

Association of human papillomavirus type 16 E2 with ChIR1:  
Implications for E2 function and the HPV life cycle

Leanne Harris

A thesis submitted to the

UNIVERSITY OF BIRMINGHAM

for the degree of

DOCTOR OF PHILOSOPHY

COLLEGE OF MEDICAL AND DENTAL SCIENCES

SCHOOL OF CANCER SCIENCES,

February 2015

UNIVERSITY OF  
BIRMINGHAM

**University of Birmingham Research Archive**

**e-theses repository**

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

## **Abstract**

Human papillomavirus (HPV) E2 is essential for transcriptional regulation of viral oncoprotein expression and the replication and persistence of episomal HPV genomes. Episomal persistence is mediated by tethering of viral genomes to host cell chromosomes during mitosis. Previous work demonstrated that interaction of E2 with the cellular DNA helicase ChIR1 is necessary for viral genome tethering. Therefore, disruption of this interaction is a potential therapeutic target for persistent HPV infections. To investigate the use of fragment-based drug discovery in the development of novel inhibitors of the E2-ChIR1 interaction, a fragment library was screened to identify those that bind E2 and several hits were identified. Concurrently, the interaction between HPV16 E2 and ChIR1 was characterised and shown to be a direct protein-protein interaction. The binding sites within E2 and ChIR1 were mapped and this information was used to identify a mutant E2 protein unable to bind ChIR1 (E2-Y131A). E2-Y131A was functionally characterised. HPV16 genomes encoding E2 wild type and Y131A were transfected into primary human keratinocytes to study the differentiation-dependent virus life cycle. Mutant genomes failed to establish genome maintenance, providing strong evidence that the interaction between HPV16 E2 and ChIR1 is necessary for the persistence of HPV infection.

## Acknowledgments

Firstly I would like to thank my supervisor Jo for her constant support and guidance over the last four years, and particularly for her willingness to take a risk and accept a student with very limited biology lab skills into her lab. Without her investment of a lot of time spent on training and answering my countless questions, all of this would not have been possible and I am truly grateful for all of her work with me. I would also like to thank all of the members of the Parish lab who have been great colleagues and friends for the last few years, along with the others in WX1.67. Thanks for being around to bounce ideas off, for the many chats over coffee and cake when things weren't going well, and for the nights out to celebrate when everything comes together. Your friendship has made this experience infinitely more rewarding.

I would also like to thank Sally for lending us her expertise (and her HFks and tissue culture space) and invaluable advice during this project. Thanks to Ben Wilcox and Jamie in the PEF, and to all others in Cancer Sciences who have at some point or another allowed me to borrow reagents and equipment, or just get some advice about ongoing experiments. Additionally, thanks to Jim Naismith, for allowing us the use of his fragment library, and to Magnus and the St Andrews NMR facility.

Finally, I would like to thank my family and friends for their support throughout this process. I know you all tune out when I start going on about western blots, but I appreciate you still asking after all this time! All the weekends spent on boardgames or camping with Kim, Sarah and the Eds, or in the pub with Dad, Ben, Josh and the rest of the village have been really valued, especially when this got hard! And last of all, I want to thank Ed for everything. His constant encouragement, patience, love and support have got me through all of the hardest times and made the good ones even better!

## Abbreviations

ATP	Adenosine triphosphate
BPV	Bovine papillomavirus
BET	Bromodomain and extraterminal domain family
BLM	Bloom's syndrome
Brd4	Bromodomain-containing protein 4
BrdU	Bromodeoxyuridine
BSA	Bovine serum albumin
CEM	Complete E medium
Chl1	Chromosome-loss 1
ChIR1	Chromosome-loss 1 related
CTD	C-terminal domain
CHX	Cycloheximide
Ctf	Chromosome transmission factor
CTP	Cytosine triphosphate
CV	Column volumes
DBD	DNA binding domain
DMSO	Dimethyl sulfoxide
DSF	Differential scanning fluorimetry
DMEM	Dulbecco's modified Eagle's Medium
DNA	Deoxyribose nucleic acid
E6AP	E6-associated protein
EBNA-1	Epstein-Barr virus nuclear antigen 1
EBV	Epstein-Barr virus
EGF	Epidermal growth factor
FA	Fanconi's anaemia
FBDD	Fragment-based drug discovery
Fen1	Flap endonuclease 1
FPC	Fork protection complex
GA	Glutathione-agarose
HFK	Human foreskin keratinocyte
HINGS	Heat-inactivated goat serum

His	Histidine
HNSCC	Head and neck squamous cell carcinoma
HP1	Heteroprotein 1
HPV	Human papillomavirus
HR	High risk
HTS	High-throughput screen
IF	Immunofluorescence
IPTG	Isopropyl $\beta$ -D-1-thiogalactopyranoside
KSHV	Karposi's sarcoma-associated herpes virus
LANA	Latency-associated nuclear antigen
LB	Luria broth
LCR	Long control region
LR	Low risk
MBP	Maltose binding protein
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
OD	Optical density
OP	Oropharyngeal
ORF	Open reading frame
Ori	Origin of replication
PAE	Early polyadenylation site
PAL	Late polyadenylation site
PE	Early promoter
PL	Late promoter
PBS	Phosphate-buffered saline
PCNA	Proliferating cell nuclear antigen
PE	Early promoter
PI	Propidium iodide
PL	Late promoter
PV	Papillomavirus
PVDF	Polyvinylidene fluoride
Rb	Retinoblastoma
RCB	Reconstitution buffer

RF	Radio frequency
RFC	Replication factor C
RPA	Replication protein A
rpm	revolutions per minute
RT	Room temperature
SCC	Sister chromatid cohesion
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SE	Size exclusion
SGS1	Small growth suppressor 1
siRNA	Short inhibitory ribose nucleic acid
SMC	Structural maintenance of chromosomes
SP1	Specificity protein 1
SPR	Surface plasmon resonance
STD	Saturation transfer difference
TAD	Transactivation domain
TBS	Tris-buffered saline
TE	Tris-EDTA
TERT	Telomerase reverse transcriptase
$T_m$	Melting temperature
TopBP1	Topoisomerase II binding protein 1
VLP	Virus-like particle
WBS	Warsaw breakage syndrome
wLOGSY	Water Ligand-Observed gradient spectroscopy
WRN	Werner's syndrome
WT	Wild type

# Table of Contents

Abstract .....	ii
Acknowledgments.....	iii
Abbreviations .....	iv
Chapter 1. Introduction.....	1
1.1 Human papillomaviruses and associated disease .....	2
1.2 HPV life cycle .....	6
1.3 E2.....	12
1.3.1 E2 structure .....	12
1.3.2 E2-dependent viral genome replication.....	15
1.3.3 Regulation of transcription .....	17
1.3.4 Maintenance of a persistent infection .....	19
1.4 ChIR1 .....	24
1.4.1 ChIR1 function: cohesion establishment.....	26
1.4.2 ChIR1 function: DNA replication .....	28
1.4.3 Warsaw breakage syndrome.....	30
1.4.4 ChIR1 function in E2-mediated PV genome maintenance .....	31
1.5 Hypothesis and aims .....	33
Chapter 2. Materials and Methods .....	35
2.1 Materials .....	35
2.1.1 Antibodies .....	35
2.1.2 Bacterial growth media and antibiotics .....	36
2.1.3 Cell culture media and reagents .....	37
2.2 Bacterial culture .....	39
2.2.1 Bacterial strains.....	39
2.2.2 Preparation of calcium competent cells .....	39
2.2.3 Transformation.....	40
2.2.4 Plasmid DNA preparation.....	40
2.2.5 Protein expression.....	41
2.3 Protein purification and analysis.....	42
2.3.1 Protein purification .....	42

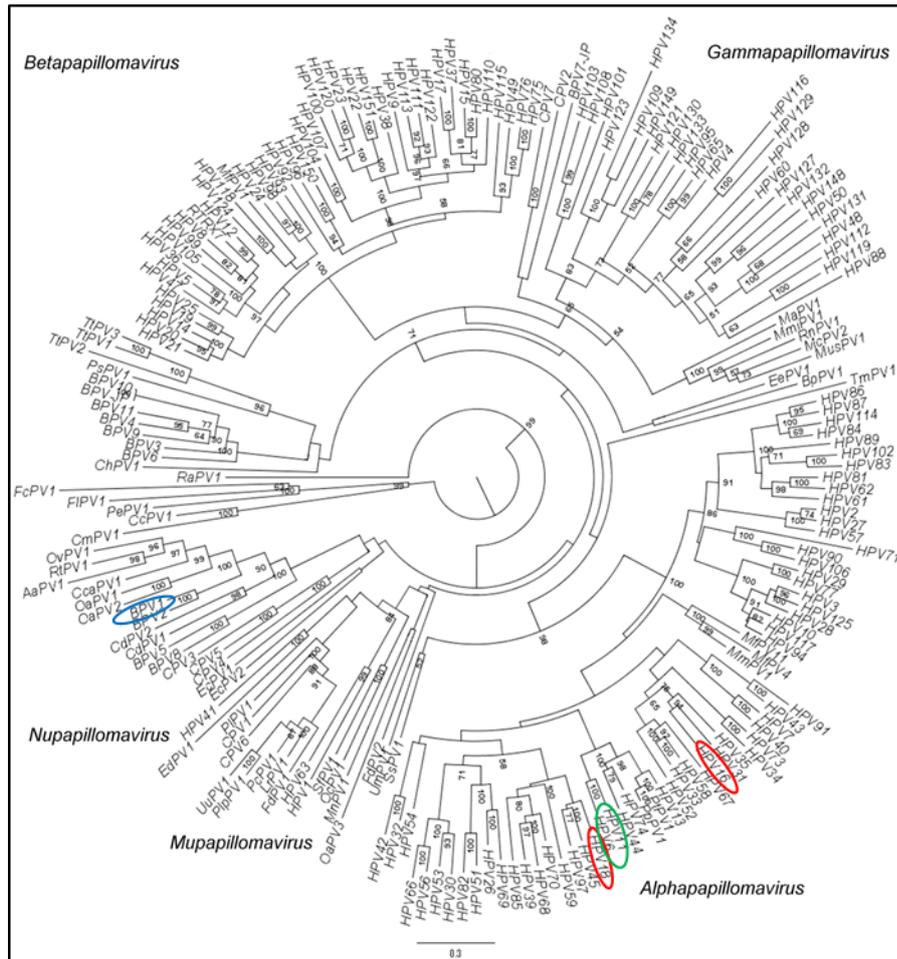
2.3.2 Pull-down assays .....	47
2.3.3 SDS-PAGE and coomassie staining .....	48
2.3.4 Western blotting .....	48
2.3.5 Bradford assay.....	49
2.4 Biophysical protein assays.....	50
2.4.1 Thermal shift assay (DSF) .....	50
2.4.2 Nuclear Magnetic Resonance (NMR) Spectroscopy.....	51
2.4.3 Hanging drop protein precipitation assay .....	52
2.4 DNA purification, analysis and cloning.....	52
2.4.1 Agarose gel electrophoresis .....	52
2.4.2 Phenol-chloroform DNA extraction.....	53
2.4.3 Restriction digest.....	54
2.4.4 Gel purification.....	55
2.4.5 Ligation (ChIR1 fragment cloning).....	55
2.4.6 Sequencing .....	56
2.4.7 Site-directed mutagenesis.....	58
2.4.8 Constructs .....	61
2.4.9 Southern blotting .....	61
2.5 Tissue culture (cell assays) .....	65
2.5.1 Cell lines and culture .....	65
2.5.2 Transient transfection .....	69
2.5.3 siRNA knockdown.....	70
2.5.4 Immunofluorescence (IF) .....	70
2.5.5 Cell synchronisation and flow cytometry.....	72
2.5.6 Co-immunoprecipitation (Co-IP) .....	74
2.5.7 Transcription assay.....	75
2.5.8 Replication assay .....	77
2.5.9 Protein lysis buffer test .....	78
2.5.10 Cycloheximide stability assay .....	78
2.5.11 Sub-cellular fractionation.....	79
2.5.12 <i>In situ</i> fractionation .....	79
2.5.13 Organotypic raft culture analysis .....	81
Chapter 3. Identifying small molecule fragments as novel binding partners of HPV11 E2 .....	83

3.1 Introduction.....	83
3.2 Aims.....	91
3.3 Results .....	92
3.3.1 Optimisation of the fluorescence-based thermal shift assay termed differential scanning fluorimetry (DSF) with purified HPV11 E2 TAD .....	92
3.3.2 Screening a library of small molecule fragments using DSF to identify novel fragments which interact with of HPV11 E2 TAD .....	104
3.3.3 Using NMR to screen a fragment library for novel fragments which bind to HPV11 E2 TAD .....	117
3.3.4 Attempting to identify fragment binding sites on HPV11 E2 TAD .....	132
3.4 Discussion .....	140
Chapter 4. Investigating the interaction between HPV16 E2 and ChIR1 .....	157
4.1 Introduction.....	157
4.2 Aims.....	160
4.3 Results .....	161
4.3.1 Investigating the E2 binding site on ChIR1 .....	161
4.3.2 Defining the ChIR1 binding site within HPV16 E2 .....	176
4.3.3 E2 mutagenesis .....	186
4.4 Discussion .....	201
Chapter 5. Characterising the HPV16 E2 mutant Y131A.....	212
5.1 Introduction.....	212
5.2 Aims.....	216
5.3 Results .....	217
5.3.1 E2 function .....	217
5.3.2 Assessment of E2 protein stability .....	223
5.3.3 Sub-cellular E2 localisation.....	229
5.3.4 Phenotypic analysis of E2 Y131A in the HPV life cycle .....	256
5.4 Discussion .....	265
Chapter 6. Conclusions and future work.....	280
References.....	285
Appendix 1. Buffers and solutions .....	295
Appendix 2. Details of NMR fragment library .....	302
Appendix 3. Details of NMR buffers screened with HPV11 E2 TAD .....	357

## **Chapter 1. Introduction**

The papillomaviruses (PV) are a large family of DNA viruses which infect both mucosal and cutaneous epithelial cells causing hyperproliferative lesions which are associated with a range of disease. To date more than 200 unique papillomaviruses have been identified, over 150 of which are known to infect humans, with multiple putative novel HPVs still being investigated [1-3]. These are categorised into distinct groups on the basis of sequence similarity, cell tropism and biological function. A phylogenetic tree illustrating the relationships between the PV genera is shown in Figure 1.1. Putative PV types are identified by sequencing, and determined to be a novel virus if they exhibit less than 90% similarity within the conserved L1 region to other known viruses [4].

The bovine papillomavirus BPV-1 has been extensively studied and is widely used as the model system for investigating the biology of PVs, particularly in determining factors leading to the development of carcinomas and understanding the interactions between the virus and host [5]. However, investigation of HPVs remains of clinical relevance due to their impact on human health.



**Figure 1.1 Phylogenetic tree illustrating the different papillomavirus genera**

Phylogenetic tree showing the relationship of 200 PVs based on sequence similarity within the L1 region, taken from [1]. Bovine papillomavirus BPV-1 is highlighted in blue, the HR HPV types 16 and 18 are highlighted in red, and the LR HPV types 6 and 11 are highlighted in green.

### 1.1 Human papillomaviruses and associated disease

The  $\alpha$ -papillomaviruses are the largest group, encompassing viruses which infect both the cutaneous epithelium (e.g. HPV2, 27, 28 etc.) and the mucosal epithelium of the anogenital and oropharyngeal tracts (e.g. 6, 11, 16, 18). Within the  $\alpha$ -PVs with a tropism for mucosal

epithelium, the viruses can be further subdivided as high-risk (HR) or low-risk (LR) on the basis of likely progression from initial lesion toward malignancy [6]. Although not generally associated with malignant progression, the LR PVs (most predominantly types 6 and 11) remain of clinical relevance as they are associated with genital warts, a highly prevalent sexually transmitted infection (STI). Genital warts, whilst not life threatening remain a significant burden on the health service as they are both highly contagious (estimated 65% transmission rate) and problematic to treat [7]. The cost of treating genital warts is estimated at between £113-139 per patient, resulting in a financial burden of approximately £17-20 million per year in the UK alone [7, 8]. The number of cases recorded has increased dramatically over time, increasing by 32% from 1995-2004 [9]. In England alone, over 80000 new cases and over 68000 recurrent cases of genital warts were reported in 2008 [7]. Treatment is at present restricted to cryotherapy or surgical interventions, or treatment with topical cytotoxic agents. Although the ablative cryotherapeutic or surgical interventions are often effective in the short term with 70-80% clearance reported, recurrence is common, with rates of 25-39% reported following multiple treatments [10]. Cytotoxic topical agents have also been used to treat genital warts, but these treatments are poorly tolerated by patients and do not target the underlying viral infection resulting in high recurrence rates [11]. Examples of the agents used include trichloroacetic acid (TCA) which can induce localised ulceration following treatment, the microtubule inhibitor podophyllin which is associated with a wide range of adverse side-effects, and the immunomodulator imiquimod [12]. Given the lack of a highly effective, well-tolerated treatments of HPV infections, the development of novel therapeutics which inhibit HPV infection or promote virus clearance is of great clinical significance.

The HR virus types are associated with the development of multiple anogenital and oropharyngeal cancers, most notable cervical cancer. HR HPV infection was identified as the major causative agent of cervical cancer [13], with >99.97% containing HPV DNA. Of these around 70% were found to be associated with HPV16 or HPV18 infection. Cervical cancer is one of the most common cancer types affecting women with around 528000 new cases and 266,000 deaths worldwide being reported in 2012 [14]. Cervical cancer is particularly a problem in less developed countries, with approximately 86% of cervical cancers occurring in women in these regions, accounting for 12% of the total female cancer burden [14].

Infection with HR HPV types is associated with other anogenital cancers, such as anal and vulval cancers, although at a significantly lower incidence. Infection is also associated with cancers at sites within the oropharyngeal (OP) tract including the tonsil and base of tongue. HPV infection is believed to be causally related to 15-20% of OP cancers, with the majority of these HPV positive tumours associated with HPV16 infection [15]. The remainder are associated with lifestyle risk factors such as smoking or alcohol consumption. In 2008 approximately 400,000 new cases of OP cancer and 223,000 deaths were reported, the majority of which occur within developed countries [16, 17]. The survival rate of head and neck squamous cell carcinomas (HNSCC) can be stratified on the basis of HPV status, with HPV positive cancers resulting in an up to 30% increase in the 5 year survival rate compared to HPV negative cancers [15, 18]. As such there is a need to develop treatments specific for HPV positive cancers as current therapeutic strategies may be too harsh in these cases [19-21].

HPV infection therefore remains a pressing concern for public health. The development of virus-like particle (VLP) based vaccines against the most prevalent HPV types is a huge step forward in the fight against HPV and the associated diseases. The vaccines are comprised of empty HPV capsids formed by self-assembly of over-expressed HPV L1 proteins. These VLPs are highly immunogenic, inducing the generation of type-specific immunodominant neutralising antibodies [22]. The L1 capsid structure is composed of highly conserved regions separated by hypervariable surface exposed loops [22]. Generated antibodies map to both conformational and linear epitopes on these surface loops and as such are highly type-specific [22]. Although some cross-neutralisation epitopes are shared between highly homologous virus pairs (e.g. 16/31, 18/45), these epitopes have been shown to be less immunogenic than type-specific epitopes, and it was therefore predicted to be unlikely that a high level of cross-protection towards other virus types would be conferred by the vaccines [23, 24].

A bivalent vaccine, Cervarix (GSK) was developed which offers protection against the two most common high-risk HPV types, 16 and 18. Similarly, a quadrivalent vaccine Gardasil (Merck) offers additional protection against the low-risk types 6 and 11. Vaccination programmes targeting pre-adolescent girls (age 12-13) along with a catch-up cohort of girls up to age 18 have been implemented in many countries. In the UK between 2008-2010 an estimated 60.4% of girls aged 12-19 had received 3 doses of the bivalent vaccine with 72.4% reported to have received at least one dose [25]. Although the absolute effects of vaccination will not be known for some time, preliminary studies on vaccinated cohorts have provided evidence that vaccination can reduce the incidence of HPV infections. Studies in a Scottish cohort of women vaccinated with Cervarix between 2009-2012 revealed a

significant reduction in the prevalence of HPV16 and 18 (13.6% after three vaccine doses compared to 29.8% in unvaccinated populations). Additionally, a significant reduction in the prevalence of the related HPV types 31, 33 and 45 was observed in vaccinated cohorts, while no change in prevalence of the other HR HPV types was observed [26].

Despite the huge advances in preventing HPV associated disease using prophylactic vaccines, there are still issues to address. Currently only girls are vaccinated and therefore boys are still at risk of HPV infection. In particular the incidence of oropharyngeal cancers in males is increasing [20, 27]. Additionally, although the quadrivalent vaccine offers protection against the 4 most prevalent HPV types, there are multiple other virus types also associated with disease, although at a significantly lower incidence [17]. There is some evidence of cross-protection from the vaccine towards other, similar virus types (e.g. HPV18 VLP cross-protection against HPV31) but the degree of protection conferred in this manner has not been fully investigated and is unlikely to fully protect against infection and HPV-induced disease. As such vaccinated individuals could remain at risk of infection with other HPV types.

## **1.2 HPV life cycle**

Papillomaviruses have a circular double-stranded DNA genome which, in a productive infection, is maintained as an extrachromosomal element within infected cells (Figure 1.2). This genome contains a long control region (LCR) containing the viral origin of replication (Ori) along with multiple binding sites for viral and cellular factors, including the viral proteins E1 and E2. The viral early promoter is also located within this region. The viral genome contains multiple overlapping open reading frames (ORFs) which encode the 8

major viral proteins: 6 proteins (E1, E2, E4, E5, E6 and E7) which are important for the multiple processes controlling viral life cycle progression, and two structural proteins (L1 and L2) which form the viral capsid [28, 29].

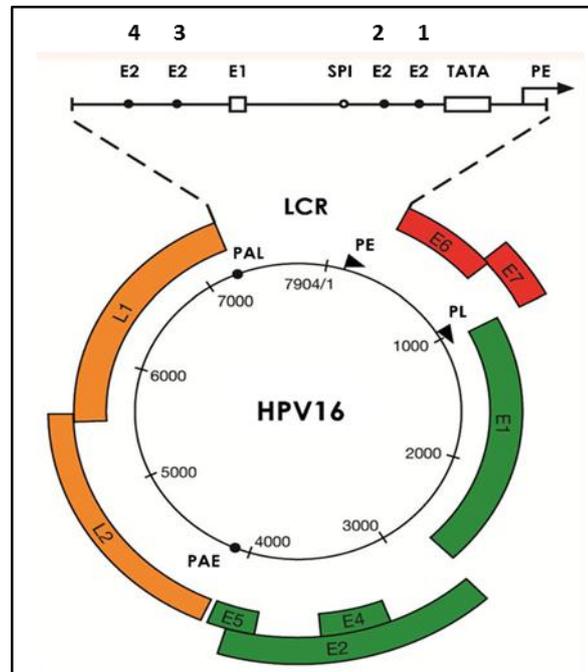


Figure 1.2 HPV16 viral genome [30]

The HPV16 circular genome is shown, with the positions of promoters and ORFs highlighted. The early p97 promoter (PE) and the late (p670) promoter (PL) are shown as arrows, and the early and late polyadenylation sites (PAE and PAL) are shown as dots. The early ORFs (red and green) are expressed from either promoter depending on the differentiation of the infected cells, while the late ORFs (orange) are expressed from the late promoter. The viral LCR (shown enlarged) includes multiple binding sites for the viral E2 protein (number 1-4), as well as binding sites for the viral E1 protein and cellular transcription factors.

The viral life cycle and the corresponding expression of the viral proteins are intimately tied together with the differentiation of the host epithelium. Normal squamous epithelium is formed of multiple stratified layers of keratinocytes. At the basement membrane there is a layer of basal cells which form the reservoir from which the epithelium is formed. Basal cells replicate asymmetrically, with one daughter cell remaining in the basal layer while the other is pushed upwards towards the surface. In the mid-layers of the epithelium, the cells begin to differentiate and exit the cell cycle, no longer actively undergoing DNA replication. As the cells migrate to the epithelium surface, they become fully differentiated, eventually being sloughed off from the surface [29]. The progression of the HPV life cycle is connected to the differentiation of the infected cells. Initial infection occurs through microabrasions in the epithelium which allow virus particles access to the basal cells. Viral entry is mediated by interactions between the viral capsid and cellular receptors including heparin sulphate proteoglycans and  $\alpha$ -integrin [31-33]. Viral decapsidation then occurs followed by L2-mediated relocalisation of the viral genomes to the cell nucleus [34]. Low level expression of the viral early proteins E1 and E2 is then required for initial amplification of the viral genome, facilitating replication of the viral genome and establishing infection at low copy number (estimates 50-200 copies per cell) within the replicating basal cells [35]. E2 is also required to upregulate the expression of the viral E6 and E7 proteins, and for partitioning of the viral genomes to ensure persistent maintenance within the nucleus of daughter cells [36, 37].

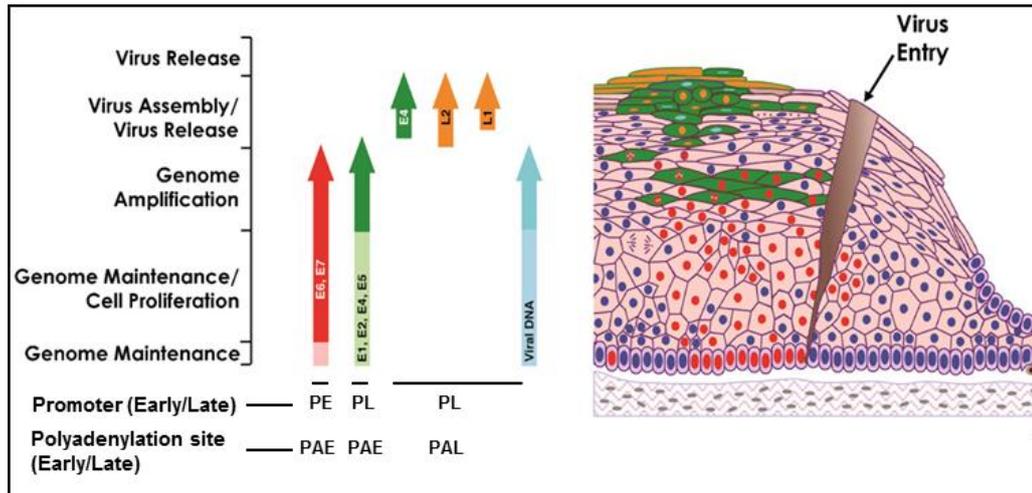


Figure 1.3 Epithelial differentiation and viral protein expression [28]

Initial HPV infection occurs through microabrasions allowing access to the basal epithelium. The viral genome is established and maintained within these basal cells as a low copy number extrachromosomal element. Asymmetric replication of the infected cells results in infected daughter cells moving upward towards the epithelial surface (red nuclei). In the lower layers, expression of the proteins required for viral genome replication and HPV-induced cell cycling are expressed (E1, E2, E6 and E7). In the mid layers, increased expression of these proteins along with E4 allows genome amplification to high episome copy numbers (green with red nuclei). Expression of the E5 protein can enhance this process, interacting with cellular factors including epithelial growth factor receptor (EGFR) and enhancing the associated growth factor pathways to support proliferation of the infected cells. As the cells progress to the upper layers, they exit the cell cycle and the viral capsid proteins L1 and L2 are expressed (yellow nuclei). The amplified viral genomes are packaged into new virions and the encapsidated virus particles are released as the terminally differentiated infected cells (yellow) are sloughed

**from the epithelium surface. The stages of the viral life cycle and the corresponding protein expression, along with promoter and polyadenylation site usage are shown to the left.**

The initial viral genome amplification and maintenance in the early phase of the virus life cycle has been shown to occur in a bidirectional fashion, with two replication forks beginning at the Ori and progressing around the genome until the forks collide, displacing the E1 double hexamer complex and allowing completion of replication [38, 39]. Following initial establishment, HPV infection causes an increased proliferation of infected cells. Normally the differentiating cells above the basal layer are removed from the cell cycle. Expression of the viral early proteins E6 and E7 cause re-entry into S-phase and continued cell cycling, allowing genome replication to continue. E7 associates with the retinoblastoma tumour suppressor protein (pRB) and its associated family members p130, p105 and p107 [40-42]. pRb functions as a block to S-phase entry, reversibly inhibiting E2F family of transcription factors. The E7 interaction with pRB prevents this inhibition, inducing premature S-phase entry [43, 44]. Additionally, in stratified epithelium, p130 acts to prevent entry of differentiating cells into cell cycle. E7 association destabilises p130 in the upper layers of the epithelium, removing this block [29].

E7-induced premature S-phase entry triggers a cascade of events resulting in p53-mediated apoptosis [45]. To prevent this and allow completion of the viral life cycle, E6 acts in concert with E7, targeting the cellular p53 response. E6 interacts with the cellular ubiquitin ligase E6AP [6, 46, 47]. The E6-E6AP complex then stably associates with p53, inducing polyubiquitination of p53 and its subsequent proteasomal degradation [44, 48]. Additionally,

the E6-E6AP complex also targets NFX1-91, a transcriptional repressor of the telomerase reverse transcriptase (TERT). E6-E6AP complex binding of NFX1-91 removes the inhibitory effect on TERT, inducing telomerase activity [49, 50]. This activity extends the lifespan of HPV infected cells, allowing cells to continue DNA replication beyond the normal limit by maintaining the telomeres of infected cells and therefore removing the signal to target cells for senescence [48]. Although necessary for completion of the productive viral life cycle, continued cell cycling and defective tumour suppression pathways result in an accumulation of genetic mutations and increased genomic instability which can eventually lead to cancer [28, 29, 51].

Interestingly the HR and LR E6 and E7 proteins function differently, potentially offering an explanation for the variation in disease progression. The LR proteins act to drive infected cells into S-phase without inducing a hyperproliferative phenotype. LR E6 has a reduced affinity for p53 and E6AP while LR E7 can target p130 but not pRB to induce cell cycling [28, 29].

Because the combined activity of E6 and E7 prevent infected cells exiting the cell cycle, HPV infection results in a thickening of the mid-layer of the epithelium. Viral genome amplification is initiated in these upper layers, resulting in upregulation of the E2 protein. High levels of E2 protein then act to repress transcription from the early promoter (PE) and activate transcription from the late promoter (PL), resulting in high levels of E1, E2, E4 and E5 protein expression following alternative splicing. The transcription of E6 and E7 is repressed, allowing cells to exit the cell cycle and initiate differentiation in response to cellular factors [52, 53]. Following cellular differentiation triggering the late, productive

phase of the viral life cycle, evidence suggests that the mode of replication switches to a unidirectional, rolling circle model, allowing for rapid amplification of the viral DNA [38, 39, 54]. A similar switch in the mode of viral genome replication has also been observed in other viruses: EBV episomes have been shown to undergo bidirectional replication during latent infection, while initiation of lytic infection is linked with a change in DNA replication to the unidirectional rolling circle mode [55].

The high levels of E2 protein expressed in the late phase of the life cycle then act to suppress polyadenylation of viral mRNAs at the early polyadenylation site (PAE) resulting in polyadenylation at the late polyadenylation site (PAL). Alternative splicing then results in the expression of the late, capsid proteins L1 and L2 in terminally differentiated cells [52, 53, 56]. New virus particles are then assembled and virions are released from the epithelium surface as terminally differentiated cells are sloughed from the surface.

### **1.3 E2**

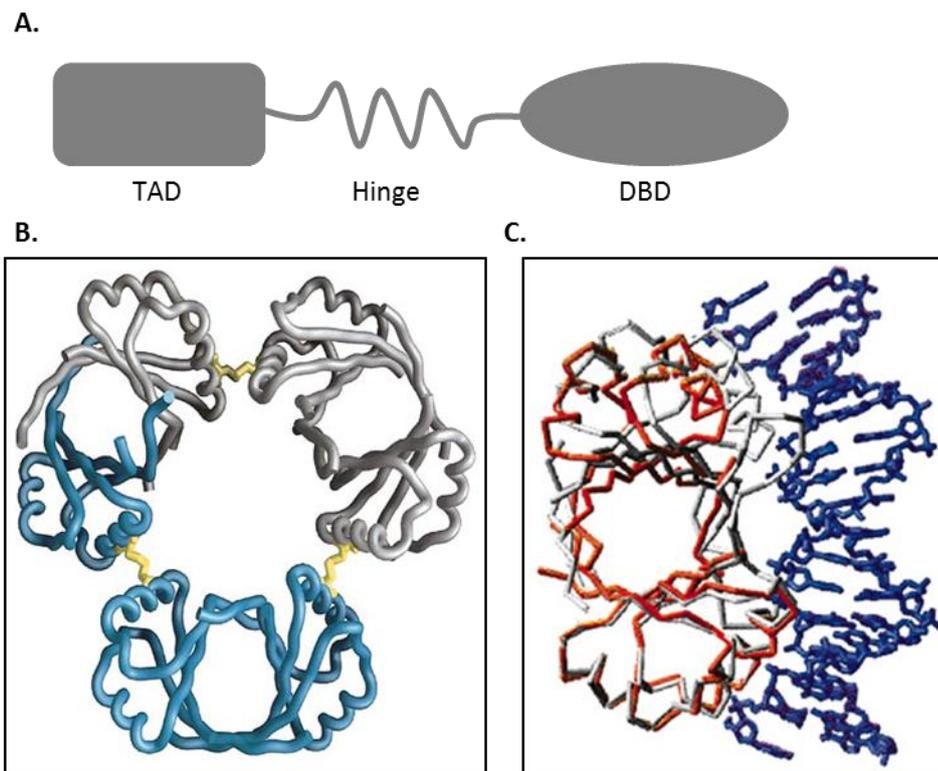
The multi-functional E2 protein is essential for initiation and progression of the viral life cycle, interacting with multiple viral and cellular proteins and performing roles necessary for viral genome replication, maintenance and amplification.

#### **1.3.1 E2 structure**

Structurally E2 is formed of two distinct domains linked by a flexible, unstructured hinge (Figure 1.4). The C-terminal DNA binding domain (DBD) forms a complex, dimeric  $\beta$ -barrel structure, with multiple  $\alpha$ -helices which form the DNA-binding surface. This unusual structure is conserved between multiple PV types, including BPV-1 and the high risk HPVs 16, 18 and 31 [57-59]. The PV E2 interaction with DNA occurs through a conserved, palindromic

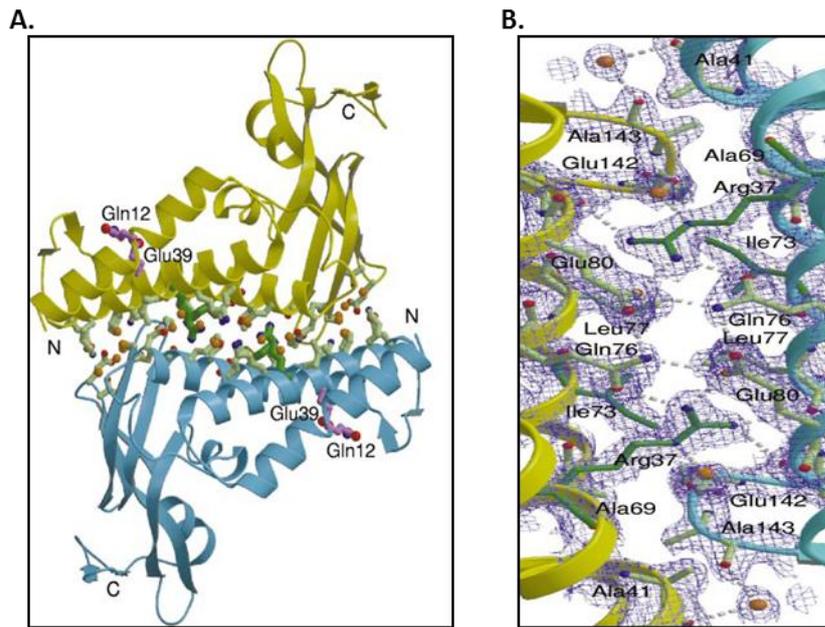
recognition sequence ACCg-(N)<sub>4</sub>-cGGT, which is repeated at multiple sites throughout the viral genome, with additional evidence suggesting that HPVs can recognise an extended recognition site AACCGN<sub>4</sub>CGGTT [60]. In particular there are multiple E2 binding sites within the viral LCR, some of which are in close proximity to the viral Ori and to the early promoter (Figure 1.2) [57, 60-63]. The affinity of E2 for these binding sites is variable, and this variation in binding is thought to be utilised to regulate E2-dependent transcription from the early promoter [64]. Displacement of cellular factors binding at, or proximal to the low affinity E2 binding sites 1, 2 and 3 by competitive binding of E2 can repress transcription from the early p97 promoter, and similarly the bend in the DNA structure induced by E2 binding can alter the conformation of the promoter region, regulating transcription [60, 63].

The N-terminal transactivation domain (TAD) is structurally very different to the DBD. It is comprised of two distinct regions, one formed of three long anti-parallel  $\alpha$ -helices (residues 1-92 in HPV16) and the other predominantly  $\beta$ -sheet (residues 110-201 in HPV16 E2) linked by a small region forming turns which act as a hinge (93-109 in HPV16 E2). In solution, the TAD forms low affinity dimers with an interface primarily between helices  $\alpha$ 2 and  $\alpha$ 3 (Figure 1.5) [65, 66]. Unlike the DBD, the TAD forms multiple protein-protein interactions with a range of both viral and cellular proteins, and it is through these interactions that E2 can mediate essential events in the viral life cycle.



**Figure 1.4 Structure of the PV E2 protein DBD**

**A. Schematic representation of the structure of E2, comprising two distinct domains linked by a flexible hinge region. The N-terminal transactivation domain (TAD) forms multiple protein-protein interactions while the C-terminal DNA-binding domain interacts with sequence specific binding sites within the viral genome. B. Dimerisation of two E2 DBD molecules (blue and grey) forms an unusual  $\beta$ -barrel structure with 4 surface exposed  $\alpha$ -helices [61]. C. E2 DBD dimers can interact with DNA through a palindromic recognition sequence *via* interactions with two of the surface helices. The crystal structure of HPV16 E2 DBD (red) is shown superimposed over BPV-1 E2 DBD (grey) bound to DNA (blue) [57].**



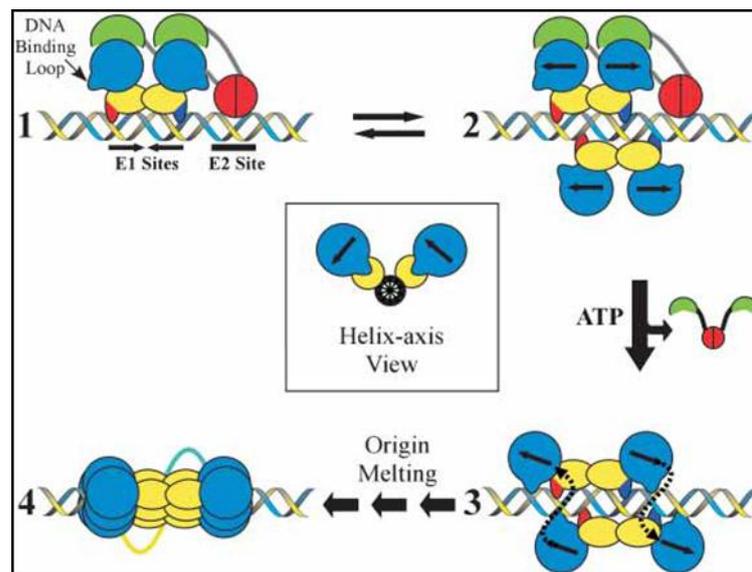
**Figure 1.5 Structure of the E2 TAD**

The E2 TAD forms low affinity dimers in solution, with two E2 TAD molecules in an antiparallel arrangement interact over a long surface. A. Crystal structure of the HPV16 E2 TAD dimer [65] with the interacting residues highlighted. The N- and C-terminal ends of each E2 TAD monomer (yellow and blue) are shown (N and C). B. An electron density model of the HPV16 E2 TAD dimer surface with the interacting residues highlighted [65].

### 1.3.2 E2-dependent viral genome replication

One of the primary functions of E2 is to facilitate the initiation of viral genome replication through an interaction with the viral helicase, E1 [67]. E1 is an ATP-dependent DNA helicase, acting to initiate episomal replication from the viral Ori. This process is dependent upon E1 binding to specific A/T rich regions within the Ori, however E1 shows limited affinity for these binding sites [68]. E2 facilitates the formation of the E1-Ori complex required for replication, forming an intermediate E1-E2 complex through an interaction of monomeric E1

with the E2 TAD. The E2 DBD then interacts with a high affinity E2 binding site adjacent to the E1 binding site within the Ori, effectively loading E1 into position [68-71]. Following E2-mediated formation of the E1-Ori complex, E2 is then displaced by an allosteric interaction of E1 with ATP [71] and further E1 monomers are recruited to assemble the active double hexamer conformation of E1 [68, 71], encircling the DNA to initiate viral genome replication.



**Figure 1.5 Schematic to illustrate the role of E2 in viral genome replication**

Interaction between E2 and E1 results in dimers of E1 (blue and yellow) and E2 (red and green) binding to the viral Ori (1). A second E1 dimer is then recruited to the Ori (2), resulting in ATP-dependent displacement of the E2 dimer and a conformational change in the E1-DNA complex (3). Recruitment of additional E1 dimers leads to assembly of the active E1 double hexameric complex which encircles the DNA double strand and initiates viral genome replication from the Ori [71].

The active E1 double hexameric complex then initiates replication by melting double stranded DNA and progressing along the newly exposed single stranded DNA, recruiting

cellular replication factors, including replication protein A (RPA), topoisomerase and polymerase  $\alpha$ -primase [54, 72].

### **1.3.3 Regulation of transcription**

Another function of E2 is in regulating the transcription of the two viral oncoproteins E6 and E7. These two proteins work in concert to alter the cell cycle progression of infected cells to allow viral genome amplification (discussed above). E2 acts to control the expression of these proteins, acting as both an activator and repressor of transcription to tightly regulate this process.

The transcription activation function of E2 is dependent on an association with bromodomain containing protein 4 (Brd4) [73]. Brd4 is a BET protein family member containing 2 bromodomains which are thought to be involved in binding to chromatin [74]. Although initially identified as an E2 interacting protein by proteomic tandem affinity purification (TAP) screening while investigating E2-dependent genome maintenance [75], more recent studies have shown another role for Brd4 within the viral life cycle through association with E2. Although not all PV E2s are dependent on Brd4 interaction for mitotic localisation (see 1.3.4), the E2-Brd4 interaction was found to be both conserved across PV types and necessary for E2-dependent transcription activation [76, 77]. One proposed mechanism for this is through the interaction of Brd4 with positive transcription elongation factor b (pTEFb) complex. Brd4 binding to pTEFb displaces an inhibitory subunit, activating the pTEFb complex to phosphorylate RNA polymerase II, activating transcription elongation. E2 recruitment of Brd4 and the associated transcription complex could therefore activate transcription from the proximal viral early promoter [78, 79].

As well as acting as an activator of viral transcription, E2 also functions to repress transcription from the viral early promoter. Unlike its transcription activation function which was found to be dependent on the Brd4 interaction, E2-dependent repression of transcription was found to be Brd4 independent [79]. Two short forms of the E2 protein which act to repress viral transcription and replication have also been identified. The 31 kDa E2-TR protein [80] is transcribed from an internal start codon within the E2 ORF, and the 28 kDa E8<sup>E2</sup> protein [81] is generated by alternative splicing of the E2 mRNA. Both proteins encode the C-terminal DBD of E2 and these regulatory forms of E2 are conserved between papillomavirus types [82, 83]. E2-TR and E8<sup>E2</sup> act to repress viral transcription by competitive binding to the E2 binding sites within the LCR. The repressor proteins can also interact with full-length E2 protein to form inactive heterodimers, repressing viral transcription by inhibiting the activity of E2 [84, 85].

This ability of E2 to both activate and repress transcription from the early promoter is proposed to be the basis for the mechanism of regulating E6 and E7 transcription. The affinity of E2 for the multiple binding sites within the LCR is variable [60], depending on the sequence of the 4 nucleotide spacer in the middle of the palindromic repeat consensus sequence. The E2 DBD does not interact directly with these amino acids, however an A-T rich spacer results in an optimum DNA conformation to allow binding of the E2 DBD DNA-recognition helices to the major grooves of the DNA [59]. At low E2 concentration, the protein will bind specifically to the high affinity binding sites upstream of the promoter. E2 binding could then activate transcription from the promoter by the recruitment of transcription factors, potentially mediated by interaction with Brd4. As the E2 protein concentrations increase, binding to the lower affinity binding sites proximal to the promoter

can occur, blocking binding of transcription factors including SP1 and transcription factor II D (TFIID) at overlapping sites and repressing transcription [86-90].

This control is necessary to prevent uncontrolled proliferation of infected cells and differentiation-dependent completion of the virus life cycle. Enhanced proliferation leads to an accumulation of DNA damage that occurs during cellular DNA replication leading to genomic instability. Loss of E2 expression, and the accompanying deregulation of E6 and E7 by integration of the viral genome into the cellular DNA, is therefore considered to be a critical step in the progression of infection towards malignancy. Random integration events can result in integration of the viral genome into cellular DNA at fragile sites within the cellular genome. Integration events which result in disruption of the E2 ORF but which leave the viral LCR and E6 and E7 ORFs intact give a growth advantage to those cells, which frequently results in cellular proliferation and the eventual progression toward malignancy [91, 92].

#### **1.3.4 Maintenance of a persistent infection**

A further function of E2 essential for the viral life cycle is the maintenance of a persistent infection. The viral genome is maintained as an extrachromosomal element in infected cells, and as such it is necessary for the viral episomes to be tethered to the cellular chromatin during mitosis in order to ensure even segregation of the episomes into daughter cells. Failure to associate the viral genomes with the cellular chromatin could result in uneven segregation of the viral DNA into the newly formed nuclei and eventual loss of episomes over time [77, 84, 86, 93-96].

Multiple DNA viruses have developed similar mechanisms for tethering episomal genomes to the cellular chromatin throughout mitosis to ensure persistence. The Epstein-Barr virus (EBV) genome is maintained as a low copy number episome in latently infected cells. The Epstein-Barr nuclear antigen 1 (EBNA1) is functionally similar to HPV E2, with multiple roles in viral genome replication initiation and transcription activation, acting through multiple interactions with key elements within the viral *oriP* [97]. Like E2, EBNA1 can associate with the viral genome through a specific DNA-binding domain and can associate with cellular chromatin associated proteins through the N-terminal domain. EBNA1 binding protein 2 (EBP2) was identified as the cellular binding partner of EBNA1 which is required for tethering of episomal genomes to mitotic chromosomes [97, 98].

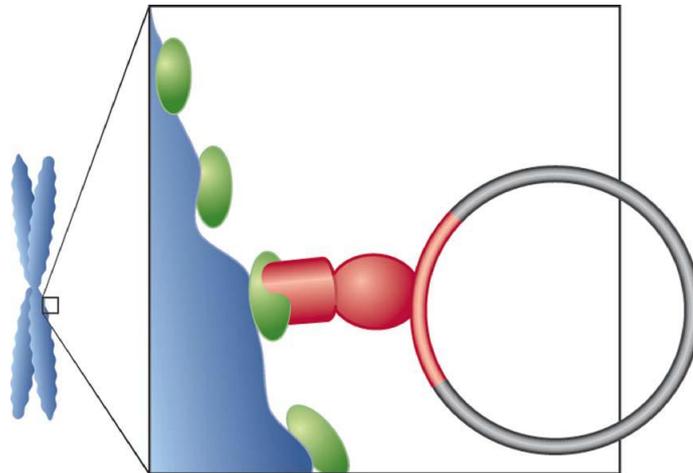
Kaposi's sarcoma-associated herpes-virus (KSHV) uses a similar tethering mechanism to maintain a latent infection. In this case latency-associated nuclear antigen (LANA) is used to mediate the tethering interaction between the viral genome and the cellular chromatin. Like EBNA-1, LANA is functionally similar to E2, interacting with both the viral genome and cellular chromatin-associated proteins. LANA interacts with a 20 nucleotide binding site within the terminal repeat region of the KSHV genome [99, 100], and both LANA and KSHV DNA are observed to colocalise in discrete foci on metaphase chromosomes. There have been multiple suggested candidates as the cellular binding partner of LANA in this process, including the bromodomain proteins Brd4 [101], RING3/Brd2 [102] and histones H2A and H2B on the nucleosome surface [103, 104]. Interestingly, Brd4 has also been suggested as one of the cellular proteins involved in the E2-mediated tethering of PV genomes, indicating the possibility of a common mechanism being utilised between viruses [75].

In papillomaviruses, E2 has been shown to act as the tether linking the viral genome and the cellular chromatin to ensure episomal maintenance throughout mitosis. PV DNA has been shown to colocalise to mitotic chromosomes throughout mitosis in the presence of E2, and further investigation demonstrated that the E2 DBD interacts with a region of the LCR containing multiple E2 binding sites termed the mini-chromosome maintenance element (MME), while the TAD is required for association with mitotic chromosomes, presumably through protein-protein interactions as the TAD does not bind directly to DNA [84, 94, 105].

One proposed candidate that mediates E2-dependent genome tethering by interaction with E2 is Brd4. Brd4 has been shown to localise with mitotic chromosomes throughout mitosis through an interaction between the bromodomains and acetylated histones H3 and H4 [106], and was shown to colocalise in discrete foci with BPV-1 E2 [75]. A region of the C-terminal domain (CTD) of Brd4 was determined to be sufficient for interaction with E2 [75], and expression of the Brd4 CTD as a competitive inhibitor of the E2 interaction resulted in a loss of BPV1 E2 localisation to mitotic chromatin. Similarly, mutation of BPV1 E2 (R37A, I73A) to disrupt the E2-Brd4 interaction [76] led to a similar loss in chromatin association. However this was not found to be a consistent interaction with all PV E2s. Although multiple PV E2 proteins were found to associate with Brd4 and be dependent on this interaction for transcription activation, alpha PV E2 proteins (types 11, 16, 31 and 57) were shown to localise in foci on mitotic chromatin (following pre-extraction stabilisation) in a Brd4 independent manner [77]. This suggests that other cellular partners may have a role in E2-mediated genome tethering, or that different mechanisms for genome segregation are employed across PV types.

Another chromatin associated protein suggested as the E2 interaction partner for genome maintenance is topoisomerase II binding protein 1 (TopBP1). TopBP1 contains multiple BRCT domains (BRCA-1 C-terminal domain) and is involved in the cellular DNA damage response. It was identified as an E2 interacting protein in a yeast 2-hybrid screen, and further investigation showed that the two proteins interact both *in vitro* and *in vivo* [107] and that this interaction is not required for the transcription activation function of E2. Immunofluorescence experiments showed that HPV16 E2 and TopBP1 colocalise in foci on mitotic chromosomes during the late stages of mitosis [108], suggesting that the E2-TopBP1 interaction could be required for the mitotic localisation of E2 and the consequent viral genome tethering. Interestingly, depletion of TopBP1 was shown to alter the sub-cellular localisation and stability of HPV16. Loss of TopBP1 resulted in an increase in E2 localisation to the chromatin following sub-cellular fractionation by sequential salt extractions [108]. The chromatin associated E2 was also shown to be more stable following TopBP1 depletion. It was suggested that this change in E2 localisation and stability could be due to E2 associating with an alternative chromatin receptor protein following TopBP1 depletion, and that this alternate complex stabilises E2 protein. This was not consistent between PV types however, with TopBP1 depletion resulting in no change in the localisation of BPV-1 E2. Further investigation of this interaction resulted in identification of a HPV16 E2 mutant (N89Y/E90V) which showed reduced affinity for TopBP1 and a reduction in the ability to facilitate E2-dependent replication [109]. The TopBP1-binding E2 mutant was shown to retain WT transcriptional activity, but when mutant genomes were used to generate human foreskin keratinocyte (HFK) cell lines, the mutant genomes were unable to establish episomal genomes. It is therefore possible that the E2-TopBP1 interaction is essential for stable

maintenance of HPV16 genomes, either by mediating the E2-dependent tethering of viral episomes to cellular chromatin, or by facilitating the initial viral genome replication required to establish low copy number episomes in infected basal cells.



**Figure 1.6 Proposed model of tethering [96]**

**Proposed mechanism persistent maintenance of viral genomes through mitosis. The E2 protein (red) acts to tether the viral genome to cellular chromatin (blue). The E2 DBD interacts with the viral genome through sequence-specific binding sites in the LCR, while the TAD interacts with cellular chromatin-associated proteins (green).**

An alternative proposal for the cellular binding partner of E2 involved in genome tethering is the DNA helicase ChR1. The interaction between E2 and ChR1 is the major focus of this thesis, therefore ChR1 function will be discussed in detail below.

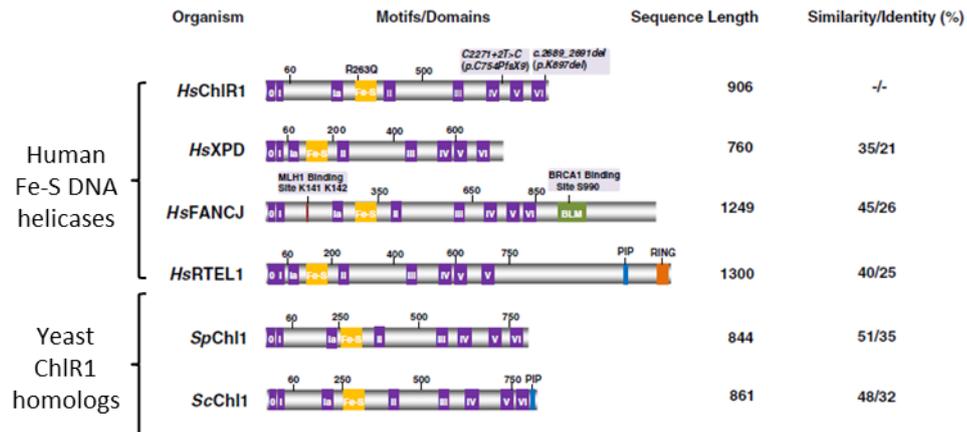
## 1.4 ChlR1

The yeast protein chromosome loss 1 (Chl1) was initially identified through genetic screens performed in *Saccharomyces cerevisiae* to find chromosome loss mutations [110]. Mutation in the CHL1 gene was shown to result in the loss of multiple chromosomes through an increased frequency of mitotic recombination events. Further research confirmed that the putative DNA helicase Chl1 is necessary for accurate chromosome transmission and that *CHL1*-null mutants exhibited large scale chromosomal missegregation and a G<sub>2</sub>/M delay which was not dependent on initiation of the RAD9 DNA damage pathway [111], suggesting that Chl1 functions through an alternate pathway to affect genome stability.

ChlR1, the human homologue of the yeast Chl1 protein was first identified from a cDNA library screen of genes differentially expressed following treatment of HaCaT cells with keratinocyte growth factor (KGF) [112]. The isolated cDNA was found to have significant homology with CHL1, showing 32% identity and 55% similarity at the amino acid level. This homology was most conserved within elements associated with helicase functionality including the Walker A and B ATP-binding motifs and the helix-turn-helix DNA binding motif [112], suggesting that ChlR1 behaves as an ATP-binding protein with helicase activity that potentially plays a role in chromosome segregation similar to that of Chl1. Further work confirmed the homology between ChlR1 and Chl1 [113], showing that ChlR1 contained the same helicase domains with the same inter-domain spacing as Chl1, indicating a conservation of function across species.

DNA helicases contain multiple conserved domains which are important for function. The Walker A and B boxes are required for ATP-binding, and the Walker B box also contains

motifs which serve to subdivide helicases on the basis of sequence similarity. One subcategory of the superfamily 2 (SF2) helicases are a group of FancJ-like helicases containing a DEAH motif in the B box and a unique iron-sulphur cluster domain [114].



**Figure 1.7 Helicase domains conserved between SF2 Fe-S DNA helicases**

**ChIR1 and the homologous yeast Chl1 proteins are related to a family of SF2 DNA helicases containing a Fe-S cluster (yellow). The position of the conserved helicase domains (purple) are shown for ChIR1, Chl1 and the related helicases XPD, FancJ and RTEL. Biologically relevant mutations in ChIR1 associated with WBS are highlighted [115].**

DNA helicases are known to contribute to genome stability through multiple pathways. Helicases are involved in DNA replication and repair pathways as well as in RNA transcription and splicing, with some helicases having multiple known functions. Helicases such as the *S. cerevisiae* small growth suppressor 1 (SGS1), a member of the RecQ subfamily which includes helicases Werner syndrome RecQ helicase-like (WRN) and Bloom syndrome RecQ helicase like (BLM), and the *Drosophila melanogaster* lodestar and are known to affect chromosome segregation in a similar manner to Chl1 [113], so it is possible that ChIR1 could

function similarly. Similarly, the DEAH-box Fe-S DNA helicases which are related to ChIR1 (Figure 1.7) are indicated to have roles in maintaining genome stability [116], suggesting that ChIR1 may also play a similar role in genomic integrity.

Biochemical analysis of the ChIR1 helicase has confirmed the activity of ChIR1 as an ATP-dependent DNA helicase which unwinds DNA preferentially in the 5' to 3' direction [114, 117]. The helicase was demonstrated to have a preference for unwinding forked duplex DNA, requiring a 5' single-strand DNA tail of approximately 15 nucleotides for maximum efficiency. ChIR1 was also shown to be capable of ATP-dependent displacement of DNA bound proteins and of unwinding complex DNA G-quadruplex (G4) structures, suggesting a potential role in unwinding areas of DNA which pose obstacles to replication fork progression and cohesion establishment [117].

#### **1.4.1 ChIR1 function: cohesion establishment**

To ensure even segregation of newly replicated DNA into daughter cells, sister chromatids are tied together in pairs to allow alignment with the metaphase spindle, before separation and segregation into daughter cells during anaphase. The cohesin complex is responsible for this tethering. Cohesin is formed of 4 protein subunits, interacting to form a ring around the DNA strands [118]. Smc1 and Smc3 subunits form heterodimers, interacting at the head and tail to form a ring around DNA. A Scc1-Scc3 dimer then interacts with the Smc1-Smc3 heads, interacting through the Scc1 subunit and strengthening the interaction [118, 119]. Following DNA synthesis, cleavage of the Scc1 subunit by separase leads to opening of the cohesin ring and separation of the sister chromatids [118, 120, 121].

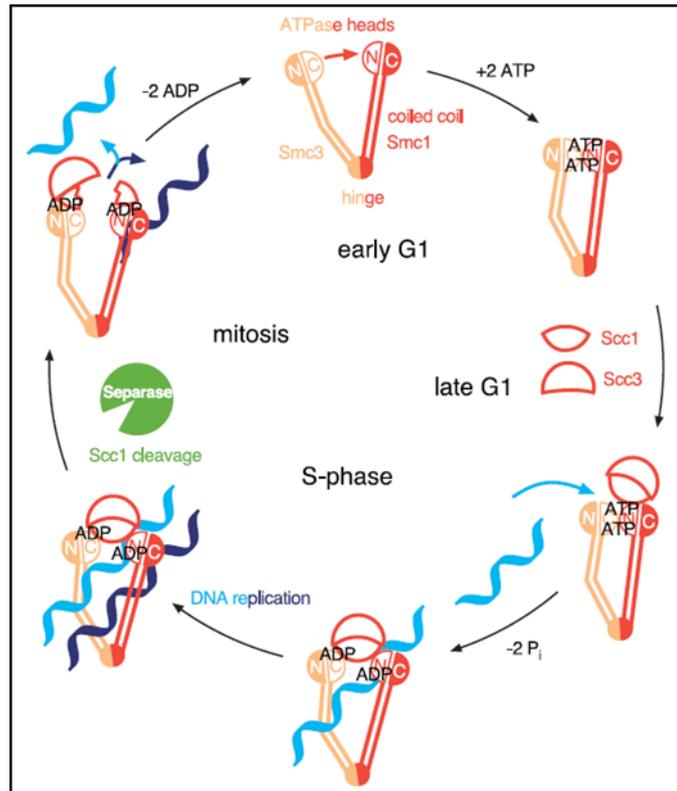


Figure 1.8 cohesion establishment schematic [121]

A model to illustrate the assembly of cohesin rings. Smc1 and Smc3 subunits interact in early G1 forming a ring linked at the heads and tails. During late G1 Scc1 and Scc3 heterodimers interact with the Smc1 head-group. In S-phase, ATP-dependent separation of the Smc1-Smc3 head-groups allows entry of DNA strands into the protein ring, leaving Scc1 joining the Smc subunits. DNA replication can progress through the cohesin rings leaving the newly synthesised sister chromatids encircled. During mitosis, separase activation cleaves Scc1, opening the cohesin ring and allowing sister chromatid separation.

Both Chl1 and the human homolog ChlR1 have been linked to a role in cohesion establishment. In yeast, Chl1 mutations have been shown to result in an abnormal cohesion phenotype similar to that displayed upon mutation of the cohesion establishment factor

chromosome transmission factor 7 (Ctf7) [110, 111, 122]. Ctf7 acts to couple cohesion establishment to ongoing DNA replication, acting as an acetyltransferase which acetylates the cohesin complex [123]. Ctf7 has also been shown to physically interact with Chl1 [122], suggesting an associated role for ChR1 in cohesion establishment. As well as the association with cohesion establishment factors, human ChR1 has been shown to associate with the cohesin subunits Smc1, Smc3 and Scc1 [124].

The localisation of ChR1 throughout mitosis appears to mirror that of the chromatin-bound cohesins, and is observed along the length of the sister chromatids early in mitosis, before being removed from the chromatid arms during metaphase [124]. Mutations which abrogate the DNA helicase activity of CHL1 result in a reduction in efficient chromosome segregation, indicating that the helicase functionality of CHL1 is required for effective cohesion [125]. Similarly, depletion of ChR1 results in defective sister chromatid cohesion [124]. Together this suggests a role for ChR1 in cohesion establishment during S-phase, which is co-incident with ongoing DNA replication.

#### **1.4.2 ChR1 function: DNA replication**

Linked to cohesion establishment, ChR1 is also indicated to have roles in transcriptional silencing, heterochromatin organisation and DNA replication at difficult sites. The yeast homolog Chl1 interacts with the cohesion establishment factors ctf18 and defective in sister chromatid cohesion 1 (Dcc1) which are involved in transcriptional silencing, and Chl1 was shown to be important for transcriptional silencing at telomeres [126]. Additionally, ChR1 is thought to effect the organisation of heterochromatin, areas of condensed and transcriptionally silent DNA [127]. This effect is potentially through the targeting of HP1

protein to methylated histone 3 (H3K9-me), allowing the loading of DNA-methyltransferases which methylate at CpG sites forming heterochromatin. HP1 is also indicated to have a role in cohesion establishment, interacting with the cohesin-loading factor NIPBL, again linking ChIR1 function to sister chromatid cohesion.

As well as cohesion-associated proteins, ChIR1 also interacts with multiple replication fork proteins including proliferating cell nuclear antigen (PCNA) and flap endonuclease 1 (FEN1) which is involved in lagging strand processing [128]. ChIR1 has also been shown to interact with the modified replication factor C (RFC) complex hCtf18-RFC, where one of the 5 RFC subunits is replaced with the cohesion establishment complex ctf18/DCC1/ctf8 [128, 129]. Interestingly, the presence of the RFC-ctf18 complex increased the processivity of ChIR1, increasing the length of duplex DNA which could be unwound. ChIR1 also associates with the replication fork protection complex subunit Timeless [130], an interaction which is implicated in stalled replication fork restart and progression. Timeless can stabilise the interaction between cohesin subunits and the chromatin, while the associated ChIR1 has been demonstrated to increase the activity of Fen1 [129]. Together with the ability of ChIR1 to unwind complex G4 structures and its interaction with the heterochromatin associated protein HP1, this suggests that ChIR1 has a functional role in facilitating DNA replication at difficult sites and that this acts in concert with its role in cohesion establishment (Figure 1.9).

