleaved His_6-TEV-tag and the His_6-TEV-Protease were incubated with 0.5-fold molar concentration of *Sortase A* and 1.5-fold molar concentration of peptide (Fluorescein or Biotin) compared to the L2 construct. Additional, 5 mM $CaCl_2$, 10 mM $MgCl_2$ and 1 mM DTT were added to the refolding buffer as these represent co-factors of *Sortase A* reaction. The first glycine residue of the peptide is cleaved by the *Sortase A* and ligated to the protein at the N-terminal glycine residues. Previous experiments showed an optimum in labeling after 2 h at 8 °C. To separate His_6-TEV-tag, His_6-TEV-Protease, His_6-SortaseA and any unligated peptide from ligated L2 construct, a 5 ml HisTrap®FF crude column was coupled online to a preparative HiLoad 16/60 Superdex S200 pg. The success of ligation was observed for Fluorescein-LPETGGRR peptide at an absorption of 499 nm and afterwards the samples were load on a reducing SDS-PAGE (section 2.2.5). Before staining the SDS-PAGE we analyzed the PAGE under UV-illumination. The degree of labeling (DoL) was quantified using the absorption measurement in the UV-VIS spectrum to determine the concentration based to the Lambert-Beer-law. Therefore, the analysis was performed by using the absorption values at 280 nm (total protein concentration) and 495 nm (fluorescein). The concentration was calculated by using the corresponding extinction coefficient for protein and fluorescein according to the Lambert-Beer-law:

$$DoL\ \% = \frac{concentration\ (peptide)}{concentration\ (total\ protein)} * 100$$

For the Biotin-LPETGGRR peptide a Streptavidin-AP conjugate assay (section 2.2.21) was used to check the general biotinylation of L2 constructs. The degree of biotinylation (DoB) was determined via a streptavidin-holdup assay and a dot blot for quantitative control. Therefore, 400 µl of Streptavidin-resin were added to a 0.22 µm filter spin column and centrifuged for

5 min at 8 °C and 5.000 g. The resin was washed for 3 times by adding 500 μl refolding buffer and resuspending the resin after centrifuging for 5 min, 8 °C at 5.000 g. Afterwards, filter column with streptavidin-resin was transferred to a new tube and 150 μl of 10 μM Biotin-L2 construct was loaded on the streptavidin-resin and incubated for 30 min at 8 °C under gentle shaking. The column was centrifuged for 5 min, 8 °C at 5.000 g and the flow-through was collected. The load and flow-through were measured at the Specord® and the degree of biotinylation was calculated:

$$DoB\ \% = 100\% - (\frac{FT_{280nm}}{Load_{280nm}} * 100)$$

### 2.2.21 Streptavidin-AP conjugate assay

A Streptavidin from *E. coli* K12 was coupled to an alkaline phosphatase (AP) [112]. The coupled Streptavidin-AP was than bound to the biotinylated L2 construct with high affinity (K_d=10^{-15} M). Alkaline phosphatase can cleave the substrate BCIP-T (5-bromo-4-chloro-3-indolylphosphate, p-toluidine salt) which is completed with an enhancer chromogen nitro blue tetrazolium (NBT). This reaction results in a formation of an insoluble blue precipitate and is visible as a well-defined spot or band on the reaction site of the membrane. Therefore, the samples were loaded on a reducing SDS-PAGE (section 2.2.5) and transferred via Western Blot (section 2.2.6) on a nitrocellulose membrane. The membrane was washed 3 times with PBS-T (0.05 % (w/v) Tween), each time 10 min and incubated in 5 ml PBS-T with 1 % (w/v) BSA and 1:2000 Streptavidin-AP conjugate for 2 h at 8 °C. Accordingly, the membrane was washed for 2 times in PBS-T for 10 min followed by washing with staining solution (10 mM Tris • HCl pH 9.5, 100 mM NaCl, 50 mM MgCl$_2$). Subsequently, the membrane was developed with 10 ml staining solution containing 200 μl NBT/BCIP and was incubated until bands were visible. The membrane was than washed with water and photographed.

2.2.22 Interaction studies of HPV16 L1 and L2

2.2.22.1 EL(2)ISA

The EL(2)ISA consists of a pull-down binding assay coupled with a ELISA. Therefore, a Streptavidin-coated 96 well plate and different constructs of biotinylated L2 were used for this assay. 100 µl of 1 µM refolded biotin-L2 were loaded on the plate and incubated for 2 h at 4 °C. Afterwards the plate was washed 3x times for 10 min, each time with 1.6 % (w/v) Brij 58 buffer. Then, 250 nM L1nav pentamer, 250 nM L1nav monomer, 250 nM CypA or 250 nM CypB were loaded to investigate direct binding to biotin-L2. After 2 h incubation at 4 °C the plate was washed 3 times with 1.6 % (w/v) Brij 58 buffer for 10 min and required antibodies for L1, CypA and CypB were added and incubated o/n at 4 °C.
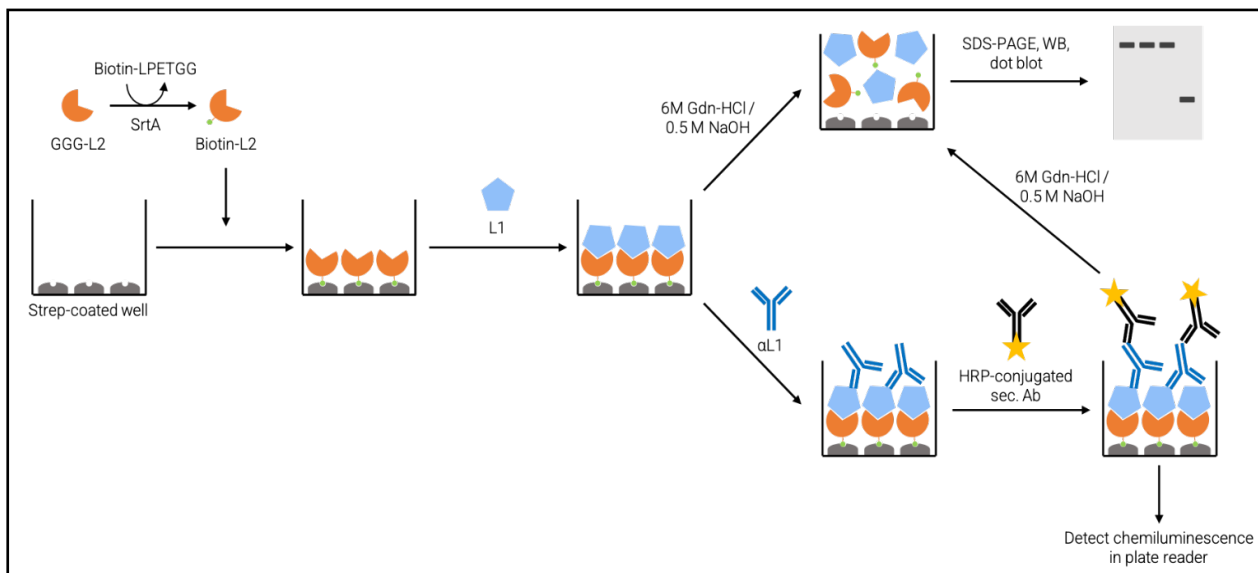
Figure 10: Scheme of EL(2)ISA. Therefore, a Streptavidin coated plate was immobilized with biotinylated L2 and afterwards incubated with potential binding partners. (Created by Samuel Maiwald)

The plate was washed afterwards 3 times for 10 min with 1.6 % (w/v) Brij 58 buffer and the samples were incubated with primary α-HPV16L1 at 4 °C. Afterwards, the plate was washed 3 times with 1.6 % (w/v) Brij 58 buffer and secondary α-mouse horseradish peroxidase was incubated for 1 h at 4 °C. Accordingly, the plate was washed again 3 times for 10 min each with 1.6 % (w/v) Brij 58 buffer and ECL substrate was added and measured at a Berthold Multimode Reader.

2.2.22.2 Anisotropy

Anisotropy was used to determine the binding affinity of interaction partners. In this method, the smaller interaction partner L2 is tagged with a fluorescein. For the anisotropy measurement, the sample is irradiated with polarized light. The emitted light is measured via a second polarizer (analyzer). Two different positions of the analyzer are used. The polarizer and analyzer are parallel to each other (I_{II}) and the polarizer and analyzer are perpendicular to each other ($I_\perp$). The anisotropy is calculated from these two measured values and the total intensity. This measurement method makes use of the fact that smaller proteins rotate faster
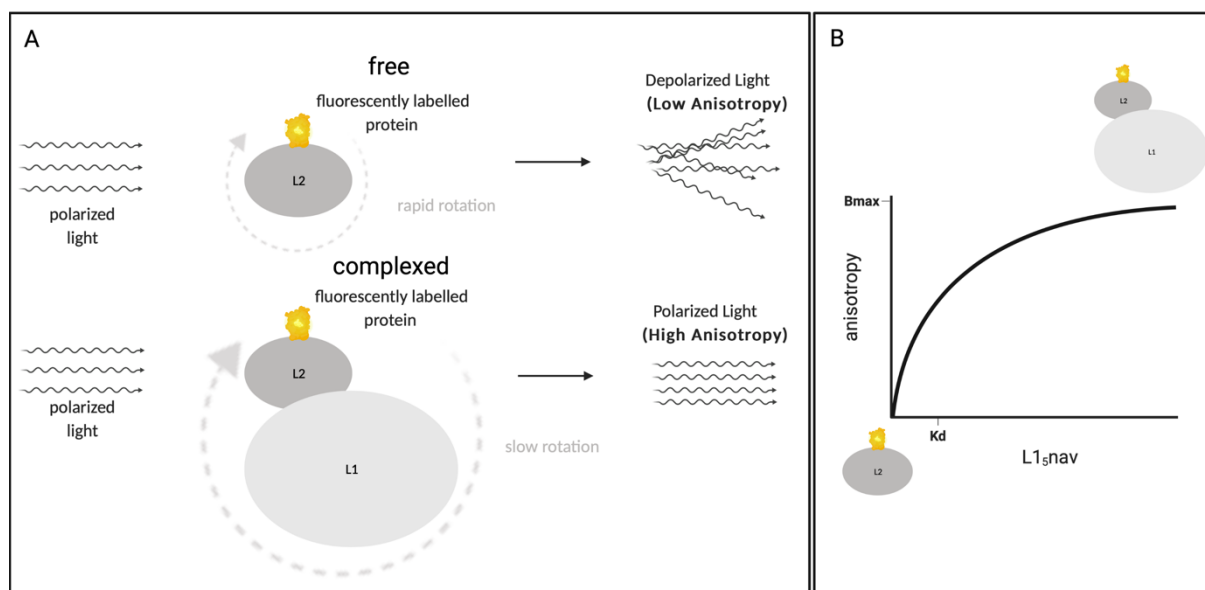
Figure 11: **Anisotropy scheme:** (A) measurement scheme of free and complexed fluorescein labeled protein. (B) Evaluation of binding affinity for anisotropy measurement (created with BioRender).

and depolarize the light more strongly than larger ones (figure 12). In this case the L2 protein was fluorescently labeled via *Sortase A* reaction (section 2.2.20). For direct binding, a constant concentration of 30 nM flourescein-L2 and unlabeled L1nav pentamer diluted 1.5-fold from 1 µM were used. All samples were diluted in refolding buffer containing 1.6 % (w/v) Brij 58. 200 µl sample for each dilution was prepared and incubated for 1.5 h to adjust the equilibrium in each sample. The incubation time to adjust the equilibrium was tested before and an elongation of incubation time did not affect the results. The fluorescence intensity was measured with a FS5 instrument with a black 1 mm quartz cuvette to control the equal concentration of labeled L2. The sample was excited at 499 nm with a band width of 10/10 and the emission was measured from 530 - 540 nm. At the same time the anisotropy was

measured in technical and experimental triplicates. The anisotropy with G-factor was calculated with the Fluoracle® software:

$$Anisotropy = \frac{I_{II} - G * I_{\perp}}{S}$$

G = device factor (ideally 1) otherwise determined independently
I_{II} = polarizer 90°, analyzer 90°
$I_{\perp}$ = polarizer 90°, analyzer 0°
S = total intensity

Additionally, to exclude any artificial binding of the fluorophore-tag to the L1nav protein a competition assay was performed. For this purpose, a sample containing 80 % of the L1-L2 complex was prepared and titrated with a 1.5-fold serial dilution of unlabeled L2 protein. The unlabeled target protein should replace the labeled target and thereby leads to decrease in anisotropy signal. Also, the direct binding assay was repeated with 150 nM L1nav and a 1.5-fold serial dilution row of Fluorescein-Peptide to investigate the non-specific interaction of Peptide and target protein. In this case the anisotropy should not increase if L1nav and the Fluorescein Peptide are not interacting. All data were analysed with GraphPad prism.

2.2.22.3 Surface plasmon resonance (SPR) – Biacore T200

The surface plasmon resonance method in general represents an optical detection process. Surface plasmons are electron oscillations which propagate parallel to a metallic surface. The electron oscillations project into the surrounding medium as so named evanescent waves. As soon as the light hits the interface between two media with different refractive indices, part of the light penetrates the second medium while the other part of the light is reflected. When the angle of the incident light becomes shallower, total reflection occurs at a certain angle, meaning that almost all of the light is reflected. The light which hits a metal surface (e.g. gold) through a prism at the angle of total reflection penetrates the metal as an electromagnetic
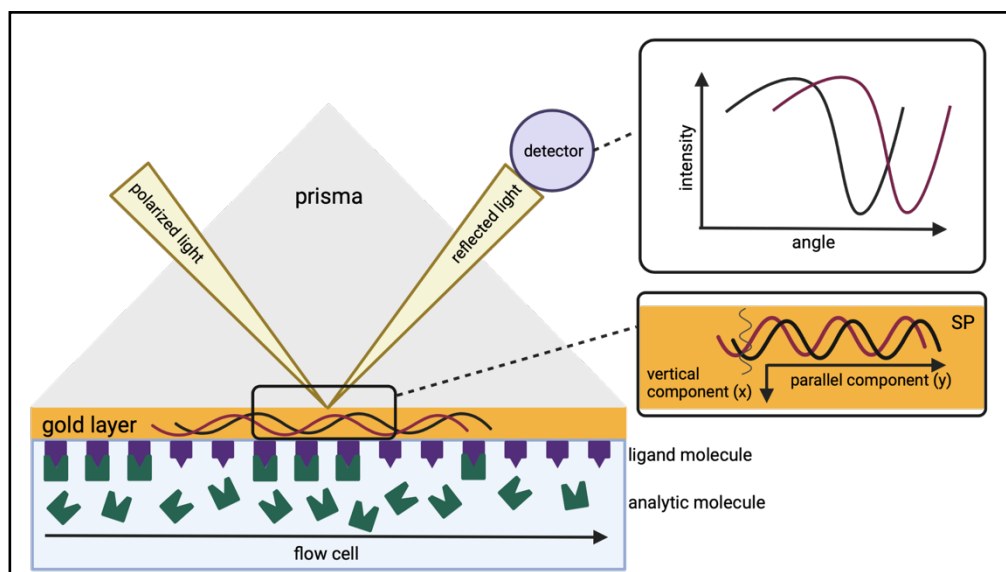
Figure 12: Scheme from Surface Plasmon Resonance (modified from Kubetschek 2021 [2] and created with BioRender)

wave - exponentially decreasing [113]. This penetrating electric field of the light can couple with the conduction electrons of the metal, thus exciting them to oscillation. As soon as the part of the field becomes parallel to the surface coincides with that of the electrons, resonance occurs (surface plasmon resonance). This means that only light at a certain angle of incidence on a metal surface can cause electron oscillation and thus generates resonance. This results in a sharp intensity minimum, a kind of "shadow" in the reflected light as part of the light energy was used to excite the electrons to oscillate. A detector measures the angle at which the intensity minimum of the reflected light occurs. As mentioned before, the angle of total reflection depends on the refractive index of the media. This changes when the mass in the environment is changed [2].
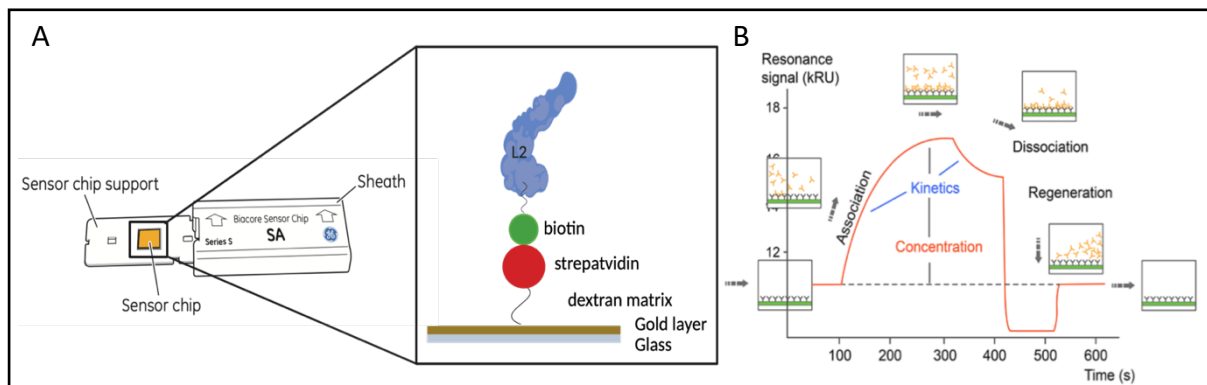
Figure 13: Live tracking of interaction studies via SPR. (A) Series S sensor Chip SA (Cytiva, Sweden) with gold layer and streptavidin coted for biotin-L2 immobilization. **(B)** SPR Sensorgram with different phases during measurement process (created with BioRender)

For live tracking of interaction studies of different molecules, the SPR method has been combined with a sensor Chip technology. This method can be performed with the Biacore T200 (Cytiva, Malborough, MA, USA) system. In practice, molecules, in this case biotinylated L2, are coupled to the gold surface Series S senor Chip SA (Cytiva, Sweden). Above this Chip the prism, which is illuminated, and the detector is located. This measures the intensity of the reflected light depending on the angle and indicates at which angle the intensity minimum occurs. If a binding of analyte (L1nav pentamer) and ligand (L2) occurs, this binding can be detected in real time. Due to an increase in mass on the Chip surface, the refractive index and thus the resonance conditions change, thus the intensity minimum occurs at a different angle of the reflected light. This angular shift can be translated into a so-called sensorgram where the SPR signal is plotted against time.

In the first phase of the sensorgram, the ligands (biotinylated L2) are immobilized on the sensor surface. This is followed by the association phase whereby the buffer with the analyte molecules is flushed across the chambers and the association of ligand and analyte is followed. After this, an equilibrium occurs, in which the analyte molecules (L1nav pentamer) are saturating the ligand (L2) and dissociate from the complex at the same time. As a result, the SPR signal remains constant. The subsequent dissociation phase indicates how fast the complex decays when rinsed with buffer. In the regeneration phase, all remaining bonds between ligand and analyte are dissolved to restore the initial state.

First, the Chip is incubated at room temperature and then washed three times for 60 sec with 1 M NaCl in 50 mM NaOH before the Chip is immobilized with the ligand. Then channel one and two are equilibrated three times for 60 sec at 15 µl/min with the L2 refolding buffer (A0 with 0.5 M L-arginine and additionally 0.005 % (v/v) Tween). Channel one is used as a systematically defined reference (harboring only pre-immobilized streptavidin and no

biotinylated protein) to subserve as a control for non-specific analyte binding. Subsequently, the biotinylated L2 in refolding buffer (0.5 M L-arginine, 0.005 % (v/v) Tween) with a concentration of 4.2 nM is immobilized in four cycles for 12 µl/min at 90 s on channel two until a resonance of 20 is reached. The immobilized and reference channel is washed three times for 60 s at 15 µl/min flow rate with MST buffer (0.16 % (w/v) Brij 58). As an analyte, L1nav pentamer is loaded on channel one and two for 120 s at 15 µl/min flow rate, followed by a dissociation phase for 180 s. After each cycle, the L1nav protein is removed with 6 M guanidine hydrochloride at 30 µl/min flow rate for 180 s and afterwards the L2 protein on channel two is refolded by incubation with refolding buffer (0.5 M L-arginine, 0.005 % (v/v) Tween) at 30 µL/min flow rate for 180 s. The SPR is also performed with the biotin-L2ΔL1bs and at different pH values.

For kinetic measurements, a 1-fold serial dilution of L1nav pentamer with a concentration range from 2.2 nM to 9.1 µM is used. The concentration series are recorded three times in a row. First, the data is analyzed using Biacore T200 Evaluation software (Cytiva) by correcting for buffer effects and refractive index changes. Then, an in-house Python script can automatically overlay and split each sensorgram. In doing so, the theoretical maximum normalized binding signal $R_{max,norm}$ is determined according to [114]:

$$R_{max,norm} = \frac{MW_{analyte}}{MW_{ligand}} R_{immob}$$

$MW_{analyte}$ and MW_{ligant} represent the molecular weight of biotinylated L2 (ligand) and L1nav pentamer (analyte). R_{immob} thus shows the immobilization level of biotinylated L2. It is assumed that the ligand is 100% active on the surface and one analyte binds to one ligand at a time. Thus, different sensorgrams can be compared directly.

3   RESULTS

3.1   Investigation of cell growth and optimization of soluble expression for different L1 constructs

As demonstrated in figure 8 different L1 full-length constructs as well as different L1 non-assembly variant constructs (L1nav) with various affinity tags are used in this thesis.

The aim was to investigate the bacterial cell growth expressing these proteins and to optimize the expression of soluble L1 protein because the final yield of purified protein depends mostly on the solubility of recombinant expressed protein. To monitor the bacterial cell growth and optimize the expression for the different L1 constructs, three different parameters were monitored: (I) different expression temperatures, since a reduced expression temperature leads to reduced protein expression and thus can affect soluble protein production [115, 116]. (II) Different fusions tags (His$_6$-SUMO-, GST- and His$_{6/10}$-tag) which can facilitate solubility and (III) the *E. coli* chaperon GroEL/ES which assists in protein folding [97, 98, 117].

The bacterial cell growth was investigated via optical density and to analyze the solubility of expressed protein, lysed sample was separated via centrifugation because large aggregates sediment in the pellet fraction and soluble protein in the supernatant.

The cell growth of the various L1 constructs (figure 8) at different expression temperatures showed no differences in bacterial cell growth except for retarded growth at 16 °C and 25 °C (figure S1). Also, the different fusion tags showed no differences in cell growth as well the co-expression of the chaperon complex GroEL/ES (figure S2) which leads to the conclusion that these factors have negligent influence on the cell growth and final yield. The expression conditions shown in figure 14 are as follows: co-expression of the chaperon GroEL/ES complex and expression at 25 °C. The GST-L1nav construct was included as a comparison because it
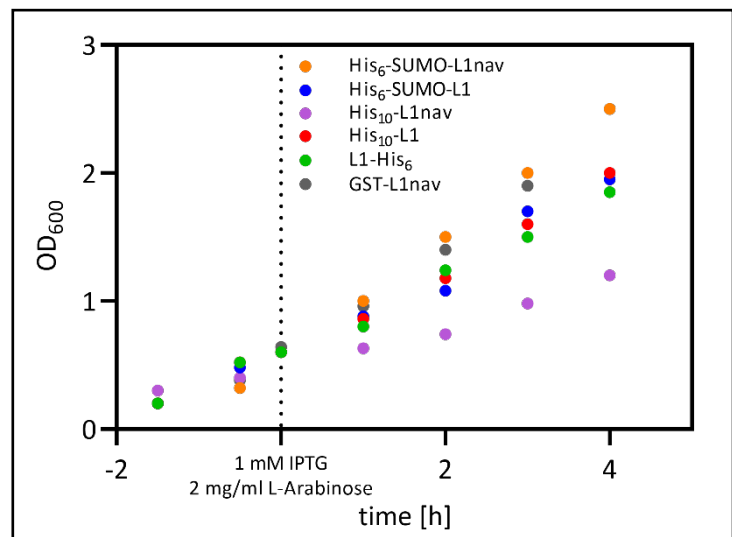
Figure 14: Growth curve of bacteria expressing different L1 constructs at 25 °C. L1 constructs containing cells were grown at 37 °C up to an OD$_{600}$ = 0.5 and afterwards the expression temperature was set to 25 °C. To start protein expression for L1nav we added 1 mM IPTG and 2 mg/ml for GroEL/ES co-expression. The bacteria were cultivated for 4 h and the growth was monitored by optical density (OD$_{600}$).

is often used in literature [103-105]. To determine the ratio between soluble and insoluble protein the samples were centrifuged at 12.000 g, 30 min and 8 °C. It is clearly demonstrated that the amount of soluble L1 full-length as well as L1nav protein in the supernatant increases as soon as the chaperone GroEL/ES complex is co-expressed. For the His$_6$-SUMO-L1nav construct less than 20 % soluble protein is located in the supernatant, however after co-expression with GroEL/ES complex the amount of soluble His$_6$-SUMO-L1nav protein increases up to 80 %. The same is shown for the His$_{10}$-L1nav construct. Less than 20 % soluble protein is localized in the supernatant, but the soluble protein content of this construct also increases to more than 50 % when GroEL/ES is co-expressed (figure 15 B). For the GST-L1nav construct 95 % soluble protein is shown, therefore this construct is not considered for further optimization [1]. Figure 15 A shows the results for the L1 full-length constructs with different
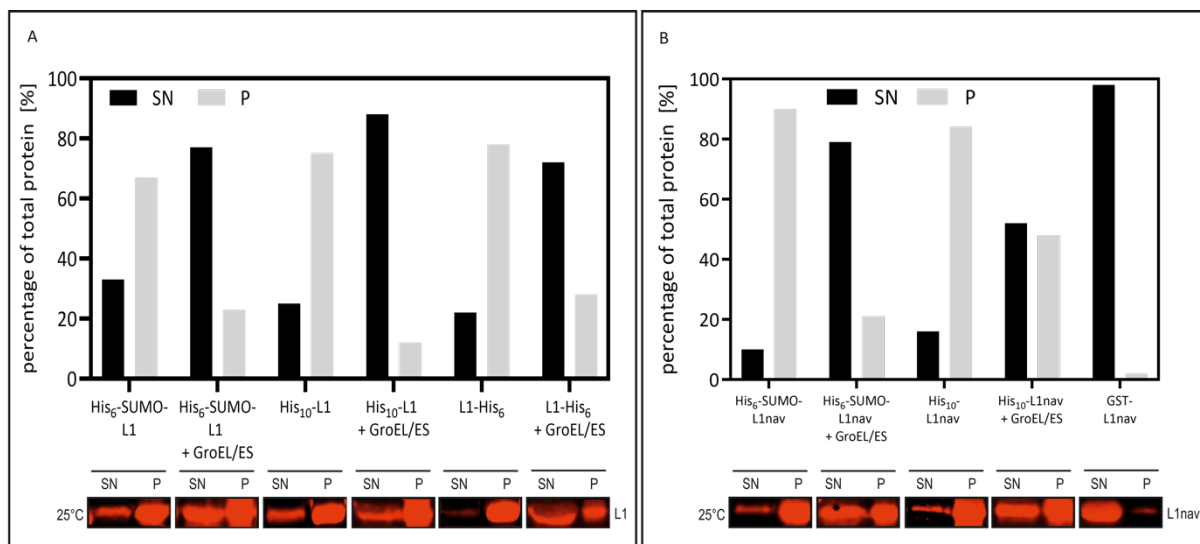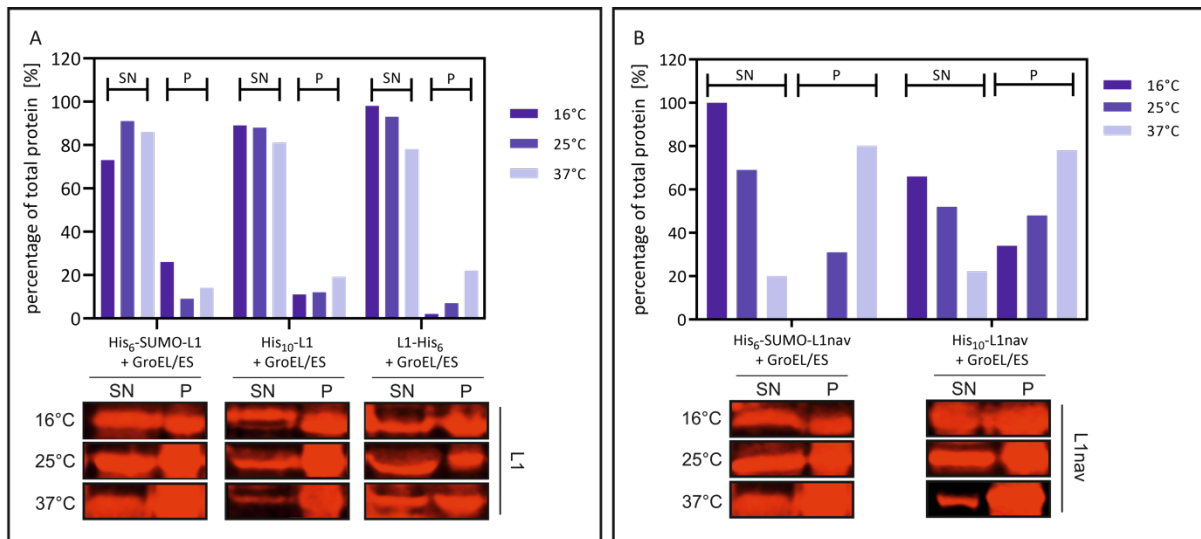

Figure 15: Solubility of recombinant L1nav protein under the influenced of GroEL/ES overexpression and N-terminal fusion tags. **(A)** L1 full-length constructs. **(B)** L1nav constructs modified according to Roos et. al. [1]. Lysed cells were centrifuged to separate insoluble (pellet, P) and soluble fraction (supernatant, SN). Corresponding to 210 µg dry biomass, supernatant and pellet were equally loaded on SDS-PAGE. The L1 protein was specifically detected using a fluorescently labeled antibody on a Western Blot (Li-Cor System) and afterwards quantified densitometrically via Li-Core Odyssey Fc software. The GST-L1nav construct show nearly 100 % soluble protein. Compare to that 10-20 % soluble protein is visible for His$_{10}$- and His$_6$-SUMO-L1nav construct. With GroEL/ES co-expression the soluble protein amount for His$_{10}$- and His$_6$-SUMO-L1nav construct were increased up to 5-fold. For the His$_6$-SUMO-, His$_{10}$- and L1-His$_6$ constructs the amount of soluble protein differs from 20-35 %. In comparison, the amount of soluble protein increases up to 3-fold with a co-expression of GroEL/ES. The total amount of protein was calculated according to the relative amount of GroEL in the supernatant and pellet (figure S3) assuming that GroEL is a highly soluble protein.

fusion tags. This shows ~ 30 % soluble His$_6$-SUMO-L1 protein, ~ 25 % soluble His$_{10}$-L1 protein and ~ 20 % soluble L1$_6$-His protein. When GroEL/ES is co-expressed, the amount of soluble protein increases up to 85 % for all constructs (His$_6$-SUMO-L1 ~ 80 %, His$_{10}$-L1 ~ 85 % and L1$_6$-His ~ 75 %).

For further optimization the influence of the expression temperature (16 °C, 25 °C and 37 °C) in combination with the GroEL/ES co-expression was tested. For the His$_6$-SUMO-L1nav and

His$_{10}$-L1nav constructs the result shown in figure 16 B are already published by Roos et. al.[1]. The amount of soluble mutated L1nav protein could be improved up to 95 % for the His$_6$-SUMO-L1nav construct and ~ 60 % for the His$_{10}$-L1nav construct (figure 16 B) at an expression

Figure 16: Solubility of recombinant L1 protein expressed under different expression temperatures. (A) L1 full-length constructs. **(B)** L1nav constructs modified according to Roos et. al. [1]. Bacterial cultures were crown at 37 °C until an OD$_{600}$ = 0.5 and afterwards protein expression was performed at 16 °C, 25 °C and 37 °C. Therefore, protein expression was induced with 1 mM IPTG for L1 constructs and 2 mg/ml L-Arabinose for GroEL/ES and the bacterial cultures were cultivated for 4 h. The total amount of protein was calculated according to the relative amount of GroEL in the supernatant and pellet (figure S3) assuming that GroEL is a highly soluble protein.

temperature of 16 °C. On the contrary, at an expression temperature of 37 °C, the amount of soluble L1nav protein decreases to ~ 20% for the His$_6$-SUMO-L1nav and His$_{10}$-L1nav construct. In figure 16 A the results for the L1 full-length protein for different expressions temperatures are shown. For the His$_6$-SUMO-L1 construct ~ 85% soluble L1 protein is localized in the supernatant at an expression temperature of 25 °C. For 16 °C, ~ 70 % and for 37 °C ~ 80 % appears the supernatant. The highest solubility for the His$_{10}$-L1 construct is shown by an expression temperature of 16 °C and 25 °C with ~ 85 % soluble L1 full-length protein. At an expression temperature of 37 °C, the amount of soluble L1 protein decreases slightly to ~ 80 %. The L1-His$_6$ construct present the highest amount of soluble L1 protein (~ 95 %) at 16 °C expression temperature. For 25 °C ~ 90 % L1 is present in the supernatant and for 37 °C ~ 80 %.

In summary these results demonstrate that the best expression conditions for soluble L1nav and L1 protein are: (I) reduced expression temperature of 16°C or 25°C, (II) solubility induced fusion tag (His$_6$-SUMO-, GST- or His$_{10/6}$-tag) and in parallel (III) a co-expression of chaperon complex GroEL/ES.

3.2 Fermentation of *E. coli*

The expression of soluble L1 and L1nav protein, in *E. coli* in flasks proved to be rather unsuccessful due to pretests (figure 19 line II). In flasks, bacterial growth occurs under non-limiting conditions which may have an influence on protein yield. Consequently, a fed-batch fermentation was established in defined and enriched M9 medium. Fermentation has the advantage that the growth of the bacteria can be controlled by the addition of
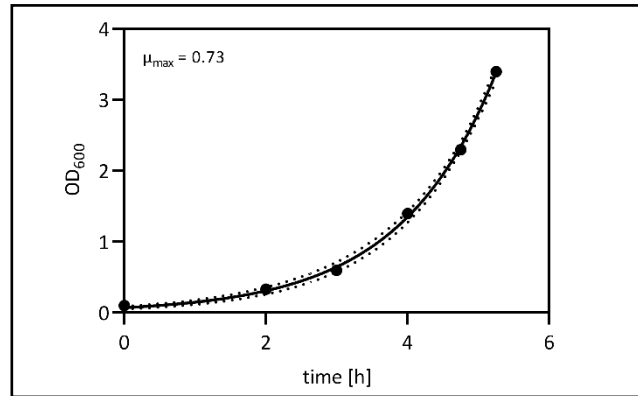
Figure 17: Determination of μ_{max} for *E. coli* fermentation containing His$_6$-SUMO-L1nav construct in BL21(DE3)pGro7 background. The cells were grown in 1 L LB medium (in 5 L flasks) at 20 °C for 4h under non-limited conditions at 100 rpm. The bacterial growth was measured via optical density OD$_{600}$ and the μ_{max} was calculated via GraphPad.

substrate. Results from yeast fermentation show an increase of the expressed protein due to substrate control and the associated reduced growth [118]. For substrate limiting fed-batch fermentation, the μ_{max} value under non-limiting conditions is required. Therefore, the bacteria were expressed in 1 L LB medium at 20 °C and 100 rpm for 4 h and cell growth was displayed by optical density (OD$_{600}$) (figure 17). The μ_{max} of 0.73 h^{-1} was calculated with an exponential growth fit via GraphPad.

3.2.1 Process of Fermentation

The fermentation was successful in terms of cell growth and product generation. The following aspects should be investigated: growth of the culture, substrate utilization and continuous expression of L1/L1nav protein.

3.2.1.1 Growth of culture

To analyze the bacterial cell growth during the fermentation process, the optical density (OD$_{600}$) of the culture was analyzed over time and substrate utilization was determined by the substrate (Glucose (g/L)) amount present in the culture via a blood glucose monitor (ACCU-CHEK, Roche). Figure 18 shows on the one hand the simulated fermentation process for cell growth (OD$_{600}$ – black dots) and feed (green dots) (figure 18 A) and on the other hand the actual growth process (OD$_{600}$) and substrate amount (sugar g/L) in the culture over time (figure 18 B). The simulated exponential bacterial growth curve (OD$_{600}$) and feed rate (g/h) shown in figure 18 A in black and green dots were calculated:

$$OD_{600} = \frac{X(g)}{\dfrac{0.3(conversion\ factor)}{Volumen(L)}}$$

$$X(g) = X_{t=0}(g) * e^{\left(\mu_{set}\left(\frac{1}{h}\right)*t_{10}(h)-t_0(h)\right)}$$

$$Volumen(L) = Volumen_0(L) + \left(\frac{feed_{10}(g)}{1000}\right)$$

$$feed(g) = feedrate\left(\frac{g}{h}\right) * \left(t_{10}(h) - t_0(h)\right) + feed_0(g)$$

$$feedrate\left(\frac{g}{h}\right) = \left(\left(\frac{\mu_{set}\left(\frac{1}{h}\right)}{Y_{XS}}\right) + m_S\left(\frac{g_S}{g_X + h}\right)\right) * \left(\frac{X_0(g)}{c_{Sfeed}\left(\frac{g}{g}\right)}\right)$$

Therefore, parameter μ_{set} with 0.1 h^{-1}, Y_{xs}= 0.5, m_s= 0.04, c_{Sfeed}= 0.3 and the calculated dry biomass ($X_{t=0}(g)$) based on the measured OD_{600} (section 2.2.9.2) were used and calculated via Microsoft Excel 2016 and displayed via GraphPad. The actual measured OD_{600}= 40 for the His$_6$-SUMO-L1nav construct after 14.5 h fermentation is displayed in figure 18 A with a red dot.
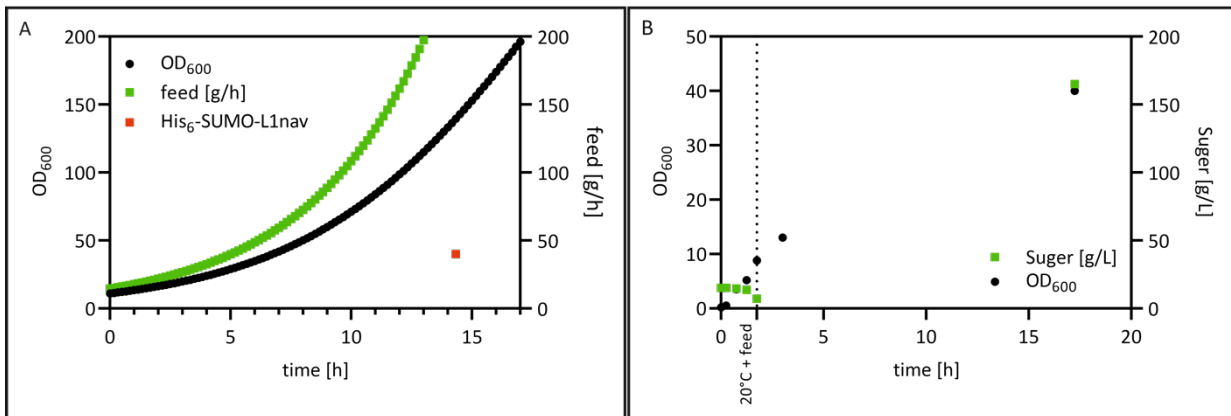
Figure 18: Calculated exponential feed/OD$_{600}$ and measured OD$_{600}$/substrate from His$_6$-SUMO-L1nav construct in BL21(DE3)pGro7. (A) Black dots: calculated OD$_{600}$ related to the pumped feed. Green dots: calculated exponential feed in g/h. Red dot: final measured OD$_{600}$ of His$_6$-SUMO-L1nav construct in BL21(DE3)pGro7 after 14.5 h fermentation at 20 °C. **(B)** Growth curve for His$_6$-SUMO-L1nav construct in BL21(DE3)pGro7 predicted in black. Amount of substrate g/L is shown in green. The whole fermentation process over 17 h for the bacterial growth and substrate amount is shown. After 2 h growth at 37 °C a minimal amount of substrate is shown, and the temperature was reduced to 20 °C and the exponential feed was launched. The recombinant protein expression was induced via feed over time starting at 2 h with 1 mM IPTG for His$_6$-SUMO-L1nav and 2 mg/ml L-Arabinose for GroEL/ES co-expression.

In figure 18 B the actual growth curve and substrate curve is shown. Thereby, the first two hours describe the bacterial growth at 37 °C and the decreasing amount of glucose. Before the exponential feed was set, the substrate in the medium needs to be used up to start the bacterial growth under substrate limited conditions. The culture was set to 20 °C because the solubility test in section 3.1 showed an increased amount of soluble L1 protein at lower expression temperatures. The optical density of the culture increases to OD_{600} = 40 after 17 h and the substrate (30 % (w/v) Glucose) to 170 g/L. The average specific growth rate was

calculated as 0.12 which fits μ_{set}= 0.1 rate. Further important parameters during the process-controlled fed-batch fermentation are: the oxygen content with 30 % ± 10 % (pO2) serves as the parameter to be controlled on the first level via air flow and pure oxygen flow. The airflow, the exitO2 and the pumping capacity of the feed pump and the base pump are displayed in figure S4 and all these addressed parameters indicate bacterial growth over time.

3.2.1.2   Yield

In all process controlled fermentations performed for the different L1 constructs produced in BL21(DE3)pGro7, an OD_{600} of 30 - 40 was achieved. This corresponds to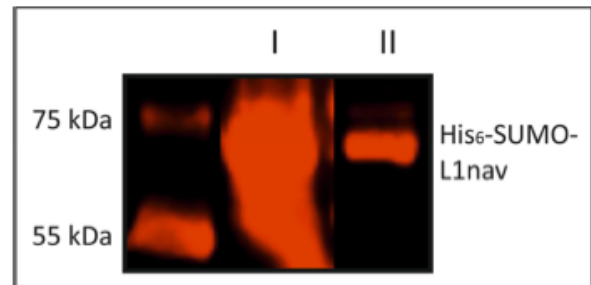 a dry biomass of 9 - 12 g per liter of medium. Compared to fermentation, the growth rate and protein expression in M9 medium in shake flask without feed is significantly lower. In the fermentation culture, the biomass is

Figure 19: His$_6$-SUMO-L1nav expression in 5 L flasks and fermenter. The construct is expressed in BL21(DE3)pGro7 o/n at 20 °C. (I) o/n expression sample from fermenter. (II) o/n expression sample from 5 L flask. Both samples were load equally corresponding to 45 µg dry biomass on an SDS-PAGE and L1nav was detected specifically on a Western Blot with a Li-Core System (Roos et. al).

about 10 times higher than in the flask (OD_{600} = 4; data not shown) and the protein expression per biomass is also 10 times higher for the L1nav construct as displayed in figure 19 [1]. Expression analysis was performed using the specific L1 antibody signal and subsequent calculation of the protein ratio (2.2.7). Overall, this result clearly demonstrates that the yield of the expressed recombinant protein is dependent on the produced biomass and the intracellularly expressed target protein.

In summary, it was possible to cultivate *Escherichia coli* reproducible in a process-controlled fed-batch fermentation and express recombinant L1 protein with a high yield (figure 19). However, we were not able to optimize the fermentation process so far that the bacteria grow under substate limited conditions. This may be related to the design of the fermentation device since the maximum gassing was too low for *E. coli* bacteria. In addition, high substrate concentrations in the medium can lead to substrate excess inhibition, depending on the microorganism. This physiological phenomenon reduces the growth rate as well. Nevertheless, we were able to express L1 recombinant protein for continuous characterization.
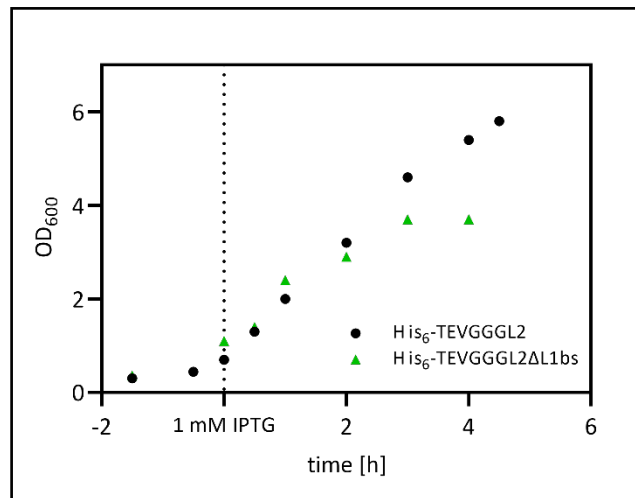
3.3 L2 expression in inclusion bodies

There are different strategies for the expression of the HPV16 L2 capsid protein. One strategy is the co-expression with the major capsid protein L1 in *E. coli* and Yeast [29, 119, 120]. Another strategy is the production of pseudovirions in mammalian cells [121]. Thereby, the L2 protein is incorporated into the viral capsid. However, the protein yield of L2 by these methods is insufficient for functional characterization of L2. In *E. coli* the expression of L2 is performed as

Figure 20: Bacterial growth curve from L2 constructs in *E. coli*. Expression for His_6-TEV-GGGL2 (black dots) and His_6-TEV-GGGL2ΔL1bs (green dots) constructs are shown. Cells were grown until OD_{600} = 0.5 - 0.7 at 37°C before protein expression was induced by 1 mM IPTG at t=0. Cells were grown for 4-5 h at 37 °C until harvested at 6.000g for 30 min at 8°C.

inclusion bodies and a high yield can be reached [77, 122]. Based on these results, the L2 constructs (His_6-TEV-GGGL2, His_6-TEV-GGGL2ΔL1bs) are expressed in *E. coli* as inclusion bodies. The expression was performed at 37 °C for 4-5 h at 100 rpm (figure 20). These growth and expression conditions are already published for a His_6-L2 construct [77]. Induction of protein expression was performed at an OD_{600} of 0.7 - 1 with 1 mM IPTG. After 4-5 h of recombinant protein expression the bacterial cell cultures were harvested and either stored at -20 °C or lysed and recombinant L2 protein was purified as described in section 3.6.
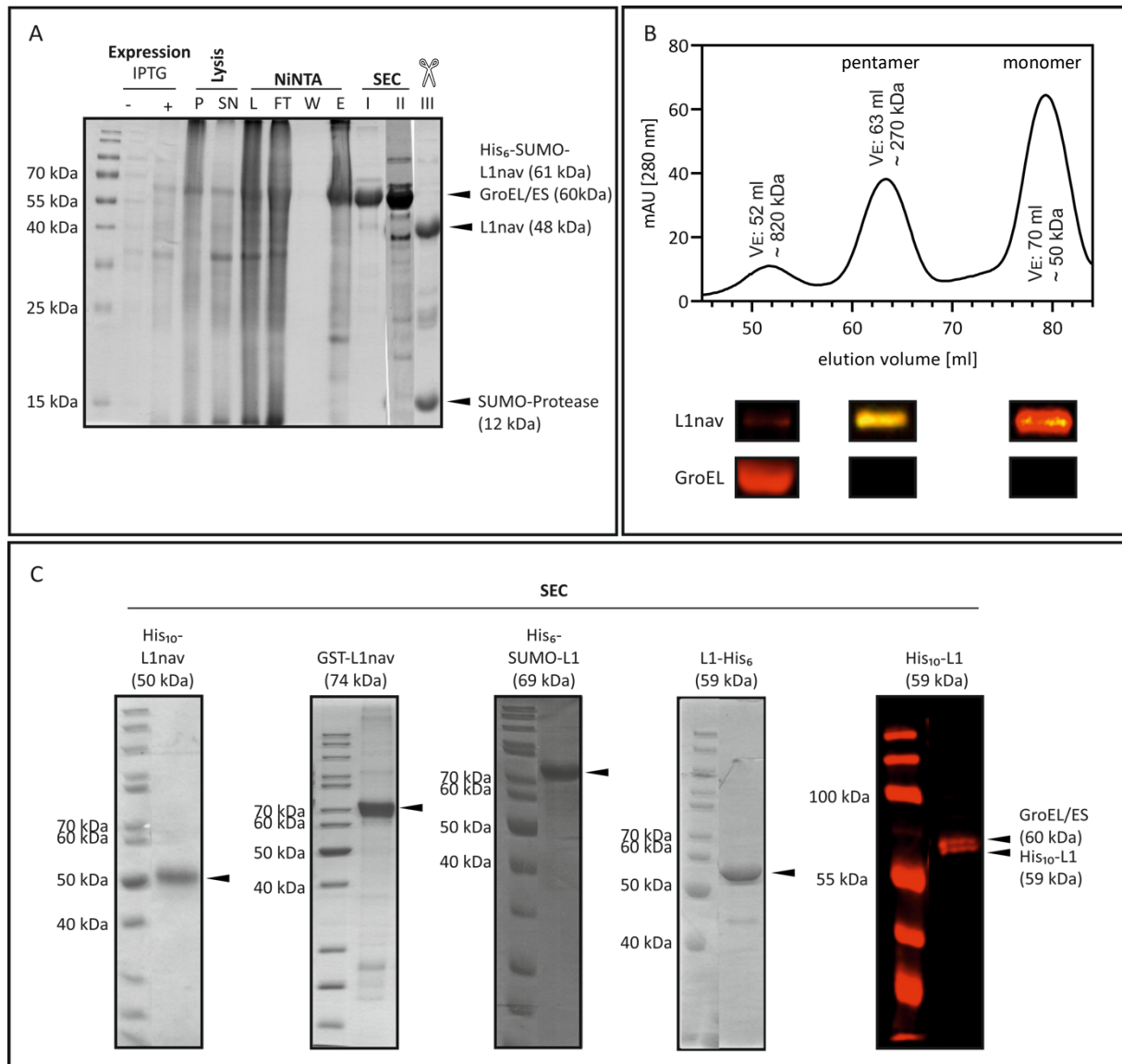
3.4 L1 constructs – purification of soluble L1 proteins

The mostly described method for L1 purification from a *E. coli* expression system is the affinity chromatography with a GST-tag [123, 124]. In addition, several publications also address purification with a His_6-tag or His_6-SUMO-tag [102]. The GST-affinity is very specific but 10-fold lower than the affinity of a His_6-tag to a NiNTA [125, 126]. For this reason, purification slows down considerably.

Figure 21: History of purification of different L1 constructs analyzed via SDS-PAGE. (A) Expression without (-) and with (+) induction of recombinant L1nav and GroEL/ES protein for 4 h at 20 °C; lysis: insoluble pellet (P) and soluble supernatant (SN) were separated by centrifugation, NiNTA: consists of load (L), flow-through (FT), wash (W) and protein elution (E); Size exclusion chromatography (SEC): final used His_6-SUMO-L1nav protein for SUMO-Protease cleavage (see SEC chromatogram 22 B); Line I shows His_6-SUMO-L1nav pentamer; Line II shows His_6-SUMO-L1nav monomer; Line III shows final cleaved L1nav pentamer with His_6-SUMO-Protease **(B)** Size exclusion chromatography (HiLoad 16/600 Superdex 200 PG column) with A0 + 1mM TCEP buffer; peak I corresponds to the size of the GroEL/ES complex ($V_E \sim 52ml \sim 820$ kDa), peak II corresponds to the His_6-SUMO-L1nav pentamer with a $V_E \sim 63ml \sim 270$ kDa and Peak III shows His_6-SUMO-L1nav monomer with a $V_E \sim 79ml \sim 50$ kDa . Complete separation of GroEL/ES and His_6-SUMO-L1nav protein is confirmed by Western Blot analysis with specific His- and L1-antibody. **(C)** Final purified L1 protein for different L1 (L1 full-length and L1nav) constructs before tag cleavage but after size exclusion chromatography (HiLoad 16/600 Superdex 200 PG column). For His_{10}-L1nav, GST-L1nav, His_6-SUMO-L1 and L1-His_6 no GroEL/ES protein is left but for the His_{10}-L1 construct which is shown via Western Blot analysis and specific fluorophore labeled Antibodies for L1 and GroEL/ES.

To investigate and optimize the purification for different L1 constructs with different affinity tags (His$_6$-tag or GST-tag) shown in figure 8 the purification of the soluble L1 full-length protein and the soluble L1 non-assembly variant was carried out via affinity chromatography and size exclusion chromatography.

After the induced expression with 1 mM IPTG (figure 21 A; expression, line +) the bacteria cells were lysed and separated in insoluble cell compartments (figure 21 A; lysis, line P) and soluble compartments (figure 21 A; lysis, line SN). The supernatant containing the soluble His$_6$-SUMO-L1nav protein was then applied on a NiNTA (figure 21 A; NiNTA). The affinity chromatography process demonstrates a prominent protein band at 60 kDa in the elution fraction of the NiNTA (figure 21 A; NiNTA, line E). This fraction was then applied on a size exclusion chromatography (SEC) for polishing the His$_6$-SUMO-L1nav protein (figure 21 A; SEC line I and II). The corresponding SEC chromatogram is displayed in figure 21 B and three peaks are visible at different elution volumes. The first peak with an elution volume of 52 ml corresponds to a protein size of 820 kDa. Further investigation with specific GroEL/ES antibody confirmed GroEL/ES protein in this peak. The second peak with an elution volume of 63 ml corresponds to a molecular weight of 270 kDa and specific L1- and His-antibody demonstrated the pentameric species of the His$_6$-SUMO-L1nav protein. In figure 21 A; SEC line I the peak was analyzed via a Coomassie stained reducing SDS-PAGE and demonstrated a band at 60 kDa indicating His$_6$-SUMO-L1nav protein. The third peak at an elution volume of 79 ml corresponds to 50 kDa and the specific L1- and His-antibody displayed the monomeric species of His$_6$-SUMO-L1nav protein. This peak was also analyzed via Coomassie stained reducing SDS-PAGE shown in figure 21 A; SEC line II and displays a prominent protein band at 60 kDa. However, neither the elution peak at 63 ml nor the elution peak at 79 ml showed a specific signal from the GroEL/ES antibody. Thus, it is demonstrated that the bound GroEL/ES protein can be separated from His$_6$-SUMO-L1nav protein by SEC. The purification process demonstrated in figure 21 A and B was modified from Roos et. al. [1]. The final purified and uncleaved protein for the different L1 constructs after SEC are shown in figure 21 C. The related purification process by affinity chromatography analyzed via reducing SDS PAGE or Western Blot analysis for the individual L1 constructs is present in the supplements (figure S5-S8).

The purification for the GST-L1nav construct was recently published and was performed as described there [127] and is not shown for this reason. Please note, for the removal of the GroEL/ES complex form the GST-L1nav construct an additional treatment with ATP/Urea is

necessary [127]. For the His_{10}-L1 no complete separation of L1 protein and GroEL/ES complex was possible (figure 21 C). The L1 constructs (except GST-L1nav construct) were purified from cell lysate in two steps. From these results, we conclude that the purification of the L1 constructs with an affinity tag followed by a polishing step via SEC was successful except for the His_{10}-L1 construct.

The challenges of purification centered on the separation of GroEL/ES and the L1 protein but the separation was successful by SEC. It was also shown that L1 exists as both a pentameric and a monomeric species. In a SEC, these two species showed clearly distinguishable peaks and can thus be separated from each other. Doing so we observed 35 % His_6-SUMO-L1nav pentamer and 65 % His_6-SUMO-L1nav monomer. The final purified proteins were then used to cleave off the fusion tag as described in section 3.5. This tag removal is crucial for subsequent functional analyses.

3.5 L1 pentamer constructs - cleavage

The removal of the fusion tag is important for subsequent functional analyses, since the sometimes very large tag, depending on the used tag, can lead to artificial results [128]. For this reason, an affinity-tag must be at least 80 % cleaved to obtain completely cleaved L1 pentamer. At a lower cleavage efficiency than 80 %, at least one affinity tag per pentamer would remain [1]. The Small Ubiquitin-like Modifier (SUMO) tag is paired with a His_6-tag for purification and can be specifically cleaved by the SUMO-Protease [106]. The SUMO-Protease cleaves without remnants of any tag-based amino acid at the N-terminus, and this allows a native N-terminal sequence of the recombinant protein [129]. For the investigation of the cleavage efficiency, various parameters such as temperature, protease:protein ratio and time were addressed.

The different fusion tags are cleaved by different proteases. GST-tag is cleaved by thrombin, His_6-SUMO-tag by SUMO-Protease and His_{10}-tag by enterokinase (figure 22 A, B, C – modified Roos et. al. [1]). For the full-length L1 construct, we observed that neither the His_{10}-tag nor
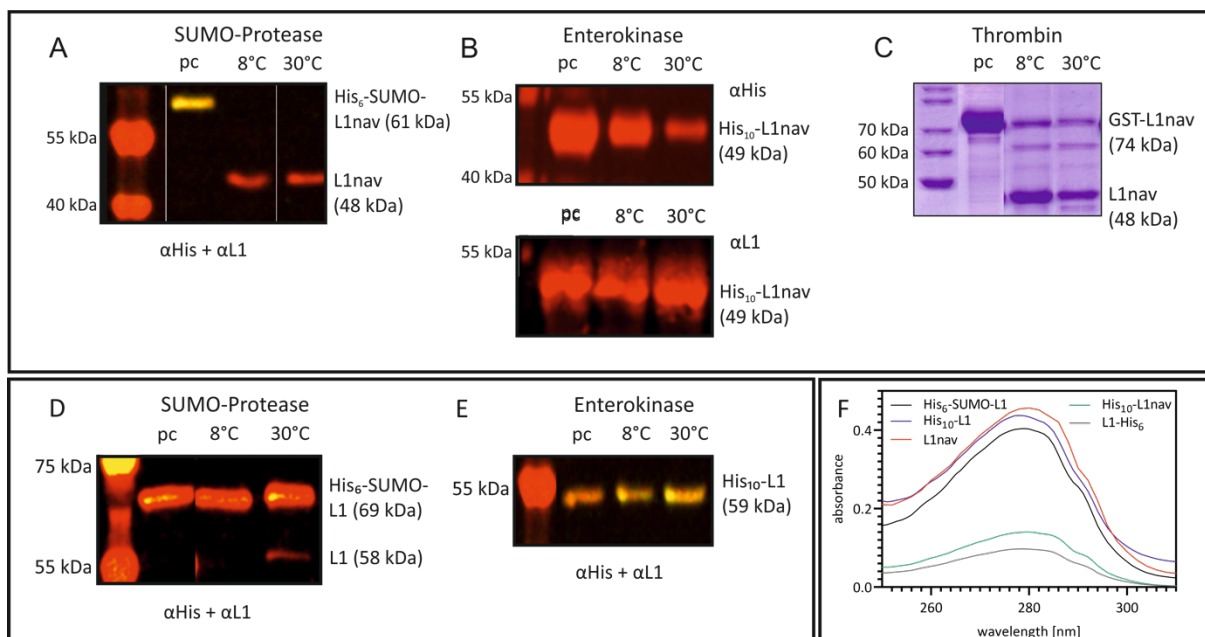
Figure 22: Cleaving efficiency of the different L1 constructs with different fusion tags. SUMO-Protease, enterokinase and thrombin were used – according to the constructs in figure 8. **A-C L1nav constructs** and **D-E L1 full-length** constructs at different temperatures (8°C and 30°C o/n). Cleavage reaction was followed by Western Blot analysis and Coomassie stained SDS-PAGE **(A)** The His_6-SUMO-L1nav construct was incubated with 1:10 mass ratio (protease:protein) SUMO-Protease. **(B)** For the His_{10}-L1nav construct the enterokinase was used with a mass ratio of 1:20 (protease:protein). Cleavage was analyzed via Western Blot analysis and specific antibodies for L1 and His. **(C)** The GST-L1nav construct was cleaved via thrombin in a mass ratio 1:8 (protease:protein). The mass ratio 1:8 corresponds to 0.025 NIHU/µg protein. **(D)** His_6-SUMO-L1 construct was cleaved with SUMO-Protease in a 1:10 mass ratio (protease:protein). **(E)** The His_{10}-L1 construct was cleaved via enterokinase with a mass ratio 1:20 (protease:protein). Only the His_6-SUMO-L1nav construct can be cleaved completely by SUMO-Protease independent of incubation temperature. **(F)** UV/VIS spectra of different L1 constructs. $A_{260}:A_{280}$ ratio of ~0.5 indicates an almost DNA free protein sample.

the His_6-SUMO-tag could be completely cleaved, even though we changed the incubation time, temperature and protease:protein ratio (figure D, E).

Compared with the manufacturer´s recommendation and previously published protocols, a comparatively high concentration was already used for both enterokinase (1:20) and thrombin (1:8). Cleavage of the GST-L1nav and the His_{10}-L1nav protein was also not completely possible. Again, different parameters (temperature and time) were optimized and a high concentration of thrombin (1:8) and enterokinase (1:20) was already used. As mentioned before, we increased the temperature from 8°C to 30°C to raise the enzymatic activity of the protease. Due to the incomplete cleavage of the L1 and L1nav proteins, the separation of cleaved and uncleaved protein is only possible to a limited extent. Due to the still present affinity-tag still present, a large proportion of the protein remained attached to the affinity columns. For this reason, the amount of cleaved protein dropped dramatically.

For the His_6-SUMO-L1nav construct, the His_6-SUMO-tag can be completely cleaved at 8 °C as well as at 30 °C with a protein:protease ratio of 1:10 (figure 22 A). Western blot analysis in figure 22 A shows no uncleaved L1nav protein. This can also be seen in a preparative scale (figure 21 A Line III) [1]. The UV-VIS spectra for all cleaved L1 proteins in figure 22 F show a A_{260}:A_{280} ratio of ~ 0.5. This implies that the recombinant L1 protein is nearly free of DNA contamination.

Separation of cleaved tag and respective protease were only achieved for His_6-SUMO-L1nav construct and since a complete cleavage is necessary because the yield of cleaved protein decreases dramatically for incompletely cleaved constructs, His_6-SUMO-L1nav construct is used for further L1nav purification. To simplify and accelerate the subsequent downstream process, we combined removal of the His_6-SUMO-tag of the His_6-SUMO-Protease and separation of the cleaved L1nav protein. For the final polishing step, a NiNTA was directly coupled to a preparative size exclusion column (HiLoad 16/600 Superdex 200 PG column). The final purified L1nav protein was then checked for purity (figure 33 A). For further experiments with the L1 full-length proteins, the purified proteins with a uncleaved $His_{6/10}$ tag (located N-terminal and C-terminal) were used (shown in figure 21 C) because an His affinity tag (6 or 10 aa) is rather small compared to GST- or His_6-SUMO-tag and thus small tag size may hinder less protein-protein interaction or assembly.

For the L1nav pentamer 5.73 mg purified protein per gram dry biomass with a purity of 95 % was produced and 10.64 mg with 97 % purity for the L1nav monomer (figure 33 A) [1].

3.6   L2 constructs – purification of insoluble L2 proteins

The purification for a His_6-L2 construct is already described by Breiner et. al. [77]. The protein purification for the His_6-TEV-GGG-L2 and His_6-TEV-GGG-L2ΔL1bs construct from inclusion bodies was carried out as shown in figure 23 A.
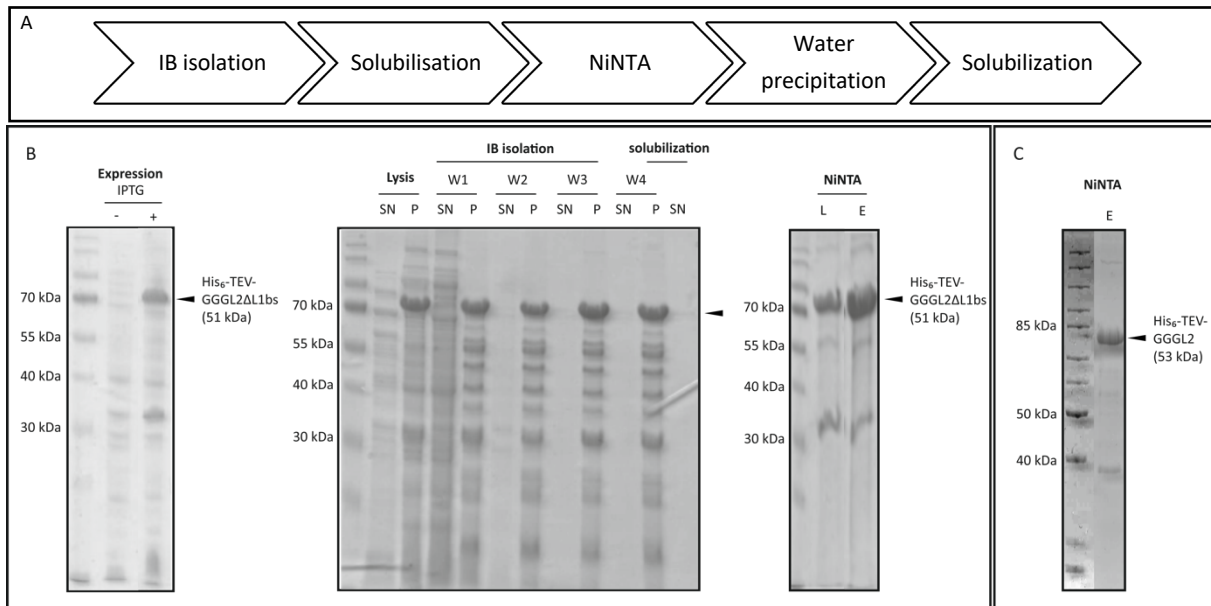
Figure 23: Overview of L2 Purification. (A) Overview of inclusion body purification of L2 constructs. **(B)** Expression: before induction (-) and after induction (+) with 1mM IPTG for 4h at 37°C for His_6-TEV-GGGL2ΔL1bs construct. Lysis: cell lysate is separated in soluble (supernatant, S) and insoluble (pellet, P) protein fraction by centrifugation. W1-W4: wash (10 mM Tris • HCl pH 7.4 at 8 °C, 200 mM NaCl, 5 mM EDTA) of inclusion bodies. NiNTA: solubilized protein pellet (6M guanidine hydrochloride, 20 mM Tris • HCl pH 7.4 at 8 °C,) under denatured conditions was loaded (L) on a NiNTA and purified His_6-TEV-GGGL2ΔL1bs protein eluted (E). **(C)** NiNTA Elution (E) of purified denatured His_6-TEV-GGGL2 protein.