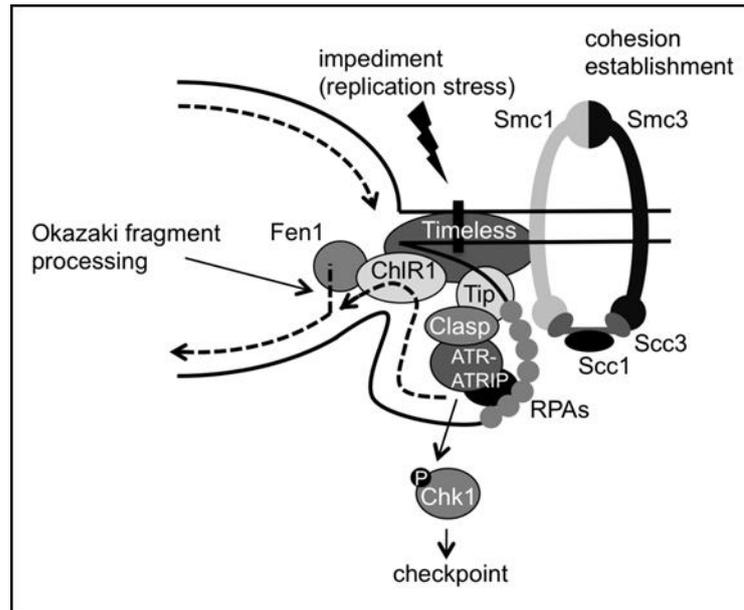


Figure 1.9 Link between replication and cohesion establishment [130]

Model to illustrate the interaction between cohesion establishment, DNA replication fork progression and the fork protection complex (FPC) [130]. The presence of DNA damage, DNA-bound proteins (including the cohesin complex) and complex DNA secondary structures can result in replication fork stalling. Single stranded DNA accumulates at these points and is coated by RPA. RPA can interact with the FPC subunit Tipin, which in turn recruits Timeless to the stalled fork. Timeless can then interact with cohesin subunits to stabilise the complex on the chromatin while the stalled fork is restarted. ChlR1 interacts with both Timeless and Fen1 to facilitate lagging strand processing, enabling progression of the stalled fork through the cohesin ring.

#### 1.4.3 Warsaw breakage syndrome

Mutations in many DNA helicases are associated with a range of diseases categorised by genome instability (Suhasini 2013). Fanconi's anaemia (FA - FancJ), Xeroderma pigmentosum (XPD), Werner's syndrome (WRN) and Bloom's syndrome (BLM) are all associated with

mutations in specific DNA helicases and are characterised by increased sensitivity to DNA damage. In 2010, a patient was identified as suffering from a novel disease which resulted in defective sister chromatid cohesion, causing a characteristic phenotype that placed the disease in a class of syndromes called cohesinopathies. A significant increase in cohesion defects, particularly premature chromatid separation following exposure to DNA damaging agents mitomycin C or camptothecin was typical of the cohesinopathies Roberts syndrome and Cornelia de Lange syndrome [131]. Increased susceptibility to chemically induced DNA damage was phenotypically similar to FA. Analysis revealed deletion of three codons within the DDX11 gene which encodes ChlR1 protein. This results in the deletion of a highly conserved lysine K897 residue. Biochemical analysis of this mutant protein has shown that ChlR1  $\Delta$ K897 is unable to bind DNA and therefore is defective as a helicase [117]. This novel syndrome that has phenotypic similarities to both a cohesinopathy and Fanconi's anaemia was called Warsaw breakage syndrome (WBS) after the domicile of the patient. Subsequently, further WBS patients have been identified that exhibit similar symptoms and have novel mutations in the DDX11 gene [132].

#### **1.4.4 ChlR1 function in E2-mediated PV genome maintenance**

ChlR1 has been shown to play an essential role in the maintenance of PV episomes through interaction with E2. Chl1 was identified as a BPV1 E2 interacting protein by a yeast two-hybrid screen, and subsequently ChlR1 was also shown to associate with both BPV1 and HPV11 E2 [95]. ChlR1 has been shown to coat the cellular chromatin while E2 forms discrete foci at the chromatin during early mitosis. E2 then remains associated with the chromatin throughout mitosis while ChlR1 was removed as cells progressed into metaphase. Interestingly, a BPV1 E2 mutant, W130R, which retains both transient replication and

transcription activity but which is defective in ChIR1 binding, showed a loss of association with mitotic chromosomes remaining diffusely cytoplasmic during mitosis. A similar phenotype was also observed in cells depleted of ChIR1 [95]. Cells transfected with BPV1 E2 W130R mutant genomes were also unable to maintain viral genomes to wild type levels over long term culture [95]. Together, these data suggest that the E2-ChIR1 interaction is required for the loading of the E2-viral DNA complex onto the mitotic chromatin. As ChIR1 does not remain chromatin associated throughout mitosis, it is likely that following the initial loading by ChIR1, E2 then binds to other chromatin associated proteins, remaining tethered during sister chromatid separation. This hypothesis is supported by further evidence showing that E2 and ChIR1 co-localise predominantly during mid-S phase [133] rather than during mitosis, suggesting that E2 utilises the interaction with ChIR1 to be positioned in close proximity to chromosomes during S-phase, allowing a subsequent interaction with an alternative chromatin associated protein (potentially Brd4 or TopBP1 as discussed above) to ensure E2-mediated tethering of the viral genomes to the cellular chromatin throughout mitosis.

## 1.5 Hypothesis and aims

**Hypothesis: The interaction between HPV E2 proteins and the cellular DNA helicase ChIR1 is required for viral genome maintenance.**

It has been well established that E2 interacts with viral genomes *via* its DBD and simultaneously associates with mitotic chromosomes through the TAD, thus tethering viral genomes to host chromosomes. Evidence from investigation with BPV-1 E2 suggests that ChIR1 functions in E2-mediated tethering of viral genomes to chromosomes prior to mitosis.

If the E2-ChIR1 interaction is direct, disruption would therefore impact on viral genome maintenance. Loss of the tethering mechanism by which viral genomes are partitioned into daughter cells following mitosis would lead to random segregation of viral genome and potentially exclusion from the nucleus, resulting in episome loss. Episome loss over time would result in a reduction in viral protein expression and a corresponding loss of the hyperproliferation and cellular transformation phenotypes observed upon productive HPV infection, potentially preventing a stable productive infecting and precluding completion of the viral life cycle.

E2 protein-protein interactions, and specifically the E2-ChIR1 interaction therefore pose an interesting novel potential target for the development of therapeutics to treat HPV infection. Validation of the E2 TAD as a target protein for drug development strategies along with further investigation into the E2-ChIR1 interaction is therefore proposed.

**Aims:**

1. Investigate the E2 TAD as a viable protein target for biophysical screening assays to develop novel small molecule therapeutics
2. Determine if a fragment based drug design strategy using HPV11 E2 TAD as a target protein can identify novel E2-binding small molecules by screening a commercial library of small molecule fragments using biophysical techniques
3. Determine whether HPV16 E2 and HPV11 E2 interact directly with ChIR1
4. Map the HPV16 E2-ChIR1 interaction site using a site-directed mutagenesis approach to identify an E2 mutant with reduced affinity for ChIR1
5. Investigate the importance of the E2-ChIR1 interaction on E2 function and localisation using a HPV16 E2 mutant which is defective for ChIR1 binding and ChIR1 knockdown experiments
6. Determine the role of the E2-ChIR1 interaction in the context of the HPV16 life cycle using a model tissue culture system

## Chapter 2. Materials and Methods

### 2.1 Materials

#### 2.1.1 Antibodies

Antibody	Host	Immuno-blotting dilution	Immuno-fluorescence dilution	Source
HPV16 E2 TVG261	Mouse	1:500	1:100	Abcam (ab17185)
HPV16 E2 TVG271	Mouse	-	1:100	Abcam (ab17190)
HPV16 E2 (poly)	Sheep	1:1000	1:100	Custom
ChIR1	Rabbit	1:500	1:100	Custom
$\beta$ -actin	Mouse	1:5000	-	Sigma
Vimentin	Rabbit	1:1000	1:100	Abcam (EPR3776)
H3	Rabbit	1:5000	1:100	Bethyl (A300-823A)
Grb2	Rabbit	1:1000	-	Cell Signalling (3972)
Orc2	Rat	1:1000	-	Cell signalling (4736)
EGFR	Rabbit	1:1000	-	Abcam (ab2430)
Lamin B1	Goat	-	1:100	Santa Cruz (sc-6217)
BrdU	Rat	-	1:200	Abcam (ab6326)
B-Actin	Mouse	1:5000	-	Sigma (A5441)
FLAG	Mouse	1:1000	1:100	Sigma (F1804)
Poly-Histidine	Mouse	1:1000	-	Abcam (ab18184)
HA.11	Mouse	1:1000	-	Covance (MMS-101P)

Mouse IgG (used for IP)	Mouse	-	-	Santa Cruz (sc-2025)
-------------------------	-------	---	---	----------------------

**Table 1.1 Details of source, dilutions required and host species for all primary antibodies used for immunoblotting and/or immunofluorescence**

*Secondary antibodies*

Western blotting

All antibodies were used at a dilution of 1:10000 in 5% milk in TBS/T, incubating for 1 hour at RT.

Goat $\alpha$ -Mouse HRP	Jackson Laboratories (115-035-003)
Goat $\alpha$ -Rabbit HRP	Jackson Laboratories (111-035-003)
Goat $\alpha$ -Rat HRP	Sigma (A5795)
Mouse $\alpha$ -Goat/Sheep HRP	Sigma (A9452)

Immunofluorescence (IF)

All antibodies were used at a dilution of 1:1000 in the buffer indicated, incubating for 1 hour at RT in the dark.

- Donkey  $\alpha$ -Mouse Alexa-Fluor 488 (Invitrogen A21202)
- Donkey  $\alpha$ -Goat Alexa-Fluor 594 (Invitrogen A11058)
- Donkey  $\alpha$ -Rabbit Alexa-Fluor 647 (Invitrogen A31573)

**2.1.2 Bacterial growth media and antibiotics**

Luria Broth (LB)

10 g LB powder (Sigma) made up in 400 mL ddH<sub>2</sub>O. Broth was autoclaved immediately and stored at RT

### LB agar

14 g LB agar powder (Sigma) was made up in 400 mL ddH<sub>2</sub>O and autoclaved immediately.

Agar was melted and cooled to approximately 60 °C before required antibiotics were added. The agar-antibiotic mixture was poured into 10 cm dishes in a sterile hood before being allowed to set at RT. Plates were stored inverted at 4 °C until required.

### Glycerol freezing solution

65% (v/v) glycerol

100 mM MgSO<sub>4</sub>

25 mM Tris pH 8.0

Make up to 100 mL with ddH<sub>2</sub>O; autoclave to sterilise before use.

### MgCl<sub>2</sub>/CaCl<sub>2</sub>

80 mM MgCl<sub>2</sub>

20 mM CaCl<sub>2</sub>

Working solution made as required from 100 mM stocks autoclaved prior to use.

## **2.1.3 Cell culture media and reagents**

### C33a growth media

DMEM (Sigma #D6249)

10 % FBS (Gibco)

### 3T3-J2 growth media

DMEM (Sigma #D6171)

10 % NBS (Gibco)

2 % L-Glutamine

### *Human foreskin keratinocytes*

#### Complete E media (CEM)

500 mL E media\*  
10 mL L-Glutamine  
2.5 mL EGF (mouse, BD 354010)

#### Organotypic raft media

500 mL E media\*  
10 mL L-glutamine

#### \*E media

300 mL DMEM D6171 (Sigma)  
160 mL Hams F12  
5 mL 100 x cocktail\*\*  
10 mL Pen/Strep (5000 U...)  
0.5 mL 1000x hydrocortisone (Sigma H0888)  
0.5 mL 1000x cholera toxin (Sigma C8052)  
25 mL FBS (Hyclone defined)  
Filter sterilised (using Corning bottles – details) and stored at 4 °C

#### \*\*100x cocktail

20 mL 0.18M adenine (Sigma A2786)  
20 mL Insulin (Sigma I6634, 5 mg/mL stock)  
20 mL Transferrin (Sigma T1147, 5 mg/mL stock)  
20 mL 2x 10<sup>-8</sup> M T3 (3,3,5-triiodo-1-thyronine, Sigma T6397)  
120 mL PBS  
Filter sterilised and stored in aliquots at -20 °C.

#### HFK cryogenic storage media

35 mL CEM  
5 mL Hyclone FBS

10 mL glycerol

Sterile filtered before use. Stored at 4 °C.

### 10 x reconstitution buffer (RCB)

0.26 M sodium bicarbonate (2.2 g)

0.2 M HEPES (4.77 g)

Dissolve in 100 mL 0.05 M NaOH

Filter sterilised, aliquoted and stored at -20 °C.

For a complete list of all other buffers and solutions used throughout, see Appendix 1.

## **2.2 Bacterial culture**

### **2.2.1 Bacterial strains**

Calcium competent *E. coli* strains BL21 for protein expression and DH5 $\alpha$  for plasmid preparation were grown in the lab and stored in aliquots at -80 °C (see 2.2.3).

Super-competent *E. coli* XL Gold (Agilent) were transformed following the manufacturer's protocol.

### **2.2.2 Preparation of calcium competent cells**

To prepare calcium competent *E. coli* strains BL21 for protein expression and DH5 $\alpha$  for plasmid preparation, a 100  $\mu$ L aliquot of bacteria was used to inoculate 5 mL LB. The culture was grown at 37 °C shaking at 200 rpm overnight. The 5 mL culture was then used to inoculate a further 200 mL LB broth and the culture was grown at 37 °C/200 rpm until OD<sub>600</sub> 0.5 was reached. The culture was cooled on ice for 20 minutes and transferred to pre-cooled 50 mL centrifuge tubes. Bacteria were pelleted by centrifugation at 4000 xg at 4 °C for 10 minutes. The bacterial pellets were resuspended in 30 mL ice-cold MgCl<sub>2</sub>/CaCl<sub>2</sub> and

centrifuged as before. Each pellet was then resuspended in 2 mL ice-cold 0.1M CaCl<sub>2</sub> plus 70 µL DMSO (Sigma) and incubated on ice for 15 minutes. A further 70 µL DMSO was added to each suspension and 100 µL aliquots were transferred to pre-cooled Eppendorf tubes. Aliquots were snap-frozen by immersion in liquid nitrogen before storage at -80 °C. Competency was tested by transforming aliquots with 10 ng, 100 ng and 1000 ng of plasmid DNA and plating onto agar plates with an appropriate antibiotic. Untransformed aliquots were also plated to check for contamination (antibiotic plate) and viability (no antibiotic).

### **2.2.3 Transformation**

100 µL aliquots of calcium competent bacteria were thawed on ice and incubated with 1-5 µL plasmid DNA for 1 hour on ice. Bacteria were then heat-shocked at 42 °C for 45 seconds and 900 µL LB broth was added. Cultures were shaken at 37 °C/210 rpm for 1 hour prior to a 200 µL aliquot being plated onto an LB-agar plate containing antibiotics for selection specific to the plasmid. Plates were incubated at 37 °C for 16 hours to allow colony growth. Single colonies were picked from plates and grown in 5 mL cultures containing LB broth supplemented with plasmid-specific antibiotics for selection at 37 °C with shaking at 210 rpm. Glycerol stocks were made from overnight cultures for long-term storage at -80 °C by adding glycerol freezing solution (see 2.1.2) at a 1:1 ratio with culture.

### **2.2.4 Plasmid DNA preparation**

5 mL (mini-prep) or 200 mL (maxi-prep) cultures were set up by inoculating LB broth supplemented with plasmid-specific selection antibiotics with a glycerol stock scraping. Cultures were incubated at 37 °C with shaking at 210 rpm overnight. Bacterial cultures were then centrifuged at 6000 xg, 4 °C, for 30 minutes to pellet cells. Plasmid DNA was extracted

from the bacterial cell pellets using DNA mini-prep or maxi-prep kits (Qiagen) according to the manufacturer's protocol. Mini-prepped DNA was dissolved in 50  $\mu$ L elution buffer (provided in the kit); maxi-prepped ethanol precipitated DNA was resuspended in 500  $\mu$ L TE buffer. Purified DNA was stored at -20 °C.

#### *DNA concentration determination*

DNA concentrations were determined using a Nanodrop 2000 spectrophotometer (Thermo), blanked against the buffer used for DNA resuspension.

### **2.2.5 Protein expression**

#### *General expression protocol*

Small starter cultures (5-100 mL) of LB with appropriate selection antibiotics were inoculated with transformed BL21 (scraping from glycerol stock or single colony picked from agar plate) and incubated at 37 °C with shaking at 210 rpm for 16 hours. Overnight cultures were used to inoculate larger expression cultures (diluting 1:100 unless otherwise stated) which were incubated at 37 °C/210 rpm until an OD<sub>600</sub> of 0.6-0.7 was reached. Protein expression was induced by addition of IPTG and the cultures were incubated for a further 4 hours. Cultures were then centrifuged at 6000 xg, 4 °C for 30 minutes to pellet the bacteria. The supernatant was removed and the pellet stored at -80 °C until required.

Protocol was used as written to express GST and GST-tagged HPV16 E2 truncations (2-50, 2-110, 2-170, 2-229, 25-229, 50-229 and 104-229).

### *Specific protocols*

His-tagged HPV11 E2 TAD and HPV16 E2 TAD WT were expressed using the above protocol with the following changes:

His-HPV11 E2 TAD expression cultures were incubated at 37 °C/210 rpm until OD<sub>600</sub> of 0.3 was reached then heated to 42 °C and incubated until OD<sub>600</sub> 0.6-0.7. The culture was then cooled to RT and 1 mM IPTG was added prior to incubation at 25 °/210 rpm for 16 hours.

HPV16 E2 TAD was expressed following the same protocol, inducing expression with 0.5 mM IPTG.

His-ChIR1 1-130 expression cultures were incubated at 37 °C/210 rpm until OD<sub>600</sub> 0.6-0.7 was reached. The culture was cooled to RT and protein expression induced by addition of 0.5 mM IPTG and incubation at 30 °C/210 rpm for 16 hours.

His-ChIR1 63-214 and His-ChIR1 63-222 expression was induced by addition of 0.05 mM IPTG and incubation at 37 °C/210 rpm for 4 hours.

MBP-His-ChIR1 63-222 expression was induced by addition of 0.3 mM IPTG and incubation at 30 °C/210 rpm for 6 hours.

## **2.3 Protein purification and analysis**

### **2.3.1 Protein purification**

#### *General lysis protocol*

Bacterial pellets were resuspended in the specified lysis buffer (3 mL per 50 mL culture) and incubated on ice for 30 minutes. Suspension was sonicated (35% amplitude, 1 minute (15 second plus 10 second intervals) per 3 mL culture) and centrifuged at 20,000 xg, 4 °C for 30 minutes. The lysate was removed from the pellet and purified as specified. Protein-specific details are below.

#### *Affinity chromatography*

His-tagged proteins were purified using an appropriate sized HisTrap FF Crude column (1 or 5 mL, GE), charged before use with NiCl<sub>2</sub> solution. Bound proteins were eluted with increasing imidazole concentration. Following elution, the column was stripped with 50 mM EDTA in PBS, washed with water and recharged.

GST-tagged proteins were purified using a 1 mL GSTrap FF column (GE) pre-packed with glutathione-sepharose. Bound proteins were eluted with increasing glutathione concentration. The column was stripped with 6M guanidine hydrochloride and washed in water following use.

Bacterial lysates were loaded onto an appropriate column pre-equilibrated in the specified buffer using a peristaltic pump flowing at 1 mL/min, collecting the flow-through. The bound proteins were then purified by step elution using a peristaltic pump, or by gradient elution using an AKTA Purifier (900/950) FPLC system connected to a UV detector to monitor protein elution by absorbance at 280 nm and a fraction collector, as specified. Following protein elution, collected fractions were analysed by SDS-PAGE and Coomassie staining or western blotting, as specified in the results.

### *Size exclusion chromatography*

Affinity column fractions determined to contain protein by SDS-PAGE were pooled and concentrated to a maximum volume of 10 mL using spin concentrators with an appropriate molecular weight cut-off at 3000 RPM for 20 minute intervals. Pooled protein was loaded onto a HiLoad 26/60 Superdex 200 gel filtration column (GE) pre-equilibrated in SE buffer, connected to an AKTA Purifier 900 FPLC system flowing at 1 mL/min. Protein was eluted using buffer C at 2.5 mL/min, monitoring protein elution by absorbance at 280 nm and collecting 3-10 mL fractions. Fractions were analysed by SDS-PAGE.

### *His-HPV11 E2 TAD and His-HPV16 E2 TAD*

Bacterial pellets resuspended in lysis buffer A (see Appendix 1) were sonicated and insoluble material was pelleted by centrifugation and the soluble lysate was removed. Crude lysate was used directly in pull down assays. HPV11 E2 TAD lysate was purified by affinity chromatography by gradient elution with the AKTA, and size exclusion chromatography.

HPV11 E2 TAD lysate was added to a 5 mL His-affinity column pre-equilibrated in lysis buffer A (with no added glycerol) using a peristaltic pump flowing at 1 mL/min. The column was then connected to the AKTA Purifier FPLC system and washed with 10 column volumes (CV) of buffer A, flowing at 4 mL/min to remove non-specific proteins. The bound protein was then eluted by increasing the concentration of elution buffer (buffer A plus 1 M imidazole) to 30% in a gradient over 10 CV, washing with 30% buffer B for 5 column volumes to ensure total elution. The column was then washed again with 5 column volumes buffer A to re-equilibrate. Fractions were collected and analysed by SDS-PAGE.

Fractions determined to contain HPV11 E2 TAD protein were pooled and purified further by SE chromatography following the above protocol. Purified HPV11 E2 TAD was dialysed into a storage buffer containing 10% glycerol at 4°C overnight. Protein was stored at -80 °C in 1 mL aliquots.

#### *His-ChIR1 1-130*

Bacterial pellets were resuspended in lysis buffer B (see Appendix 1), sonicated, and soluble protein was removed by centrifugation. Crude lysate was used directly in pull down assays.

#### *His-ChIR1 63-214, 63-222*

Bacterial pellets were resuspended in lysis buffer C (see Appendix 1), sonicated, and soluble protein was removed by centrifugation. Crude lysate was used directly in pull down assays.

#### *MBP-His-ChIR1 63-222*

Bacterial pellets were resuspended in lysis buffer D (see Appendix 1), sonicated and soluble protein was removed. Protein was purified by His-affinity chromatography. The column was washed with 3 column volumes (CV) lysis buffer before the protein lysate loaded at a flow rate of 1 mL/min using a peristaltic pump. The column was washed before bound proteins were eluted with 60 mM, 100 mM and 250 mM imidazole in lysis buffer, collecting 3 mL fractions.

Collected fractions were analysed by SDS-PAGE followed by Coomassie staining or western blotting. Fractions identified to contain MBP-His-ChIR1 63-222 protein were pooled and the imidazole removed by 5x volume buffer exchange using spin concentrators with a 5000 Da molecular weight cut off (MWCO). His-TEV protease was added to the MBP-His-ChIR1 63-222

at a final dilution of 1:100 and incubated for 16 hours at 4 °C. The cleaved MBP-His tag, uncleaved protein and the His-TEV protease were then removed from the cleaved ChIR1 63-222 by further His-affinity chromatography. The protein mixture was loaded onto the column at 1 mL/min, and the flow through collected and reapplied to the column a total of three times. The column was washed with 3x 3 CV lysis buffer and bound proteins were eluted in 5 mL 250 mM imidazole. The collected fractions were analysed by SDS-PAGE followed by Coomassie staining or western blotting.

#### *GST-ChIR1 peptides (182-193, 194-213)*

Pellets of bacteria expressing GST and GST-ChIR1 peptides were resuspended in lysis buffer E and sonicated. Soluble proteins were removed by centrifugation, and the crude lysate was used directly in pull-down assays.

#### *GST-HPV16 E2 truncations*

Bacterial pellets expressing GST and GST-HPV16 E2 truncations were resuspended in lysis buffer F and sonicated. Soluble proteins were removed by sonication and the insoluble pellet was resuspended in 1 mL Triton buffer (see Appendix 1) and incubated at 4 °C for 15 minutes. Insoluble proteins were pelleted by centrifugation and soluble lysate was removed. The pellet was washed with 2x 1 mL 20 mM Tris pH 8.0, and the final pellet was resuspended in 1 mL urea buffer (see Appendix 1) and incubated on ice for 90 minutes. Insoluble material was pelleted by centrifugation and the lysate was removed and added dropwise into 14 mL refolding buffer. The protein was incubated at 4 °C overnight to allow refolding. The refolded protein was then used in pull-down assays (see below).

### 2.3.2 Pull-down assays

For *in vitro* pull-down assays, protein expression was induced and bacteria lysed according to the specified protocols above. Protein lysate (1 mL per 100  $\mu$ L resin) was added to His- or GST-resin (determined by affinity tag, allowing 15  $\mu$ L resin in 50% slurry per pull-down sample), pre-equilibrated by washing in lysis buffer (3 x 500  $\mu$ L washes). Lysate was incubated with resin for 2 hours at 4  $^{\circ}$ C to allow specific binding of the tagged over-expressed protein to the resin. Resin was pelleted by centrifugation (500 xg, 4  $^{\circ}$ C, 5 minutes) then washed 3x with 500  $\mu$ L lysis buffer to remove unbound proteins. The resin was resuspended in an equal volume of lysis buffer and a 10  $\mu$ L sample was removed for analysis by SDS-PAGE to confirm isolation of the tagged protein.

C33a cells were transfected and lysed following the protocol laid out for co-immunoprecipitation (co-IP). 15  $\mu$ L protein-bound resin slurry was mixed with 300  $\mu$ L cell lysate and an equal volume of IP binding buffer (see Appendix 1), and incubated at 4  $^{\circ}$ C for 16 hours, mixing by gentle rotation. The resin was pelleted by centrifugation (500 xg, 4  $^{\circ}$ C, 5 minutes) and the unbound lysate was removed. The resin was washed 3 x with 500  $\mu$ L IP wash buffer and the resin was resuspended in an equal volume 2x SDS loading buffer. Samples were boiled at 95  $^{\circ}$ C for 10 minutes before analysis by SDS-PAGE and western blotting for the pulled down proteins.

Details of the proteins combined in pull-down assays are included in the results.

### **2.3.3 SDS-PAGE and coomassie staining**

Proteins were separated by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Gels were poured as required with the percentage of acrylamide in the resolving gel determined by the size of the protein to be resolved. Stacking gels were consistently poured at 6% (See Appendix 1).

Unless specified in the results, samples were made up containing 20  $\mu$ L protein sample and 4  $\mu$ L 6x loading buffer. Samples were boiled at 95 °C for 10 minutes and centrifuged for 10 seconds prior to loading. Gels were loaded with 20  $\mu$ L sample unless otherwise specified along with 5  $\mu$ L of a pre-stained protein ladder (PageRuler Plus, Biorline). Gels were run in SDS running buffer at 120 V for 30 minutes to allow samples to pass into the resolving gel and then the voltage was increased to 160 V for 60 minutes or until the proteins were sufficiently resolved as determined by the ladder.

Resolved proteins were analysed by staining with Coomassie (Fermentas Brilliant Blue) or by western blotting as specified. To Coomassie stain proteins the gel was washed with water to remove residual SDS buffer before immersing in Coomassie stain for a minimum of 1 hour at room temperature. Excess staining of the gel matrix was removed by washing in water and the gel was imaged for analysis.

### **2.3.4 Western blotting**

For western blotting the resolved proteins were transferred to PVDF membranes for immunoblotting. SDS gels were removed from between glass plates and were equilibrated in transfer buffer. Gels were layered in a stack with sponges and filter paper along with PVDF

membrane cut to the required size. Transfers were carried out at 100 V, 400 mA for 75 minutes in transfer buffer.

Following transfer membranes were blocked in 5% milk in TBS/T for 1 hour at room temperature. The membranes were then incubated with the specified primary antibodies at the dilutions specified for 1-16 hours as specified. Following primary antibody incubation, membranes were washed with TBS/T for 5x 5 minute washes to remove unbound antibody. The membranes were then incubated in the required secondary antibody diluted 1:10000 in 5% milk for 1 hour at room temperature. Following this incubation the membranes were again washed with TBS/T (5x 5 minute washes) and the proteins were detected by adding ECL (SuperSignal West Dura Extended Duration Solution, Thermo) to the membranes and visualising using a Fusion FX100 chemiluminescence camera.

### **2.3.5 Bradford assay**

Bradford assay reagent was diluted (1:5) with PBS. Protein standards were made by serial 1:1 dilution of 2 mg/mL bovine serum albumin (BSA, Sigma) to make standards ranging from 1 – 0.065 mg/mL. 10 µL cell lysate or purified protein solution was added to each well of a (clear/optical) 96 well plate, along with the concentration standards. Each sample was run in triplicate, diluting samples 1:10-1:100 as necessary to fit within the standard curve. 200 µL of the diluted Bradford reagent was added to each well and mixed. The plate was read using a Biorad iMARK plate reader at 595 nm.

A standard curve was generated from the average of the protein standard data, fitting to a linear curve. The sample concentrations were determined by fitting to the equation of the straight line generated.

## 2.4 Biophysical protein assays

### 2.4.1 Thermal shift assay (DSF)

#### *Fragment screening*

Purified HPV11 E2 TAD in 20 mM phosphate pH 7.4, 200 mM NaCl buffer was concentrated to 4 mg/mL using spin concentrators. Thermal shift assays were set up in 96 well plates with each well containing a final concentration of 8  $\mu$ M protein. Small molecule fragments (see appendix 2 A.1 for details) were added to a final concentration of 10 mM along with 50  $\mu$ L of 50 mM EPPS pH 8.5 buffer containing 2x Sypro Orange (from 5000x stock – Invitrogen). Each well was then made up to a final volume of 100  $\mu$ L with EPPS buffer.

Plates were heated to 96 °C at 1 °C/min and Sypro Orange fluorescence was monitored using a Stratagene Mx3005 RT-PCR machine. Data was analysed using a DSF analysis excel workbook V3 from Frank Nieson, SGC, Oxford [134, 135] and GraphPad Prism V5. Data obtained from experimentally obtained fluorescence curves were fitted to the Boltzmann sigmoidal equation to find the melting temperature ( $T_m$ ).

This process was repeated with selected fragments (as specified in results) using 5  $\mu$ M protein/well in different buffers: 50 mM EPPS pH 8.5; 50 mM MOPS pH 7.2; 50 mM HEPES pH 7.5 and 100 mM Tricine pH 8, 100 mM NaCl, 10 mM DTT.

#### *Buffer screening*

Purified HPV11 E2 TAD at 2.5 – 5 mg/mL stock concentration was diluted to the specified concentration (results) to a final volume of 10  $\mu$ L in the specified buffer. A further 10  $\mu$ L of buffer containing 2x Sypro Orange was added to each well and mixed well. The plate was

then heated as previously specified and the  $T_m$  determined for each condition. Details of the buffer composition and concentrations used are provided in the results.

#### *NMR buffer screening*

Purified HPV11 E2 TAD was concentrated to 2.5 mg/mL (100  $\mu$ M) using spin concentrators. 2  $\mu$ L Sypro Orange (5000 x) was added to 120  $\mu$ L protein solution. 2  $\mu$ L of this stock was added to each well of a 96 well plate along with 18  $\mu$ L of the specified buffer. Each well was mixed and the plate centrifuged for 2 minutes at 400 xg. The plate was then heated to 96 °C at 1 °C/min using a Stratagene Mx3005p RT-PCR machine set to excitation wavelength of 492 nm, emission wavelength of 568 nm and a gain of 1, 4 and 8. From the obtained dissociation curve, the initial lowest point on the curve was taken to be the  $T_m$  point.

The initial screen was carried out using a pre-designed set of buffers and additives used for NMR experiments. Subsequent buffers were determined based on the DSF results, combining optimal buffer and additive conditions to maximise protein stability (see appendix for buffer details).

#### **2.4.2 Nuclear Magnetic Resonance (NMR) Spectroscopy**

Purified protein was concentrated to approximately 190  $\mu$ M (5 mg/mL HPV11 E2 TAD) using spin concentrators (MWCO 15 kDa; 1800 xg, 4 °C, 30 minute spins). NMR samples were made up by combining protein for a final concentration of 20  $\mu$ M with 6  $\mu$ L of pre-prepared fragment cocktail (see below) made up to 750  $\mu$ L with NMR buffer in an NMR tube.

Fragment cocktails: Small molecule fragments were made up to 1M in  $d_6$ -DMSO. Fragments were cocktailled in sets of 8 determined based on the spectral overlap of the compounds, at

a final concentration of 125 mM each. Each fragment is present in the NMR sample at a concentration of 1 mM.

NMR experiments were carried out using a Bruker Avance 500 NMR spectrometer with a Bruker Ultrashield 500 MHz magnet using a TXI 5 mm inverse triple resonance probe. The NMR spectrometer was connected to a 60 slot automated sample changer and was controlled by a 3 channel Bruker Avance console. 1D proton and water suppression experiments were carried out initially before STD and wLOGSY spectra were obtained for each sample in turn. Raw data was processed using TopSpin and then fragment binding was assessed by comparing experimental spectra with a database of fragment reference spectra using Bruker Amix software.

#### **2.4.3 Hanging drop protein precipitation assay**

1  $\mu$ L purified HPV11 E2 TAD (5 mg/mL) was added to a circular coverslip. 1  $\mu$ L of buffer was added to the protein and the coverslip was inverted over a well containing 1 mL of buffer. An airtight seal was generated by addition of grease around the outside of the well before coverslip positioning. Plates were incubated at RT, protein precipitation was assessed by light microscopy at 2 and 7 days.

### **2.4 DNA purification, analysis and cloning**

#### **2.4.1 Agarose gel electrophoresis**

Agarose gels were made by dissolving the required amount of agarose powder in 1 x TBE buffer and heating until a homogenous solution was achieved. The solution was allowed to cool slightly before ethidium bromide was added (2.5  $\mu$ L/50 mL gel; 5  $\mu$ L/500 mL gel). The solution was then poured into a gel caster containing the required comb and was allowed to

set at room temperature. The set gel was then transferred to a suitable tank and submerged in 1 x TBE buffer. 5x DNA loading buffer was added to DNA samples and the mixture was loaded onto the gel, along with 5  $\mu$ L of an appropriate DNA ladder (hyperladder 1kb or 100bp). The agarose gels were run at 80 mA unless otherwise stated. DNA was visualised using a UV transilluminator.

#### **2.4.2 Phenol-chloroform DNA extraction**

##### *DNA purification*

An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1, Sigma) was added to DNA and the mixture vortexed for 10 seconds before centrifugation (16000 xg, 4 °C, 10 minutes). The upper aqueous layer was removed to a clean tube and the DNA precipitated by addition of 2 volumes 100% ethanol and 0.1 volumes 3 M sodium acetate pH 5.2. and incubation at -20 °C overnight. DNA was pelleted by centrifugation (20000 xg, 4 °C, 10 minutes) and the supernatant was removed. The DNA pellet was air-dried before resuspension in sterile TE buffer. DNA concentration was determined as described.

##### *DNA extraction from HFK cell pellets*

HFK cell pellets (see 2.5.1.2) were thawed on ice and resuspended in 3 mL DNA lysis buffer. RNase A was added to a final concentration of 400 ng/mL, the suspension was vortexed for 5 seconds and then incubated at 37 °C for 30 minutes. Proteinase K was then added to a final concentration of 50  $\mu$ g/mL along with 60  $\mu$ L 10% SDS and the resulting mixture was incubated at 37 °C overnight.

The DNA was then sheared by slowly passing through an 18 gauge needle 10 times. The sheared DNA was then extracted twice with 6 mL phenol: chloroform: isoamyl alcohol (25:24:1) (Sigma) then twice with chloroform:isoamyl alcohol (24:1) (Sigma), in each case inverting the extraction 8 times and centrifuging at 1800 xg for 5 minutes before removing the aqueous DNA containing layer to a fresh tube.

Following extraction, 0.3 M sodium acetate (pH 5.2) and 6 mL ice-cold ethanol were added to the DNA suspension and the mixture was incubated at -20 °C overnight.

The ethanol precipitated DNA was centrifuged at 1800 xg for 30 minutes and the ethanol was removed. The DNA pellet was washed with 5 mL 70% ethanol and centrifuged for a further 15 minutes at 1800 xg. After removing the ethanol wash, the DNA pellet was washed with a further 1 mL 70% ethanol and transferred to an Eppendorf tube before centrifuging at 10000 xg for 10 minutes. The ethanol supernatant was removed and the DNA pellet air dried for 2 minutes before resuspension in 100 µL TE buffer (see DNA prep) and incubating at 37 °C for 1 hour. The DNA concentration and purity was determined using a Nanodrop and the DNA stored at -20 ° until required.

#### **2.4.3 Restriction digest**

For cloning ChIR1 peptides into a pGEX-4T-1 plasmid, 2 µg plasmid DNA was digested with 1 µL each *EcoRI* and *XhoI* enzymes in buffer H (Promega). Reactions were made up to 20 µL with nuclease free H<sub>2</sub>O and were incubated at 37 °C for 2 hours. Digested plasmid was purified by gel extraction.

For Southern blotting, pUC19-HPV16 114/K was digested with *BamHI* to linearise the genome. To generate copy number controls, 2 µg mini-prepped plasmid DNA was digested

with 1  $\mu$ L *Bam*HI (Promega) in buffer H for 2 hours at 37 °C. Digested DNA was stored at -20 °C until required. To generate DNA for radiolabelled probes, 1  $\mu$ g pUC19-HPV16 114/K DNA was digested with 1  $\mu$ L *Bam*HI overnight at 37 °C. The 8kb HPV16 genome fragment was gel purified and stored at -20 °C until required.

#### 2.4.4 Gel purification

Ethidium bromide-stained DNA bands were visualised on a UV transilluminator and excised from the gel using a clean scalpel. DNA was extracted from the agarose using a GenElute gel purification kit (Sigma) following the manufacturer's protocol. DNA was eluted into 50  $\mu$ L TE buffer and the concentration determined using a Nanodrop prior to storage at -20 °C.

#### 2.4.5 Ligation (ChIR1 fragment cloning)

Fragments of ChIR1 were ligated into a pGEX-4T-1 plasmid to allow production of GST-tagged peptides. Primers were designed with ends complementing the *Eco*RI and *Xho*I restriction sites in the pGEX plasmid (Table 2.3) 20  $\mu$ L of each forward and reverse primer (100  $\mu$ M) were combined and heated to 95 °C for 2 minutes then cooled to RT to form double-stranded DNA inserts (~1ng/ $\mu$ L). 0.5  $\mu$ g pGEX vector was digested with *Eco*RI and *Xho*I in Promega buffer H for 2 hours at 37 °C before running on a 0.8% agarose gel. The linearised plasmid DNA was gel purified using a GenElute gel extraction kit (Sigma) following the included protocol and the concentration determined using a Nanodrop 1000 as previously described.

Name	T <sub>m</sub> /°C	Primer sequence
ChIR1 182-193 Fw	59	AAT TCC CAG AAG CTG AAC GTT TAG AAC AAC TTG AAT CAT GAC

ChIR1 182-193 Rv	61	TCG AGT CAT GAT TCA AGT TGT TCT AAA CGT TCA GCT TCT GGG
ChIR1 194-213 Fw	67	AAT TCG GAG AAG AAG AAC TTG TGC TTG CTG AAT ATG AAT CAG ATG AAG AAA AGA AGG TGG CTT CAC GTT GAC
ChIR1 194-213 Rv	68	TCG AGT CAA CGT GAA GCC ACC TTC TTT TCT TCA TCT GAT TCA TAT TCA GCA AGC ACA AGT TCT TCT TCT CCG
ChIR1 239-256 Fw	66	AAT TCT TAG CAC AAT TCG TAC ACG AAG TAA AGA AGT CAC CAT TCG GAA AGG ACG TAC GTT GAC
ChIR1 239-256 Rv	67	TCG AGT CAA CGT ACG TCC TTT CCG AAT GGT GAC TTC TTT ACT TCG TGT ACG AAT TGT GCT AAG

**Table 2.3 Primer sequences used to clone ChIR1 peptides into pGEX-4T-1 vector**

To ligate the ChIR1 fragments into the pGEX vector, 100 ng digested pGEX DNA was combined with 4 ng annealed primer insert, 10 x ligation buffer and a T4 DNA ligase (details), made up to 20  $\mu$ L with water. Ligation reactions were then incubated at RT overnight.

The ligation reactions were used to transform DH5 $\alpha$  as previously described. Single colonies were picked and DNA extracted following the mini-prep protocol previously discussed. DNA was digested with *Eco*RI and *Xho*I as previously described to confirm ligation by looking for the insert dropout. Promising clones were sequenced to confirm the presence of the ChIR1 fragment insert. The sequenced plasmid DNA was then used to transform BL21 as previously discussed for bacterial protein expression and use in *in vitro* binding assays.

#### **2.4.6 Sequencing**

Sequencing was carried out by University of Birmingham Sequencing facility. Sequencing reactions were made up containing 400 ng DNA, 1.1  $\mu$ L primer (0.3 pM) made up to 11  $\mu$ L with nuclease free H<sub>2</sub>O. 10  $\mu$ L of reaction mix was loaded for sequencing.

HPV16 genomes were sequenced using an overlapping set of primers covering the entire genome (Table 2.4).

ID	Anneals at HPV16 nucleotide	Primer sequence
E6 111	111	AGG ACC CAC AGG AGC GAC CC
LateProm	649	GAC AGC TCA GAG GAG GAG GA
E1 1250	1250	GCG AAG ACA GCG GGT ATG GCA
E1 2158 Fw	2158	AGG GTA GAT GAT GGA GGT GAT TGG
E1 2158 Rv	2316	GAT TTA CCT GTG TTA GCT GCA CCA
E2 CTCF Fw	2853	GGA AAC ACA TGC GCC TAG AAT GTG C
E2 CTCF Rv	2950	TGA TAC AGC CAG TGT TGG CAC C
E4	3407	CAC TCC GCC GCG ACC CAT AC
E5	3936	ACG TCC GCT GCT TTT GCT TGT GT
L2	4419	CAG GGT CGG GTA CAG GCG GA
L2 CTCFbs Fw	5123	AGG CGT ACT GGC ATT AGG TAC AGT
L2 CTCFbs Rv	5309	AGG TAA GGC TGC ATG TGA AGT GGT
L1 CTCFbs Fw	6039	TGC AGC AAA TGC AGG TGT GGA T
L1 CTCFbs Rv	6157	TGG GGA TCC TTT GCC CCA GTG T
L1	7045	ACA AGC AGG ATT GAA GGC CAA ACC A
L1 Rv	7121	AGA GGT AGA TGA GGT GGT GGG TGT
Enh	7555	CCA AAT CCC TGT TTT CCT GA

**Table 2.4 – primers used for sequencing the HPV16 genome**

HPV16 DNA extracted from HFKs was sequenced using E2 CTCF forward and reverse primers. HPV16 E2 mutants were sequenced using MDP234 (forward) and MDP236 (reverse) primers. ChIR1 peptides were sequenced using a pGEX-F primer.

ID	Primer sequence
MDP234	TCC TAC AGC TCC TGG GCA ACG
MDP236	AGA AGT CAG ATG CTC AAG GG
pGEX-F	GGG CTG GCA AGC CAC GTT TGG TG

**Table 2.5 Primer sequences used to sequence plasmids. All primers are shown in 5' – 3' orientation**

#### 2.4.7 Site-directed mutagenesis

Specified template DNA was obtained by mini-prep as described. Site directed mutagenesis was carried out using primers designed to introduce single or multiple point mutations into the HPV16 E2 sequence (see primer details in Table 2.6) using *Pfu* Ultra II HS DNA polymerase (Agilent). To generate E2 mutants in the plasmid pCMV-16E2, PCR reactions were set up containing 5  $\mu$ L *Pfu* Ultra II 10x buffer, 200  $\mu$ M dNTPs, 0.2  $\mu$ M primer (forward and reverse), 100 ng DNA template and 1  $\mu$ L DNA polymerase made up to a final volume of 50  $\mu$ L with water. PCR reactions were carried out as follows.

PCR cycles

1. 95 °C, 120 seconds (hot start for enzyme activation)
2. 95 °C, 20 seconds
3. 50 °C, 30 seconds
4. 72 °C, 180 seconds → back to step 2, repeat x2
5. 95 °C, 30 seconds

6. 60 °C, 30 seconds
7. 72 °C, 180 seconds → back to step 5, repeat x27
8. 72 °C, 180 seconds
9. Cool to 4 °C

A sample of the PCR reaction (3 µL PCR reaction, 1 µL 5x DNA loading buffer, 1 µL water) was then run on a 1% agarose gel at 400 mA for 90 minutes and PCR products were visualised by UV detection of intercalated ethidium bromide. The PCR reaction products were then subjected to *DpnI* digestion to remove input template DNA by adding 3 µL *DpnI* (New England Biolabs) and incubating at 37 °C for 2 hours. DNA fragments were removed using a PCR clean-up kit (Sigma) according to manufacturer's instructions and the newly amplified plasmid DNA was used to transform super-competent *E.coli* XL10-Gold (Agilent), following the manufacturer's protocol. Plasmid DNA was obtained by mini-prep as described previously and the plasmid DNA was sequenced to confirm mutagenesis.

To clone the Y131A mutant into the pMEP-16E2 vector, the same PCR reaction mix and PCR cycles were used as described for E2 mutagenesis.

To clone the Y131A mutant into the pUC19-HPV16 114/K plasmid, PCR reactions containing 10 ng of plasmid DNA (same protocol as above) were set up. PCR cycles as follows:

1. 95 °C, 60 seconds
2. 95 °C, 50 seconds
3. 60 °C, 50 seconds
4. 68 °C, 11 minutes → back to step 2, repeat x18
5. 68 °C, 7 minutes

PCR products were analysed and purified as previously described.

Mutant	Template	Plasmid	Primer Sequence (5'-3')	Tm / °C
E118A	HPV16 E2	pCMV-16 E2	catggatatacagtggcagtgcagtttgatggagacata tgcaatacaatgc	69
E118A/T116A	E118A		cgatgtataaaaaaacatggatatgcagtggcagtgc gtttgatggagacatatgc	70
K177A	HPV16 E2		cgaacatattttgtgcagtttaaagatgatgcagaagcat atagtaaaaaataaagatgggaagttcatgcggg	70
D124A	HPV16 E2		gtgcagtttgatggagccatatgcaatacaatgcattata caaactgg	67
E118A/D124A	E118A		Gtgcagtttgatggagccatatgcaatacaatgcattat acaactgg	67
E118A/T116A/D 124A	E118A/ T116A		Gtgcagtttgatggagccatatgcaatacaatgcattat acaactgg	67
D173A	HPV16 E2		Gaaggaatacgaacatattttgtgcagtttaaagctgat gcagaaaaatatagtaa	65
D173A/K177A	K177A		Tattttgtgcagtttaaagctgatgcagaagcatatagta aaaataaagtatgg	67
Y131A	HPV16 E2		Gatggagacatatgcaatacaatgcatgctacaaactg gacacatatatat	66
H130R/Y131A	H130R		Gatggagacatatgcaatacaatgcgtgctacaaactg gacacatatatat	67
Y131A	HPV16 114/K	pUC19	Gatggagacatatgcaatacaatgcatgctacaaactg gacacatata	66
Y131A	pMEP- HPV16 E2	pMEP4	tgcaataccatgcatgccaccaattggaccac	66

**Table 2.6 Primer sequences used for site-directed mutagenesis. In each case the reverse primer is the reverse complement of the given sequence.**

#### 2.4.8 Constructs

Plasmid	Codes/contains	Source	Tag	Antibiotic
pCMV-16E2	HPV16 E2	Iain Morgan, University of Virginia, USA	-	Amp
pCMV-tag2A	ChIR1	Jill Lahti, St Jude Children's Hospital, USA	FLAG	Kan
	HPV16 E1	Mart Ustav, Estonia	HA	Amp
pCMV	<i>Renilla</i> Luciferase	Promega	-	Amp
pFLOri	HPV31 Ori (replication)	J. Archambault	-	Amp
ptk	6-E2 firefly (transcription)	Iain Morgan	-	Amp
pUC19	HPV16 114/K genome	Ethel-Michele de Villiers, DKFZ, Germany	-	Amp
pMEP4	HPV16 E2 (codon optimised)	A. McBride, NIH Bethesda, USA	FLAG	Amp

**Table 2.7 Details of the plasmid constructs used**

#### 2.4.9 Southern blotting

*DNA preparation*

5 µg DNA extracted from HPV16 genome-containing HFKs (see Section 2.4.3) was digested with either high-fidelity *Bam*HI (to cut within the HPV genome) and *Dpn*I (to remove input DNA), or *Hind*III and *Dpn*I (to cut outside the HPV genome) in a 20 µL reaction at 37 °C overnight. In both cases the buffer for the specific restriction enzyme was used (Promega buffer E). Copy number controls were generated by digesting maxi-prepped pUC19-HPV16 with *Bam*HI for 2 hours at 37 °C and diluting to the required concentration (equivalent to 5, 50 or 500 copies/cell) with salmon sperm DNA.

5x DNA loading buffer was added to the digested sample DNA and the copy number controls which were analysed by agarose gel electrophoresis, running on a 0.8% gel for ~20 hours at 50 V (in TBE containing 10 µL/L ethidium bromide). The presence of DNA on the gel was determined by UV detection of incorporated ethidium bromide.

### *Transfer*

The agarose gel was washed twice in 500 mL 0.25M HCl for 20 minutes at RT to depurinate the DNA followed by two 30 minute RT washes in 500 mL 0.4 M NaOH to neutralise the acid and denature the DNA. The DNA was then transferred to a nylon membrane by capillary action.

A stack was made by layering 4 pieces of filter paper (Whatman TM 3MM cut to 24 x 33 cm) over a glass plate suspended above a reservoir containing 2 L 0.4 M NaOH, with the ends of the filter paper resting in the solution. The agarose gel was then placed in the centre of the filter paper. A nylon membrane (Gene Screen Plus, Perkin Elmer) cut to fit the gel and pre-soaked in NaOH was placed over the gel, followed by a further 4 pieces of filter paper (cut to 21 x 23.5 cm). At each stage any air bubbles were removed by rolling. A stack of absorbent

paper towels (~ 10 cm height) was placed on top of the filter paper, followed by a glass plate and a weight. Finally the reservoir was covered over with Saran wrap separating the top and bottom filter paper layers. The stack was then left overnight at RT to allow transfer of DNA from the gel to the membrane to occur.

After 18 hours the paper towels and top layers of filter paper are removed. The gel and membrane were flipped over so the gel is on top and the position of the wells was marked onto the membrane with pencil. The gel was checked for remaining un-transferred DNA using a UV transilluminator before disposal. The DNA was cross-linked to the membrane using a Stratalinker UV cross-linker 90 (Stratagene) before wrapping in Saran wrap and storing at -20 °C.

#### *Radiolabelled DNA probe preparation*

Maxi-prepped pUC19-HPV16 was digested with *Bam*HI and the linearised HPV16 genome was gel purified as described (2.4.6). 50 ng of the purified DNA was diluted into a final volume of 45 µL with TE buffer and the DNA was denatured by heating to 100 °C for 10 minutes before cooling in ice-water slurry for 2 minutes. The denatured DNA was added to a Ready To Go DNA labelling beads-dCTP kit (Amersham Biosciences) following the manufacturer's instructions. After resuspension of the labelling beads in the denatured DNA, 50 µCi [ $\alpha$ -<sup>32</sup>P]-dCTP (Perkin Elmer) was added and the mixture was incubated at 37 °C for a further 30 minutes. The labelled probe was then purified using a radiolabeled probe purification kit (Illustra Probe Quant G-50 microcolumn (Amersham Biosciences) following the supplied protocol. The purified probe was stored at -20 °C until required.

#### *Hybridisation*

200  $\mu$ L of a 10 mg/mL stock of salmon sperm DNA (Invitrogen) was denatured by heating to 100 °C for 10 minutes before cooling in ice-water slurry for 2 minutes. The denatured DNA was added to 10 mL 1x hybridisation buffer (see 2.1.6) to make a pre-hybridisation buffer. The nylon membrane containing the transferred DNA was placed onto a square of gauze pre-soaked in water. The membrane and gauze were then rolled up and placed into a hybridisation tube (Hybaid) along with the pre-hybridisation buffer. The canister was then incubated in a rotating hybridisation oven at 42 °C for 1 hour. To prepare the hybridisation solution, a further 200  $\mu$ L aliquot of 10 mg/mL salmon sperm DNA was added to the radiolabelled probe and the mixture was denatured by heating to 100 °C for 10 minutes prior to immersion in an ice-water slurry for 2 minutes. The denatured probe DNA mixture was added to a further 10 mL of 1x hybridisation buffer. The pre-hybridisation buffer was removed from the membrane-containing canister and the hybridisation solution containing the probe was added. The canister was then incubated in the rotating oven at 42 °C overnight.

### *Detection*

Following hybridisation with the radiolabelled probe, the hybridisation solution was discarded and the membrane removed from the hybridisation tube. The membrane was submerged in a wash buffer comprising 2 x SSC + 0.1% w/v SDS (see 2.1.6) and a sponge was used to wipe away excess hybridisation solution. The buffer was discarded and the membrane washed in a further 400 mL high stringency buffer 1 for 15 minutes at RT, before repeating this wash with a further 500 mL of buffer 1. The wash buffer was removed and the membrane was then subjected to 2 x 15 minute washes at RT with each of wash buffers 2

and 3 in succession. The membrane was then washed in pre-heated wash buffer 4 at 55 °C for 30 minutes. The excess SDS was removed by rinsing the membrane in 2x SSC buffer and the membrane was wrapped in Saran wrap. Radiolabelled DNA was detected by exposing the membrane to autoradiography film (Amersham Biosciences) in a cassette containing an intensifying screen at -20 °C. Film was developed after 24 and 72 hours exposure using a Xograph Healthcare Compact 4x automatic processor. The membrane was also used to expose a PhosphorImager screen for 48 hours before the plate was scanned using a Typhoon PhosphorImager.

## **2.5 Tissue culture (cell assays)**

### **2.5.1 Cell lines and culture**

HPV negative cervical carcinoma C33a cells were maintained in Dulbecco's modified eagle medium (DMEM) containing 4500 mg/L glucose, L-glutamine, sodium pyruvate and sodium bicarbonate (Sigma), supplemented with 10% foetal bovine serum (FBS) and were routinely passaged at a 1:10 ratio every 3-4 days.

3T3-J2 murine fibroblasts were maintained in DMEM containing 4500 mg/L glucose, 25 mM HEPES and sodium bicarbonate (Sigma D6171) supplemented with 10% newborn calf serum (NBCS) and 2% L-glutamine and were routinely passaged at a 1:10 ratio every 4-5 days. Confluent plates of 3T3-J2s were harvested by trypsinising and resuspending in growth media. Cells were counted and centrifuged (200 xg, 3 minutes) and the media was removed. Pelleted cells were resuspended in complete E medium to a final density of  $2 \times 10^6$  cells/mL.

Resuspended cells were then subjected to 40 grays of gamma-irradiation and were stored at 4 °C for up to 5 days.

#### *2.5.1.1 pMEP cell lines*

pMEP4-16 E2 was obtained from Alison McBride (NIH, USA). pMEP4-16 E2 Y131A was generated by site-directed mutagenesis (see 2.4.9 for details). C33a cells were seeded at a density of  $3 \times 10^6$  cells in 10 cm dishes and were transfected with 5 µg DNA (WT or Y131A pMEP-16 E2) using XtremeGene HP as detailed in 2.5.2. 24 hours post-transfection media was changed and supplemented with 200 mg/mL hygromycin B (concentration determined by kill curve on untransfected C33a cells). Transfected cells were maintained in hygromycin containing media changed every 2-3 days. E2 expression was induced by addition of 150 µM zinc chloride to growth media 4-6 hours prior to harvesting (specific concentrations of ZnCl<sub>2</sub> and induction times were experimentally determined, as discussed in Chapter 5).

#### *2.5.1.2 Primary human foreskin keratinocytes*

HFK lines containing WT or mutant HPV16 genomes were generated from primary HFKs from three donors designated Jerry, Nigel and Freddie.

#### *Preparing HPV genomes*

HPV16 genome containing plasmids (WT and Y131A E2, pUC19 vector – see mutagenesis for details of cloning) were digested by adding 10 µg DNA to 1 µL *Bam*HI (Promega) with 5 µL 10x buffer (E) and the reaction volume was made up to 50 µL with nuclease-free water. The reaction was gently vortexed and centrifuged for 10 seconds and then incubated at 37 °C in a water bath overnight. The digestion was checked by running 1 µL (200 ng) of the reaction on

a 1% agarose gel. Once digestion had run to completion the *Bam*HI enzyme was heat-inactivated by incubating at 65 °C for 20 minutes.

To the inactivated reaction was then added 669 µL nuclease-free water, 180 µL 5x ligation buffer (Invitrogen #Y90001) and 1 µL T4 DNA ligase (high concentration, New England Biolabs #M0202S). The reaction was gently vortexed and centrifuged for 10 seconds before incubating at 16 °C in a water bath in a cold room overnight.

The religated DNA was centrifuged for 30 seconds and divided between two tubes. 600 µL isopropanol and 180 µL 5 M NaCl were added to each tube and the reaction was mixed before incubating at -20 °C overnight. The reactions were then centrifuged at 20000 xg for 30 minutes at 4 °C to pellet the precipitated DNA. The supernatant was removed and the DNA pellets were washed with 1 mL 70% ethanol before centrifuging for a further 15 minutes. The supernatant was carefully removed and the pellet allowed to air dry briefly. The pellets were then resuspended in 12 µL TE buffer and 1 µL was ran on a 1% agarose gel to check the religation. The remaining DNA was stored at -20 °C until required.

#### *Transfection and selection*

Early passage HFKs were plated onto 6 cm dishes at a density of  $5 \times 10^6$  per dish and were grown to around 50-60% confluency in complete E media. Cells were transfected with 1 µg religated HPV DNA and 1 µg neomycin-resistant marker plasmid, pSV<sub>2</sub>Neo using XtremeGene HP (see below). The following day transfected cells were trypsinised and re-plated onto 10 cm dishes in complete E media supplemented with  $2 \times 10^6$  irradiated 3T3-J2 feeder cells that had been seeded the previous day and allowed to settle. The next day, the media was changed and G418 (100 µg/mL) was added to select for transfected cells. Cells were grown

under selection for 8 days, changing the media and adding fresh irradiated 3T3-J2 feeder cells every two days (and adding fresh G418 every 2 days).

Following selection colonies were allowed to grow out, with fresh media being added every 2 days and fresh irradiated 3T3-J2 cells added as necessary until colonies were approximately 1-2 cm in diameter. Colonies were then pooled and cells maintained as described below.

### *Maintenance*

HFK lines were seeded at a density of  $1 \times 10^5$ - $5 \times 10^5$  cells onto plates pre-seeded with  $2 \times 10^6$  irradiated 3T3-J2 feeder cells in 10 mL complete E media. Media was changed every two days and fresh feeder cells were added as necessary (minimum of every 4-5 days). Cells were grown to ~80% confluency before trypsinising and counting. At each passage, cells were reseeded to continue growing as before, vials of cells were frozen down for liquid nitrogen storage ( $5 \times 10^5$ - $1 \times 10^6$  cells per vial) in HKF freezing media and cell pellets of up to  $1 \times 10^6$  cells were stored at  $-80$  °C for DNA extraction.

### *Organotypic raft culture*

To prepare collagen plugs for organotypic raft culture, confluent dishes of 3T3-J2 cells were harvested and counted.  $2 \times 10^6$  3T3-J2 cells per plug were pooled and centrifuged (200 xg, 5 mins) to pellet. A mixture of 10 x DME (0.3 mL/plug sterile filtered prior to use) and RCB (0.3 mL/plug, sterile filtered before use) was used to resuspend the cell pellet. 2.4 mL rat tail collagen (diluted to 4 mg/mL in 0.02 M acetic acid) was added per plug using an ice-cold stripette, mixing well. 1M NaOH (sterile filtered prior to use) was added dropwise to correct

the pH of the mixture, until a red/orange colour was achieved. The solution was mixed well and 3 mL was added to each 3 cm dish as required. The plugs were then incubated at 37 °C for 30 minutes before 2 mL CEM was added. The plugs were then kept at 37 °C.

After 24 hours, HFks were harvested and  $2 \times 10^6$  cells/plug were pelleted by centrifugation. The pelleted cells were resuspended in 2 mL CEM/plug. The media was removed from the plugs and the HFks seeded dropwise. The media was changed every 24 hours until it was yellow (around day 3 post seeding). At this point, the plug was lifted onto a mesh grid suspended in a 10 cm dish. Raft media was added to the dish until level with the top of the mesh leaving the plug at the liquid-air interface. The media was then changed every 2 days. At day 13 post-lifting, the media was removed and substituted with fresh media supplemented with 20  $\mu$ M BrdU. After allowing BrdU incorporation to proceed for 8 hours, the raft was fixed by flooding with 3.7% formaldehyde (Sigma 37% stock) in DMEM. The fixed rafts were then paraffin embedded and sectioned for analysis (ProPath Ltd, Hereford).

### **2.5.2 Transient transfection**

Plasmid DNA obtained by maxi-prep as previously described (Section 2.4.1) was transfected into cells following the general protocols below. For experiments requiring multiple plates to be transfected with the same DNA, the total required DNA, media and reagent were pooled and the resulting transfection complex was divided evenly between plates.

#### *XtremeGene HP (Roche)*

Plasmid DNA (as specified in results) was mixed with serum-free DMEM (150  $\mu$ L for 6 well plates, 200  $\mu$ L for 6 cm<sup>3</sup> plate, 500  $\mu$ L for 10 cm<sup>3</sup> plate). XtremeGene HP was added dropwise directly into the DMEM-DNA solution to a ratio of 2  $\mu$ L reagent per 1  $\mu$ g DNA (2:1) and the

solution was mixed by gentle agitation before incubation at room temperature for 15 minutes. The transfection complex was then added dropwise to seeded cells.

#### *Lipofectamine 2000 (Life Technologies)*

Plasmid DNA (as specified) was mixed with 200  $\mu\text{L}$  serum-free DMEM. In a separate tube Lipofectamine was mixed with 200  $\mu\text{L}$  serum-free DMEM to a ratio of 3  $\mu\text{L}$  reagent per 1  $\mu\text{g}$  DNA (3:1) and was incubated at room temperature for 5 minutes. The Lipofectamine-containing media was added dropwise to the DNA with agitation and air was bubbled through the mixture before incubating for 30 minutes at room temperature. The transfection complex was then added dropwise to seeded cells.

#### **2.5.3 siRNA knockdown**

Suspension of  $4 \times 10^6$  C33a or C33a pMEP-16 E2 cells were centrifuged to pellet (200  $\times g$ , 3 minutes) and the medium was removed. Cells were washed with 1 mL PBS before resuspending in 100  $\mu\text{L}$  Ingenio electroporation solution (Mirus). 20  $\mu\text{L}$  siRNA (20  $\mu\text{M}$  stock concentration; commercial non-targeting control siRNA (Dharmacon) or siRNA targeted to the 3' UTR of ChlR1) was added mixed by pipetting before transfer to an electroporation cuvette. Cells were electroporated using an Amaxa nucleofector (program S-005). Cells were resuspended in 850  $\mu\text{L}$  DMEM and the final cell suspension was divided between plates as required for further experiments at  $1 \times 10^6$  cells per 6  $\text{cm}^3$  plate,  $2 \times 10^6$  cells per 10  $\text{cm}^3$  plate or  $4 \times 10^6$  cells per 15  $\text{cm}^3$  plate.