We reached a final yield of ~ 40 mg purified and solubilized protein, for the His_6-TEV-GGGL2 (figure 23 C) and His_6-TEV-GGGL2ΔL1bs (figure 23 B) construct, per liter culture medium in shaking flasks. The purification process for His_6-TEV-GGGL2 is shown in figure S9. Despite L2 with a size of 55 kDa, reducing SDS-PAGE demonstrates the typical running behavior of L2 with a prominent band at 75 kDa and a less prominent band at ~ 40 kDa [25, 26, 130, 131]. We obtained as measured densometrically via ImageJ a protein purity of > 95 % for the His_6-TEV-GGGL2 protein and 91 % for the His_6-TEV-GGGL2ΔL1bs protein and the tyrosine-based UV-absorption measurement showed a
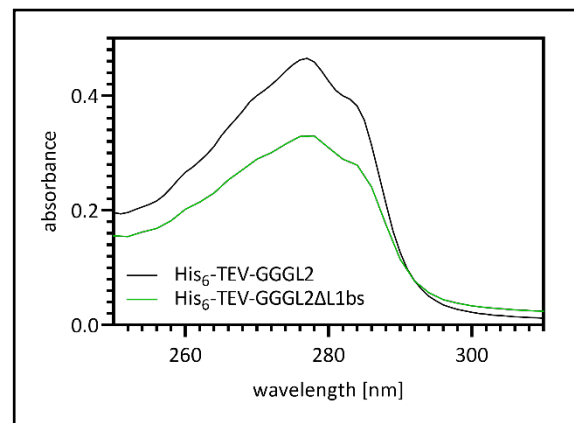
Figure 24: Characterization of denatured and purified His_6-TEV-GGGL2 constructs. UV/VIS spectra for the L2 constructs**:** (black line) His_6-TEV-GGGL2 construct and (green line) His_6-TEV-GGGL2ΔL1bs measured in solubilization buffer (6 M guanidine hydrochloride, 20 mM Tris • HCl pH 7.4 at 8°C). The ratio A_{260}:A_{280} = ~ 0.5 indicated a protein sample almost free of DNA. According to the Lambert-Beer-law the protein concentration was calculated with the A_{280nm} and extinction coefficient (L*mol^{-1}*cm^{-1}).

maximum absorbance at 274 nm (figure 24) for denatured L2 proteins. Corresponding to the 260 and 280 nm (A_{260}:A_{280}) absorbance with a ratio of ~ 0.5 the protein is free of nucleic acids. The purified solubilized L2 proteins were stored at 4°C and further used for characterization due to stability analysis via refolding studies and temperature transmission measurements, secondary structure analysis per circular dichroism and N-terminal TEV-tag cleavage and labeling/biotinylation.

3.7   L2 constructs – cleavage and labeling/biotinylation

3.7.1   TEV-tag cleavage and separation

The L2 constructs expressed as inclusion bodies were purified under denaturing conditions (6 M guanidine hydrochloride, 20 mM Tris • HCl pH 7.4 @ 8°C) using affinity chromatography (NiNTA) as described in section 3.6. The L2 constructs are N-terminal tagged with a His_6-TEV tag. This TEV-tag is specifically cleaved by the TEV protease [132] and a GGG motif is thereby released N-terminally. The cleavage scheme for the His_6-TEV-tag for the L2 constructs is shown in figure 25 A. To cleave the L2 constructs, the solubilized constructs were refolded in A0 buffer (200 mM NaCl, 50 mM Tris • HCl pH 7.4 @ 8 °C, 5 % (w/v) Glycerol) containing 0.5 M L-arginine. Afterwards the His_6-TEV-tag is cleaved off via TEV-Protease (figure 25 B, C) and this cleavage exposes a glycine rich N-terminal region. For His_6-TEV-tag cleavage, the His_6-TEV-GGGL2 and His_6-TEV-GGGL2ΔL1bs constructs were incubated with a 1:5 (TEVProtease : L2 protein) ratio because previously experiments for enzyme activity described best cleavage results for this ratio (described by Christine Zarges) (figure
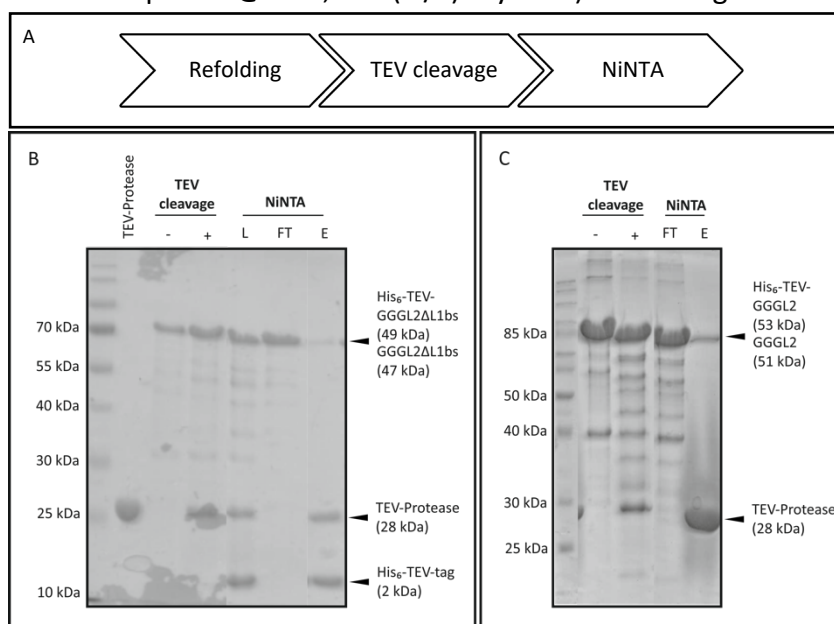
Figure 25: Cleavage reaction of TEV-Protease for refolded TEV-tagged L2 constructs in A0 buffer containing 0.5 M L-arginine. **(A)** scheme of cleavage reaction and purification. **(B)** cleavage and purification process of the refolded His_6-TEV-GGGL2ΔL1bs (49 kDa) construct. TEV-Protease: purified Protease; TEV cleavage: reaction before (-) and after (+) TEV-cleavage o/n with a mass ratio 1:5 at 8 °C (protease:protein); NiNTA: Load (L) of cleaved sample, cleaved GGGL2ΔL1bs protein (47 kDa) in flow-through (FT), elution of TEV-Protease and cleaved TEV-tag (E) analyzed via Coomassie stained SDS-PAGE. **(C)** cleavage and purification process of the refolded His_6-TEV-GGGL2 (53 kDa) construct. TEV cleavage: cleavage reaction before (-) cleavage and after (+) cleavage o/n at 8 °C. NiNTA: the cleaved GGGL2 (51 kDa) protein is shown in the flow through (FT) and the elution of TEV-Protease and cleaved TEV-tag in E. All in all, the TEV cleavage of the TEV tag and the subsequent purification was successful for both L2 constructs.

25 B and C; TEV cleavage line -and +). Separation of His$_6$-TEV-tag, His$_6$-TEV protease, and cleaved L2 protein was performed via a NiNTA (figure 25 B; NiNTA). The His$_6$-TEV-tag and the His$_6$-TEV-protease remain attached to the column due to the His$_6$-tag and the cleaved L2 protein is localized in the flow-through (figure 25 B, C line FT). The successful cleavage for the full-length L2 construct is demonstrated in figure 25 C and for the L2ΔL1b construct in figure 25 B. The cleaved and separated L2/L2ΔL1bs protein is displayed in figure 25 B and C; NiNTA line FT. The cleaved and refolded protein can now either be used for labeling and biotinylation or can be denatured via dialyzing against H$_2$O and solubilization in 6 M guanidine hydrochloride, 20 mM Tris • HCl pH 7.4 @ 8°C and afterwards stored at 4 °C.

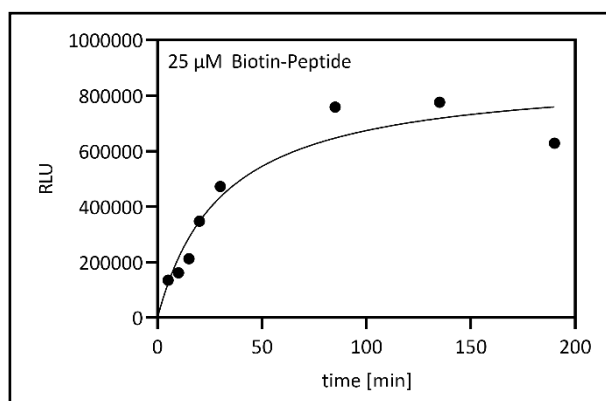### 3.7.2 Labeling/biotinylation and separation

The released GGG motif is necessary to label the L2 constructs with a fluorescein- or biotin-tag N-terminally. The labeling is carried out via *Sortase A* reaction and for this purpose, a recombinant protein with a Gn-tag at the N-terminus must be produced [111]. The second ligation partner, e.g., fluorescein peptide, must have a LPxTG recognition motif at the C-terminus [92, 109]. The coupling of the fluorescein or biotin N-terminally should not interfere with the interaction if the N-terminus is not involved in binding.

The cleaved L2 protein is incubated with *Sortase A* and fluorescein-LPETGGRR-peptide or biotin-LPETGGRR-peptide to label or biotinylate the protein. To achieve an optimum biotinylation with the biotin-LPETGGRR-peptide, a small-scale kinetics was performed first as shown in figure 26. Therefore, biotinylation was followed over time via Streptavidin conjugated AP reaction (section 2.2.21) and analyzed densitometrically via ImageJ. For curve fitting evaluation the Equation: One site – Specific binding (GraphPad) was used:

$$Y = \frac{B_{max} * X}{(K_d + X)}$$

This fit demonstrates a completed *Sortase A* reaction at ~100 min for the biotin-LPETGGRR-peptide. For the fluorescein-

Figure 26: Small scale biotinylation kinetic for GGGL2 construct with *Sortase A* in A0 buffer containing 0.5 M L-arginine. 20 μM GGGL2 were incubated with 25 μM biotin-Peptide and 4 μM *Sortase A* and samples were taken over time. The biotinylation of GGGL2 was displayed via Streptavidin conjugated AP and densometrically analyzed by ImageJ.

LPETGGRR-peptide this optimization has already been done by Christine Zarges. The labeling/biotinylation scheme for the L2 constructs is demonstrated in figure 27 A. To label and

biotinylate the refolded L2 proteins (A0 buffer containing 0.5 M L-arginine), *Sortase A* in a 5:1 (L2 : *Sortase A*) ratio was used for 3 h at 8°C. Labeling and biotinylation of the L2 constructs was followed by separation of labeled/biotinylated L2 and free peptide via size exclusion. The Superose® 6 10/300 column was used for the fluorescein labelled L2 (figure 27 B) and the HiLoad 16/600 Superdex® 200 PG column (figure 27 C) was used for the biotinylated L2 constructs. The SEC chromatogram in figure 27 B for the fluorescein labeled L2 (FL-L2) shows protein at 280 nm absorbance and fluorescein at 499 nm. The peak I at 16 ml shows an increase of the 280 nm and 499 nm signal. The peak II at 19 ml demonstrates only an increase
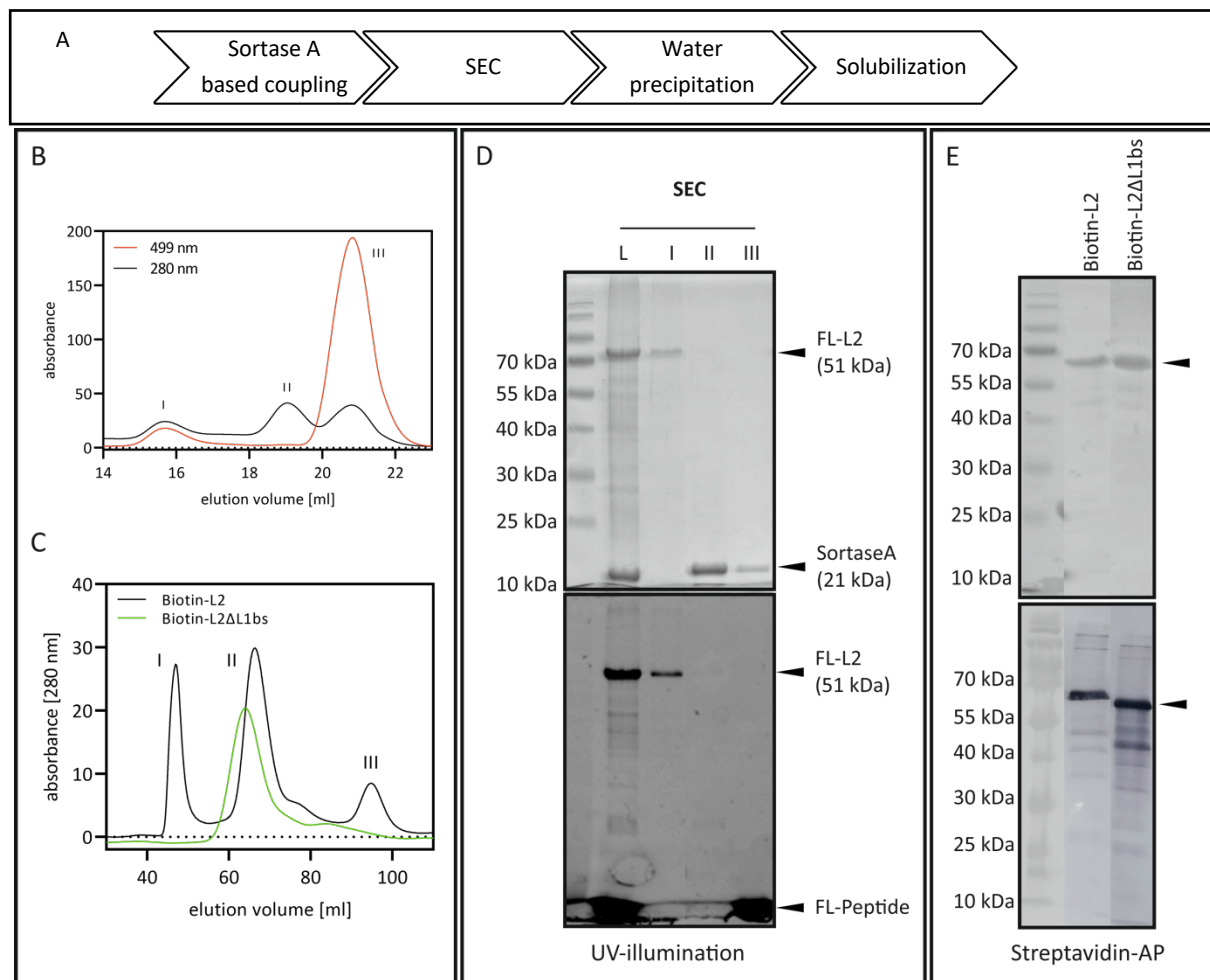
Figure 27: *Sortase A* **based labeling for refolded L2 constructs with biotin-LPETGGRR-peptide and fluorescein-LPETGGRR-peptide. (A)** scheme for labeling. **(B)** Size exclusion chromatogram (Superose® 6 10/300, CV: 24 ml, sample load: 185 µl, flowrate: 0.5 ml/min) for FL-L2 protein: 280 nm (black line) displays protein and 499 nm (red line) fluorescein. Peak I (I) at 15.8 ml shows a similar signal for protein and fluorescein. Peak II (II) at ~19 ml displays just a protein signal. Peak III (III) at 21 ml shows a high fluorescein signal and a little protein signal. **(C)** Size exclusion chromatogram (HiLoad 16/600 Superdex® 200 PG column, CV: 120 ml, sample load: 5 ml, flowrate: 1 ml/min) for biotin-L2 (Peak II: black: V_E = ~ 67ml) and biotin-L2ΔL1bs (Peak II: green: V_E = ~ 64ml). Peak I of biotin-L2 run demonstrates aggregates and Peak III *Sortase A*. **(D)** SDS-PAGE from SEC (Superose 6 10/30) purification process for fluorescein labeled L2. Coomassie stained PAGE in upper part and UV-illumination in bottom part. Load of labeled and unseparated FL-L2 (L); first elution peak (I): Both the SDS PAGE Coomassie stained and the UV illuminated results show a band above 70 kDa. This running behavior is unique for L2; Line II (II) displays elution peak II and the Coomassie stained SDS-PAGE indicates a band at ~15 kDa, this size fits to the *Sortase A*. No signal is shown in UV-illumination; In Line III (III) the third elution peak shows free fluorescein-peptide and little amount of *Sortase A*. **(E):** Coomassie stained SDS PAGE and Western Blot with Streptavidin-AP conjugation from final SEC (HiLoad 16/600 Superdex® 200 PG column) purification step. For biotin-L2 and biotin-L2ΔL1bs a prominent band at 70 kDa is visible in Coomassie stained SDS-PAGE and Western Blot incubated with Streptavidin-AP conjugation.

in the 280 nm signal and the peak III at 21 ml displays only an increase in the 499 nm signal. All three peaks were analyzed by SDS-PAGE. The SDS-PAGE was first analyzed via UV illumination and afterwards Coomassie stained (figure 27 D). For the first peak a clear band of > 70 kDa is visible under UV-illumination as well as in the Coomassie stained SDS-PAGE. This result fits to the characteristic running behavior of the L2 protein [23, 25, 26]. In addition, the elution volume of 16 ml fits L2 protein.
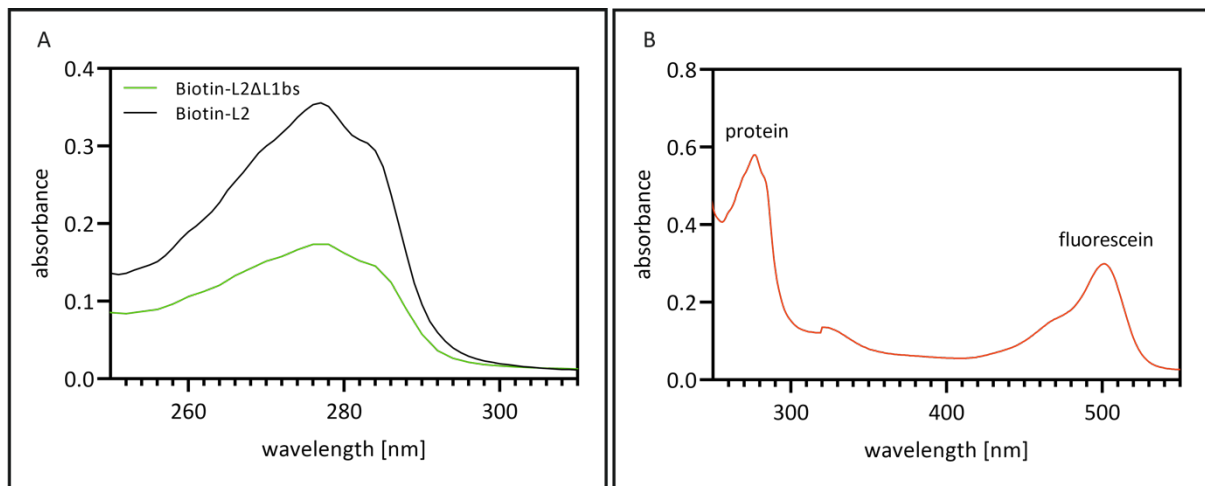
Figure 28: UV/VIS spectra for labeled L2 constructs. **(A)** UV/VIS Spectra from denatured biotin-L2 (black line) and denatured biotin-L2ΔL1bs (green line). Measurement was performed in solubilization buffer (6 M guanidine hydrochloride, 50 mM Tris • HCl pH 7.4 @ 8 °C). **(B)** Spectra for denatured fluorescein labeled L2 protein (FL-L2) in solubilization buffer. The aromatic amino acids of the proteins and of the coupled fluorophore fluorescein absorb in the UV-range (Maximum at 274 nm) and in the visible range (Maximum at 499 nm), respectively. Based on the absorbance value for 280 nm and 499 nm and the extinction coefficient the protein concentration as well as the degree of labelling (DoL) can be determined as demonstrated in table 11.

By the UV-illumination a band at > 70 kDa was visible. We can assume that the L2 is labeled with the fluorescein-LPETGGR-peptide. The second peak corresponds to the *Sortase A* enzyme (figure 27 D line II) and the third peak shows the free fluorescein-LPETGGR-peptide. According to this procedure L2 could be successfully labeled with fluorescein-LPETGGRR-peptide. The final labeled product is demonstrated in figure 27 D SEC; Line I and there is no free fluorescein-LPETGGRR-peptide visible as well as *Sortase A*. We determined the degree of labeling (DoL) based on the UV/VIS absorption values (shown in figure 28 B) of 280 nm and 499 nm and their corresponding extinction coefficients (table 11). The theoretical extinction coefficient of L2 protein was calculated based on the amino acid sequence via ProtParam tool ExPASy. In contrast, the extinction coefficient for fluorescein was measured and calculated using different concentrations of free fluorescein-LPETGGRR-peptide (done by Christine Zarges).

Table 11: extinction coefficient and measured absorbance (via UV/VIS spectra) of L2 protein and labelled L2 protein with fluorescein-LPETGGRR-peptide

	extinction coefficient	absorbance
L2 protein (280 nm)	32908 M^{-1} cm^{-1}	280 nm = 0.6
fluorescein-LPETGGRR-peptide (499 nm)	22962 M^{-1} cm^{-1}	499 nm = 0.3

$$DoL \ \% = \frac{\frac{0.3}{22962 \ L*mol^{-1}*cm^{-1}}}{\frac{0.6}{32908 \ L*mol^{-1}*cm^{-1}}} * 100 = 72\%$$

As already mentioned, we biotinylated the TEV-cleaved GGGL2 and GGGL2ΔL1bs constructs. Therefore, the *Sortase A* reaction was performed in a 5:1 (L2 : *Sortase A*) ratio for 3 h at 8°C. A preparative HiLoad 16/600 Superdex® 200 PG column was used for the separation of biotin-LPETGGR-labeled L2 and unlabeled biotin-LPETGGRR-peptide. The first peak in the Biotin-L2 run (figure 27 C black line) demonstrates probably protein aggregates due to the light scattering signal at 320 nm (data not shown) and the elution volume of 45 ml. In addition, the Coomassie stained reducing SDS PAGE and the Western Blot analysis via Streptavidin-AP conjugated assay (data not shown) showed a prominent biotinylated protein band at ~ 70 kDa. The peak II at approx. 67 ml for the Biotin-L2 shows a signal at 280 nm (figure 27 C, black line). For the Biotin-L2ΔL1bs the protein elutes at 64 ml (figure 27 C, green line) and the elution peaks for the two L2 constructs were analyzed via a Coomassie stained reducing SDS-PAGE and a Streptavidin-AP conjugated assay displayed in figure 27 E. A distinct band at 70 kDa can be seen in the Coomassie stained reducing SDS PAGE and the Streptavidin-AP conjugated assay for both elution peaks. This indicates the presence of biotinylated-L2 in both samples. The third peak of the Biotin-L2 protein elutes at ~ 95 ml and corresponds to *Sortase A* (data not shown).

The degree of biotinylation (DoB) was determined with a hold up assay. Thereby, biotinylated L2 was incubated on a streptavidin-matrix in a filter tube and centrifuged. Therefore, it is important that the saturation range of the streptavidin-matrix is not exceeded. This was tested via a dilution series. The protein concentration of the input and flow-through were analyzed via a UV/VIS spectrum due to 280 nm absorption and the DoB was calculated as described in section 2.2.20:

$$DoB \ for \ Biotin - GGGL2 \ \% = 100\% - \left(\frac{0.01}{0.13} * 100\right) = 92\%$$

$$DoB \ for \ Biotin - GGGL2\Delta\text{L1bs} \ \% = 100\% - \left(\frac{0.13}{0.28} * 100\right) = 46\%$$

For the His_6-TEV cleavage 10 mg His_6-TEV-GGGL2 and His_6-TEV-GGGL2ΔL1bs protein was used. After cleavage, labeling/biotinylation and separation 3-4 mg (table 12) final product was obtained. In summary for both reactions, labeling as well as biotinylation and separation of educt and product was successful. However, 60-70 % of the L2 protein used was lost via the cleavage, labeling and biotinylation process. The degree of biotinylation and the labeling efficiency for the different GGGL2 constructs are important for further experiments (table 12).

Table 12: labeling and biotinylation efficiency for L2 constructs and yield of labeled and biotinylated L2 proteins.

Construct	yield labeled/biotinylated protein	degree of labeling or biotinylation
FL-L2	4 mg	72 %
Biotin-L2	3.5 mg	92 %
Biotin-L2ΔL1bs	3 mg	46 %

3.8 Characterization of GGGL2 construct

3.8.1 Stability for GGGL2

For further experiments, the denatured GGGL2 had to be refolded. This refolding is necessary to continue with functional L2 protein because denatured proteins are normally not functional. This refolding process is already described for L2 in A0 buffer containing 0.5 M L-arginine and was performed via rapid dilution [77]. Therefore, L2 can be refolded without significant amounts of precipitate. Furthermore, Breiner et. al. analyzed the refolded protein in A0 buffer containing 0.5 M L-arginine due to stability over time, secondary structure, size, DNA binding and membrane interaction studies. Taken together, these results indicate that the refolded L2 is functional. Arginine is known for its protein stabilizing effect, but this can also lead to the prevention of protein-protein interaction [133-135]. To overcome the problem that arginine can prevent protein-protein interaction, we have tested another detergent
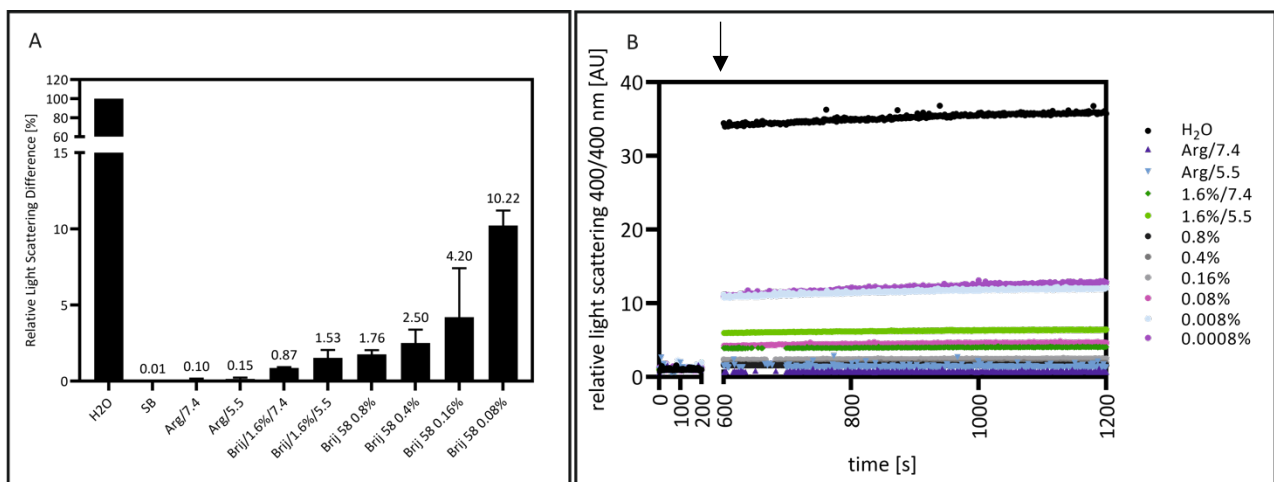
Figure 29: Refolding screen of 1µM solubilized and reduced GGGL2 in different buffers at 8°C via rapid dilution. The experiments were performed in triplicates. **(A)** As a signal of aggregation, the light scattering (90-degree ancle) at 400 nm was monitored over time in a 10 mm cuvette. The higher the light scattering appears the more inefficient is the refolding of L2 protein. After 1200 s of protein addition the samples reached a plateau phase in light scattering. To calculate the relative light scattering, the first step is to average the buffer signal from 0-200 s and then normalize it. Subsequently, the values of 200-1200 s were relativized against the normalized buffer value (shown in B). For A values, the difference from the relativized buffer value of 0s and 1200s was calculated and plotted against the positive control H_2O (values shown in A). Subsequently, the values of 200-1200s were relativized against the normalized buffer value (shown in B). For A values, the difference from the relativized buffer value of 0s and 1200s was calculated and plotted against the positive control H_2O (shown in A). H_2O sample displays complete protein aggregation. In SB buffer (6 M guanidine hydrochloride, 20 mM Tris • HCl pH 7.4 @ 8 °C) no aggregation occurs. A0 buffer pH 7.4 with 0.5 M L-arginine shows lowest L2 aggregation. With decreasing pH to 5.5 aggregation increases a little. Additionally, different concentrations of Brij 58 buffer (1.6 % (w/v)-0.08 % (w/v)) were tested. With decreasing Brij concentration the light scattering increases which indicates protein aggregation. **(B) Refolding kinetic** over time. From 0 – 200 s relative buffer signal is displayed from different buffers. From 600 – 1200 s the light scattering at 400 nm from refolded protein is shown.

called Brij 58 in different concentrations for protein stabilization [136]. For this purpose, solubilized protein was refolded in different refolding buffers by rapid dilution (1 µM GGGL2 protein) and this reaction was monitored via light scattering to analyze aggregation (figure 29). Two controls are needed to evaluate the aggregation of L2 in different buffers. The negative control of 6 M guanidine hydrochloride displays no light scattering which leads to the
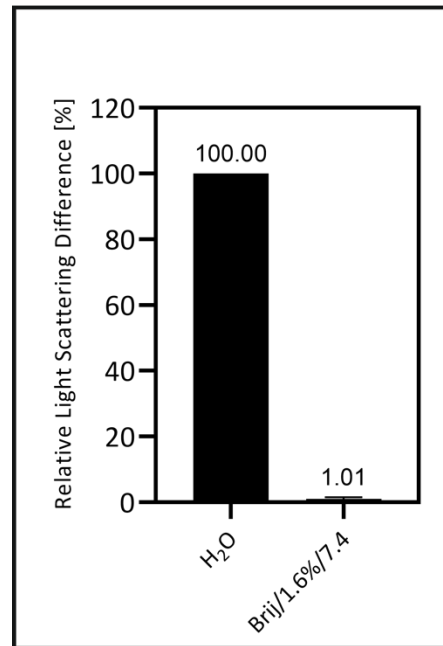
assumption that L2 is not aggregating during rapid dilution because no refolding is taking place. On the contrary, the positive control H_2O displays a high signal of light scattering which indicates a complete aggregation of L2 during the refolding process as already published [77]. In addition to the previously published buffer conditions (200 mM NaCl, 50 mM Tris • HCl pH 7.4 @ 8 °C, 5 % (w/v) Glycerol and 0.5 M L-arginine), we tested A0 buffer containing 0.5 M L-arginine at pH 5.5 and various concentrations of Brij 58 % (w/v) at different pH values. Thereby, A0 buffer pH 5.5 containing 0.5 M L-arginine showed no light scattering and thus no difference to A0 buffer pH 7.4 with 0.5 M L-arginine (figure 29 A). For the different Brij 58 concentrations (0.08 % (w/v)-1.6 % (w/v)) we observe, the higher the concentration of Brij 58, the less light scattering is visible. It is important to note that we determined the critical micelle concentration under the applied conditions, which is 0.08 % (w/v) Brij 58, (200 mM NaCl, 50 mM Tris • HCl pH 7.4 @ 8 °C, 5 % (w/v) Glycerol) (figure S 10). This result is in line with the literature data [137]. In figure 29 B the refolding kinetic for the different samples is demonstrated. The first 200 s of the measurement display the light scattering of the buffer signal and the arrow at 600 s indicates the addition of protein and the associated increase in light scattering relative to the buffer signal at the first 200 s. To reach a plateau phase the samples were incubated for 1200 s. The results clearly demonstrate the highest relative light scattering at 400 nm with ~ 35 in water (black dots) after protein addition. It is also visible here that the relative light scattering at 400 nm increases with decreasing Brij 58 concentration.

The results indicate that a Brij 58 concertation of at least 1.6 % (w/v) is required to prevent aggregation of 1 µM L2. For this reason, following experiments with refolded L2 were performed using A0 buffer containing 1.6 % (w/v) Brij 58 to stabilize L2.

### 3.8.2 Influence of fluorescein tag on stability of FL-L2

For further experiments not only GGGL2 is used but also fluorescein labeled L2 (FL-L2). Therefore, it is important to know whether this fluorescein tag affects the stability of L2. To make sure that fluorescein labeled L2 can also be refolded and stabilized in A0 buffer pH 7.4 containing 1.6 % (w/v) Brij 58, FL-L2 was tested (figure 30) via rapid dilution in the same set up as GGGL2 (3.8.1). It is clearly visible that the light scattering of FL-L2 with 1.0 % is as low as the aggregation of unlabeled GGGL2 with ~ 0.9 % (figure 29 A). This result suggests that the fluorescein label has no effect on the refolding stability of FL-L2 in A0 buffer pH 7.4 containing 1.6 % (w/v) Brij 58 (figure 30). This result demonstrates no light scattering for FL-L2 in A0 buffer pH 7.4 containing 1.6 % (w/v) Brij 58 and therefore can be used for further experiments. The biotinylated L2 constructs were not tested in A0 buffer

Figure 30: Refolding screen of 1 µM solubilized and reduced FL-L2 in A0 buffer pH 7.4 containing 1.6 % (w/v) Brij 58 at 8 °C via rapid dilution. As a signal of aggregation, the light scattering at 400 nm was monitored over time for 1200 s. The values were calculated as described in figure 29. A0 buffer containing 1.6 % (w/v) Brij 58 was tested because refolding of unlabeled GGGL2 showed lowest light scattering signal.

containing 1.6 % (w/v) Brij 58 because these constructs were not used in this buffer for further experiments.

3.8.3 T_{trans} – temperature stability for GGGL2 and FL-L2

For further experiments it is important to know the stability for GGGL2 and labeled GGGL2 (FL-L2). To analyze the stability for GGGL2 in different buffers, the samples were denatured thermally. The stability measurements for His$_6$-L2 in CP buffer containing 0.5 M L-arginine buffer have already
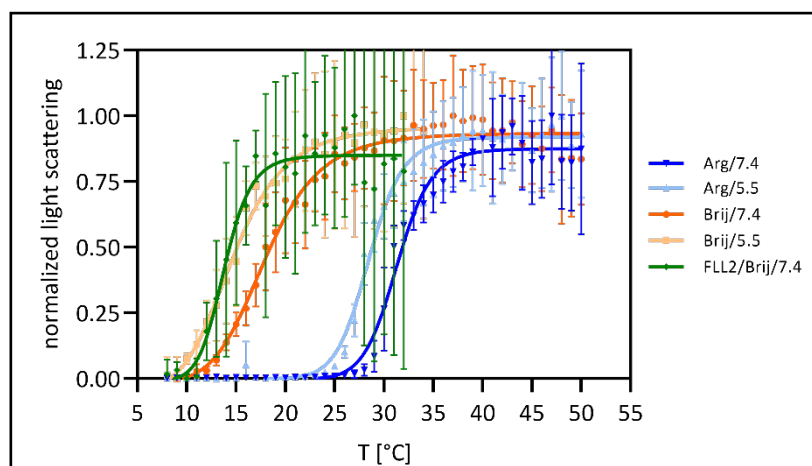
Figure 31: Stability measurement of refolded L2 by temperature induced aggregation. 1 µM of refolded GGGL2 or FL-L2 was heated in different CP buffers containing either 0.5 M L-arginine or 1.6 % (w/v) Brij 58 at pH 7.4 or 5.5. As a degree of aggregation, the light scattering at 275 nm was monitored at different temperatures.

been published [77]. To find out the stability of GGGL2 in CP buffer with either pH 5.5 (11 mM

Na_2HPO_4, 4.4 mM $C_6H_8O_7$, 200 mM NaCl, 5 % (w/v) Glycerol) or pH 7.4 (18 mM Na_2HPO_4, 1 mM $C_6H_8O_7$, 200 mM NaCl, 5 % (w/v) Glycerol) and additional 1.6 % (w/v) Brij 58 or 0.5 M L-arginine, samples were thermally denatured as displayed in figure 31. For fitting the data via GraphPad the [Agonist] vs. response – variable slope (four parameters) fit was used. The results show that CP buffer containing 0.5 M L-arginine with pH 7.4 prevents aggregation up to 32 °C. This fits to the already published data for His$_6$-L2 by Breiner et. al. [77]. For the CP buffer containing 0.5 M L-arginine at pH 5.5 aggregation occurs at ~ 28 °C. CP buffer containing 1.6 % (w/v) Brij 58 pH 7.4 prevents aggregation up till 18 °C and for pH 5.5 up to ~ 14 °C. Additionally, we tested FL-L2 in CP buffer containing 1.6 % (w/v) Brij 58 pH 7.4. Aggregation occurs at ~ 14 °C. This result indicates that FL-L2 is less stable in CP buffer with 1.6 % (w/v) Brij 58 pH 7.4 than unlabeled GGGL2. In table 13 the results for the different measurements and the associated confidence interval are demonstrated.

Table 13: T_M measurement results for different A0 buffers containing either arginine or Brij 58 and GGGL2 as well as FL-L2. Additionally, calculated confidence interval is provided for every measurement. T_M [°C] demonstrates the inflection point in degree.

	T_M [°C]	GGGL2 0.5 M L-arginine	GGGL2 1.6 % (w/v) Brij 58	FL-L2 1.6 % (w/v) Brij 58
pH	5.5	28.5 ± 0.5	14.7 ± 2.6	-
pH	7.4	31.4 ± 1.1	18.0 ± 1.0	13.8 ± 2.8

Furthermore, it can be shown that the stability of GGGL2 decreases in CP buffer containing 1.6 % (w/v) Brij 58 compared to 0.5 M L-arginine independently from the pH. Additionally, the pH value displays a higher stability for GGGL2 at pH 7.4 in CP buffer containing 1.6 % (w/v) Brij 58 and 0.5 M L-arginine versus a decreasing pH value (pH 5.5). These measurements demonstrate the strong stability dependence of L2 on the buffer composition (table 13). From these results, we can conclude that attention must be paid to whether buffer for L2 (A0 with arginine or Brij 58) is used. Depending on the buffer, the temperature must be adjusted to ensure the stability of L2. For this reason, all further measurements of L2 will be performed at 8 °C.

3.8.4 Circular Dichroism - Secondary structure analysis of GGGL2

The refolded L2 protein needs to be analyzed in respect of the secondary structure. This provides information about the refolding of L2 in A0 buffer pH 7.4 @ 8 °C with 1.6 % (w/v) Brij 58 and allows comparison to published and previous refolding conditions. To analyze the secondary structure of refolded soluble GGGL2 protein we performed a far-UV circular dichroism (CD)-spectroscopy. For refolded and denatured 16L2 without GGG motif
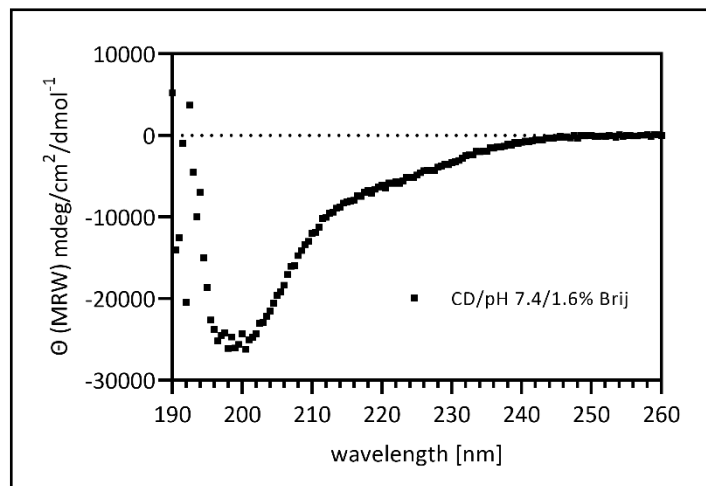
Figure 32: The result of the circular dichroism measurement for refolded GGGL2. Therefore, the far-UV circular dichroism spectra from 1 µM refolded GGGL2 in A0 buffer pH 7.4 containing 1.6 % (w/v) Brij was measured. To ensure linearity of the measurement, only the measured values below 600 V are considered. In this case from 260-195 nm. The secondary structure of GGGL2 protein over all is rather unstructured.

the secondary structure based on CD data is already published [77]. In figure 32 the CD-spectra of refolded GGGL2 protein in A0 buffer pH 7.4 containing 1.6 % (w/v) Brij 58 is shown. For the previously published A0 buffer containing 0.5 M DL-arginine, the refolded protein shows a minimum at 204 nm and an **Θ**MRW ~10.000 mdeg/cm^2/dmol^{-1} at 220 nm. Here, **Θ**MRW indicates molar ellipticity and for the calculation the average molecular weight of an amino acid was used [138]. The CD spectrum in A0 buffer pH 7.4 containing 1.6 % (w/v) Brij 58 shows a minimum at 200 nm and a signal of ~ 8.000 **Θ**MRW mdeg/cm^2/dmol^{-1} at 220 nm (figure 32). Thus, the spectra in A0 buffer containing DL-arginine and Brij 58 differ just slightly.

To ensure linearity in the measurement, only the results from 195-260 nm can be used for analysis, since below 195 nm the Volt signal increases above 600. For this reason, it is not possible to make an accurate prediction on the secondary structure of GGGL2 for A0 buffer containing 1.6 % (w/v) Brij 58 because firstly the information below <195 nm is missing and secondly the measured spectrum is an unusual shape with a minimum at 200 nm and a partial minimum at 220 nm. However, the results show that the secondary structure of refolded GGGL2 do not differ between A0 buffer containing arginine and Brij. For this reason, refolded GGGL2 can also be used in A0 buffer containing 1.6 % (w/v) Brij 58 for further experiments.

3.9 Characterization of purified L1nav construct

3.9.1 Purity of L1nav monomer and pentamer protein

The purity level of the purified protein plays a significant role in biochemical characterization and continuing interaction studies. Contaminants or impurities in the protein sample can distort or influence results. To avoid this, the purity of the target proteins must be investigated in terms of protein contamination and DNA contamination.
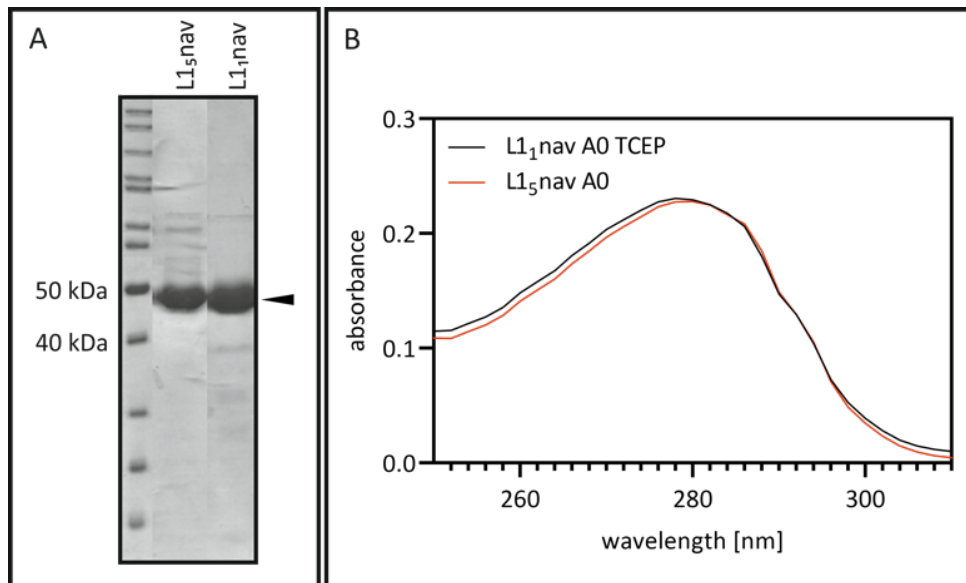
Figure 33: Characterization of purified L1nav protein. (A) Coomassie stained SDS-PAGE of 10 μg purified L1nav monomer (L1$_1$nav) and L1nav pentamer (L1$_5$nav). Densiometric analysis (performed with ImageJ) suggest a purity of 95 % for L1$_5$nav and 97 % for L1$_1$nav. **(B)** UV/VIS spectra of L1nav pentamer and L1nav monomer measured in A0 buffer (200 mM NaCl, 50 mM Tris • HCl pH 8 @ 8 °C, 5 % (w/v) Glycerol with or without 1 mM TCEP) with a 260:280 ratio of 0.5 indicating no DNA impurities. Final protein concentration was calculated according to the Lambert-Beer-law at 280 nm and an extinction coefficient of 63.425 L/mol/cm.

To determine the purity of final purified L1nav protein after tag removal a Coomassie stained reducing SDS-PAGE was performed and analyzed as shown in figure 33 A. Therefore, 10 μg of purified L1$_5$nav (L1nav pentamer) and L1$_1$nav (L1nav monomer) protein was loaded and for L1$_5$nav (figure 33 A line L1$_5$nav) as well as L1$_1$nav (figure 33 A line L1$_1$nav) a prominent protein band at ~ 50 kDa occurs. The size of the protein band fits to the calculated size of ~48 kDa for the L1nav protein. The purity for L1$_5$nav was calculated densometrically via ImageJ with > 95 % and for L1$_1$nav with > 97 % based on the result of the reducing SDS-PAGE demonstrated in figure 33 A. Furthermore, the proteins were tested due to DNA impurities. For this purpose, the UV/VIS spectra for L1nav monomer (black line figure 33 B in A0 buffer with 1 mM TCEP) and pentamer (red line figure 33 B in A0 buffer) was analyzed due to the ratio of the 260:280 nm absorption. The ratio of 0.5 (260:280) indicated no DNA impurities. The results

demonstrate a protein purity of over > 95% and no DNA contamination, allowing the proteins ($L1_1$nav/$L1_5$nav) to be used for further biochemical characterization and interaction studies.

3.9.2 Oligomeric state of L1nav monomer and pentamer protein

During protein purification of the L1nav construct, two prominent protein species were revealed. These have already been described in literature [103]. These two species appear to be His_6-SUMO-L1nav monomer and pentamer according to size exclusion chromatography analysis before tag cleavage and removal (section 3.4). For more detailed examination of the two protein species, the His_6-SUMO tag was removed (section 3.5) and protein purity was determined (section 3.9.1). To analyze the oligomeric state of the purified L1nav protein, a size exclusion chromatography (SEC) was performed under reducing and non-reducing conditions. These buffer conditions were chosen because disulfide bonds between cysteines do not form under reducing conditions and do under non-reducing conditions. These disulfide bonds may play a role in the stability of the individual protein species. The theoretical molecular weight for L1nav monomer is 48 kDa and for L1nav pentamer 240 kDa. To investigate the oligomeric state of L1nav monomer and pentamer a size exclusion chromatography on a Superose® 6 10/300 column was performed after dialyzing the different samples in reducing and non-reducing buffer over night at 4 °C (figure 34).
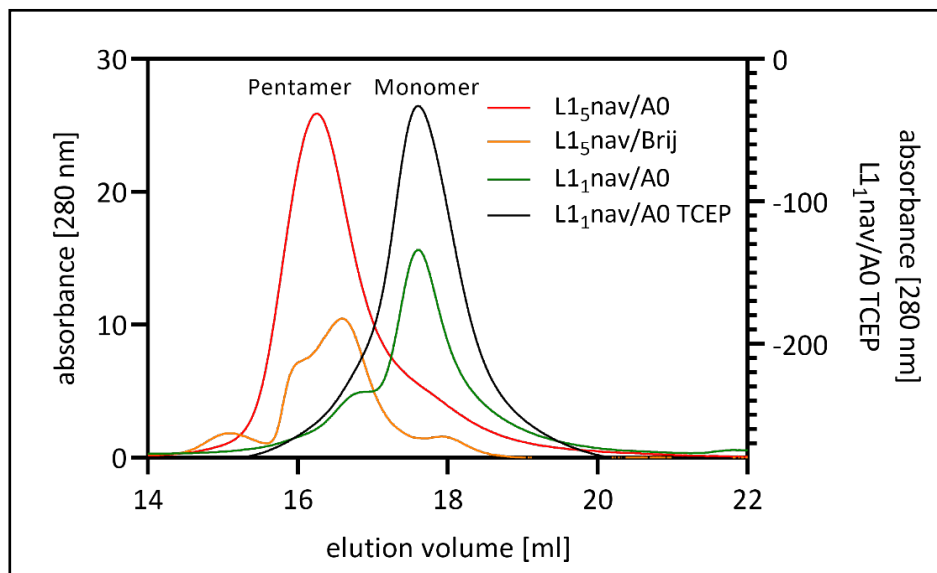
Figure 34: Size exclusion chromatography of L1nav pentamer and monomer at different buffer conditions. 185 µl sample with a protein concentration of 4.2 µM $L1_5$nav protein and 204 µM $L1_1$nav protein was load on a Superose® 6 10/300 column and the runs were performed with a flowrate of 0.5 ml/min. Under non reducing conditions (A0 buffer pH 8 @ 8°C) L1nav pentamer (red curve) elutes at 16.2 ml which corresponds to 209 kDa. L1nav monomer elutes at 17.9 ml which corresponds to 49 kDa and a shoulder is formed at 16.6 ml (green curve). Under reducing conditions (A0 buffer pH 8 @ 8°C + 1 mM TCEP) one elution peak at 17.9 ml occurs (black curve). For the A0 buffer at pH 8 @ 8 °C containing 1.6 % (w/v) Brij 58 the chromatogram looks not homogeneously (orange curve). One small peak at 15 ml and 17.9 ml is visible and one larger peak with two maxima is shown at 16 ml and 16.8 ml. This indicates different oligomeric states.

To represent reducing conditions, A0 buffer (200 mM NaCl, 50 mM Tris • HCl pH 8 @ 8 °C and 5 % (w/v) Glycerol) with 1 mM TCEP was used as dialyzing and running buffer. The result for the L1nav monomer displayed in Figure 34 black line shows an elution peak at 17.6 ml which correlates with a molecular weight (MW) of ~50 kDa. Under non-reducing conditions (figure 34 green line), an elution peak at 17.6 ml and a shoulder at 16.2 ml are shown. The elution volume of 16.2 ml correlates with a MW of 209 kDa. For the L1nav pentamer run an elution peak at 16.2 ml occurs under non-reducing conditions (figure 34 red line) which also correlates with a MW of 209 kDa. The size exclusion chromatography indicates a monomeric form of L1nav under reducing conditions and a pentamer form for L1nav under non-reducing conditions.

For continued interaction studies of L1nav and L2, it is important that L1nav is present as a pentamer. The characterization of L2 has shown that it is stable in A0 buffer containing 1.6 % (w/v) Brij 58. For this reason, we have studied the running behavior of L1nav pentamer also in A0 buffer containing 1.6 % (w/v) Brij 58 to be sure that the running behavior of L1 does not change and we can therefore assume that structurally nothing changes for L1 due to the micelles in the Brij buffer. The L1nav pentamer in A0 buffer pH 8 @ 8 °C containing 1.6 % (w/v) Brij 58 demonstrates (figure 34 organe line) a more heterogeneous elution chromatogram compared to the A0 buffer without Brij 58. For this reason, the L1$_5$nav protein used in A0 buffer containing 1.6 % (w/v) Brij 58 was examined via size exclusion before each experiment and the peak at 16.5 ml was pooled and used for further experiments.

From these results we can conclude that L1nav monomer as well as L1nav pentamer can be separated and the buffer condition plays a major role for the oligomeric state of the L1nav protein. To further investigate the size of the different oligomeric states of L1nav we performed an analytical ultracentrifugation displayed in section 3.9.3.

3.9.3 Size determination and structure analysis of L1nav pentamer under different buffer conditions via AUC/TEM

3.9.3.1 AUC measurement under reducing and non-reducing conditions

The characterization of the purified L1nav protein includes the determination of the protein size. L1nav protein occurs in different oligomeric states under reducing and non-reducing conditions as demonstrated in section 3.9.2. To determine the protein size of L1nav pentamer under different buffer conditions, an analytical ultracentrifugation was performed. The analytical ultracentrifugation for L1nav protein consists of two different measurements.
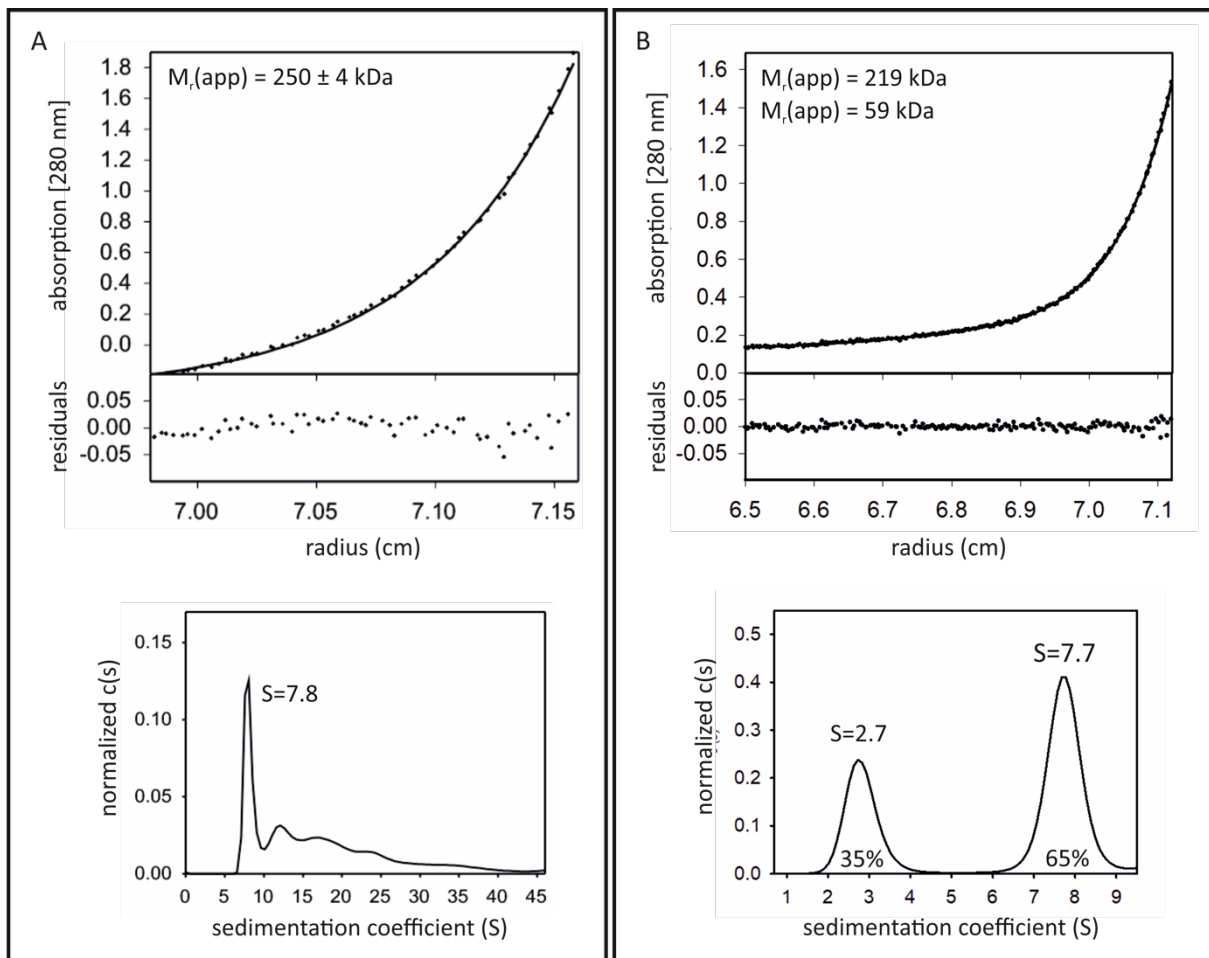
Figure 35: Oligomerization of purified L1nav protein under non-reducing (A) and reducing (B) conditions analyzed via analytical ultracentrifugation. Top graph: sedimentation equilibrium A280 measured at 8.000 g until equilibrium was reached. The distribution of the protein in equilibrium was measured at 280 nm (closed circles) and fitted according to the protein species. The lower panel shows the residuals. Bottom graph: sedimentation velocity measurement at 40.000 rpm and 8 °C. The absorption at 280 nm was measured every 10 min and the experimental data were analyzed via SEDFIT. The calculated c (s) distribution was plotted as a function of the sedimentation coefficient (S). **(A)** 0.38 mg/ml L1nav pentamer in A0 buffer (A0: 200 mM NaCl, 50 mM Tris•HCl pH 7.4 @ 8 °C, 5 % Glycerol). The size of the protein is calculated as MW$_{app}$ = 240 ± 4 kDa. The major species is shown with a s value of 7.8. The measured data are present in figure S12 **(B)** 0.38 mg/ml L1nav in A0 buffer with 1 mM TCEP measured. The size of the protein is calculated as MW$_{app}$ = 219 kDa and 59 kDa. The major species is shown with a s value of 7.7 and 2.7. The measured data are present in figure S12.

The sedimentation equilibrium measurement for the L1nav pentamer protein under non-reducing conditions (A0 buffer: 200 nM NaCl, 50 mM Tris • HCl pH 8 @ 8 °C, 5 % (w/v) Glycerol) indicates a MW of 250 kDa ± 4 kDa (figure 35 A top graph). Additionally, the analysis
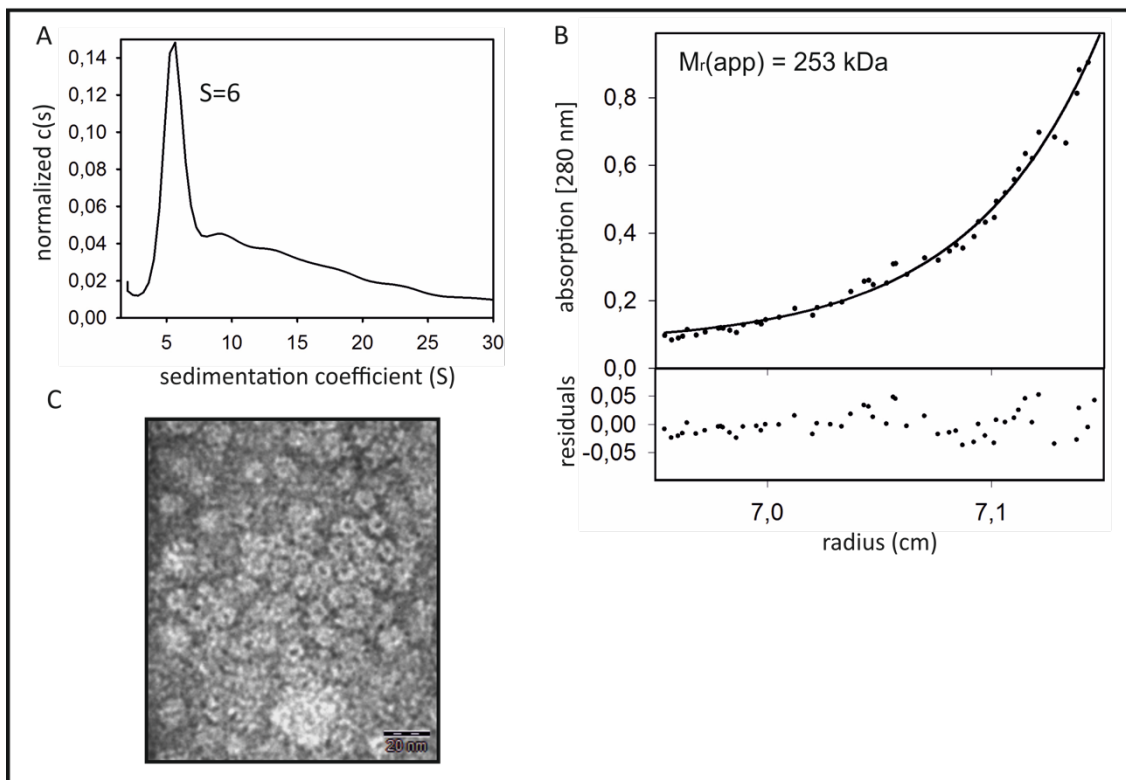
of the sedimentation velocity (AUC-SV) (figure 35 A bottom graph) reveals a sedimentation coefficient of s = 7.8 which aligns to the result of the homologous polyomavirus capsid protein VP1 pentamer with similar pentameric structure and molecular weight [5, 17] indicating pentamer formation for L1nav protein under non-reducing conditions. During L1nav protein purification two different protein species were demonstrated in the size exclusion (section 3.4) under reducing condition (A0 buffer containing 1 mM TCEP). Therefore, L1nav pentamer sample after tag removal under reducing conditions was measured via AUC. The sedimentation velocity (AUC-SV) measurement displays a 35 % species (s = 2.7) and a 65 % species (s = 7.8). Here, the analysis of the sedimentation equilibrium demonstrated two species with a size of 219 kDa and 59 kDa. The sizes of the two protein species suggest a pentamer species and a monomer species under reducing condition for the L1nav protein. Under non-reducing conditions, the data demonstrate a major species with a size of 250 kDa and this result fit the assumed size of the L1nav pentamer. Over all these results are in line with the assumed size of L1nav pentamer and monomer. Furthermore, the L1nav monomeric pentameric distribution of 65 % and 35 % reveals the same result as in section 3.4.

3.9.3.2   AUC measurement in A0 buffer with 1.6 % (w/v) Brij 58 and TEM analysis

As already demonstrated in section 3.9.3.1 is the analytical ultracentrifugation a proper method to characterize particle properties such as molecular weight (MW). For further experiments it is important to investigate the MW of the L1nav pentamer as well as the oligomeric state under different buffer conditions. In section 3.9.3.1. we already investigated the molecular weight of L1nav pentamer under reducing and non-reducing conditions. However, the measurement of L1nav pentamer under non-reducing conditions with additional 1.6 % (w/v) Brij 58 is still missing. This measurement is important since it was shown in section 3.9.2 that L1nav pentamer in A0 buffer containing 1.6 % (w/v) Brij 58 has a heterogeneous elution profile due to the size exclusion chromatography.

Figure 36: Condition of purified L1nav protein under non-reducing conditions (A0 buffer containing 1.6% (w/v) Brij 58) analyzed via analytical ultracentrifugation. (A) sedimentation velocity at 40.000 rpm and 8 °C with a L1nav pentamer concentration of 0.38 mg/ml in A0 buffer (A0: 200 mM NaCl, 50 mM Tris • HCl pH 7.4 @ 8 °C, 5 % (w/v) Glycerol) containing 1.6 % (w/v) Brij 58 is shown. The absorption at 280 nm was measured every 10 min and the experimental data were analyzed via SEDFIT. The calculated c (s) distribution was plotted as a function of the sedimentation coefficient. The major species is shown with a s value of 6. **(B)** sedimentation equilibrium of 0.38 mg/ml L1nav pentamer in A0 buffer containing 1.6 % (w/v) Brij 58 measured at 8.000 g until equilibrium was reached. The distribution of the protein was measured at 280 nm, visible in the closed circles. The size of the protein is calculated as MW_{app} = 253 kDa. The measured data are present in figure S12. **(C)** Transmission electron microscopy of 0.1 mg/ml negatively stained L1nav pentamer. The morphology indicates a pentameric species.

The results are demonstrated in figure 36 A and B. The sedimentation velocity shows a main peak at s = 6 under non reducing conditions with additional 1.6 % (w/v) Brij 58 (figure 36 A). The sedimentation equilibrium demonstrated in figure 36 B determine the molecular weight

of the protein with 253 kDa. These results indicate L1nav pentamer protein with a size of 253 kDa.
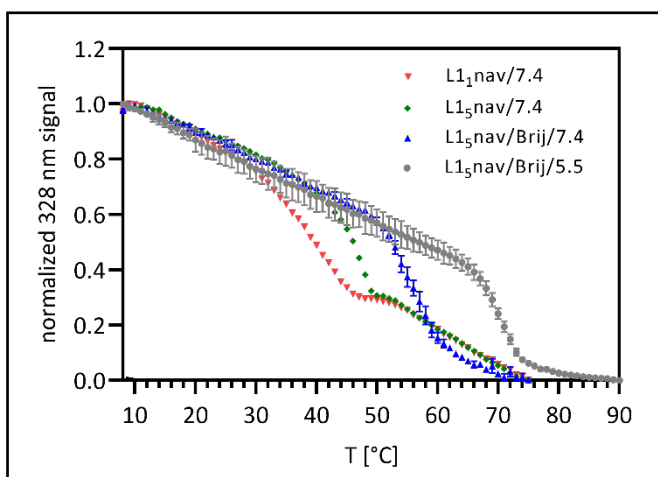
To investigate the pentamer morphology under non-reducing conditions, a transmission electron microscopy was performed. Therefore, 0.1 mg/ml L1nav pentamer in A0 buffer (200 mM NaCl, 50 mM Tris • HCl pH 8 @ 8 °C, 5 % (w/v) Glycerol) was negatively stained and analyzed by transmission electron microscopy. The result in figure 36 C demonstrate a pentameric species for the analyzed L1nav protein. All orthogonally used methods (AUC, TEM and SEC) indicate a pentamer species under non-reducing conditions for the molecular weight and morphology.

3.9.4 T_{trans} — temperature stability of L1nav monomer and pentamer

The purified L1nav protein will be used for various interaction studies. For this reason, it is important to know how stable L1nav monomer and pentamer are in different buffer compositions. The buffer conditions for the different oligomeric states of L1nav (monomer/pentamer) were selected according to the following methods (Surface plasmon resonance) for interaction studies. To investigate the temperature stability, we denatured the samples thermally and measured the intrinsic protein fluorescence at 328 nm under different buffer conditions. For the L1nav pentamer the non-reducing CP buffer with either pH 5.5 (11 mM Na_2HPO_4, 4.4 mM (Citrate) $C_6H_8O_7$, 200 mM NaCl, 5 % (w/v) Glycerol) or pH 7.4 (18 mM Na_2HPO_4, 1 mM $C_6H_8O_7$,

Figure 37: Stability measurement of L1nav pentamer ($L1_5nav$) and monomer ($L1_1nav$) by temperature induced aggregation. 0.05 mg/ml of L1nav monomer or pentamer was heated in different buffers (CP buffer pH 5.5: 11 mM Na_2HPO_4, 4.4 mM $C_6H_8O_7$, 200 mM NaCl, 5 % (w/v) Glycerol) (CP buffer pH 7.4: 18 mM Na_2HPO_4, 1 mM $C_6H_8O_7$, 200 mM NaCl, 5 % (w/v) Glycerol) containing additional 1.6 % (w/v) Brij 58. Intrinsic protein fluorescence at 328 nm is measured.

200 mM NaCl, 5 % (w/v) Glycerol) and additional 1.6 % (w/v) Brij 58 was used. We monitored the aggregation of thermally induced unfolding in figure 37. The evaluation was done as demonstrated and calculated in figure S 11. It is shown, that L1nav pentamer starts aggregating at ~ 45 °C in CP buffer pH 7.4 and the L1nav monomer at ~ 38 °C. For CP buffer pH 7.4 containing 1.6 % (w/v) Brij 58, L1nav pentamer starts aggregating at ~ 55 °C and the L1nav pentamer in CP buffer pH 5.5 containing 1.6 % (w/v) Brij 58 at ~ 69 °C. The results show that

the L1nav pentamer is slightly more stable than the L1nav monomer in CP buffer pH 7.4. Furthermore, there is 10 °C difference in stability with and without 1.6 % (w/v) Brij 58 in CP buffer pH 7.4 for the L1nav pentamer. The L1nav pentamer protein seems to be most stable in CP buffer pH 5.5 containing 1.6 % (w/v) Brij 58. The results for L1nav pentamer in CP buffer pH 5.5 and pH 7.4 were measured in experimental triplicates. As an overview all measured buffer conditions and results are demonstrated in table 14. We also included the calculated standard division as described in figure S 11.