#### **2.5.4 Immunofluorescence (IF)**

C33a cells were seeded at a density of  $1 \times 10^6$  into 6  $\text{cm}^3$  plates containing sterile coverslips. To look at protein localisation in cells synchronised to enrich for mitotic populations the

coverslips were pre-treated for 1 hour with poly-lysine (0.0005% stock solution diluted 1:25 with PBS = 0.00002%). Seeded cells were transfected with the 2.5 µg of specified plasmid DNA using Lipofectamine 2000 as described. 48 hours post-transfection the coverslips were transferred to 6 well plates and were gently washed with 1 mL PBS. The cells were then fixed with 4% formaldehyde in PBS for 15 minutes at room temperature followed by washing with PBS. Where stated, cells were pre-extracted prior to fixation by incubating in pre-extraction buffer for 30 seconds. The medium was removed from the remaining cells and dishes were washed with PBS. Cells were scraped in 100 µL urea lysis buffer and incubated on ice for 30 minutes before sonicating at 35% amplitude for 10 seconds. Lysates were centrifuged at 20000 xg for 10 minutes at 4 °C and the supernatant was removed for western blot analysis.

### Staining

Fixed cells were permeabilised by addition of 0.2% Triton-x-100 for 10 minutes before blocking for 1 hour in HINGS block solution as specified. Coverslips were then inverted onto 50 µL specific primary antibodies in block solution (Table 2.1) for 1 hour. The coverslips were then returned to 6 well plates and were then washed 3 times with PBS and incubated with secondary antibodies in block solution for 1 hour in the dark. The coverslips were washed again and then were incubated with Hoechst stain (10 mg/mL stock diluted 1:5000 in PBS) for 10 minutes in the dark before washing. Stained coverslips were mounted onto microscope slides with Fluoroshield mounting solution (Sigma) and allowed to set before storing at -20 °C. Stained cells were then analysed by fluorescence microscopy.

### *In-situ fractionation*

Fixed coverslips were stained following the above protocol with the following exceptions. Coverslips were blocked in 3% BSA in PBS for 1 hour at room temperature before incubation with antibodies. All antibodies were added at the specified dilution in PBS. Incubation with primary antibodies was carried out at 4 °C overnight.

### *pMEP-16 E2 IF*

IF with pMEP-16 E2 cells was carried out as described with some alterations to the above protocol. Where necessary, ChIR1 siRNA knockdown was carried out as described prior to  $1.5 \times 10^6$  cells being seeded into 6 cm<sup>3</sup> dishes containing sterile coverslips. Cells were transfected as required with 2 µg plasmid DNA using Lipofectamine 2000 (2.5.2) 24 hours post-knockdown. E2 expression was induced by addition of ZnCl<sub>2</sub> to a final concentration of 200 µM and coverslips were removed and cells harvested after 6 hours induction. Coverslips were treated as described above with the exception of the block solution being substituted for 3% BSA in PBS. All antibodies were subsequently diluted in PBS and incubation with primary antibodies was carried out for 16 hours at 4 °C (Table 2.1).

### **2.5.5 Cell synchronisation and flow cytometry**

Cells were synchronised to allow investigation of any link between the cell cycle and HPV16 E2 localisation and interactions with ChIR1 and the cellular chromatin.

### *Double thymidine block*

Cells were seeded and transfected as required following the protocols described above. 2 mM thymidine was added to the cells approximately 8 hours post-transfection to block

overnight. After 16 hours, thymidine-containing media was removed and the cells washed twice with warm PBS before fresh media was added. After 8-9 hours 2 mM thymidine was added and the cells blocked for a further 16 hours. Cells were then harvested directly (G1/S) or were released into fresh media for 3 hours (mid-S) or 8-9 hours (mitosis).

To harvest cells, the medium was removed and retained. Cells were washed and trypsinised before resuspending in the retained medium. The cells were counted, divided as needed and pelleted by centrifugation (200 xg, 3 minutes).

#### *Nocodazole block*

As an alternative to double thymidine block, cells were treated with nocodazole to induce arrest in early mitosis. Cells were seeded and transfected as previously before treating with 2 mM thymidine. After 24 hours the cells were released into fresh media for 3 hours before treating with 100 ng/mL nocodazole in DMSO. After 12 hours the cells were harvested as previously described.

#### *Flow cytometry*

Cell synchrony was analysed by flow cytometry.  $1 \times 10^6$  harvested cells were pooled and pelleted by centrifugation at 200 xg, then washed with PBS. The PBS was removed and the cells fixed by adding 1 mL ice-cold 70% ethanol dropwise while gently vortexing to mix. Fixed cells were stored at 4 °C until required.

To allow the cell cycle to be analysed, the ethanol was removed from the fixed cells and the cell pellet was resuspended in a mixture of propidium iodide (50 µg/mL) and RNase A (5

µg/mL) made up to 500 µL in PBS. The cells were incubated in the dark at 4 °C for a minimum of 2 hours before diluting to 1:4 in PBS and analysing by flow cytometry.

Flow cytometry was carried out with a Cyan ADP flow cytometer (Dako Cytometrics) using a 488 nm blue laser controlled using Summit 4.3 software, acquiring data until 50000 counts were reached. Data was analysed using FlowJo, gating for single cell populations on an asynchronous control sample and transferring the gating to all other samples. Cell cycle was determined by generating PI profiles (histograms of count vs. PE-Texas Red or FL3 Lin).

#### **2.5.6 Co-immunoprecipitation (Co-IP)**

$1.5 \times 10^6$  C33a cells were seeded into 10 cm dishes. After 24 hours, cells were transfected with a total of 4 µg plasmid DNA comprising combinations of E2 and FLAG-ChIR1 DNA made up with empty 3xFLAG vector as specified in the results, using XtremeGene HP as described. 24 hours post-transfection the media was removed and cells were washed with PBS before scraping in 1 mL PBS. Cells were pelleted by centrifugation at 500 xg for 3 minutes at 4 °C and the PBS was removed. Cells were resuspended in 300 µL IP lysis buffer (see below) and were incubated on ice for 30 minutes before sonicating at 40% amplitude for 10 seconds. The lysates were centrifuged at 20000 xg for 10 minutes at 4 °C and the supernatant was removed for co-immunoprecipitation.

Sufficient protein-G sepharose to allow for 10 µL of 50% slurry per reaction was equilibrated by washing twice with 500 µL IP binding buffer. The resin was subjected to a pulse-spin for 5 seconds to pellet between washes. The wash was removed and the equilibrated resin was resuspended with IP binding buffer to make 50% (v/v) slurry. Co-Immunoprecipitation reactions were set up by adding 10 µL of the resin slurry to 200 µL IP binding buffer. 2 µL of

the specified antibody was added to each reaction followed by 200  $\mu\text{L}$  of the required lysate. The IP reactions were then incubated at 4  $^{\circ}\text{C}$  with gentle rotation for 16 hours.

After incubation the resin was pelleted by pulse-spin as before and the unbound supernatant was removed. The resin was washed 3 times with 500  $\mu\text{L}$  IP wash buffer by 5 x gentle inversions followed by pulse-spin to pellet the resin and allow the supernatant to be removed. After washing the resin was resuspended in 10  $\mu\text{L}$  6x-SDS loading buffer and boiled at 95  $^{\circ}\text{C}$  for 10 minutes. The entire IP sample was used for SDS-PAGE and western blot analysis, along with 10% input samples prepared from the remaining lysate.

### **2.5.7 Transcription assay**

C33a cells were seeded into 6 well plates at  $2.5 \times 10^5$  cells per well. Transfections were set up in triplicate using XtremeGene as described. A total of 1  $\mu\text{g}$  DNA for each triplicate set (Table 2.8) was transfected in each well as specified.

After 48 hours the media was removed and plates were washed with PBS before the cells were harvested by scraping in 150  $\mu\text{L}$  passive lysis buffer (PLB, supplied in Promega Dual Luciferase Assay Kit). Cells were centrifuged at 20000  $\times g$  for 10 minutes 4  $^{\circ}\text{C}$  and the lysate was removed. Firefly luciferase and *renilla* luciferase activities were determined using the Promega Dual Luciferase assay kit according to the manufacturer's protocol. In each well of a white walled 96 well plate 10  $\mu\text{L}$  lysate was combined with 100  $\mu\text{L}$  luciferase assay substrate and the Firefly luciferase activity was measured over 10 seconds using a 96-well plate luminometer. 100  $\mu\text{L}$  Stop 'n' Glo reagent was then added to each well and the *renilla* luciferase activity was measured. Remaining lysate was analysed by SDS-PAGE and western blotting as discussed.

	Reporter	<i>Renilla</i> luciferase	WT E2	Y131A E2	Salmon sperm DNA
1	100	10	-	-	223
2	100	10	10	-	213
3	100	10	50	-	173
4	100	10	100	-	123
5	100	10	200	-	23
6	100	10	-	10	213
7	100	10	-	50	173
8	100	10	-	100	123
9	100	10	-	200	23

**Table 2.8 Transfected DNA for transcription assay (ng DNA/well)**

6 well plates were seeded with C33a cells at  $2.5 \times 10^5$  cells per well. A total of 333 ng DNA (combination of ptk6-E2 firefly, *renilla* and E2 made up with salmon sperm DNA (Invitrogen)) was transfected into each well using XtremeGene as previously described. 48 hours post-transfection cells were harvested by scraping in 100  $\mu$ L passive lysis buffer (Promega Dual Luciferase assay kit). Lysates were centrifuged at 20000 xg for 10 minutes at 4 °C to pellet any insoluble material. 10  $\mu$ L lysate was used in a luciferase assay to detect Firefly and *renilla* luciferase activity. Remaining lysate was used to determine E2 protein levels by SDS-PAGE and western blotting.

### 2.5.8 Replication assay

C33a cells were seeded into 6 well plates at  $2.5 \times 10^5$  cells per well. Transfections were set up in triplicate using XtremeGene as described. A total of 2190 ng DNA for each triplicate set (Table 2.9) was transfected, with the total transfected DNA normalised with salmon sperm DNA.

After 48 hours the media was removed and plates were washed with PBS before the cells were harvested by scraping in 150  $\mu$ L passive lysis buffer (PLB, supplied in Promega Dual Luciferase Assay Kit). Cells were centrifuged at 20000 xg for 10 minutes 4 °C and the lysate removed. Firefly and *renilla* luciferase activities were determined using the Promega Dual Luciferase assay kit according to the manufacturer's protocol, as described for the transcription assay above. Remaining lysate was analysed by SDS-PAGE and western blotting.

	Reporter	<i>Renilla</i> luciferase	HA-E1	E2	Salmon sperm DNA
1	25	5	-	-	700
2	25	5	600	-	100
3	25	5	-	10	690
4	25	5	-	10	690
5	25	5	600	10	90
6	25	5	600	25	75
7	25	5	600	50	50
8	25	5	600	100	-

**Table 2.9 Transfected DNA for replication assay (ng DNA/well)**

### **2.5.9 Protein lysis buffer test**

C33a cells were seeded into 10 cm<sup>3</sup> dishes at a density of 1x10<sup>6</sup> and were transfected with 1 µg WT or Y131A E2 DNA with XtremeGene following given protocol. Cells were harvested by scraping in 300 µL PBS and the cells pelleted by centrifugation (500 g, 3 minutes, 4 °C). The PBS was removed and the cells lysed by resuspension in either PLB or ura lysis buffer 2 (300 µL). Insoluble material was removed by centrifugation (20000xg, 10 minutes, 4 °C) and the total protein in the cell lysates was quantified by Bradford assay. E2 protein levels were analysed by SDS PAGE and western blotting.

### **2.5.10 Cycloheximide stability assay**

C33a cells were seeded into 15 cm plates at a density of 6x10<sup>6</sup> and were transfected with plasmid DNA expressing E2 as specified using XtremeGene HP (following above protocol). After 24 hours transfected cells were trypsinised and reseeded into 6 cm plates at a density of 1x10<sup>6</sup> cells per well. After a further 24 hours cells were treated with cycloheximide to a final concentration of 10 µg/mL. At the specified time points post-cycloheximide treatment, media was removed from plates and cells were washed with PBS before harvesting by scraping in 200 µL urea lysis buffer. Scraped cells were incubated on ice for 30 minutes before sonicating at 35% amplitude for 10 seconds. Lysates were centrifuged at 20000 xg for 10 minutes at 4 °C and the supernatant was removed. The total protein concentration was determined by Bradford assay (2.3.4) and 20 µg total protein for each sample was analysed by SDS-PAGE and western blotting. E2 levels were determined by densitometry.

### 2.5.11 Sub-cellular fractionation

C33a cells were seeded into 10 cm dishes at a density of  $3 \times 10^6$  cells and were transfected with 1  $\mu\text{g}$  E2-expressing plasmid DNA using XtremeGene HP as described. After 48 hours, cells were trypsinised and  $5 \times 10^6$  cells were pelleted by centrifugation at 500 xg for 3 minutes. The media was removed and the cells were washed with PBS and again pelleted by centrifugation. The cells were then subjected to fractionation by sequential lysis in specific buffers using a sub-cellular fractionation kit (ThermoScientific) according to the manufacturer's protocol, following the instructions as written for a 20  $\mu\text{L}$  cell pellet. 20  $\mu\text{L}$  samples of each fraction were analysed by SDS-PAGE and western blotting with antibodies specific to markers of the various sub-cellular fractions (Table 2.10) along with E2 and ChIR1.

Fraction	Marker
Cytoplasmic (CE)	GRB2
Nuclear soluble (NE)	Orc2
Chromatin bound (CB)	Histone 3 (H3)
Cytoskeletal/pellet (PE)	Vimentin

**Table 2.10 List of the proteins used as markers for sub-cellular fractionation**

### 2.5.12 *In situ* fractionation

$6 \times 10^6$  C33a cells were seeded into 15 cm plates containing 6-10 coverslips and were transfected with 6  $\mu\text{g}$  DNA (HPV16 E2 WT or Y131A) using 3:1 Lipofectamine 2000 as described above. Alternatively,  $4 \times 10^6$  C33a cells were treated with control or ChIR1 siRNA as described above before seeding into 15 cm plates containing 6-10 coverslips. After 24 hours siRNA treated cells were transfected with 6  $\mu\text{g}$  HPV16 E2 WT DNA using 3:1 Lipofectamine 2000 as described above.

48 hours post-transfection coverslips were removed to 6 well plates and were gently washed with 1 mL ice cold PBS to remove media. Coverslips were sequentially extracted following a modification of a published protocol [136]. All coverslips were fixed with 1 mL 4% formaldehyde at 4 °C for 30 minutes after extraction, then were washed twice with 1 mL PBS and were stored in PBS at 4 °C.

The first coverslip representing the whole cells was removed to a separate 6 well plate and was fixed as stated. A second coverslip was extracted with 200 µL TES buffer for 1 minute and room temperature and the whole cell extract was removed and stored on ice.

The remaining coverslips were treated with 200 µL CSK buffer containing 0.1% Triton-x-100 on ice for 1 minute. The cytoplasmic and nuclear soluble protein extract was removed and the coverslips were gently washed twice with ice-cold PBS. One coverslip representing the nuclear fraction was removed and fixed as above. The remaining coverslips were treated with 200 µL CSK buffer containing 0.5% Triton-x-100 for 20 minutes at 4 °C. The nuclear protein extract was then removed and the coverslips were again gently washed. One coverslip representing the chromatin fraction was removed and fixed. The remaining two coverslips were treated with 100 µL CSK buffer containing 100 µL DNase I (100 µg/mL in water) for 30 minutes at 37 °C and the chromatin fraction was removed. The coverslips were again washed gently with PBS and one was fixed to represent the nuclear matrix fraction. The final coverslip was treated with 200 µL TES buffer for 1 minute at room temperature and the nuclear matrix extract fraction was removed. The fixed coverslips were stained for immunofluorescence following the standard protocol outlined above (2.5.5).

### **2.5.13 Organotypic raft culture analysis**

For histology, raft sections were hemotoxylin and eosin stained by Propath, Ltd. The stained raft sections were imaged by light microscopy.

#### *Immunofluorescence*

##### Slide preparation – ALTER

Paraffin embedded raft sections were prepared for immunofluorescence by agitated low temperature epitope retrieval (ALTER). Slides were immersed in HistoClear (National Diagnostics) for 10 minutes at RT. Excess HistoClear was drained and the slides transferred to 100% industrial methylated spirits (IMS) for 5 minutes. The slides were then washed 3 times in water before placing in a solution of 0.3% hydrogen peroxide in water for 15 minutes. Following this the slides were washed a further 3 times in water before transferring to 0.1 mM EDTA/ 0.01 % Tween 20 buffer and preheated to 60 °C, stirring at 600 rpm. The slides were incubated at 60 °C overnight before cooling to room temperature. The slides were stored in PBS until required for immunofluorescence.

##### Immunofluorescence

Prepared slides were blocked by adding 300 µL HINGS block solution to the surface of the slide and incubating at RT for 1 hour. The slides were drained to remove block and 250 µL primary antibody diluted in HINGS block was added to each slide before incubating at 4 °C overnight. The primary antibody was drained and the slides washed (3x 5 min) in PBS. 250 µL secondary antibody diluted in HINGS block was added to the slides and they were then incubated at 37 °C in the dark for 1 hour. The antibody was removed and the slides washed

(3x 5 min) in PBS. 10  $\mu$ L Hoechst (5 mg/mL stock) was added to the final PBS wash to stain DNA. The slides were removed from the wash and excess PBS was drained. Mounting solution was added to the slides (Fluoroshield, Sigma) and a coverslip placed on top. The mount was allowed to set before slides were stored at -20 °C in the dark. Slides were imaged as described previously.

## **Chapter 3. Identifying small molecule fragments as novel binding partners of HPV11 E2**

### **3.1 Introduction**

#### *Strategies for treating HPV infection*

HPV infection is associated with disease ranging from benign cutaneous and genital warts to multiple anogenital and oropharyngeal carcinomas [28] (detailed in introduction). Treatments for HPV-associated disease currently target the outward symptoms rather than the underlying infection. Common treatments for genital warts include treatments with chemical agents such as trichloroacetic acid (TCA) or cryotherapeutic or surgical removal of the warts. These can be effective in the short term but recurrence is common, requiring multiple treatments [9, 11]. Some success has been observed using immune modulators to stimulate the host immune response to respond to HPV infection. Imiquimod acts to stimulate interferon and cytokine release from keratinocytes by acting as an agonist of TLR7 (toll-like receptor 7), while polyphenon E acts to alter signalling pathways, inhibiting telomerase activity and inducing cell-cycle arrest. These treatments have both shown some efficacy against genital warts but recurrence is still an ongoing problem [11, 12]. Alternatively, an antiviral against cytomegalovirus (CMV), cidofovir, can be used to treat low risk HPV infection, particularly in cases of recurrent respiratory papillomatosis (RRP). Cidofovir is a monophosphate analogue of deoxycytidine (dCTP), acting to inhibit dCTP incorporation into viral DNA and prevent elongation of DNA [11]. Although this cannot cure the infection, it does control the symptoms during continuous treatment.

Although the development of prophylactic vaccines against the most common HPV types is a significant step in preventing HPV-associated disease, alternative therapeutic strategies to target the underlying viral infection are still necessary. Small molecule inhibitors which directly target various functions of the HPV life cycle are therefore of great interest.

The HPV E1 protein is the only viral protein possessing enzymatic activity. E1 acts as both an ATPase and DNA helicase and is essential for viral genome replication [28, 30]. As such the development of E1 inhibitors has been investigated. A high throughput screen (HTS) of a compound library with recombinant HPV6 E1 identified an initial lead compound which inhibited the ATPase activity of E1. Structure activity relationship (SAR) development of this hit led to the development of a series of compounds, resulting in an E1 ATPase inhibitor with low nanomolar activity. Analysis of these compounds showed they act through an allosteric binding site on E1, binding reversibly to inhibit ATP binding and hydrolysis [137, 138]. The ATPase activity of E1 is required for the assembly of the E1 double hexamer, the active form necessary for the initiation of viral genome replication.

Alternatively, the E1-E2 interaction has also been investigated as a potential target to inhibit viral replication. Screening a compound library for molecules which inhibited the cooperative binding of HPV11 E1 and E2 to DNA containing the HPV11 Ori identified an initial hit compound which interacts reversibly with the E2 transactivation domain (TAD) [137]. Structural analysis of the compound interaction with E2 followed by extensive SAR development resulted in two potent lead compounds with differing structural properties which inhibit the E1-E2 interaction in both *in vitro* and *in vivo* assays [139-141].

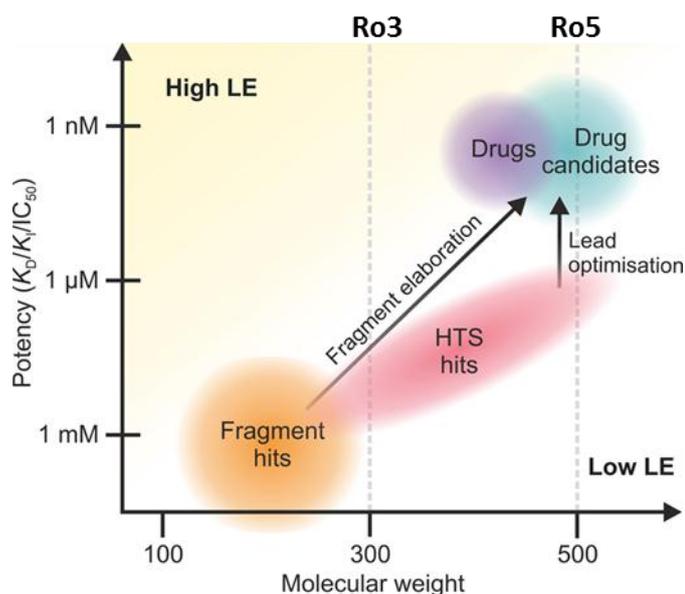
This work shows that the E2 TAD is a viable target for drug development. As well as the interaction with E1, which is necessary for the initiation of viral genome replication, E2 TAD interacts with multiple cellular proteins. Many of these interactions are essential for the viral life cycle, as E2 uses these to regulate viral transcription and mediate viral persistence (discussed in introduction). In particular, the interaction between E2 and the cellular DNA helicase ChIR1 is proposed to be required for the partitioning of viral genomes between daughter cells. Disrupting this protein-protein interaction could have a profound effect on the ability of the virus to establish and maintain a persistent infection, thus limiting the potential progression towards malignancy.

#### *Fragment based drug design (FBDD)*

Since ~1990, rational drug design methods have been used to generate lead compounds for drug development. Advances in combinatorial chemistry allowing for the rapid synthesis of vast libraries of compounds, along with high throughput screening (HTS) techniques allow libraries of up to  $10^6$  compounds to be screened for activity against biological targets. Analysis of generated libraries and the resulting successful drug molecules identified gave rise to a series of rules governing the physico-chemical properties of drug-like molecules. These rules are adhered to in >90% of orally active drugs, with drugs falling outside of these rules generally belonging to categories containing structural motifs which allow drugs to mimic substrates for natural transporters, thereby avoiding issues with absorption, including antibiotics, antifungals and vitamins. Successful drug molecules were found to have a molecular weight of  $\leq 500$  Da, a calculated lipophilicity (CLogP determined on ratio of octanol solubility versus aqueous solubility) of  $\leq 5$ , a total number of hydrogen bond donors

of  $\leq 5$  and a total number of hydrogen bond acceptors of  $\leq 10$ . Together these are referred to as the rule of 5 [142]. One of the serious drawbacks to the HTS technique is tied to the fundamental properties of the screen. In order for library compounds to be active in the bioassays used in HTS screening, they have to have properties approaching the limit of the rule of 5, similar to drug-like molecules. However this dramatically limits the potential for optimisation from initial hits toward lead compounds. Optimisation of hits to improve the activity and pharmacokinetics of the molecules generally requires increasing both the molecular weight and lipophilicity of the compound. If the initial hit is already approaching the limit set out by the rule of 5, increasing these parameters further can have an adverse effect on the solubility and bioavailability of the resulting drug-like molecule [143].

To circumvent these restrictions that prevent ready optimisation of initial hits to drug-like molecules, screening a library comprising smaller compounds with a molecular weight of between 100-250 Da was proposed. Screening smaller fragments of molecules in place of drug-like compounds would allow for easy optimisation without compromising the physico-chemical properties of the resulting drug-like molecules (Figure 3.1, [144]). A modification of the Lipinski rules was developed to facilitate the designing of fragment libraries. Fragments with a molecular weight of  $\leq 300$  Da, a ClogP  $\leq 3$  and a total of  $\leq 3$  of each hydrogen bond acceptors and donors were found to be the most successful in further development towards drug-like molecules. These modified rules were termed the rule of 3 [145, 146].



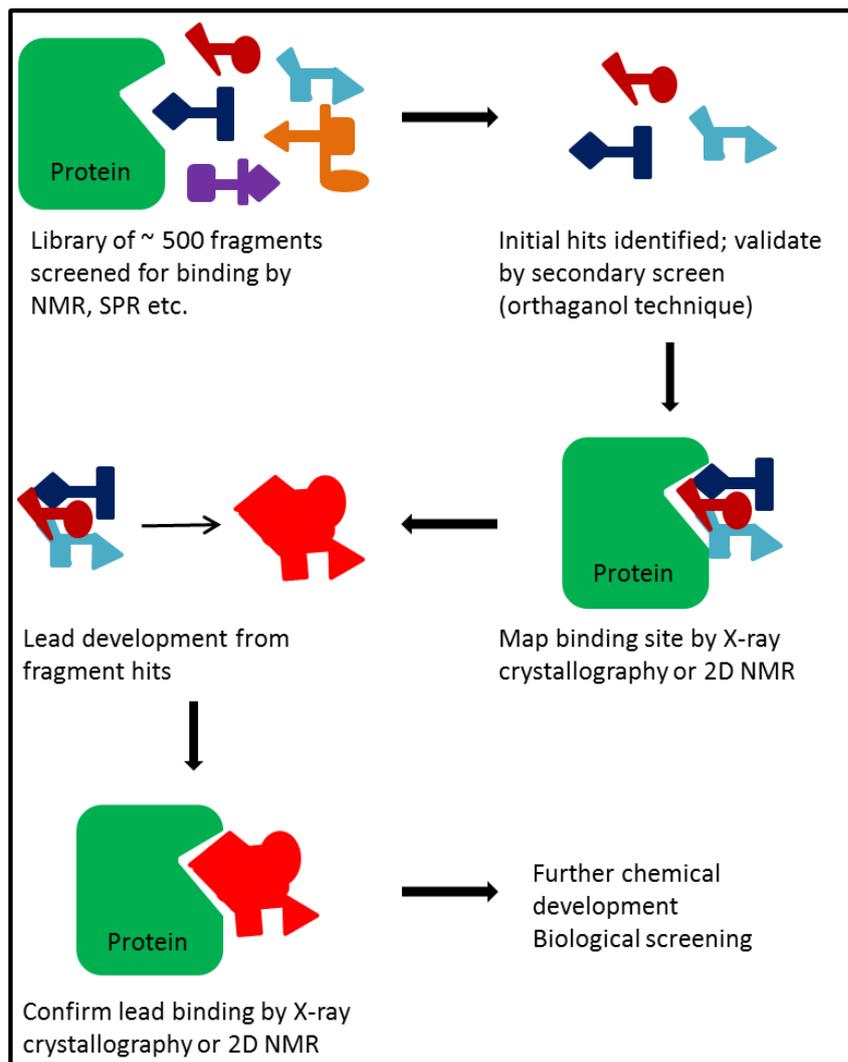
**Figure 3.1** The association between molecular weight and potency of hits from HTS and FBDD [144]. HTS hits occupy a region between 300-500 Da with a typical potency in the low micromolar range. There is little room for optimisation toward drug candidates without exceeding the Ro5 parameters. Fragment hits have a lower molecular weight and potency, allowing for increased flexibility in elaboration toward drugs and reducing the risk of exceeding the Ro5 limits.

Along with the additional flexibility in optimising initial hits to lead compounds, the fragment based drug discovery (FBDD) approach has other advantages over traditional HTS (Figure 3.2). Libraries of up to  $10^6$  small molecules within the limits imposed by the rule of 5 can be screened using HTS. However it has been estimated that there are in the region of  $10^{60}$  possible compounds with a molecular weight of 300-500 Da [144, 147]. In contrast there are only an estimated  $10^7$  possible compounds made of up to 11 atoms which fall within the rule of 3 parameters [148]. This means that screening a much smaller library of fragments, generally around  $10^3$ , allows a far greater area of chemical space and diversity to be

investigated than the traditional HTS methods, resulting in novel chemical moieties with affinity for the biological target being identified.

The small size of the fragment molecules used in these screens pose their own set of restrictions however. Fragments are highly unlikely to bind to protein targets with the same affinity as drug-like molecules and so exhibit low activity. This means fragment hits are unlikely to be detected in traditional biological assays and that alternatives must be employed [149]. In place of biological activity assays, biophysical assays can instead be used. Structural biology techniques including X-ray crystallography of protein targets with fragment libraries [149], protein NMR [150], mass spectroscopy, surface plasmon resonance (SPR) and thermal shift assays [144] have all been used to screen fragment libraries to identify binding fragments as hits. Although fragments tend to bind with lower affinity than drug-like molecules, these highly sensitive techniques are able to identify low-affinity interactions. Repeating library screens with multiple assays allows fragment hits to be identified with increasing confidence. Additionally, structural assays like X-ray crystallography and 2D/3D protein NMR can provide information about the binding site of fragment hits, allowing for rational structure-based development of the hits into drug-like lead compounds. Given the small size of the fragments, it is possible for multiple fragments to bind to adjacent sites on a protein target, mapping out a larger binding site. As such, fragment hits can be elaborated into larger drug-like molecules which fit multiple pockets within a binding site increasing the affinity and specificity of the drug. There are three primary methods that are used for fragment development which take advantage of this structural information – fragment merging where multiple binding fragments in the same region are used to model a new compound containing the features of both; fragment linking,

where multiple fragment binding to adjacent sites are linked chemically to form a new molecule containing each fragment moiety fastened to a scaffold; and fragment growing where a set of fragments which bind to the same site are used as the basis for synthesis of a set of chemical analogues which can be used in structure-activity relationship (SAR) analyses to sequentially improve the activity of the developing drug-like molecule. FBDD has been proven a successful approach to novel drug development. In 2011 the first drug developed using these techniques, Zelboraf, was FDA-approved for use to treat late-stage melanoma [144, 151].



**Figure 3.2. Schematic to illustrate the process of FBDD.**

### 3.2 Aims

The aims of this project were centred on using FBDD techniques to identify novel small molecules which bind to HPV11 E2 TAD. The main aims were as follows:

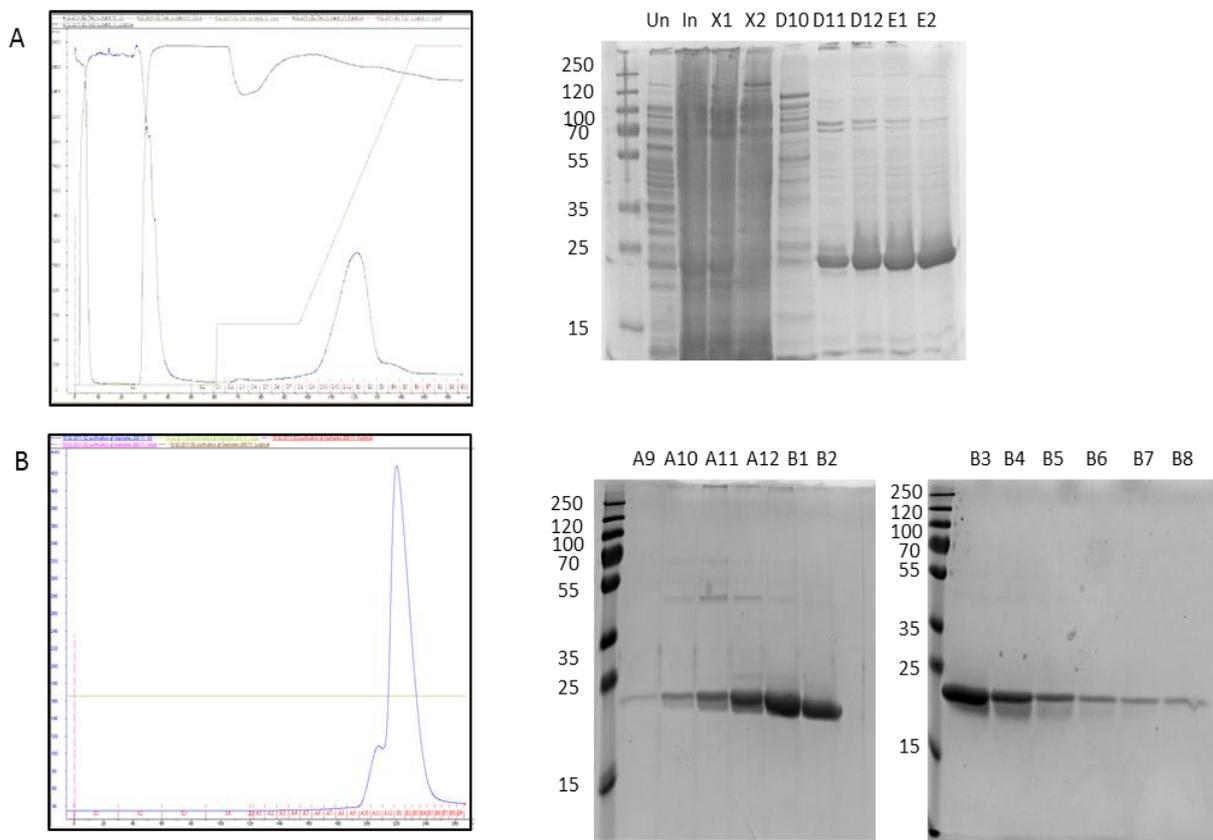
1. To develop suitable conditions to use *in vitro* purified HPV11 E2 TAD as a target protein for FBDD in biophysical screens including fluorescence-based thermal shift assays and NMR
2. To use the optimised thermal shift and NMR based assays to screen a commercial fragment library for binding to HPV11 E2 TAD as a means to identify novel binding fragments as a potential start for inhibitor development, and
3. To investigate the possibility of using HPV11 E2 TAD in 2D/3D NMR assays to obtain structural information about the site of fragment binding

### 3.3 Results

#### 3.3.1 Optimisation of the fluorescence-based thermal shift assay termed differential scanning fluorimetry (DSF) with purified HPV11 E2 TAD

##### *HPV11 E2 transactivation domain purification*

HPV11 E2 TAD was expressed as an N-terminally 6x-His tagged fusion protein from *E. coli* strain BL21 transformed with a pET15b vector using IPTG induction. Bacterially expressed protein was then purified using His-affinity chromatography, isolating His-tagged proteins onto a column charged with Ni<sup>2+</sup> ions. SDS-PAGE and Coomassie staining was used to identify protein-containing fractions (Figure 3.3A); these were then pooled and further purified by size exclusion chromatography, again purity was confirmed by SDS-PAGE and Coomassie staining (Figure 3.3B). HPV11 E2 TAD was isolated at >95% purity, assessed by Coomassie staining, at typical yields of approximately 10 mg per litre of culture.



**Figure 3.3 Purification of HPV11 E2 TAD.**

Expression of HPV11 E2 TAD in bacterial culture was induced by addition of IPTG. Bacterial cell pellets were lysed and protein purified by chromatography. A. UV trace (AKTA, blue line) showing purification of HPV11 E2 TAD by nickel affinity chromatography. Fractions (labelled along the top of the gels) were analysed by SDS-PAGE followed by Coomassie staining. Fractions determined to contain protein of the correct size were pooled and further purified by size exclusion (SE) chromatography. B. UV trace (blue line) showing SE purification of HPV11 E2 TAD and the corresponding Coomassie stained SDS gels.

### *Thermal shift assay – buffer screen*

Initially, a DSF thermal shift assay was employed to screen the fragment library for interaction with purified HPV11 E2 TAD. DSF is based around the principle of protein denaturation once a critical melting temperature ( $T_m$ ) is reached. Purified proteins are incubated with Sypro Orange, a hydrophobic dye, along with a buffer alone or with a small molecule of interest. The protein mixture is then gradually heated to 95 °C using a real time PCR machine and Sypro Orange dye fluorescence is monitored with each incremental temperature increase. Once the  $T_m$  of the protein is reached, it unfolds and exposes hydrophobic regions which would normally be buried within the protein structure. The dye can then bind to the newly exposed hydrophobic regions, removing it from the solution and preventing the quenching of fluorescence by the water in the solution. This leads to a spike in fluorescence intensity which correlates with the protein denaturation. For a normal two-state equilibrium system, protein unfolding results in a sigmoidal fluorescence curve which can be fitted using the Boltzmann equation to calculate the  $T_m$  of the protein. Normalising the resulting curve for fluorescence intensity results in a bell-shaped curve, highlighting the shift in protein  $T_m$  observed with changing conditions (Figure 3.4).

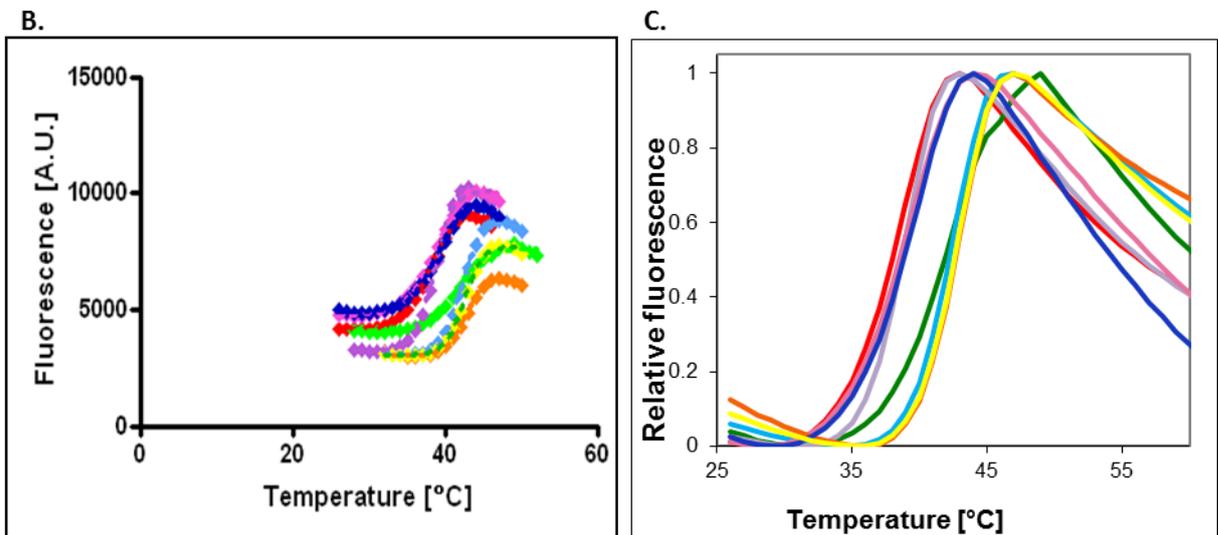
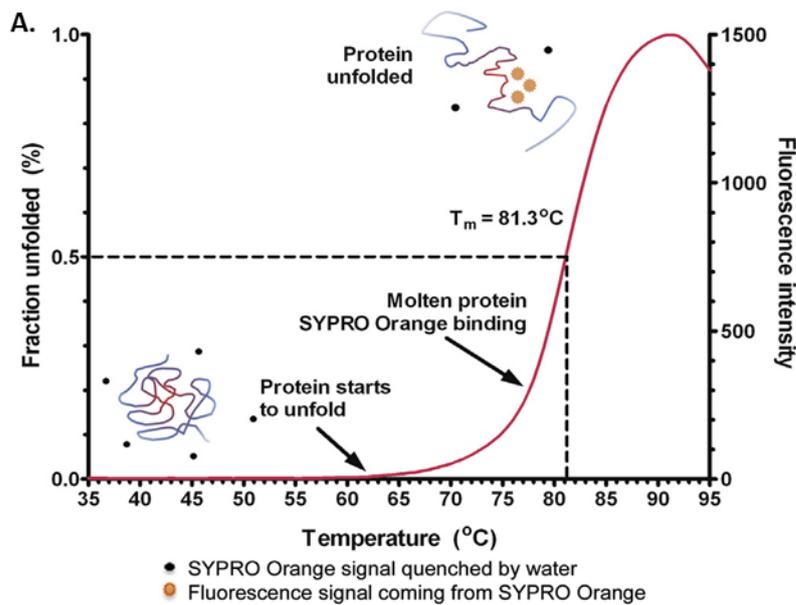
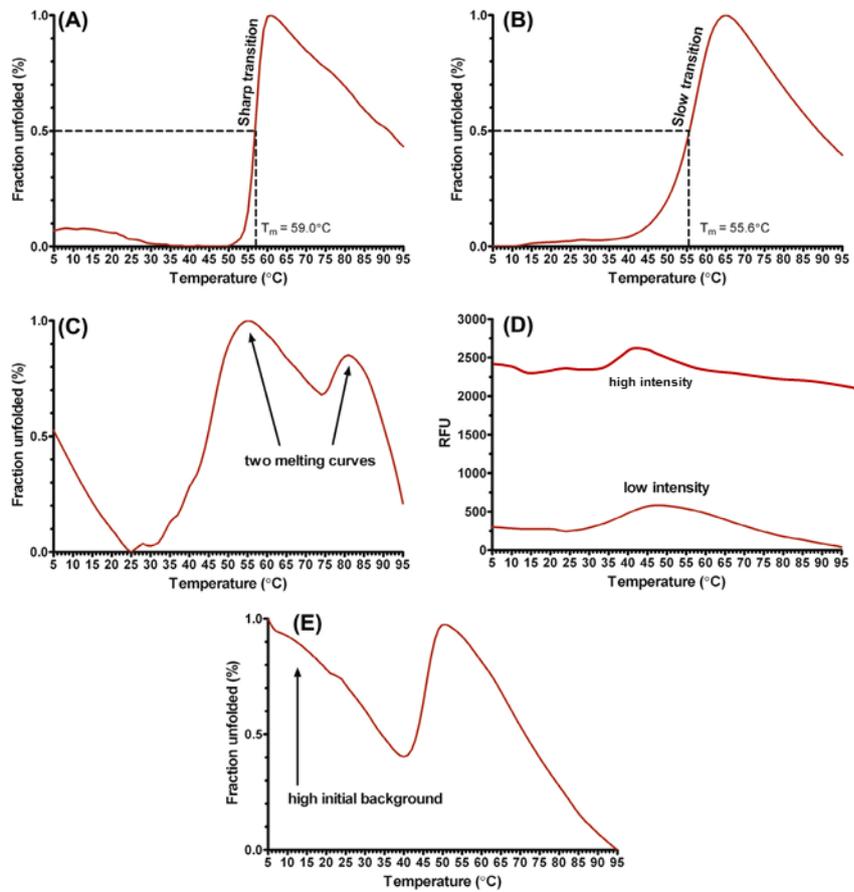


Figure 3.4 Representative curves to illustrate DSF assay

A. Representative fluorescence intensity curve to illustrate the DSF assay, taken from [152]. Protein is incubated with Sypro Orange dye. At low temperatures the protein structure is stable and Sypro Orange fluorescence is quenched by the bulk water. As the temperature increases the protein begins to unfold, exposing previously buried hydrophobic regions. Sypro Orange can bind to these regions and the corresponding increase in fluorescence intensity is recorded. The protein  $T_m$  is determined as the mid-

point of the transition between minimum and maximum fluorescence intensity. B. Examples of sigmoidal fluorescence intensity curves obtained with HPV11 E2 TAD in varying buffers. Due to variation in the fluorescence intensity between samples, comparing the shift in transition temperature is difficult. C. Normalising the maximum fluorescence intensity of each sample to 1 allows the change in  $T_m$  between samples to be assessed more easily.

In some situations, the expected sigmoidal fluorescence curve is not observed. This can be due to the effects of added small molecules or co-factors influencing the protein stability, or due to variations in the protein between wells. Protein aggregation or precipitation can result in small transitions in fluorescence intensity, and direct interactions between the protein or added small molecule and the fluorescent dye can induce high initial fluorescence intensity which can mask transitions. Protein dimerization can result in biphasic melting curves, again obscuring the transition and determination of  $T_m$ . High levels of background fluorescence from the protein or added small molecules can also influence the fluorescence curve obtained. Examples of the anomalous fluorescence profiles which can be obtained are shown in Figure 3.5.



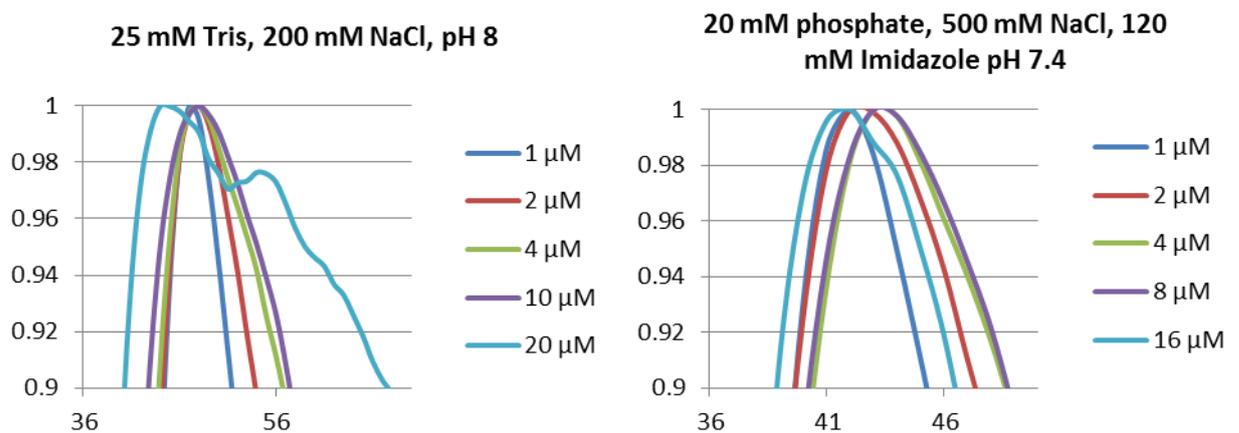
**Figure 3.5 Examples of different fluorescence profiles from DSF (modified from [152])**

Representative curves to illustrate different fluorescence profiles which can be observed from DSF assays. A. A sharp transition curve resulting from rapid melting of the protein over a small temperature range, B. a slow transition curve indicating good thermal stability of the protein resulting in slow denaturation over a wider temperature range, C. a biphasic melting profile with two distinct transitions. This suggests the presence of a multi-domain or dimeric protein. D. Limited transition in fluorescence intensity. This can be observed with a high initial fluorescence suggesting protein aggregation or unfolding early, or a direct interaction between the protein and dye. Small transitions can also be seen with low fluorescence intensity, indicating protein aggregation or precipitation preventing

**Sypro Orange binding, or the lack of a hydrophobic core within the protein structure. E. High background fluorescence due to protein aggregation or partial unfolding, or interference from hydrophobic buffer components**

Proteins in solution can be stabilised by optimising buffer conditions or by adding small molecule ligands or peptides which directly bind to the protein, leading to a detectable positive shift in the protein  $T_m$ . As this assay can be carried out using widely available real time PCR technology on a high-throughput scale, this makes it useful for rapidly assessing large sets of buffers or molecules for their effect on protein stability.

Using this assay, the buffers routinely used for HPV11 E2 TAD purification (Tris and phosphate) were tested to determine the stability of HPV11 E2 TAD in these buffers with a view to identifying the best conditions for fragment screening assays. The fluorescence data collected was used to plot Boltzmann curves and calculate the protein  $T_m$  using a series of Microsoft Excel workbooks and a GraphPad Prism template as published by Niesen [135]. Shown are representative curves for each buffer condition used (Figure 3.6).



**Figure 3.6 HPV11 E2 TAD stability is dependent on buffer conditions.**

**Normalised curves showing fluorescence intensity against temperature for increasing concentrations of HPV11 E2 TAD in A. Tris buffer and B. phosphate buffer, both used for purification. T<sub>m</sub> is determined as the mid-point of the transition in fluorescence intensity. Peak fluorescence intensity is normalised to 1.**

In both the Tris- and phosphate-based buffers routinely used for purification, the stability of HPV11 E2 TAD decreased with protein concentration. This suggests that HPV11 E2 TAD could aggregate at high concentrations, reducing the observed stability. The protein stability in general was shown to be dependent on the buffer conditions used. The reduction in protein stability with increasing concentration was particularly observed in Tris-buffer, with the observed T<sub>m</sub> decreasing by 3.85 °C as the protein concentration was increased from 1 μM to 20 μM. In the phosphate buffer however, the increase in protein concentration had little effect on the protein stability, with the observed T<sub>m</sub> only changing by 0.21 °C going from 1 μM to 16 μM protein concentration. This is potentially due to the Tris buffer being less able to accommodate the increasing protein-protein interactions occurring with an increase in

protein concentration while the phosphate buffer is able to consistently stabilise these inter-protein effects. Alternatively, as the Tris-buffer had a greater effect on protein stability overall, resulting in a  $T_m$  for HPV11E2 TAD of 40.81 °C at 1  $\mu$ M protein, compared to just 36.66 °C in the phosphate buffer at the same protein concentration, it is possible that in phosphate buffer the protein is less stable and therefore more prone to aggregation even at low concentrations. In the Tris-buffer however, HPV11 E2 TAD is more stabilised at low concentrations and could therefore be protected from aggregation until the protein concentration reaches a critical level and outweighs the stabilising effects from the buffer.

25 mM Tris pH 8, 200 mM NaCl		20 mM phosphate pH 7.4, 500 mM NaCl, 170 mM imidazole	
HPV11 E2 TAD concentration / $\mu$ M	$T_m$ / °C	HPV11 E2 TAD concentration / $\mu$ M	$T_m$ / °C
1	40.81	1	36.66
2	40.46	2	36.39
4	39.75	4	36.72
10	38.44	8	36.70
20	36.96	16	36.45

**Table 3.1 Calculated melting temperatures ( $T_m$ ) of HPV11 E2 TAD at increasing concentration in two buffers.**

A further screen of 10 different buffers of covering a pH range of 5-9.5 was then carried out to identify suitable conditions for fragment screening by determination of the  $T_m$  of HPV11 E2 TAD at 6.5  $\mu$ M. Each buffer was tested at two concentrations (50 mM or 100 mM), alone

or supplemented with 100 mM NaCl or 10 mM dithiothreitol (DTT), or both. NaCl and DTT are routinely used in protein purification and downstream experiments to stabilise proteins in solution [153-155], as such their addition to the buffer was used to see if cumulative stabilising effects could be detected with the buffers screened.

The resulting HPV11 E2 TAD  $T_m$  values obtained (Table 3.2) showed a dramatic effect of pH on protein stability. Optimum protein stability was observed in the two buffers tested at pH 7.5, MOPS and HEPES, with good stability being observed across the pH range 7-8.5 (Figure 3.7). Increasing the buffer concentration from 50 mM to 100 mM had little effect on the protein stability, but the addition of NaCl increased the  $T_m$  in multiple buffers. DTT had a smaller effect and only resulted in an increase in protein stability with buffers at the pH extremes tested. The combination of NaCl and DTT again increased the protein stability but there was only a small additive effect observed over that of NaCl alone.

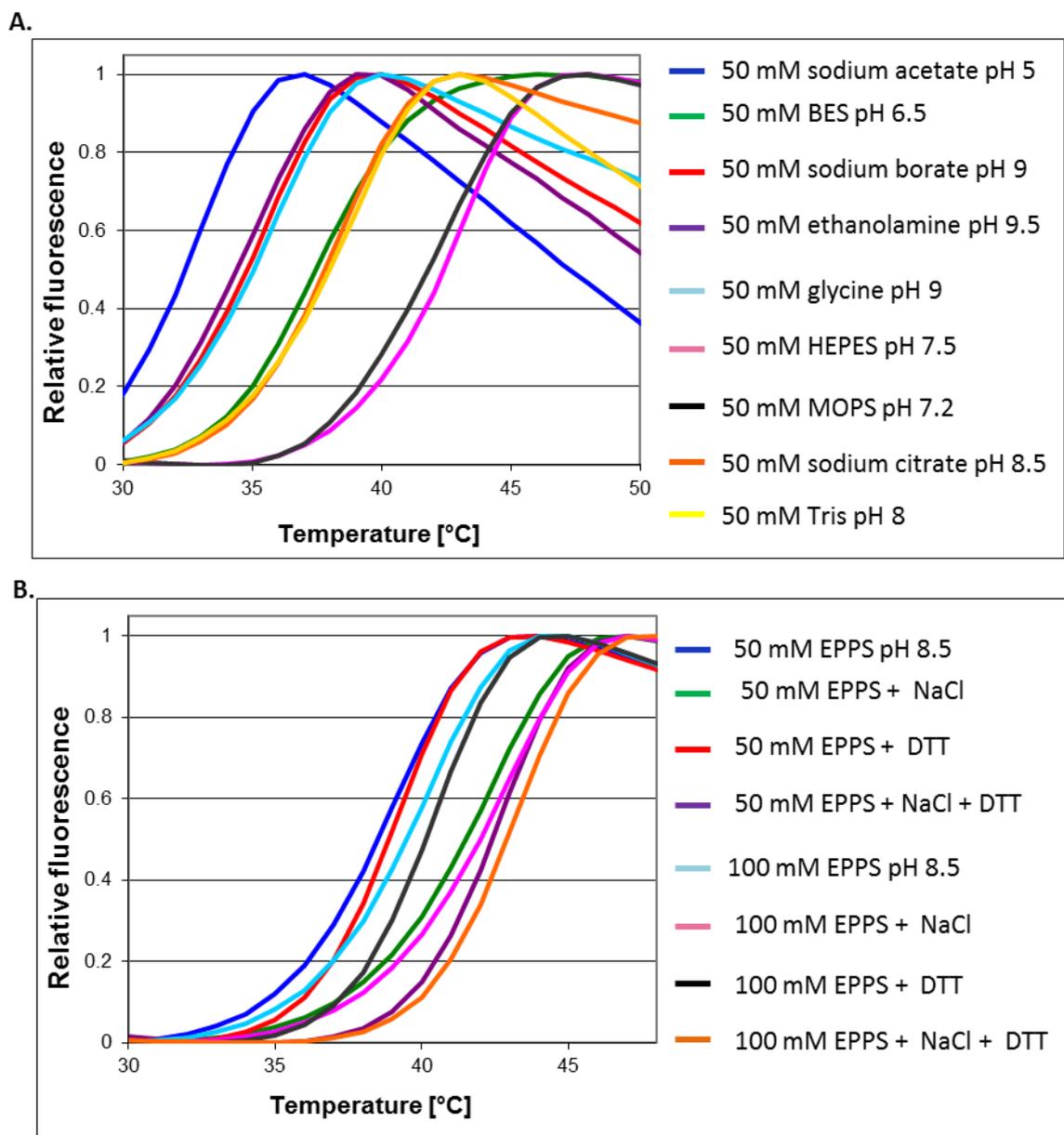
Generally the buffers inducing the highest protein stability (MOPS, HEPES) gave consistent  $T_m$  values across the tested conditions, displaying little additive effect from altering the buffer composition. However, the buffers in which HPV11 E2 TAD was initially the least stable were more influenced by addition of NaCl and DTT but the resulting  $T_m$  was still lower, suggesting that the effect of stabilising additives is less significant than that of the initial buffer pH. The greatest increase in protein  $T_m$  observed following the addition of NaCl alone or in combination with DTT was seen in buffers at the mid-range pH, such as tricine, EPPS and glycine. These buffers were within the optimum pH range to ensure reasonable protein solubility while still allowing small cumulative stabilising effects from additives. This suggests that either these buffers are more susceptible to variation due to additives, or that

a baseline level of protein stability is required initially before the additional increases in stability can be observed following changes to the buffer composition.

	50 mM	50 mM + NaCl	50 mM + DTT	50 mM + NaCl + DTT	100 mM	100 mM + NaCl	100 mM + DTT	100 mM + NaCl + DTT
Sodium Acetate pH 5	32.22	33.72	32.08	33.59	32.69	33.9	32.25	34.05
BES pH 6.5	37.45	37.82	37.32	37.59	38.31	38.15	38.27	38.05
Imidazole pH 7	-	41.06	40.08	41.07	-	41.27	38.91	41.81
HEPES pH 7.5	42.28	41.23	41.55	41.41	41.00	41.79	40.82	41.98
MOPS pH 7.5	41.70	40.54	41.08	40.87	40.16	41.39	40.30	41.49
Tricine pH 8	39.25	41.26	39.36	41.87	38.93	41.71	39.36	43.07
EPDS pH 8.5	38.41	41.38	38.8	42.33	39.41	41.84	40.03	42.82
Glycine pH 9	34.91	38.19	36.57	41.00	35.41	37.78	36.55	40.23
Sodium borate pH 9	34.59	37.13	36.40	39.44	35.62	37.54	37.46	39.92
Sodium ethanolamine pH 9.5	34.25	35.46	35.72	37.97	34.23	35.52	39.97	38.40

**Table 3.2  $T_m$  values calculated for HPV11 E2 TAD in varying buffer conditions.**

Purified HPV11 E2 TAD (6.5  $\mu$ M) was combined with the buffers listed and Sypro Orange dye. Samples were gradually heated over time and the fluorescence intensity was recorded with each incremental increase in temperature.  $T_m$  values were calculated using open access analysis tools [134, 135].



**Figure 3.7 Representative normalised DSF curves for HPV11 E2 TAD in buffer screen.**

Normalised curves showing Sypro Orange dye fluorescence intensity against temperature for HPV11 E2 TAD in a range of buffers.  $T_m$  is determined as the mid-point of the transition in fluorescence intensity. **A.** HPV11 E2 TAD thermal stability in 50 mM buffers covering a range of pH from 5-9.5. **B.** HPV11 E2 TAD thermal stability in EPPS based-buffers.

On the basis of these results, 50 mM EPPS pH 8.5 was selected as the buffer to be used in further thermal shift assays to screen for small molecule fragments binding to HPV11 E2 TAD (Figure 3.7). The fragments included in the library used were all designed in accordance with Lipinski's rule of 3 and as such are considerably smaller than traditional drug-like molecules. As such it is unlikely that their binding to the protein target would have a large effect on protein stability and so the buffer conditions must be selected to maximise the chances of observing these weak-binding events having a limited impact on protein stability. If the protein is already very stable (for example in HEPES or MOPS buffers) it is unlikely that the limited impact of fragment binding on protein stability would be observed under these conditions. Conversely, if the protein is not sufficiently stabilised by the buffer then there is an increased risk of protein aggregation or unfolding throughout the assay, potentially obscuring the small stabilising effects of fragment binding. To balance these opposing issues, the mid-range 50 mM EPPS buffer was selected for use in further assays.

### **3.3.2 Screening a library of small molecule fragments using DSF to identify novel fragments which interact with of HPV11 E2 TAD**

#### *Thermal shift assay*

As an alternative to screening vast numbers of drug-like molecules for binding interactions with HPV11 E2 TAD, a library of 508 small molecule fragments produced by Maybridge Ltd (Thermo Fisher Scientific, UK) was instead screened for their potential to interact with E2 (436 of which are included in an NMR screen described later). Although a relatively small number of compounds, these highly diverse fragments allow a large area of chemical space to be screened for interacting motifs [146, 148, 156]. Screening fragments in place of drug-

like small molecules allows any fragment hits to be readily combined or modified and developed into drug-like lead structures while avoiding many potential issues with pharmacodynamics and pharmacokinetic properties, but it also means that any interactions are likely to be of lower affinity and therefore more difficult to detect. Details of the individual fragments are included as an appendix (Table A.1).

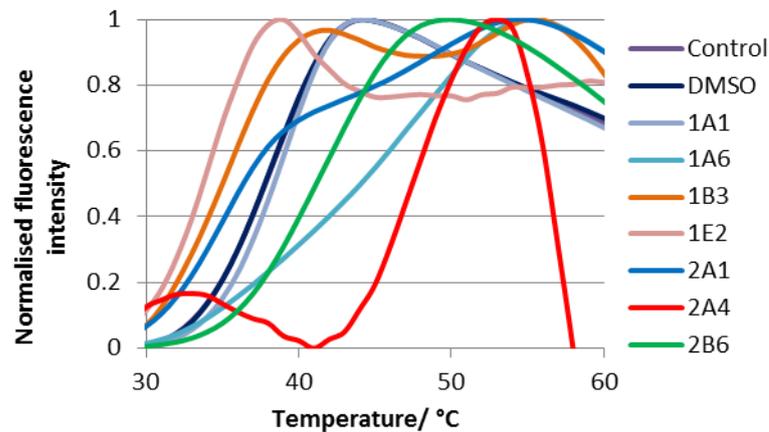
As before, purified HPV11 E2 TAD was diluted into EPPS buffer before diluting 1:1 into EPPS buffer containing Sypro Orange dye, resulting in a final protein concentration in each well of 8  $\mu$ M. Each fragment in turn was added at a final concentration of 10 mM from a DMSO stock. The thermal shift assay was carried out as previously described and the resulting fluorescence intensity curves generated were again used to calculate  $T_m$  values. The  $T_m$  for HPV11 E2 TAD alone was compared to that in the presence of DMSO to ensure the fragment solvent was not impacting protein stability prior to the screen. In each plate HPV11 E2 TAD plus DMSO was used as the control, giving consistent  $T_m$  values averaging 38.01  $^{\circ}$ C (SD = 0.91, range 36.85 – 39.52  $^{\circ}$ C). Fragment hits were determined as those giving a  $T_m$  value greater than 2  $^{\circ}$ C above the protein with DMSO control for the plate [157].  $T_m$  shifts calculated relative to the DMSO control for the plate for each fragment are included in the appendix (Table A.1).

From this assay, 29 fragments (5.71%) were shown to increase the  $T_m$  of HPV11 E2 TAD by more than 2  $^{\circ}$ C, however some of these fragments led to a drastic alteration in the shape of the fluorescence curve. Rather than exhibiting the expected fluorescence profile (Figure 3.4), in some wells the protein displayed a very high initial fluorescence and a limited transition, indicative of protein instability or aggregation, or the fragment interacting directly with the

dye (Figure 3.5). These altered fluorescence curves prevent an accurate assessment of the effect of the fragment on protein  $T_m$  being determined and as such these fragments were discounted. This left 14 fragments (2.76%) which appeared to interact directly with HPV11 E2 TAD and stabilise the protein structure without adversely affecting the unfolding profile (Table 3.3). Examples of the fluorescence curves obtained are shown in Figure 3.8.

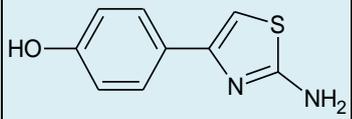
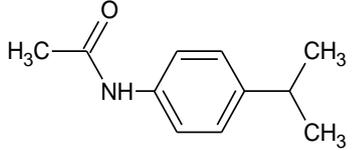
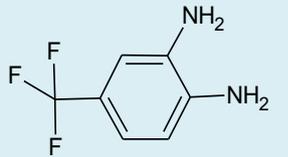
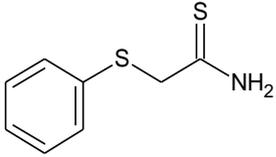
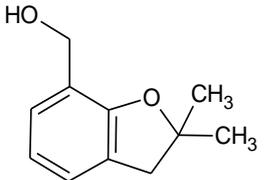
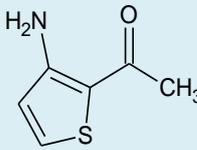
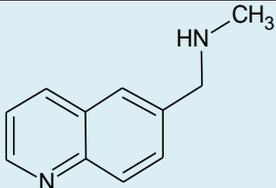
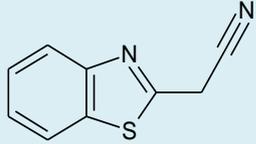
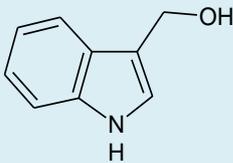
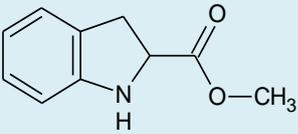
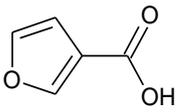
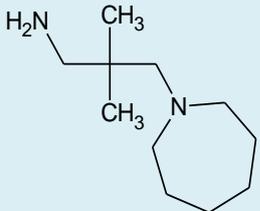
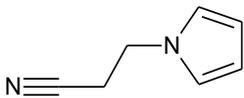
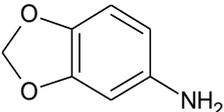
A further 31 (6.10%) fragments decreased the observed  $T_m$  by more than 2 °C. Again, some of these fragments resulted in altered curves indicative of protein instability or aggregation. Discounting these fragments, 14 fragments (2.76%) appeared to be specifically binding and destabilising the protein (Table 3.4).