Table 14: T_M [°C] results for different CP buffer conditions containing Brij and at different pH of L1nav monomer/pentamer with calculated standard division

		$L1_1$nav CP	$L1_5$nav CP	$L1_5$nav CP + 1.6 % (w/v) Brij 58
pH	5.5	-	-	69.7 ± 0.1
pH	7.4	38.2 ± 0.2	45.8 ± 0.2	55.5 ± 0.2

All in all, the results demonstrate a higher temperature stability for L1nav monomer/pentamer compared to L2 (table 13) in CP buffer pH 7.4 and 5.5 containing 1.6 % (w/v) Brij 58. Moreover, the results show that $L1_5$nav (L1nav pentamer) aggregates 7.6 °C later than $L1_1$nav (L1nav monomer) in CP buffer at pH 7.4 and therefore this result suggests more stability of L1nav pentamer. In addition, the measurement of L1nav pentamer in CP buffer pH 7.4 with and without 1.6 % (w/v) Brij 58 demonstrates a difference of 9.7 °C. This result indicates that L1nav pentamer is more stable in CP buffer pH 7.4 containing 1.6 % (w/v) Brij 58. In contrast, L1nav pentamer in CP buffer pH 5.5 containing 1.6 % (w/v) Brij 58 shows a difference of 14.2 °C compared to CP buffer pH 7.4 containing 1.6 % (w/v) Brij 58. This result suggests that L1nav pentamer is more stable at pH 5.5 in CP buffer containing 1.6 % (w/v) Brij 58. For all measurements the standard division is between 0.1 - 0.2 °C as demonstrated in table 14. The results indicate that further measurements and biochemical characterizations for L1nav monomer and pentamer at 8 °C are possible since the L1nav protein is stable at this temperature.

3.9.5 Circular Dichroism – Secondary structure analysis of L1nav

The protein secondary structure provides information about the three-dimensional form of local segments of the analyzed protein. To analyze the secondary structure of L1nav monomer and pentamer protein we performed a far-UV circular dichroism (CD)-spectroscopy. In figure 38 the different measured CD-spectra for L1nav pentamer and monomer in different A0 buffers (A0: 200 mM NaCl, 50 mM Tris • HCl pH 7.4 @ 8 °C, 5 % Glycerol) with pH 7.4 or 5.5 containing 0.16 % (w/v) Brij 58 are demonstrated.
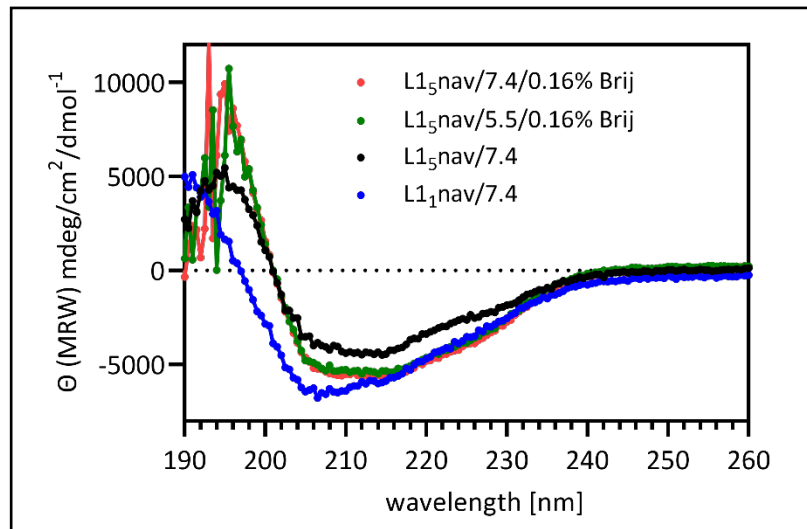
Figure 38: The secondary structure of L1nav monomer (L1$_1$nav) and L1nav pentamer (L1$_5$nav). Therefore, the far-UV circular dichroism spectra from 0.15 mg/ml L1nav monomer and pentamer in A0 buffer (A0: 200 mM NaCl, 50 mM Tris • HCl pH 7.4 @ 8 °C, 5 % Glycerol) with and without 0.16 % (w/v) Brij 58 was measured for 64 accumulations in a 1 mm cuvette. In blue L1nav monomer in A0+ 1 mM TCEP buffer at pH 7.4 is demonstrated. In red L1nav pentamer in A0 buffer at pH 7.4 containing 0.16 % (w/v) Brij 58 is shown. In green L1nav pentamer in A0 buffer at pH 5.5 containing 0.16% (w/v) Brij 58 is displayed and in black L1nav pentamer in A0 buffer pH 7.4 is demonstrated. To ensure linearity of the measurement, only the measured values below 600 V are considered. In this case from 260-190 nm.

In figure 38 all measured spectra for L1nav protein are demonstrated. The spectra for L1nav pentamer in different buffers show two minima at 215 nm and 206 nm with ΘMRW ~ 5.500 mdeg/cm^2/dmol^{-1} as well as a maximum at 196 nm. The results of the three L1nav pentamer measurements in different buffers demonstrate the same spectra although the spectra for L1nav pentamer in A0 buffer with pH 7.4 seems to be shifted upwards. This shift can be explained by a deviating protein concentration or the 0.16 % (w/v) Brij 58. Nevertheless, the shape of the three spectra is the same, indicating no change in secondary structure. For the L1nav monomer in A0 + 1 mM TCEP buffer pH 7.4 the spectrum shows a minimum at 205 nm with ΘMRW ~ 7.000 mdeg/cm^2/dmol^{-1} and a maximum at 195 nm with ΘMRW ~ 10.000 mdeg/cm^{-1}. This result demonstrates a different spectrum for L1nav

monomer than L1nav pentamer in A0 buffer pH 7.4 and suggests that there are structural differences between L1nav monomer and pentamer. These results are in line with already published data for L1 monomer and pentamer [103]. We can conclude from these results, that the secondary structure of L1nav pentamer does not change with decreasing pH and additional 0.16 % (w/v) Brij 58 detergent.

3.10 VLP assembly of full-length L1 constructs

So far, we have only characterized the L1 non-assembly variant (L1nav) for further interaction studies with L2. However, we also wanted to investigate the interaction of L1 full-length protein and L2 in virus-like particles. For this interaction studies we first needed a suitable amount of VLP´s. The assembly of HPV16L1 virus-like particles (VLPs) has already been published [19, 36]. Likewise, the assembly of VLP´s in polyomavirus is very well described and published [139, 140]. Based on these published assembling conditions, we prepared the
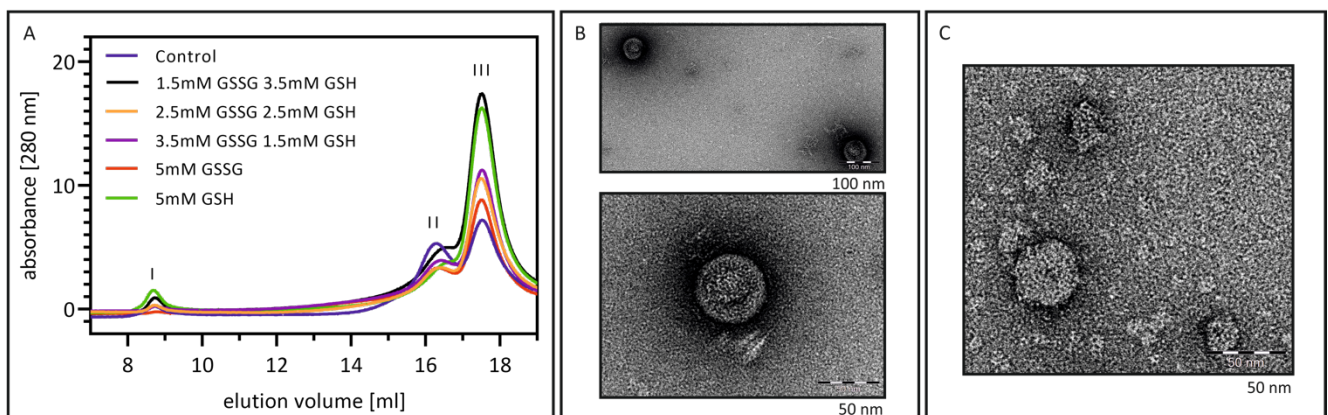
Figure 39: Assembly of HPV16L1 virus like particles in assembly buffer (AB: 11 mM Na$_2$HPO$_4$, 4.4 mM C$_6$H$_8$O$_7$, 1 M NaCl, pH 5.5) with an additional Redox-System. (A) Size exclusion chromatography (Superose® 6 10/300) of assembled samples to separate different assembly products and educts. Therefore, 185 µl sample was load and a flowrate 0.5 ml/min was applied. Peak I at ~ 8.7 ml demonstrate the HPV16 VLPs, Peak II at ~ 16.2 ml shows the pentameric species and Peak III at ~ 17.8 ml the monomeric species of HPV16L1 at different redox conditions. **(B)** Transmission electron microscopy of 0.1mg/ml negatively stained L1-His$_6$ VLPs (Peak I). **(C)** Transmission electron microscopy of 0.1 mg/ml negatively stained His$_{10}$-L1 VLPs.

appropriate assembling buffer for our needs (AB: 11 mM Na$_2$HPO$_4$, 4.4 mM C$_6$H$_8$O$_7$, 1 M NaCl, pH 5.5). Furthermore, the L1 monomer was used as the starting product for the assembly, as the monomer amount was higher after size exclusion chromatography than the pentamer amount (result size exclusion chromatography for L1 full-length protein not shown). In addition, the L1 pentamer concentration was too low for VLP´s assembly. For assembly, 1 mg/ml L1-His$_6$ monomer was dialyzed against the assembly buffer for 72 h at 8 °C with gentle stirring. An additional redox system with different GSSG and GSH (reduced and oxidized Glutathione) concentrations was also applied. The redox system can oxidize or reduce cysteines and for this reason serves as an assembly helper. To characterize the efficiency of

the different assembly conditions, a size exclusion chromatography was performed on a Superose® 6 10/300 and demonstrated in figure 39 A. The results of the individual size exclusions run all demonstrate three different peaks. In addition, the absorption of 280 nm for every peak is demonstrated in table 15.

Table 15: Results for 280 nm absorbance for the three different peaks of the size exclusion chromatography demonstrated in figure 39 for the different buffer conditions. Additionally, the calculated percentage distribution for every peak is listed.

	absorbance at 280 nm in mAU		
	8.7 ml [VLP´s]	16.2 ml [$L1_5$-His_6]	17.8 ml [$L1_1$-His_6]
Control	0.3 (2 %)	5.4 (42 %)	7.2 (56 %)
GSH:GSSG (5:0)	1.8 (8.3 %)	3.7 (17 %)	16.2 (74.7 %)
GSH:GSSG (3.5:1.5)	0.9 (4 %)	4.9 (21 %)	17.4 (75 %)
GSH:GSSG (2.5:2.5)	0.3 (2.1 %)	3.5 (24.5 %)	10.5 (73.4 %)
GSH:GSSG (1.5:3.5)	0.2 (1.5 %)	3.9 (25.3 %)	11.3 (73.2 %)
GSH:GSSG (0:5)	0.2 (1.6 %)	3.3 (26.8 %)	8.8 (71.6 %)

Peak III at 17.8 ml elution volume demonstrates 56 % - 75 % left over L1-His_6 monomer protein after assembly. Judging from the results of section 3.9.2, the elution volume fits L1nav monomer with a calculated MW of 50 kDa. This result indicates a semi successful assembly of L1-His_6 monomer to VLP´s. Peak II at elution volume of 16.2 ml indicates up to 42 % L1-His_6 pentamer species depending on the assembly buffer condition. Again, judging from the results of section 3.9.2, the elution volume matches L1nav pentamer with a calculated MW of ~ 209 kDa. Thereby the control, assembly buffer without GSSG and GSH, demonstrates the highest amount with 42 % of L1-His_6 pentamer species.

The peak I at 8.7 ml elution volume displays with high probability VLP´s. This can be assumed on the one hand because the small elution volume (8.7 ml) indicates large particles and on the other hand an absorption at 320 nm (data not shown) indicates light scattering. This light scattering signal suggests VLPs. The VLP yield after assembly of L1-His_6 monomers are rather low, with the highest assembly efficiency of 8 % in assembly buffer containing 5 mM GSH. These results are shown in figure 39 A as a green line and the efficiency was calculated in table 15. For the other assembly conditions, the assembly of VLP's is even more inefficient with a 1.5 % - 4 % yield. Due to this inefficient assembly the concentration of VLPs do not lend themselves to further characterization experiments. However, after assembly of the L1-His_6

monomer and a continuing size exclusion chromatography the concentration of the pooled peak I contain the VLP´s with 0.1 mg/ml was suitable for negative stained transmission electron microscopy (TEM) (figure 39 B and C). TEM could be obtained showing L1-His$_6$ (figure 39 B) and His$_{10}$-L1 (figure 39 C) spherical particles of ~ 50 nm in diameter similar to the size of HPV16 VLP´s [6, 141]. We conclude that assembly with L1 monomer is possible but inefficient. Assembly of L1 full-length pentamer into VLP´s was also not possible because the amount of purified L1 pentamer was too low. The amount of assembled VLP´s is too low and hampers further experiments.

3.11 Interaction studies between L1nav and L2

Different methods can be used to analyze L1nav L2 protein interaction. In order to find suitable methods for interaction analysis, the two interaction partners were biochemically characterized beforehand. This characterization revealed for L2 in section 3.8.3 a temperature stability of about 30 °C in A0 buffer containing 0.5 M L-arginine and between 18 – 14 °C in A0 buffer containing 1.6 % (w/v) Brij 58 buffer. These results demonstrate that methods performed at room temperature (RT) are not suitable for interaction studies with L2, which tends to aggregate at RT. Another important role is played by the buffer. Although L2 is more stable in A0 buffer with 0.5 M L-arginine temperature wise, it cannot be used because L-arginine can not only stabilize proteins but also prevent protein interactions [142]. Therefore, the interaction studies must be performed in A0 buffer containing 1.6 % (w/v) Brij 58. However, it was proven (data not shown) that L2 is stable only at low concentrations in A0 buffer containing 1.6 % (w/v) Brij 58 compared to A0 buffer containing 0.5 M L-arginine. For this reason, only methods where low protein concentrations can be used and detected are considered.

The L1 non-assembly mutant (L1nav) was used as interaction partner for L2. In addition, the L1 full-length protein in form of virus like particles (VLP's) should also be used. However, the yield after assembling L1 monomers into VLP's was too low for further interaction studies and could therefore not be used. In addition, the amount of purified L1 full-length pentamer was too low for assembly into VLP's.

As an overview we displayed the applied methods which did not lead to any results in table 16.

Table 16: Used methods for investigating L1nav L2 interaction and the occurred problems.

method	problem
size exclusion chromatography (SEC)	The elution volumes of L1 and L2 are too similar, therefore a differentiation of L1, L2 and L1/L2 complex was not possible with the used columns: Superose® 6 10/300, Superdex® 200 pg 16/60 Superdex® 200 increase 10/30. In addition, L2 in A0 buffer containing 1.6 % (w/v) Brij 58 is stable only at low concentrations which are not detectable with the UV detector of the ÄKTA system.
Native Page	For stabilization of L2, 1.6 % (w/v) Brij 58 is required. It seems that this high number of micelles has an influence on the protein running behavior at the Native Page and therefore no sharp bands occurred.
Microscale Thermophoresis (MST)	The MST measurement is performed at 20 °C, where L2 protein aggregated.
Analytical Ultracentrifugation (AUC)	In A0 buffer containing 0.5 M L-arginine no interaction between $L1_5$nav and Fl-L2 was observed. In contrast, in A0 buffer containing 1.6 % (w/v) Brij 58 both sedimentation velocity and sedimentation equilibrium showed aggregation of L1/L2 complex samples over time and therefore could not be evaluated.

3.11.1 EL(2)ISA as a qualitative scan

To demonstrate the interaction of HPV16 L1nav and L2 at the biochemical level with purified proteins, we used a modified ELISA assay as a qualitative scan. Therefore, 100 µl of 1 µg refolded and biotinylated L2 (A0 buffer containing 0.5 M L-arginine) was immobilized on a streptavidin-coated plate. To ensure that the biotin-L2 bound to the streptavidin plate, immobilized biotin-L2 was first incubated with Alexa Fluor® 594 conjugated specific HPV16L2
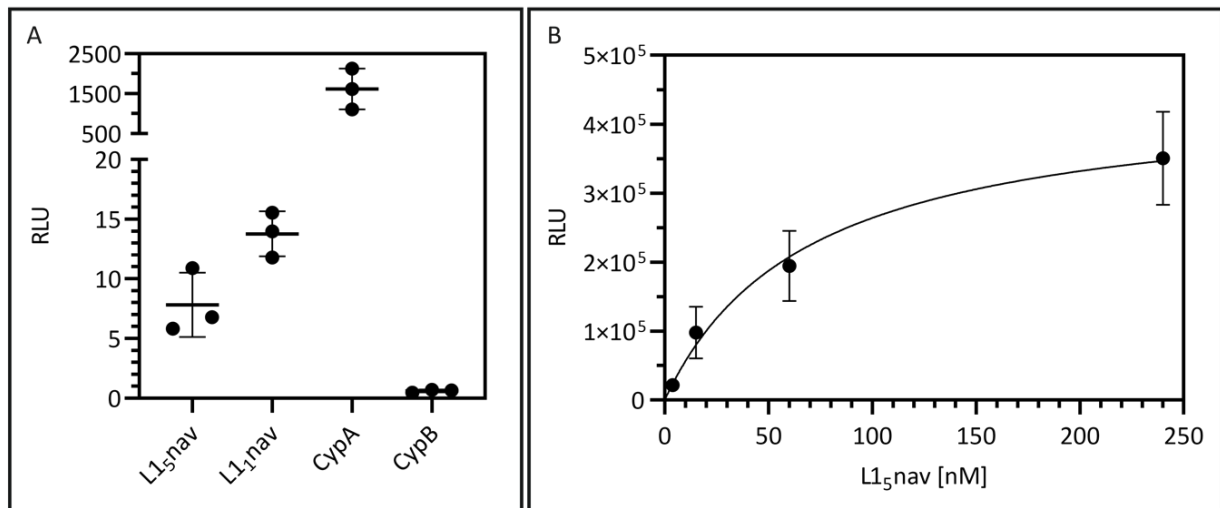
Figure 40: Qualitative scan of HPV16L2 binding partners analyzed via ELISA. (A) pre scan with L1$_5$nav and L1$_1$nav as well as CypA and CypB. Therefore, 100 µl of 1 µg refolded (A0 buffer pH 7.4 containing 0.5 M L-arginine) biotin-L2 were loaded on the Streptavidin-coated plate and incubated for 2 h at 4 °C. Afterwards, immobilized and refolded biotin-L2 was washed 3x with A0 buffer containing 1.6% (w/v) Brij 58 and 250 nM L1nav pentamer and monomer as well as CypA and CypB in A0 buffer pH 7.4 containing 1.6 % (w/v) Brij 58 were incubated for 2 h at 4 °C on immobilized biotin-L2. The proteins were analyzed via the specific primary antibodies and the enzymatic activity of the Horseradish Peroxidase (HRP). **(B)** To determine a binding affinity range of L1$_5$nav and L2, a 1-fold serial dilution of L1$_5$nav (starting concentration 240 nM) was incubated on immobilized L2. Data were fitted in GraphPad with an Equation: One site – Total binding fit.

antibody as a control and then analyzed as demonstrated in figure S 13. The results display a clear signal of the specific HPV16L2 antibody and thus immobilized biotin-L2 on the streptavidin plate. The immobilized L2 was then incubated with potential binding partners including L1 monomer and pentamer plus two purified cyclophilins (CypA and CypB) as cellular binding partners (kindly provided by PD Dr. Cordelia Schiene-Fischer) for 2 h at 4 °C. To analyze the binding, the particular proteins were incubated with the appropriate primary antibodies and afterwards with the Horseradish Peroxidase (HRP). For the analysis the enzymatic activity of the HRP was measured. To correct the HRP signal the unspecific protein signal was subtracted.

The results in figure 40 A show a binding from L1$_5$nav, L1$_1$nav and CypA to the immobilized L2. No binding to L2 is evident for the CypB protein. Please note, that the RLU of different added proteins cannot be compared quantitatively, because different antibodies were used and epitopes of L1$_5$nav and L1$_1$nav might have different accessibilities. To obtain initial indications

of the strength of binding affinity between L1nav pentamer and L2, we incubated immobilized L2 with different L1$_5$nav concentrations. The experiment was performed with L1nav pentamer only, as only this was used for continued interaction studies.

The specific binding affinity of L2 and L1nav pentamer is demonstrated in figure 40 B. Via Graph Pad a One site – specific binding fit was performed and demonstrated a specific binding affinity (K$_d$) with ~ 70 nM ± 58 nM:

$$Y = \frac{B_{max} * X}{(K_d + X)}$$

An ELISA assay can yield three different types of data output: quantitative, qualitative and semi-quantitative. For figure 40 A we receive a qualitative result which tells us, L1nav pentamer/monomer or CypA/CypB is binding to L2 or not. For figure 40 B we obtain a quantitative data output. However, it should be noted that it is not possible to determine exactly how much L2 is immobilized in each well on the plate and thus a quantitative analysis of the K$_d$ must be interpreted with caution. For this reason, the method can be used as a qualitative method for screening of possible binding partners of L2 and as a quantitative method to determine the specific binding affinity. However, there are more appropriate quantitative methods determining the binding affinity of different proteins.

3.11.2 Anisotropy measurement of L1nav pentamer and L2

An orthogonal method to investigate the binding affinity of L1 and L2 proteins on a quantitative level is the anisotropy measurement. The assay was performed as described in section 2.2.22.2 and different purified proteins are needed. For the measurement FL-L2, L1nav pentamer, unlabeled GGGL2 and FL-peptide were used.
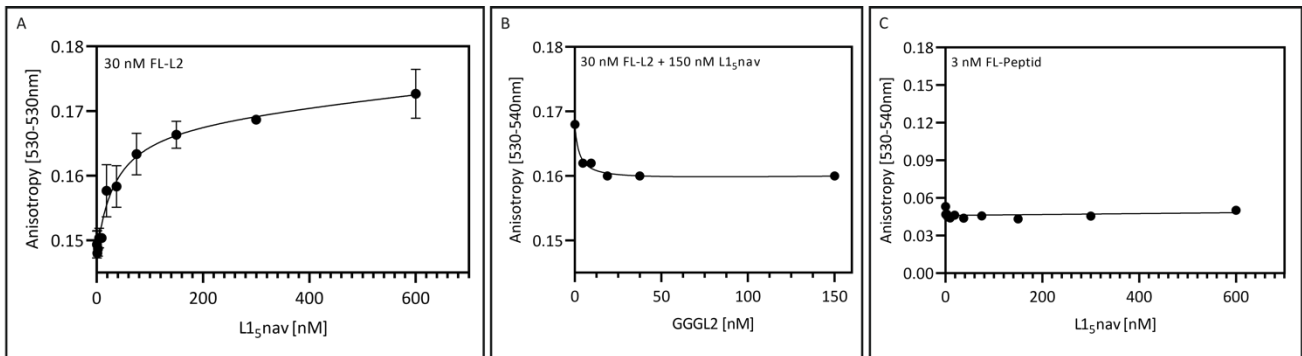
Figure 41: Binding experiment measured via anisotropy. (A) direct binding measurement of 30 nM FL-L2 with a 1-fold serial dilution of L1nav pentamer with a starting concentration of 600 nM incubated for 1.5 h at 4 °C. **(B)** competition measurement performed with 80 % complex saturation. At first 30 nM FL-L2 with 150 nM $L1_5$nav was incubated for 1.5 h at 4 °C to form a complex. Afterwards, unlabeled GGGL2 with a starting concentration of 150 nM and a 1-fold serial dilution was incubated with the complex for 1.5 h at 4 °C. **(C)** unspecific binding measurement of 3 nM FL-peptide with a 1-fold serial dilution of L1nav pentamer starting with a concentration of 600 nM. Samples were incubated for 1.5 h at 4 °C. All measurements were measured in A0 buffer containing 1.6 % (w/v) Brij 58 (A0: 200 mM NaCl, 50 mM Tris • HCl pH 7.4 @ 8 °C, 5 % (w/v) Glycerol).

In figure 41 A the change of anisotropy for the direct binding of FL-L2 and L1nav pentamer is shown. Therefore a 2-fold serial dilution of L1nav pentamer with a starting concentration of 600 nM was incubated with 30 nM of labeled L2 and the measurements were performed in experimental triplicates.

The change in Anisotropy was rather small. To compensate for measurement deviations, the mean value of the anisotropy values from 530-540 nm was used and plotted against the associated L1nav pentamer concentration. However, the small increase in the anisotropy signal, does not indicate reliable values upon complex formation. For this data a one site – specific binding fit was applied (formula demonstrated in section 3.11.1). The result demonstrates a binding affinity from L1nav pentamer and L2 of K_d = 33 nM (± 32 nM). The confidence interval of the fit is very high due to the variability of the B_{max} (maximum binding in the same unit as Y) value. In addition, it looks like saturation has not yet been reached at 600 nM L1nav pentamer. Additionally, the measured anisotropy data demonstrate a large variance between the different measurements.

Secondly a competition assay was performed to measure the relative binding affinity and investigate the reversibility of the L1/L2 complex. Furthermore, high-affinity interaction which are difficult to measure directly can also be quantified via a competition assay as well as

nonfluorescent ligands [143, 144]. In figure 41 B the results are demonstrated. For this purpose, a L1/L2 complex with 80 % saturation, consisting of 30 nM FL-L2 and 150 nM L1nav pentamer, was prepared and afterwards 2-fold serial dilution of unlabeled GGGL2 with a maximum concentration 150 nM was prepared and added to the L1/L2 complex. The data were processed as described for the direct binding assay and the binding affinity resulted in K_d = 2.5 nM (± 25nM). The result indicates a reversibility of the complex formation. The binding affinity of the competition experiment appears more than ten times stronger between L1 and L2 compared to the direct binding measurement which indicates a negative influence of the fluorescein-tag on the binding ability of FL-L2 to $L1_5$nav. However, the signal decrease in the competition measurement does not return to the initial value of 0.15. This incomplete decrease in the anisotropy value demonstrates an incomplete replacement of FL-L2 to GGGL2. This could be related to aggregation of GGGL2 or a replacement is prevented by the fluorescein tag or the containing micelles due to the Brij detergent.

To exclude non-specific binding between L1nav pentamer and the fluorescein tag, 3 nM fluorescein-peptide was incubated with a 1-fold serial dilution of L1nav pentamer at a starting concentration of 600 nM (figure 41 C). The result indicates no interaction of fluorescein-peptide with the L1nav pentamer protein. The various results of the different anisotropy measurements and the calculated binding affinities including the associated confidence interval are demonstrated in table 17.

Table 17: Results from anisotropy measurement for direct binding and competition assay with calculated binding affinity (K_d) and associated confidence interval (95 % CI).

Measurement	K_d	95% CI
direct binding	33 nM	± 32 nM
competition assay	2.5 nM	± 25 nM

The results of the anisotropy measurement demonstrate for the direct binding (33 nM) as well as for the competition assay (2.5 nM) a binding of L1nav pentamer and FL-L2/GGGL2. Therefore, the anisotropy data support the EL(2)ISA data where immobilized biotin-L2 is shown to interact with L1nav pentamer with a binding affinity of ~ 70 nM. However, it should be noted that this method is not suitable for the interaction studies of L1 and L2 because L2 tends to aggregate and the direct binding measurement result demonstrates that the increase in the anisotropy signal of 0.02 is rather small and therefore not reliable. In addition, the competition assay does not indicate complete competition of FL-L2 to GGGL2.
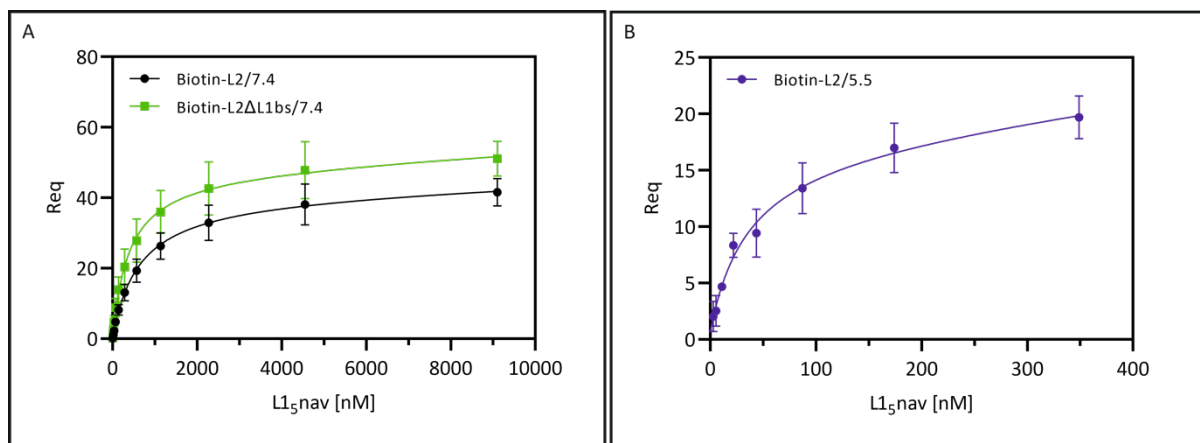
3.11.3 Surface plasmon resonance (SPR) measurement of L1nav pentamer and L2

Because of the high aggregation tendency of L2 we applied a second orthogonal method to determine the binding affinity of the L1 and L2 interaction. This method is called surface plasmon resonance (SPR) measurement and is commonly used to analyze the binding kinetics of proteins in real time. The advantage of this method is the immobilization of biotin-L2 which can lead to the prevention of L2 aggregation.

The measurement was performed as described in section 2.2.22.3, briefly biotinylated L2 was immobilized on a streptavidin coated Chip and afterwards incubated with different L1nav pentamer concentrations. Additionally, to biotin-L2 a L2 mutant which lacks of the L1 binding site, according to analogy to literature data, was analyzed (Biotin-L2ΔL1bs). Furthermore, the measurements for biotin-L2 were performed in CP buffer pH 7.4 and 5.5 containing 0.16 % (w/v) Brij 58. The pH value in CP buffer is not temperature dependent and was therefore used to ensure the correct pH value at 25 °C.

To calculate the binding affinity based on the different SPR measurements, we need the response unit where the equilibrium is reached (Req - original sensorgrams figure S 14). For the analysis, the response unit at 115 s (Req) was plotted against the different L1nav pentamer concentrations as demonstrated in figure 42. Afterwards, the fitting of the data was performed via GraphPad with a One site - total binding fit applied (formula demonstrated in section 3.11.1).

Figure 42: Surface plasmon resonance measurement. Measurements were performed at 25 °C and all plotted Y-data show the response unit in equilibrium at 115 s (associated sensorgrams displayed in figure 14). **(A)** Measurements performed in CP buffer (CP buffer pH 7.4: 18 mM Na_2HPO_4, 1 mM $C_6H_8O_7$, 200 mM NaCl, 5 % (w/v) Glycerol) pH 7.4 containing 0.16 % (w/v) Brij 58 for biotin-L2 (black dots) and biotin-L2ΔL1bs (green dots) with a 1-fold serial dilution of L1nav pentamer, starting concentration 9100 nM. Data were fitted via GraphPad one site – total binding fit. **(B)** Result for biotin-L2 (blue dots) in CP buffer pH 5.5 (CP buffer pH 5.5: 11 mM Na_2HPO_4, 4.4 mM $C_6H_8O_7$, 200 mM NaCl, 5 % (w/v) Glycerol) containing 0.16 % (w/v) Brij 58, performed with a 1-fold serial dilution of L1nav pentamer with a start concentration of 2790 nM. Due to measurement errors, the last three measurement points had to be excluded. Data were fitted via GraphPad one site – total binding fit.

In figure 42 A the results for biotin-L2 (black dots) and biotin-L2ΔL1bs (green dots) in CP buffer pH 7.4 containing 0.16 % (w/v) Brij 58 are shown. The black line demonstrates the data fit for biotin-L2 and the green line for biotin-L2ΔL1bs. Calculated with this fit, the binding affinity of biotin-L2 to L1nav pentamer is 620 nM and for biotin-L2ΔL1bs 392 nM. The difference in response unit can be explained by the fact that two different sections of a Chip were used for the measurements. It is possible that for the biotin-L2 less ligand was loaded onto the Chip than for biotin-L2ΔL1bs and for this reason, less L1nav pentamer could bind. In figure 42 B the result for biotin-L2 (blue dots) in CP buffer pH 5.5 containing 0.16 % (w/v) Brij 58 is demonstrated. The blue line displays the one site – total binding fit with a calculated binding affinity for biotin-L2 and L1nav pentamer at pH 5.5 of 36 nM. Compared to the measurements in CP buffer pH 7.4 containing 0.16 % (w/v) Brij 58 (figure 42 A) the response unit is reduced by half. This is related to the fact that the immobilized Chip section containing biotin-L2 was previously used for another measurement (biotin-L2 in CP buffer pH 7.4 containing 0.16 % (w/v) Brij 58). Thus, more immobilized ligand (biotin-L2) was washed down and less L1nav pentamer could bind.

In table 18 the calculated binding affinities with the associated 95 % confidence interval (CI) are demonstrated.

Table 18: Results from SPR measurement with calculated binding affinity (K_d) and associated confidence interval (95 % CI).

Sample	K_d	95 % CI
biotin-L2 pH 7.4	620 nM	± 296 nM
biotin-L2 pH 5.5	36 nM	± 69 nM
biotin-L2ΔL1bs pH 7.4	392 nM	± 229 nM

The large binding range, visible by the high 95 % confidence interval of the applied One site – total binding fit, for biotin-L2 pH 5.5 measurement, can be explained because the equilibrium is not reached at 348.75 nM. This gives the fit many degrees of shaping and thus a large range in the confidence interval. To better fit the data, one could fix the B_{max} value. However, by fixing the B_{max} value the fit shifts and the fit aligns to fewer measured data points. To improve the measurement, one would have to perform a new measurement with higher L1nav pentamer concentrations. Nevertheless, the data indicate a significantly higher binding affinity of L1 and L2 at pH 5.5.

The data suggest binding between L1nav pentamer and L2 at pH 7.4. Furthermore, the measurement demonstrates a ~ 17 times higher binding affinity of the L1nav pentamer protein to L2 at a lower pH of 5.5. This could be related to the fact that L1nav pentamer seems to be more stable at lower pH as demonstrated by the temperature stability measurements in section 3.9.4. An increase in stability may indicate that protein interaction can be more easily achieved, as any flexible interaction regions are stabilized and thus more easily accessible.

In contrast, for the biotin-L2ΔL1bs mutant low or no binding of L1nav pentamer was expected but we did not observe a large change in binding affinity compared to the full length L2 protein at a pH of 7.4. On the contrary, with a binding affinity of 392 nM, the L2 mutant appears to bind L1nav pentamer better than the L2 full length protein. This result suggests that there may be a second L1 binding site located on the L2 protein.

4   DISCUSSION AND OUTLOOK

The main goal of this work was to investigate the interaction of recombinant L1 and L2 *in vitro* due to various environmental conditions (pH) and L2 full-length as well as a L2 variant leaking the putative L1 binding domain at the C-terminus (L2ΔL1bs). Furthermore, we screened L2 with a modified EL(2)ISA for further potential binding partners. This protein interaction could be demonstrated by three different methods and the binding affinity quantitatively determined.

For the analysis of the L1 L2 interaction we needed purified and biochemical/biophysical characterized protein.

To produce both recombinant L1 full-length protein and L1nav protein and to characterize them after purification, *E. coli* was used as an expression system and a fed-batch fermentation was established [1]. High yields of soluble L1nav and L1 recombinant protein were obtained after purification with an N-terminal His_6-fusion tag. After tag-removal the purified recombinant L1 and L1nav protein were characterized due to stability, oligomeric state and secondary structure.

The second subordinate goal of this thesis was the labeling and biotinylation of HPV16 L2 for further biochemical analysis and protein interaction studies. Therefore, recombinant His_6-tagged TEV-GGGL2 was expressed in *E. coli* in inclusion bodies and purified via affinity chromatography (NiNTA). The subsequent analyses were able to gain insights into the refolding conditions, stability and secondary structure of the minor structure protein L2.

4.1 Recombinant L1 full-length and L1nav protein production in *Escherichia coli*

It has been described several times that the expression of soluble L1 capsid protein in *E. coli* is a problem regarding solubility [1, 96, 102, 104]. Nevertheless, prokaryotic expression systems are a popular choice for the expression of recombinant proteins because the *E. coli* expression systems have a high yield of heterologous protein, easy culturing, fast growing, low production costs and high space-time yield (amount of synthesized product per volume pack per day). To overcome the solubility problem of recombinant L1 protein in *E. coli* different parameters can be optimized: (I) reduced expression temperature [115, 116], (II) co-expression of a GroEL/ES chaperone system [1, 96, 98], (III) solubility inducing fusions tags [1, 4, 102] and (IV) fusion-tag removal.

Lowering the expression temperature enhances the solubility of the recombinantly produced proteins [145-148] because at lower temperatures the cell processes are slower and this leads to reduced rates of transcription, translation and cell deviation [149]. This increases cultivation times to obtain high biomass. However, it increases protein yield by providing more time for protein folding and oligomerization and finally reducing protein aggregation [150, 151]. In addition, many proteases are less active at reduced temperatures and this results in a reduction of degradation of proteolytically sensitive proteins [152-155]. To further increase solubility, the chaperone complex GroEL/ES was co-expressed as a second optimization strategy, what did not influence overall biomass yield (figure 14, S2). This chaperones generally assist in protein folding [97] and protein assembly [96]. The GroEL/ES system detects misfolded proteins and provides kinetic support for refolding. Therefore, the activation barrier for intermediate states of refolding process in an ATP-dependent manner is reduced [97]. As a third optimization option to increase solubility, different solubility inducing fusion-tags (GST- and SUMO-tag) were used (figure 8). It has been shown that in particular a N-terminal GST-tag increases the solubility of recombinant L1 protein significantly [127] as well as other fusions proteins including the SUMO fusion-tag. For fused L1 protein [124] as well as for other fusion proteins, this tag has a solubility-mediating effect [156]. All applied fusion tag show minor effects on bacterial cell growth and solubilities between about 50 % to 100 % which can also be achieved without any tag (data not shown). These results suggest that in this context the solubility-inducing SUMO-tag plays a rather minor role and the reduction of the expression temperature as well as the co-expression of GroEL/ES significantly increases the solubility. However, a tag-free protein purification did not result in sufficient and pure L1 protein. Hence

the tag-removal is an additional parameter influencing the final protein yield post purification. The successful cleavage and removal of the especially larger fusions-tags (e.g., GST- or SUMO-tag) plays an important role. The problem: the large size of the GST- , SUMO-tag [128, 156] can prevent interactions between proteins, assembly or medical applications. For the L1-pentamers as well as the full-length protein a high cleavage efficiency is necessary because just 20 % remaining fusion tag on one L1 pentamer, would mean that statistically every pentamer has one fusion tag left. Conversely, this means that the cleavage efficiency must higher than 80 % to get completely cleaved L1 pentamers [1]. For the GST-L1nav construct the purification was performed as already published [127] but the GST-tag cleavage was not successful (figure 22) [1]. The same result for incomplete cleavage could be shown for the His_{10}- fusion tag from the L1nav and L1 full-length L1 construct. This could be due to the rather small recognition site of the enterokinase and its prevented access in the His_{10}-L1nav/His_{10}-L1 protein. Due to the small $His_{6/10}$-tag the uncleaved constructs can be used for further experiments without any tag cleavage and removal. To overcome the inefficient fusion-tag cleavage we tested an alternative fusion construct His_6-SUMO-L1nav/His_6-SUMO-L1 harboring a N-terminal His_6-SUMO-fusion. For His_6-SUMO-L1 construct we observed incomplete cleavage (figure 22). One reason could be the longer N-terminus of the L1 construct. Even though the purification and cleavage were performed under reducing conditions to avoid assembly, it is possible that the longer N-terminus is too flexible or hidden that cleavage is difficult.

In contrast the His_6-SUMO-tag from the His_6-SUMO-L1nav construct was cleaved successfully by a His_6-SUMO-Protease (figure 22) and removed via an affinity chromatography column (NiNTA). This complete cleavage can occur because the SUMO-Protease recognizes the whole SUMO-tag instead of just a small recognition site [129]. Furthermore, the SUMO-tag was fused N-terminal with a His_6-tag for faster purification of the fusion construct and the SUMO-Protease can be produced in *E. coli* [157]. In contrast to the SUMO-Protease, thrombin, like enterokinase, recognizes only a small recognition site of the GST-tag and this could explain the incomplete cleavage for the GST-L1nav construct.

The expression strategy with co-expression of GroEL/ES [117] and the fusion with the His_6-SUMO-tag [156] was already reported before. Here, we combined these strategies and transferred it to a process-controlled fed-batch fermentation (published in Roos et. al. [1]). This process increased on the one hand the biomass yield by 10-fold and on the other hand

the protein yield per biomass by 10-fold (figure 18). Process controlled fermentation with regulation of oxygen content and substrate limitation has been shown to be a good strategy in previous fermentations of VP1 [158]. Probably because slowed and controlled growth by substrate limitation and reduced temperature increases protein solubility and therefore, we achieve a total yield of 343.6 mg final purified L1nav pentamer and 638.4 mg L1nav monomer per fermentation (in 5 L M9 medium with OD_{600} = 40) [1].

4.2   L1nav characterization

After the successful purification we needed to analyze the cleaved and pure L1nav construct. At first L1nav was characterized due to the purity of the recombinant protein. The SDS-PAGE analysis showed > 95 % pure L1nav pentamer protein and > 97 % pure monomer L1nav protein as displayed in figure 33 A. The same is true for DNA impurities, the proteins are almost free of DNA impurities due to a 260:280 ratio of ~ 0.5 as shown in figure 33 B. It is possible to purify this in *E. coli* expressed construct in a L1nav pentameric or monomeric form as already reported [103] and in total we were able to purify 5.73 mg L1nav pentamer and 10.64 mg L1nav monomer per gram dry biomass. This yield for purified L1nav protein is rather high for comparison of published data [1]. To investigate the oligomeric state of the L1nav protein a size exclusion chromatography under different conditions was performed and it was shown that under reducing conditions L1nav elicited at a size of ~ 50 kDa. This result is consistent with the size of an L1nav monomer. As soon as the L1nav monomer is dialyzed against non-reducing buffer, an elution shoulder appears. This suggests that the oligomerization of L1nav proteins changes under non-reducing conditions (figure 34). For the L1nav pentamer under non-reducing conditions the result shows a size of ~ 250 kDa which fits to the already published data [1]. The situation is different for the L1nav pentamer in A0 buffer pH 7.4 containing 1.6 % (w/v) Brij 58. The run looks more heterogeneous compared to the run in A0 buffer (non-reducing conditions). Everything from oligomers to monomers is present but the main part demonstrates L1nav pentamer. This could be due to the micelles which in some way influence the interactions of the individual monomers. This result emphasizes that we must pay attention to the buffer composition when using the L1nav pentamer and monomer. If necessary, analyze all the samples for further experiments on a size exclusion chromatography again to be able to determine the oligomerization state.

To confirm the results from the analytical size exclusion, an additional analytical ultracentrifugation has been performed. The results for reducing buffer indicate two distinct

species which have the size of an L1nav monomer with 59 kDa (65 %) and L1nav pentamer with 219 kDa (35 %). Under non-reducing conditions, only one distinct species with the size of an L1nav pentamer with 250 kDa is seen, although smaller numbers of larger species, presumably oligomers, are also seen. For the measurement in A0 buffer pH 7.4 containing 1.6 % (w/v) Brij 58 under non-reducing conditions the result looks similar. These results confirm that the state of oligomerization differs under various buffer conditions.

To find out the stability of L1nav monomer and L1nav pentamer, we monitored the aggregation of thermally induced unfolding (figure 37) under different conditions. It can be clearly shown that the L1nav pentamer is more stable than the monomer in CP pH 7.4 buffer. The addition of 1.6 % (w/v) Brij 58 also increases the stability of the L1nav pentamer. This may be related to the fact that detergents increase the stability of proteins [159]. L1nav pentamer is most stable when the pH is also lowered to 5.5 (figure 37). This result is difficult to reconcile with the assumption that uncoating of the virus in the endosome occurs at low pH [62, 75]. This stabilization may also play a role in viral entry at the cell membrane, as published results have shown, the pH in infected cervix tissue differs than in healthy tissue. This study indicates a 30 % greater risk of infection with multiple HPV types when pH increases [160]. For this reason, elevated pH could cause L1 protein to destabilize, allowing the N-terminal explosion of the L2 protein and afterwards entry into the cell.

To ensure that the purified protein folds correctly, the secondary structure was measured via circular dichroism and compared with previously published data [103]. The correct folding guarantees a functional protein which can be used for further experiments. The results of the measurement promote that the secondary structure of L1nav pentamer in A0 buffer and A0 buffer containing 0.16 % (w/v) Brij 58 buffer pH 7.4 and pH 5.5 look similar which indicates no change in the secondary structure under different buffer conditions. In contrast, the L1nav monomer looks different. However, this result also fits the already published data [103]. In summary, analytical size exclusion chromatography, analytical ultracentrifugation, and TEM demonstrated that the L1 non-assembly variant forms pentameric subunits under non-reducing buffer conditions.

4.3   L1 full-length characterization

For the functionality of the L1 full-length protein, L1 was assembled into VLPs via a redox system. The analytical size exclusion showed a small amount of VLPs and the TEM images show particles with a size of round about 55 nm [127]. These results would fit to HPV16L1 VLPs, but the assembly is too inefficient and further characterization was not possible.

This low efficiency of VLP assembly may be related to the fact that L1 monomer was used as the starting product and L1 monomers still contain an N- or C-terminal $His_{6/10}$-tag as the fusion-tag cleavage for the L1 full-length protein was rather unsuccessful (section 4.2). Although this His-tag is very small, it can still interfere or prevent assembly of the monomeric subunits. In addition, the literature demonstrates the assembly of VLP's from pentameric L1 subunits [139, 161]. Both the L1 full-length protein as a pentameric subunit as well as assembled VLP's would be good further options for investigating L1 and L2 interaction. Until now it is not known how L2 and L1 interact with each other and which parts of the L1 and L2 protein are involved. In addition, it is not known whether L1 full-length protein or VLP's stabilize L2 and the N- or C-terminus as well as the h4 loop, which were removed in the L1nav mutant, play a role.

4.4 Recombinant L2 protein expression and purification

The expression of L2 constructs (His$_6$-TEV-GGGL2, His$_6$-TEV-GGGL2ΔL1bs) in *E. coli* as inclusion bodies and the high yield purification is already described. Purification of L2 as demonstrated in figure 23 was performed under denatured conditions (6 M guanidine hydrochloride) and via affinity chromatography (NiNTA) [77]. The results demonstrate a distinct band at ~ 75 kDa and one at ~ 40 kDa and this running behavior is characteristic for L2 [25, 26, 130, 131]. Additionally, the sample is almost free of DNA according to the A$_{260}$:A$_{280}$ ratio (figure 24).

The denatured L2 is not functional and was therefore refolded in various detergents and buffers which has already been described [77]. To ensure enzymatic activity for cleavage and labeling/biotinylation, the L2 sample must be refolded in A0 buffer pH 7.4 with 0.5 M L-Arginine since TEV-Protease and *Sortase A* are not expected to be active under denaturing conditions.

4.5 *Sortase A* as a "Tool-Box"

As already described, protein tag-fusion via sortase-mediated ligation (SML) by *Sortase A* is a popular and simple tool [92, 111, 162]. The applications are very diverse, as demonstrated in this work for two applications (labeling and biotinylation of L2). Before we could apply sortase-mediated ligation, the N-terminal His$_6$-TEV-tag had to be successfully cleaved with TEV protease (figure 25) and removed afterwards to reveal a GGG-motif.

The results for the SML with fluorescein-LPETGGRR peptide in figure 27 clearly demonstrate a N-terminal labeling of the L2 protein and the degree of labeling (DoL) was determined spectroscopically by absorption measurement (UV/VIS spectrum – figure 28) according to the Lambert-Beer-law and reveals 72 %. The problem is that the extinctions coefficient (ε) for the fluorescein-peptide was determined via free fluorescein-peptide and it is not known whether immobilization of fluorescein peptide alters the extinctions coefficient.

The SML for the biotin-LPETGGRR peptide in figure 27 demonstrates a successful biotinylation for the L2 protein with a different degree of biotinylation (DoB) for biotin-L2 (92 %) and for biotin-L2ΔL1bs (46 %). The rather low DoB for biotin-L2ΔL1bs could be a result of steric hindrance, high mobility of the N-terminal region of L2 or maybe poor accessibility. Therefore, *Sortase A* could not sufficiently catch the GGG-motif. For improving the labeling rate, the incubation time could be extended, or a higher molecular excess of peptide could be used [162, 163].

All in all, the labeling or biotinylation of L2 via sortase-mediated ligation was successful but 60-70 % (table 12) of the L2 protein was lost due to the high number of purification steps. One could try to run the His$_6$-TEV-tag cleavage and the *Sortase A* reaction sequentially or simultaneously without a purification step. However, it is important to note that *Sortase A* catalyzes both the linkage of the substrates and the dissolution of the bond, and therefore attention must be paid to the reaction equilibrium.

4.6 L2 refolding and characterization

L2 purification from inclusion bodies was performed under denaturing conditions and for this reason L2 must be refolded afterwards into its native structure. For A0 buffer containing 0.5 M L-Arginine this has already been successfully established and described [77]. However, pre liminary results of interaction studies show no interaction of L1 and L2 in A0 buffer pH 7.4 containing 0.5 M L-arginine (data not shown). This may be related to the fact that arginine stabilizes proteins but can also hinder protein-protein interaction [142]. For this reason, other detergents had to be considered for the stabilization of L2. According to literature the detergent Brij 58 appears to be a good candidate [159] and was therefore tested according to the critical micelle concentration [164]. A Brij concentration of 1.6 % (w/v) with pH 7.4 and 5.5 showed lowest aggregation of GGGL2 protein. These data match those already published for A0 buffer containing 0.5 M L-Arginine [77]. To make sure that the fluorescein tag does not affect the stability of L2, tagged L2 (FL-L2) in A0 buffer containing 1.6 % (w/v) Brij 58 pH 7.4 was also examined (figure 30). Again, the results show little aggregation, which supports the stability of FL-L2.

To further characterize the stability of refolded L2 in A0 buffer containing Arginine or Brij 58, the thermally induced aggregation of GGGL2 was analyzed. For A0 buffer pH 7.4 containing 0.5 M L-Arginine these data are already published and serve as comparative values [77]. The values in A0 buffer pH 7.4 with 0.5 L-M arginine (~ 31 °C) agree with the already published values and for pH 5.5 the stability slightly dropped to ~ 28.5 °C. In A0 buffer pH 7.5 and 5.5 with 1.6 % (w/v) Brij 58 the thermally induced aggregation results look quite different (figure 31). The data indicate that stability of GGGL2 is increased nearly by half in A0 buffer containing 1.6 % (w/v) Brij 58 compared to A0 buffer with 0.5 M L-Arginine. To investigate if the fluorescein tag has an influence on the L2 stability, the FL-L2 was analyzed via thermally induced aggregation A0 buffer pH 7.4 containing 1.6 % (w/v) Brij 58. The results demonstrate that again the stability is reduced from 18 °C without fluorescein tag to 13.8 °C. This result

suggests that the fluorescein tag has an influence on the temperature stability of L2 in A0 buffer containing 1.6 % (w/v) Brij 58. These temperature stability data are very important to determine which temperatures can be used in subsequent experiments to avoid aggregation of L2.

Finally, the secondary structure of L2 was characterized by a circular dichroism (CD) measurement to ensure that L2 is properly folded. For His$_6$-L2 in A0 buffer pH 7.4 with 0.5 M L-Arginine, these data have been described previously [77]. For A0 buffer containing pH 7.4 1.6 % (w/v) Brij buffer the data in figure 32 demonstrate a similar secondary structure of GGGL2 as already published for His$_6$-L2 [77]. It can be concluded that Brij has no effect on the secondary structure of GGGL2 compared to arginine and that both buffer compositions are suitable for refolding L2. However, it must be mentioned that GGGL2 in Brij buffer is stable only at low concentrations (data not shown).

4.7   L1-L2 interaction

As already described in the introduction, previous studies demonstrate an interaction between HPV16L1 and HPV16L2 *in vitro* and *in vivo* [3, 6, 29, 87, 91]. However, it has not yet been possible to characterize this interaction further with respect to the binding affinity or under different chemical conditions.

4.7.1   EL(2)ISA as a "screening tool"

This established and modified ELISA (EL(2)ISA) serves on the one hand to analyze the interaction of immobilized L2 with different purified binding partners an on the other hand to screen for unknown binding partners from cell lysate with combined mass spectrometry. In this work, we focused on the purified interaction partners of L2. Therefore, refolded biotinylated L2 was immobilized on a streptavidin plate and incubated with potential binding partners. The results in figure 40 demonstrate a clear binding of L1nav pentamer and monomer with L2. In addition, binding of CypA was shown, but not for CypB. CypA is a cytosolic protein, although nuclear localization and secretion were described as well [165]. CypB is located on the extracellular membrane and mediates the exposure of L2 N-terminus [166]. In addition, previously described results demonstrate that Cyps facilitate the release of L1 from the L1/L2 complex [62]. Our pre liminary binding results suggest no binding of L2 and CypB. However, these results are contradictory because it has already been shown that a CypB binding site (aa 90 - 110) is located on the L2 protein [62]. This contrary result could be possible since we used a CypB protein leaking aa 1 – 34 at the N-terminus. However, to be

able to make a clear statement, this interaction would have to be characterized further. With this established EL(2)ISA assay it is possible to analyze the binding partners of L2 in a qualitative and quantitative way. In addition, this assay could be used to investigate potential binding partners for other biotinylated and immobilized proteins as well.

Applying the EL(2)ISA assay the binding affinity of L2 and L1 was determined with (K_d) ~ 70 nM ± 58 nM. This method can give a first impression of the binding affinity of the proteins for further investigation.

4.7.2   L1-L2 interaction

To analyze the L1 L2 interaction *in vitro*, there were several experimental challenges to overcome. Firstly, detergent had to be used and secondly, all measurements had to be made at below 13 degrees for L2 stability. In addition, two orthogonally used methods for L1 and L2 interaction were established to confirm results from EL(2)ISA measurement.

To study the interaction of L1nav pentamer and L2 on a quantitative level we established a binding assay based on fluorescence polarization called anisotropy and a surface plasmon resonance (SPR) measurement.

The direct binding for the anisotropy measurement demonstrates a K_d of 33 ± 32 nM (figure 41 A). However, the anisotropy signal increase from 0.15 for only labeled L2 to 0.17 for the complex and this difference is very small and does not fit to the literature data [167, 168]. Moreover, such a small increase is untypical for the assumed complex size. This small increase may be related to the micelles present in the A0 buffer containing Brij 58. But also, the unstructured L2 can give a very high anisotropy signal already at the beginning, if FL-L2 is present rather linearly and thus the rotation is very slow resulting in an already high anisotropy signal. Also, the increasing protein concentration of L1nav pentamer could cause a slight increase of the anisotropy signal in the direct binding measurement due to the increasing viscosity in the sample. Thus, the small signal change is also susceptible to measurement errors. For this reason, the average value of 530-540 nm was used for the calculation of the anisotropy signal. These results suggest that the method is not suitable for the study of L1 L2 integration. For this reason, we have established a second method (SPR) to investigate the direct binding of L1 and L2. After evaluation of the data the K_d for biotin-L2 with L1nav pentamer pH 7.4 was 620 nM (± 296 nM). This result demonstrates an ~ 18-fold lower binding affinity compared to the result of the anisotropy measurement at pH 7.4. In contrast, the K_d was found to be 17-fold stronger with 36 nM (± 69 nM) for pH 5.5. This

difference in binding affinity can be explained on the one hand by the increased stability of L1nav pentamer with decreasing pH. The stability data of the temperature transmission measurements clearly demonstrate that the lower the pH, the more stable the L1nav pentamer and this could lead to a stabilization of eventual flexible parts of L1nav pentamer which play a role in L1 and L2 interaction.

In addition, a biotin-L2ΔL1bs variant was measured. In this L2 variant, the previously published L1 binding side was removed. However, it must be taken into account that the putative L1 binding site was analyzed on HPV11 but the sequence of L2 is generally highly conserved and an alignment demonstrated putative L1 binding site for HPV16 L2 [29]. Moreover, it is not precisely known whether HPV16 L2 binds L1 via this putative L1 binding site. According to current knowledge, this mutant should interact significantly poorly with L1nav pentamer. However, the measured K_d of 392 nM (± 229 nM) demonstrate similar binding affinity as the L2 full-length protein at pH 7.4. This result suggests that there may be a second L1 binding site on the L2 of HPV16 protein located. To confirm this assumption, different L2 mutants need to be tested in future experiments.

In the SPR measurement, it appears that the detergent has no effect on the binding affinity of L1 and L2. However, to completely exclude this assumption, the detergent should be carefully removed and a new SPR measurement should be performed.

To support the results for the anisotropy measurement and to exclude non-specific binding of the fluorescein-tag, we had to perform a competition assay and a measurement of the direct binding with the free fluorescein peptide and L1nav. The results for the competition assay demonstrate a replacement of FL-L2 to GGGL2 with decreasing anisotropy signal (figure 41 B) and thereby specific binding can be displayed. Normally, the anisotropy signal decreases almost to the signal of unbound labeled protein alone in this case 0.15 (FL-L2). However, the anisotropy signal decreased to just 0.16. This result indicates that FL-L2 is still bound, due to dynamic equilibrium between FL-L2 and GGGL2 binding to L1nav pentamer. From these measured data, the calculated binding affinity is 2.5 nM (0 ± 28nM). These differences in the measurements can have various reasons. On the one hand, the micelles of the A0 buffer containing Brij can hinder an exchange of labeled and unlabeled GGGL2 and increase the viscosity of the buffer and the unstructured L2 may also play a role [169, 170]. Furthermore, the evaluation of binding affinities demonstrates a 13-fold stronger binding for the competition assay than for the direct binding. This result suggests that the fluorescein tag has

a negative influence for the L1nav pentamer and L2 interaction, however it should be noted that the competition measurement data is not reliable.

To exclude the unspecific binding of fluorescein-peptide to L1nav pentamer a direct binding via anisotropy measurement was performed and displayed no unspecific binding of fluorescein-tag and L1nav pentamer (figure 41 C) and therefore L1 and L2 interaction must be specific. As discussed above, this experimental set up for the anisotropy measurements does not seem to be optimal for our purposes although a binding of L1nav pentamer and FL-L2 could be confirmed and the binding affinity fits the EL(2)ISA data in section 4.7.1. For this reason, the SPR results were used as reliable data. Both orthogonally used methods display a L1 L2 interaction. The advantage of the anisotropy measurement is that L2 is not immobilized and thus the interaction can be measured in solution. However, non-immobilized L2 is also more unstable. For this reason, the SPR measurement is more suitable. By immobilizing L2, it is more stable and the amount of detergent can be reduced by a factor of 10. Also, the results demonstrate a real time kinetic of the binding affinity of L1 and L2 and therefore more information are provided in one measurement.

4.7.3 Biological relevance of the data in relation to viral entry

With respect to the above discussion, I conclude 3 major findings: (I) L2 interacts with CypA but not CypB, (II) alternative or additional L1 binding site of L2 and (III) lowering the pH increases binding affinity of L1 and L2 in parallel to thermal stability of L1.

The first finding (I) includes the interaction of L2 with CypA but not CypB. It is already known that Cyps play an important role in the viral entry of HPV16 and the dissociation of L1 and L2 at an acidic pH. The results for CypB conflict with previously published data. Thereby, a binding site of CypB is located on the L2 protein [62]. That our data does not fit the literature may be related to the fact that we did not use full-length CypB for the interaction studies (aa 34-216) and it is not known whether other co-factors are involved in the L2 CypB interaction. CypA facilitates membrane translocation [171] and furthermore CypA assists in the translocation of C2 toxin of *Clostridium botulinum* across membranes of acidified endosomes into the cytosol [172]. These published data of *Clostridium botulinum* suggest that the interaction of L2 and CypA may occur in the endosome. However, this assumption would require further experiments.

The second observation (II) includes an alternative or additional L1 binding site of L2 because a L2 mutant (Biotin-L2ΔL1bs) which, according to the literature, should no longer bind L1 [29] demonstrates a similar binding affinity than full-length L2. The L1 L2 interaction plays a role in the capsid, endosome and during viral entry. In the endosome, the L2 transmembrane domain (TEM) is known to interact with the endosomal membrane [28, 77] and thus the L2 N-terminus is located inside the endosome and the C-terminus is located in the cytosol according to a topology model [173]. Thus, the L1 binding site on L2 would no longer be accessible to L1. From this hypothesized model, we came to believe that a second L1 binding site might possibly be present on the L2 N-terminus. To investigate an additional L1 binding site various L2 mutants could be tested due to the binding affinity.

The third (III) and last finding includes an increasing binding affinity of L1 and L2 with decreasing pH. Therefore, it is important to discuss these results according to the different environments during viral entry. The stability results reveal with increasing pH the stability of L1nav pentamers decreases. These results are in line with already conducted studies. It was found that the pH in infected cervical tissue increases with HPV infection [160] and this decreases the stability of L1 which can lead to a conformational change being favored and the N-terminus of L2 is than exposed [60-62].

In addition, the pH also changes in the endosome where the pH drops to 5.5. This acidification together with cyclophilin interaction is important in virus uncoating [62, 75]. However, the acidification in the endosome contrasts with the stability data of L1nav pentamer. Moreover, L1 and L2 indicate a stronger binding affinity at pH 5.5 (36 nM) compared to 7.4 (620 nM). This suggests that other cellular factors such as cyclophilins have a greater influence.

4.8 Outlook

Regarding the interaction of L1/L2 the still remaining questions include: (I) the role of L1nav monomer, (II) the stoichiometry of the L1 L2 complex, (III) additional L1/L2 binding sites, (IV) the influence of CypA interaction with L2 and (V) tertiary structure of L1 L2 interaction. It is not yet known whether the monomer species is an artifact of *E. coli* production or whether the L1 monomer plays a role during viral entry. To exclude an artifact of *E coli* production, VLPs produced in yeast would have to be disassembled and analyzed according to the oligomeric state of L1 (SEC, AUC). In addition, the L1nav monomer interaction with L2 via SPR would need to be analyzed under different buffer conditions to compare the binding affinities of L1nav pentamer and monomer.

Previously published results demonstrate a 5:1 stoichiometry of the L1 L2 complex, indicating a potential amount up to 72 L2 bindings sites in the virion [29] However, it is questionable whether 72 L2 monomers are present per virion. To find out whether the stoichiometry of the L1 L2 complex has an influence on the viral entry, this should be investigated under different conditions (pH) and with further interaction partners (cellular factors).

At this time, a C-terminal L1 binding site on L2 is known [29]. However, our results indicate that either an additional binding site is located on L2 or the published L1 binding site does not fit for HPV16. An eventual second L1 binding site on the L2 protein should be investigated via truncated L2 variants. In addition, it is still unknown where the L2 binding site is located on the L1 protein. First published results demonstrate an L1 L2 interaction in the conserve 306-loop of L1 [91]. However, it is not known if this is the only L2 binding site on L1. For investigations, one would need to remove the 306-loop from L1 and analyze the binding affinity to L2 with SPR. In addition, structural data would display where the L1 L2 interaction takes place.

The EL(2)ISA results demonstrated an interaction of L2 and CypA. This interaction must now be further characterized according to binding affinity via SPR or the influence of CypA interaction on the L2 structure via Circular Dichroism. In addition, it is important to investigate whether the presence of L1 has an influence on the L2 CypA interaction and if so, which one? In addition, when L2 and CypA interact in the endosome, does this interaction affect the membrane interaction of L2? This question may be answered by a previously published liposome binding assay [77].

Discussion and Outlook

Overall, we also performed structural analysis to obtain atomic information about the L1/L2 complex. We applied either X-ray crystallography or CryoEM (in collab with Prof. Hipp and Niklas Bayer/Nils Rustmeier). Here we used L1nav pentamers complex with L2 full length or selected peptides. The experiments are still ongoing and thus no results are available yet. These results could provide further information about the L1/L2 interaction on a structural level. All in all, it is important to understand the viral entry of HPV16 also on the biochemical level to be able to answer important open questions.

5   CONCLUSION

The expression and characterization of purified L1nav pentamer and monomer protein as well as L2 is partly published [1, 77]. For further characterization of the L2 L1nav pentamer binding the L2 was labelled and the binding affinity was characterized *in vitro* with purified proteins. The interaction was verified via established anisotropy measurement and reversibility was demonstrated with a K_d of ~ 33 nM. The artificial binding caused by fluorescein was also excluded and the K_d for the competition assay was 13-fold stronger with 2.5 nM. This difference indicates a negative influence of the fluorescein-label on the binding to L1nav pentamer (figure 41).

Furthermore, the interaction was verified and compared via SPR. Here, the binding of L2 and L1nav pentamer was compared under different conditions. In CP buffer pH 7.4 with 0.16 % (w/v) Brij 58 the K_d is 620 nM compared to 36 nM, which is 17-fold stronger at pH 5.5. This is probably related to the stabilization of L1nav pentamer at a lower pH. In addition, a L2ΔL1bs was also tested. This mutant with a K_d of 392 nM demonstrates similar binding affinity to L1nav pentamer as the L2 full-length protein. This suggests a second L1 binding site on L2 (figure 42).

Now, the binding affinity for the L1nav pentamer and L2 is validated in three *in vitro* methods however the SPR probably demonstrates the most reliable binding data.

Nevertheless, the functional role, structure and stoichiometry is still elusive and needs to be investigated further.

6   LITERATURE

1.   Roos, N., et al., *Optimized production strategy of the major capsid protein HPV 16L1 non-assembly variant in E. coli.* Protein Expr Purif, 2020. **175**: p. 105690.

2.   Kubetschek, N., *Die Oberflächenplasmonenresonanzspektroskopie.* Treffpunkt Forschung - Biologie unserer Zeit, 2021.

3.   S.Chen, X., T. Stehle, and S. C.Harrison, *Interaction of polyomavirus internal protein VP2 with the major capsid protein VP1 and implications for participation of VP2 in viral entry.* The EMBO Journal, 1998. **17**(12): p. 3233-3240.

4.   Chen, X.S., et al., *Structure of Small Virus-like Particles Assembled from the L1 Protein of Human Papillomavirus 16.* Molecular Cell, 2000. **5**: p. 557-567.

5.   Stehle, T., et al., *Structure of murine polyomavirus complexed with an oligosaccharide receptor fragment.* Nature, 1994. **369**: p. 160-163.

6.   Buck, C.B., et al., *Arrangement of L2 within the papillomavirus capsid.* J Virol, 2008. **82**(11): p. 5190-7.

7.   Tommasino, M., *The human papillomavirus family and its role in carcinogenesis.* Semin Cancer Biol, 2014. **26**: p. 13-21.