Of the remaining fragments, the majority tested in this assay (322, 63.39%) gave a  $T_m$  shift of less than 1 °C relative to the DMSO control on the plate. A further 24 fragments (4.72%) displayed no transition in fluorescence intensity, suggesting enough protein destabilisation or aggregation to cause immediate binding of the dye and preventing any  $T_m$  from being calculated.

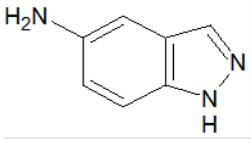
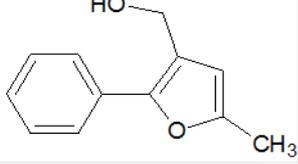
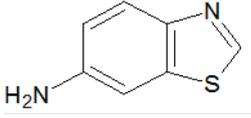
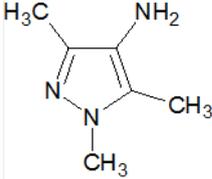
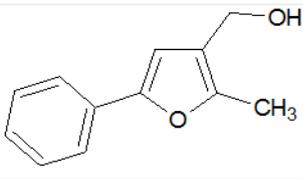
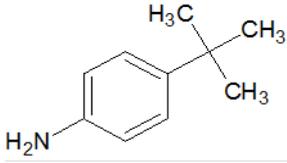
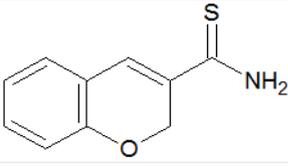
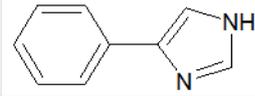
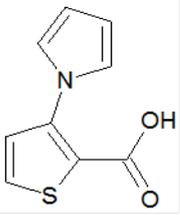
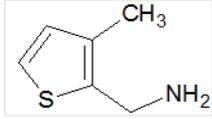
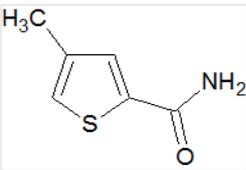
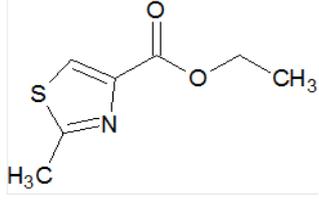
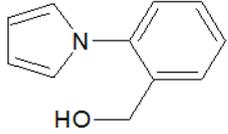


**Figure 3.8 Normalised DSF curves for selected fragments**

HPV11 E2 TAD was incubated with small molecule fragments and Sypro Orange dye. The protein mixture was heated to 95 °C at 1 °C/min and the Sypro Orange fluorescence intensity was recorded at each incremental temperature increase. Normalised fluorescence curves are shown for HPV11 E2 TAD alone, with DMSO or with a selection of fragments. Addition of DMSO (black line) had no effect on HPV11 E2 TAD  $T_m$  (purple line), as did fragment 1A1 (pale blue). Fragments 1A6 (teal) and 2B6 (green) showed an increase in HPV11 E2 TAD  $T_m$ , while fragment 1E2 (pink) showed a decrease in  $T_m$ . Fragments 1B3 (orange) and 2A1 (blue) gave biphasic melting profiles indicative of protein dimerization or stabilisation of an intermediate transition state and fragment 2A4 (red) showed a high initial fluorescence intensity and small transition indicative of protein instability or direct interaction between the protein or fragment and Sypro Orange dye. These curves are representative of the range of melting profiles obtained in the DSF assay.

Fragment ID	Structure	T <sub>m</sub> shift / °C	Fragment ID	Structure	T <sub>m</sub> shift / °C
1A6		7.37	4D1		4.87
1F5		2.9	4E4		3.35
2A11		2.03	4F1		2.93
2B3		14.1	4G1		8.93
2B6		3.4	4G7		6.78
2G1		6.94	4H8		10.7
3E10		2.31	5G1		2.78

**Table 3.3** Fragment hits from the thermal shift assay which increased the T<sub>m</sub> of HPV11 E2 TAD by >2 °C compared to the DMSO control without adversely affecting the fluorescence curve. Highlighted fragments (blue) were selected for further analysis.

Fragment ID	Structure	T <sub>m</sub> shift/°C	Fragment ID	Structure	T <sub>m</sub> shift/°C
1E2		-3.36	3E8		-3.59
2B4		-4.98	3H11		-2.57
2G6		-4.12	5E4		-2.96
2G9		-3.31	6B1		-2.42
2H8		-5.87	6D1		-2.39
3D1		-3.97	6E1		-3.53
3E2		-2.75			

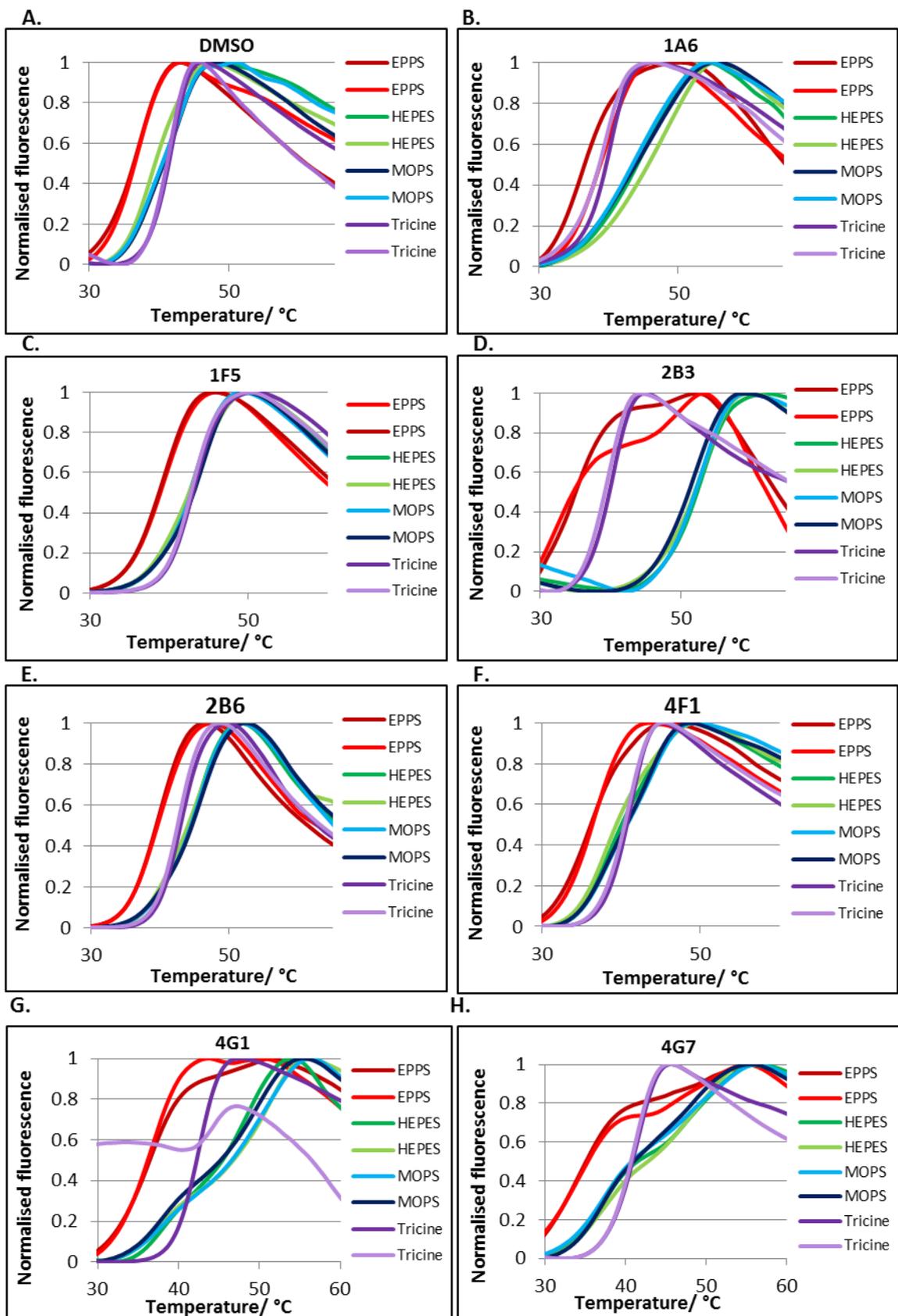
**Table 3.4** Fragment hits from the thermal shift assay which decreased the T<sub>m</sub> of HPV11 E2 TAD by >2 °C compared to the DMSO control without adversely affecting the fluorescence curve.

*Thermal shift – do buffer conditions influence the effects of small molecule fragments on protein stability?*

From the 14 fragment hits which appeared to be binding and stabilising HPV11 E2 TAD structure, as indicated by an increase in the protein  $T_m$ , 8 fragments (highlighted in Table 3.3) were selected to assess further. These fragments were selected to include a diverse collection of structural motifs and cover a range of induced  $T_m$  shifts from 2.9 – 14.1 °C. Another two fragments, 5D4 and 6H6 were included in the set for further testing. These fragments also resulted in a  $T_m$  shift of >2 °C (5D4 = 2.53; 6H6 = 7.39) but displayed altered fluorescence curves consistent with a biphasic dissociation, protein aggregation or fragment insolubility (5D4), or protein instability or fragment interactions with the dye (6H6). This was to determine if these results were specific to the fragment and therefore consistent between screens, or if the results were anomalies due to variability in protein stability between wells.

Having already demonstrated the effects of buffer choice and pH on the protein stability as recorded by the thermal shift assay, the 10 fragments shown to affect HPV11 E2 TAD stability were re-assessed for their influence on HPV11 E2 TAD stability in 3 alternate buffers (tricine, MOPS and HEPES), along with repeating the screen in EPPS. The tricine buffer had previously behaved similarly to EPPS buffer, resulting in similar  $T_m$  values which increased with the addition of NaCl and DTT. The other buffers, HEPES and MOPS, gave consistently high  $T_m$  shifts in the previous screen but were not affected by the addition of NaCl or DTT, suggesting that potentially only specific binding events that stabilise the protein structure would be observed in these systems. Retesting the fragments in these buffers would allow

determination of any cumulative stabilising effect of fragment binding to the protein on top of the effects of a stabilising buffer, or conversely confirm that these small effects are lost in more stabilising conditions.



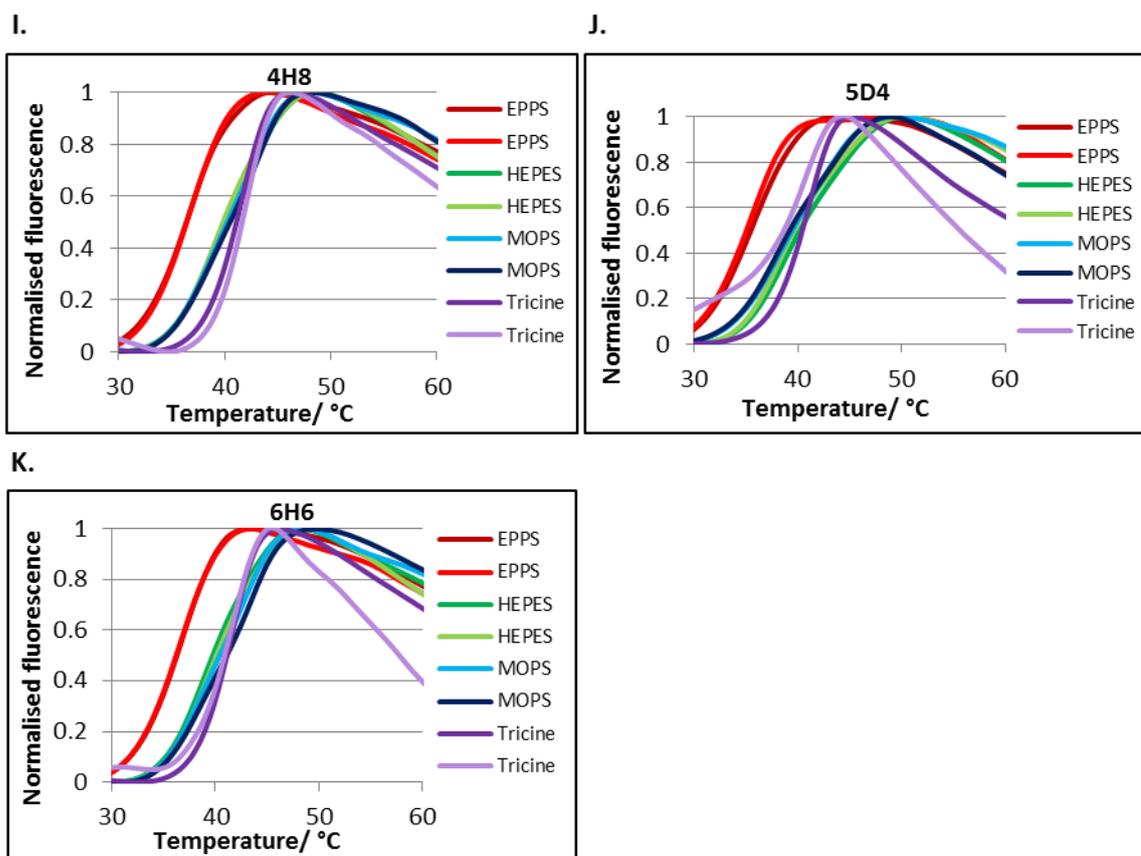


Figure 3.9 DSF curves for HPV11 E2 TAD with selected fragments in different buffers

HPV11 E2 TAD in different buffers was incubated with DMSO or small molecule fragments and Sypro Orange dye, in duplicate to confirm consistency. The protein mixture was heated to 95 °C and Sypro Orange fluorescence was recorded at each incremental temperature increase. Normalised fluorescence curves are shown for HPV11 E2 TAD in 4 different buffers with A. DMSO, B. 1A6, C. 1F5, D. 2B3, E. 2B6, F. 4F1, G. 4G7, H. 4H8, I. 5D4 and J. 6H6. Calculated  $T_m$  values are shown in Table 3.5.

ID	50 mM EPPS pH 8.5			50 mM HEPES pH 7.5			50 mM MOPS pH 7.5			100 mM tricine pH 8		
	A	B	Av. $\Delta T_m$	A	B	Av. $\Delta T_m$	A	B	Av. $\Delta T_m$	A	B	Av. $\Delta T_m$
DMSO	36.27	36.44	-	40.61	40.34	-	40.49	39.42		41.29	40.97	
1A6	36.45	38.78	1.03	44.12	43.59	3.38	44.18	45.94	5.1	39.61	38.61	-2.02
1F5	38.9	38.72	2.45	42.73	42.97	2.37	42.46	42.54	2.54	43.0	42.76	1.75
2B3	34.57	26.05	-6.05	52.04	51.13	11.11	52.28	51.95	12.16	39.77	39.18	-1.65
2B6	39.42	39.6	3.15	44.35	44.6	4.00	44.1	44.01	4.05	42.94	42.5	1.59
4F1	36.15	36.35	-0.11	40.88	40.68	0.30	40.19	39.53	-0.1	40.66	40.35	-0.62
4G1	36.18	36.12	-0.21	46.85	45.39	5.64	45.0	47.2	6.14	42.17	-	42.17
4G7	33.23	30.36	-4.56	40.76	41.01	0.41	41.62	42.93	2.27	40.72	40.61	-0.46
4H8	36.33	36.38	0	40.28	40.52	-0.08	39.95	39.88	-0.04	41.06	41.59	0.20
5D4	35.44	34.93	-1.17	39.56	39.3	-1.05	40.34	39.79	0.11	40.33	38.88	-1.52
6H6	36.28	36.18	-0.13	40.35	40.99	0.19	39.81	40.12	0.01	41.03	40.85	-0.19

**Table 3.5  $T_m$  results for each fragment with 4 buffers**

Each buffer condition was tested in duplicate (A and B) and the average  $T_m$  shift compared to the DMSO control is recorded (Av  $\Delta T_m$ ). Fragments resulting in an increase in  $T_m >1$  °C are highlighted in green and fragments resulting in a decrease in  $T_m >1$  °C are highlighted in red.

Although, for the most part, the  $T_m$  values for the 10 fragments were consistent between replicates in the assay (with the exception of 4G1 in tricine B which resulted in a fluorescence curve indicative of protein aggregation and was therefore discounted), the  $T_m$

values obtained in the EPPS buffer were very different to those obtained from the previous assay. Previously all 10 fragments increased the  $T_m$  of HPV11 E2 TAD by  $>2$  °C compared to the DMSO control in EPPS buffer, however in the second assay, performed in identical buffer conditions, only two of the fragments (1F5 and 2B6) still showed this positive  $T_m$  shift to a similar value as before (Figure 3.9, Table 3.5). Fragment 1F5 gave a  $T_m$  shift of 2.9 °C in the initial screen. In the second screen with EPPS buffer it gave a similar shift of 2.45 °C. This increase in HPV11 E2 TAD  $T_m$  was consistent across the buffers used; in HEPES the  $\Delta T_m$  was 2.37 °C, and in MOPS the  $\Delta T_m$  was 2.54 °C. Similarly, fragment 2B6 gave an initial  $T_m$  shift of 3.4 °C and a second shift of 3.15 °C in EPPS buffer. Again this increase was consistent across the HEPES (4.00 °C) and MOPS (4.05 °C) buffers tested.

A further two fragments (2B3 and 4G7) showed a large, negative  $T_m$  shift in direct contrast to the previous assay. The remaining 6 fragments had little effect on the calculated  $T_m$ . The two fragments which had previously resulted an increase in protein  $T_m$  along with altered dissociation curves indicative of protein instability or fragment insolubility (5D4 and 6H6) did not show the same effects; this time normal dissociation curves were obtained suggesting that the problems previously observed are limited to issues with the individual well, with protein aggregation or unfolding occurring due to local variations between wells rather than being inherently due to the effect of the fragment on the protein.

Interestingly, the 6 fragments (1A6, 2B3, 4G1, 4G7, 4H8 and 6H6) which had previously resulted in the highest positive shifts in  $T_m$  in the initial fragment library screen (Table 3.3) gave the least consistent results when the thermal shift assay was repeated in EPPS buffer. Although an increase in  $T_m$  was observed for some of these fragments in HEPES and MOPS

buffers (1A6, 2B3 and 4G1), the differences between screens could be due to variation in protein stability between wells. The very large increases in protein  $T_m$  could be due to non-specific binding events resulting in a high degree of variation between experiments.

Tricine buffer had previously been shown to increase the stability of HPV11 E2 TAD to the greatest extent. In the experiments shown in Table 3.5, tricine buffer again gave the highest  $T_m$  value in the DMSO control. In this buffer system, the two fragments shown to consistently increase the  $T_m$  of HPV11 E2 TAD to  $>2$  °C still showed a positive shift but to a lesser extent (1F5 = 1.75 °C; 2B6 = 1.59 °C).

Fragment 2B3 gave a very high initial  $T_m$  shift of 14.1 °C, but in the second screen in EPPS buffer, this fragment resulted in a large negative  $T_m$  shift of -6.05 °C. The values obtained in HEPES buffer (11.11 °C) and MOPS buffer (12.16 °C) however were fairly consistent with the initial value obtained from the first screen. Similarly fragment 4G7 also gave a high initial shift (6.78 °C) followed by a large negative shift (-4.56 °C) in the second EPPS screen. In this case, the values obtained in HEPES (0.41 °C) and MOPS (2.27 °C) are not consistent with the initial finding. In both cases the fragment had a limited negative effect on the protein  $T_m$  in tricine.

Fragment 1A6 still increased HPV11 E2 TAD  $T_m$  in EPPS buffer in the second screen but to a much lesser extent than in the initial screen (1.03 vs 7.37 °C). In the HEPES and MOPS buffers the increase in E2  $T_m$  was greater (3.38 °C in HEPES and 5.1 °C in MOPS), but in tricine buffer a negative  $T_m$  shift of -2.02 °C was observed.

These results highlight the inconsistency of the DSF screen for identifying fragment binding. Buffer choice for the screen had a dramatic impact on the results obtained, and these were

not consistent across repeated assays. Individual variation between wells resulted in protein aggregation or unfolding in some cases obscuring the effects of the added fragments, and these variations were found to be independent of the fragment. Although a fast method of screening large libraries of compounds for protein binding, an alternative screen is required to validate the resulting DSF hits as genuine HPV11 E2 TAD interacting compounds.

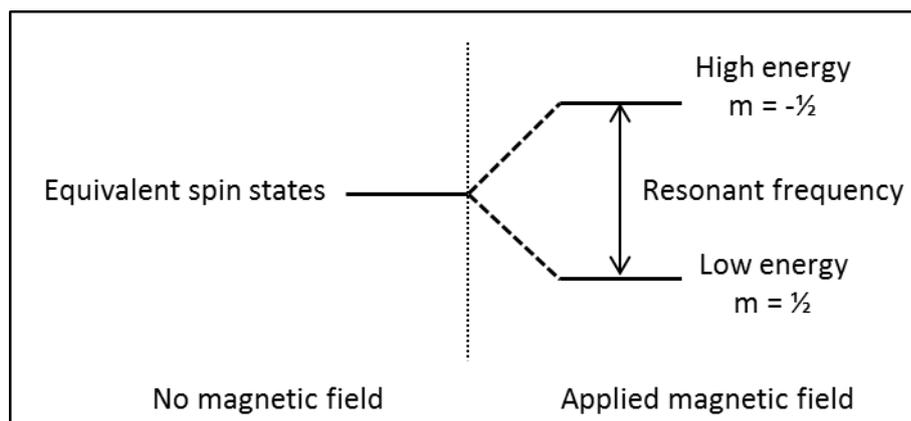
### **3.3.3 Using NMR to screen a fragment library for novel fragments which bind to HPV11 E2 TAD**

*NMR analysis of 10 fragments shown to bind HPV11 E2 TAD by thermal shift assay*

As an alternative to DSF screening described above, nuclear magnetic resonance (NMR) spectroscopy is routinely employed to determine whether small molecules bind to a protein in solution.

NMR spectroscopy works by utilising the magnetic properties of specific NMR active nuclei within molecules to gather information about the chemical state and surrounding environment of each nucleus in turn. As such the NMR spectrum for a given molecule is unique, acting as a molecular fingerprint which allows individual nuclei to be identified and assigned.

NMR active nuclei have a total number of neutrons and protons which is odd, giving them an intrinsic magnetic moment, and are designated spin  $\frac{1}{2}$ . When a magnetic field is applied to the sample, the nuclei have two available spin states,  $-\frac{1}{2}$  (low energy) where the spin is aligned with the applied magnetic field and  $+\frac{1}{2}$  (high energy) where the spin is opposed to the magnetic field.



**Figure 3.10 Diagram of the spin state energy levels.**

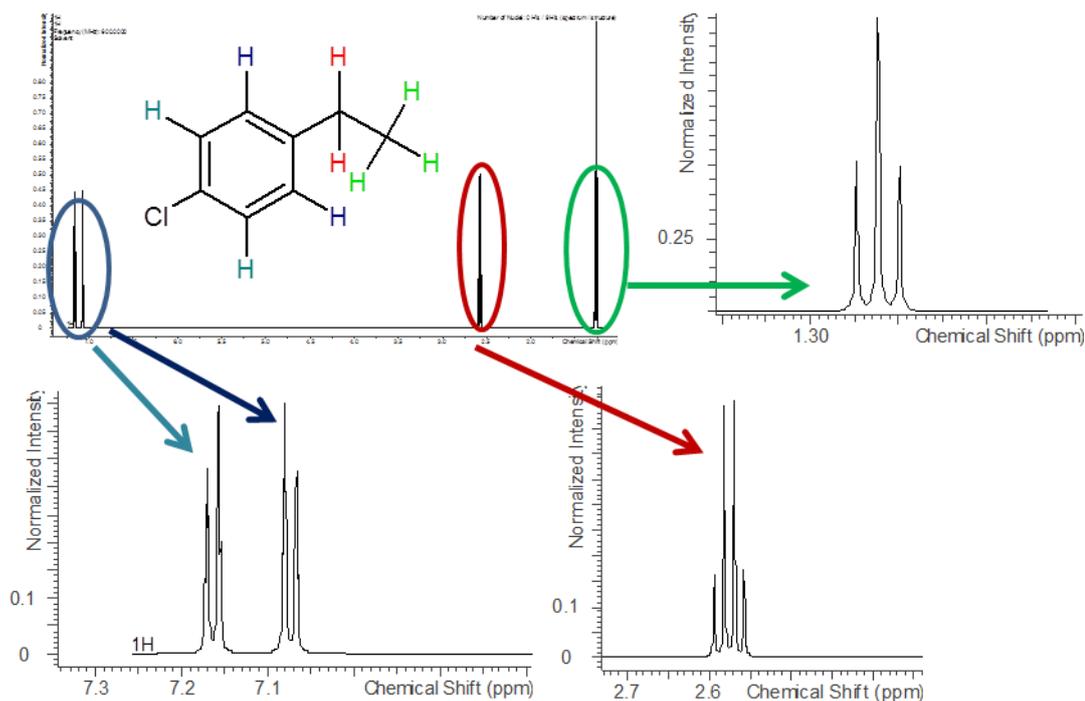
**NMR active spin  $\frac{1}{2}$  nuclei exist in two equivalent spin states. When a magnetic field is applied the spin states separate to a low and high energy state. At equilibrium, the nuclei will reside in both spin states with a greater number populating the lower energy level. When energy is applied that is equal to the resonant frequency separating the levels, nuclei can be excited to the higher energy level, eventually saturating the system with equal numbers at both levels. NMR spectroscopy takes advantage of the change in spin states to provide information about the environment of the nuclei.**

When nuclei are excited by a pulse of radio frequency (RF) electromagnetic radiation equivalent to the Larmor precession frequency (the velocity of angular precession of nuclei around the applied magnetic field) they can be excited to move up to the higher energy level (Figure 3.10). At equilibrium, more nuclei will reside at the lower energy level but by pulsing with RF radiation the sample can be saturated, resulting in equal numbers of nuclei in each spin state. Although the spin states are identical for all common nuclei within a molecule, the resonant frequency (the absorbed RF required to move between spin states) detected for each nucleus varies due to effects from neighbouring nuclei leading to distinctly different

signals being detected. It is these differences which are due to the unique environment of each nucleus within a molecule that result in a unique NMR spectrum for a given molecule (Figure 3.11).

Excited nuclei can move back down to the lower energy level by relaxation. Fluctuations in the magnetic field due to the motion of the sample molecule (lattice) cause nuclei to return to equilibrium through a process called spin-lattice relaxation ( $T_1$ ). Spin-spin relaxation ( $T_2$ ) is an additional process whereby neighbouring nuclei can exchange quantum states. Cross-relaxation occurs when the relaxation processes of large molecules are transferred to interacting small molecules, while unbound small molecules tumble too quickly in solution for effective relaxation to occur.

Related to this is the Nuclear Overhauser Effect (NOE), the transfer of nuclear spin polarisation from one nuclear spin population to another via cross-relaxation. Large molecules such as proteins tumble slowly in solution and as such experience large negative NOEs. Smaller molecules however tumble rapidly and experience much smaller positive or negative NOEs. When a fragment binds to the protein it brings nuclei into close spatial proximity and allows the fragment to adopt the characteristics of the protein. When the fragment dissociates, it carries the transferred large negative NOE into the bulk solution where it can be detected as a change in the fragment spectrum.



**Figure 3.11** Representative  $^1\text{H}$  NMR spectrum of 4-chloroethylbenzene.

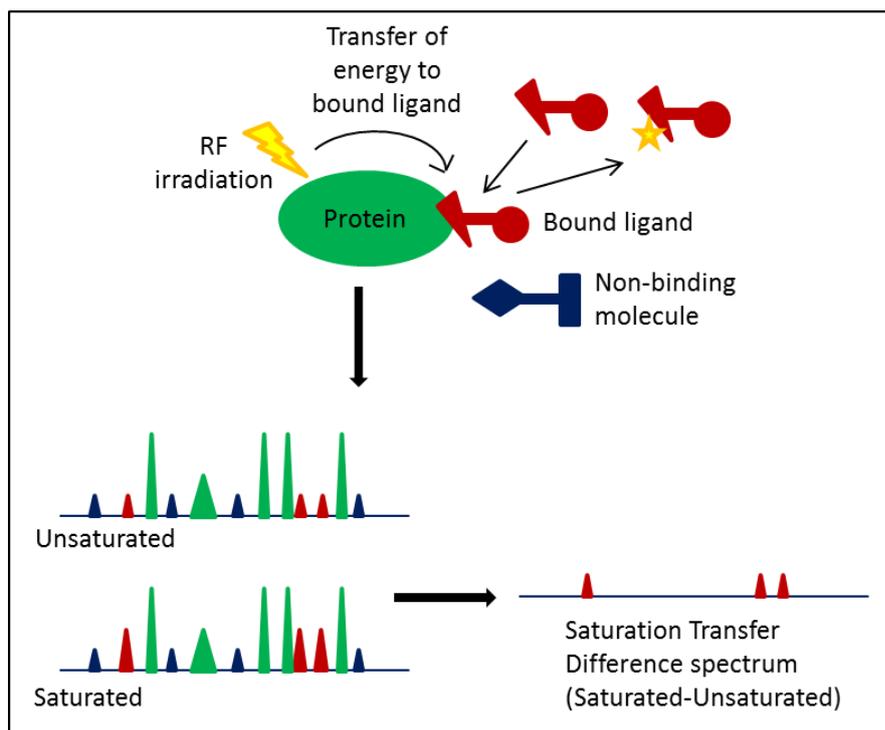
The  $^1\text{H}$  NMR spectrum of 4-chloroethylbenzene shows 4 distinct sets of peaks corresponding to the 4 unique proton environments within the molecule. The methyl group protons (green) are observed as a triplet with an intensity of 3H, because there are 3 identical H atoms in these position and they are influence by the neighbouring ethyl protons (red). Similarly, the ethyl protons are observed as a quartet with an intensity of 2H, showing two identical nuclei being affected by the 3 protons on the adjacent methyl carbon (green). The benzene group is para-substituted resulting in a plane of symmetry and therefore the protons opposite each other are equivalent. As such the 4 aromatic protons (teal, blue) give rise to 2 distinct signals with an intensity of 2H. Both are doublets, showing that they are influenced by the proton on the adjacent carbon atom, but one signal is shifted due to its proximity to the neighbouring chlorine atom (teal). The effects

**of the surrounding atoms on the signal from each  $^1\text{H}$  nucleus in turn results in a unique spectrum for any given molecule.**

To confirm that the 10 fragments initially identified by DSF as stabilising the HPV11 E2 TAD structure were binding to the protein, two NMR experiments were employed, saturation transfer difference (STD) and water-ligand observed via gradient spectroscopy (wLOGSY). Both STD and wLOGSY are indirect assays looking at changes in the fragment spectra following binding to the protein resulting from a transfer of NOEs to the bound fragment.

In STD NMR, the protein is selectively saturated by a RF pulse. When fragments bind reversibly to the protein, NOEs are transferred from the saturated protein to the fragment. A pre-saturation NMR spectrum is then subtracted from a post-saturation spectrum, removing the signals from any unbound fragments and the protein. The remaining signals are those of fragments which had bound to the protein and therefore experienced a change in NOE/relaxation which can be detected (Figure 3.12).

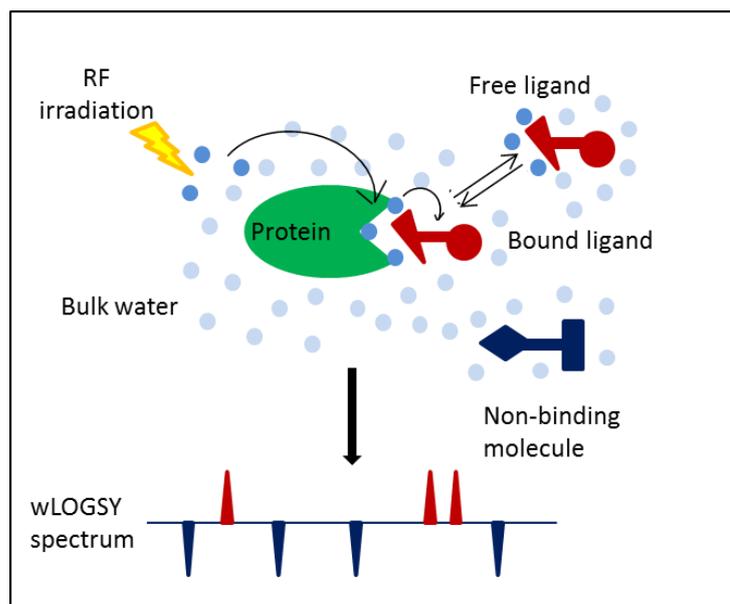
wLOGSY experiments operate on a similar principle. In this case the bulk water is selectively irradiated by RF pulse targeted to the resonance frequency of the water molecules. At the protein-ligand interface, the saturated water can transfer energy to the bound fragment by NOE transfer. Because any bound small molecule is acting as part of the large protein molecule, this results in a large, negative NOE. As before, the fragment can then dissociate and carry this NOE back into the bulk solution where it can be detected. The resulting NMR spectra obtained will show the signals from all fragments present in solution, however the phase of those which were bound and gained transferred NOEs will be opposite to the phase of non-binding fragments, allowing easy identification (Figure 3.13).



**Figure 3.12 Schematic to illustrate the principle of STD NMR.**

Protein is selectively irradiated by RF pulse targeted to a resonance frequency within the protein, away from the signals of the small molecule fragments. The energy (NOE) can be transferred to reversibly bound ligands (red), which carry the NOE back into solution where the changes can be detected.

To facilitate the analysis of the experimentally obtained NMR data, a database was generated by members of the Naismith group (University of St Andrews, UK) using Bruker Amix software containing the reference spectrum for each fragment. This dataset was used to identify the fragments which gave positive results in the NMR screen along with ensuring that there was no spectral overlap between fragments being screened in combination during later experiments.

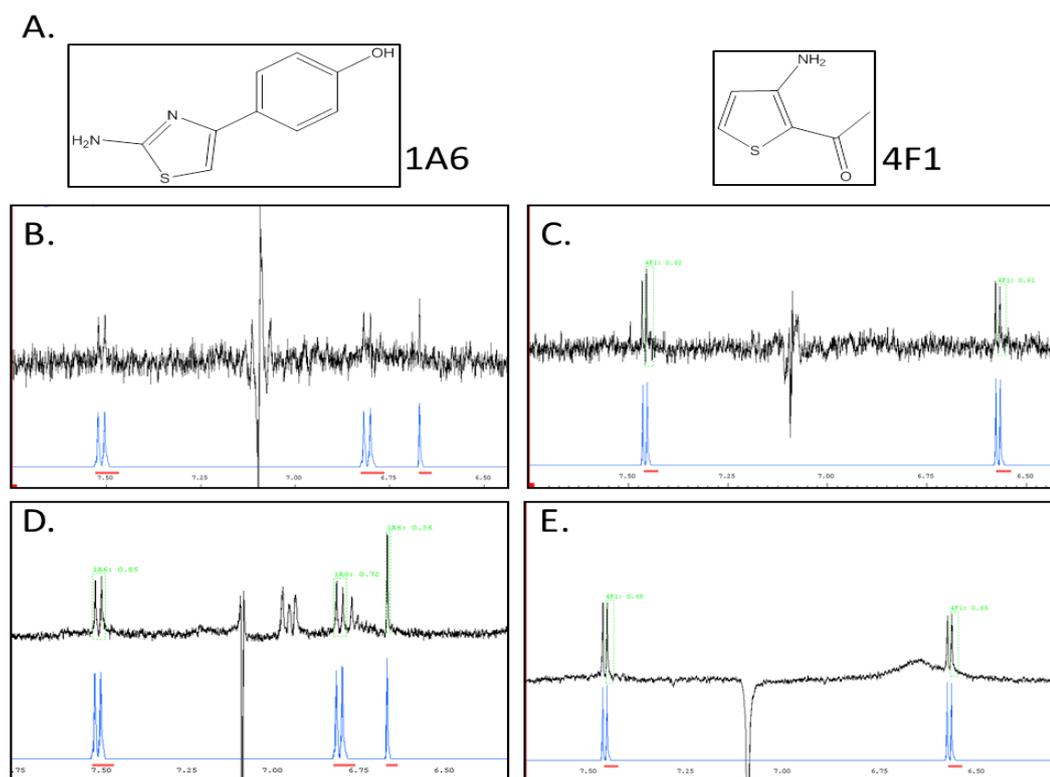


**Figure 3.13 Schematic to illustrate the principle of wLOGSY NMR.**

**The bulk water is selectively irradiated by a pulse of RF radiation. The energy is transferred to the protein and the bound ligands (red) by NOE transfer. When the fragments dissociate they carry the transferred energy back into the bulk solution where this can be detected as a change in phase of the fragment spectrum following processing.**

Using this database, the 10 fragments identified from the DSF screen were combined in pairs selected to minimise spectral overlap in order to reduce both the amount of protein and the instrument time required. Purified HPV11 E2 TAD at a final concentration of 20  $\mu\text{M}$  was combined with 4 mM fragments in a phosphate buffer containing 10%  $\text{d}_2\text{O}$ . NMR spectra were obtained using a 500 MHz NMR machine (Bruker Avance 500 with a Bruker Ultrashield 500 MHz magnet and a TXI probe), acquiring data for a water suppression and  $^1\text{H}$  spectrum initially then an STD and a wLOGSY. Raw data was processed using TopSpin before analysis using Amix and the reference dataset mentioned above. Of the 10 fragments used in these experiments only two were identified in both the STD and wLOGSY experiments as

interacting with HPV11 E2 TAD (Figure 3.14). These were fragments 1A6 which had previously been shown to increase HPV11 E2 TAD  $T_m$  consistently in multiple buffers over two screens (Table 3.3, 3.5) and 4F1 was shown to increase  $T_m$  in the first DSF screen (Table 3.3) but not in the repeat screen in multiple buffers.



**Figure 3.14 STD and wLOGSY NMR identification of fragments binding to HPV11 E2 TAD.**

Mixtures of small molecule fragments were screened for binding to purified HPV11 E2 TAD by STD and wLOGSY NMR. Experimentally obtained spectra were compared to reference spectra for the fragments. A. Structures of the two fragment hits, 1A6 and 4F1. Experimental STD spectrum (black) overlaid with reference spectrum (blue) for 1A6 (B) and 4F1 (C), and wLOGSY spectrum for 1A6 (D) and 4F1 (E).

For both fragments the STD spectra obtained (Figure 3.14 B and C) clearly show peaks matching those in the corresponding reference spectra. For 1A6 two doublet peaks are visible corresponding to the two different hydrogen environments on the para-substituted phenol group. The remaining singlet peak corresponds to the thiazole ring proton. In the spectrum for 4F1, the two doublet peaks visible in the experimental spectra correspond to the two thiophene protons.

As mentioned, the STD spectra only show signals from fragments after binding to the irradiated protein and dissociation back into the bulk solvent. The wLOGSY spectra only show positive fragment peaks after protein binding. As the peaks corresponding to both 1A6 and 4F1 could be clearly observed in both experiments this is a good indication that the fragments are reversibly binding to the protein.

#### *NMR screen of fragment library*

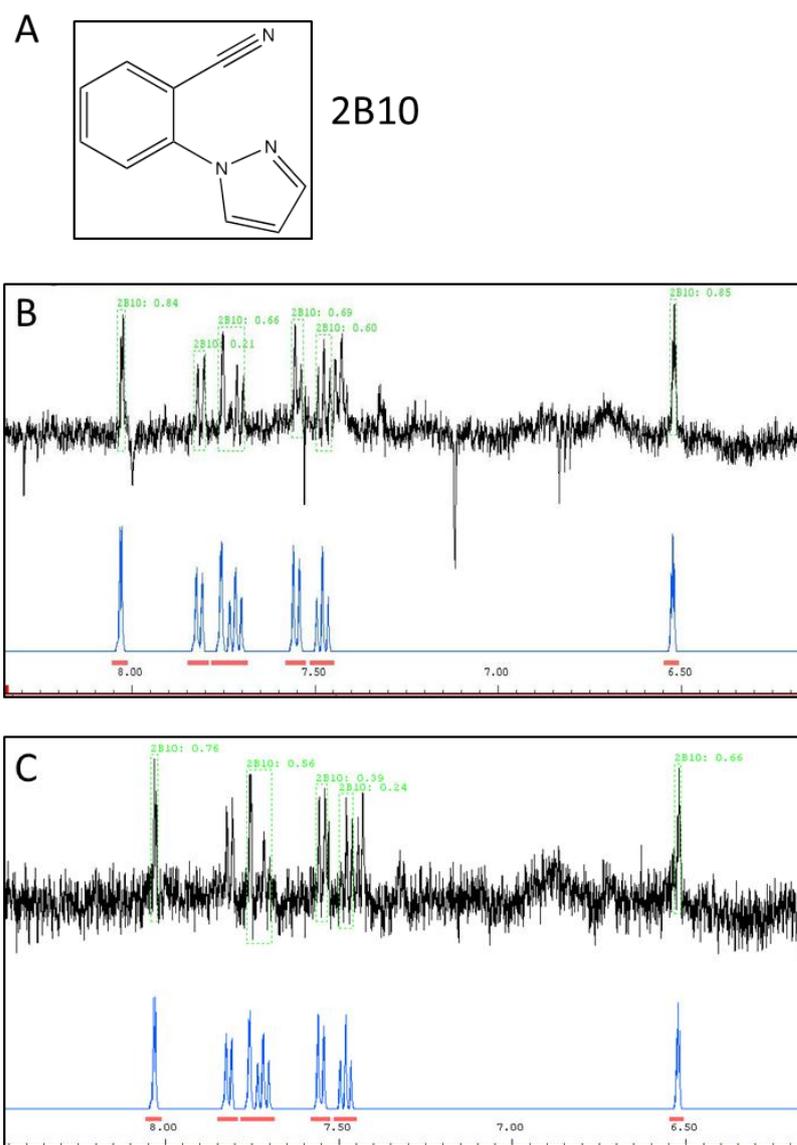
Given that the results from the initial thermal shift screen were not readily corroborated by the NMR experiments, the same NMR experiments were then used to screen the fragment library to look for interacting hits. To allow the entire fragment library to be screened with the minimum of protein to attempt to circumvent problems with protein stability and to reduce instrument time requirements, the individual fragments were cocktailed into sets of 8 containing 125 mM of each fragment. The reference dataset described above was used by the Naismith group (University of St Andrews, UK) to design the cocktails ensuring that there was no spectral overlap between the fragments in each set.

Cocktailing the fragments in this way allowed the 436 fragments previously determined to be compatible with the NMR assays by the Naismith group to be screened in 55

experiments, acquiring both STD and wLOGSY spectra for each. 750  $\mu$ L NMR samples were made up containing 20  $\mu$ M HPV11 E2 TAD and a final concentration of each cocktailed fragment of 1 mM. From this screen, 29 fragments were identified as binding to HPV11 E2 TAD in both the STD and wLOGSY screen (see representative data in Figure 3.15, comprehensive list of fragment hits in Table 3.6), with a further 11 fragments being identified as hits in the wLOGSY experiments alone.

Of these hits, only one had previously given rise to a positive shift in HPV11 E2 TAD  $T_m$  compared to the DMSO control, 4F1. The majority of the other hits had shown very little effect on protein stability in the thermal shift assay. The other fragment which had previously been positively identified as a hit in both the thermal shift and NMR experiments, 1A6, was not picked up in this NMR screen.

Details of the NMR results are included in the appendix table of the fragment structures.



**Figure 3.15 Representative spectra identifying small molecule fragments binding to HPV11**

**E2 TAD**

Cocktails of 8 small molecule fragments were screened for binding to HPV11 E2 TAD by STD and wLOGSY NMR. A. Structure of one identified fragment hit, 2B10. Experimentally obtained wLOGSY spectrum (B. black) and STD spectrum (C. black) overlaid with the reference spectrum for 2B10 (blue). Matching peaks are highlighted in green.

Fragment ID	T <sub>m</sub> shift/ °C	Fragment ID	T <sub>m</sub> shift/ °C	Fragment ID	T <sub>m</sub> shift/ °C
1C1	0.19	3C10 <sup>a</sup>	-1.64	4A11	0.54
1C10	1.16	3D1 <sup>c</sup>	-3.97	4B7 <sup>d</sup>	-0.12
1D10 <sup>d</sup>	1.35	3D6 <sup>a</sup>	-0.35	4B8	N/A
2B10 <sup>d</sup>	-1.62	3D11	-0.27	4E1	-1.02
2B11	0.02	3E2 <sup>c,d</sup>	-2.75	4F1 <sup>b,d</sup>	2.93
2D1 <sup>a</sup>	0.90	3E7	-0.63	4G10 <sup>a</sup>	-1.10
2F6 <sup>d</sup>	-0.43	3F1 <sup>d</sup>	-0.74	5A4 <sup>a</sup>	* 8.51
2H4	*46.38	3F5	-0.66	5B9 <sup>d</sup>	-0.95
2H7 <sup>a</sup>	0.13	3F11 <sup>d</sup>	-0.72	5C2	-1.25
3A8 <sup>a</sup>	-1.25	3G7	-0.81	5C7 <sup>a</sup>	-0.96
3A11	-0.66	3H10 <sup>a</sup>	-1.03	6C1	-1.56
3B3	-0.60	4A1 <sup>d</sup>	0.43	6H1 <sup>a</sup>	-1.46
3B9	-0.41	4A4	1.10		
3C1	-0.31	4A7	* -2.9		

**Table 3.6 Fragments identified as binding HPV11 E2 TAD by NMR.**

Fragments were screened for binding to HPV11 E2 TAD by STD and wLOGSY NMR. Fragment hits are included along with T<sub>m</sub> shift compared to the DMSO control sample in the previous thermal shift assay screen of the library. <sup>a</sup>Fragments identified by wLOGSY only (not STD); <sup>b</sup>identified as a hit by DSF assay; <sup>c</sup>resulted in a negative T<sub>m</sub> shift of >2 °C in previous DSF assays; <sup>d</sup>Fragments selected for further analysis on the basis of published

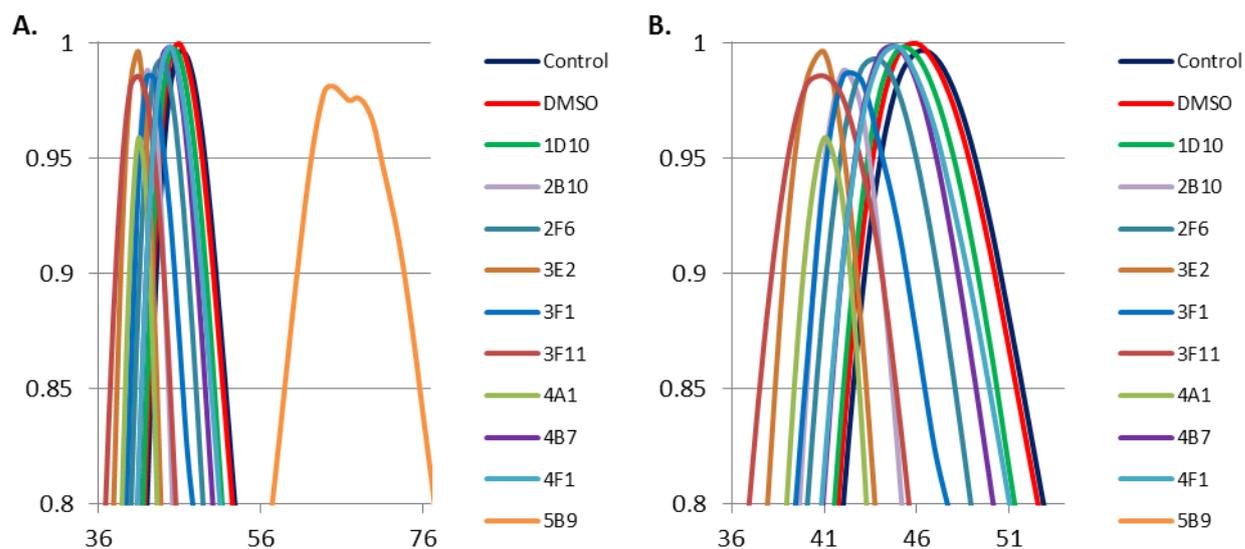
**cytotoxicity data [158], structural diversity and commercial availability. \*DSF assay resulted in altered fluorescence curve profile**

Previously published work by [158] investigated the fragment library for cytotoxicity in HeLa and A549 cells. For further downstream development of the fragment hits towards potential drug leads, having starting fragments which are already determined to have low levels of cytotoxicity would be advantageous. Additionally, assessing the fragment hits for cytotoxicity would allow for in cell assays to look at the effects on the protein in cell.

From the fragments which showed low level cytotoxicity, 10 were selected which are both structurally diverse and commercially available (Table 3.6). These fragments were used for further analysis.

*Screening 10 fragment hits from NMR screen by thermal shift assay*

Given the variability of the thermal shift assay data and the dependence on buffer conditions as previously shown, the 10 fragments selected from the NMR screen were again used in the thermal shift assay, this time using the NMR buffer to look for any affect the binding may have on the protein stability under these conditions.



**Figure 3.16 Normalised fluorescence curves showed little stabilisation of HPV11 E2 TAD from selected small molecule fragments.**

**A.** The  $T_m$  of purified HPV11 E2 TAD in combination with fragments selected from the NMR library screen was determined by thermal shift assay. Normalised fluorescence curves are shown. Protein  $T_m$  is determined as the mid-point of the fluorescence intensity transition using Boltzmann's equation. **B.** Close up of the normalised DSF curve peaks to allow visualisation of the differences in  $T_m$  between similar fragments. Fragment 5B9 result is excluded from this view.

	1	2	3	4	5	6	7	8	Av	Av $\Delta T_m$
Control	40.17	39.22	39.22	39.19	39.36	39.09	39.13	39.29	39.33	-
DMSO	39.48	39.16	39.17	39	39.14	39.09	38.99	39.02	39.13	-
4A1	37.48	35.85	37.39	37.1	36.61	36.64	36.86	37.66	36.95	-2.18
2F6	36.69	36.87	37.25	37.55	37.76	37.83	38.3	38.08	37.54	-1.59
4F1	38.11	38.4	38.23	37.82	38.81	38.68	37.78	38	38.23	-0.90
4B7	38.15	38.24	38.36	38.06	38.87	38.41	38.27	38.21	38.32	-0.81
1D10	39.28	39.15	39.39	38.93	38.88	38.54	38.91	38.83	38.99	-0.14
3F11	34.67	34.33	33.42	33.48	33.44	34.07	34.48	34.49	34.05	-5.08
2B10	38.23	38.33	-	-	37.28	37.47	37.03	37.65	37.67	-1.47
3E2	35.63	36.19	35.52	-	36.03	35.65	35.62	35.92	35.79	-3.34
5B9	53.5	47.34	54.1	44.51	47.57	43.53	51.61	54.36	49.57	10.43
3F1	36.41	36.48	36.34	36.45	36.16	37.63	36.33	36.8	36.58	-2.56

**Table 3.7 HPV11 E2 TAD  $T_m$  values from DSF assay with NMR fragment hits**

DMSO or small molecule fragments were added to HPV11 E2 TAD with Sypro Orange dye and the protein solution was heated to 95 °C, recording the Sypro Orange fluorescence at each temperature point. The protein  $T_m$  was determined from the obtained fluorescence curves for each sample. Each fragment was used in 8 replicates and the average  $T_m$  shift relative to DMSO (Av  $\Delta T_m$ ) was determined. Samples resulting in a decrease in  $T_m$  are highlighted in red, samples resulting in an increase in  $T_m$  are highlighted in green.

From this assay, only one fragment was shown to have a significant stabilising effect on HPV11 E2 TAD in these conditions, 5B9 despite not giving a positive result in other screens. The only fragment which had previously given a consistently positive shift in  $T_m$ , 4F1 now showed no effect on HPV11 E2 TAD stability.

With the variable reliability of the hits identified from the thermal shift and NMR screens, carrying out a more in-depth binding assay would both confirm the binding interaction and allow the binding sites on the protein target to be mapped.

### **3.3.4 Attempting to identify fragment binding sites on HPV11 E2 TAD**

#### *2D/3D NMR – buffer screening*

2D NMR is a widely used method for assessing the binding sites of small molecules on protein targets. By generating labelled protein ( $^{15}\text{N}$  or  $^{13}\text{C}$ ), NMR can be used to identify distinct signals from individual amino acids within the protein. When fragments bind to specific residues in the protein, this binding affects the signal from those residues leading to a change in the observable spectrum. By identifying the specific residues involved in the interaction, the binding site within the protein can thus be mapped.

To identify the specific residues with an altered spectrum following fragment binding, the specific amino acid peaks have to be assigned. This is done using 3D NMR with protein labelled with both  $^{15}\text{N}$  and  $^{13}\text{C}$ . Both the 2D and 3D NMR experiments require purified protein of a high concentration which will remain stable in an NMR compatible buffer for long periods of time. Unfortunately, HPV11 E2 TAD precipitates rapidly in the usual sodium phosphate buffer used in 2D and 3D NMR experiments, making it unsuitable to be used for long acquisition periods. A particular problem observed with HPV11 E2 TAD is that the

protein is particularly prone to precipitation in low salt buffers. The presence of high concentrations of chloride ions however is problematic for NMR ( $\text{Cl}^-$  counteracts the effects of the cryoprobe leading to peak broadening) [159] therefore an alternative stabilising buffer is necessary for long term NMR experiments.

Although giving highly variable results with determining the effects of fragments on the protein, the thermal shift assay used previously is more commonly used for screening buffers for effects on protein stability. To attempt to identify suitable buffers to be used for further NMR experiments, the DSF assay was again employed to screen multiple buffers with various NMR compatible additives. Purified HPV11 E2 TAD at 100  $\mu\text{M}$  was combined with 18  $\mu\text{L}$  buffer plus Sypro Orange for a final protein concentration of 10  $\mu\text{M}$  and the thermal shift assay was again used to determine the  $T_m$  of the protein.

	Water	Bis-Tris pH 6	Na Cacodylate pH 6.5	MES pH 6.5	Na phosphate pH 7	MOPS pH 7	HEPES pH 7.5	Tris pH 8
-	40.23	38.27	40.22	39.23	42.22	40.32	40.28	41.27
500 mM NaCl	45.23	44.27	45.28	45.27	46.26	45.23	46.30	45.23
150 mM NaCl	43.23	41.23	43.25	41.23	44.27	43.28	43.25	43.27
150 mM KCl	41.22	39.23	41.23	39.23	43.22	41.30	42.27	42.27
150 mM NH <sub>4</sub> Cl	42.22	41.23	42.23	41.23	44.27	42.23	43.25	43.27
150 mM MgCl <sub>2</sub>	45.25	44.27	45.28	45.27	44.27	45.25	46.30	45.23
150 mM NH <sub>4</sub> SO <sub>4</sub>	40.23	38.27	40.22	39.23	42.22	40.32	40.28	41.25
150 mM NaCl, 2 mM CHAPS	42.23	39.23	41.23	40.22	42.22	41.30	42.27	41.25
150 mM NaCl, 0.5 mM TMAO	44.27	43.22	44.28	43.23	45.28	44.27	44.22	44.23
150 mM NaCl, 50 mM Arg-Glu	39.25	42.22	43.25	42.22	45.28	44.27	44.25	44.25
150 mM NaCl, 10% glycerol	45.25	43.22	45.28	44.30	45.28	44.27	45.23	45.25
150 mM NaCl, 200 mM sucrose	44.27	42.22	44.28	43.23	45.27	44.27	44.25	45.25

**Table 3.8. HPV11 E2 TAD T<sub>m</sub> in buffer screen 1**

Sypro Orange dye was added to HPV11 E2 TAD in different buffers and the protein mixture was heated to 95 °C at 1 °C/min. Sypro Orange fluorescence was recorded at each incremental temperature increase and protein T<sub>m</sub> values were calculated from the

**resulting fluorescence curve. Samples with a  $T_m$  of  $>46$  °C are highlighted in green, 44-46 °C in blue, 42-44 °C in purple and 40-42 °C in orange.**

The initial buffer screen was carried out using a panel of buffers and additives commonly used in NMR and crystallography assays known to stabilise protein structures. As expected, the addition of 500 mM NaCl increased the stability of HPV11 E2 TAD in all the buffers tested by 4-6 °C, but this level of  $\text{Cl}^-$  is incompatible with NMR. Lowering the concentration of NaCl to 150 mM, a level which is compatible with NMR, still stabilised the protein but reduced the increase in  $T_m$  to 2-3 °C. Some additives only had a small effect on the protein stability (ammonium sulphate, KCl, CHAPS) but the combinations of 150 mM NaCl with 200 mM sucrose or 10% glycerol increased the stability to levels equivalent to 500 mM NaCl alone. Trimethylamine-N-oxide (TMAO) also had a good stabilising effect but is not an ideal additive for NMR experiments and so was discounted for further assays. Arginine-glutamate also had an additional stabilising effect on the protein above that of the NaCl alone. The sodium phosphate, MOPS, HEPES and Tris buffers were consistently the most stabilising, and so these were selected for further analysis.

Subsequent rounds of buffer screening combined these 4 buffers (phosphate, MOPS, HEPES and Tris) with varying combinations of  $\text{MgCl}_2$ , sucrose and arginine-glutamate, additives which were shown to increase protein stability and which are also compatible with NMR analysis. Low concentrations of NaCl which are compatible with NMR were also included in the screen, to allow comparison of the stabilisation induced by other additives with that due to the known stabilisation from NaCl. After each buffer set was tested, the combinations

resulting in the highest protein stability were used to devise the next buffer set for testing.

Details of all the buffer combinations screened are given in the appendix (Tables A2-A5).

	-	10 mM MgCl <sub>2</sub> / 20 mM sucrose	20 mM MgCl <sub>2</sub> / 50 mM sucrose	50 mM MgCl <sub>2</sub> / 100 mM sucrose	100 mM MgCl <sub>2</sub> / 150 mM sucrose	150 mM MgCl <sub>2</sub> / 200 mM sucrose	200 mM MgCl <sub>2</sub> / 250 mM sucrose	250 mM MgCl <sub>2</sub> / 500 mM sucrose
Sodium phosphate pH 7 200 mM sucrose	42.23	43.27	43.25	44.22	44.2	45.22	45.25	45.25
Sodium phosphate pH 7 200 mM sucrose 50 mM RE	42.23	43.23	43.25	43.25	43.22	44.22	43.22	45.25
Sodium phosphate pH 7 100 mM MgCl <sub>2</sub>	43.25	43.28	43.25	43.25	43.22	43.22	44.23	43.25
Sodium phosphate pH 7 100 mM MgCl <sub>2</sub> 50 mM RE	43.25	43.28	43.25	43.25	43.22	44.2	44.23	46.27
MOPS pH 7 200 mM sucrose	42.25	42.25	43.25	44.23	45.22	46.22	46.23	46.27
MOPS pH 7 200 mM sucrose 50 mM RE	41.23	41.25	41.2	41.23	42.23	42.23	42.23	43.25
MOPS pH 7 100 mM MgCl <sub>2</sub>	44.25	44.28	44.25	44.23	45.22	45.2	45.28	45.25
MOPS pH 7 100 mM MgCl <sub>2</sub> 50 mM RE	44.25	44.28	44.25	44.23	45.22	45.2	45.28	46.27
HEPES pH 7.5 200 mM sucrose	41.22	42.25	42.23	44.23	44.2	45.2	46.27	46.27
HEPES pH 7.5 200 mM sucrose 50 mM RE	40.25	41.25	40.22	41.22	41.22	41.25	41.23	42.22
HEPES pH 7.5 100 mM MgCl <sub>2</sub>	44.25	44.27	44.23	44.23	45.22	45.2	45.28	45.3
HEPES pH 7.5 100 mM MgCl <sub>2</sub> 50 mM RE	44.25	44.27	44.23	45.23	45.22	45.2	46.27	47.27

**Table 3.9 HPV11 E2 TAD T<sub>m</sub> values from buffer screen 4**

Sypro Orange dye was added to HPV11 E2 TAD in different buffers containing various concentrations of sucrose, MgCl<sub>2</sub> and arginine-glutamate (RE) and the protein mixture was heated to 95 °C at 1 °C/min. Sypro Orange fluorescence was recorded at each incremental temperature increase and protein T<sub>m</sub> values were calculated from the resulting

fluorescence curve. To each sucrose containing buffer (left), the stated concentration of  $\text{MgCl}_2$  was added (top), and to each  $\text{MgCl}_2$  containing buffer the stated concentration of sucrose was added. Samples with a  $T_m$  of  $>46$  °C are highlighted in green, 44-46 °C in blue, 42-44 °C in purple and 40-42 °C in orange.

After 4 successive rounds of screening buffer combinations, the MOPS and HEPES buffers remained the most consistently stabilising. HPV11 E2 TAD had the highest recorded  $T_m$  in HEPES buffer containing 100 mM  $\text{MgCl}_2$ , 50 mM RE and 500 mM sucrose. MOPS and HEPES buffers containing 200 mM sucrose and 150 – 250 mM  $\text{MgCl}_2$  were also highly stabilising conditions. The addition of RE led to a reduction in protein stability until high concentrations of  $\text{MgCl}_2$  or sucrose were added to counter this effect. Sodium phosphate buffer had lower stabilising effect on HPV11 E2 TAD, requiring high concentrations of  $\text{MgCl}_2$  or sucrose to be added to reach the same  $T_m$  as observed in the MOPS and HEPES buffers.

Although addition of high concentrations of sucrose or  $\text{MgCl}_2$  were able to increase the  $T_m$  of HPV11 E2 TAD, at these concentrations the buffer is no longer ideal for NMR. High sucrose concentrations increase the buffer viscosity, while increasing  $\text{Cl}^-$  ion concentrations negatively affect the NMR cryoprobe and distort the acquired results. For the long data acquisition times required for 2D and 3D NMR experiments, the concentration of these components must be finely balanced against protein stability to optimise conditions which both allow the protein to remain stable and in solution for the entirety of the acquisition period, and do not have a negative impact on the quality of the data collected. On these grounds, three buffers were selected which meet these criteria, 25 mM MOPS pH 7 with 200 mM sucrose and 150 mM  $\text{MgCl}_2$ , 25 mM HEPES with 150 mM sucrose and 100 mM  $\text{MgCl}_2$ ,

and 25 mM sodium phosphate pH 7 with 200 mM sucrose and 150 mM MgCl<sub>2</sub>. The Tris purification buffer was also included in this screen for comparison (25 mM Tris pH 7.5 with 500 mM NaCl) and the DSF assay was repeated with 8 replicates of each buffer to allow consistency to be assessed (Table 3.10).

	25 mM MOPS pH 7 200 mM sucrose 150 mM MgCl <sub>2</sub>	25 mM HEPES pH 7.5 150 mM sucrose 100 mM MgCl <sub>2</sub>	25 mM sodium phosphate pH 7 200 mM sucrose 150 mM MgCl <sub>2</sub>	25 mM Tris pH 7.5 500 mM NaCl
1	46.23	44.23	44.23	44.25
2	47.25	44.25	44.25	44.25
3	47.27	44.23	44.23	44.23
4	46.25	44.23	44.23	44.23
5	47.25	44.23	44.23	44.23
6	46.23	44.23	44.23	44.22
7	46.23	44.25	44.25	44.25
8	45.27	44.27	44.27	44.27
Av	46.50	44.24	44.24	44.24

**Table 3.10 HPV11 E2 TAD T<sub>m</sub> values from final DSF buffer screen**

Sypro Orange dye was added to HPV11 E2 TAD in different buffers and the protein mixture was heated to 95 °C at 1 °C/min. Sypro Orange fluorescence was recorded at each incremental temperature increase and protein T<sub>m</sub> values were calculated from the resulting fluorescence curve. Results from 8 replicates and an average T<sub>m</sub> for each buffer are shown. Samples with a T<sub>m</sub> of >46 °C are highlighted in green and 44-46 °C in blue.