8.   de Villiers, E.M., et al., *Classification of papillomaviruses.* Virology, 2004. **324**(1): p. 17-27.

9.   Bernard, H.U., et al., *Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments.* Virology, 2010. **401**(1): p. 70-9.

10.   *Monographs on the Evaluation of Carcinogenic Risks to Humans.* IARC 2012. **100B**: p. 1-441.

11.   Münger, K. and P.M. Howley, *Human papillomavirus immortalization and transformation functions.* Virus Research, 2002. **89**: p. 213-228.

12.   Muñoz, N.M.D., et al., *Epidemiologic Classification of Human Papillomavirus Types Associated with Cervical Cancer.* The NEW ENGLAND JOURNAL of MEDICINE, 2003. **348;6**(6): p. 518-527.

13.   Zanier, K., et al., *Structural Basis for Hijackingof Cellular LxxLL Motifs byPapillomavirus E6 Oncoproteins.* Science, 2013. **339**: p. 694-698.

14.   de Martel, C., et al., *Worldwide burden of cancer attributable to HPV by site, country and HPV type.* Int J Cancer, 2017. **141**(4): p. 664-670.

15.   Parkin, D.M. and F. Bray, *Chapter 2: The burden of HPV-related cancers.* Vaccine, 2006. **24 Suppl 3**: p. S3/11-25.

16.   Doorbar, J., et al., *The biology and life-cycle of human papillomaviruses.* Vaccine, 2012. **30 Suppl 5**: p. F55-70.

17.   Baker, T.S., et al., *Structures of bovine and human papillomaviruses Analysis by cryoelectron microscopy and three dimensional image reconstruction.* Biophysical Journal, 1991. **60**: p. 1445-1456.

18.   B.L., T., et al., *Novel structural features of bovine papillomavirus capsid revealed by a three-dimensional reconstruction to 9 A resolution.* Nat Struct Biol., 1997 **4**(5): p. 413-420.

19.   Kirnbauer, R., et al., *Papillomavirus L1 major capsid protein self-assembles into virus-like particles that are highly immunogenic.* Medical Sciences, 1992. **89**: p. 12180-12184.

20.   Liddington, R.C., et al., *Structure of simian virus 40 at 3.8-A resolution* Nature, 1991. **354**: p. 278-284.

21.   Müller, M., et al., *Chimeric Papillomavirus-like Particles.* Virology, 1997. **234**: p. 93-111.

22.   Modis, Y., B.L. Trus, and S.C. Harrison, *Atomic model of the papillomavirus capsid.* The EMBO Journal, 2002. **21**: p. 4754-4762.

23.   Wang, J.W. and R.B. Roden, *L2, the minor capsid protein of papillomavirus.* Virology, 2013. **445**(1-2): p. 175-86.

24.   Conway, M.J., et al., *Cross-neutralization potential of native human papillomavirus N-terminal L2 epitopes.* PLoS One, 2011. **6**(2): p. e16405.

25.   Doorbar, J. and P.H. Gallimore, *Identification of Proteins Encoded by the L1 and L2 Open Reading Frames of Papillomavirus 1a.* Journal of Virology, 1987. **61**: p. 2793-2799.

26.   Komly, C.A., et al., *The L2 Open Reading Frame of Human Papillomavirus Type 1a Encodes a Minor Structural Protein Carrying Type-Specific Antigens.* Journal of Virology 1986. **60**: p. 813-816.

27.   Bronnimann, M.P., et al., *Furin Cleavage of L2 during Papillomavirus Infection: Minimal Dependence on Cyclophilins.* J Virol, 2016. **90**(14): p. 6224-34.

28.   Bronnimann, M.P., et al., *A transmembrane domain and GxxG motifs within L2 are essential for papillomavirus infection.* J Virol, 2013. **87**(1): p. 464-73.

29.   Finnen, R.L., et al., *Interactions between Papillomavirus L1 and L2 Capsid Proteins.* Journal of Virology, 2003. **77**(8): p. 4818-4826.

30.   Conway, M.J., et al., *Overlapping and independent structural roles for human papillomavirus type 16 L2 conserved cysteines.* Virology, 2009. **393**(2): p. 295-303.

31.   P. Heino, et al., *Human papillomavirus type 16 capsids expose multiple type-restricted and type-common antigenic epitopes* Journal of General Virology, 1995. **76**: p. 1141-1153.

32.   Day, P.M., et al., *Mechanisms of human papillomavirus type 16 neutralization by l2 cross-neutralizing and l1 type-specific antibodies.* J Virol, 2008. **82**(9): p. 4638-46.

33.   Kines, R.C., et al., *The initial steps leading to papillomavirus infection occur on the basement membrane prior to cell surface binding.* Proc Natl Acad Sci U S A, 2009. **106**(48): p. 20458-63.

34.   Kamper, N., et al., *A membrane-destabilizing peptide in capsid protein L2 is required for egress of papillomavirus genomes from endosomes.* J Virol, 2006. **80**(2): p. 759-68.

35.   Signorelli, C., et al., *Human papillomavirus 9-valent vaccine for cancer prevention: a systematic review of the available evidence.* Epidemiol Infect, 2017. **145**(10): p. 1962-1982.

36.   Jansen, K.U., et al., *Vaccination with yeast-expressed cottontail rabbit papillomavirus (CRPV) virus-like particles protects rabbits from CRPV-induced papilloma formation* Vaccine, 1995. **13**(16): p. 1509-1514.

37.   Wang, R., et al., *Human papillomavirus vaccine against cervical cancer: Opportunity and challenge.* Cancer Lett, 2020. **471**: p. 88-102.

38.   Chintala, R.V. and A. Bhambhani, *HPV VACCNE FORMULATIONS COMPRISING ALUMNUMADJUVANT AND METHODS OF PRODUCING SAME* 2014.

39.   Hu, Y.M., et al., *Safety of an Escherichia coli-expressed bivalent human papillomavirus (types 16 and 18) L1 virus-like particle vaccine: an open-label phase I clinical trial.* Hum Vaccin Immunother, 2014. **10**(2): p. 469-75.

40.   Varsani, A., et al., *Transient expression of Human papillomavirus type 16 L1 protein in Nicotiana benthamiana using an infectious tobamovirus vector.* Virus Res, 2006. **120**(1-2): p. 91-6.

41.   Kim, H.J., et al., *One-step chromatographic purification of human papillomavirus type 16 L1 protein from Saccharomyces cerevisiae.* Protein Expr Purif, 2010. **70**(1): p. 68-74.

42.   Bazan, S.B., et al., *Expression and characterization of HPV-16 L1 capsid protein in Pichia pastoris.* Arch Virol, 2009. **154**(10): p. 1609-17.

43.   Cox, M.M., *Recombinant protein vaccines produced in insect cells.* Vaccine, 2012. **30**(10): p. 1759-66.

44.   Schellenbacher, C., et al., *HPV-Impfstoffe – zugelassene Vakzinen und experimenteller RG1-VLP-Impfstoff der nächsten Generation.* hautnah, 2021. **20**(3): p. 155-160.

45. Schellenbacher, C., R. Roden, and R. Kirnbauer, *Chimeric L1-L2 virus-like particles as potential broad-spectrum human papillomavirus vaccines.* J Virol, 2009. **83**(19): p. 10085-95.

46. Gambhira, R., et al., *A protective and broadly cross-neutralizing epitope of human papillomavirus L2.* J Virol, 2007. **81**(24): p. 13927-31.

47. Schellenbacher, C., et al., *Efficacy of RG1-VLP vaccination against infections with genital and cutaneous human papillomaviruses.* J Invest Dermatol, 2013. **133**(12): p. 2706-2713.

48. Schellenbacher, C., R.B.S. Roden, and R. Kirnbauer, *Developments in L2-based human papillomavirus (HPV) vaccines.* Virus Res, 2017. **231**: p. 166-175.

49. *Aktuelles aus der KV-Impfsurveillance–Impfquoten ausgewählter Schutzimpfunge in Deutschland.* Epidemiologisches Bulletin - RKI, 2018.

50. Pelkmans, L. and A. Helenius, *Insider information: what viruses tell us about endocytosis.* Current Opinion in Cell Biology, 2003. **15**(4): p. 414-422.

51. Day, P.M. and J.T. Schiller, *Chapter 12: Early events in the papillomavirus life cycle. from Papillomavirus research: From natural history to vaccines and beyond. From Papillomavirus research: from natural history to vaccines and beyond.* Caister Academic Press Campo SM, 2006: p. 175-192.

52. Horvath, C.A.J., et al., *Mechanisms of cell entry by human papillomaviruses: an overview.* Virology Journal, 2010. **7**.

53. Surviladze, Z., A. Dziduszko, and M.A. Ozbun, *Essential roles for soluble virion-associated heparan sulfonated proteoglycans and growth factors in human papillomavirus infections.* PLoS Pathog, 2012. **8**(2): p. e1002519.

54. Johnson, K.M., et al., *Role of heparan sulfate in attachment to and infection of the murine female genital tract by human papillomavirus.* J Virol, 2009. **83**(5): p. 2067-74.

55. Abban, C.Y. and P.I. Meneses, *Usage of heparan sulfate, integrins, and FAK in HPV16 infection.* Virology, 2010. **403**(1): p. 1-16.

56. Knappe, M., et al., *Surface-exposed amino acid residues of HPV16 L1 protein mediating interaction with cell surface heparan sulfate.* J Biol Chem, 2007. **282**(38): p. 27913-22.

57. Dasgupta, J., et al., *Structural basis of oligosaccharide receptor recognition by human papillomavirus.* J Biol Chem, 2011. **286**(4): p. 2617-24.

58. Schowalter, R.M., D.V. Pastrana, and C.B. Buck, *Glycosaminoglycans and sialylated glycans sequentially facilitate Merkel cell polyomavirus infectious entry.* PLoS Pathog, 2011. **7**(7): p. e1002161.

59. Selinka, H.C., et al., *Further evidence that papillomavirus capsids exist in two distinct conformations.* J Virol, 2003. **77**(24): p. 12961-7.

60. Richards, R.M., et al., *Cleavage of the papillomavirus minor capsid protein, L2, at a furin consensus site is necessary for infection.* Proc Natl Acad Sci U S A, 2006. **103**(5): p. 1522-7.

61. Day, P.M., D.R. Lowy, and J.T. Schiller, *Heparan sulfate-independent cell binding and infection with furin-precleaved papillomavirus capsids.* J Virol, 2008. **82**(24): p. 12565-8.

62. Bienkowska-Haba, M., et al., *Cyclophilins facilitate dissociation of the human papillomavirus type 16 capsid protein L1 from the L2/DNA complex following virus entry.* J Virol, 2012. **86**(18): p. 9875-87.

63. Pakula, R., et al., *Syndecan-1/CD147 association is essential for cyclophilin B-induced activation of p44/42 mitogen-activated protein kinases and promotion of cell adhesion and chemotaxis.* Glycobiology, 2007. **17**(5): p. 492-503.

64. Fischer, G., et al., *Cyclophilin and peptidyl-prolyl cis-trans isomerase are probably identical proteins.* Nature, 1989. **337**: p. 476-478.

65. Cerqueira, C., et al., *Kallikrein-8 Proteolytically Processes Human Papillomaviruses in the Extracellular Space To Facilitate Entry into Host Cells.* J Virol, 2015. **89**(14): p. 7038-52.

66. Samperio Ventayol, P. and M. Schelhaas, *Fluorescently Labeled Human Papillomavirus Pseudovirions for Use in Virus Entry Experiments.* Curr Protoc Microbiol, 2015. **37**: p. 14B 4 1-22.

67. EVANDER, M., et al., *Identification of the alpha 6 Integrin as a Candidate Receptor for Papillomaviruses.* Journal of Virology, 1997. **71**(3): p. 2449-2456.

68. Sonnenberg, A., et al., *Integrin alpha 6/beta 4 complex is located in hemidesmosomes, suggesting a major role in epidermal cell-basement membrane adhesion.* The Journal of Cell Biology, 1991. **113**(4): p. 907-917.

69. Scheffer, K.D., et al., *Tetraspanin CD151 mediates papillomavirus type 16 endocytosis.* J Virol, 2013. **87**(6): p. 3435-46.

70. Dziduszko, A. and M.A. Ozbun, *Annexin A2 and S100A10 regulate human papillomavirus type 16 entry and intracellular trafficking in human keratinocytes.* J Virol, 2013. **87**(13): p. 7502-15.

71. Wustenhagen, E., et al., *The Cytoskeletal Adaptor Obscurin-Like 1 Interacts with the Human Papillomavirus 16 (HPV16) Capsid Protein L2 and Is Required for HPV16 Endocytosis.* J Virol, 2016. **90**(23): p. 10629-10641.

72. Kawana, Y., et al., *Human papillomavirus type 16 minor capsid protein l2 N-terminal region containing a common neutralization epitope binds to the cell surface and enters the cytoplasm.* J Virol, 2001. **75**(5): p. 2331-6.

73. Spoden, G., et al., *Human papillomavirus types 16, 18, and 31 share similar endocytic requirements for entry.* J Virol, 2013. **87**(13): p. 7765-73.

74. Pereira, R., Hitzeroth, II, and E.P. Rybicki, *Insights into the role and function of L2, the minor capsid protein of papillomaviruses.* Arch Virol, 2009. **154**(2): p. 187-97.

75. Selinka, H.C., T. Giroglou, and M. Sapp, *Analysis of the infectious entry pathway of human papillomavirus type 33 pseudovirions.* Virology, 2002. **299**(2): p. 279-287.

76. Stephen DiGiuseppe, M.B.-H., Lucile G. M. Guion,, Timothy R. Keiffer, and M. Sapp, *Human Papillomavirus Major Capsid Protein L1 Remains Associated with the Incoming Viral Genome throughout the Entry Process.* Journal of Virology, 2017. **91**(16).

77. Breiner, B., et al., *Refolding and in vitro characterization of human papillomavirus 16 minor capsid protein L2.* Biol Chem, 2019. **400**(4): p. 513-522.

78. Florin, L., et al., *Identification of a dynein interacting domain in the papillomavirus minor capsid protein l2.* J Virol, 2006. **80**(13): p. 6691-6.

79. Sapp, M.J., *HPV virions hitchhike a ride on retromer complexes.* Proc Natl Acad Sci U S A, 2013. **110**(18): p. 7116-7.

80. Popa, A., et al., *Direct binding of retromer to human papillomavirus type 16 minor capsid protein L2 mediates endosome exit during viral infection.* PLoS Pathog, 2015. **11**(2): p. e1004699.

81. Lipovsky, A., et al., *Genome-wide siRNA screen identifies the retromer as a cellular entry factor for human papillomavirus.* Proc Natl Acad Sci U S A, 2013. **110**(18): p. 7452-7.

82. Laniosz, V., et al., *Human papillomavirus type 16 infection of human keratinocytes requires clathrin and caveolin-1 and is brefeldin a sensitive.* J Virol, 2009. **83**(16): p. 8221-32.

83. DiGiuseppe, S., et al., *The nuclear retention signal of HPV16 L2 protein is essential for incoming viral genome to transverse the trans-Golgi network.* Virology, 2014. **458-459**: p. 93-105.

84. Schneider, M.A., et al., *Identification of the dynein light chains required for human papillomavirus infection.* Cell Microbiol, 2011. **13**(1): p. 32-46.

85. Pyeon, D., et al., *Establishment of human papillomavirus infection requires cell cycle progression.* PLoS Pathog, 2009. **5**(2): p. e1000318.

86. Mamoor, S., et al., *The high risk HPV16 L2 minor capsid protein has multiple transport signals that mediate its nucleocytoplasmic traffic.* Virology, 2012. **422**(2): p. 413-24.

87. Lowe, J., et al., *Evolutionary and structural analyses of alpha-papillomavirus capsid proteins yields novel insights into L2 structure and interaction with L1.* Virol J, 2008. **5**: p. 150.

88. DAN H. BAROUCH, S.C.H., *Interactions among the Major and Minor Coat Proteins of Polyomavirus.* Journal of Virology, 1994. **68**(6): p. 3982-3989.

89.   Belnap, D.M., et al., *Conserved features in papillomavirus and polyomavirus capsids.* J Mol Biol, 1996. **259**(2): p. 249-63.

90.   Gabler, F., et al., *Protein Sequence Analysis Using the MPI Bioinformatics Toolkit.* Curr Protoc Bioinformatics, 2020. **72**(1): p. e108.

91.   Goetschius, D.J., et al., *High resolution cryo EM analysis of HPV16 identifies minor structural protein L2 and describes capsid flexibility.* Sci Rep, 2021. **11**(1): p. 3498.

92.   Schmohl, L. and D. Schwarzer, *Sortase-mediated ligations for the site-specific modification of proteins.* Curr Opin Chem Biol, 2014. **22**: p. 122-8.

93.   Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nat Methods, 2009. **6**(5): p. 343-5.

94.   Potapov, V., et al., *Comprehensive Profiling of Four Base Overhang Ligation Fidelity by T4 DNA Ligase and Application to DNA Assembly.* ACS Synth Biol, 2018. **7**(11): p. 2665-2674.

95.   Anton, B.P., et al., *Complete Genome Sequence of Escherichia coli BE104, an MC4100 Derivative Lacking the Methionine Reductive Pathway.* Microbiol Resour Announc, 2019. **8**(28).

96.   Pan, D., et al., *Enhanced expression of soluble human papillomavirus L1 through coexpression of molecular chaperonin in Escherichia coli.* Protein Expr Purif, 2016. **120**: p. 92-8.

97.   Georgescauld, F., et al., *GroEL/ES chaperonin modulates the mechanism and accelerates the rate of TIM-barrel domain folding.* Cell, 2014. **157**(4): p. 922-934.

98.   Weissman, J.S., et al., *GroEL-Mediated Protein Folding Proceeds by Multiple Rounds of Binding and Release of Nonnative Forms* Cell 1994. **78**: p. 693-702.

99.   Li, Z., M. Nimtz, and U. Rinas, *The metabolic potential of Escherichia coli BL21 in defined and rich medium.* Microbial Cell Factories, 2014.

100.   H.E.Reiling, H.Laurila, and A.Fiechter, *Mass culture of Escherichia coli: Medium development for low and high density cultivation of Escherichia coli B/r in minimal and complex media* Journal of Biotechnology, 1985. **2**: p. 191-206.

101.   SAUER, U., et al., *Metabolic Flux Ratio Analysis of Genetic and Environmental Modulations of Escherichia coli Central Carbon Metabolism.* Journal of Bacteriology 1999. **181**(21): p. 6679-6688.

102.   Chen, Y., et al., *Human papillomavirus L1 protein expressed in Escherichia coli self-assembles into virus-like particles that are highly immunogenic.* Virus Res, 2016. **220**: p. 97-103.

103.   Zheng, D.D., et al., *In vitro monitoring of the formation of pentamers from the monomer of GST fused HPV 16 L1.* Chem Commun (Camb), 2013. **49**(76): p. 8546-8.

104.   Yuan, H., et al., *Immunization with a pentameric L1 fusion protein protects against papillomavirus infection.* J Virol, 2001. **75**(17): p. 7848-53.

105.   Bishop, B., J. Dasgupta, and X.S. Chen, *Structure-based engineering of papillomavirus major capsid l1: controlling particle assembly.* Virol J, 2007. **4**: p. 3.

106.   Butt, T.R., et al., *SUMO fusion technology for difficult-to-express proteins.* Protein Expr Purif, 2005. **43**(1): p. 1-9.

107.   Schuck, P., *Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling.* Biophysical Journal 2000. **78**: p. 1606-1619.

108.   J.E., T., C. S., and W. D.S., *Expression and Purification of Soluble His6-Tagged TEV Protease. In: Doyle S.A. (eds) High Throughput Protein Expression and Purification. Methods in Molecular Biology.* Humana Press, 2009. **498**.

109.   Kruger, R.G., et al., *Analysis of the Substrate Specificity of the Staphylococcus aureus Sortase Transpeptidase SrtA.* Biochemistry, 2004. **43**(6): p. 1541-1551.

110.   Schneewind, O., A. Fowler, and K.F. Faull, *Structure of the Cell Wall Anchor of Surface Proteins in Staphylococcus aureus.* Science 1995. **268**.

111.   Popp, M.W., et al., *Sortagging: a versatile method for protein labeling.* Nat Chem Biol, 2007. **3**(11): p. 707-8.

112.   JC., B. and D. F., *Expression of Recombinant Alkaline Phosphatase Conjugates in Escherichia coli. In: Balbás P., Lorence A. (eds) Recombinant Gene Expression. Methods in Molecular Biology.* Humana Press, 2004. **267**.

113.   Tudos, A.J. and R.B.M. Schasfoort, *Chapter 1. Introduction to Surface Plasmon Resonance*, in *Handbook of Surface Plasmon Resonance*. 2008. p. 1-14.

114.   Karlsson, R., et al., *Practical aspects concerning direct detection of low molecular weight analytes using BIACORE. .* Methodology 1–3, 1997.

115.   VASINA, J.A. and F. BANEYX, *Recombinant Protein Expression at Low Temperatures under the Transcriptional Control of the Major Escherichia coli Cold Shock Promoter cspA.* APPLIED AND ENVIRONMENTAL MICROBIOLOGY, 1996. **62**: p. 1444-1447.

116.   *Tips for Optimizing Protein Expression and Purification.* ROCKLAND IMMUNOCHEMICALS, INC., 2021.

117.   Machida, S., et al., *Overproduction of beta-glucosidase in active form by an Escherichia coli system coexpressing the chaperonin GroEL/ES.* FEMS Microbiology Letters, 1998. **159**: p. 41-46.

118.   Cos, O., et al., *Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast Pichia pastoris under different promoters: a review.* Microb Cell Fact, 2006. **5**: p. 17.

119. SASAGAWA, T., et al., *Synthesis and Assembly of Virus-like Particles of Human Papillomaviruses Type 6 and Type 16 in Fission Yeast Schizosaccharomyces pombe* Virology, 1995. **206**: p. 126-135.

120. Buonamassa, D.T., et al., *Yeast coexpression of human papillomavirus types 6 and 16 capsid proteins.* Virology, 2002. **293**(2): p. 335-44.

121. Buck, C.B. and C.D. Thompson, *Production of papillomavirus-based gene transfer vectors.* Curr Protoc Cell Biol, 2007. **Chapter 26**: p. Unit 26 1.

122. Fay, A., et al., *The positively charged termini of L2 minor capsid protein required for bovine papillomavirus infection function separately in nuclear import and DNA binding.* J Virol, 2004. **78**(24): p. 13447-54.

123. Zhang, J., et al., *Soluble expression, rapid purification, and characterization of human interleukin-24 (IL-24) using a MBP-SUMO dual fusion system in Escherichia coli.* Appl Microbiol Biotechnol, 2015. **99**(16): p. 6705-13.

124. Xu, J., et al., *Self-assembly of virus-like particles of canine parvovirus capsid protein expressed from Escherichia coli and application as virus-like particle vaccine.* Appl Microbiol Biotechnol, 2014. **98**(8): p. 3529-38.

125. Knecht, S., et al., *Oligohis-tags: mechanisms of binding to Ni2+-NTA surfaces.* J Mol Recognit, 2009. **22**(4): p. 270-9.

126. *Binding Kinetics of Glutathione-S-Transferase (GST) Tagged Proteins Using OpenSPR.* Nicoya, 2019.

127. Xiaojiang S. Chen, G.C., Stephen C. Harrison, Robert L. Garcea, *Papillomavirus Capsid Protein Expression in Escherichia coli: Purification and Assembly of HPV11 and HPV16 L1.* J. Mol. Biol., 2001. **307**: p. 173-182.

128. Harper, S. and D.W. Speicher, *Purification of proteins fused to glutathione S-transferase.* Methods Mol Biol, 2011. **681**: p. 259-80.

129. Malakhov, M.P., et al., *SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins.* Journal of Structural and Functional Genomics, 2004. **5**: p. 75-86.

130. XIAN WEN JIN, et al., *Identification of L2 Open Reading Frame Gene Products of Bovine Papillomavirus Type 1 Using Monoclonal Antibodies* J. gen. Virol. , 1989. **70**: p. 1133-1140.

131. RIPPE, R.A. and W.J. MEINKE, *Identification and Characterization of the BPV-2 L2 Protein* Virology, 1989. **171**: p. 298-301.

132. Raran-Kurussi, S., et al., *Removal of Affinity Tags with TEV Protease.* Methods Mol Biol, 2017. **1586**: p. 221-230.

133. Tsumoto, K., et al., *Role of Arginine in Protein Refolding, Solubilization, and Purification.* Biotechnology Progress, 2008. **20**: p. 1301–1308.

134.   Baynes, B.M., D.I.C. Wang, and B.L. Trout, *Role of Arginine in the Stabilization of Proteins against Aggregation.* Biochemistry, 2005. **44**: p. 4919-4925.

135.   Tischer, A., et al., *L-arginine hydrochloride increases the solubility of folded and unfolded recombinant plasminogen activator rPA.* Protein Sci, 2010. **19**(9): p. 1783-95.

136.   Yue, L., et al., *Brij-58, a potential injectable protein-stabilizer used in therapeutic protein formulation.* European Journal of Pharmaceutics and Biopharmaceutics, 2020. **146**: p. 73-83.

137.   Patist, A., et al., *On the measurement of critical micelle concentrations of pure and technical-grade nonionic surfactants.* Journal of Surfactants and Detergents volume, 2000. **3**: p. 53–58

138.   Schmid, F.X. and T.E. Creighton, *Protein Structure: A Practical Approach.* Oxford; New York : IRL Press at Oxford University Press, 1989: p. 251-286.

139.   Le, D.T. and K.M. Muller, *In Vitro Assembly of Virus-Like Particles and Their Applications.* Life (Basel), 2021. **11**(4).

140.   Li, T.C., et al., *Characterization of self-assembled virus-like particles of Merkel cell polyomavirus.* PLoS One, 2015. **10**(2): p. e0115646.

141.   Xiaojiang S. Chen, et al., *Structure of Small Virus-like Particles Assembled from the L1 Protein of Human Papillomavirus 16.* Molecular Cell, 2000. **5**: p. 557–567.

142.   Arakawa, T., et al., *Suppression of protein interactions by arginine: a proposed mechanism of the arginine effects.* Biophys Chem, 2007. **127**(1-2): p. 1-8.

143.   Pollard, T.D., *A guide to simple and informative binding assays.* Mol Biol Cell, 2010. **21**(23): p. 4061-7.

144.   Invitrogen, *Technical Resource Guide Fluorescence Polarization.* **Fourth Edition**.

145.   Shirano, Y. and D. Shibata, *Low temperature cultivation of Escherichia coli carrying a rice lipoxygenase L-2 cDNA produces a soluble and active enzyme at a high level* FEBS, 1990. **271**: p. 128-130.

146.   Kataeva, I., et al., *Improving Solubility of Shewanella oneidensis MR-1 and Clostridium thermocellum JW-20 Proteins Expressed into Esherichia coli.* Journal of Proteome research, 2005: p. 1942-1951.

147.   Volonte, F., et al., *Optimization of glutaryl-7-aminocephalosporanic acid acylase expression in E. coli.* Protein Expr Purif, 2008. **61**(2): p. 131-7.

148.   Piserchio, A., R. Ghose, and D. Cowburn, *Optimized bacterial expression and purification of the c-Src catalytic domain for solution NMR studies.* J Biomol NMR, 2009. **44**(2): p. 87-93.

149. Chou, C.P., *Engineering cell physiology to enhance recombinant protein production in Escherichia coli.* Appl Microbiol Biotechnol, 2007. **76**(3): p. 521-32.

150. Vasina, J.A. and F. Baneyx, *Expression of Aggregation Prone Recombinant Proteins at Low Temperatures: A Comparative Study of the Escherichia coli cspA and tac Promoter Systems.* Protein Expression and Purification 1997. **9**: p. 211–218.

151. Sahdev, S., S.K. Khattar, and K.S. Saini, *Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies.* Mol Cell Biochem, 2008. **307**(1-2): p. 249-64.

152. Spiess, C., A. Beil, and M. Ehrmann, *A Temperature-Dependent Switch from Chaperone to Protease in a Widely Conserved Heat Shock Protein.* Cell, 1999. **97**: p. 339–347.

153. Hunke, S. and J.M. Betton, *Temperature effect on inclusion body formation and stress response in the periplasm of Escherichia coli.* Mol Microbiol, 2003. **50**(5): p. 1579-89.

154. Pinsach, J., et al., *Influence of process temperature on recombinant enzyme activity in Escherichia coli fed-batch cultures.* Enzyme and Microbial Technology, 2008. **43**(7): p. 507-512.

155. Francis, D.M. and R. Page, *Strategies to optimize protein expression in E. coli.* Curr Protoc Protein Sci, 2010. **Chapter 5**: p. Unit 5 24 1-29.

156. Ulrich, H.D., *SUMO Protocols.* Humana Press, 2009. **497**.

157. Reverter, D. and C.D. Lima, *Preparation of SUMO proteases and kinetic analysis using endogenous substrates.* Methods Mol Biol, 2009. **497**: p. 225-39.

158. Simon, C., *Virus-analoge Partikel des Polyomavirus: Produktion in Hefe und biophysikalische Charakterisierung.* Naturwissenschaftlichen Fakultät I Biowissenschaften der Martin-Luther-Universität Halle-Wittenberg, 2012.

159. Krause, M., R. Rudolph, and E. Schwarz, *The non-ionic detergent Brij 58P mimics chaperone ejects.* FEBS, 2002. **532**: p. 253-255.

160. Clarke, M.A., et al., *A large, population-based study of age-related associations between vaginal pH and human papillomavirus infection.* BMC Infect Dis, 2012. **12**: p. 33.

161. Wei, M., et al., *N-terminal truncations on L1 proteins of human papillomaviruses promote their soluble expression in Escherichia coli and self-assembly in vitro.* Emerg Microbes Infect, 2018. **7**(1): p. 160.

162. Sarpong, K. and R. Bose, *Efficient sortase-mediated N-terminal labeling of TEV protease cleaved recombinant proteins.* Anal Biochem, 2017. **521**: p. 55-58.

163. Kobashigawa, Y., et al., *Attachment of an NMR-invisible solubility enhancement tag using a sortase-mediated protein ligation method.* J Biomol NMR, 2009. **43**(3): p. 145-50.

164. Joos, P. and J.K. Hunsel, *Adsorption kinetics of micellar Brij 58 solutions.* Chemistry, 1988.

165. RYFFEL, B., et al., *Distribution of the cyclosporine binding protein cyclophilin in human tissues.* Immunology, 1991. **72**: p. 399-404.

166. Bienkowska-Haba, M., H.D. Patel, and M. Sapp, *Target cell cyclophilins facilitate human papillomavirus type 16 infection.* PLoS Pathog, 2009. **5**(7): p. e1000524.

167. Owicki, J.C., *Fluorescence Polarization and Anisotropy in High Throughput Screening: Perspectives and Primer.* Journal of Biomolecular Screening, 2008. **5**: p. 297-306.

168. *Establishing and optimizing a fluorescence polarization assay.* Molecular Devices. **1/21 2355B**.

169. Borst, J.W., et al., *Effects of Refractive Index and Viscosity on Fluorescence and Anisotropy Decays of Enhanced Cyan and Yellow Fluorescent Proteins.* Journal of Fluorescence, 2005. **15**(2): p. 153-160.

170. Tanford, C. and J.G. Buzzell, *The Viscosity of Aqueous Solutions of Bovine Serum Albumin between pH 4.3 and 10.5.* The Journal of Physical Chemistry, 1956. **60**(2): p. 225-231.

171. Dmochewitz, L., et al., *Role of CypA and Hsp90 in membrane translocation mediated by anthrax protective antigen.* Cell Microbiol, 2011. **13**(3): p. 359-73.

172. Kaiser, E., et al., *Cyclophilin A facilitates translocation of the Clostridium botulinum C2 toxin across membranes of acidified endosomes into the cytosol of mammalian cells.* Cell Microbiol, 2009. **11**(5): p. 780-95.

173. Campos, S.K., *Subcellular Trafficking of the Papillomavirus Genome during Initial Infection: The Remarkable Abilities of Minor Capsid Protein L2.* Viruses, 2017. **9**(12).