In this final screen of NMR compatible buffers, the HEPES and sodium phosphate buffers showed no increase in HPV11 E2 TAD T<sub>m</sub> compared to the Tris purification buffer, each

averaging 44 °C. The Tris buffer is known to stabilise HPV11 E2 TAD to reasonable levels, but the high concentration of NaCl is not ideal for long term NMR experiments. Identifying NMR compatible buffers which achieve comparable stability suggests that these conditions could potentially be used for 2D and 3D NMR experiments. The MOPS buffer showed an increase in HPV11 E2 TAD  $T_m$ , averaging 46.5 °C. The 8 replicates of each condition were consistent throughout the plate. As these 3 buffers were all capable of stabilising HPV11 E2 TAD to levels comparable to the high salt purification buffer in the short term DSF assay, the long term stability of HPV11 E2 TAD at room temperature was then assessed in these buffers using a hanging drop precipitation assay. Purified protein 200  $\mu$ M (1  $\mu$ L) was mixed with 1  $\mu$ L buffer and the drop was suspended over a 1 mL reservoir of buffer and sealed. The drops were allowed to sit at room temperature and were examined by light microscopy after 2 days and 1 week. Unfortunately, all the buffer combinations tested did not prevent protein precipitation from occurring under these conditions, making it unlikely that the protein would remain in solution to allow sufficient data acquisition for 3D NMR.

### 3.4 Discussion

The HPV E2 protein plays multiple functional roles in the viral life cycle. The TAD interacts with a range of viral and cellular proteins and many of these interactions are essential for virus replication, transcription and viral genome maintenance. As such, the E2 TAD is a valuable potential target for novel small molecule inhibitors to treat HPV infections [137, 139, 160, 161]. The use of small molecule fragments in biophysical screens to identify start points for further drug development is a relatively novel and highly powerful technique. Exploiting sensitive screening methods such as thermal shift, NMR and surface plasmon resonance (SPR) can identify fragments which bind to purified proteins and develop lead compounds for further assessment.

Fluorescence-based thermal shift assays (such as DSF) were initially developed as a tool for screening compound libraries for binding to protein targets [162]. The assay is relatively quick and inexpensive and utilises existing real time PCR technology in a high-throughput screen to rapidly assess large numbers of compounds. It has also been shown to be an effective predictor of the binding affinity of small molecules to proteins when compared to results generated using other biophysical screening techniques including isothermal calorimetry (ITC), dynamic light scattering (DLS), differential scanning calorimetry (DSC), circular dichroism (CD) spectroscopy and fluorescence polarisation along with functional enzyme assays [152, 162, 163]. It is known that the binding of specific ligands and non-specific co-factors can induce a conformational stabilisation of protein structures, and DSF exploits this link between ligand-binding and protein denaturation. Protein stability is associated with the Gibbs free energy of unfolding,  $\Delta G_U$ , a temperature-dependent measurement related to the enthalpy and entropy of the system ( $G = H - TS$ ).  $\Delta G_U$  decreases

as temperature increases, reaching an equilibrium state where half of the protein is unfolded at a temperature designated  $T_m$ . Although many proteins do not undergo a reversible two-state unfolding reaction, assuming that equilibrium thermodynamics apply allows information about ligand binding to be collected. The stabilising effects of ligands binding to the protein increase the  $\Delta G_U$  and can lead to an increase in the observed  $T_m$  of the protein by a combination of enthalpic and entropic effects along with binding affinity [135, 152].

To determine the best conditions for screening small molecules for protein binding by DSF, the conditions first require optimisation. If the buffer conditions are too stabilising, the small effects contributed by binding of low-affinity fragments are unlikely to be detected, however if the buffer conditions do not stabilise the protein sufficiently, the risk of protein aggregation or precipitation during the assay increases. A balance between these two extremes is required to optimise the possibility of detecting weak binding events of the fragments to the protein. To identify suitable conditions for screening small molecule fragments for binding to HPV11 E2 TAD by DSF, the fluorescence-based thermal shift assay was first utilised for the high-throughput screening of buffer conditions to identify optimal stabilising conditions for HPV11 E2 TAD [152, 157]. Following the same rationale as drug screening assays, a range of buffer conditions and additives can be readily screened to determine conditions supporting maximum protein stability, solubility and homogeneity as required for downstream experiments.

These experiments showed that the buffer choice played a major role in determining protein stability, as suggested by Norlund. Interestingly, initial tests comparing Tris and phosphate

buffers showed that protein concentration can affect stability differentially between buffer systems.

Further screening showed that buffer pH was critical to stability, with HPV11 E2 TAD displaying optimum stability in the range pH 7-8. This optimum pH range was expected based on published work on E2 [66] and the calculated pI, 6.14. Increasing the ionic strength of the buffer by adding NaCl also played a role in stability, suggesting that the electrostatic interactions within the protein are important [164]. The addition of DTT had a minor effect, suggesting that aggregation *via* disulphide bond formation is potentially a problem for HPV11 E2 TAD stability but overall this is less significant than the effects of other interactions. From these results, a buffer resulting in a mid-range protein  $T_m$  that was influenced by addition of the co-factors mentioned was selected to use in downstream fragment screening assays in the hope that small stabilising effects from the binding fragments wouldn't be masked by the buffer effects while maintaining sufficient protein stability throughout the assay to minimise problems with protein aggregation or insolubility.

Although the DSF assay has been used to good effect in screening compound libraries for binding to target proteins, the fragment library used here comprises much smaller molecules that are unlikely to have a high binding affinity for the protein resulting in much smaller effects on the  $T_m$ . As the variation in the  $T_m$  recorded in control samples between plates varied by up to 1 °C in my experiments (Average  $T_m$  = 38.01 °C; standard deviation = 0.31 °C), a threshold  $\Delta T_m$  of 2 °C was applied to increase the chances of removing effects due to sample variation [157]. Although this does risk missing weak binding effects by some fragments, any high-affinity binders would potentially induce a larger effect on protein

stability. The vast majority of the fragments screened had little effect on the  $T_m$  of HPV11 E2 TAD but this could be due to low affinity binding having little effect on the protein stability instead of necessarily indicating no binding interaction. A total of 322 fragments (63.4%) resulted in a  $T_m$  of within 1 °C of the DMSO control, while a further 97 fragments (19.1%) resulting in a  $T_m$  shift of between 1-2 °C relative to the DMSO control. While it is possible that some of these fragments were specifically binding to HPV11 E2 TAD and imposition of the threshold  $T_m$  shift of 2 °C results in fragment hits being missed, differentiating these from the fragments which do not bind and are simply showing an increase in  $T_m$  due to the observed variation in protein stability between wells is not reliable with this assay.

In multiple wells, the fluorescence profile obtained deviated from the expected sigmoidal curve associated with protein unfolding. In some wells, high initial fluorescence intensity was observed which could be caused by the inherent fluorescence activity of the fragment, direct interaction between the Sypro Orange dye used in the assay and the fragment, or protein aggregation or instability. Some wells displayed no transition in fluorescence intensity, which could be caused by protein instability, fragment-Sypro Orange interactions or binding of the compound to both the native and unfolded protein forms, thus lowering the  $\Delta T_m$ . In some cases multi-phasic dissociation curves were observed rather than the expected two-state curve, again suggesting protein instability or fragment aggregation or oligomerisation. Examples of the curve shapes are given in Figure 3.5. These fragments were removed from the assay results for analysis as any positive binding effects would be masked by the other issues mentioned [134, 135, 152, 162].

There is also a danger of false positive results when using a thermal shift assay to screen fragment libraries. Non-specific binding of a fragment to multiple sites on the protein could artificially increase the  $\Delta T_m$ , as could weak, entropically-driven binding to unfolded proteins. As such the identified hits cannot be ranked by binding affinity from this assay. Decay of compounds stored in DMSO can also be an issue. Therefore, a secondary assay is necessary to confirm the results from the thermal shift assay. Despite this potential problems however, DSF remains a powerful tool for library screening. The ability to rapidly screen large libraries of compounds with readily available technology requiring minimal purified protein is a huge advantage over traditional screening, and as such fluorescence based thermal shift assays are becoming established as a primary screen to rapidly identify initial hits for further assessment using more time- and resource-intensive structural assays like NMR, SPR and X-ray crystallography. Fluorescence based thermal shift assays have been successfully used in this manner to screen large fragment libraries for binding to proteins before using other screens to validate these hits and identify leads for further development [144, 163, 165, 166].

Of the 508 fragments screened, 14 caused an increase in HPV11 E2 TAD  $T_m$  above the imposed threshold and without displaying any obvious alterations in curve from any of the extraneous issues described above. From these, 10 fragments were selected for further validation on the basis of structural diversity, cellular cytotoxicity [158] and commercial availability. In addition, two fragments which had given altered fluorescence intensity profiles were also included to investigate whether the unusual fluorescence profiles were consistent in different buffer conditions.

Repeating the thermal shift assay with selected hits highlighted the inconsistency of the assay with small molecule fragments. 8 of the 10 fragments selected to retest in multiple buffers (highlighted in Table 3.3) had originally increased the  $T_m$  of HPV11 E2 TAD in EPPS buffer without adversely affecting the unfolding profile, but when this was repeated only two of the 10 fragments showed a consistent increase in  $T_m$  (1F5 and 2B6). These two fragments gave consistent shifts in EPPS and an even greater shift in the more stabilising MOPS and HEPES buffers. This suggests that the increase in protein stability caused by fragment binding and optimised buffer conditions can act additively, resulting in a cumulative increase in protein stability greater than that induced by either factor alone. In the tricine buffer which had previously been shown to stabilise HPV11 E2 TAD to the greatest extent (Table 3.2), there was still a positive  $T_m$  shift with these two fragments but this was much smaller. This could indicate an integral threshold for protein stability above which the small additional stabilising effects of fragment binding cannot be observed, or more likely given the small difference in HPV11 E2 TAD  $T_m$  observed between the HEPES, MOPS and tricine buffers (approximately 40 °C in HEPES and MOPS compared to approximately 41 °C in tricine), that the cumulative stabilising effects of fragment binding are influenced by the buffer conditions.

The two fragments which consistently resulted in a positive shift in  $T_m$  had both previously increased HPV11 E2 TAD  $T_m$  to >2 °C, but compared to other fragments the recorded shift in  $T_m$  was at the lower end of the range observed (see Table 3.3). In contrast, some of the fragments (4G7, 4H8, 6H6) which had previously resulted in very high positive  $T_m$  shifts when the library was screened using the DSF assay (Table 3.3) did not result in the same positive shifts when the assay was repeated (Table 3.5, Figure 3.9). It is possible that the very high

shifts previously observed with some fragments were not specifically due to fragment binding to HPV11 E2 TAD and could instead be due to variations in protein or fragment stability between wells. Alternatively this could be due to a direct interaction between the fragment and Sypro Orange, or due to fragment binding to both native and unfolded protein, but as the results are not consistent between screens this is unlikely to be the case. The  $T_m$  results obtained with all 10 fragments from the two screens are summarised in Table 3.11.

A further three fragments (2B3, 1A6 and 4G1) still resulted in a positive change in  $T_m$  in the MOPS and HEPES buffers despite not giving a consistent change in the EPPS buffer. This suggests that when incubated in more stabilising buffer conditions, the protein remains stable long enough for the fragments to bind and further stabilise the protein, resulting in a positive shift in  $T_m$  upon fragment binding. However there does appear to be a threshold for this effect since incubation in the most stabilising buffer resulted in a very small increase in protein stability with some fragments and no effect from others. Given the lack of consistency with the EPPS buffer between screens, it is possible that a greater level of initial protein stability and solubility is required for this assay and that in the less stabilising EPPS buffer a greater degree of protein variability occurs between wells leading to different results. The consistently higher results obtained in the more stabilising MOPS and HEPES buffers adds to this suggestion, indicating that there is a balance to be struck between solubilising the protein sufficiently for the assay and not stabilising it so far as to miss the small additional stabilities from binding fragments. Nonetheless, these experiments highlighted several fragments that have the potential to bind and stabilise HPV11 E2 TAD that were worthy of downstream validation (summarised in Table 3.11).

	EPPS screen 1	EPPS screen 2	50 mM HEPES pH 7.5	50 mM MOPS pH 7.5	100 mM tricine pH 8
ID	$\Delta T_m$	Av. $\Delta T_m$	Av. $\Delta T_m$	Av. $\Delta T_m$	Av. $\Delta T_m$
1A6	7.37	1.03	3.38	5.1	-2.02
1F5	2.90	2.45	2.37	2.54	1.75
2B3	14.1	-6.05	11.11	12.16	-1.65
2B6	3.4	3.15	4.00	4.05	1.59
4F1	2.93	-0.11	0.30	-0.1	-0.62
4G1	8.93	-0.21	5.64	6.14	42.17
4G7	6.78	-4.56	0.41	2.27	-0.46
4H8	10.7	0	-0.08	-0.04	0.20
5D4	2.53	-1.17	-1.05	0.11	-1.52
6H6	7.39	-0.13	0.19	0.01	-0.19

**Table 3.11 Summary of HPV11 E2 TAD  $T_m$  with selected fragments in different buffers**

NMR spectroscopy experiments based on the principle of NOE transfer were used to validate the fragment hits identified by the thermal shift assays. STD and wLOGSY experiments are both indirect assays to look at fragment binding to the protein relying on transfer of saturation to the bound ligands. Initially, 10 fragments were selected for further analysis by NMR based on performance in the thermal shift experiments discussed above. Of these, only 2 were identified as positive hits by NMR showing poor correlation between the two assays. This is not entirely unexpected, thermal shift based assays are known to have problems with reproducibility, and due to the weak nature of the fragment binding, fragments which do

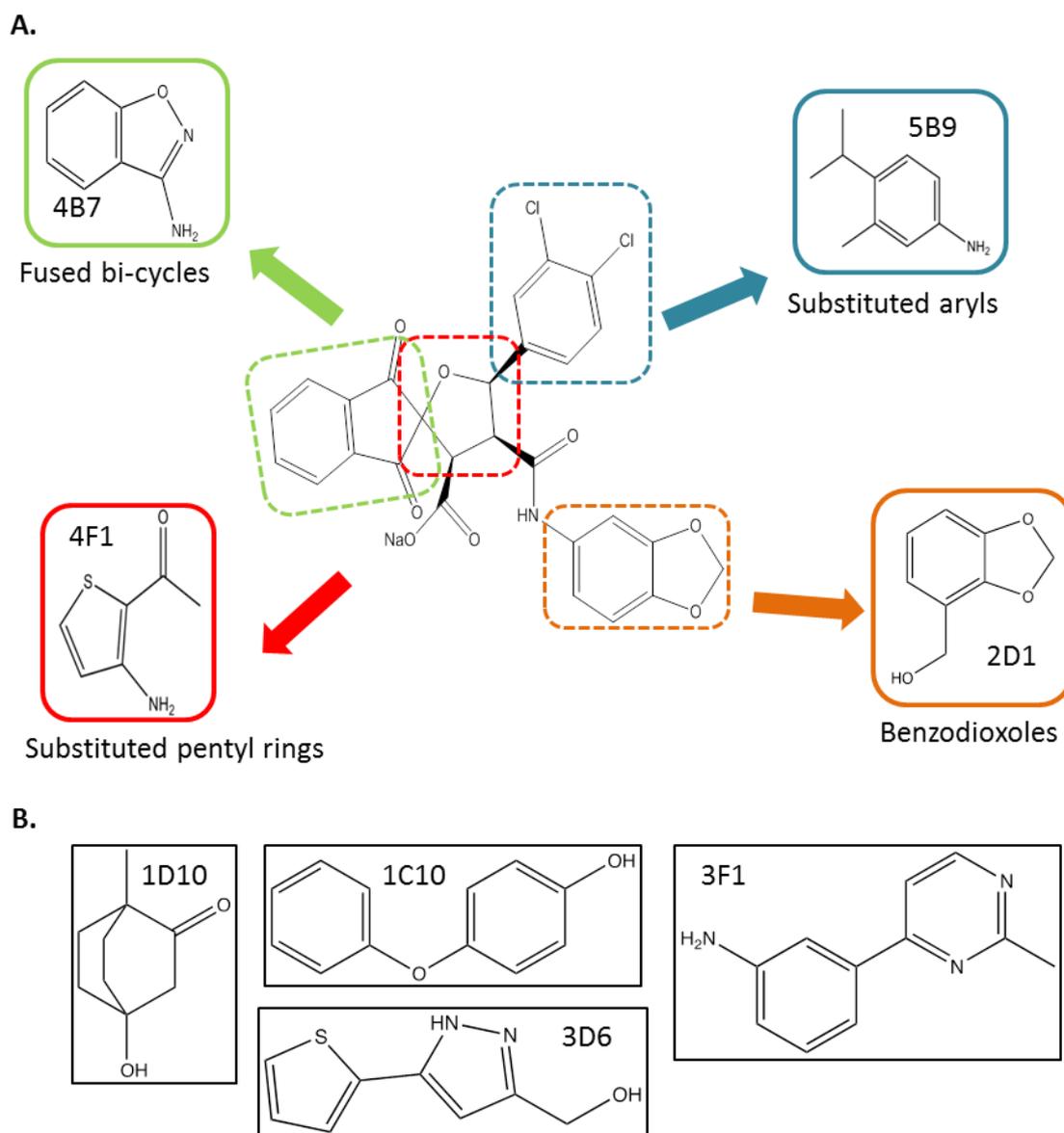
bind to the protein but which exert a weak effect on protein stability can be missed [144, 166].

Due to the poor correlation between the thermal shift experiments and the downstream NMR analysis, the entire library of 436 fragments was screened using two NMR-based assays, STD and wLOGSY. 29 fragments were identified as binding to HPV11 E2 TAD in both sets of NMR experiments (6.65%) with a further 11 identified in the wLOGSY experiments only. Both assays are based around the principle of transferred NOEs from binding events resulting in an altered fragment spectrum [167]. wLOGSY is the more sensitive of the two however, as bridging water molecules between protein and ligands and proton exchange events between water molecules at the protein-ligand interface and bound ligands with reactive protons allow more efficient transfer of irradiation to low affinity reversibly binding fragments [168-170]. It is therefore possible that low-affinity binding interactions which are missed by the STD-NMR assay can be identified by wLOGSY.

A hit rate of approximately 3-10% is expected when screening large numbers of fragments with these assays, [143, 144] unless a library is specifically designed to optimise binding to a particular target. Although the fragments that were identified as potential binders in the NMR experiments did not correlate with those identified in the thermal shift screens, the consistency between the two NMR assays employed provides evidence that NMR-based screening approaches are more reliable when used to detect low affinity binding events. Nonetheless, one fragment (4F1) was observed to interact with HPV11 E2 TAD in both the NMR experiments and the thermal shift experiments (DSF – Figure 3.9, Tables 3.3, 3.5; NMR Figure 3.14, Table 3.6). The other fragment (1A6), which had previously been shown to bind

by NMR and which also gave a consistently positive  $T_m$  shift was no longer identified as binding (DSF – Figure 3.9, Tables 3.3, 3.5; NMR Figure 3.14). This difference could be due to fragment degradation due to long term storage in DMSO with degradation products no longer binding to HPV11 E2 TAD. Alternatively, as the initial screen of fragment 1A6 with HPV11 E2 TAD was carried out with fragments in pairs and the second screen with cocktailed fragments, it is possible that 1A6 was binding to the same protein site as another, stronger binding fragment within the cocktail and therefore was out-competed for binding leading to a negative result.

The fragment hits were then assessed against the structure of a published small molecule inhibitor of the E1-E2 interaction to look for similar structural motifs indicative of binding to the same site on the protein.



**Figure 3.13 Comparison between structural motifs in known E2 inhibitor and fragment hits from NMR screen**

**A.** Structure of a molecule previously shown to bind HPV11 E2 [137, 140, 141] with key structural motifs highlighted. Examples of fragments identified as binding using NMR assays are shown with structural similarity to the previously characterised E2 inhibitor highlighted. **B.** Examples of fragment structures identified as binding to HPV11 E2 TAD using NMR assays which vary from the structural motifs contained in known E2 inhibitors.

Breaking down the structure of a published E2 interacting compound [140] into the component structural motifs allowed comparison between the fragment structures identified. Of the 30 hits, 20 were composed of structural moieties observed in the known inhibitor, divided into fused bicycles (5), substituted aryls (4), benzodioxoles (6) and substituted pentyls (5) (Figure 3.13). This suggests that it is possible that the identified fragment hits are binding into the same E2 pocket utilised by the published inhibitor, somewhat validating the assay results. Interestingly, a small subset of the hit fragments did not bear any structural similarity to the other groups, indicating that these groups may potentially be binding at alternative sites on the protein (Figure 3.13 B).

Using this analysis of structural motifs along with published cytotoxicity data [158] and commercial availability of the fragments, 10 of the fragments hits were selected and the thermal shift assay was repeating using the NMR buffer conditions. As buffer conditions were shown to influence protein stability dramatically, the observed effect of the fragments would also be expected to vary with buffer. Repeating the thermal shift assay in NMR buffer, the only fragment hit which had previously led to a positive  $T_m$  shift (4F1) now had little effect on protein stability. Conversely, fragment 5B9 which had shown little effect in previous assays now gave a positive shift in protein  $T_m$ . This again highlights the variability of the thermal shift assay as conditions change and brings into question its use as a screening tool unless variables such as buffer pH and composition are controlled for initially and not varied in down-stream screening. Although a fast technique requiring minimal purified protein, in this case the thermal shift assay appears an ineffective primary screening technique, although it would potentially still be of use for down-stream determination of the effects of fragment hits on the protein under carefully managed conditions.

Having identified multiple fragments covering a range of diverse structural motifs as binding to HPV11 E2 TAD by NMR (Figure 3.13), structural information about the binding site of these fragments on the protein is required for further downstream development. The NMR experiments utilised for screening work around the principle of transferred NOEs, and as such can only identify binding fragments without informing about the site of binding events. Traditionally, X-ray crystallography techniques have been employed to obtain this information. Protein crystals can be soaked with fragments prior to analysis, or combinations of protein and fragments can be co-crystallised and the resulting crystals analysed by X-ray diffraction. Depending on the quality of the crystal, this can provide valuable insight into the binding site of small molecules, however as this data is obtained with crystallised protein some detail can be lost. Attempts were made to co-crystallise HPV11 E2 TAD with the selected 10 fragment hits following published protocols [66], however in this case no co-crystals suitable for analysis by X-ray diffraction could be obtained. Alternatively, 2- and 3-dimensional NMR experiments can be used to look at small molecule binding to protein in solution.

2D NMR works by generating a spectrum looking at the signals corresponding to two different types of nuclei within a molecule. By synthesising  $^{15}\text{N}$  labelled protein and running NMR experiments calibrated to look at both  $^1\text{H}$  and  $^{15}\text{N}$  signals, a spectrum can be generated which shows each unique amino acid residue within the protein as a distinct point. Binding of a small molecule fragment to a protein would result in a shift in the peaks corresponding to amino acids within the binding site, allowing these to be mapped. However, for this data to be analysed the NMR spectrum of the protein must be assigned, with the individual peaks mapped to the corresponding amino acid within the sequence. This can be done by acquiring

3D NMR spectra, labelling protein with both  $^{15}\text{N}$  and  $^{13}\text{C}$  and looking at the signals from each  $^1\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$  nucleus in turn to match the spectrum to each amino acid residue in the protein sequence in turn. This is a complex process requiring a huge amount of data, and as such the data acquisition time is extensive. This means that the protein of interest must be soluble in an NMR-compatible buffer for long periods of time. Any degradation or precipitation of the protein during this time would impact negatively on data quality and prevent an accurate spectrum being obtained.

Given the observed problems with HPV11 E2 TAD stability in traditional NMR buffer, in order to carry out 3D NMR experiments and map the protein spectrum to allow downstream 2D NMR analysis of fragment binding sites, the buffer conditions used required further optimisation to maximise protein stability. Despite the drawbacks discussed with using thermal shift assays to screen for protein-binding fragments, this assay is able to provide valuable information about the effect of buffer composition on protein stability. As such it is an ideal technique for screening large sets of buffers to identify favourable conditions to allow for the long data acquisition required for structure determination by 3D NMR [134, 135, 152, 157].

For the initial buffer screen (Table 3.8), a set of buffers routinely used in NMR and X-ray crystallography experiments were tested alone and in combination with multiple co-factors known to stabilise some proteins. These co-factors included various concentrations of salts, detergents, glycerol and sucrose. As had previously been shown (Table 3.5), the buffers in the pH range 7-8 had the largest stabilising effect on HPV11 E2 TAD. The addition of high concentrations of NaCl increased the protein stability as had been previously shown, but a

much lower concentration of  $\text{MgCl}_2$  was able to induce a similar level of protein stability. This is advantageous for NMR as high concentrations of  $\text{Cl}^-$  ions within the buffer affect the quality of the acquired data. Sucrose, glycerol, TMAO and the amino acid salt arginine-glutamate (RE) were also shown to increase protein stability, increasing the  $T_m$  of HPV11 E2 TAD by approximately 2-3 degrees in each of the pH 7-8 buffers tested (Table 3.8). From this screen the 4 buffers shown to have the greatest stabilising effect on HPV11 E2 TAD (sodium phosphate pH 7, MOPS pH 7, HEPES pH 7.5 and Tris pH8) along with the co-factors also shown to positively affect protein stability ( $\text{NaCl}$ ,  $\text{MgCl}_2$ , sucrose and RE) were retested in different combinations over 4 screens (screens 1 and 4 summarised in Tables 3.8 and 3.9; screens 2 and 3 detailed in Appendix Tables A.2-5;). Although glycerol and TMAO had also been shown to increase the protein stability in the initial screen, these buffer additives were not carried forwards for downstream screening. Addition of glycerol to the NMR buffer increases the viscosity which can have a negative effect on data acquisition. TMAO is a small molecule which can readily decompose to form trimethylamine, a nucleophilic compound which can readily interact with other small molecules and could potential affect the resulting spectra obtained in fragment-binding assays.

Following four screens, the four buffers shown to increase HPV11 E2 TAD stability to the greatest extent were retested in replicates of 8 (Table 3.10). Of these four buffers, the most stabilising was found to be MOPS pH 7 containing a combination of  $\text{MgCl}_2$  and sucrose. This buffer resulted in an increase in HPV11 E2 TAD  $T_m$  of approximately 6 °C compared to protein in MOPS buffer alone, and of approximately 4 °C compared to the normal NMR buffer, sodium phosphate pH 7.5. The  $T_m$  values obtained were consistent for all buffers

across multiple repeats, demonstrating that the thermal shift assay is a viable tool to use for the rapid assessment of buffer conditions to determine optimum protein stability.

The observed increase in protein  $T_m$  in these our stabilising conditions was promising, however the thermal shift assay is a short term assay looking at the thermal stability of a protein as the temperature is artificially increased over a fixed range. For long-term data acquisition NMR experiments, the protein is required to remain in homogenous solution for up to two weeks at room temperature. To test the ability of these optimised buffers to stabilise HPV11 E2 TAD at high concentrations (5 mg/mL) at room temperature over a long time period, hanging drop precipitation assays were carried out. This operates on a similar principle to protein crystallography – a small volume of concentrated protein solution is suspended over a well containing a reservoir of the buffer. The well is sealed to prevent evaporation and the protein is incubated at room temperature and examined microscopically at regular intervals to monitor protein precipitation. Ideally, no protein precipitation would be observed under these conditions over a two week period for a buffer to be considered suitable for use in downstream 2D and 3D NMR experiments. However, despite the high thermal stability of HPV11 E2 TAD exhibited in the optimised buffer conditions in thermal shift assays, a high degree of protein precipitation was observed in all conditions after less than a week at room temperature. This suggests that while thermal shift is a valuable tool for examining the thermal stability of proteins, additional assessment is required to determine the long term effects of the conditions on protein stability. The NMR-compatible buffers tested were not able to prevent E2 precipitation in the long term and therefore further analysis of fragment binding to HPV11 E2 TAD was not pursued.

Despite the inability to find a buffer which stabilised HPV11 E2 TAD sufficiently to allow structural NMR analysis of the fragment binding sites to be investigated, a lot of information about the thermal stability and solubility of HPV11 E2 TAD was obtained and this will be of use for future work. Additionally, both thermal shift and NMR-based fragment screens have been optimised to identify fragments which bind to HPV11 E2 TAD, and the data obtained from these screens can potentially be used at a later date to continue to develop novel small molecule inhibitors of E2.

One of the initial aims behind this work was to identify novel small molecule fragments which bind to HPV11 E2 TAD and could potentially inhibit the E2-ChIR1 protein-protein interaction. As further development of these identified binding fragments could not be continued using structural NMR to investigate the fragment binding sites, an alternative approach was required. Investigating the E2-ChIR1 protein interaction in depth would provide valuable information which could then later be applied to the fragment screening assay protocols discussed here to identify fragments which specifically interact at the site of the E2-ChIR1 interaction and act to inhibit this. As such, the remainder of the work presented here (Chapters 4 and 5) describes a detailed analysis of the E2-ChIR1 interaction.

## **Chapter 4. Investigating the interaction between HPV16 E2 and ChR1**

### **4.1 Introduction**

The papillomavirus E2 protein plays an integral role in multiple processes that are essential for the virus life cycle. The two distinct domains of E2, the C-terminal DNA-binding domain (DBD) and the N-terminal transactivation domain (TAD) act in concert to facilitate processes such as viral genome replication and regulation of viral transcription. The DBD binds to the viral genome through sequence specific binding sites within the LCR, while the TAD can interact with multiple viral and cellular proteins. Disrupting either the E2-DNA or E2-protein interactions could negatively impact on the viral life cycle.

Viral genome replication is dependent on an interaction between the E2 TAD and the viral DNA helicase E1. While the E2 DBD binds to high affinity binding sites proximal to the viral origin of replication (Ori) [171], the E2 TAD interacts with E1, effectively loading the E1 helicase onto the Ori and initiating viral genome replication [68]. Mutation of a highly conserved amino acid residue located on the E1-binding surface, glutamic acid 39 to alanine (E39A), has been shown to abrogate this interaction and effectively inhibit viral genome replication while leaving the transcription activation function of E2 intact, thus demonstrating that the replication and transcription functions of E2 are separable [172, 173].

Similarly, the ability of E2 to activate transcription can be abrogated by point mutations within the E2 TAD which disrupt essential protein-protein interactions. E2 acts to regulate

the transcription of the viral oncoproteins E6 and E7 from the early promoter within the viral LCR through a negative feedback loop. Binding of the E2 DBD to high affinity binding sites within the LCR allows the E2 TAD to recruit cellular factors required for transcription. When high levels of E2 protein are present, the E2 DBD can bind to lower affinity binding sites which overlap with the binding sites of necessary transcription factors such as SP1, displacing them from the LCR and acting to repress viral transcription [88, 89]. The interaction between the E2 TAD and the bromodomain containing protein Brd4 has been shown to be essential for E2-dependent transcription activation [73, 75, 76, 174]. A mutation of the E2 TAD amino acid isoleucine 73 to alanine (I73A) and to a lesser extent mutation of arginine 37 to alanine (R37A) abrogates the interaction between E2 and Brd4, resulting in a protein which retains the ability to facilitate viral genome replication but which is no longer able to activate viral transcription [76].

E2-dependent tethering of the viral genomes to cellular chromatin to facilitate the partitioning of replicated genomes into the nucleus of newly formed daughter cells during mitosis also takes advantage of the discrete interactions formed by the E2 DBD and TAD. While the E2 DBD binds to the viral genome through the LCR, the E2 TAD interacts with cellular chromatin associated proteins, effectively tethering the viral genomes to cellular chromosomes throughout mitosis. The DNA helicase ChR1 has been proposed as a cellular partner of E2 involved in this tethering process [95]. ChR1 is known to associate with mitotic chromosomes and has been shown to be required for cohesion establishment [124]. Both ChR1 and the yeast homolog CHL1 were identified as BPV-1 E2 interacting proteins [95] and further investigation showed that E2 and ChR1 colocalise on chromatin early in mitosis, but that co-localisation is lost as cell progress through mitosis, indicating that ChR1 is important

for recruitment of E2 to chromatin, but not for tethering E2 as cells progress through mitosis. Fluorescence resonance energy transfer (FRET) experiments yielded further information, showing that the interaction is dynamic throughout the cell cycle and indicating that a direct protein-protein interaction is formed rather than E2 and ChIR1 associating through a larger complex [133]. Mapping of the ChIR1 binding site on E2 interaction indicated that two regions of BPV-1 E2 TAD contained ChIR1 binding sites; amino acids 50-74 and 113-134 [95]. Further investigation led to identification of a point mutation in BPV-1 E2 TAD, tryptophan 130 to arginine (W130R), which abrogated the interaction with ChIR1. While E2 W130R was shown to be able to stimulate E2-dependent transcription and replication, this mutant was unable to associate with mitotic chromosomes, and resulted in an impaired ability to maintain episomal genomes in transfected cells. Together these data suggest a role for the E2-ChIR1 interaction in the tethering of viral genomes to mitotic chromosomes. If this interaction is conserved across papillomavirus types, an analogous mutation in HPV16 E2 could potentially result in a similar phenotype. As the BPV-1 model system is not ideal for analysis of the life cycle of HPV as the mouse fibroblast cells used cannot be induced to differentiate in the lab, identification of a point mutation in HPV16 E2 analogous to the BPV-1 E2 W130R mutant would allow the role of the E2-ChIR1 interaction within the viral life cycle to be fully investigated.

## 4.2 Aims

Building on the published work looking at the interaction between BPV-1 E2 and ChIR1 [95, 96, 133], the aims of this work were to investigate the interaction between HPV16 E2 and ChIR1 and to:

1. Map the binding site of HPV16 E2 within a unique domain of ChIR1
2. Conversely map the ChIR1 binding site within HPV16 E2 TAD using a series of E2 truncations
3. Use site-directed mutagenesis to generate a mutant of HPV16 E2 which exhibits a reduced affinity for ChIR1 to allow the role of the HPV16 E2-ChIR1 interaction to be investigated in the context of the viral life cycle.

## 4.3 Results

### 4.3.1 Investigating the E2 binding site on ChIR1

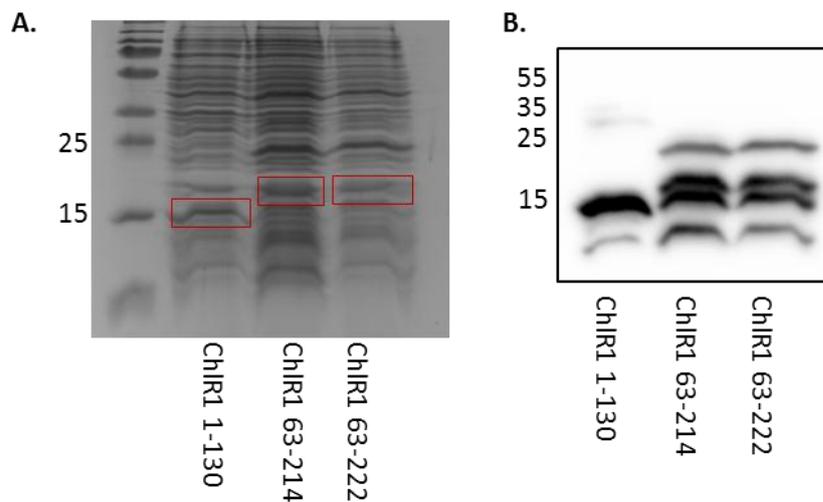
ChIR1 and the *S. cerevisiae* homolog CHL1 have been shown to bind to the N-terminal domain of BPV-1 E2 [95]. However, E2 is not known to interact with any of the other SF2 DNA helicases such as FancJ or XPD which are structurally related to ChIR1. Sequence alignment of human ChIR1 and ChIR2 and the yeast homolog CHL1 (identified from *S. cerevisiae* and *S. pombe*) with the related iron-sulphur cluster superfamily 2 DNA helicase XPD was therefore analysed to determine if there were differences between the sequences. This revealed a unique domain located at the N-terminus of ChIR1 and CHL1 between helicase domains I and Ia (Figure 4.1, full sequence alignment included in Appendix A.6). As this unique domain of ChIR1 is not contained in the related XPD helicase that is not thought to bind E2, it was postulated that this domain is important for the E2-ChIR1 interaction.

	I			
hChr1R1	1	MANETQKVGAIHFFPFFPYSIQEDFMAELYRVLEAGK-IGIFESPTGTGKSLSLICGALSWLDRFEQKKREEEARLLEGTGTGPLHDEKDESLCLSSCE	99	
hChr1R2	1	MANETQKVGAIHFFPFFPYSIQEDFMAELYRVLEAGK-IGIFESPTGTGKSLSLICGALSWLDRFEQKKREEEARLLEGTGTGPLHDEKDESLCLSSCE	99	
Spombe	1	MCHSKEVKFKTNFHHFYPYDIQLEFMRSLYSSISDGK-IGIFESPTGTGKSLSLICASLTWLDEHG-----GVLLEDNEK-----SNDNKSNTSSKIP	88	
Scerev	1	---MDKKEYSETFYHPYKPYDIQVQLMETVYRVLSGKKAIAIESPTGTGKTLSLICATMTWRMKN-----ADIFTRMETNIKTNEDDSENLSDDPEP	90	
TacXPD	1	-----MOKSYG-----VAIESPTGSGKTIIMALKSALOYSSERK-----	33	
		*****.***. . . .		
hChr1R1	100	GAAGTPRPAEPWVTQFVQKKEERDLVDRLEKAEQARRKQREERLQQLQHRVQLKYAAKRLRQEE-----EERENLLRLSREMLETG	181	
hChr1R2	100	GAAGTPRPAEPWVTQFVQKKEERDLVDRLEKAEQARRKQREERLQQLQHRVQLKYAAKRLRQEE-----EERENLLRLSREMLETG	181	
Spombe	89	DWLEQDLKIQKDLVKETHARLEQR--LEEIRKNQSRKNQMSNSTTYHRETKRRNINAEASTS-----DNCNNSNTSVDPMDEYL	168	
Scerev	91	DWVIDTYRKSVLQEKVDDLNDYKHH--LNEINTTSCQKLTMCDDLKHEGRYKSVDPRLKRRKRGARHLDVDLSEEQDFIPRPYESDSENNDTSKSTRGGRI	188	
TacXPD	34	-----	33	
		1 a		
hChr1R1	182	PEARLEQLESGEEELVLAIEYEDSEKKNVASRVEDDEDDLEEEHITKIYCSRTHSQLAQFVHEVKK-----SFFGKDVRLLVLSGRQNLQVNV---	269	
hChr1R2	182	PEARLEQLESGEEELVLAIEYEDSEKKNVASRVEDDEDDLEEEHITKIYCSRTHSQLAQFVHEVKK-----SFFGKDVRLLVLSGRQNLQVNV---	269	
Spombe	169	VTAEYTMPSTSEQSEDLFN--NGYSSKVSSELLRKLSPDNEKPPIVQKIYFTSRTHSQLOQLVQEIKK-----LNNQTFSTPIRVVSLASRKNLQVNV---	257	
Scerev	189	SDKDYKLSSEINSQIITILLDKIDGKVSRRDPNNGDRFDVTNQNP---VKIYVASRTYSQLGQFTSQLLRPSFFSSFRKVPDEKVKYLPKAKKQQLQVNV---	282	
TacXPD	34	-----LKVLVLRVTRNSQEEQVIKELRS-----LSSTMKIRAIEMQGRVNMCIILYRM	79	
		*. . . * * * * *		
		. . . . .		

Figure 4.1 Chr1 sequence alignment with XPD shows a unique domain at N-terminus

**Alignment of the N-terminal sequences of human ChIR1 and ChIR2, and the yeast homolog CHL1 with XPD. Helicase domains I and Ia are highlighted in blue. Cysteine residue included in the iron-sulphur cluster is highlighted in yellow.**

Three different truncations which contain all or part of the unique, N-terminal domain of human ChIR1 were ligated into a protein expression vector downstream of a tobacco etch virus (TEV)-cleavable N-terminal His-tag within the Parish laboratory by Laura McFarlane-Majeed. These encoded amino acids 1-130, 63-214 and 63-222 of human ChIR1. These regions of ChIR1 were selected for expression on the basis of molecular modelling studies which indicated that amino acids 1-130 contain helicase domain I and inaccessible amino acids which are unlikely to bind E2, while amino acids 63-222 contain the unique ChIR1 domain which is predicted to protrude from the core helicase domains. The shorter region encoding amino acids 63-214 was also cloned into the expression vector because amino acids 214-222 are highly acidic and it was thought that inclusion of this peptide could destabilise the truncated protein. Following confirmation of the cloning by sequencing, these expression constructs were used to transform *E.coli* BL21 cells and protein was over-expressed by isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) induction (protein expression, lysis and purification conditions are detailed in Materials and Methods). Bacterial pellets were lysed and His-ChIR1 proteins were isolated by affinity purification with His-affinity nickel-agarose resin (Figure 4.2).



**Figure 4.2 Expression and purification of ChIR1 fragments**

BL21 cells were transformed with His-ChIR1 fragments 1-130, 63-214 and 63-222. Protein expression was induced by addition of IPTG and bacterial cultures were incubated at 37 °C for 4 hours. Bacterial pellets were resuspended in lysis buffer and lysed by sonication. Samples of lysates were analysed by SDS-PAGE and Coomassie staining. B. Bacterial lysates were incubated with nickel-agarose resin at 4 °C for 2 hours. Unbound lysate was removed and the resin washed 3 times to remove non-specific binding proteins. A sample of the bound proteins were analysed by SDS PAGE and western blotting with an anti-His antibody.

All three fragments of ChIR1 were expressed and isolated on nickel-agarose resin. ChIR1 1-130 ran as a single band at the expected molecular weight of approximately 15 kDa, while ChIR1 63-214 and 63-222 were observed as doublets running at ~16 and 17 kDa respectively. Despite attempts to optimise the purification of these proteins, it has not been possible to separate these two close-running bands. With the similar molecular weight and inability to

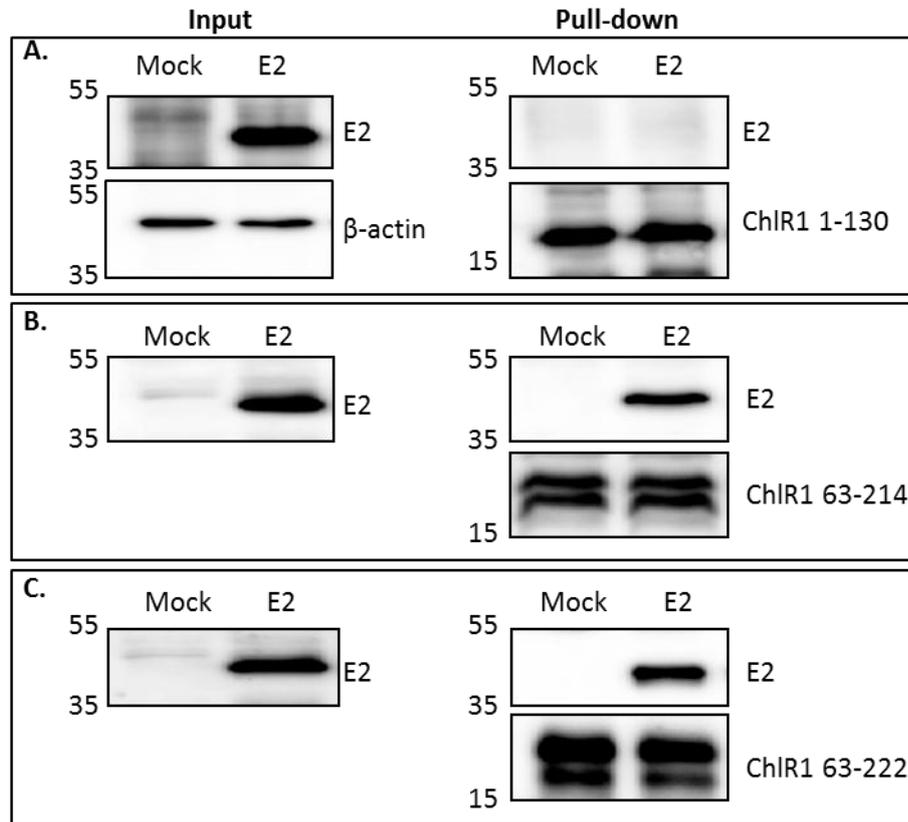
separate these proteins by chromatography, it was assumed that the lower band is a cleavage product derived from degradation of the ChIR1 protein (upper band).

To investigate the interaction of these ChIR1 fragments with HPV16 E2, the ChIR1 fragments immobilised on Nickel-agarose resin were incubated with lysate of C33a cells transfected with HPV16 E2. Resin was washed to remove unbound lysate and the bound proteins were analysed by western blotting for E2 and ChIR1 (Figure 4.3).

Using this *in vitro* pulldown assay, HPV16 E2 was found to associate with ChIR1 63-214 and 63-222 but not with ChIR1 1-130. This narrows the E2 binding site on ChIR1 to a 151 amino acid region within the unique domain of ChIR1 located at the N-terminus between helicase domains I and Ia.

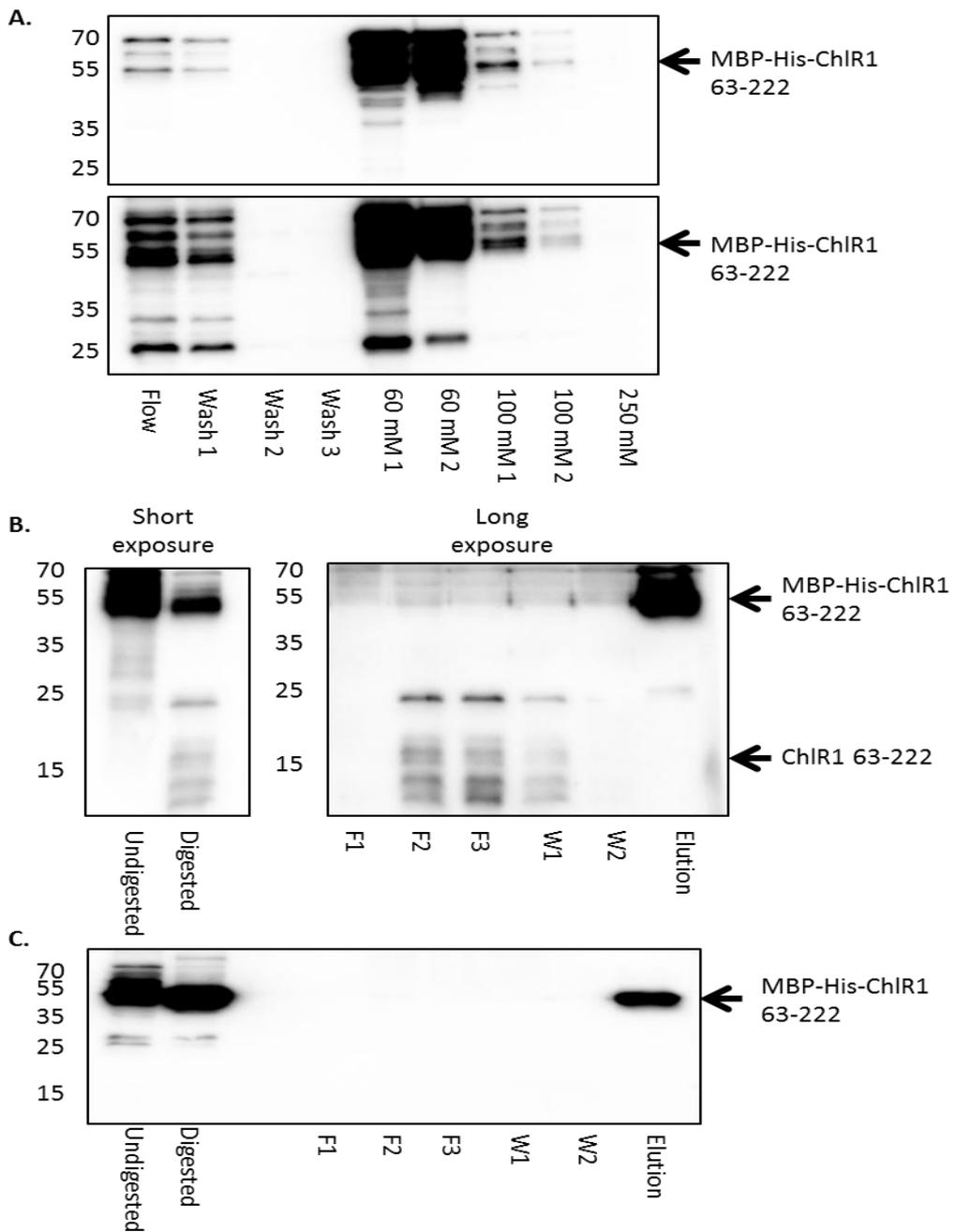
To determine if there was a direct interaction between ChIR1 63-222 and HPV16 E2, the pulldown assay described above was repeated using purified His-HPV16 E2 TAD in the place of HPV16 E2 expressing cell lysate. The ChIR1 fragment containing amino acids 63-222 had previously been cloned into a maltose-binding protein (MBP)-His-tag vector containing a TEV-cleavage site between the histidine tag and the multiple cloning site to allow the MBP-His tag to be removed by cleavage with TEV-protease (cloning carried out by Laura McFarlane-Majeed, Parish lab). MBP-His-ChIR1 63-222 was over-expressed by IPTG induction of bacterial culture and the tagged protein was purified by chromatography with a histidine-affinity column. Bound proteins were eluted with increasing imidazole concentration. The eluted ChIR1 proteins were dialysed to remove the imidazole and the MBP-His-tag was removed by cleavage with TEV-protease. The cleaved tag was then removed from the protein by affinity chromatography; the cleaved ChIR1 fragments are no longer able to bind

to the histidine-affinity column and were therefore isolated in the unbound fraction. Fractions were analysed by SDS-PAGE and western blotting with ChR1- and His-specific antibodies to identify fractions containing untagged ChR1 63-222 (Figure 4.4).



**Figure 4.3 HPV16 E2 binds to ChR1 within region 63-214.**

C33a cells were transfected with salmon sperm DNA (mock) or HPV16 E2. A sample of cell lysate was reserved for analysis and the remainder was incubated with His-ChR1 fragments immobilised on nickel-agarose resin. The unbound lysate was removed and the resin was washed to remove non-specific binding. The bound proteins were analysed by western blotting with HPV16 E2 and His (ChR1) specific antibodies. A. Input (left) and pulled down E2 (right) with ChR1 1-130, B. Input (left) and pulled down E2 (right) with ChR1 63-214 and C. Input (left) and pulled down E2 (right) with ChR1 63-222.



**Figure 4.4 Purification of TEV cleaved ChIR1 63-222**

**A.** MBP-His-ChIR1 63-222 transformed BL21 were grown in culture and protein over-expression was induced by addition of IPTG followed by incubation with shaking at 37 °C for 4 hours. Bacterial pellets were lysed and MBP-His-ChIR1 63-222 was purified by affinity chromatography with a histidine-affinity column, eluting with increasing imidazole.

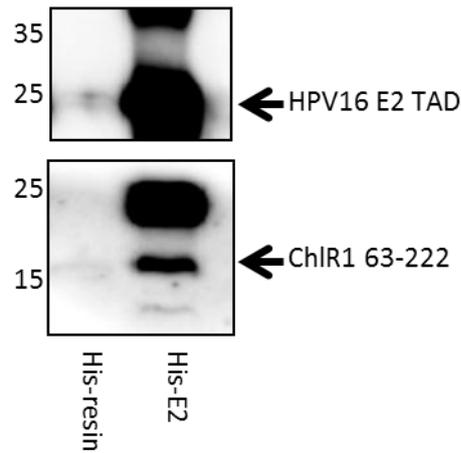
Collected fractions were analysed by western blotting with His- (top) and ChIR1- (bottom) specific antibodies. The imidazole was then removed from the first fraction eluted with 100 mM imidazole (labelled 100 mM 1) by buffer-exchange using a spin-concentrator with a 5000 Da molecular weight cut-off and the resulting protein was concentrated to 1 mL. The concentrated protein was incubated with His-TEV protease (3 µg TEV protease added to 400 µg MBP-His-ChIR1 63-222 protein) at 4 °C for 16 hours. B, C. The digested protein was purified by chromatography with a His-affinity column. Flow through fractions (F1-3) were collected followed by 2 wash fractions (W1 and W2). Bound proteins were then eluted with 250 mM imidazole buffer. Samples of the undigested and digested protein along with each collected fraction were analysed by SDS PAGE and western blotting with B. ChIR1-specific antibody and C. His-specific antibody.

Western blotting the fractions obtained from the initial purification of MBP-His-ChIR1 63-222 found that the highest concentration of relatively clean protein was obtained from the first 100 mM imidazole elution, with the expected band observed at ~55 kDa (Figure 4.4 A). Although multiple bands were observed in each elution fraction, the obtained MBP-His-ChIR1 protein was sufficiently isolated to allow TEV-cleavage of the tag followed by further rounds of purification to obtain the ChIR1 protein alone for downstream experiments.

Following TEV-cleavage of the MBP-His tag and removal of the tag by affinity chromatography, the cleaved ChIR1 63-222 protein was observed as a band at ~17 kDa in the flow through fractions when western blotted with a ChIR1 antibody (Figure 4.4 B). As previously observed (Figure 4.2), multiple bands were detected when western blotting for ChIR1, indicating potential degradation of the cleaved ChIR1 protein resulting in multiple

smaller products. The cleavage was not complete, with a large amount of ChIR1 remaining tagged and therefore binding to the column until elution with 250 mM imidazole, however western-blotting with Histidine antibody did not detect any protein in the flow through fractions, confirming that the His-tag had been successfully removed from the protein in these fractions (Figure 4.4 B and C).

To determine if the interaction previously observed between ChIR1 63-222 and HPV16 E2 is a direct, protein-protein interaction, His-HPV16 E2 TAD was over-expressed by IPTG induction of *E.coli* BL21 transformed with expression construct and incubated at 25 °C for 16 hours. Bacterial pellets were lysed by resuspension in lysis buffer containing lysozyme followed by sonication. Bacterial lysate was incubated with nickel-agarose resin at 4 °C for 2 hours before unbound lysate was removed and the resin washed to remove non-specific binding proteins. Purified TEV-cleaved ChIR1 63-222 was added to the resin-bound His-HPV16 E2 TAD and incubated at 4 °C for 16 hours. The unbound ChIR1 63-222 was then removed and the resin washed before bound proteins were analysed by SDS PAGE and western blotting for HPV16 E2 and ChIR1 (Figure 4.5).



**Figure 4.5 Direct interaction between purified HPV16 E2 TAD and ChIR1 fragment 63-222**

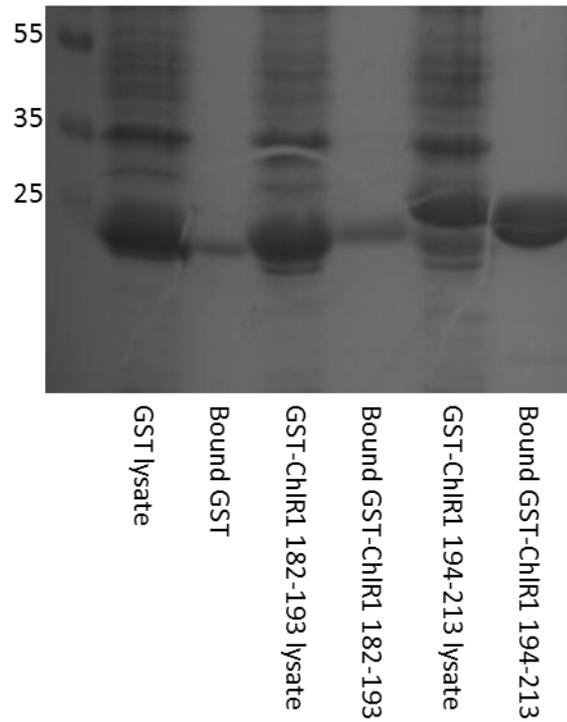
Expression of HPV16 E2 TAD in BL21 *E. coli* was induced by addition of 0.5 mM IPTG and incubation with shaking at 25 °C for 16 hours. The bacterial cells were then lysed and the lysate incubated with nickel-affinity resin for 2 hours at 4 °C. Unbound lysate was removed and the resin was washed three times. Purified ChIR1 63-222 (Figure 4.4 B, lane F2) was then added to nickel-agarose resin alone or resin with bound HPV16 E2 TAD and was incubated at 4 °C for a further 16 hours. Unbound ChIR1 63-222 was removed and the resin washed to remove non-specific binding proteins. The protein bound to the resin was analysed by SDS-PAGE and western blotting with HPV16 E2 and ChIR1 specific antibodies. No Coomassie gel of the HPV16 E2 TAD protein is included as no bands were visible following staining.

The pulldown of purified ChIR1 63-222 with purified His-HPV16 E2 TAD showed a direct interaction between the two proteins. This is the first evidence showing that E2 directly binds to ChIR1 rather than the interaction being mediated by other proteins as part of a

larger complex. Furthermore, these experiments provide evidence that E2 associates with the unique domain of ChIR1 between amino acids 130-214.

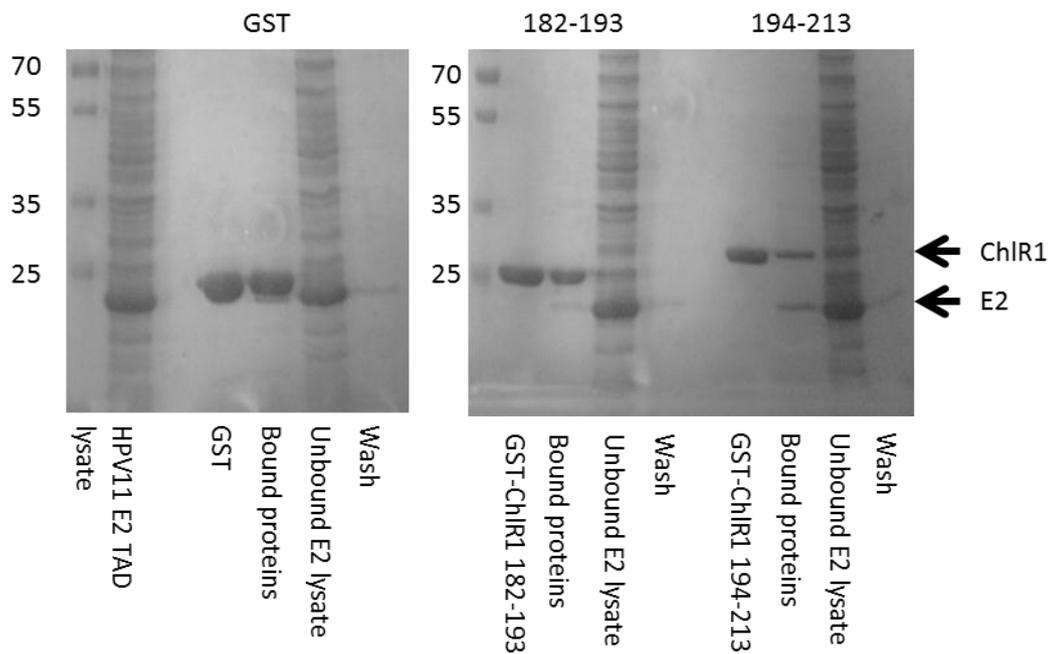
To further investigate the E2 binding site on ChIR1, two peptides of the E2 binding region predicted by modelling to form helical structures, covering amino acids 182-193 and 194-213 were selected. These peptides were cloned into a pGEX-4T-2 expression vector by ligation of annealed primer pairs that encode the peptide sequences into the digested vector. Positive clones were confirmed by sequencing. BL21 *E. coli* were then transformed with each construct in turn and the ChIR1 peptides were expressed as GST-fusion proteins by induction with IPTG for 4 hours at 37 °C. Following induction of expression, bacteria were lysed and the GST-peptides were isolated by incubation of the lysate with glutathione-agarose (GA) resin (Figure 4.6). The GST-tag can recognise and bind with the GA, allowing the tagged proteins to be immobilised on the resin and removed from the bulk lysate.

To investigate binding of the ChIR1 peptides to E2, the GA-resin immobilised peptides were incubated with HPV11 E2 TAD bacterial lysate. The expression and purification of HPV11 E2 TAD is discussed in Chapter 3 (Figure 3.3). Although the previous work investigating the E2 binding site on ChIR1 was carried out using HPV16 E2, HPV11 E2 was instead used here. Unlike HPV16 E2 TAD, HPV11 E2 TAD is well expressed, stable and readily purified from BL21 and for this reason was used in this experiment. Following incubation of GA-resin bound GST-ChIR1 182-193 or 194-213 with HPV 11 E2 TAD, the unbound proteins were removed by rigorous washing and the bound proteins were analysed by SDS PAGE followed by Coomassie staining (Figure 4.7).



**Figure 4.6 Expression and purification of GST-ChIR1 peptides**

BL21 were transformed with GST and GST-ChIR1 182-193 and 194-213 expression plasmids in turn. Protein expression was induced by addition of 1 mM IPTG and shaking at 37 °C for 4 hours. Bacteria were pelleted by centrifugation and lysed by sonication. Lysates were added to GA-resin and incubated for 2 hours at 4 °C before unbound proteins were removed by repeated washing. Samples of the protein lysate and the resin-bound proteins were analysed by SDS PAGE and Coomassie staining.



**Figure 4.7 HPV11 E2 TAD binds to GST-ChIR1 peptide 194-213**

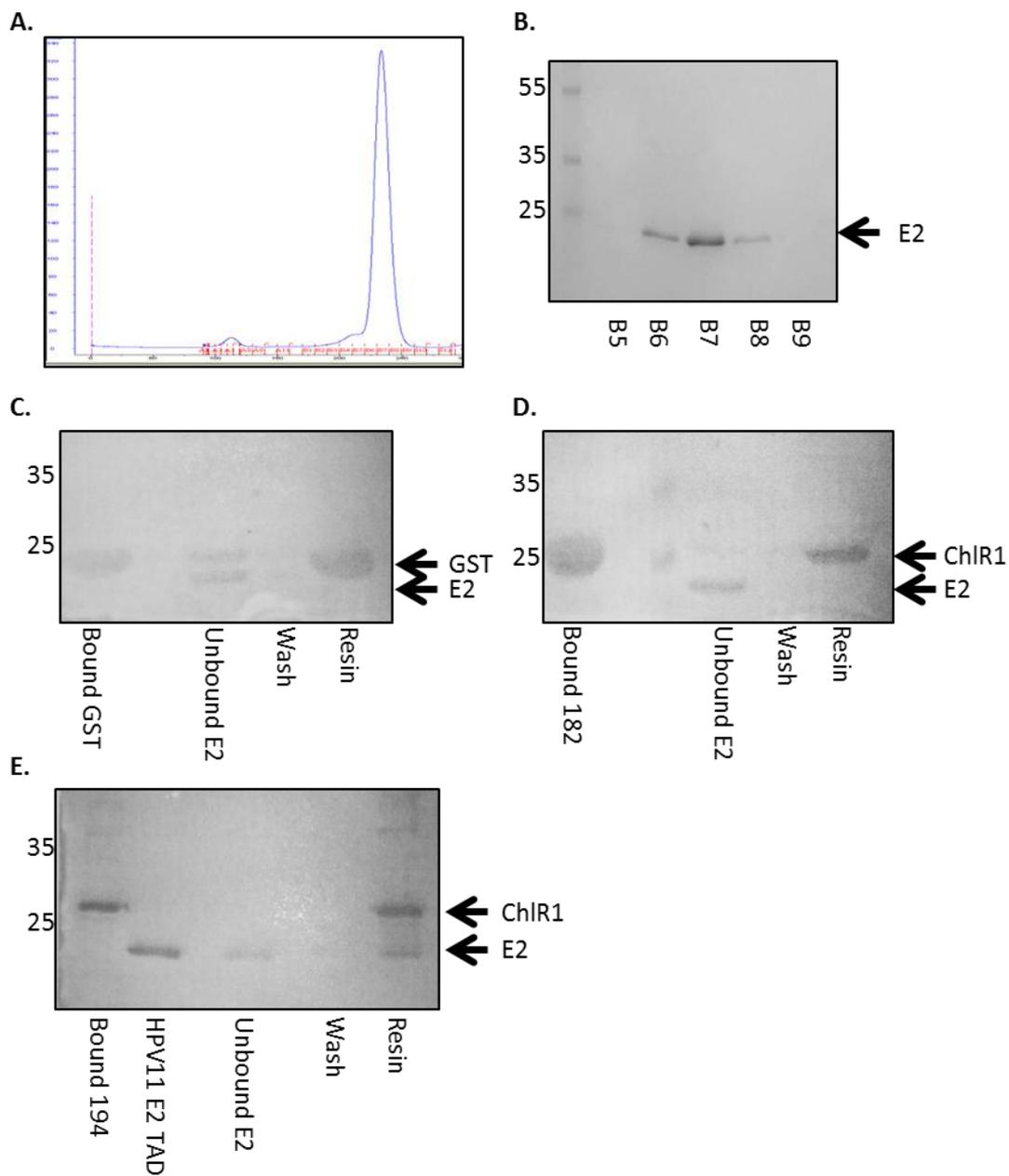
Expression of HPV11 E2 TAD was induced by addition of 1 mM IPTG and incubation at 25 °C for 16 hours. Bacterial cells were pelleted by centrifugation and lysed by sonication. HPV11 E2 TAD lysate was added to GST or GST-ChIR1 peptides bound to GA-resin and incubated for 4 hours at 4 °C. Unbound proteins were removed by rigorous washing and samples of the GA-resin before (GST/GST-ChIR1) and after incubation (bound proteins), the unbound protein lysate and the wash fraction were analysed by SDS-PAGE and Coomassie staining.

When HPV11 E2 TAD lysate was incubated with GST or GST-ChIR1 182-193 immobilised on GA resin, very little E2 protein was observed to bind, however more bound E2 protein was detected when incubated with immobilised GST-ChIR1 194-213. This suggests that the major binding site of HPV11 E2 TAD is within the amino acid region 194-213 of ChIR1, although binding within the amino acid region 182-193 is also observed to a lesser extent (Figure 4.7).

Minimal binding of HPV11 E2 TAD to GST was also observed, indicating that the assay conditions used have not been fully optimised, and therefore that the binding to ChR1 182-193 could be a non-specific interaction of HPV11 E2 TAD with the resin or GST tag.

As HPV11 E2 TAD can be readily purified from BL21 lysate by affinity chromatography (see Figure 3.3, Figure 4.8 A, B), the pull-down was repeated as before with purified protein. GA-resin bound GST-ChR1 peptides were incubated with purified HPV11 E2 TAD. Unbound protein was again removed by repeated washing and the bound proteins were analysed by SDS PAGE and Coomassie staining (Figure 4.8).

As had been previously demonstrated (Chapter 3), HPV11 E2 TAD was purified to >95% purity as determined by SDS PAGE and Coomassie staining by chromatography (Figure 4.8 A, B). Incubation of purified His-HPV11 E2 TAD with GA-resin bound GST or GST-ChR1 182-193 resulted in minimal binding of HPV11 E2 TAD (Figure 4.8 C and D). However, a protein band of approximately 23 kDa, which matches that observed for input purified HPV11 E2 TAD was pulled down with GST-ChR1 194-213. This suggests that amino acids 194-213 of ChR1 are sufficient for binding to HPV11 E2 TAD, providing evidence that the E2 binding site is located within this 19 amino acid region. The lack of binding observed with the adjacent ChR1 peptide 182-193 provides further evidence that the E2 binding site is specifically located within this narrow region of ChR1 rather than involving the entire unique ChR1 domain. Taken together, these experiments show that both HPV16 and HPV11 E2 TAD interact directly with purified fragments of the unique domain of ChR1, supporting the hypothesis that the E2-ChR1 interaction is a direct protein-protein interaction that is conserved between papillomavirus types.



**Figure 4.8 Direct interaction between HPV11 E2 TAD and ChIR1 194-213**

His-HPV11 E2 TAD was expressed in BL21 by induction with IPTG. E2 was purified from lysed bacterial cells by affinity chromatography followed by size exclusion chromatography. A. UV AKTA trace showing size exclusion purification of HPV11 E2 TAD (blue line). B. Coomassie stained SDS gel showing fractions of eluted protein from E2

column corresponding to the peak observed in the UV trace. Purified HPV11 E2 TAD was added to GA-resin bound GST (C), GST-ChIR1 182-193 (D) and GST-ChIR1 194-213 (E) in turn and incubated for 4 hours at 4 °C. Unbound E2 protein was removed and the resin washed vigorously to remove non-specific binding protein. C, D, E. Samples of the bound GST protein, unbound E2, wash fraction and the proteins which remain bound to the resin were analysed by SDS-PAGE and Coomassie staining. Bands of the expected size corresponding to the GST or GST-ChIR1 and E2 proteins are indicated. Minimal E2 protein is observed binding to the resin-bound GST or GST-ChIR1 182-193 proteins. In contrast, a band corresponding to bound E2 is observed with the resin-bound GST-ChIR1 194-213 protein suggesting an interaction.

#### **4.3.2 Defining the ChIR1 binding site within HPV16 E2**

Having determined that the HPV16 E2-ChIR1 interaction is a direct protein-protein interaction and that the E2 binding region within ChIR1 exists between amino acids 130 and 214, with strong evidence that the interaction could be limited to a 19 amino acid binding site (194-213), the interaction was further investigated to define the binding site of ChIR1 with HPV16 E2 TAD. Initially a series of HPV16 E2 TAD truncations were obtained from Professor Iain Morgan (University of Virginia, USA), each with an N-terminal GST tag to allow affinity purification of the expressed proteins (Figure 4.9).

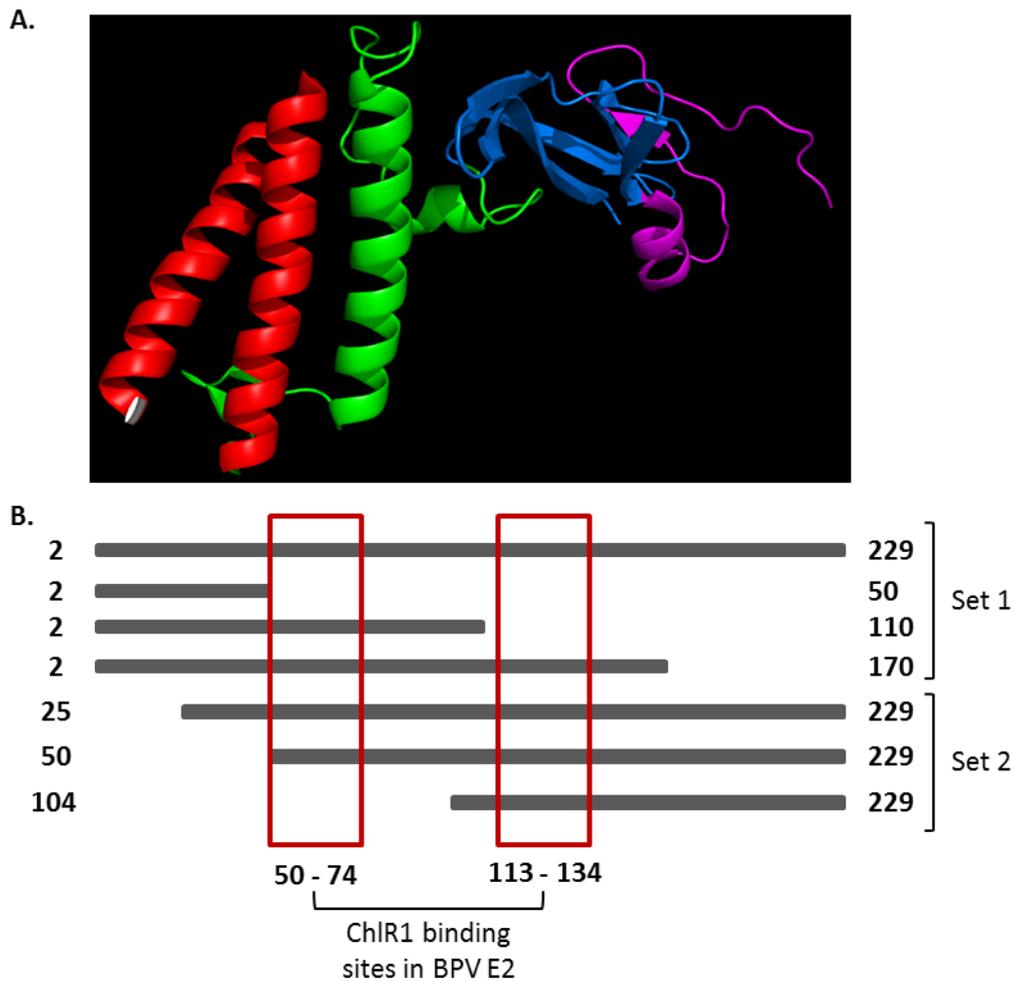


Figure 4.9 HPV16 E2 TAD truncation mutants used to map the ChR1 binding site.

A. Crystal structure of HPV16 E2 TAD, generated from the published PDB file (ID 2NNU, Antson). Truncated domains are highlighted (amino acids 2-50 = red, 51-110 = green, 111-170 = purple and 171-202 in blue. Amino acids 203-229 are not included in the crystal structure). B. Schematic to illustrate the regions of the HPV16 E2 TAD covered by the various truncations expressed and purified. The truncations were separated into two sets, those containing the intact N-terminus of HPV16 E2 TAD with sequential C-terminal truncations (set 1) and those with an intact C-terminus and N-terminal truncations (set 2).

**The full-length HPV16 E2 TAD (aa 2-229) is included in set 1. Highlighted by red boxes are the regions of BPV-1 E2 TAD which were identified as ChR1 binding sites [95].**

Each HPV16 E2 TAD truncation construct was in turn used to transform *E. coli* strain BL21. Transformed *E. coli* were incubated in growth media supplemented with selection antibiotics until an optical density at 600 nm ( $OD_{600}$ ) of 0.6-0.7 was reached. Protein expression was then induced by addition of 1 mM IPTG and incubation with shaking at 37 °C for 4 hours. Bacterial pellets were lysed by resuspension in a standard lysis buffer and sonication (detailed in Materials and Methods) and the lysate and pellet fractions were analysed by SDS-PAGE and Coomassie staining. Coomassie stained SDS gels showed that for the full-length GST-HPV16 E2 TAD and the set 1 truncation mutants, the majority of the over-expressed protein remained within the insoluble pellet fraction and was therefore not accessible for purification and downstream experiments.

To solubilise the over-expressed proteins, an alternative lysis procedure was implemented. As previously, pellets were resuspended in standard lysis buffer and sonicated and the soluble fraction was removed by centrifugation. The insoluble pellet was then washed with a 2% Triton-X-100-containing buffer to solubilise membrane proteins, followed by a further wash in Tris-buffer to remove the excess Triton-X-100. The remaining insoluble material was then resuspended in a second lysis buffer containing 8M urea. These harsh lysis conditions act to denature protein structures, solubilising the proteins contained within the pellet. The urea lysate was then removed from the remaining insoluble pellet and added dropwise to large volume of a refolding buffer. Details of all buffers are included in the Materials and

Methods. This process allows proteins solubilised by the urea buffer to be refolded, allowing downstream experiments to proceed [175].

The refolded soluble proteins were then concentrated to approximately 1 mL using spin-columns containing a suitable molecular weight cut-off filter. To isolate the GST-tagged HPV16 E2 TAD truncation mutants, the concentrated refolded soluble protein fraction was added to glutathione-agarose (GA) resin. The GA-resin and bound proteins were pelleted by centrifugation and the unbound proteins within the lysate were removed. The resin was then washed twice with PBS/T to remove any non-specific binding proteins. The protein fractions obtained from these lysis and purification steps were then analysed by SDS-PAGE and Coomassie staining (Figure 4.10).

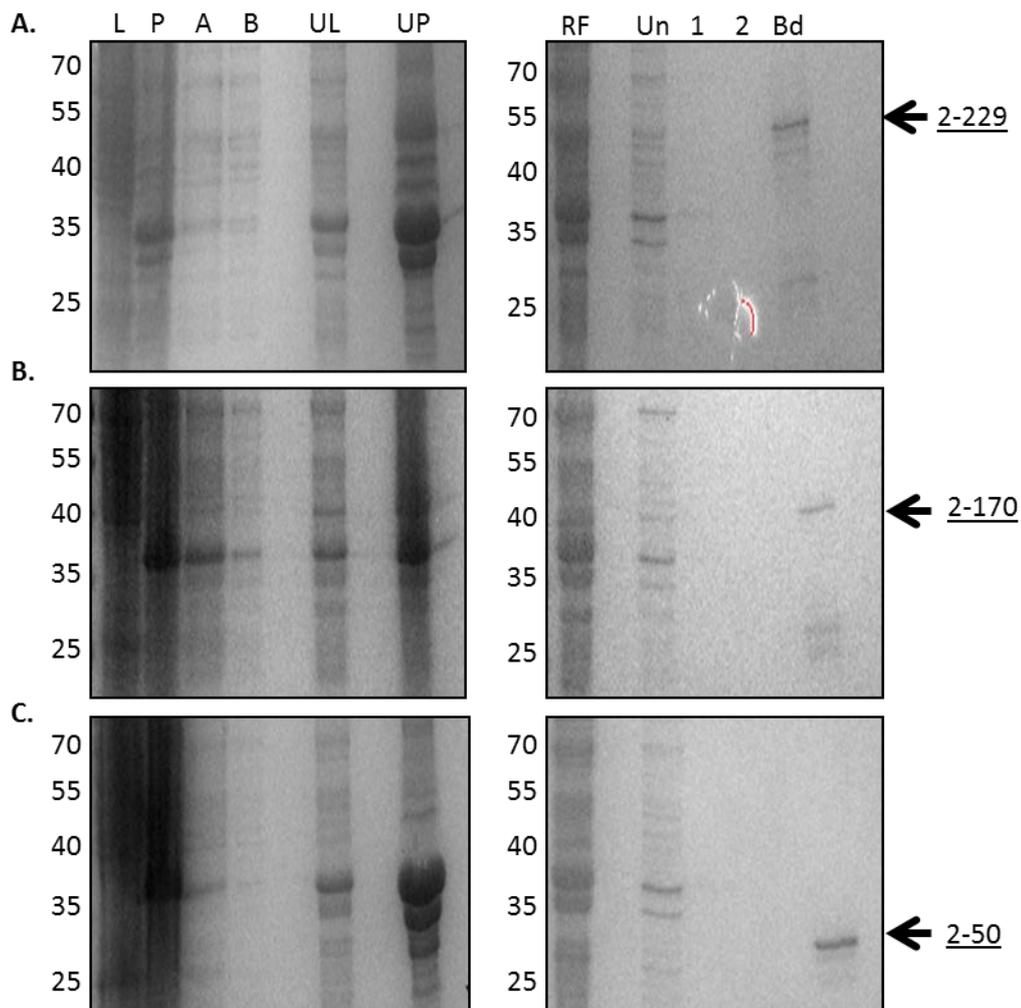


Figure 4.10 Purification of set 1 HPV16 E2 TAD truncation mutants.

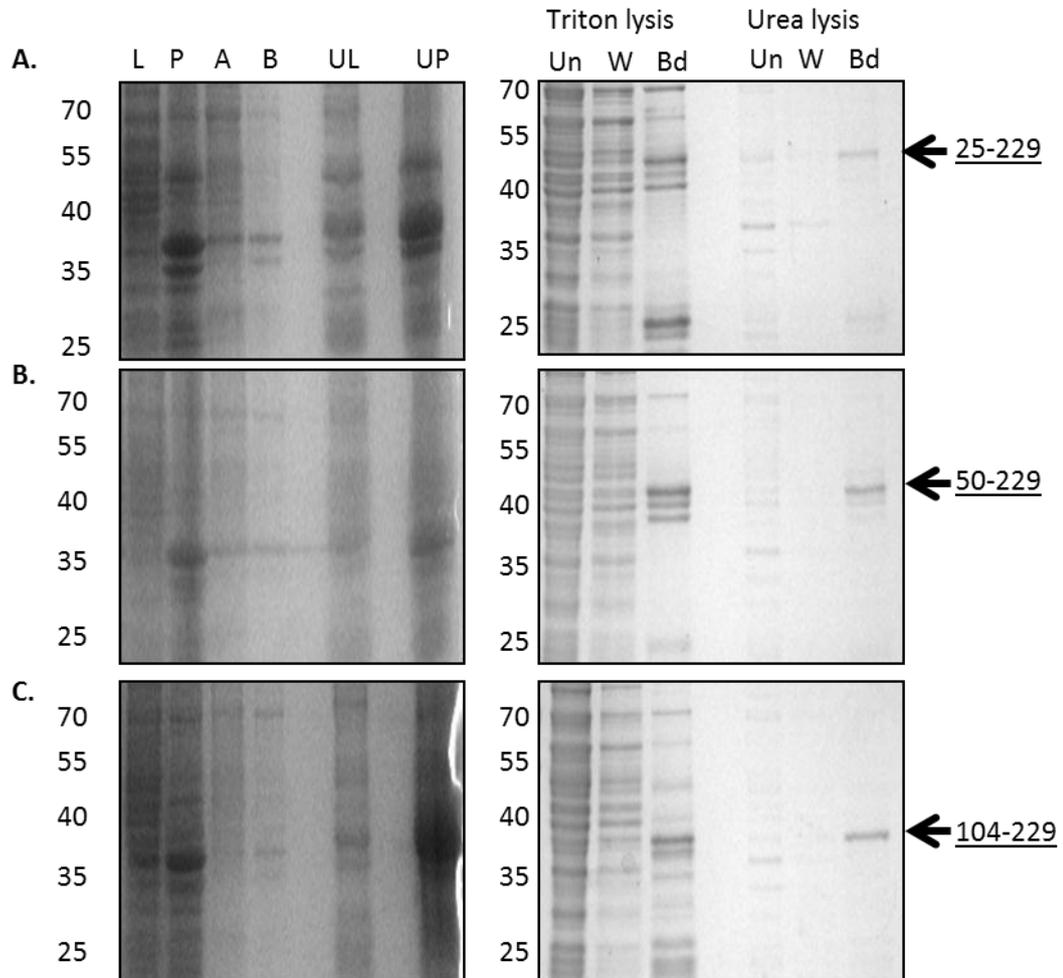
GST-HPV16 E2 TAD truncations were expressed in *E. coli* BL21 following IPTG induction. Bacterial pellets were resuspended in lysis buffer (1 mL per 50 mL culture) and lysed by sonication. Soluble lysate fraction (L) was removed from the insoluble pellet (P), and the pellet fraction was washed with a Triton-x-100 containing buffer (A) followed by a Tris-buffer (B). The pellet was then resuspended in urea lysis buffer (UL) and the lysate containing solubilised proteins was removed from the remaining pellet. The urea lysate was then added to a large volume of refolding buffer. The refolded soluble protein fraction (RF) was then concentrated and added to glutathione-agarose resin (GA) to isolate the

refolded GST-HPV16 E2 TAD truncations. Unbound protein fraction (Un) was removed and the GA resin was washed twice with PBS/T (1, 2). A sample of each fraction along with a sample of the GA-resin bound proteins (Bd) was analysed by SDS-PAGE and Coomassie staining. A. Coomassie stained SDS gels showing lysis (left) and purification (right) fractions for full-length GST-HPV16 E2 TAD, 2-229. B. Coomassie stained SDS gels showing lysis (left) and purification (right) fractions for GST-HPV16 E2 TAD 2-170. C. Coomassie stained SDS gels showing lysis (left) and purification (right) fractions for GST-HPV16 E2 TAD 2-50. Molecular weight standards are shown on the left of each gel.

Analysis of the Coomassie stained SDS gels obtained showed that for each GST-HPV16 E2 TAD truncation, a protein of the correct size could be observed that was isolated on the GA resin following the double-lysis, refolding and purification protocol explained above. Although urea solubilises proteins by disrupting the structure, allowing aggregated, insoluble proteins to become soluble, for the GST-tag to be active and associate with the GA resin it must be correctly refolded. The GST-fusion proteins bound with high affinity to the GA resin, suggesting that the GST-HPV16 E2 TAD truncations have refolded following dilution in refolding buffer in order to be isolated using this affinity purification technique.

The set 2 GST-HPV16 E2 TAD mutants behaved differently to the set 1 proteins. These mutants contained the intact C-terminus of the TAD with truncations at the N-terminus (Figure 4.9). When bacterial pellets containing over-expressed protein were lysed by resuspension in standard lysis buffer and sonication, some of the GST-HPV16 E2 TAD protein was solubilised. The lysate was removed from the pellet and insoluble proteins were extracted following the same protocol as used with the set 1 mutants. GA resin was then

used to isolate the GST-HPV16 E2 TAD truncations from both the first (Triton) and second (urea) lysates as before. Samples of the lysis and purification fractions were again analysed by SDS PAGE and Coomassie staining (Figure 4.11).



**Figure 4.11 Purification of set 2 HPV16 E2 TAD truncation mutants.**

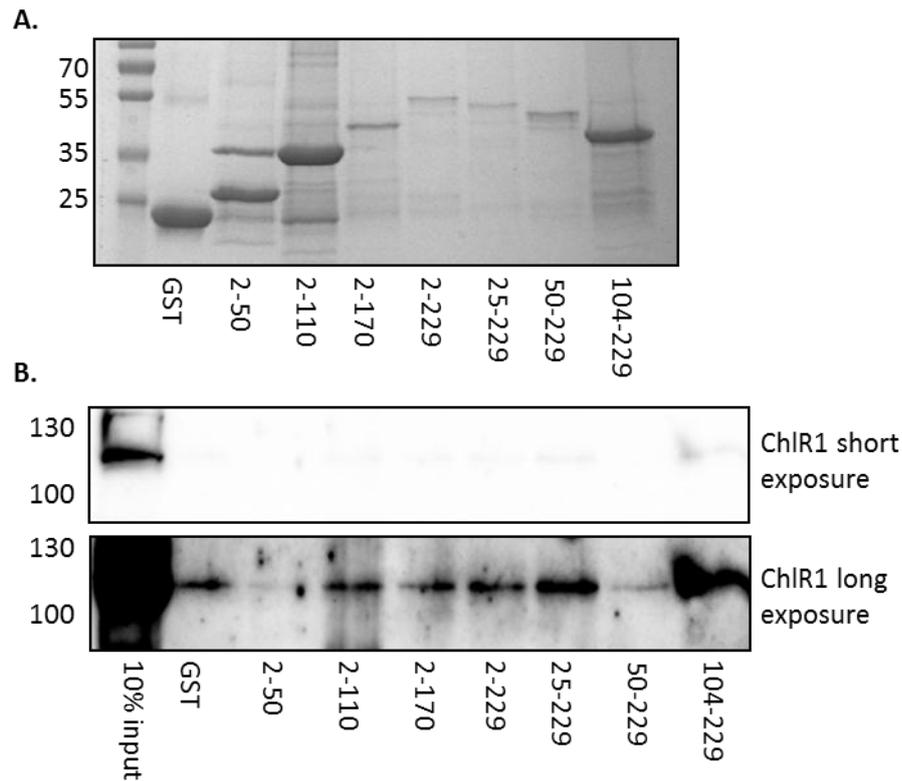
GST-HPV16 E2 TAD truncations were expressed in *E. coli* BL21 following IPTG induction. Bacterial pellets were resuspended in (vol) lysis buffer and lysed by sonication. The soluble lysate fraction (L) was removed from the insoluble pellet (P), and the pellet fraction was washed with a Triton-x-100 containing buffer (A) followed by a Tris-buffer (B). The pellet was then resuspended in a urea lysis buffer and the lysate containing solubilised proteins

was removed from the remaining pellet. The urea lysate was then added to a large volume of refolding buffer. After 16 hours the refolded soluble protein fraction (RF) was concentrated for purification. Both the triton- and urea-lysate were added to glutathione-agarose resin (GA) to isolate the refolded GST-HPV16 E2 TAD truncations. Unbound protein fraction (Un) was removed and the GA resin was washed twice with PBS/T (1). A sample of each fraction along with a sample of the GA-resin bound proteins (Bd) was analysed by SDS-PAGE and Coomassie staining. A. Coomassie stained SDS gels showing lysis (left) and purification (right) fractions for GST-HPV16 E2 TAD 25-229. B. Coomassie stained SDS gels showing lysis (left) and purification (right) fractions for GST-HPV16 E2 TAD 50-229. C. Coomassie stained SDS gels showing lysis (left) and purification (right) fractions for GST-HPV16 E2 TAD 104-229. Molecular weight standards are indicated on the left of each gel.

Proteins corresponding to the correct size were observed in the GA-resin bound fraction for each mutant following purification from the Triton-extracted lysates, however there were also multiple bands observed from non-specific binding of other proteins within the lysate to the resin. Although faint bands of the correct size are also observed in the resin-bound fraction of the urea-extracted lysate purification, these fractions are much cleaner with little other staining observed. The reduction in band intensity is likely due to some of the over-expressed protein being solubilised by the first lysis step and therefore being removed from the pellet prior to urea lysis. However, given the increase in isolated protein purity following urea lysis, all HPV16 E2 TAD truncations from both set 1 and 2 were purified in this way for downstream experiments.

The interaction between the HPV16 E2 TAD truncations and ChIR1 was then investigated using an *in vitro* GST pull-down assay. C33a cells were transfected with FLAG-ChIR1 and cell lysate was incubated with the HPV16 E2 TAD truncations immobilised on GA-resin. Unbound proteins were removed by extensive washing and bound proteins were analysed by western blotting for ChIR1 (Figure 4.12).

The GST pulldown assay found that FLAG-ChIR1 was pulled down by all of the HPV16 E2 TAD truncations with the exception of 2-50 and 50-229. ChIR1 was also detected in the GST control lane, which could be due to a non-specific interaction between ChIR1 and the GST tag, or more likely (as GST-E2 2-50 and GST-E2 50-229 did not pull-down ChIR1 protein), could be spill-over from the adjacent 10% input sample. The greatest amount of binding was detected with E2 truncations 25-229 and 104-229, suggesting that the strongest ChIR1 binding site is situated within the C-terminal 125 amino-acid region of the E2 TAD. This correlates with data obtained looking at the interaction between BPV-1 E2 TAD and ChIR1 which identified a ChIR1 binding site between amino acids 113-134 [95] (Figure 4.9). Interestingly there was only minimal ChIR1 binding detected with E2 truncation 50-229. This could be due to the relatively low levels of that E2 truncation immobilised on the GA resin, but as similar levels of E2 truncations 2-229 and 25-229 were also observed this is unlikely. It is also possible that truncation 50-229 was not correctly folded following the urea lysis protocol, or that the protein was degraded as lower molecular weight bands visible on the Coomassie stained gel would suggest. Although the successful immobilisation of the expressed protein on the GA-resin suggests the GST tag was correctly folded, the E2 domains may not have been fully folded and as such the binding epitopes could have been masked.



**Figure 4.12 HPV16 E2 truncations pull-down FLAG-ChIR1**

**A.** GST-HPV16 E2 TAD truncations were over-expressed in *E. coli* BL21 and purified by a two-step triton-urea lysis followed by refolding and immobilisation on GA-resin. A sample of the resin-bound proteins were analysed by SDS-PAGE followed by Coomassie staining.

**B.** C33a cells were transfected with FLAG-ChIR1. Cell lysate was incubated with GST-HPV16 E2 TAD truncations immobilised on GA-resin. Unbound lysate was removed and the resin washed to remove non-specific binding proteins. The resin-bound proteins were then analysed by SDS-PAGE followed by western blotting with a FLAG-specific antibody to detect FLAG-ChIR1. 10% input of cell lysate extracted from FLAG-ChIR1 transfected C33a cells used in the experiment is also shown. Molecular weight standards are shown to the left.

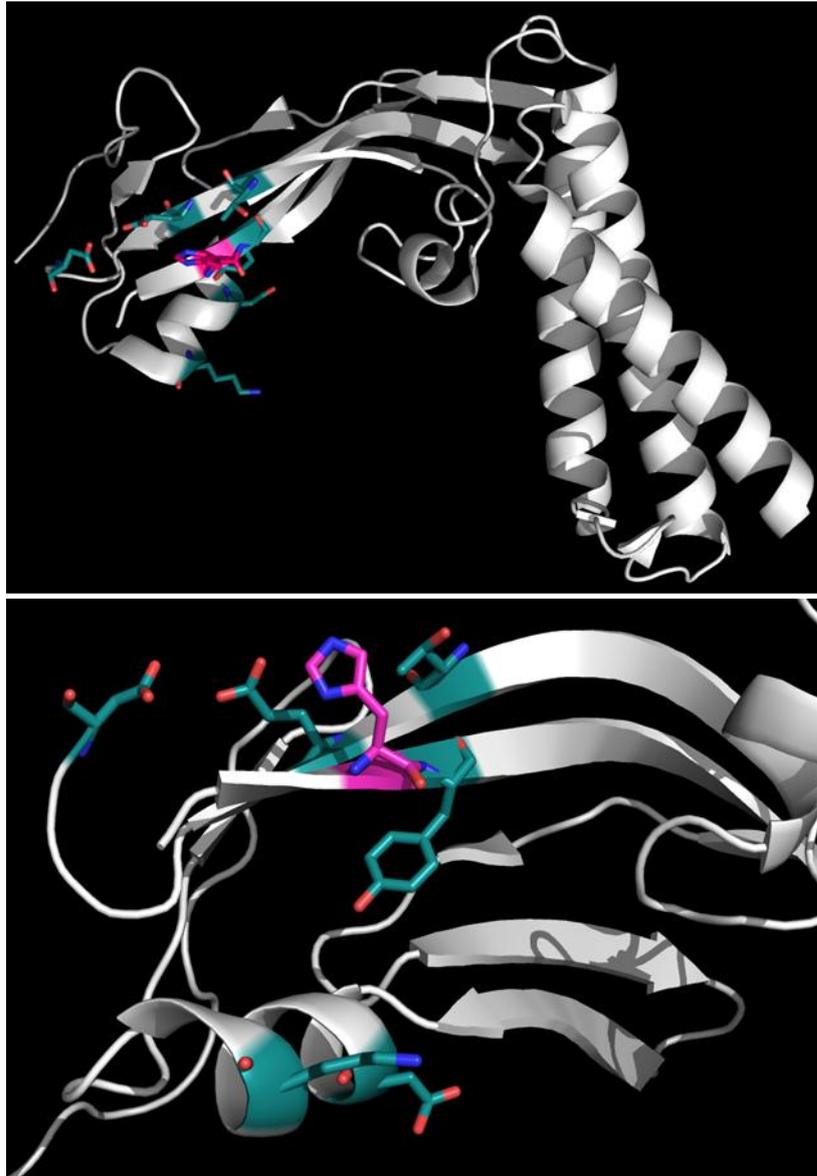
There was also very minimal ChIR1 binding observed with E2 truncation 2-50. This also aligns with the data shown with BPV-1 E2, which showed a second potential ChIR1 binding site is located between amino acids 50-74. If both binding sites are conserved between BPV and HPV16 E2, truncation 2-50 is the only E2 mutant which should not interact with ChIR1.

Given that the data obtained from the pulldown with HPV16 E2 TAD truncations does not disagree with the binding sites identified in BPV-1 E2, it therefore follows that a mutation in HPV16 E2 comparable to the BPV-1 E2 W130R mutation could have a similar effect in disrupting the protein-protein interaction, and therefore prevent the localisation of E2 to mitotic chromosomes, potentially abrogating E2-mediated tethering of viral genomes.

#### **4.3.3 E2 mutagenesis**

The critical amino acid residue in BPV-1 E2 for association with ChIR1 was found to be tryptophan 130 (W130). Mutation of this tryptophan residue to an arginine (W130R) led to abrogation of the E2-ChIR1 interaction. This residue is not conserved between papillomavirus types; in HPV16 a histidine residue is present at this position. However there are multiple residues within the C-terminal region of the E2 TAD which are highly conserved between virus types (Figure 4.13). Analysis of the crystal structure of HPV16 E2 TAD revealed that some of these residues reside on the protein surface and therefore could potentially be involved in protein-protein interactions. In addition to the histidine at position 130, six surface exposed residues which are reasonably conserved between papillomavirus types were selected for mutagenesis (Figure 4.14). These are highlighted in Figures 4.13 and 4.14.





**Figure 4.14 Crystal structure of HPV16 E2 TAD highlighting mutagenesis sites.**

**Crystal structure of HPV16 E2 TAD (PDB ID 2NNU) [65], with the amino acid residues selected for mutagenesis highlighted (green). The histidine residue at position 130 (pink) which was critical for the interaction between BPV-1 E2 and ChIR1 is also shown.**

The residues selected for mutagenesis were as follows: threonine 116 (T116), glutamic acid 118 (E118), aspartic acid 124 (D124), histidine 130 (H130), tyrosine 131 (Y131), aspartic acid

173 (D173), and lysine 177 (K177). Each residue was mutated to alanine (A) by site-directed mutagenesis, with the exception of H130 which was mutated to arginine (R) to mimic the W130R mutation in BPV-1 E2 found to abrogate ChIR1 binding. Purified HPV16 E2-expressing plasmid DNA was used in PCR reactions with primers designed to incorporate amino acid substitutions. PCR amplified DNA products were used to transform *E.coli* strain DH5 $\alpha$ , and plasmid DNA was subsequently extracted from bacterial culture and sequenced using an E2-specific primer (described in Materials and Methods). Each selected residue was mutated individually or in combination, with sequential rounds of mutagenesis being used to add additional mutations to the constructs already generated. A list of the E2 mutants generated is included in Table 4.1.

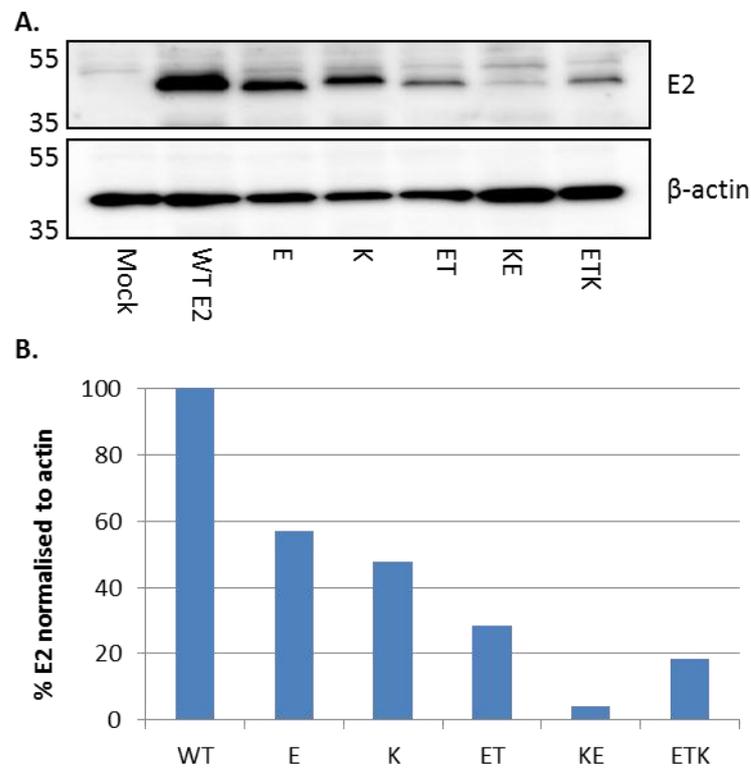
	<b>Mutant ID</b>	<b>Set</b>	<b>Residues mutated</b>
1	H	2	H130R
2	E	1	E118A
3	K	1	K177A
4	ET	1	E118A/T116A
5	EK	1	E118A/K177A
6	EKT	1	E118A/K177A/T116A
7	HE	2	H130R/E118A
8	HET	2	H130R/E118A/T116A
9	HK	2	H130R/K177A
10	HKE	2	H130R/K177A/E118A
11	HKET	2	H130R/K177A/E118A/T116A
12	D124	3	D124A
13	Y	3	Y131A
14	D173	3	D173A
15	ED	3	E118A/D124A
16	HY	3	H130R/Y131A
17	DK	3	D173A/K177A

**Table 4.1 List of HPV16 E2 mutants generated.**

Sequencing with both forward and reverse primers allowed the entirety of the E2 sequence to be checked. For each mutant generated the desired site mutation was identified with no additional changes to the E2 coding sequence.

After the presence of each HPV16 E2 mutation was confirmed by sequencing, the binding interaction between the E2 mutants and the ChR1 fragment previously determined to be sufficient for E2 binding (63-222, Figure 4.3) was investigated by *in vitro* pulldown assays. As

before, His-ChR1 63-222 was over-expressed in bacterial culture and affinity purified using nickel-agarose resin. C33a cells were transfected with HPV16 E2 WT or mutant expressing plasmid in turn, and the transfected C33a cell lysates were incubated with His-ChR1 63-222 protein immobilised on nickel-agarose resin. Unbound protein lysate was then removed and the resin was washed to remove non-specific binding. Samples of the C33a lysates and the resin-bound proteins were then analysed by western blotting with E2 and ChR1 specific antibodies (Figures 4.15 – 4.20).



**Figure 4.15 Expression of HPV16 E2 mutant set 1 in C33a cells.**

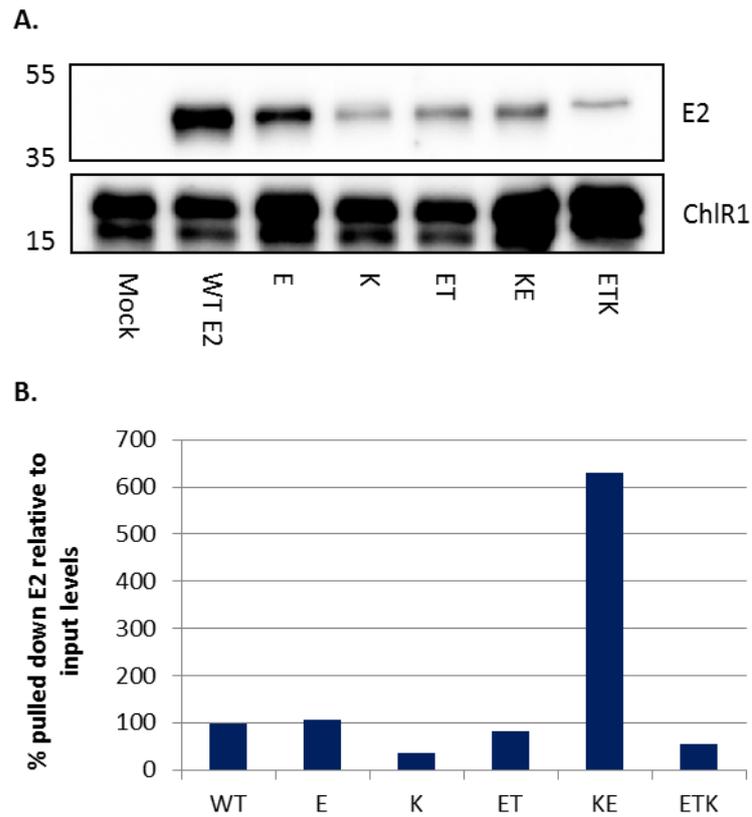
**A. C33a cells were transfected with HPV16 E2 WT or set 1 mutants. Cells were lysed and samples were analysed by SDS PAGE and western blotting with antibodies specific to**

**HPV16 E2, and  $\beta$ -actin as a loading control. B. HPV16 E2 mutant set 1 expression was determined by densitometry and normalised to  $\beta$ -actin.**

All of the set 1 E2 mutants were expressed to a lower level than WT E2 when normalised to the  $\beta$ -actin loading control (Figure 4.15). In particular, the KE mutant was only detected at 4% of the WT levels. The other mutants containing multiple point mutations also expressed poorly, with ET at 28% and ETK at 18% of WT levels. The single mutants E and K were both expressed to ~50% WT levels under these conditions.

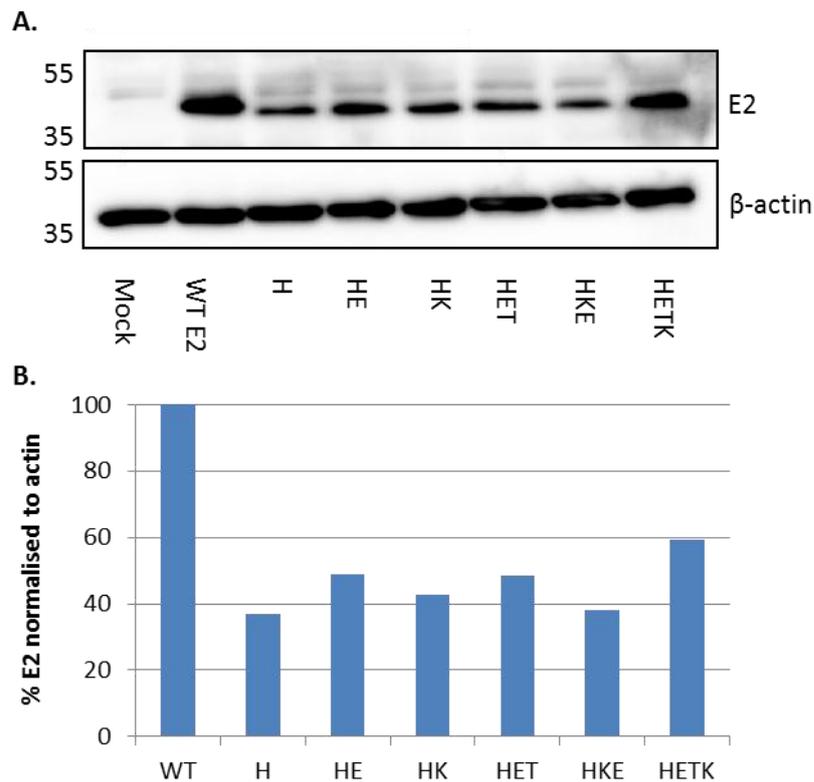
The affinity of the set 1 HPV16 E2 mutants for ChIR1 was determined by pulldown of E2 from the C33a cell lysates by ChIR1 63-222 immobilised on nickel affinity-resin (Figure 4.16). Bound HPV16 E2 protein was quantified by densitometry and normalised to the input protein levels to allow comparison between WT E2 and the mutants. Mutant E was pulled down at levels comparable to WT, and mutant ET showed only a 17% reduction in affinity. Mutant K showed the greatest reduction in affinity for ChIR1, with 64% less E2 being pulled down compared to WT. The triple mutant ETK was pulled down to 56% the level of WT. Interestingly the KE mutant which had been shown to be expressed to much lower levels than WT appeared to bind 6.3 x more effectively to ChIR1 than WT. These results were consistent when repeated with ChIR1 63-214.

This experiment was repeated with the set 2 (Figure 4.17, 4.18) and set 3 (Figure 4.19, 4.20) HPV16 E2 mutants.



**Figure 4.16. Pulldown of HPV16 E2 mutant set 1 with ChIR1 63-222.**

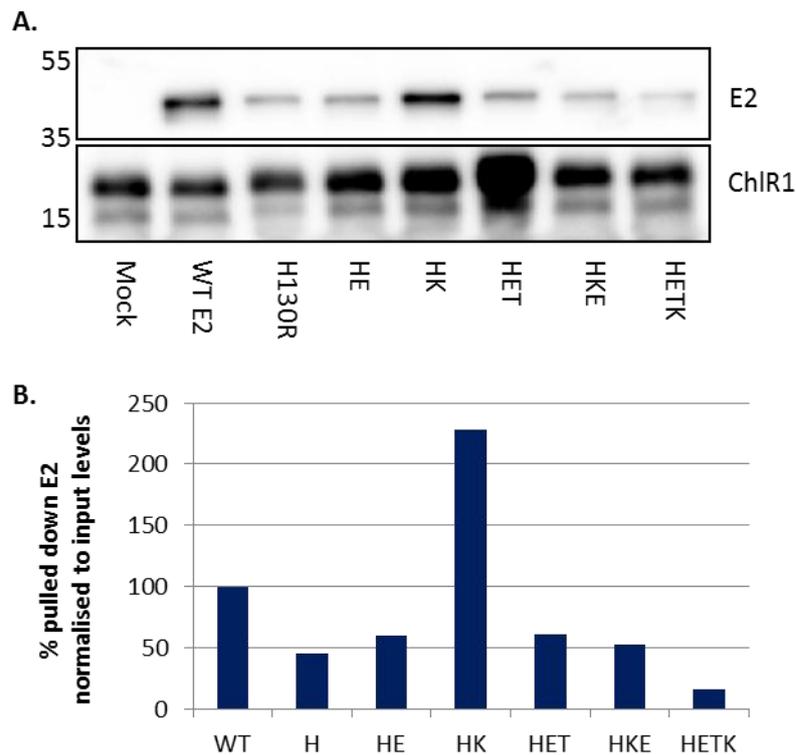
**A.** HPV16 E2 mutant set 1 cell lysates were incubated with ChIR1 63-222 immobilised on His resin. Unbound lysate was removed and resin was washed. Bound proteins were analysed by SDS PAGE and western blotting with antibodies specific to E2 and ChIR1. **B.** Total pulled down E2 was quantified by densitometry, shown relative to input levels normalised to  $\beta$ -actin.



**Figure 4.17 Expression of HPV16 E2 mutant set 2 in C33a cells.**

**A.** C33a cells were transfected with HPV16 E2 WT or set 2 mutants. Cells were lysed and samples were analysed by SDS PAGE and western blotting with antibodies specific to HPV16 E2, and  $\beta$ -actin as a loading control. **B.** HPV16 E2 mutant set 2 expression was determined by densitometry, normalising to  $\beta$ -actin.

The set 2 HPV16 E2 mutants were also consistently expressed to a lower level than WT E2 after normalisation to the  $\beta$ -actin loading control. Each mutant was expressed to between 36% and 59% of the WT levels. The H130R mutant which is the equivalent to the BPV-1 W130R mutant that was critical for the BPV-1 E2-ChlR1 interaction was expressed at the lowest levels, at 36% of WT level. The quadruple mutant HETK was expressed best under these conditions, at 59% of WT levels.



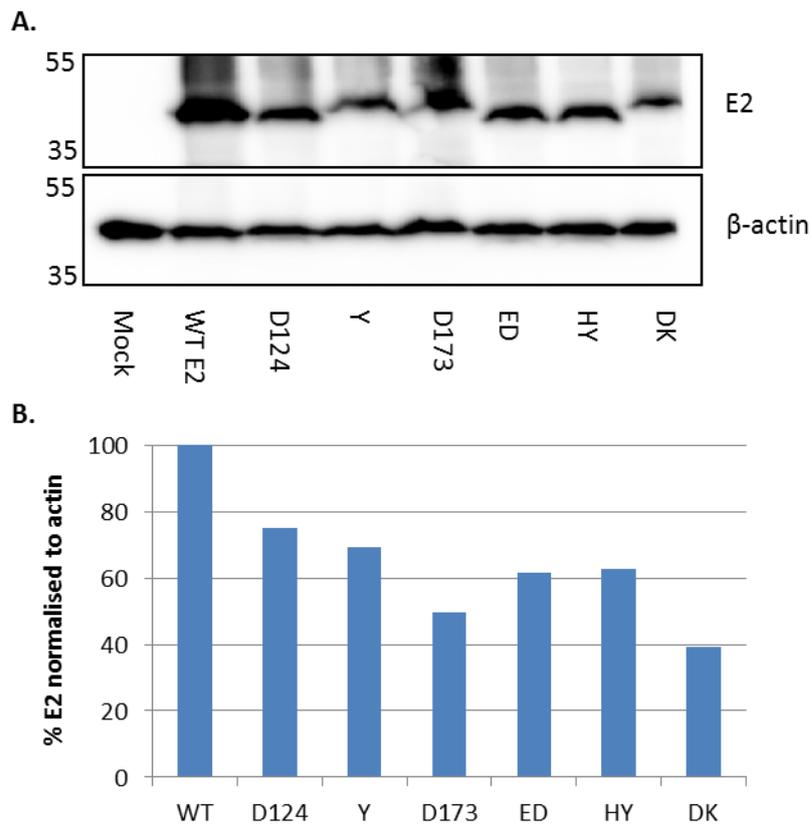
**Figure 4.18 Pulldown of HPV16 E2 mutant set 2 with ChIR1 63-222.**

**A.** HPV16 E2 mutant set 2 cell lysates were incubated with ChIR1 63-222 immobilised on His resin. Unbound lysate was removed and resin was washed. Bound proteins were analysed by SDS PAGE and western blotting with antibodies specific to E2 and ChIR1. **B.** Total pulled down E2 was quantified by densitometry, shown relative to input levels normalised to  $\beta$ -actin.

As with the set 1 mutants, the majority of the set 2 mutants showed a reduction in affinity for ChIR1 and were pulled down by ChIR1 63-222 to a lesser extent than WT E2. HPV16 E2 H130R showed a reduction in ChIR1 affinity of 55% compared to WT, however in BPV-1 E2 the analogous W130R mutation resulted in near total abrogation of the E2-ChIR1 interaction. This suggests that potentially an alternative site is critical for the ChIR1 interaction with

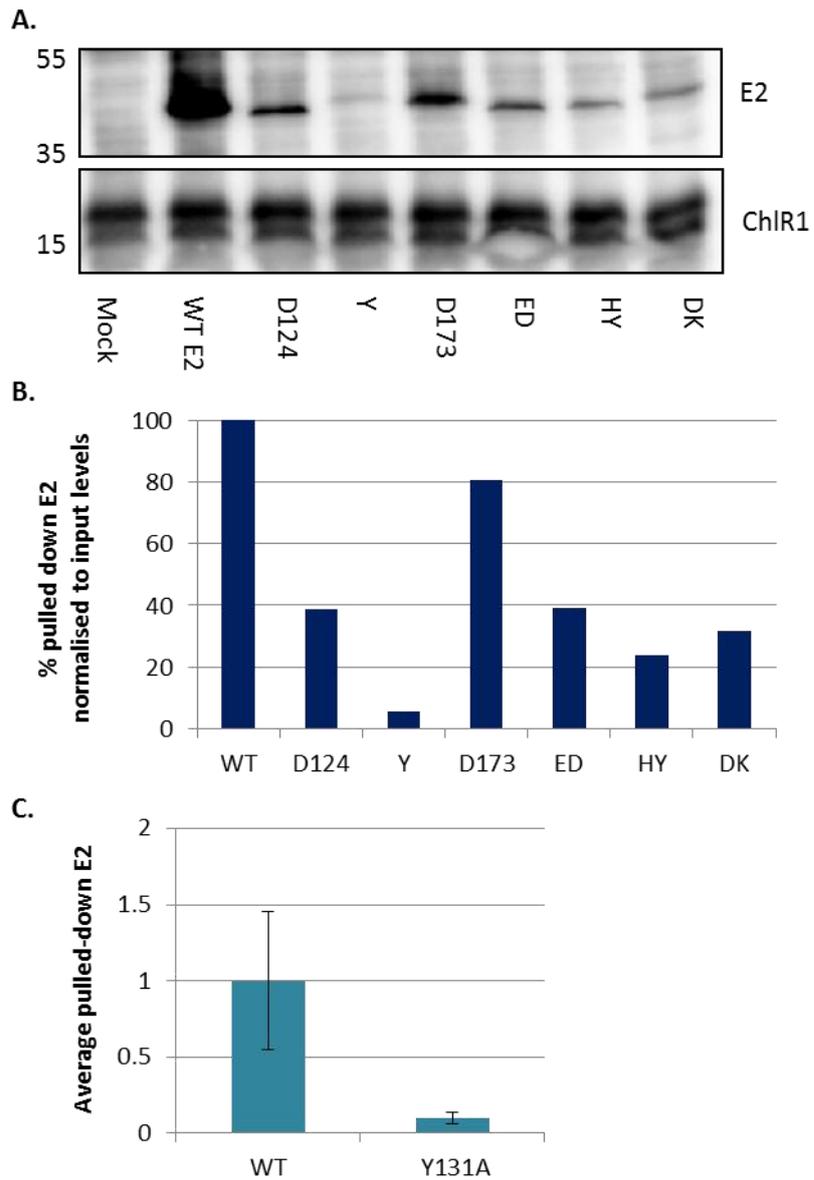
HPV16 E2. Mutants HE, HET and HKE all showed a similar affinity for ChIR1, between 40-50% of WT levels. The quadruple mutant HETK had the greatest effect on the interaction with ChIR1 showing a reduction in pulled down E2 of 84% compared to WT levels. This suggests that potentially rather than a single amino acid residue being critical for the E2-ChIR1 interaction, multiple residues form a binding surface and therefore multiple mutations within the surface are required to disrupt this interaction. As mutant HETK was also expressed to the highest levels relative to WT compared to the other mutants, this could potentially be used in downstream experiments to investigate the role of the E2-ChIR1 interaction in the HPV life cycle.

The set 3 HPV16 E2 mutants were all expressed to reasonable levels in C33a cells (Figure 4.19). The mutation of residue D173 had the greatest impact on protein expression, as the lowest expression levels were observed with the two proteins containing this mutation, DK (39% of WT levels) and D173 (49%). The other mutants were all expressed to >60% of WT levels, with the single mutations D124 and Y131 having the least effect on protein expression (75% and 69% respectively).



**Figure 4.19 Expression of HPV16 E2 mutant set 3 in C33a cells.**

**A.** C33a cells were transfected with HPV16 E2 WT or set 3 mutants. Cells were lysed and samples were analysed by SDS PAGE and western blotting with antibodies specific to HPV16 E2, and  $\beta$ -actin as a loading control. **B.** HPV16 E2 mutant set 3 expression was determined by densitometry, normalising to  $\beta$ -actin.



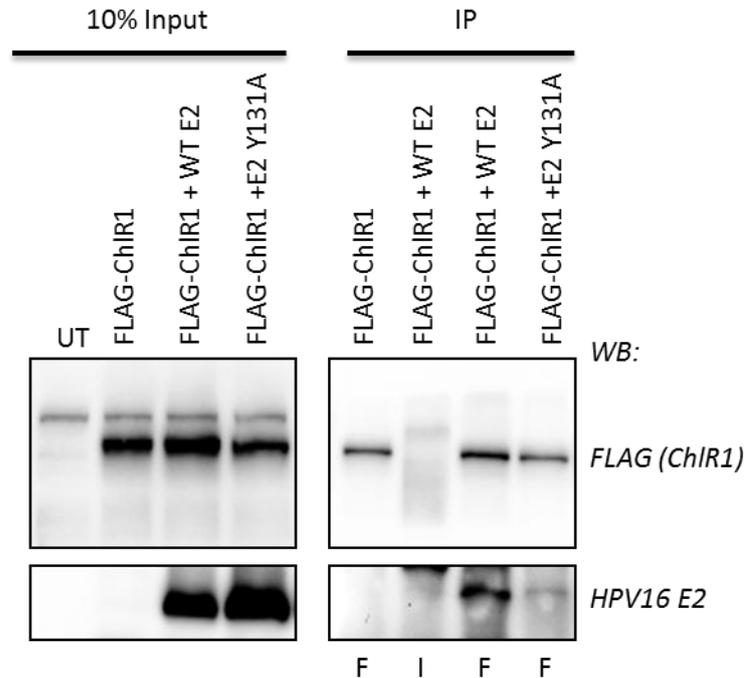
**Figure 4.20 Pulldown of HPV16 E2 mutant set 3 with ChIR1 63-222**

**A.** HPV16 E2 mutant set 3 cell lysates were incubated with ChIR1 63-222 immobilised on His resin. Unbound lysate was removed and resin was washed. Bound proteins were analysed by SDS PAGE and western blotting with antibodies specific to E2 and ChIR1. **B.** Total pulled down E2 was quantified by densitometry, shown relative to input levels normalised to  $\beta$ -actin. **C.** The pulldown assay was repeated with C33a cells transfected

**with HPV16 E2 WT and mutant Y131A. The input protein levels were normalised to the  $\beta$ -actin loading control and the total pulled down protein relative to the normalised input levels was quantified by densitometry. The average of three repeats is shown. Error bars represent standard error.**

As the binding of His-ChIR1 63-222 to set 3 of the HPV16 E2 mutants showed that only 5% binding was detected with mutant Y131A compared to WT E2. The HY double mutant also showed a reduction in ChIR1 binding (24% of WT) but to a lesser extent, indicating that multiple mutations at this site have the effect of somehow partially restoring the ability of ChIR1 to bind to this site. With the exception of the D173 mutant which was pulled down at 80% of WT levels, the other set 3 mutants all showed a 60-70% reduction in ChIR1 affinity compared to WT. Of all 17 HPV16 E2 mutants tested in this *in vitro* pulldown assay with ChIR1 63-222, Y131A remains the mutant which showed the greatest reduction in affinity for ChIR1. To confirm this, the pulldown assay was repeated with HPV16 E2 WT and Y131A a total of three times and the relative amount of bound E2 protein again quantified by densitometry (Figure 4.20 C). On average, HPV16 E2 Y131A showed a 90% reduction in affinity for ChIR1 compared to WT HPV16 E2, a comparable result to that obtained with the BPV-1 E2 mutant W130R. Interestingly, as well as being adjacent to amino acid 130, Y131 is a highly conserved residue between papillomavirus types (BPV-1, HPV-16, -18, -6 and -11). It is therefore possible that this residue is important for the interaction of E2 with ChIR1 across other HPV types. Equally, it is possible that the abrogation in ChIR1 binding observed with the BPV-1 E2 mutant W130R is partially due to disruption of binding to the neighbouring tyrosine residue rather than being a specific effect localised to the mutated residue.

To confirm the reduction in ChIR1 affinity observed with HPV16 E2 Y131A, C33a cells were transfected with FLAG-ChIR1 and WT or Y131A E2 and the protein-protein interaction was assessed by coimmunoprecipitation with FLAG-specific antibodies (Figure 4.21).



**Figure 4.21 HPV16 E2 mutant Y131A shows reduced affinity for ChIR1 by co-immunoprecipitation**

C33a cells were transfected with FLAG-ChIR1 and HPV16 E2 WT or Y131A. Cell lysates were incubated with protein-G-sepharose and a FLAG antibody (F), or a non-specific IgG antibody (I) as a control. Unbound lysate was removed and the bound proteins were analysed by SDS PAGE and western blotting with FLAG and HPV16 E2 antibodies. The image shown is representative of three independent experiments.

FLAG-ChIR1 and both WT and Y131A E2 were expressed to good levels in C33a cells. Minimal HPV16 E2 WT was pulled down with a non-specific IgG antibody, while WT E2 was detected

when a FLAG antibody was used, indicating that the interaction with FLAG-ChIR1 was necessary for E2 binding. Although a low level of HPV16 E2 Y131A was pulled down with the FLAG antibody, this was to a much lower level than WT. This supports the suggestion that the Y131A mutation disrupts the interaction between HPV16 E2 and ChIR1.

#### **4.4 Discussion**

Sequence alignment of the human ChIR1 and yeast CHL1 proteins with the related superfamily 2 DNA helicase XPD revealed a unique domain of ChIR1 located at the N-terminus of the protein between helicase domains I and Ia. To date there is no evidence suggesting that the papillomavirus E2 proteins interact with XPD, or other related SF2 family members; it was therefore hypothesised that the E2 binding site on ChIR1 is located within this unique domain. Modelling of the protein structure indicates that this domain would form a surface exposed protrusion from the helicase domains and that it could therefore be accessible for protein-protein interactions. To investigate this, a series of constructs encoding regions of this unique domain were cloned into bacterial expression vectors and the subsequently expressed proteins were used to map the HPV16 E2 binding site. Initially three fragments of the N-terminus of ChIR1 were expressed, encoding amino acids 1-130, 63-214 and 63-222. The residues 215-222 were omitted from the smaller fragment in an attempt to optimise protein stability, as these residues are highly acidic and therefore could negatively impact protein folding in solution when expressed as a truncated protein. These three ChIR1 fragments were all successfully expressed as His-tagged proteins and could be isolated on nickel-affinity resin to allow downstream binding assays (Figure 4.2).

Interestingly, ChIR1 63-222 was better expressed and easier to isolate than the shorter fragment 63-214 despite concerns about the acidic region affecting protein stability, however both fragments were routinely observed as multiple bands upon analysis by SDS PAGE and Coomassie staining or western blotting. These bands could not be separated sufficiently to allow for identification by mass spectrometry, and it is assumed that the lower bands are due to degradation of the ChIR1 protein resulting in smaller protein species being present.

*In vitro* pull-down assays carried out with the immobilised ChIR1 fragments and HPV16 E2 demonstrated that E2 can associate with amino acids 63-222 and 63-214 of ChIR1, while no E2 was observed to interact with amino acids 1-130. This supports the suggestion that the E2 binding site on ChIR1 is localised to the unique domain between the N-terminal helicase domains, and shows that the amino acid region 131-214 is sufficient for HPV16 E2 binding. However, as the larger fragment ChIR1 63-222 was more readily expressed and purified than the smaller protein (63-214), it was this ChIR1 fragment which was selected for use in further assays.

The interaction between ChIR1 63-222 and HPV16 E2 was also shown to be a direct protein-protein interaction by repeating the *in vitro* pull down assay with bacterially expressed His-HPV16 E2 TAD immobilised on nickel-affinity resin and incubated with purified ChIR1 63-222. Purification of HPV16 E2 TAD remains challenging, despite attempts to optimise expression and purification conditions, however enough protein can be detected in bacterial lysates to allow for use in binding assays. Therefore association of His-HPV16 E2 TAD with ChIR1 was initially detected by incubation of total bacterial lysate, following induction of E2 protein

expression, with nickel-affinity resin. ChIR1 63-222 was purified as an MBP-His fusion protein and the affinity tag subsequently removed by TEV cleavage and further affinity purification. This purified and cleaved ChIR1 truncation was then incubated with the His-HPV16 E2 TAD bound resin. Binding of ChIR1 63-222 was observed with E2-bound resin but not with resin alone.

The presence of the TEV-cleavage site within the MBP-His-ChIR1 63-222 expression vector allowed the expressed ChIR1 protein to first be removed from bulk lysate by affinity purification, then eluted and dialysed before TEV cleavage to remove the MPB-His tag. A second round of affinity purification was then utilised to remove the cleaved tag from the protein, allowing separation of the purified ChIR1 63-222. Although the initial affinity purification of MBP-His-ChIR1 63-222 was not complete with multiple bands being detected by western blot (Figure 4.4), sufficient ChIR1 protein was isolated to allow TEV cleavage of the tag. The TEV-cleavage was also not complete, with a large amount of uncleaved ChIR1 protein still visible. However, subsequent affinity purification was able to separate both the cleaved tag and the uncleaved tagged-ChIR1 protein from the cleaved ChIR1 63-222, as demonstrated by western blotting with both ChIR1- and His-specific antibodies. As had been previously observed, the eluted ChIR1 63-222 protein was detected as multiple bands, indicating possible degradation of the protein in solution resulting in multiple smaller protein products being observed. However, sufficient ChIR1 63-222 remained (detected as a band at approximately 17 kDa) for use in binding assays with HPV16 E2 TAD. When the pull-down assay was repeated in this way, ChIR1 63-222 was observed to bind to His-HPV16 E2 TAD immobilised on nickel-agarose resin, but was not observed to bind to the resin alone (Figure 4.5). This both supports the suggestion that amino acid region 131-214 of the unique

ChIR1 domain is sufficient for interaction with HPV16 E2, and provides the first evidence that this interaction is a direct protein-protein interaction between these two proteins. Previous data did not preclude the possibility that E2 and ChIR1 interact indirectly as part of a larger protein complex, and while it is still possible that the two proteins are involved in a larger complex, this data shows that the two proteins do interact physically. This is further supported by published data using fluorescence resonance energy transfer (FRET) techniques to measure the interaction between BPV-1 E2 TAD and ChIR1 [133]. As the FRET assay requires interacting proteins to be in very close proximity, the positive FRET signal between two tagged proteins suggests a direct protein-protein interaction, although the possibility of an indirect reaction still remains. The interaction studies described in this chapter provide strong evidence that the interaction between HPV16 E2 TAD and ChIR1 is direct, also suggesting that the protein interaction could therefore be targeted and disrupted by mutations within the protein structure, or by external factors including small molecule or peptide inhibitors of binding.