7 SUPPLEMENTS

7.1 Materials and Methods

7.1.1 Primers

Table S1: Used constructs and primers

Construct	Name	Sequence 5´-3´
His$_6$-SUMO-L1nav	Vector_pETSUMOadapt_f Vector_pETSUMOadapt_r Pet_coecpent_r Pet_coecpent_f	agacaagcttaggtatttattcggcgcaa accaatctgttctctgtgagcctcaa ttgcgccgaataaatacctaagcttgtctttacaggcccgcttgcag ttgaggctcacagagaacagattggtggtagcgcggttgttagc
His$_6$-SUMO-L1	Vector_pETSUMOadapt_f Vector_pETSUMOadapt_r	agacaagcttaggtatttattcggcgcaa accaatctgttctctgtgagcctcaa
His$_{10}$-L1	Cloned by Julian Gidda	-
His$_{10}$-L1nav	Cloned by Julian Gidda	-
L1-His$_6$	Cloned by Julian Gidda	-
GST-L1nav	gift Dr. Bärbel Blaum	-
His$_6$-TEV-GGGL2	Bought by GenScript	-
His$_6$-TEV-GGGL2$\triangle$L1bs	L2 ΔL1bs rev L2 ΔL1bs fwd	CACCAGCGGAATGTTG CGTAAGCGTCGTAAGC

7.1.2 Proteins

Table S2: Used purified protein

	Protein	Tag	Purification
	SUMO-Protease	His$_6$-Tag	Affinity Chromatography
	SortaseA	His$_6$-Tag	
	TEV-Protease	His$_6$-Tag	
HPV16L1	His$_6$-SUMO-L1nav	His$_6$-Tag	
	His$_6$-L1nav	His$_6$-Tag	
	GST-L1nav	GST-Tag	
	His$_{10}$-L1	His$_6$-Tag	
	L1-His$_6$	His$_6$-Tag	
	His$_6$-SUMO-L1	His$_6$-Tag	
HPV16L2	His$_6$-TEV-GGGL2	His$_6$-Tag	
	His$_6$-TEV-GGGL2$\triangle$L1bs	His$_6$-Tag	

7.2 Results

7.2.1 Temperature
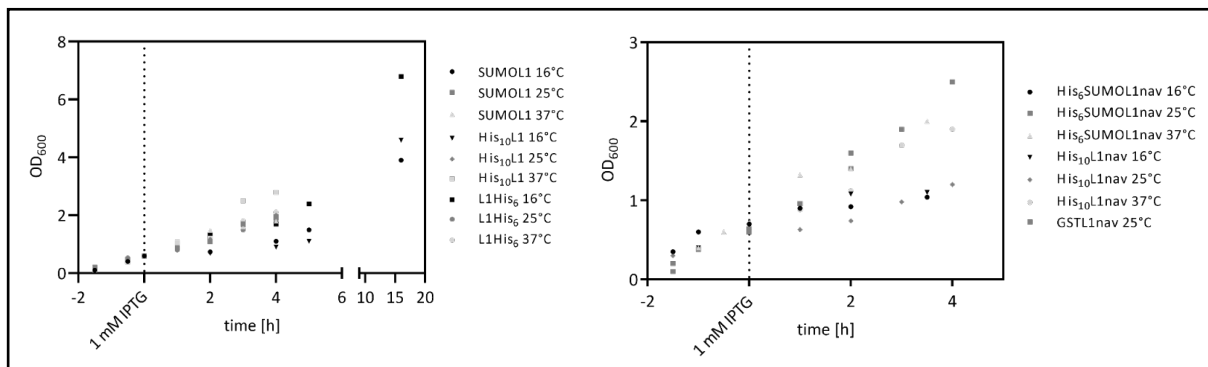
Figure S1: The growth curve of BL21(DE3) bacteria with different L1 constructs at different expression temperatures is shown. L1 constructs Cells were grown at 37°C up to an OD_{600} = 0.5 and afterwards the expression temperature was set to 16 °C, 25 °C and 37 °C. To start protein expression for L1nav by adding 1 mM IPTG. The bacteria were cultivated for 4h and the growth was monitored by optical density (OD_{600}).

7.2.2 GroEL/ES
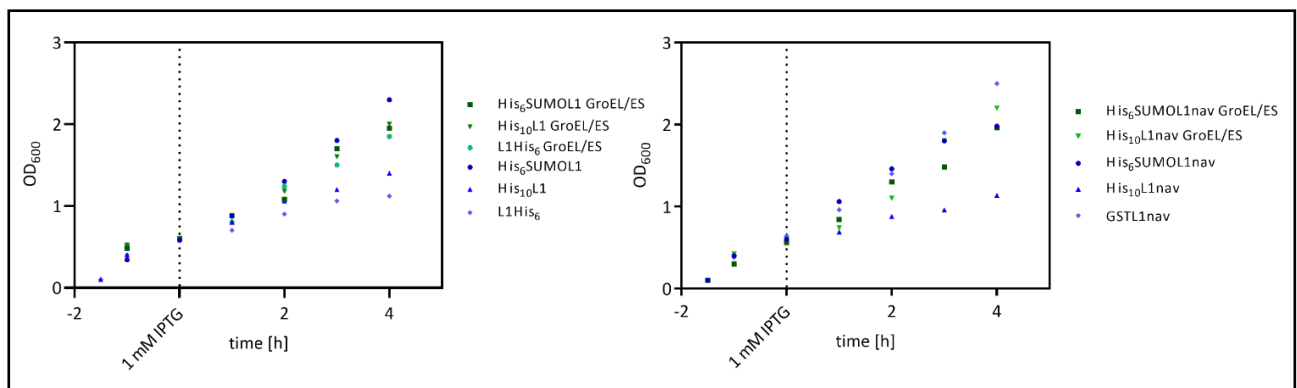
Figure S2: The growth curve of BL21(DE3)/BL21(DE3)pGro7 bacteria with different L1 constructs is shown. L1 constructs Cells were grown at 37°C up to an OD_{600} = 0.5 and afterwards the expression temperature was set to 25 °C To start protein expression for L1nav by adding 1 mM IPTG and 2 mg/ml for GroEL/ES co-expression. The bacteria were cultivated for 4h and the growth was monitored by optical density (OD_{600}).

7.2.3 Efficiency of cell lysis
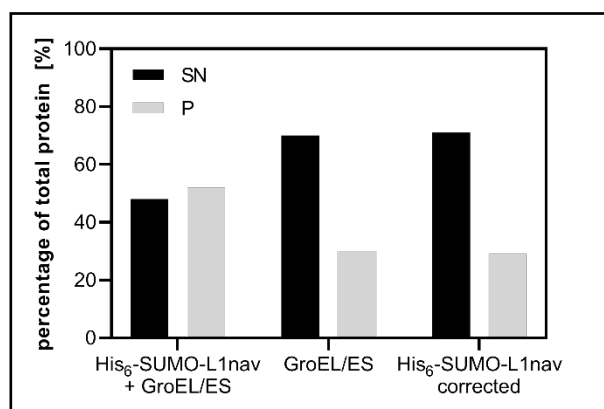
Figure S3: Correction of the detected amount of L1nav in the soluble fraction by cell lysis efficiency. This is shown for His$_6$-SUMO-L1nav protein at an expression temperature of 25°C. The cells were lysed and centrifuged to separate the soluble fraction (supernatant, SN) from the insoluble fraction (pellet, P). Supernatant and pellet were equally loaded on SDS-PAGE corresponding to 210 µg dry biomass. L1nav and GroEL were specifically detected after Western blot using a fluorescent labeled antibody with (Li-Cor System) and quantified densitometrically with the Li-Core Odyssey Fc software. Cell lysis efficiency was calculated according to the relative amount of GroEL in the supernatant and pellet assuming that GroEL is a highly soluble protein [1].

$$\%relative\ L1nav\ (100\ \%\ lysis) = \frac{\%relative\ L1nav\ in\ SN}{\%relative\ GroEL\ in\ SN} * 100$$

7.2.4 Fermentation Process
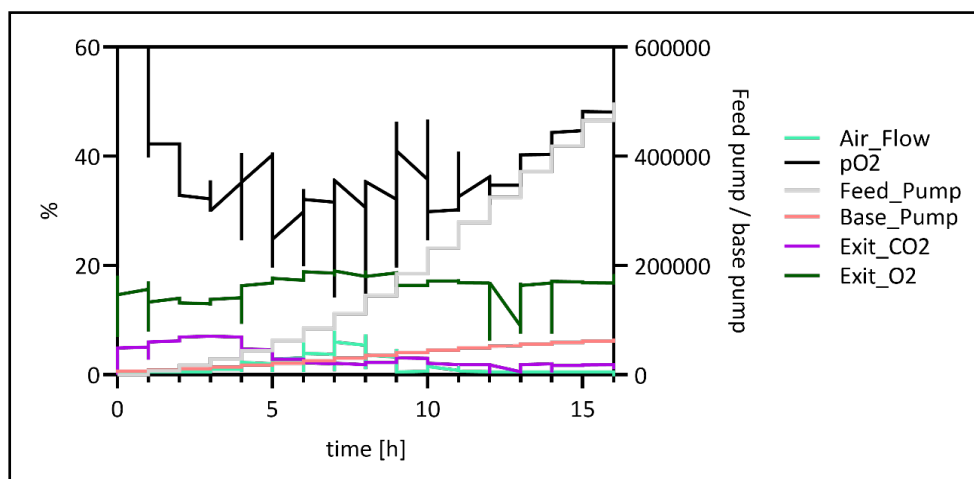
Figure S4: Different important parameters during the process-controlled feed-batch fermentation. The process is controlled by a cascade. The oxygen content with 30 % ± 10 % (pO2) serves as the parameter to be controlled on the first level via air flow and pure oxygen flow. The pO2 curve is shown in black and the values vary between 20 % and 40 % as set. The airflow is predicted in green and an increase with a maximum is shown at round about 7 - 8h. Depending on this, the value of exitO2 present in dark green remains fairly constant over the entire process. However, the value of exitCO2 predicted in pink is significantly higher in the first 5 hours and then decreases steadily over the process. The pumping capacity of the feed pump and the base pump is shown in gray and light red. The feed pump shows an exponential curve, and the base pump shows a continuously rising curve. All parameters shown indicate growth of bacteria over time.

7.2.5 Purification – L1 constructs
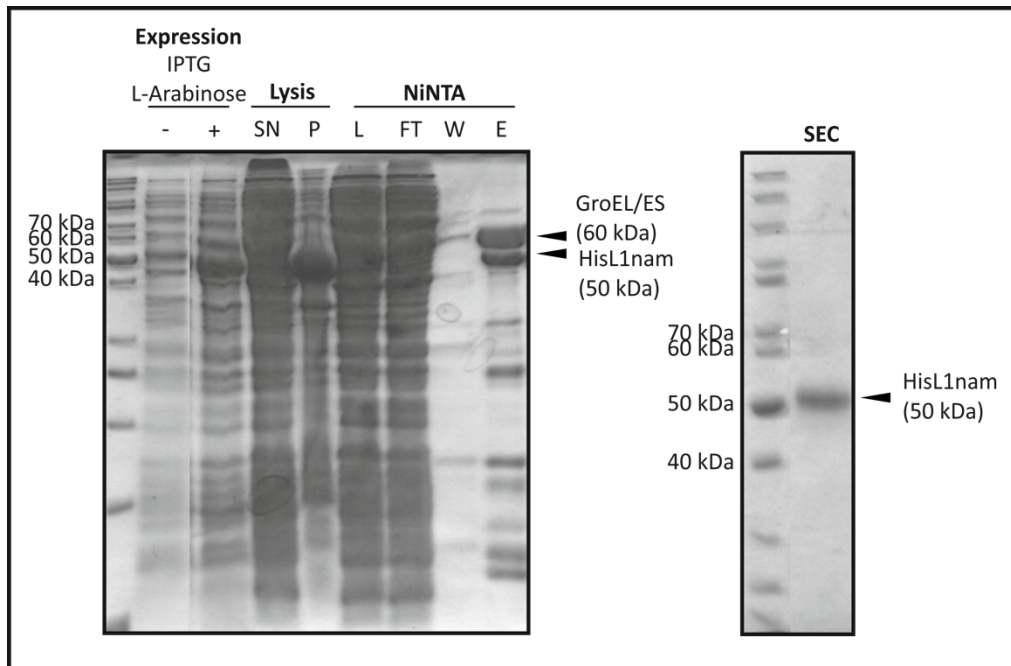
Figure S5: History of purification of His₁₀-L1nav construct analyzed via SDS-PAGE. (A): Expression without (-) and with (+) induction of recombinant L1nav and GroEL/ES protein for 4 h at 20 °C; lysis: insoluble pellet (P) and soluble supernatant (SN) were separated by centrifugation, NiNTA: consists of load (L), flow-through (FT), wash (W) and protein elution (E), size exclusion result from purified protein.
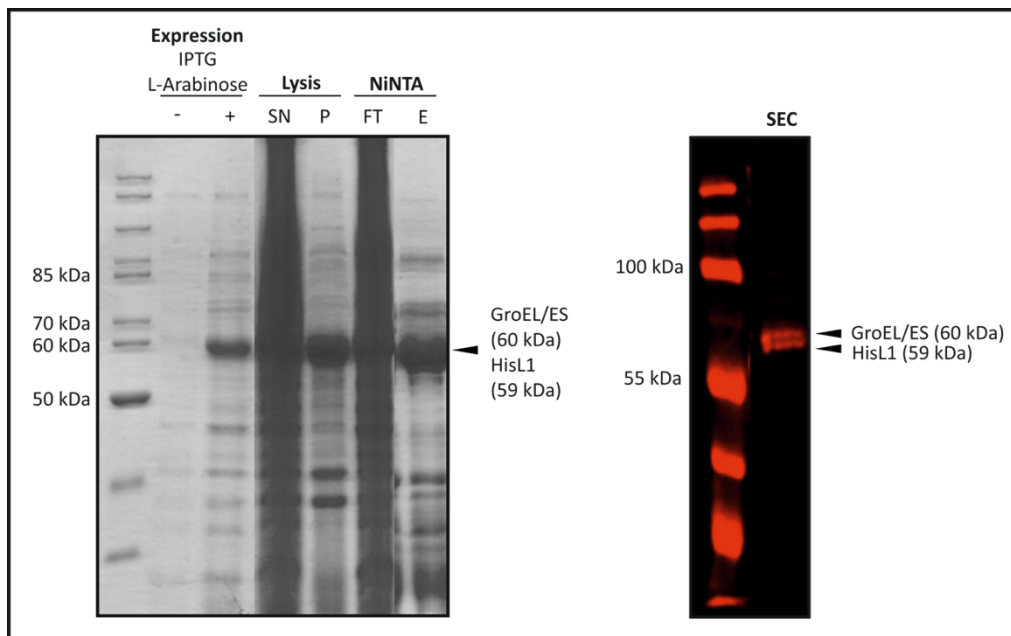
Figure S6: History of purification of His₁₀-L1 construct analyzed via SDS-PAGE and Western Blot. (A): Expression without (-) and with (+) induction of recombinant L1nav and GroEL/ES protein for 4 h at 20 °C; lysis: insoluble pellet (P) and soluble supernatant (SN) were separated by centrifugation, flow-through (FT) and protein elution (E), size exclusion result from purified protein.
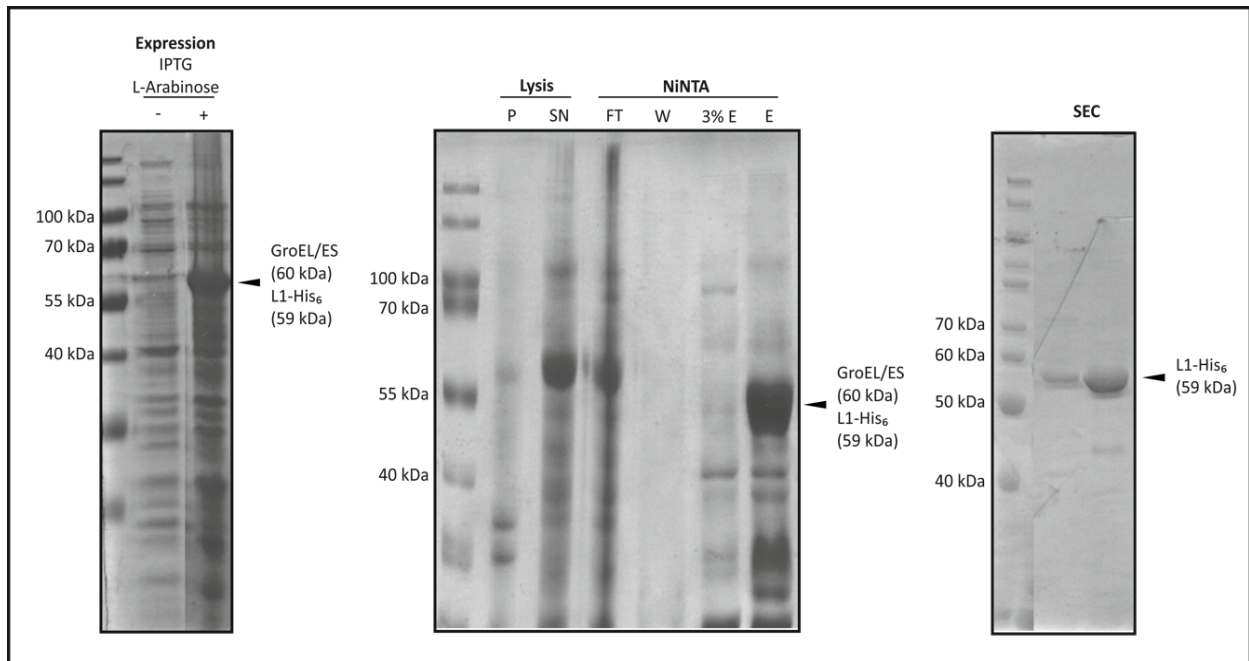
Figure S7: History of purification of L1-His₆ construct analyzed via SDS-PAGE. (A): Expression without (-) and with (+) induction of recombinant L1nav and GroEL/ES protein for 4 h at 20 °C; lysis: insoluble pellet (P) and soluble supernatant (SN) were separated by centrifugation, flow-through (FT), wash (W), 3% pre-elution and protein elution (E), size exclusion result from purified protein.
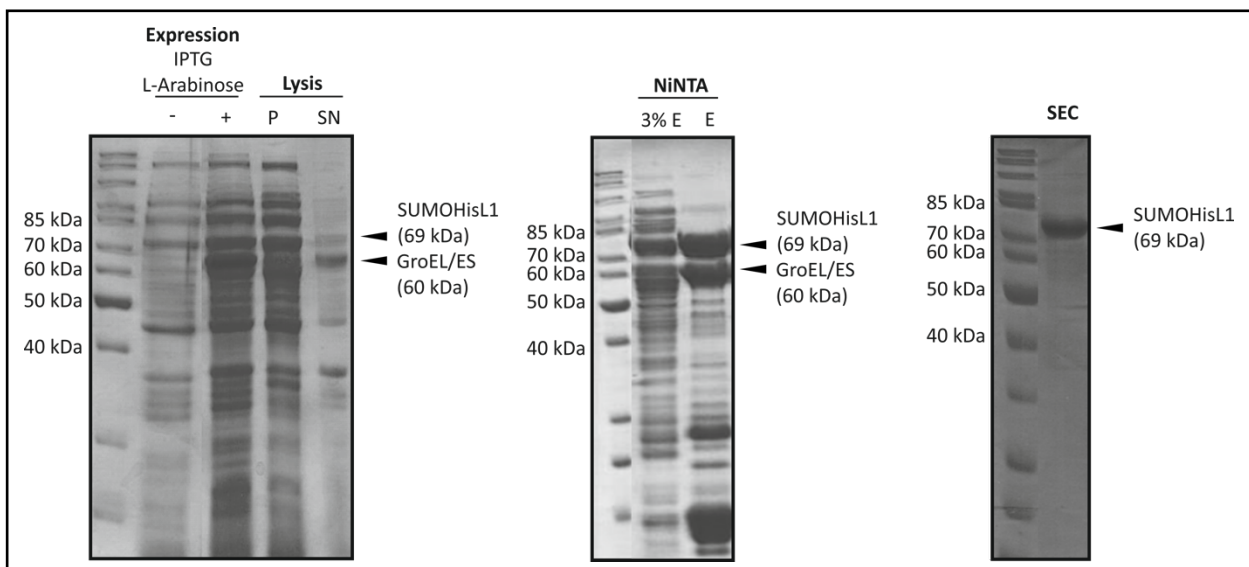
Figure S8: History of purification of His₆-SUMO-L1 construct analyzed via SDS-PAGE and Western Blot. (A): Expression without (-) and with (+) induction of recombinant L1nav and GroEL/ES protein for 4 h at 20 °C; lysis: insoluble pellet (P) and soluble supernatant (SN) were separated by centrifugation, 3% pre-elution (3%) and protein elution (E), size exclusion result from purified protein.

7.2.6 Purification His$_6$-TEV-GGGL2
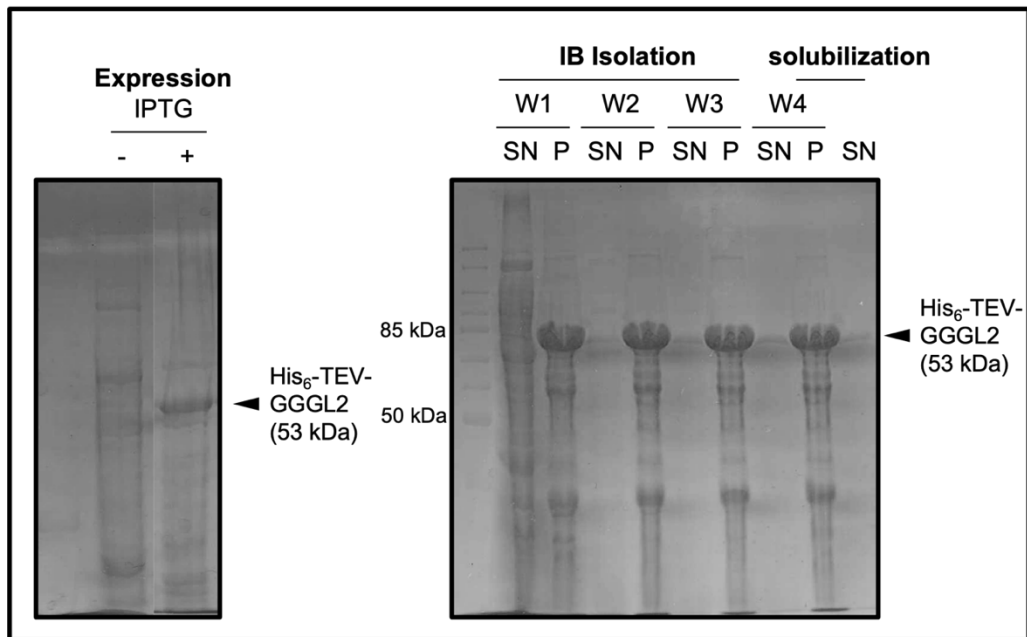
Figure S9: Expression and Purification of His$_6$-TEV-GGGL2 construct. Expression: before induction (-) and after induction (+) with 1mM IPTG for 4h at 37°C for His$_6$-TEV-GGGL2ΔL1bs construct. **(B)** W1-W4: wash of inclusion bodies.

7.2.7 Critical Micelle Concentration of Brij (CMC)
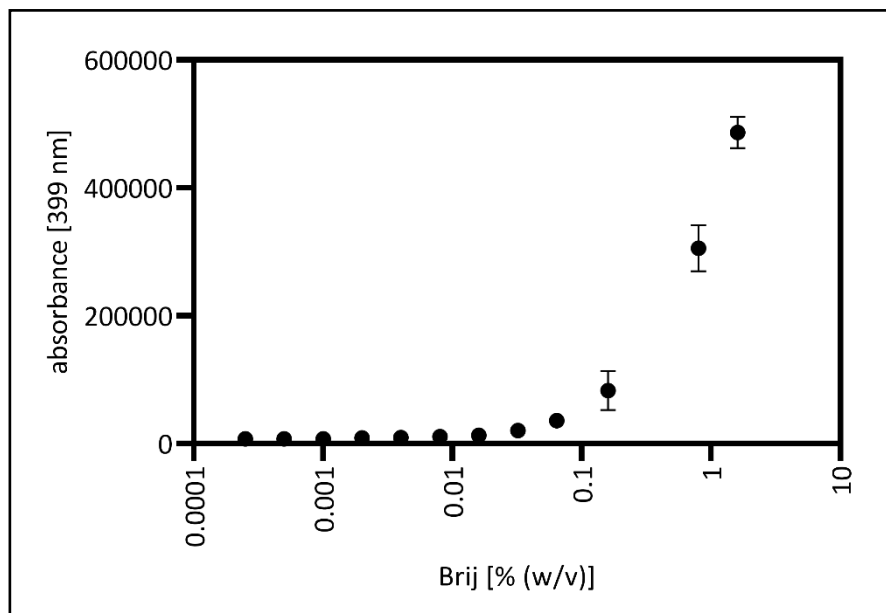
Figure S10: Determination of the critical micelle concentration (CMC) of Brij buffer. Therefore, the light scattering of different Brij concentration was measured. The CMC is localized at 0.08% Brij.

7.2.8   T$_{trans}$ – Evaluation of the L1 measurements
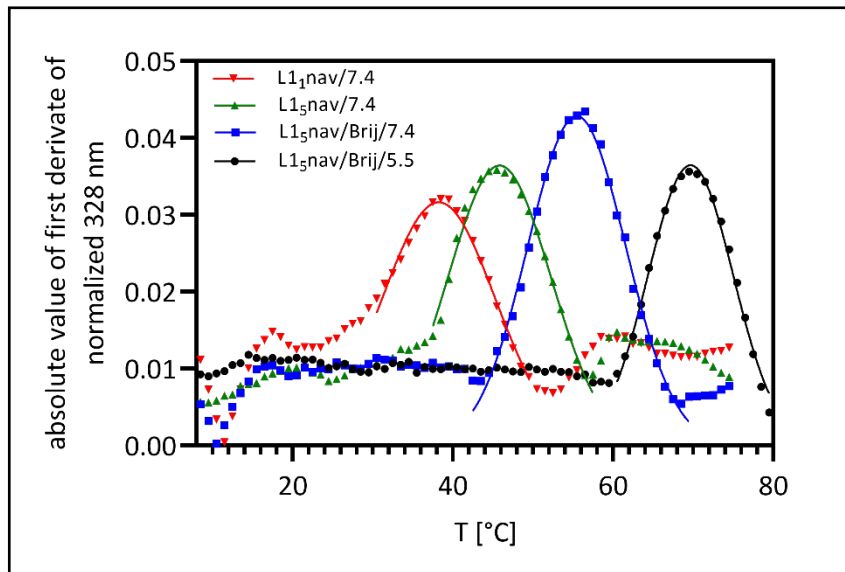
Figure S11: T$_{trans}$ evaluation. To determine the stability of L1nav pentamer and monomer under different buffer conditions, we normalized the data calculated the absolute value and fitted with a gaussian normal distribution. Pentamer in 1.6 % Brij pH 7.4 is stable until 56 °C for 1.6 % Brij pH 5.5 until 70 °C. In A0 L1nav pentamer is stable until 46 °C. On the contrary L1nav Monomer is stable until 38 °C

7.2.9   Measured data of the sedimentation velocity of the AUC
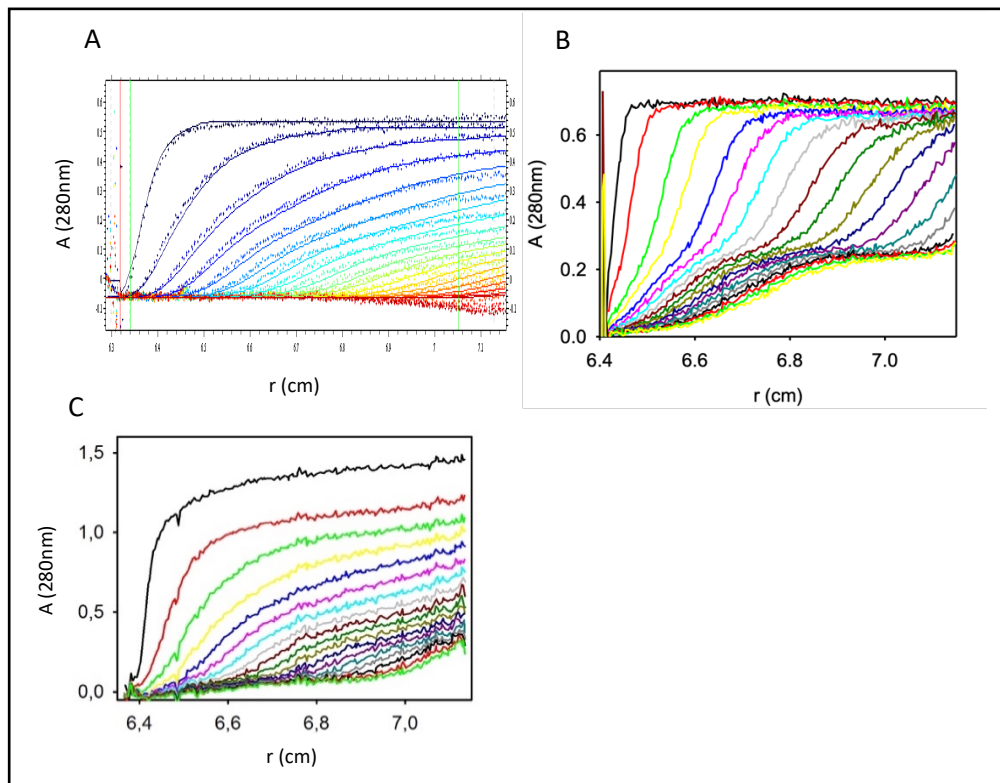
Figure S12: original data from sedimentation velocity measurement from figure 36 and 27. **(A)** L1nav pentamer under non-reducing conditions, fitted data present in figure 36 A. **(B)** L1nav pentamer under reducing conditions, fitted data present in figure 36 B. **(C)** L1nav pentamer under non-reducing conditions containing 1.6% Brij, fitted data displayed in Figure 37 B

7.2.10 ELISA – immobilized biotin-L2
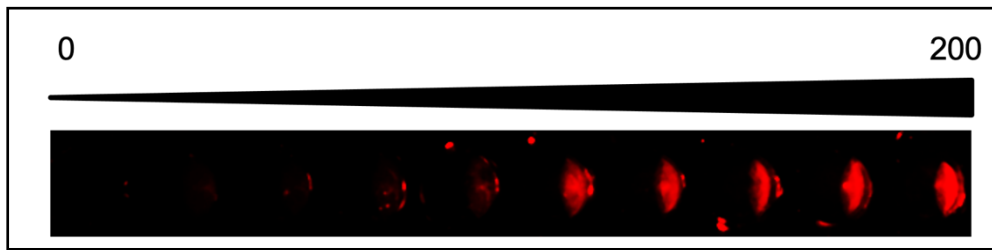
Figure S13: Immobilized biotin-L2 in Streptavidin plate. Different Biotin-L2 concentration form 0 – 200 µg/ml were incubated for 2 h at 4 °C and afterwards incubated with HPV16L2 antibody conjugated with Alexa Fluor®594

7.2.11 SPR – Sensograms from the different SPR measurements
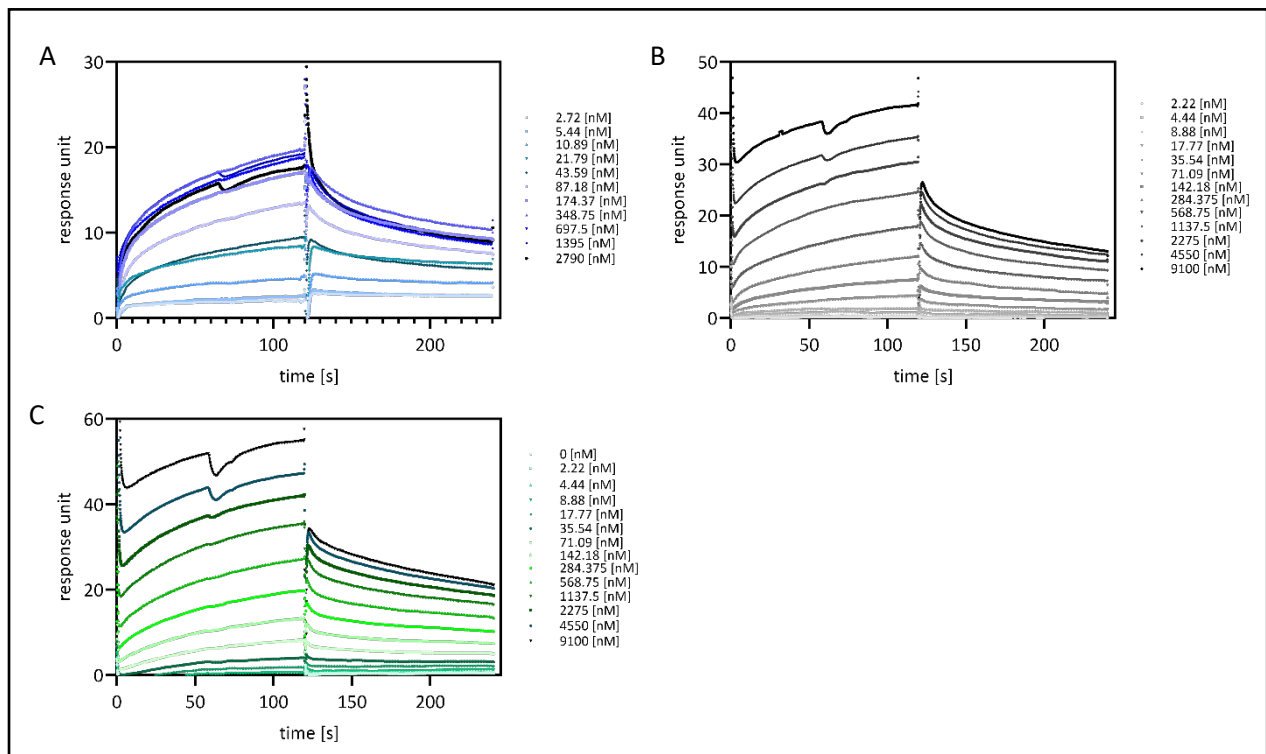
Figure S14: Sensograms from different SPR measurements. (A) Sensorgram from Biotin-L2 measurement in CP buffer pH 5.5 containing 0.16 % (w/v) Brij 58. **(B)** Sensorgram from Biotin-L2 measurement in CP buffer pH 7.4 containing 0.16 % (w/v) Brij 58. **(C)** Sensorgram from Biotin-L2 ΔL1bs measurement in CP buffer pH 7.4 containing 0.16 % (w/v) Brij 58. All measurements were performed at 25 °C.

8   PUBLICATIONS

Parts of this work have been published as follows:

Breiner, B., et al., *Refolding and in vitro characterization of human papillomavirus 16 minor capsid protein L2.* Biol Chem, 2019. **400**(4): p. 513-522.

Roos, N., et al., *Optimized production strategy of the major capsid protein HPV 16L1 non-assembly variant in E. coli.* Protein Expr Purif, 2020. **175**: p. 105690