Domain mapping experiments showed that the binding site of HPV16 E2 is within an 83 amino acid region of ChIR1. To map this interaction further, an additional two peptides of ChIR1 (182-193 and 194-213) were cloned into bacterial expression vectors and expressed as GST-fusions. These peptides fall within the identified E2 binding site and structural modelling has predicted that both of these peptides form helical structures. To repeat the *in vitro* binding assay with these ChIR1 peptides, HPV11 E2 was used in place of HPV16 E2. Unlike the previous experiment where the ChIR1 tag could be removed by TEV-cleavage, the pGEX expression plasmid used to clone the ChIR1 peptides as GST-fusion proteins does not include a similar TEV-cleavage site and therefore the tag could not be removed in this case. The GST

tag is a large protein, and in solution could potentially sterically hinder binding of E2 TAD on the significantly smaller ChIR1 peptides. Immobilising the GST-ChIR1 proteins through an interaction between GST and glutathione agarose would optimise the chances of the ChIR1 peptide being accessible for binding. This requires the E2 protein to be purified and added to the immobilised GST-ChIR1 peptides to investigate the binding interaction. As previously stated, HPV16 E2 TAD is expressed to low levels in BL21 and is problematic to purify at high concentrations. HPV11 E2 TAD however is readily expressed and purified (See Chapter 3), and as such it was decided to use HPV11 E2 TAD in this case. The *in vitro* binding assay was repeated with GST-ChIR1 peptides and HPV11 E2 TAD lysate (Figure 4.7) and with purified HPV11 E2 TAD (Figure 4.8). HPV11 E2 was observed to bind to GST-ChIR1 193-214, whereas minimal binding was observed with GST alone or with GST-ChIR1 182-193. This suggests that the E2 binding site on ChIR1 can be narrowed down to this 19 amino acid region. This was observed using both HPV11 E2 TAD lysate and purified E2 protein, adding further support to the suggestion that the E2-ChIR1 interaction is a direct, physical interaction between the two proteins. These results also suggest that the E2-ChIR1 interaction is conserved between papillomavirus types, as HPV16 and HPV11 E2 TAD were shown to interact directly with regions of ChIR1 located within the unique N-terminal domain.

In concert with mapping the E2 binding site on ChIR1 to a small region within the N-terminus of the protein, the binding site of ChIR1 on the HPV16 E2 TAD was also investigated. A series of HPV16 E2 TAD truncation protein-expression constructs were obtained (Figure 4.9) falling into two sets, those with sequential truncations at the N-terminus with an intact C-terminus (set 1) and *vice versa* (set 2). The truncations were expressed in turn as GST-fusion proteins. Expression of the set 1 truncations revealed that the majority of the expressed protein was

insoluble and was retained in the bacterial pellet following resuspension in a Triton-X-100 containing buffer and lysis by sonication. In an attempt to solubilise the GST-HPV16 E2 proteins, a second lysis protocol was implemented following removal of the soluble proteins, with high concentrations of urea being added to the insoluble protein pellet. Urea solubilises insoluble proteins by denaturing the protein structure in a similar manner to thermal denaturation [176]. Soluble proteins can then be removed from the remaining insoluble material and refolded by slow dilution of the urea lysate into a large volume of refolding buffer. Dilution of the urea removes the denaturing influence and allows protein structure to reform, however it is necessary to find a balance between refolding and protein aggregation. At high protein concentrations the number of intermolecular interactions is increased and can result in incorrect folding and protein aggregation. Slow addition of the unfolded proteins into the refolding buffer allows the initial stages of protein refolding to occur at low concentrations, minimising this risk [175, 177]. Dialysis of the refolded protein then allows the diluted urea to be removed from the refolded proteins, preventing further denaturation.

When the urea denaturation and refolding protocol was used with the GST-HPV16 E2 truncation proteins, more of the E2 protein was detected in the soluble fraction. This protein could be successfully immobilised on GST-affinity resin (Figure 4.10), indicating that the protein had refolded correctly following removal of the urea. If the GST protein had not been correctly refolded, the interaction with the glutathione-agarose would not occur and the proteins could not be isolated utilising this high affinity interaction with any efficiency. Despite the evidence that the GST protein refolded correctly and the implication that the associated HPV16 E2 truncation was also refolded however, the refolding of the E2 truncations is implied rather than proven.

The set two HPV16 E2 truncations were observed to be more soluble than the set one proteins. This suggests that the problem of protein insolubility is linked to the truncation of the protein, with an intact N-terminus resulting in more soluble protein. However use of the urea denaturation and refolding protocol discussed did result in an increase in soluble protein and an increase in the purity of the immobilised proteins following affinity purification. As such the urea lysis and protein refolding technique was used with all GST-HPV16 E2 truncations, resulting in all 7 proteins being successfully immobilised on GA-resin for downstream binding assays with ChIR1.

When the immobilised E2 truncations were used in a binding assay with FLAG-ChIR1, the result was inconclusive (Figure 4.12). FLAG-ChIR1 was observed to interact with all E2 truncations with the exception of 2-50 and 50-229. As these two proteins together cover the entirety of the E2 TAD, this result was somewhat unexpected. Despite a high concentration of immobilised protein (as seen as a large band on Coomassie-stained SDS gel, Figure 4.12 A), no ChIR1 was observed to bind to E2 truncation 2-50. This correlates with published data concerning the BPV-1 E2 interaction with ChIR1 which did not identify a binding site within the N-terminal 50 amino acids of E2. ChIR1 was observed to bind to the remaining set 1 mutants (2-110, 2-170) and the full length E2 TAD (2-229) as well as to two of the set 2 truncations (25-229 and 104-229). This again corresponds with expected data, as each of these truncations includes one or both of the amino acid regions identified as binding ChIR1 in BPV-1 E2 (50-73 and 113-134). The expression of the set 1 truncations was varied, with the shorter two proteins (2-50 and 2-110) being observed on Coomassie stained gel as bands comparable in intensity to the GST control. 2-170 and 2-229 were expressed to a lesser extent, as shown by fainter bands on the gel (Figure 4.12 A). Despite this discrepancy in

immobilised protein concentrations, E2 truncations 2-110, 2-170 and 2-229 all resulted in similar amounts of bound ChIR1 protein (Figure 4.12 B). Set 2 E2 truncation 25-229 was also observed to be immobilised at a similar low concentration, however more ChIR1 binding was observed than that bound to E2 2-229 and the set 1 truncations. Similarly, more ChIR1 was detected as binding to the final set 2 E2 truncation 104-229 compared to the concentration of E2 protein. This suggests that the ChIR1 binding site located toward the C-terminus of the E2 TAD is the stronger binding site, but that multiple ChIR1 binding sites are located within the E2 TAD structure. Despite the lack of sequence homology between BPV-1 E2 and HPV16 E2 (Figure 4.13), the proteins share a high degree of structural similarity so it is not without reason to suggest that the protein binding sites might be similar between papillomavirus types. These data do not disagree with published data investigating the ChIR1 interaction with BPV-1 E2 TAD showing that there are two regions within the E2 TAD which were sufficient for ChIR1 binding, amino acids 50-73 and 113-134, suggesting the possibility that these sites are conserved across papillomavirus types.

The lack of ChIR1 binding to E2 truncation 50-229 remains anomalous. On the basis of binding to other, overlapping fragments there should be a ChIR1 binding site located within this protein truncation. Although not expressed and immobilised at a similar concentration to some of the other truncations, the protein levels are not dissimilar to those observed with truncations 25-229 and 2-229, both of which are shown to bind ChIR1 and therefore the lack of binding observed in this case cannot be attributed to low protein levels. It is possible that this particular E2 truncation was not folded correctly following urea denaturation. Although the immobilisation of the protein on the GA-resin required the GST tag to be correctly folded, it is possible that in this case the E2 protein was not folded correctly. Alternatively,

the GST-tag may also not have refolded optimally, leading to the lower degree of isolated protein being observed.

Published data investigating the BPV-1 E2 TAD interaction with ChIR1 resulted in the identification of a point mutation in E2, W130R, which abrogated the ChIR1 interaction and resulted in a change in the E2 localisation to mitotic chromosomes. Identification of a similar mutant in HPV16 E2 would allow this interaction to be studied further using a model system to study the viral life cycle [178]. Sequence alignment of BPV-1 E2 with the most prevalent HPV types, HPV-16, -18, -11 and -6 (Figure 4.13) revealed multiple conserved residues throughout the protein, however the amino acid analogous to BPV-1 E2 W130 is not conserved. In HPV16 E2, a histidine residue is instead located at this position. If the E2-ChIR1 interaction is conserved between papillomavirus types, it is possible that this interaction proceeds through interaction at one of these conserved sites. When the crystal structure of HPV16 E2 TAD (Figure 4.14) [65] was examined, six amino acid residues were identified as surface exposed and within the predicted binding surface of ChIR1. These amino acids were mutated to alanine using site-directed mutagenesis, both singly and in combination to generate a panel of HPV16 E2 mutants. H130 was also mutated to mimic the W130R mutation found to be critical for BPV-1 E2 binding to ChIR1. Following extensive mutagenesis, 17 HPV16 E2 mutants were generated and confirmed by sequencing.

The HPV16 E2 mutants were then expressed in C33a cells and the binding affinity for ChIR1 63-222 (previously shown to be sufficient for the E2-ChIR1 interaction) was determined for each mutant by *in vitro* pull-down assay. The mutants were divided into 3 sets: set 1 included combinations of mutations at residues T116, E118 and K177; set 2 included the

same mutations in the presence of the H130R mutation; and set 3 included combinations of mutations at residues D124, Y131A and D173 in combination with the other listed residues (see Table 4.1 for a list of generated mutants). The HPV16 E2 mutants were expressed to varying levels, although all were detected by western blotting at a lower level than HPV16 E2 WT. Mutants D124A and Y131A were expressed to the highest levels relative to WT E2, while the double mutant K177A/E118A (KE) was expressed to the lowest level relative to WT. The expressed E2 mutants were used in binding assays with ChIR1 63-222 and the total pulled down E2 relative to the input levels (which had been normalised to  $\beta$ -actin loading control) was calculated by densitometry. The majority of the mutants tested showed a small reduction in ChIR1 binding affinity compared to WT, although in some cases the mutant protein was observed to have an increased affinity for ChIR1. Mutants KE and HK (H130R/K177A) were found to bind to ChIR1 to a much greater extent than WT E2 relative to the protein expression levels. Some mutants did exhibit reduced affinity for ChIR1 compared to WT E2 however. Mutants K177A, HETK (H130R/E118A/T116A/K177A) and Y131A showed the greatest reduction in ChIR1 affinity, with the lowest levels of binding observed with Y131A. This was of great interest, as Y131A is the amino acid adjacent to the residue at position 130 which was necessary for the BPV-1 E2-ChIR1 interaction. Unlike the residue at 130 however, Y131 is conserved across all papillomavirus types investigated (Figure 4.13), and is therefore of interest as a potentially conserved site of ChIR1 binding between PV types. It is possible that the efficacy of the BPV-1 E2 W130R mutation is due to disrupting protein interactions with the adjacent tyrosine residue by introducing steric or electronic interference. Reduction of ChIR1 binding with HPV16 E2 mutant Y131A was then confirmed by co-immunoprecipitation (co-IP) with FLAG-ChIR1. Immunoprecipitating with a FLAG-

specific antibody resulted in a reduction in HPV16 E2 Y131A protein detected compared to HPV16 E2 WT.

For these reasons, the HPV16 E2 mutant Y131A was selected for use in further investigation of the E2-ChIR1 interaction. The effect of loss of the E2-ChIR1 interaction in the virus life cycle was studied using this point mutation (Chapter 5).

## Chapter 5. Characterising the HPV16 E2 mutant Y131A

### 5.1 Introduction

The like other viruses which maintain episomal genomes within infected cells, papillomaviruses have developed mechanisms to ensure even segregation of viral genomes between daughter cells during mitosis. The Epstein-Barr virus (EBV) and Kaposi's sarcoma-associated herpes virus (KSHV) have been shown to use similar tethering mechanisms, utilising the ability of viral proteins (EBNA-1 in EBV and LANA in KSHV) to interact with both the viral genomes and cellular proteins to facilitate viral genome partitioning [97-100]. In papillomaviruses the E2 protein acts as a tether, interacting with the viral genome through the DNA binding domain (DBD) and cellular chromatin-associated proteins through the transactivation domain (TAD) [84, 94, 96].

Different PV types have been shown to behave differently in this mechanism. There is evidence to suggest that the E2 proteins from different PV genera can localise differently with mitotic chromatin [93], suggesting the possibility that multiple cellular chromatin receptors of E2 are utilised in this tethering mechanism. BPV-1 E2 and other animal PVs including cottontail rabbit (CRPV), rabbit oral (ROPV), canine oral (COPV), European elk (EEPV) deer papillomavirus (DPV) as well as some human PVs with tropism for cutaneous epithelium from different genera ( $\mu$ -HPV1a,  $\beta$ -HPV8 and  $\gamma$ -HPV4) have been shown to localise in discrete foci on mitotic chromosomes in IF experiments. The  $\beta$ -papillomavirus HPV8 showed E2 localisation specific to the pericentromeric regions of mitotic chromosomes, in contrast to the less-ordered pattern of foci observed with BPV-1 E2, suggesting alternative mechanisms for tethering are in use. The  $\alpha$ -papillomaviruses (HPV16, -

11, -31 and -57) again displayed a different pattern of localisation, with chromatin localised foci detected in both prophase and telophase cells but not in metaphase [77, 93], supporting the suggestion that alternative tethering mechanisms are utilised across PV types. As the E2 TAD has been shown to be essential for the localisation of  $\alpha$ -PVs to the chromatin, it is likely that the regulation of this tethering is through the protein-protein interactions formed by the TAD.

There have been several cellular proteins proposed to act as the chromatin-associated receptor of E2 to mediate this tethering. The bromodomain containing protein Brd4 was identified as an E2 interacting protein by tandem affinity purification (TAP) proteomics screen [75]. Brd4 and BPV-1 E2 have been shown to colocalise at mitotic chromosomes throughout mitosis [76, 77], although the same patterns of mitotic foci were not observed with  $\alpha$ -papillomavirus E2s [77, 95]. Mutagenesis of BPV-1 E2 identified mutations in E2 which abrogated the protein interaction with Brd4: arginine 37 to alanine (R37A) and isoleucine 73 to alanine (I73A). Investigation with R37A/I73A E2 demonstrated that the interaction of Brd4 with E2 is both necessary for BPV-1 E2 localisation with mitotic chromatin, and that this association correlates to the transcription activation function of E2, however the replication activity of E2 remains unimpaired. Further investigation showed that although the interaction of E2 with Brd4 is necessary for the transcription activation function of E2, the association of the  $\alpha$ -PV E2s with mitotic chromosomes is likely to be regulated through an alternate cellular chromatin receptor.

Another proposed candidate as the cellular binding partner of E2 implicated in viral genome tethering is topoisomerase II binding protein 1 (TopBP1). TopBP1 was identified as a HPV16

E2 interacting protein by yeast-2-hybrid screen, and further investigation showed that like the E2-Brd4 interaction, the E2-TopBP1 interaction is mediated through the E2 TAD [107]. The TopBP1 interaction is not required for the transcription activation function of E2 [107, 108], however the interaction has been shown to enhance the replication activity of E2, again demonstrating that the functional activities of E2 are separable. HPV16 E2 and TopBP1 were observed to co-localise at mitotic chromosomes during late mitosis, and interestingly, depletion of TopBP1 resulted in an alteration in E2 localisation resulting in increased protein stability and an increase in localisation to the cellular chromatin [108], suggesting that a change in the interaction of E2 with TopBP1 results in a change in the localisation of E2 to alternative cellular chromatin complexes which alters the protein stability. Mutagenesis of HPV16 E2 identified a TAD mutation, asparagine 89 to tyrosine (N89Y) and glutamic acid 90 to valine (E90V), with reduced affinity for TopBP1. This mutant E2 shows significantly reduced replication activity compared to WT E2, but retains the ability to activate transcription [109]. When primary human foreskin keratinocytes (HFKs) were transfected with HPV16 genomes containing the HPV16 E2 mutant N89Y/E90V, the mutant HPV16 genome lines failed to establish episomal genomes. Together, this suggests that the E2-TopBP1 interaction is involved in the E2-dependent cellular chromatin tethering mechanism of PVs.

The DNA helicase ChlR1 is proposed to be important for loading E2 onto the cellular-chromatin protein complex required for viral genome tethering. ChlR1 was identified as an interacting partner of BPV-1 E2 by yeast-2-hybrid screen, and the interaction was confirmed by co-immunoprecipitation [95]. Domain mapping of the interaction demonstrated the presence of two domains within the BPV-1 E2 TAD which are required for ChlR1 binding

(amino acids 50-73 and 113-134). Fluorescence resonance energy transfer (FRET) experiments with E2 and ChIR1 demonstrated that the two proteins are in very close proximity, suggesting a direct interaction [133] and revealed that the E2-ChIR1 interaction is cell cycle dependent, occurring primarily in mid-S phase. Mutagenesis of BPV-1 E2 TAD revealed a point mutation at amino acid tryptophan 130 (W130R) which resulted in a decrease in the affinity of E2 for ChIR1, with no reduction in the functional activity of E2. Mutant E2-BPV-1 genome containing mouse fibroblasts were unable to establish episomes [95], providing evidence to support the suggestion that the E2-ChIR1 interaction is required for the E2-dependent tethering of viral genomes to mitotic chromosomes.

The identification of a HPV16 E2 mutant (Y131A) deficient in binding ChIR1 (see Chapter 4) provided an opportunity for further investigation into the biological function of the E2-ChIR1 interaction. The activity and localisation of HPV16 E2 Y131A were characterised using assays designed to measure various activities of E2 in cell-based assays. Using a HPV-negative cervical carcinoma-derived keratinocytes (C33a), initial experiments were carried out to characterise the function and localisation of mutant E2 compared to wild type (WT). Following on from this, primary human foreskin keratinocytes (HFKs) were transfected with HPV16 genomes encoding WT or mutant Y131A E2 protein and selected to generate HFK lines containing HPV genomes, which were used to assess the effect of the E2-ChIR1 interaction in the context of the viral life cycle.

Building from the published data investigating the interaction between BPV-1 E2 and ChIR1, we hypothesised that the loss of ChIR1 interaction will result in an impaired ability to

maintain viral genomes within primary cell lines due to a loss of the E2-dependent tethering of viral genomes to cellular chromatin, resulting in episome loss over time.

## **5.2 Aims**

Having identified a mutant of HPV16 E2 (Y131A) which is deficient in ChIR1 binding (See Chapter 4), the aims of this work were to:

- 1.** Determine the effect of the ChIR1 binding mutation Y131A on the replication and transcription activation functions of HPV16 E2
- 2.** Investigate the impact of reduced ChIR1 affinity on the stability and subcellular localisation of HPV16 E2
- 3.** Use a primary cell model system to investigate the ability of HPV genomes containing the E2 mutant Y131A to establish and maintain genomes within the context of the viral life cycle

## 5.3 Results

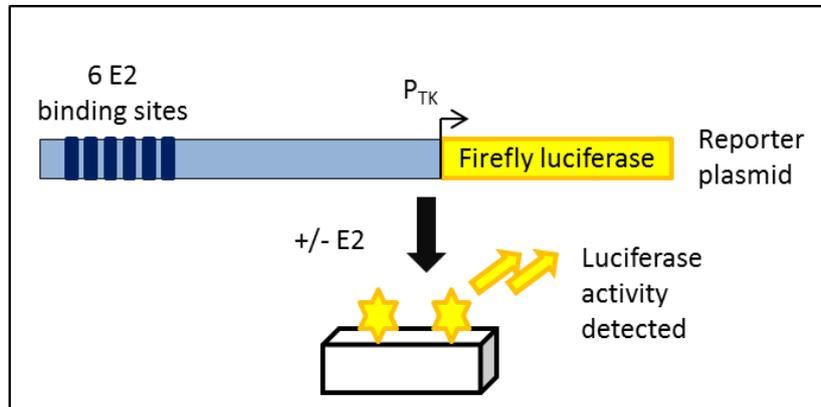
### 5.3.1 E2 function

The HPV life cycle is dependent on the E2 protein. E2 mediates viral genome replication by facilitating loading of the E1 helicase to the viral origin of replication (Ori), and regulates early gene transcription by both activating and repressing viral transcription [28]. E2 function is therefore important in the regulation of virally-induced cell proliferation and the replication and amplification of viral genomes. As such, mutations in the E2 protein could have a dramatic impact on the viral life cycle by affecting either of these functions. Introduction of the Y131A mutation into HPV16 E2 significantly reduces the protein affinity for ChIR1 (Chapter 4), but in order to use this mutant E2 to investigate the importance of the E2-ChIR1 interaction in the life cycle it was necessary to ascertain whether this mutant E2 protein functions normally to both mediate transcription and facilitate viral genome replication.

#### *Transcription activation*

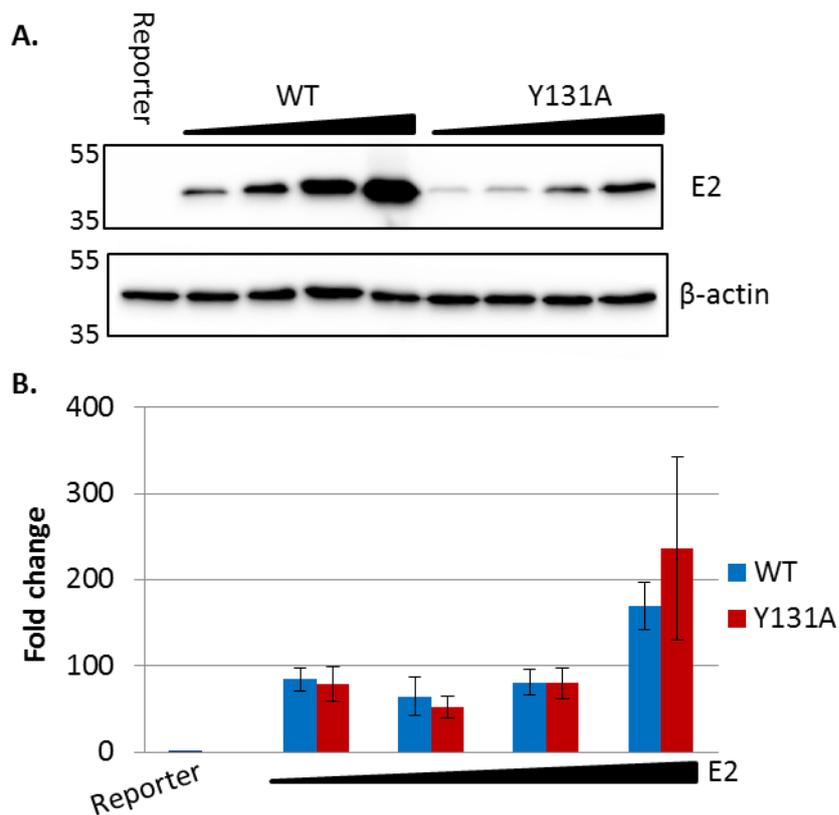
To determine whether the Y131A E2 mutation alters the ability of E2 to activate transcription, a luciferase-based reporter assay was used [133]. C33a cells were transfected with a reporter plasmid (ptk-6E2-Luc) containing the Firefly luciferase open reading frame (ORF) downstream of a thymidine kinase promoter. Six E2 binding sites are situated immediately upstream to the promoter, allowing regulation of Firefly luciferase transcription by E2 (Figure 5.1). Cells were co-transfected with increasing amounts of WT or Y131A E2 along with a plasmid that constitutively expressed *renilla* luciferase from a cytomegalovirus promoter (pCMV-RLuc) as an internal control. E2-dependent transcription activation was

then directly quantified by measuring the luciferase activity using a commercial kit (Figure 5.2).



**Figure 5.1 A schematic representation of the E2-dependent transcription activation assay**

A reporter plasmid containing six E2 binding sites upstream of a thymidine kinase promoter controlling expression of firefly luciferase was transfected into C33a cells along with increasing amounts of an E2 expression plasmid and a plasmid which constitutively expresses with *renilla* luciferase as an internal control. After cell lysis, luciferase activity was measured using a Dual-Luciferase assay kit.



**Figure 5.2 HPV16 E2 Y131A is able to activate transcription to WT levels**

**A.** C33a cells were transfected with WT or Y131A E2 (100 – 1000 ng) along with ptk-6E2-Luc reporter plasmid and *renilla* luciferase. Representative western blots showing the levels of E2 protein expression and  $\beta$ -actin as a loading control are shown. Molecular weight standards are indicated on the left. **B.** The E2-dependent Firefly luciferase activity detected was normalised to the *renilla* luciferase activity. The mean and standard error of three independent repeats is shown.

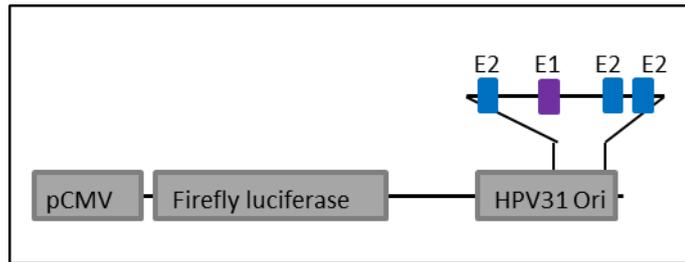
E2 protein expression levels were assessed by western blot analysis of a sample of the cell lysates used to measure the luciferase activity. Y131A E2 was consistently expressed to a lower level than WT E2. Despite this, no significant differences in Firefly luciferase activity

were observed with increasing E2 protein concentration. This suggests that the mutant E2 protein is able to activate transcription to similar levels as WT E2 and provides evidence that neither the disrupted E2-ChIR1 interaction or the Y131 residue are required for the transcription activation activity of E2.

#### *E2 dependent replication*

A similar luciferase-based assay was used to determine if the Y131A mutation and the corresponding loss of the E2-ChIR1 interaction had an effect on E2-dependent viral replication. This assay was originally developed by J. Archambault [179] and uses a plasmid construct that contains the minimal origin of replication (Ori) from HPV31 downstream of a Firefly luciferase reporter gene which is independently controlled by a CMV promoter (pFLOri) (Figure 5.3). C33a cells were transfected with this reporter along with HPV16 E1 and increasing amounts of WT or Y131A E2. It has previously been shown that the luciferase activity is proportional to the E1-E2 dependent replication of the plasmid [179].

C33a cells were transfected with pFLOri, HA-HPV16 E1 and increasing amounts of WT or Y131A E2, along with a plasmid with constitutively expresses *renilla* luciferase as an internal control. Samples of the cell lysates were analysed by western blotting with antibodies for HPV16 E2 and the HA-tag for E1 along with  $\beta$ -actin as a loading control to assess E1 and E2 protein expression (Figure 5.4).

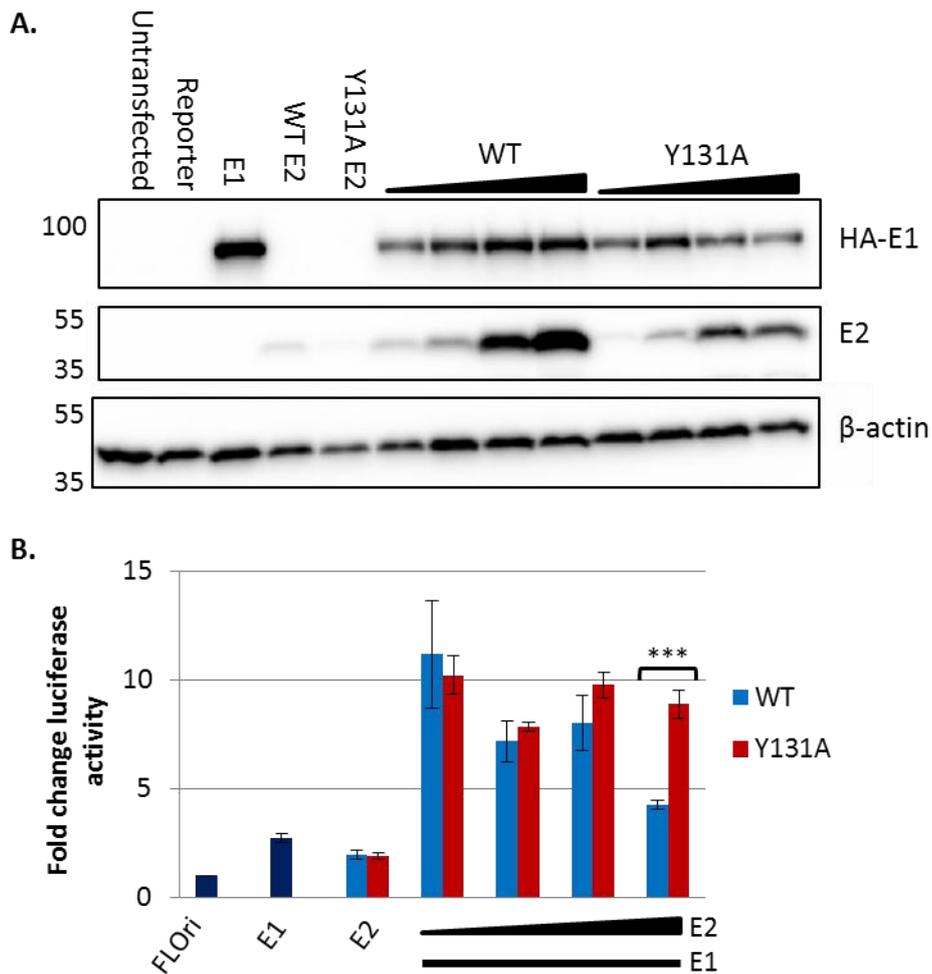


**Figure 5.3 Schematic representation of the luciferase-based E2-dependent replication assay reporter.**

**The reporter plasmid used in replication assays consists of the Firefly luciferase gene under control of a CMV promoter. The plasmid includes the HPV31 minimal origin of replication (Ori), containing E1 (purple) and E2 (blue) binding sites.**

Consistent with the previous results (Figure 5.2), Y131A E2 was expressed at a lower level than WT E2. However, as observed in the transcription assays described above, there was no significant difference in the Firefly luciferase activity detected at the lower concentrations of transfected E2. Transfection with both WT and Y131A E2 led to a 5-10 fold increase in luciferase activity compared to pFLOri alone. At the highest concentration of E2 however, there is a reduction in Firefly luciferase activity in WT transfected cells compared to the lower concentrations whereas in the Y131A cells the luciferase activity remains high.

Nonetheless, these data show that Y131A E2 is able to facilitate replication of the pFLOri reporter in this system to WT levels at low concentrations and shows no loss of this activity over an increasing concentration range. This indicates that the Y131A mutation and the corresponding reduction in E2-ChR1 interaction do not adversely affect the replication activity of HPV16 E2 which is essential for the virus life cycle.



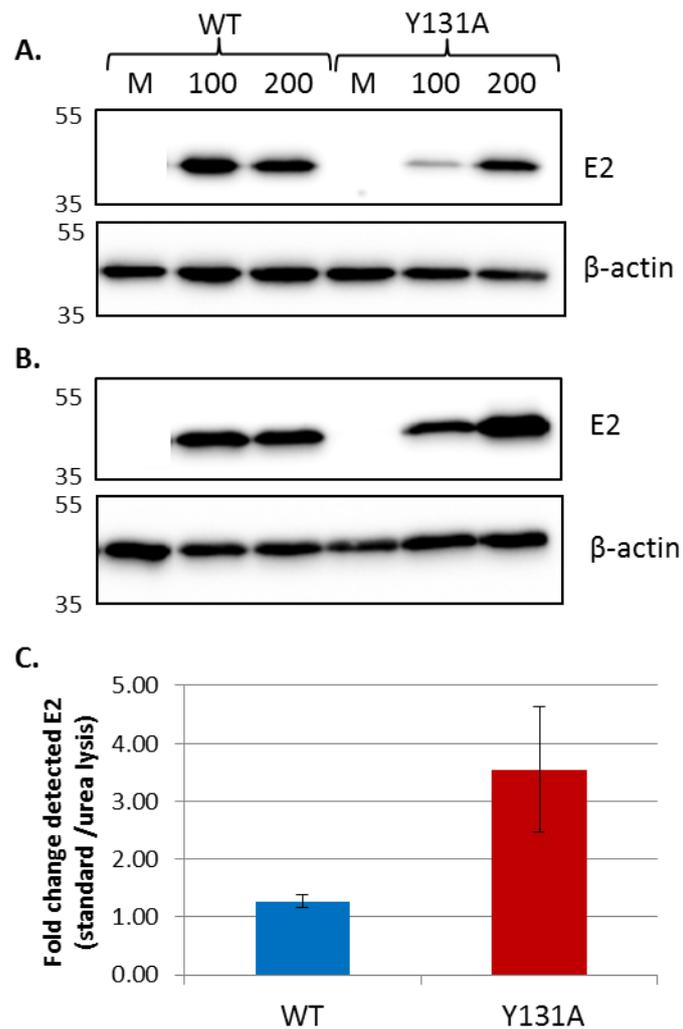
**Figure 5.4 HPV16 E2 Y131A is replication competent**

A. C33a cells were transfected with a HPV31-Ori containing plasmid (pFLOri, see figure 3.3) along with plasmids expressing *renilla* luciferase, HA-tagged HPV16 E1 and increasing amounts of WT or Y131A E2 (10-100 ng). Representative western blots of cell lysates show expression of HA-E1, E2 and  $\beta$ -actin as a loading control. Molecular weight standards are shown on the left. B. The Firefly luciferase activity in the absence and presence of increasing WT or Y131A E2 was detected using a commercially available kit. An average of three independent experiments is shown. Error bars represent standard error. Significance was determined using a two-tailed, unpaired student's T test. \*\*\*=  $p < 0.001$ .

### 5.3.2 Assessment of E2 protein stability

Having established that Y131A E2 is functionally comparable to WT E2 in activating transcription and replication, another important consideration was the stability of the mutant protein. Lower expression levels of Y131A compared to WT E2 protein were observed by western blot in the previous experiments. It was therefore hypothesised that the mutant protein was less stable than WT E2. If Y131A E2 was being removed from the cell by degradation, any effect on the HPV lifecycle could be attributed to this reduction in accessible E2 rather than specifically loss of the E2-ChIR1 interaction.

Initially, C33a cells transfected with HPV16 E2 WT or Y131A expression plasmids were lysed with either the passive lysis buffer (PLB) used with the dual-luciferase assay kit in the previous transcription and replication assays, or with a lysis buffer containing 8 M urea. Cell lysates were analysed by western blotting for E2, and the amount of E2 detected in the urea lysed samples (normalised to  $\beta$ -actin loading control) was determined compared to the E2 level in PLB lysed cells (Figure 5.5).



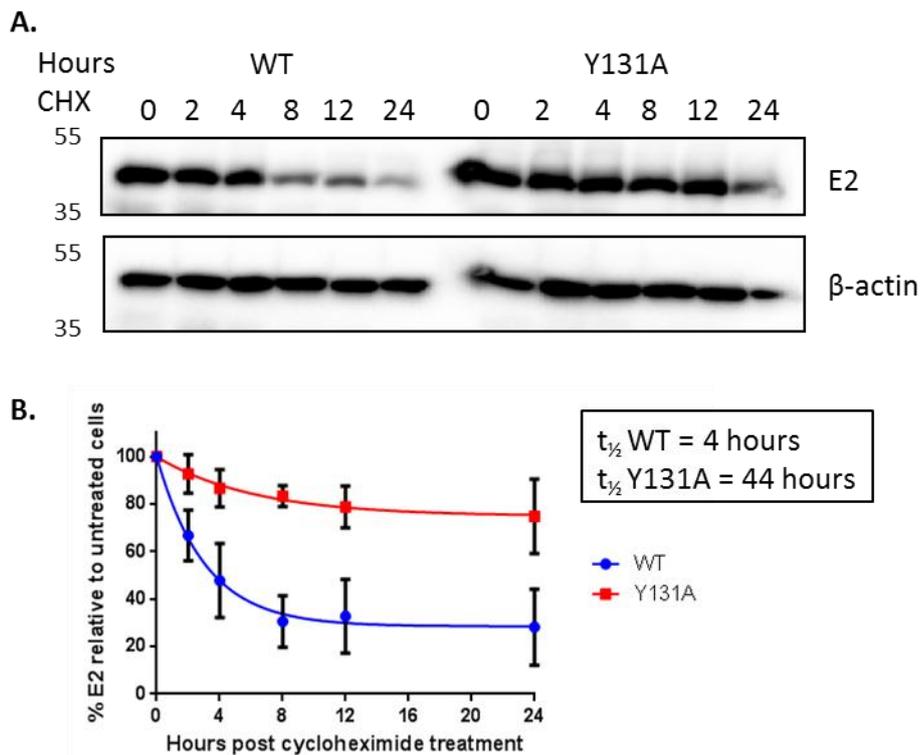
**Figure 5.5 Varying lysis conditions result in increased HPV16 Y131A E2 detected**

C33a cells were transfected with HPV16 E2 WT or Y131A expression plasmids and lysed with A. PLB or B. urea lysis buffers. Lysates were western blotted with HPV16 E2 specific antibodies and normalised to  $\beta$ -actin loading control. Representative blots are shown; molecular weight standards are indicated to the left of each blot. C. The fold change in E2 protein levels in urea lysate relative to PLB lysate (200 ng E2 plasmid transfected) was calculated by densitometry. The mean of three independent repeats is shown; error bars represent standard deviation.

Western blot analysis showed a reduction in Y131A E2 protein detected in PLB lysed cells compared to WT E2 (Figure 5.5 A), while in urea lysed cells the WT and Y131A E2 protein levels were comparable (Figure 5.5 B). Quantification of E2 protein by densitometry showed that urea lysis resulted in a 1.3 fold increase in WT E2 protein detected compared to the levels in PLB lysate, while the Y131A E2 protein detected was increased by an average of 3.5 fold in urea lysate compared to PLB lysate. This suggests that although the expression of WT and Y131A E2 protein from the transfected expression plasmid is similar, the Y131A E2 protein is not readily accessible for solubilisation by the PLB lysis buffer. This suggests a possible shift in mutant E2 protein localisation within the cell which alters the protein solubility, and could affect protein stability.

To determine the stability of exogenously expressed E2, C33a cells were transfected with WT or Y131A E2 and treated with the protein synthesis inhibitor cycloheximide (CHX) 24 hours post-transfection. CHX is a glutarimide small molecule which has been shown to inhibit protein synthesis by binding preferentially to the tRNA E-site on the ribosome large sub-unit and blocking the binding of deacetylated tRNA [180]. This stops translocation of tRNA and halts ongoing translation by preventing further elongation [181].

Following treatment with CHX, transfected cells were harvested and lysed using a urea lysis buffer (see methods) at interval time points and the E2 protein levels were determined by western blotting and densitometry (Figure 5.5). E2 protein levels were normalised to  $\beta$ -actin loading control. The E2 levels at each time point were then analysed using a non-linear regression curve which was used to calculate the half-life ( $t_{1/2}$ ) of each protein.



**Figure 5.6 Mutation of HPV16 E2 at Y131 results in increased protein stability**

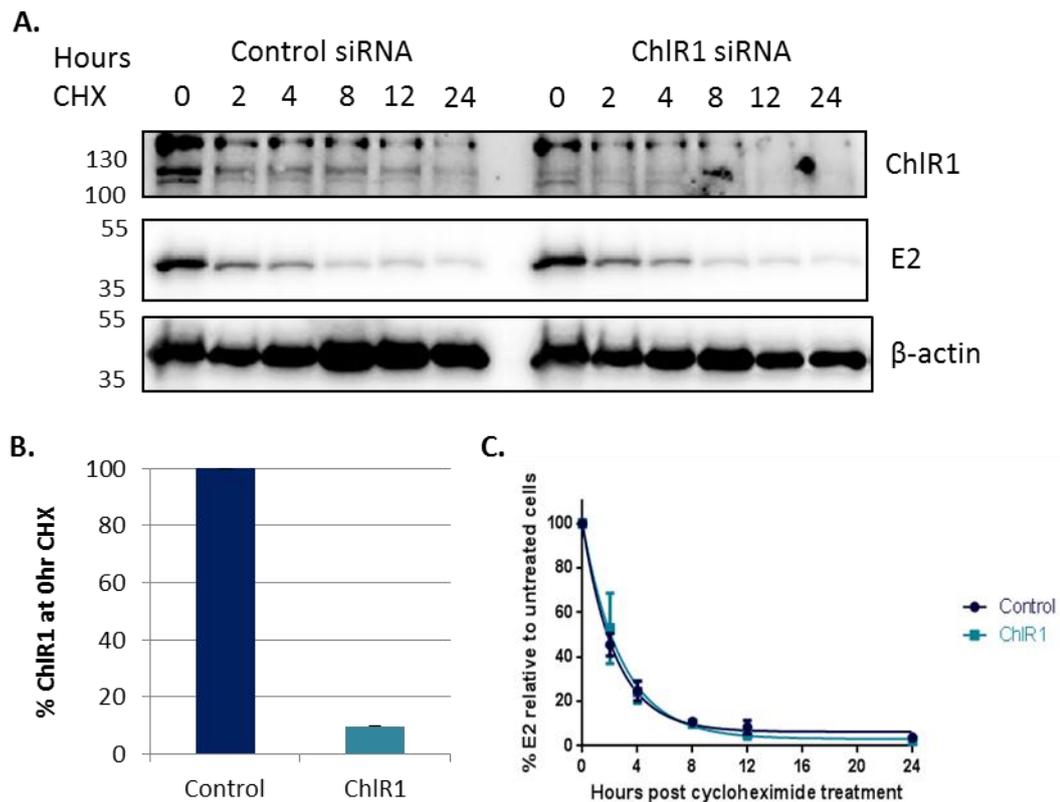
**A.** Cells transfected with HPV16 E2 WT or Y131A were treated with (10  $\mu\text{g}/\text{mL}$ ) cycloheximide and harvested at the time points shown, lysing with a urea lysis buffer. Cell lysates were analysed by western blotting using E2 specific antibodies, with  $\beta$ -actin as a loading control. Representative blots are shown. Molecular weight standards are shown at the left of each blot. **B.** E2 protein levels were determined by densitometry in three independent repeats and normalised to  $\beta$ -actin detected on the same membrane. Data were plotted as a non-linear regression curve showing E2 degradation over time and used to calculate protein half-life. Error bars represent standard error.

Following the analysis of three independent experiments, Y131A E2 was found to have a significantly increased  $t_{1/2}$  compared to WT E2 protein. WT E2 consistently had a  $t_{1/2}$  of

approximately 4 hours as is consistent with published data [182, 183], however Y131A E2 showed a  $t_{1/2}$  of approximately 44 hours, an over 10 fold increase compared to WT E2 (Figure 5.6).

This observed increase in protein stability could be due to the reduction in E2 affinity for ChR1 leading to relocalisation of the protein and subsequent increased interactions with alternative protein complexes within the cell. E2 interactions with other viral [182] and cellular proteins [108] have previously been shown to increase the stability of E2 in this way. Alternatively, this could be an effect induced by mutation of the Y131 site affecting a different process. Phosphorylation of E2 has previously been shown to have an important role in the cell-cycle dependent stability of E2 protein [184, 185]. If the tyrosine residue at position 131 is phosphorylated, mutation at this site could alter protein degradation.

To investigate this further, C33a cells were depleted of ChR1 by transfection with ChR1-specific siRNA or non-targeting control siRNA prior to transfection with WT E2 (as detailed in Materials and Methods). The cells were then treated with CHX and again were harvested at time points to analyse the protein stability over time (Figure 5.7).



**Figure 5.7 ChIR1 depletion does not alter the stability of HPV16 E2.**

C33a cells were transfected with control or ChIR1 specific siRNA before transfecting with HPV16 E2 expression plasmid. Transfected cells were treated with cycloheximide and harvested at the time points shown. A. Cell lysates were analysed by western blotting with ChIR1 and HPV16 E2 specific antibodies, and  $\beta$ -actin was used as a loading control. One representative set of blots are shown. Molecular weight standards are shown on the left of each blot. B. Densitometry was used to determine ChIR1 knockdown efficiency using the 0 hour sample. Shown is a mean quantification of 3 independent experiments. Error bars represent standard error of the mean. C. Densitometry was used to calculate E2 protein levels in three independent repeats. Data were plotted as a non-linear regression showing E2 protein degradation over time. Error bars represent standard error of the mean.

ChIR1 knockdown efficiency was determined by densitometry of western blots of the untreated (0 hour CHX) samples and normalised to  $\beta$ -actin. The ChIR1 knockdown efficiency was, on average, approximately 90% compared to control siRNA-treated cells. However, assessment of E2 protein stability revealed no difference between control and ChIR1-depleted cells. In both cases the protein  $t_{1/2}$  was slightly lower than in the WT E2 samples not treated with siRNA (Figure 5.7) and averaged approximately 3 hours compared to 4 hours in the untreated cells.

### **5.3.3 Sub-cellular E2 localisation**

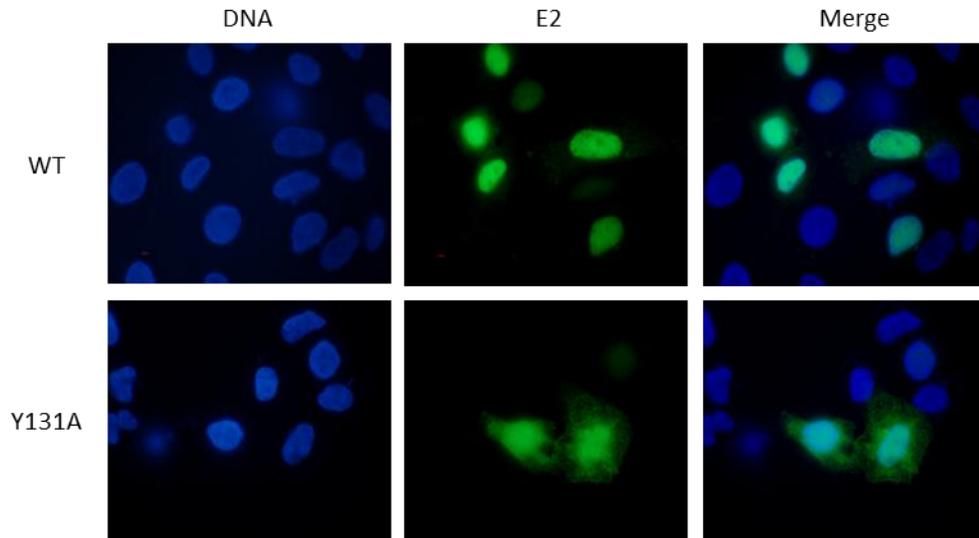
The BPV-1 E2-ChIR1 interaction has been shown to be important for E2 localisation, and previous studies suggest that this ChIR1-mediated localisation of E2 to the mitotic chromatin is involved in tethering the viral genome to sister chromatids to ensure even nuclear segregation during mitosis [95, 96]. Loss of the E2-ChIR1 interaction could therefore potentially lead to relocalisation of E2 within the cell.

### **Immunofluorescence**

#### *Immunofluorescence in asynchronous transfected cells*

Initially the localisation of WT and Y131A E2 was studied by immunofluorescence (IF). C33a cells growing on coverslips were transfected with WT or Y131A E2 expression plasmids. Coverslips were harvested and the cells formaldehyde fixed before staining with E2-specific antibodies to allow visualisation of the protein localisation within the cells (Figure 5.8). DNA

was stained with Hoechst 33342 (DNA-specific bis-benzamide intercalating dye) to provide a visual reference point within the cells.



**Figure 5.8 HPV16 E2 Y131A shows altered sub-cellular localisation**

**Cells transfected with WT or Y131A E2 were grown on coverslips. Cells were fixed in 4% formaldehyde before staining with HPV16 E2 (green) and Hoechst (DNA, blue). Images shown are representative of multiple experiments.**

In transfected C33a cells, WT E2 localisation was observed as intense nuclear staining with diffuse cytoplasmic staining visible in some cells, as is consistent with the published literature [124, 133, 186]. However, in cells transfected with the E2 mutant Y131A there is a marked difference in the protein localisation. Although the intense nuclear staining seen in WT cells was conserved there was an increase in punctate, cytoplasmic staining in many of the Y131A E2 positive cells (although some cells showed the majority of E2 protein confined to the nuclear compartment as observed with WT E2). This change in staining is indicative of

a shift in protein localisation, potentially showing increased interactions with elements of the cytoskeleton.

#### *E2 localisation in mitotic cells*

Analysis of the analogous BPV-1 E2 mutant W130R by IF showed a loss of E2 localisation to mitotic chromosomes, supporting the hypothesis that the E2-ChIR1 interaction is important for viral genome tethering [95, 96, 133]. It was therefore hypothesised that the HPV16 E2 mutant Y131A would similarly be unable to associate with mitotic chromosomes due to loss of ChIR1 binding.

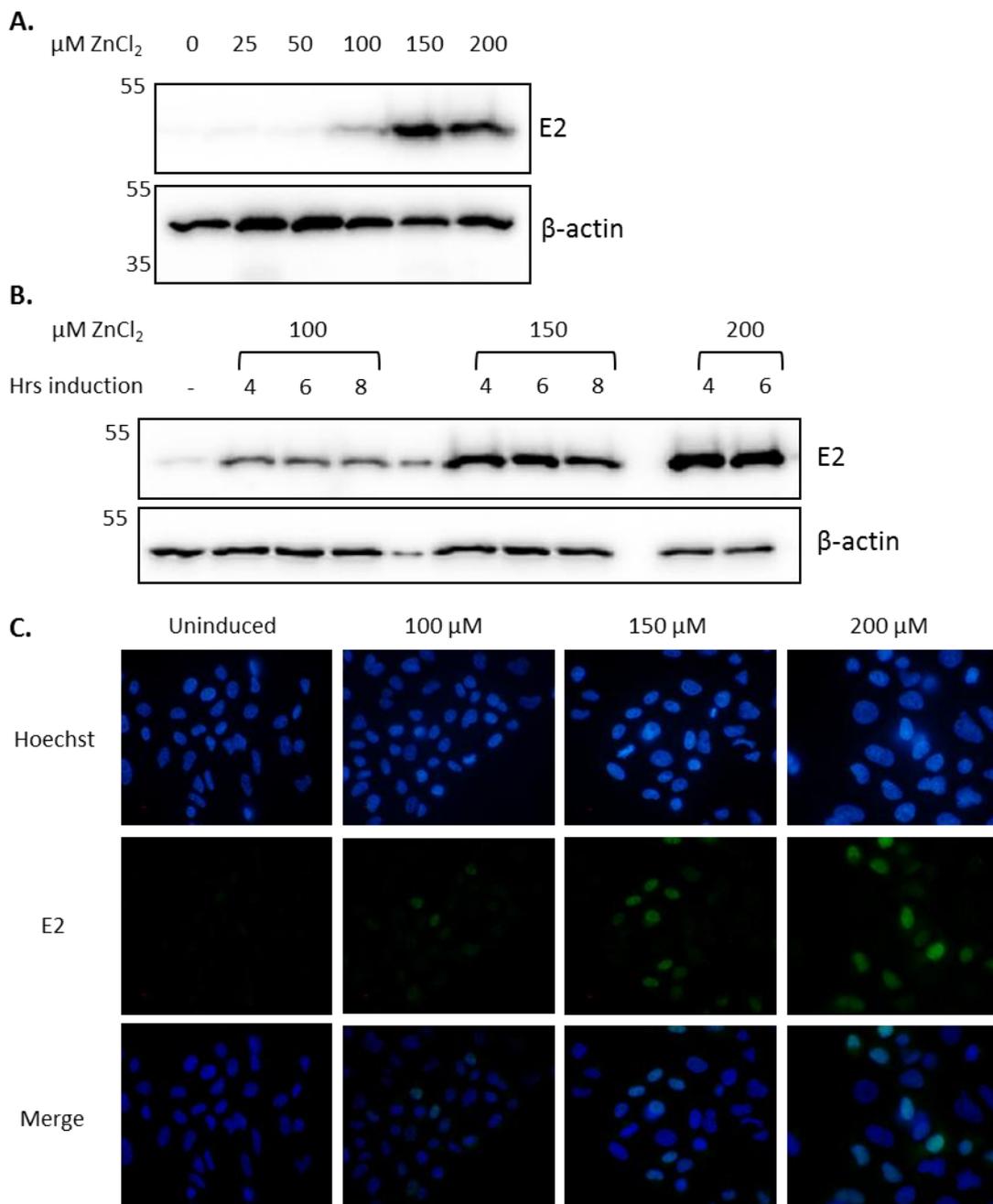
C33a cells growing on coverslips were transfected with WT or Y131A E2 expression plasmids. To increase the mitotic cell population, transfected cells were subjected to double thymidine block to synchronise cells at the G<sub>1</sub>/S boundary. Cells were then released for 8 hours to allow mitotic entry as observed by microscopic analysis and were formaldehyde fixed prior to staining. Unlike BPV-1 E2 which is observed to localise in foci on chromosomes throughout mitosis, HPV16 E2 is only observed in chromatin-associated foci at early and late stages of mitosis [93, 108]. Therefore, determining the localisation of the  $\alpha$ -HPV E2s in mitotic cells is a challenging process. Unlike  $\beta$ -HPV types (e.g. HPV8) where E2 can be observed as foci localised to the chromatin, HPV16 E2 is difficult to observe even after pre-extraction to remove cytoplasmic protein and reveal previously masked epitopes [93].

Although double thymidine block and release of cells prior to harvest resulted in an increase in mitotic cell population, the proportion of these at each stage of mitosis was relatively low. This, coupled with low transfection efficiencies, resulted in insufficient E2-positive mitotic

cells to allow any conclusions to be drawn. Treating transfected cells with the microtubule inhibitor nocodazole to block cells in early mitosis also provided similar results.

To counter the issues with low transfection efficiencies and therefore increase the total number of E2 positive cells within a sample, stable inducible cell lines were established. The Y131A mutation was cloned into a pMEP-16 E2 vector (from Alison McBride, NIH, USA) which expresses a FLAG-tagged codon-optimised HPV16 E2 under control of a metallothionein IIa promoter and contains a hygromycin B resistance gene for selection. C33a cells were transfected with pMEP-16E2 WT and Y131A and following hygromycin B selection, colonies were grown up to generate stable cell lines which were grown in hygromycin B-supplemented medium to ensure cells maintained the pMEP plasmid as an extra-chromosomal element. HPV16 E2 expression was then induced cells by treatment with zinc chloride to activate transcription from the metallothionein promoter.

The pMEP-16E2 WT cell line was used to optimise the induction conditions by varying the concentration of zinc chloride and the time induced for and analysing E2 expression levels by western blot and IF (Figure 5.9).



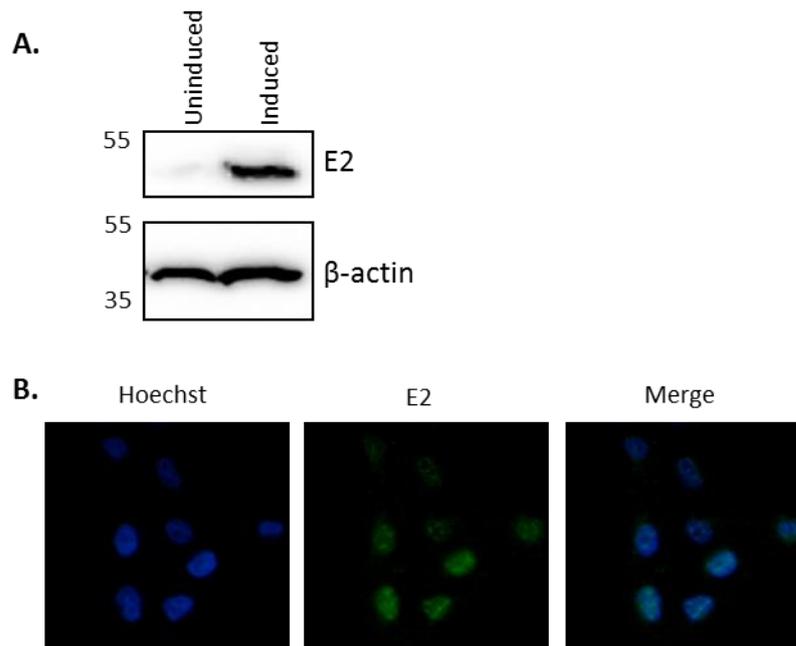
**Figure 5.9 Optimisation of the induction of HPV16 E2 expression from a stable cell line**

**A.** Stable pMEP-16E2 WT cells were treated with increasing concentrations of zinc chloride for 4 hours to induce E2 expression. Cell lysates were analysed by western blot for E2, using  $\beta$ -actin as a loading control. Molecular weight standards are shown to the left of each blot. **B.** pMEP-16E2 WT cells were treated with increasing concentrations of zinc

**chloride for 4, 6 or 8 hours before harvesting to induce E2 expression. E2 expression was determined by western blot. C. pMEP-16E2 WT cells were induced with increasing zinc chloride for 4 hours. Cells were fixed and stained for E2 (green) and DNA (blue) and analysed by IF.**

Initially cells were treated with increasing zinc chloride for 4 hours (Figure 5.9 A). Western blot analysis showed that the expressed E2 increased with ZnCl<sub>2</sub>, with 150-200 µM giving the highest expression. Cells were then treated with 100, 150 or 200 µM ZnCl<sub>2</sub> for 4, 6 and 8 hours to further optimise induction. Western blots (Figure 5.9 B) showed that again the level of E2 expression increased with the ZnCl<sub>2</sub> concentration but to the detriment of cell viability. Cells treated with 100 µM ZnCl<sub>2</sub> did not appear to suffer any significant increase in cell death whereas the majority of those treated with 200 µM ZnCl<sub>2</sub> had died before the 8 hour time point. Cells treated with increasing ZnCl<sub>2</sub> were also analysed by IF, staining with an E2 specific antibody. As increasing the induction time from 4 hours had no significant effect on the E2 expression (Figure 5.9 B), only the 4 hour time point was used. E2 staining can be detected in cells induced with each concentration of ZnCl<sub>2</sub> but the intensity of the visible staining increased with the ZnCl<sub>2</sub> concentration (Figure 5.9 C). Interestingly, despite the cells being a stable line maintaining the pMEP-16E2 WT plasmid required to allow continuous culture in hygromycin B supplemented media, E2 staining is not visible in every cell.

Weighing up the detected protein expression levels against the effects of the ZnCl<sub>2</sub> induction on cell viability, treating with 150 µM ZnCl<sub>2</sub> for 4 hours was determined to be the optimum condition for E2 induction. These conditions were then used with the pMEP-16E2 Y131A cell line to check the protein expression levels by western blot and IF (Figure 5.10).



**Figure 5.10 Induction of HPV16 E2 Y131A expression from a stable cell line**

**A.** pMEP4-HPV16 E2 Y131A cells were treated with 150  $\mu$ M ZnCl<sub>2</sub> for 4 hours to induce protein expression. Induced and uninduced cell lysates were analysed by SDS PAGE and western blotting for HPV-16 E2 and  $\beta$ -actin loading control. Molecular weight markers are indicated to the left of each blot. **B.** Induced pMEP4-HPV16 E2 Y131A cells grown on coverslips were fixed with 4% formaldehyde and stained with HPV16 E2 specific antibodies. DNA was Hoechst staining. Cells were analysed by fluorescence microscopy.

Despite the relatively poor staining compared to transiently transfected cells, having optimised the induction conditions for E2 expression, the pMEP cells were used to look at the localisation of E2 in mitotic cells. As previously, cells were enriched for mitotic populations by double thymidine block and release. E2 expression was induced by treatment with 150  $\mu$ M ZnCl<sub>2</sub> 4 hours prior to harvest, however the combination of exposure to

hygromycin, thymidine and zinc chloride led to excessive cell death before cells could be harvested and stained.

As IF experiments were not proving successful in allowing E2 localisation to be determined, alternative approaches became necessary. In place of looking at the E2 localisation in whole cells, subcellular fractionation techniques were employed to sequentially extract different protein compartments and analyse the distribution.

### **Subcellular fractionation**

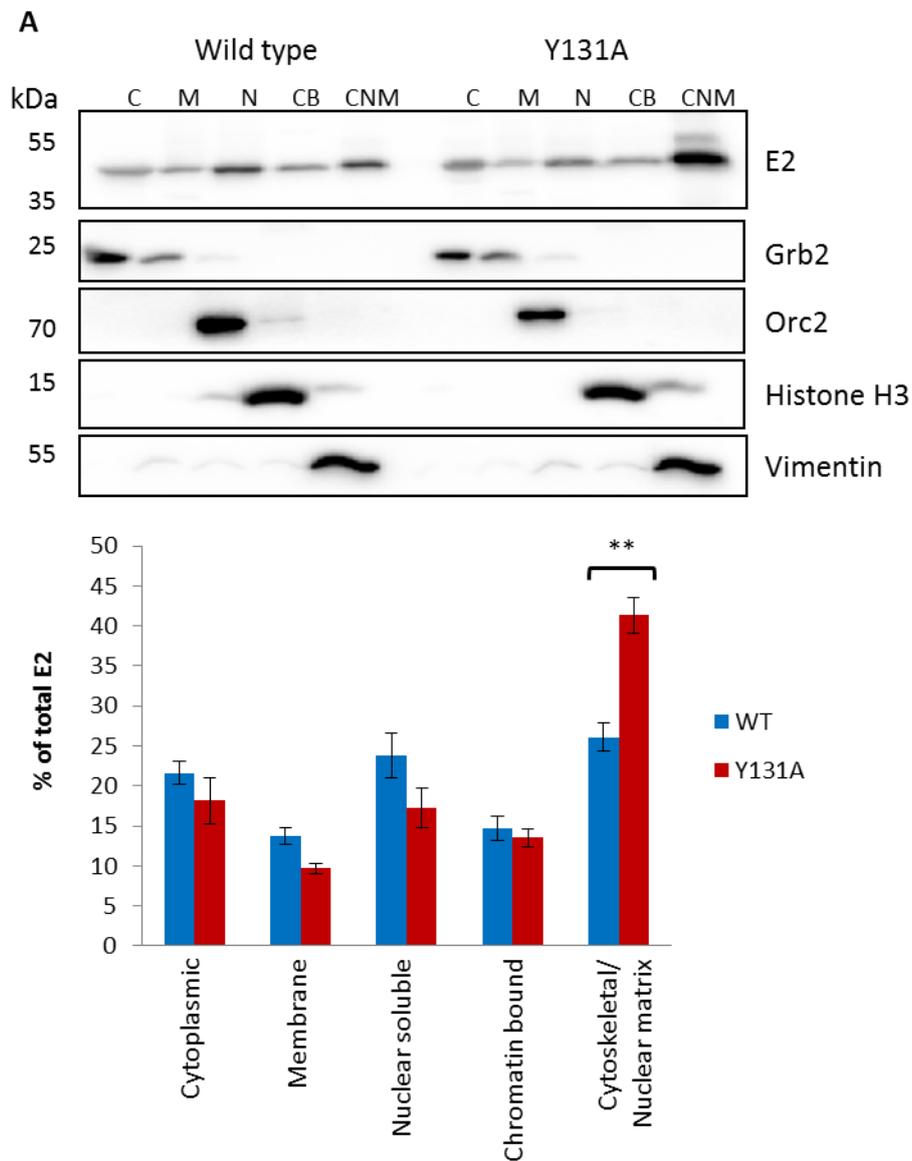
#### *Comparison of the sub-cellular localisation of WT and Y131A E2*

Although the visualisation of E2 protein localisation in mitotic cells was problematic, the IF carried out in asynchronous cells showed a marked difference in the localisation of WT and Y131A E2 protein (Figure 5.8). While WT E2 was predominantly nuclear in the majority of cells, the Y131A protein showed an increased punctate cytoplasmic localisation. To investigate this altered localisation in more detail, C33a cells were transfected with WT or Y131A E2 expression plasmids and a commercial sub-cellular fractionation kit was used to sequentially extract the different sub-cellular fractions. This method sequentially removes the cytoplasmic protein fraction (C), membrane (M), nuclear soluble (N), chromatin associated (CB) and finally the cytoskeletal and nuclear matrix (CNM) fractions. Samples were then analysed by western blotting for E2 along with markers of the various subcellular fractions to confirm the fractionation efficiency (Figure 5.11).

Growth factor receptor-bound protein 2 (Grb2) was used as a marker for the cytoplasmic fraction, origin recognition complex subunit 2 (Orc2) as a nuclear soluble protein marker, the

nucleosome structure protein histone H3 as a chromatin associated protein marker and the intermediate filament vimentin was used as a cytoskeletal and nuclear matrix marker [187, 188]. Western blots revealed good separation of these marker proteins into the expected fractions, indicating that the fractionation was successful. Representative blots are shown in Figure 5.11.

Densitometry analysis of the E2 western blots allowed the percentage of total E2 in each fraction to be determined. Comparing the distribution of WT and Y131A E2 in three independent experiments demonstrated a reduction in the level of nuclear soluble Y131A E2, and a significant increase in Y131A E2 localised to the cytoskeletal/nuclear matrix fraction.



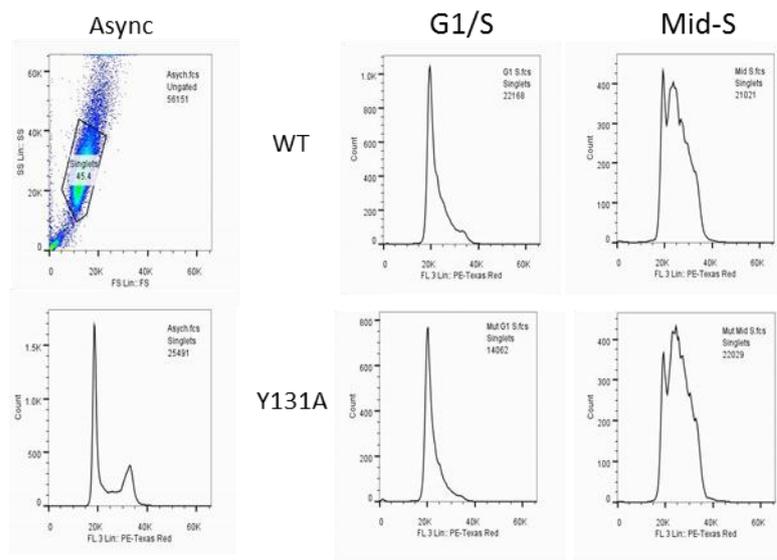
**Figure 5.11 HPV16 E2 Y131A has increased localisation to cytoskeletal and nuclear matrix fractions and decreased association with the nuclear soluble fraction**

A. C33a cells were transfected with WT or Y131A E2 expression plasmids. Cells were sequentially extracted with buffers to separate proteins localised to each cellular compartment before detecting by western blot using antibodies for E2 and markers for each subcellular fraction (Grb2 = cytoplasmic, Orc2 = nuclear soluble, Histone H3 = chromatin associated, vimentin = cytoskeletal). Representative blots are shown. Fractions

are cytoplasmic (C), membrane (M), nuclear soluble (N), chromatin bound (CB) and cytoskeletal/nuclear matrix associated (CNM). B. The amount of fractionated E2 was quantified by densitometry. E2 in each fraction is displayed as a percentage of the total fractionated E2. Shown is an average of three independent repeats. Error bars represent standard error; significance was calculated using a 2-tailed, unpaired student's T-test. \*\* =  $p < 0.01$

#### *Analysis of cell-cycle dependent E2 localisation*

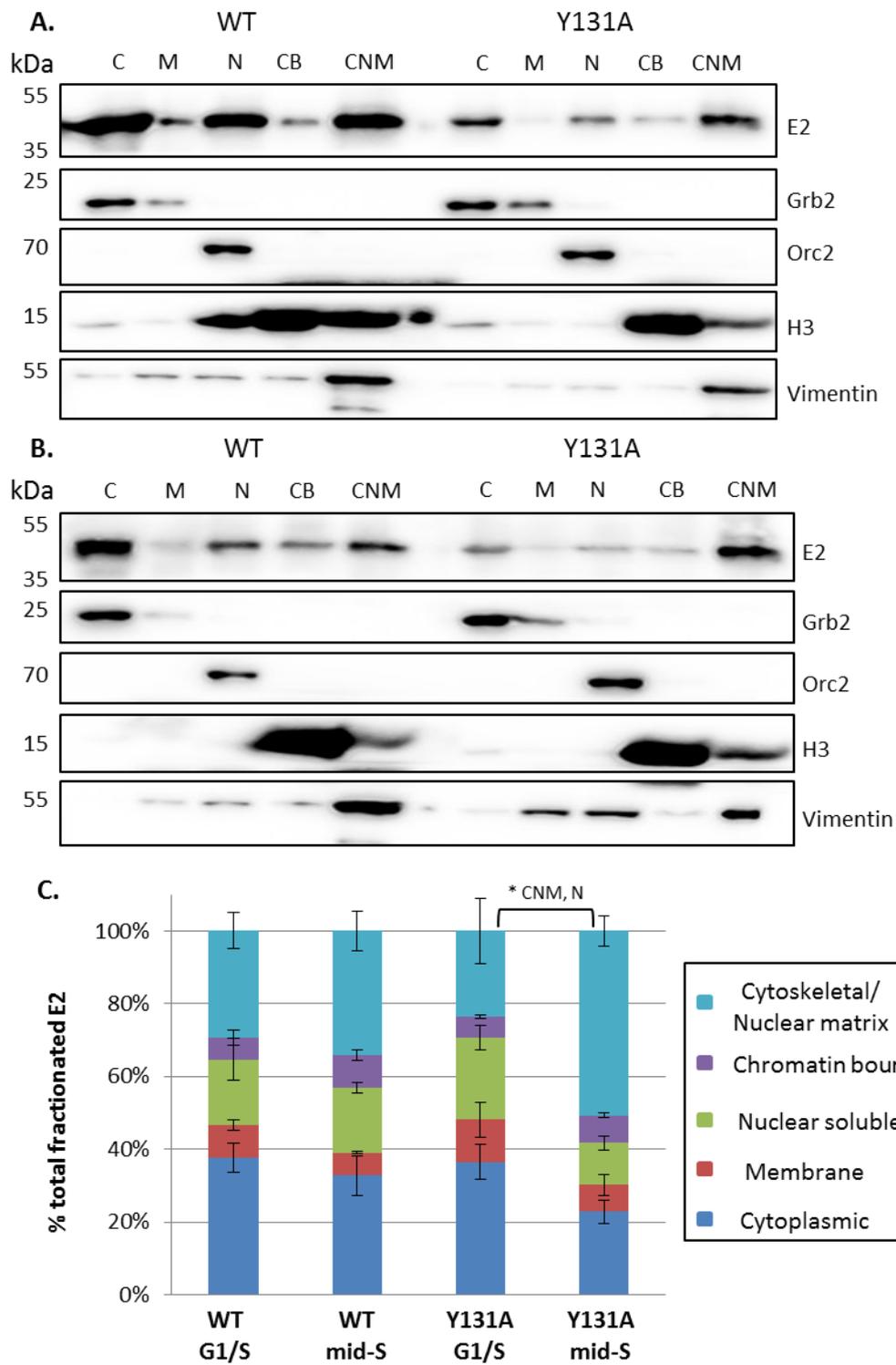
The BPV-1 E2-ChR1 interaction has previously been shown to occur predominantly within mid-S phase [133]. As such it would be expected that any changes in E2 localisation due to a loss of this interaction would be observed primarily at this point in the cell cycle. To study the cell cycle dependence of the observed change in Y131A E2 sub-cellular localisation, transfected C33a cells were synchronised at the G<sub>1</sub>/S boundary by double thymidine block, or mid-S phase by double thymidine block followed by a three hour release prior to fractionation. A sample of the harvested cells was reserved to confirm synchrony by propidium iodide (PI) staining and analysis by flow cytometry (Figure 5.12). PI fluorescence profiles obtained after gating around the singlet population showed good synchrony to the required phase when compared to an asynchronous cell population.



**Figure 5.12 Synchronisation of cells at the G<sub>1</sub>/S boundary and mid-S phase**

**Propidium iodide fluorescence of cell populations obtained by flow cytometry with cells transfected with WT or Y131A HPV16 E2 and synchronised by double thymidine block (G<sub>1</sub>/S) and release (mid-S).**

The remaining cells were biochemically fractionated as previously described and analysed by western blotting for E2 and the marker proteins described above. Representative western blots are shown in Figure 5.13.



**Figure 5.13 Cell-cycle dependent sub-cellular localisation of HPV16 E2 WT and Y131A**

**Representative western blots showing the fractionation of WT and Y131A E2 in A. G<sub>1</sub>/S and B. mid-S phase synchronised cells, along with markers of each subcellular fraction. Molecular weight standards are indicated on the left of each blot. Fractions are cytoplasmic (C), membrane (M), nuclear soluble (N), chromatin bound (CB) and cytoskeletal/nuclear matrix associated (CNM). C. The percentage of WT and Y131A E2 protein in each fraction was determined by densitometry. The average of three independent experiments is shown. Error bars represent standard error; significance was determined using a 2-tailed, unpaired student's T-test, \* p<0.05.**

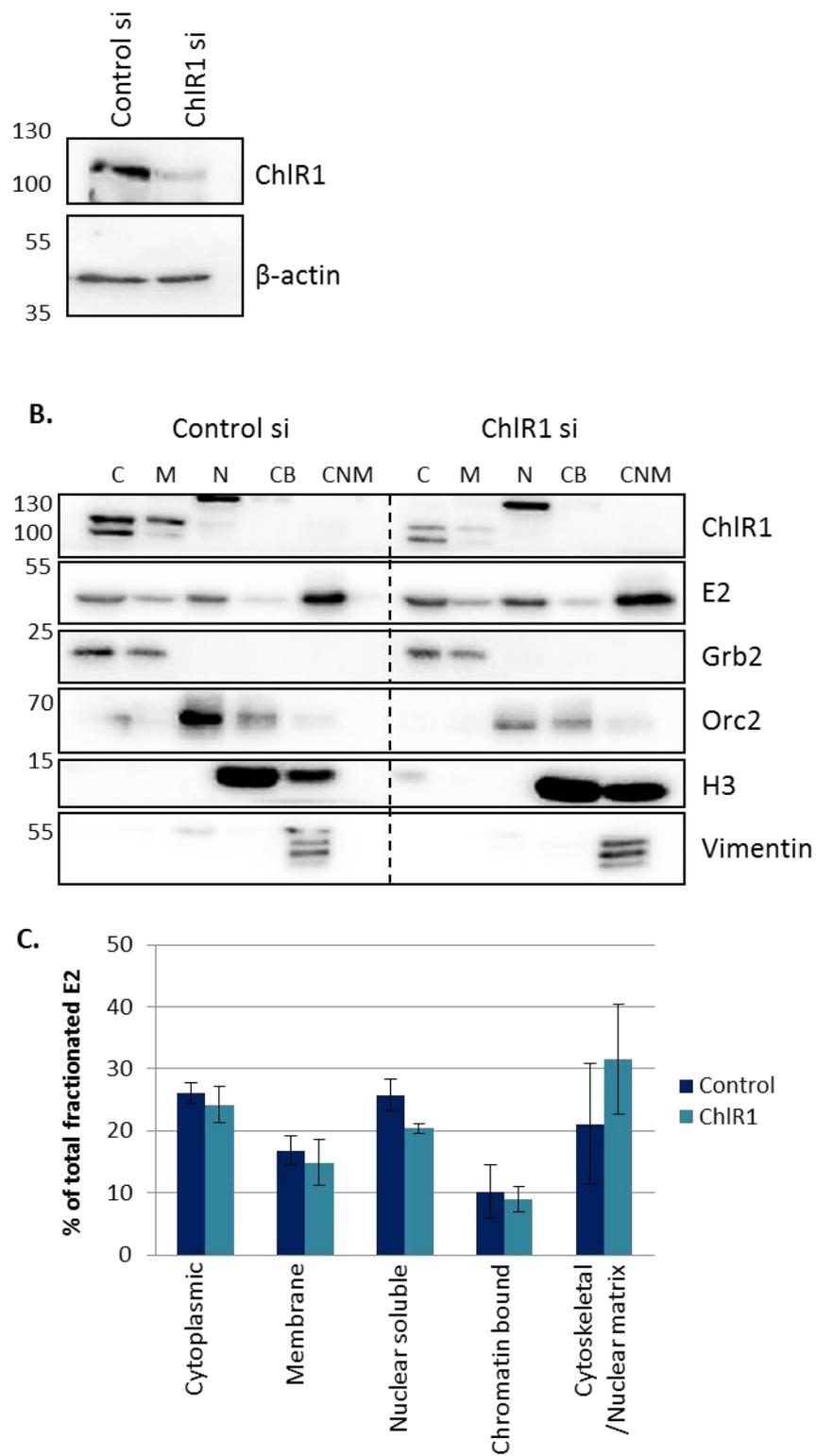
As previously shown, western blots for the marker proteins showed reasonable separation into the expected fractions. The percentage of the total fractionated E2 in each fraction was determined by densitometry and the average of three experimental repeats is shown in Figure 5.13 C.

Comparison of cells synchronised at the G<sub>1</sub>/S boundary and at mid-S phase revealed no change in the localisation of WT E2. In addition, the distribution of WT E2 between fractions remained consistent with that observed in the previous fractionation experiments using asynchronous cells (Figure 5.11). However, in G<sub>1</sub>/S synchronised cells, Y131A E2 displays a similar distribution to WT E2, but in cells synchronised to mid-S phase there is a significant change in Y131A E2 localisation. A significant reduction in nuclear soluble protein and an increase in cytoskeletal/nuclear matrix protein was observed, again consistent with the change in distribution observed in asynchronous cells.

*Investigating the sub-cellular localisation of HPV16 E2 in cells depleted of ChIR1*

To further investigate whether the change in Y131A E2 sub-cellular localisation is caused by the loss of E2-ChIR1 interaction, subcellular fractionation experiments were carried out to determine the sub-cellular localisation of WT E2 in cells depleted of ChIR1.

C33a cells were treated with control or ChIR1 specific siRNA prior to transfection with WT E2 expression plasmid. Cells were harvested and fractionated as before, and the fractionation efficiency was determined by western blot analysis of the fraction markers. ChIR1 knockdown efficiency was also determined by western blot analysis of whole cell lysate with a ChIR1-specific antibody, normalised to the  $\beta$ -actin loading control. Sub-cellular E2 localisation was determined by western blot analysis of the biochemical fractions and the percentage of total fractionated E2 in each fraction was determined by densitometry (Figure 5.14).



**Figure 5.14 Increased cytoskeletal/nuclear matrix localisation of HPV16 E2 following ChR1 depletion**

C33a cells were treated with control or ChIR1 specific siRNA 24 hours prior to transfection with HPV16 E2 expression plasmid. Cells were harvested and biochemically fractionated. A. Representative western blots showing ChIR1 expression in whole cell extract. ChIR1 levels were quantified by densitometry and normalised to  $\beta$ -actin. B. Harvested cells were fractionated and samples were analysed by western blotting for ChIR1, E2 and markers of the subcellular fractions. A representative set of blots is shown. Molecular weight standards are indicated on the left of each blot. C. Fractionated E2 was quantified by densitometry. The average of three independent repeats is shown. Error bars represent standard error of the mean.

In control siRNA treated cells, WT E2 displayed a similar distribution pattern to that previously observed. When cells were depleted of ChIR1, a change in the localisation of WT E2 was observed as compared to control cells. As seen with Y131A E2, there was a decrease in nuclear soluble WT E2 protein following ChIR1 depletion with a concomitant increase in cytoskeletal/nuclear matrix associated E2 protein. In this case however the difference between control and ChIR1 siRNA treated cells was not significant (Nuclear soluble  $p = 0.06$ , CNM  $p = 0.24$ ), despite the same trend being observed.

#### *Cell cycle dependence of the localisation of HPV16 E2 in cells depleted of ChIR1*

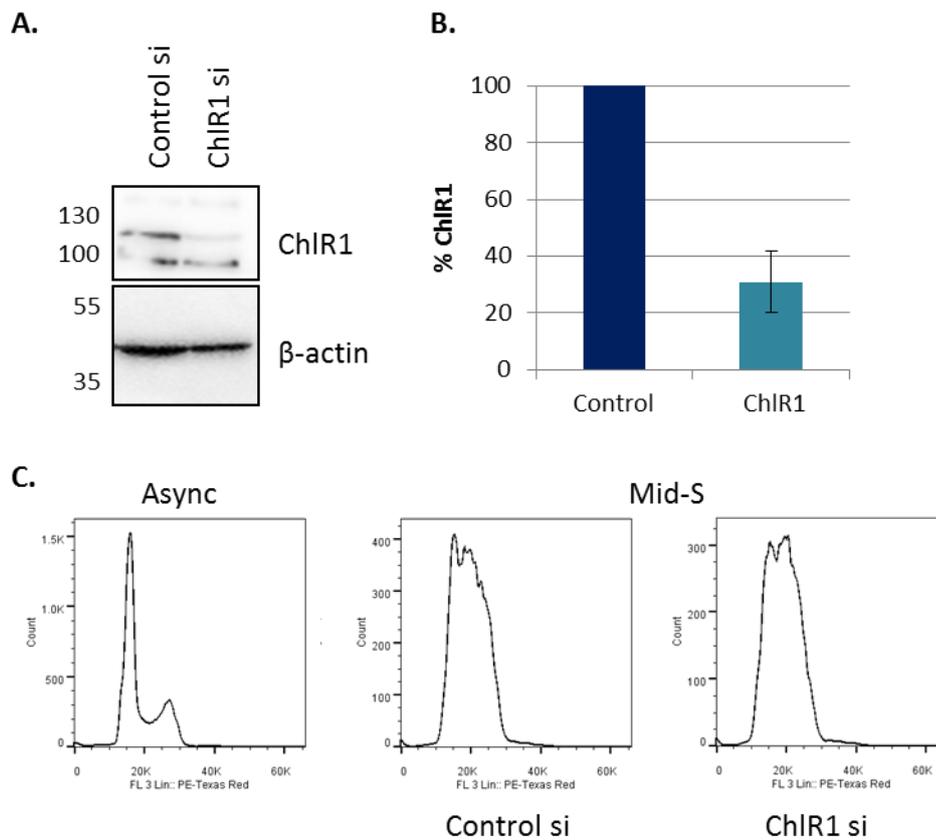
As the shift in Y131A E2 localisation was observed to be cell cycle dependent and primarily occurring in mid-S phase cells, this was again repeated in cells depleted of ChIR1. C33a cells were treated with control or ChIR1 siRNA and transfected with WT E2 before synchronising to mid-S phase by double thymidine block followed by a three hour release prior to

harvesting. ChIR1 knockdown efficiency was determined by western blot and densitometric analysis of whole cell extract normalised to loading control. Over three experiments the ChIR1 knockdown efficiency was on average approximately 70% compared to control siRNA treated cells (Figure 5.15), a reduction in knockdown efficiency of approximately 20% compared to the previous experiments (Figure 5.14).

A sample of the harvested cells was reserved to confirm synchrony by PI staining and flow cytometry as previously described. The PI fluorescence profiles obtained showed good enrichment of the mid-S phase population compared to an asynchronous control sample (Figure 5.15).

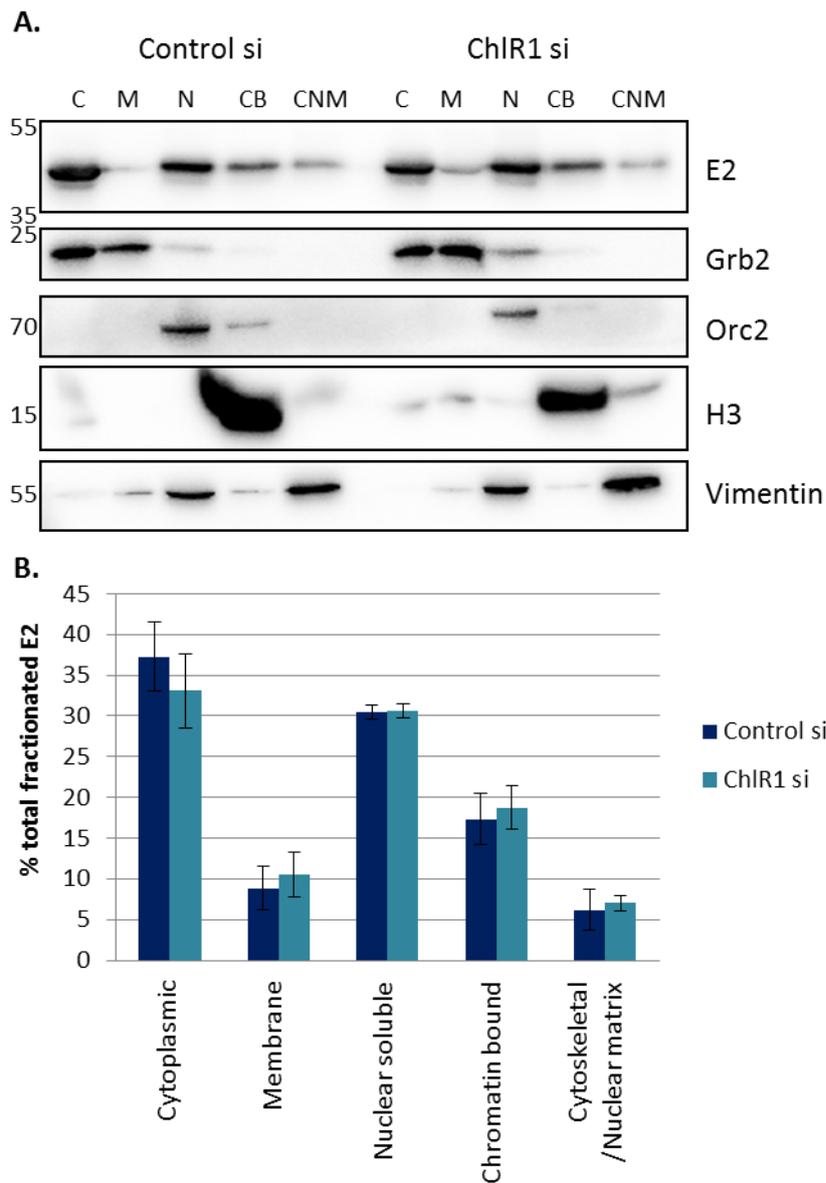
The remaining cells were biochemically fractionated as before and analysed by western blotting for the markers of the various fractions. As previously described, the distribution of E2 protein within each fraction was determined as a percentage of total fractionated E2 by densitometry (Figure 5.16). These experiments revealed that there was no difference in the observed distribution of WT E2 protein between sub-cellular fractions in mid-S phase synchronised cells treated with control or ChIR1 siRNA. In both ChIR1- and control siRNA transfected cells, an approximately equal proportion of the E2 protein was localised to the cytoplasmic and nuclear soluble fractions. Unlike in previous experiments (Figures 5.11, 5.13 and 5.14), very little E2 was localised to the cytoskeletal and nuclear matrix fraction in both control and ChIR1 siRNA transfected cells.

This possibly contributes to there being no apparent change in E2 localisation following ChIR1 depletion despite this having been previously observed in asynchronously growing cells.



**Figure 5.15 Mid-S phase synchronisation of cells treated with control or ChIR1 siRNA**

**A.** C33a cells were treated with control or ChIR1 specific siRNA before transfection with HPV16 E2 WT expression plasmid and synchronised at mid-S phase by double thymidine block followed by a three hour release. Cells were harvested 48 hours post-transfection and whole cell extracts were analysed by western blotting. A representative set of blots is shown. Molecular weight standards are indicated to the left. **B.** ChIR1 knockdown in three independent repeats was calculated by densitometry; data shown represent the mean and standard error. **C.** Cell synchrony was confirmed by flow cytometry. Ethanol fixed cells were stained with propidium iodide and cell cycle analysis was carried out by flow cytometry. Representative PI profiles are shown.



**Figure 5.16 ChR1 depletion does not affect HPV16 E2 localisation in mid-S phase synchronised cells.**

C33a cells treated with control or ChR1-specific siRNA were transfected with HPV16 E2 expression plasmid and synchronised at mid-S phase. A. Sub-cellular compartments were sequentially extracted and the fractions were analysed by western blotting with antibodies against E2 and markers of the fractions (Grb2 = cytoplasmic, Orc2 = nuclear

soluble, Histone H3 = chromatin associated, vimentin = cytoskeletal). Representative blots are shown. Fractions are cytoplasmic (C), membrane (M), nuclear soluble (N), chromatin bound (CB) and cytoskeletal/nuclear matrix associated (CNM). Molecular weight standards are indicated at the left of each blot. B. Fractionated E2 from three repeats was quantified by densitometry. The data represent the mean and standard error.

### ***In situ* fractionation**

The HPV16 E2 mutant Y131A has been shown to increase localisation of the E2 protein to the cytoskeletal/nuclear matrix associated sub-cellular fraction. A similar shift in the sub-cellular localisation of WT E2 in asynchronous cells depleted of ChR1 was also observed. To further assess this change in E2 localisation an *in situ* fractionation IF technique [136] was utilised.

C33a cells growing on coverslips were transfected with WT or Y131A E2 expression plasmids. Harvested coverslips were washed with sequential buffers to extract different protein fractions in turn. One coverslip was removed after each extraction step and the remaining material was formaldehyde fixed. Each experiment resulted in four coverslips to be stained for IF analysis: a whole cell sample fixed before any extractions; an intact nuclear sample fixed after the cytoplasmic and loosely held nuclear proteins are removed; a chromatin sample fixed after the nuclear soluble proteins are removed; and a nuclear matrix sample fixed after the chromatin has been disrupted by DNase treatment and the chromatin associated proteins are removed. The *in situ* fractionation protocol is summarised in Figure 5.17.

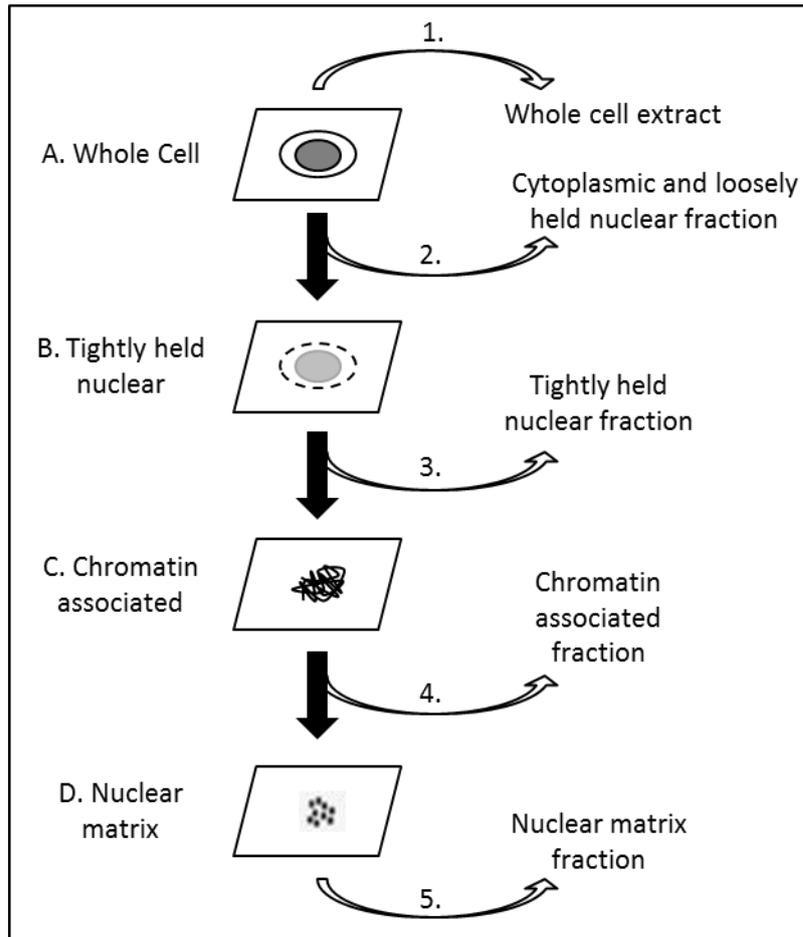


Figure 5.17 Schematic diagram of the *in situ* fractionation technique (adapted from [136]).

C33a cells were grown on coverslips and transfected as required. One coverslip was used to remove whole cell extract for western blot analysis (1.). A second coverslip was removed and fixed for IF as whole cell (A.). The remaining coverslips were treated with CSK-1 to remove cytoplasmic proteins (2.) before one coverslip was removed and fixed (B.). The remaining coverslips were then treated with CSK-2 to extract the nuclear fraction and a coverslip was removed and fixed (C.). Remaining coverslips were then treated with CSK-3 to digest DNA and extract the chromatin associated fraction and a coverslip was fixed for IF (D.). The final coverslip was treated with TES buffer to extract the nuclear

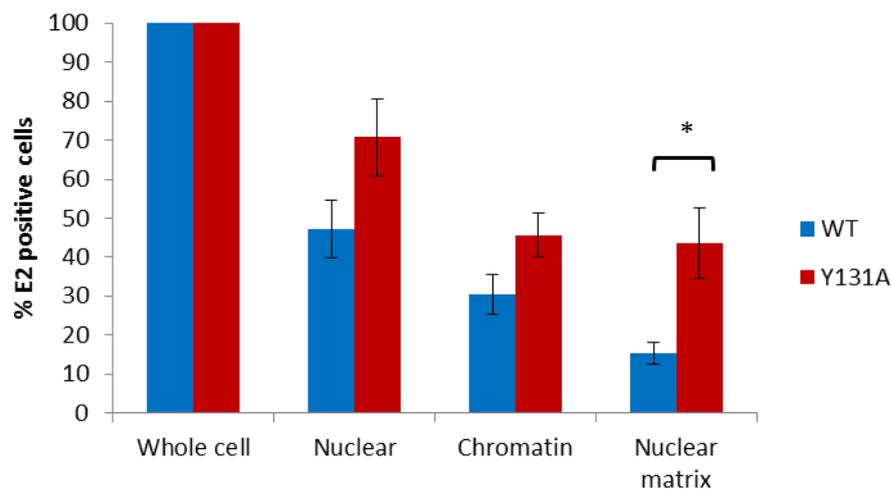
**matrix fraction. All coverslips were fixed with 4% formaldehyde for 30 minutes at 4 °C before staining.**

The formaldehyde fixed coverslips were stained with antibodies specific to HPV16 E2 and a marker of the nuclear matrix Lamin B1, and the DNA was stained with Hoechst. The number of E2 positive cells remaining in each sample after the preceding extractions was then determined by IF and the percentage of E2 positive cells in comparison to the number of E2 positive cells on a coverslip harvested from the same tissue culture dish prior to fractionation was determined. For each fraction >500 cells were counted and scored for the percentage of E2 positive cells remaining, normalised to the whole cell fraction. In cells transfected with WT HPV16 E2, removing the cytoplasmic and nuclear soluble protein fraction led to a 53% reduction in the number of E2 positive cells counted in the intact nuclear sample. Extracting the tightly held nuclear proteins led to a further 17% reduction in the number of E2 positive cells counted in the chromatin containing samples. Disruption of the chromatin by DNase treatment and subsequent removal of the chromatin associated proteins then resulted in a further 15% reduction in E2 positive cells, with only 15% of the E2 positive cells counted in the whole cell sample showing E2 remaining associated with the nuclear matrix (Figure 5.18).

In contrast, the numbers of Y131A E2 positive cells counted after each fractionation were increased compared to WT E2 positive cells. Removal of the cytoplasmic and nuclear soluble proteins led to a 29% reduction in E2 positive cells compared to the whole cell sample. This is a 24% increase in Y131A E2 positive cells compared to WT E2 containing cells at the same point, suggesting a reduction in cytoplasmic and nuclear soluble associated mutant E2 as

previously observed with the subcellular fractionation experiments (Figure 5.11). Removing the tightly held nuclear protein fractions led to a further reduction in E2 positive cells of 25%, similar to the 17% reduction observed with WT E2, suggesting that although the total number of E2 positive cells remaining was higher with Y131A E2 compared to WT (46% compared to 30%), this is not due to an increase in nuclear associated protein but rather a cumulative effect from remaining protein in the other fractions. Finally, disruption of the chromatin and removal of the associated proteins had very little effect on the numbers of E2 positive cells counted with only a 2% reduction being observed. This was a significant difference compared to the 15% reduction in WT E2 positive cells following chromatin disruption, showing that there is an increase in Y131A E2 association with the nuclear matrix which remains following fractionation leading to a 29% increase in the remaining E2 positive cells counted.

To determine whether the increased nuclear matrix association of Y131A E2 compared to WT E2 was specifically due to the loss of ChIR1 interaction, the *in situ* fractionation and immunofluorescence experiments were repeated to determine the localisation of WT HPV16 E2 in cells depleted of ChIR1. The experimental scheme was as described previously (Figure 5.17), however ChIR1 knockdown efficiency was determined by western blotting for ChIR1 and a  $\beta$ -actin loading control in the whole cell lysate sample for each experiment (Figure 5.19). ChIR1 levels in cells treated with control of ChIR1 siRNA were determined by densitometry and normalised to the loading control. An average knockdown efficiency in the region of 90% was observed.



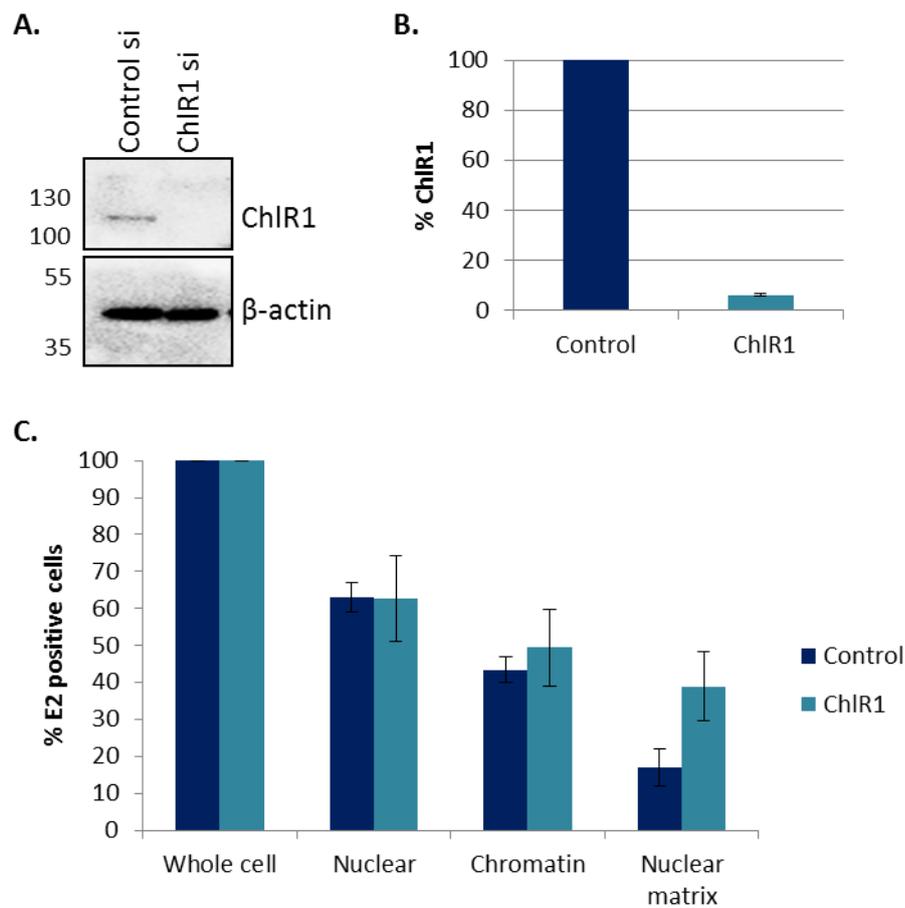
**Figure 5.18 Increased association of HPV16 E2 Y131A with the nuclear matrix**

C33a cells grown on coverslips were transfected with HPV16 E2 WT or Y131A expression plasmid. E2 localisation was determined by IF following *in situ* fractionation. Three independent repeats were quantified. For each experiment >500 cells per fraction were counted and expressed as percentage of E2 positive cells in each fraction normalised to whole cells harvested from the same transfection. Error bars represent standard error of the mean. Significance was determined using a 1-tailed student's T-test; \* $p < 0.05$ .

In cells treated with control siRNA, removal of the cytoplasmic and loosely held nuclear protein led to a 37% reduction in the number of WT E2 positive cells (Figure 5.19). This reduction in E2 positive cells was the same in ChR1 siRNA treated cells, indicating that there is no change in E2 association with the cytoplasmic protein fraction upon ChR1 depletion as was observed in the previous subcellular fractionation experiments (Figure 5.18). Removal of the tightly held nuclear proteins resulted in a further 20% reduction in the number of E2 positive control siRNA treated cells, and a similar 14% reduction in E2 positive ChR1 siRNA treated cells. This shows little difference in the association of E2 with nuclear proteins as has

previously been observed when comparing WT and Y131A E2 localisation. Finally, chromatin disruption and extraction of the chromatin-associated proteins led to a further 26% reduction in E2 positive control cells but only a 10% reduction in E2 positive ChIR1 siRNA treated cells. This is a similar trend to that observed when comparing WT and Y131A E2 localisation, but in this case the difference in E2 positive cells remaining in the nuclear matrix fraction in control and ChIR1 siRNA treated cells did not reach statistical significance (NM fraction  $p = 0.053$ ).

Together, these data show that the HPV16 E2 mutant Y131A retains functional activity (both transcription activation and replication) compared to WT E2, but displays a cell-cycle dependent change in localisation with increased association to the nuclear matrix observed along with a corresponding increase in protein stability. Although not reaching statistical significance, similar trends in the localisation of HPV16 E2 WT in cells depleted of ChIR1 were also demonstrated, supporting the suggestion that loss of the E2-ChIR1 interaction results in a shift in E2 protein localisation.



**Figure 5.19 Increased association of HPV16 E2 shows with the nuclear matrix following ChIR1 depletion.**

C33a cells were treated with control or ChIR1 siRNA before seeding onto coverslips. Cells were transfected with HPV16 E2 WT expression plasmid and coverslips were harvested for *in situ* fractionation. A. Whole cell extracts were analysed by western blotting to confirm ChIR1 knockdown. Representative blots are shown. Molecular weight standards are indicated on the left of each blot. B. ChIR1 knockdown was quantified from three independent repeats by densitometry, normalised against  $\beta$ -actin. Data shows the mean ChIR1 levels in cells treated with ChIR1 siRNA expressed as a percentage of ChIR1 level in control siRNA treated cells. Error bars represent standard error of the mean. C. IF was used

to quantify the percentage of E2 positive cells remaining following fractionation, relative to whole cell. Three independent repeats were quantified and in each experiment >500 cells were counted per fraction. Error bars represent standard error of the mean.

#### **5.3.4 Phenotypic analysis of E2 Y131A in the HPV life cycle**

##### **Generation of primary keratinocyte lines containing WT or mutant (E2 Y131A) HPV16 genomes**

The HPV life cycle is intricately linked to the differentiation status of the host epithelium. Upon initial infection of the basal cells, the virus genome is amplified and infection is established at approximately 50-200 copies per cell [28]. As infected cells move up through the epithelium, expression of the early proteins E6 and E7 drive cells into S phase and allow for amplification of the genomes to several hundred copies. As the epithelial cells differentiate the late promoter is activated and the late proteins are expressed, leading to encapsidation of the viral genomes and new virion production. E2 plays a vital role in this process, regulating the transcription of the viral E6 and E7 proteins required for viral genome amplification and triggering the switch to the late phase of the viral life cycle by regulating use of the viral polyadenylation sites [56].

In order to recapitulate the complex life cycle of HPV, a model system using primary human foreskin keratinocytes (HFKs) was developed [178]. Cell lines can be generated which retain episomal HPV genomes. These cells can be maintained in monolayer culture to study the early stages of the viral life cycle or they can be encouraged to differentiate by growth in organotypic raft culture, allowing the late stages of the life cycle to be studied.

To investigate the role of the E2-ChIR1 interaction in the context of the viral life cycle, the E2 Y131A mutation was cloned into a pUC19 vector containing the HPV16 genome clone 114/K [189] (obtained from Ethel-Michele de Villiers, DKFZ, Germany). The cloned genome was sequenced using overlapping primers before restriction digest with *Bam*HI to excise the viral genome from the vector backbone. Complete digestion was confirmed by gel electrophoresis before the digested genome was religated and ethanol precipitated to produce purified, recircularised HPV genomes.

Primary human foreskin keratinocytes from two independent donors (Nigel and Jerry) were transfected with the recircularised WT or mutant HPV16 genomes along with a neomycin resistance marker. Following G418 selection, colonies were pooled and expanded, establishing 4 cell lines containing the HPV genomes – Nigel WT, Nigel mutant, Jerry WT and Jerry mutant. These cell lines were cultured in monolayer for multiple passages, collecting cell pellets at each passage to allow for downstream analysis. Cell pellets were lysed and the DNA extracted to allow comparison of the HPV genome copy number and physical status between the WT and mutant HPV lines.

#### **Determination of HPV copy number in HFK lines**

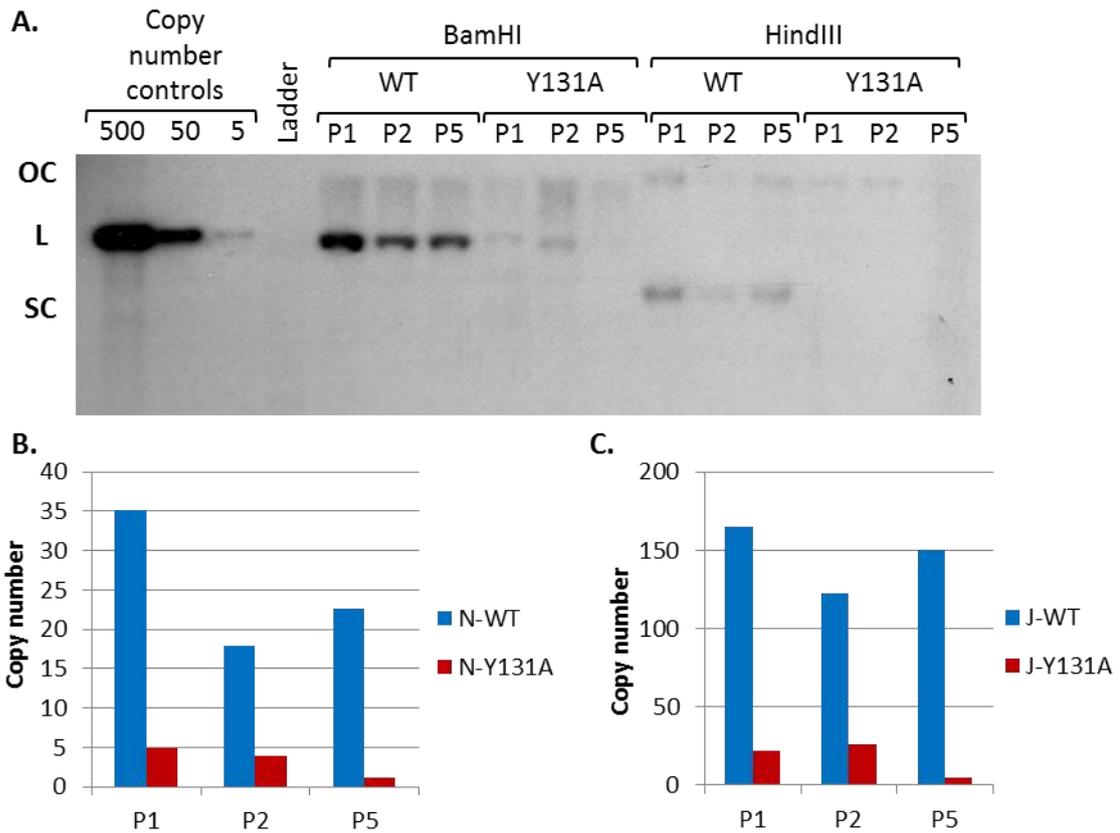
Following initial HPV16 infection, the viral genome replication is established and genomes maintained at ~50-200 copies per cell [28]. The preceding work using transient transfection of HPV16 E2 has shown that loss of ChIR1 interaction results in a corresponding change in E2 protein localisation but that the functional activity (transcription activation and replication) of E2 is unaffected. If the hypothesis that the E2-ChIR1 interaction is important for the tethering of viral genomes to chromatin and therefore the nuclear segregation of genomes

into daughter cells is correct, loss of this interaction and the associated re-localisation of E2 would lead to a progressive reduction in the viral genome copy number in primary keratinocytes.

To test this hypothesis, DNA was extracted from WT and mutant HFK lines at passage 1, 2 and 5 to look at the maintenance of episomal HPV genomes in each line. Hirt extracted DNA was digested with restriction enzymes that cut within the viral genome (*Bam*HI) to linearise episomal DNA, or with an enzyme which does not cut the viral genome but does cut the human genome to reveal any integrated viral DNA (*Hind*III). The digested DNA samples were analysed by Southern blotting using a HPV16 specific <sup>32</sup>P-labelled probe (Figure 5.21).

Copy number control samples were generated by dilution of HPV16 genome DNA obtained by restriction digest from the pUC19-HPV16 plasmid. Controls for 5, 50 and 500 copies were run alongside the HFK-extracted DNA samples for analysis (Figure 5.20). The hybridised radiolabelled probe was detected by autoradiography after 24 and 72 hours (not shown) exposure. This was carried out with DNA samples extracted from all 4 HFK lines. The Southern blot shown is representative of the results from both donors.

The copy number for each linearised sample was calculated by densitometry and compared to the 50 copies per cell control. In both donor lines, there was a marked reduction in the copy number of the mutant genome compared to the WT. This reduction was apparent from the passage 1 (P1) samples and the HPV copy number was not recovered to WT levels by passage 5 (P5), suggesting that the mutant genomes fail to establish to WT levels even after extended passage.

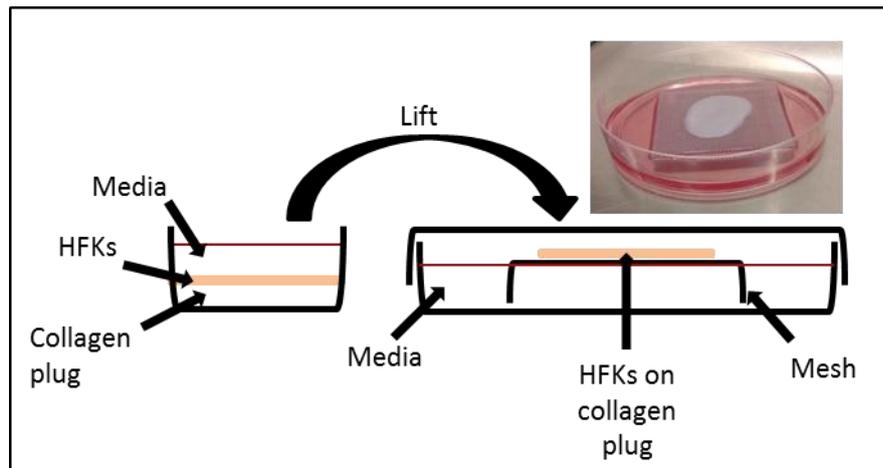


**Figure 5.20 Reduction in HPV16 mutant genome copy number compared to WT genome copy number in HFK lines**

DNA extracted from HFK lines containing HPV16 WT or Y131A-E2 mutant genomes was analysed by southern blotting. Extracted DNA from cells at P1, 2 and 5 was digested with specific (*Bam*HI) or non-specific (*Hind*III) restriction enzymes and the HPV genomes were analysed by Southern blotting with a <sup>32</sup>P-labelled HPV16 specific probe. A. HPV16 DNA was detected by autoradiography. Shown is a representative 24 hour exposure of samples from one donor (Nigel). B. The copy number of WT and mutant HPV genomes in HFKs from donor 1 (Nigel) was determined by densitometry with the linearised samples normalised to the 50 copies per cell control sample. C. The copy number of WT and mutant HPV

genomes in HFKs from donor 2 (Jerry) was determined by densitometry with the linearised samples normalised to the 50 copies per cell control sample.

To further investigate the function of the E2-ChIR1 interaction, the WT and Y131A E2 mutant HFK lines generated were grown in organotypic raft culture as a physiologically relevant model to recapitulate the viral life cycle (Figure 5.21). As the HPV16 mutant genome-containing lines were demonstrated to be unable to establish episomes at high copy number compared to WT genome lines (Figure 5.20), it was hypothesised that raft cultures grown from these mutant HFK lines would not display the increased proliferative phenotype which has been previously demonstrated with HPV16 WT genome containing lines.



**Figure 5.21 Schematic illustration of organotypic raft culture.**

HFKs are seeded onto a collagen plug containing J2 feeder cells. After the HFKs monolayer becomes confluent (3-4 days) the collagen plug is lifted onto a mesh grid, suspending it at the liquid-air interface. This encourages differentiation of the HFK cells allowing the HPV life cycle to be recapitulated.

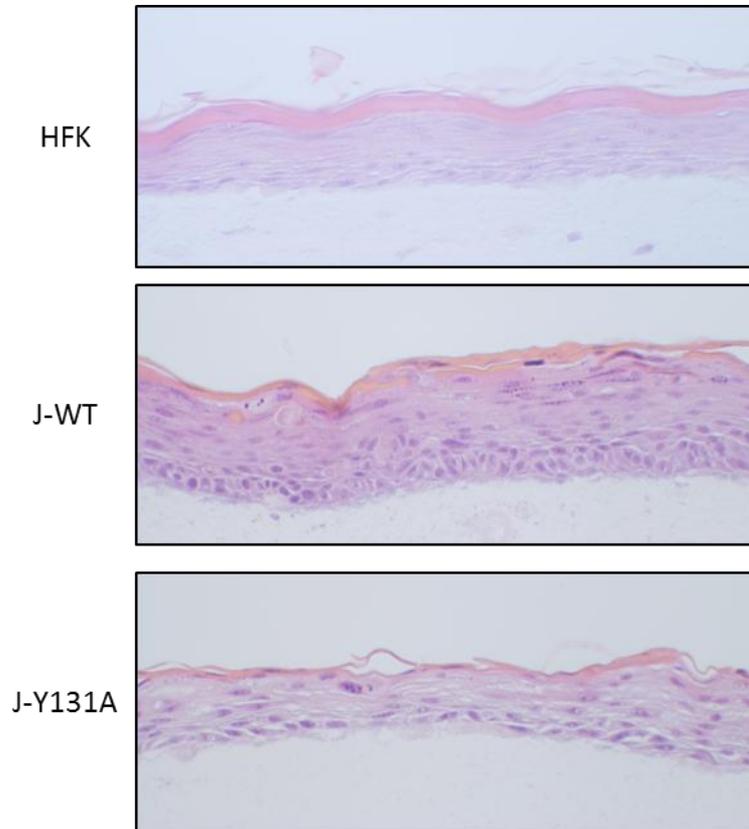
HFKs containing the WT or mutant HPV16 genomes were seeded onto collagen plugs containing fibroblast feeder cells. The HFKs were grown to confluency before the plug was lifted onto a mesh grid and suspended at the liquid-air interface. Growing the HFKs at this interface encourages cellular differentiation and recapitulates the normal epithelial differentiation, thus allowing the late stages of the HPV life cycle to be studied. After 13 days growth, the rafts were formaldehyde fixed, paraffin embedded and sectioned for analysis. Two rafts were grown for each WT and mutant cell line, along with normal untransfected HFKs from each donor as controls.

Initially raft sections were stained with hematoxylin and eosin (H&E) to look at the morphology of the rafts by light microscopy. Representative images of raft sections from one donor are shown (Figure 5.22).

When grown in organotypic raft culture, normal untransfected HFKs form a thin raft with no obvious proliferation in the suprabasal layers. In contrast, rafts containing WT HPV16 genomes (Jerry) displayed a hyperproliferative phenotype typified by thickening of the raft mid-section (suprabasal layer). This is consistent with the expected observation for HPV16 containing rafts [1], resulting from an increase in cells migrating upwards from the basal layer and E6/E7 protein expression driving cells to keep cycling into S-phase to allow for amplification of the viral genomes prior to virion production in terminally differentiated cells.

When cells containing Y131A E2 HPV16 genomes were grown in organotypic raft culture, the raft morphology appeared to revert to that exhibited by untransfected HFKs rather than retain the phenotype observed in WT rafts, with the thickening of the raft mid-section no

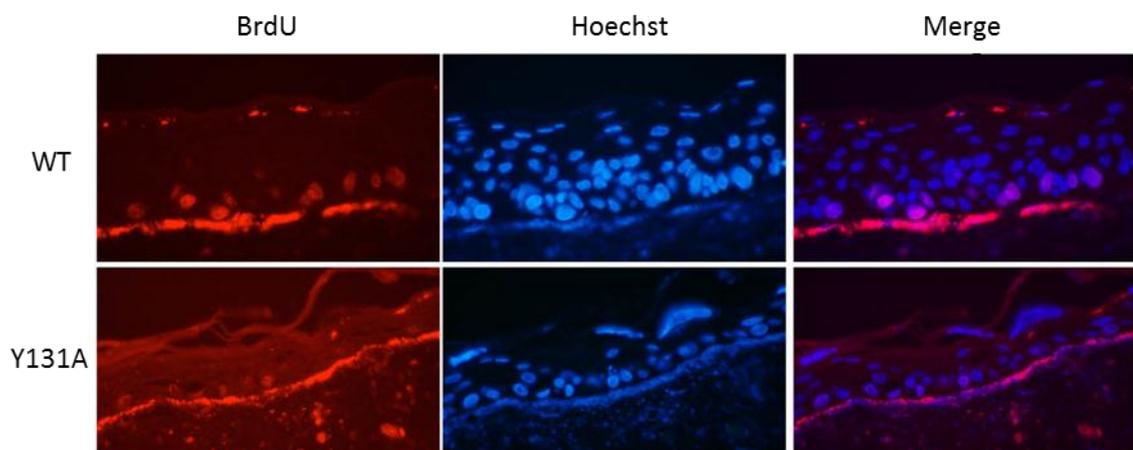
longer observed. This observed phenotype is consistent with a reduction in HPV copy number as seen in the southern blot analysis carried out on the cells prior to rafting.



**Figure 5.22 HFKs containing mutant HPV16 genomes exhibit changes in raft morphology compared to WT cell lines.**

Sections of formaldehyde fixed paraffin embedded rafts grown from HFK lines from two donors containing WT and Y131A-E2 mutant HPV16 genomes were H&E stained for analysis by light microscopy. Shown are representative raft sections of A. untransfected HFKs, B. WT HPV16 genome containing HFKs and C. Y131A-E2 HPV16 genome containing HFKs from donor 2 (Jerry).

WT and mutant HFK lines from both donors were grown in this way, however the WT rafts from one donor (Nigel) did not fully differentiate and so cannot be used for comparison with the mutant lines. However, the rafts produced from the Nigel mutant line show a similar phenotype to that shown with the Jerry mutant line, although this data is preliminary and without the corresponding WT rafts from the same donor no conclusions can be drawn from this without follow-up work at a later date.



**Figure 5.23 BrdU staining of HFK rafts**

**Organotypic raft cultures of HPV16 WT and Y131A E2 mutant genome containing HFK lines were incubated with BrdU for 8 hours prior to fixation. Sections of formaldehyde fixed paraffin embedded rafts were stained with BrdU-specific antibodies, and DNA was Hoechst stained. Shown are representative images of raft sections of WT- and Y131A-genome containing HFK lines from donor 2 (Jerry).**

Prior to fixation, the organotypic raft cultures were incubated with the nucleotide analogue bromodeoxyuridine (BrdU) for 8 hours. BrdU can be incorporated into newly synthesised

DNA, and as such cells undergoing active DNA replication can be identified by IF staining of raft sections with BrdU-specific antibodies (Figure 5.23).

Comparison of the BrdU staining in HPV16 WT and Y131A E2 genome-containing rafts showed a reduction in the number of actively proliferating cells in the mutant rafts compared to WT. In WT rafts, BrdU positive cells can be identified in the basal and supra-basal layers of the raft, while very few BrdU positive cells can be identified in the mutant rafts. Shown (Figure 5.23) are representative images for raft sections taken from lines generated in HFKs from one donor (Jerry). As discussed above, the Nigel WT line did not fully differentiate and therefore cannot be used for analysis, but this preliminary data does suggest that the loss of E2-ChIR1 interaction is negatively affecting the proliferation of transfected HFK-lines grown in organotypic raft culture.

## 5.4 Discussion

The E2-ChIR1 interaction has been proposed to play an important role in the maintenance of viral genomes in infected cells. The DNA helicase ChIR1 has been isolated as a cellular binding partner of E2, suggested to facilitate the E2-mediated tethering of viral genomes to cellular chromatin. In BPV, the E2-ChIR1 interaction was identified and demonstrated to affect the localisation of E2 to the cellular chromatin [95]. Interestingly, the association of BPV-1 E2 and ChIR1 could be abrogated by a W130R mutation in E2, leading to a corresponding loss of chromatin localisation of E2. The interaction between E2 and ChIR1 appears to be conserved with HPV16 and a similar mutation in HPV16 E2, Y131A led to significant reduction in ChIR1 association (Chapter 4). This mutation could therefore be used to assess the role of the E2-ChIR1 interaction in the HPV life cycle.

### *E2 function*

The E2 protein plays multiple key roles in the HPV life cycle. Mutations in E2 could therefore affect any of these functions in turn. To investigate whether loss of ChIR1 interaction induced by mutation Y131A affected E2-dependent transcription and replication, luciferase based assays using artificial reporter plasmids were used. Together, the data from the transcription and replication assay show that the HPV16 E2 mutant Y131A retains WT E2 functionality. It is able to both activate transcription and facilitate replication from reporter plasmids designed to mimic the early stages of the viral life cycle, suggesting that the E2-ChIR1 interaction is not involved in these processes.

To study E2 dependent transcription activation, a reporter plasmid (ptk-6E2) [190] containing 6 cloned E2 binding sites upstream of a thymidine kinase promoter which drives

expression of firefly luciferase was used. Co-transfecting this reporter with increasing amounts of WT or Y131A E2 expression plasmid allowed this system to be used in a luciferase assay to determine the amount of E2 dependent transcription of the firefly luciferase gene. In this way the detected Firefly luciferase activity has been shown to be proportional to the E2 concentration, until a saturating concentration is reached resulting in a plateau in the Firefly luciferase activity. Along with the Firefly luciferase expressing reporter plasmid, cells were also cotransfected with a constitutively active *Renilla* luciferase expression plasmid to act as an internal control. The *renilla* luciferase expression can be used to normalise results against transfection efficiency, allowing comparisons of Firefly luciferase activity to be drawn between varying conditions. Co-transfection with HPV16 E2 WT or Y131A expression plasmids increased the detected Firefly luciferase activity compared to the reporter alone (Figure 5.2), and no difference in induced Firefly luciferase activity was detected between WT and Y131A E2 over a range of concentrations. This demonstrates that the loss of ChIR1 interaction due to the E2 mutation Y131A has no effect on the E2-dependent transcription activation of reporter plasmid, indicating that the E2-ChIR1 interaction is not required for the transcription activation function of E2.

In a similar manner, a reporter plasmid encoding the Firefly luciferase expression gene controlled through a CMV promoter and the HPV31 minimal origin of replication (pFLOri, Figure 5.3) was used in a similar luciferase-based assay to determine the effect of the HPV16 E2 mutation Y131A on E2-dependent replication. This artificial system developed by J. Archambault [179] allows the initial stages of transient viral replication to be studied outside the confines of the life cycle model. As with the above transcription assay, a *renilla* luciferase expression plasmid was co-transfected along with the pFLOri reporter plasmid to act as an

internal control and correct for any variation in transfection efficiency between samples. At the lower concentrations of transfected E2, no difference in the Firefly luciferase activity was detected between WT and Y131A E2 transfected cells, showing a comparable ability to facilitate replication of the reporter plasmid. At the highest concentration of transfected E2 however there is a significant reduction in Firefly luciferase activity with WT E2 compared with Y131A E2. The observed drop-off in luciferase activity following transfection with the highest concentration of WT E2 is not consistent with the published data [179], where a steady increase in Firefly luciferase was detected with increasing E2. However, HPV16 E2 Y131A is able to mediate replication of the reporter plasmid to WT levels at low concentrations, and is able to retain this activity at higher concentrations to a greater extent than WT E2. This is potentially due to the consequences of the loss of ChIR1 interaction rather than a specific function of that protein-protein interaction. Data investigating the BPV-1 E2 interaction with ChIR1 indicated that the interaction was cell-cycle dependent, occurring predominantly during S-phase [133]. Interaction with ChIR1 would therefore potentially relocate E2 into a chromatin-associated complex, reducing the pool of free E2 available to associate with E1 and initiate replication from the viral Ori. Loss of the ChIR1 interaction and the subsequent disruption in E2 localisation to the chromatin could therefore potentially result in an increased cellular pool of E2 available to facilitate viral replication, leading to a sustained level of Firefly luciferase activity.

#### *E2 protein stability*

Although HPV16 E2 Y131A was demonstrated to be functionally active, western blotting consistently revealed lower levels of protein detected compared to WT E2 when equal

concentrations of expression plasmids were transfected. A possible explanation for this was that the mutant protein was unstable and rapidly degraded within cells, or relocalised to alternative, less soluble complexes.

Lysis of C33a cells transfected with WT or Y131A E2 with either the passive lysis buffer (PLB) used in the functional luciferase assays or with a strongly denaturing urea lysis buffer revealed differences in the solubility of the two E2 proteins. Urea buffers solubilise proteins by denaturation of the secondary structure and can therefore allow proteins which are either aggregated or stabilised in protein complexes to be detected. HPV16 E2 WT protein levels detected by western blotting were only marginally affected by the change in lysis buffer, suggesting that the majority of the expressed E2 protein is soluble in the cells. However, urea lysis of HPV16 E2 Y131A transfected cells revealed an average 3.5 fold increase in detected protein compared to PLB lysis. This suggests that while the expression of both WT and Y131A E2 from the transfected expression plasmids may be similar, a larger proportion of the mutant E2 protein is not readily solubilised within cells compared to WT. This could be due to the mutant protein aggregating in cells, or it could be due to a change in E2 localisation to alternative protein complexes which stabilise E2 following loss of the E2-ChIR1 interaction. Published data has shown that the interaction of E2 with protein complexes can increase the stability of E2 within cells [108].

To further investigate the effect of Y131 mutation on E2 protein stability, cycloheximide chase time course experiments were used. Linear regression analysis of the E2 protein levels following CHX treatment showed that while WT E2 displayed a  $t_{1/2}$  consistent with published data [109, 182], an approximately 10 fold increase in Y131A E2 protein  $t_{1/2}$  compared to WT

E2 was observed. It was hypothesised that depletion of ChIR1 would have the same effect on E2 protein stability as mutation to prevent binding to ChIR1. However, when the CHX experiment was repeated studying the stability of HPV16 E2 WT in cells depleted of ChIR1, there was consistently no change in E2 stability following ChIR1 depletion. It is therefore possible that the increase in stability observed with the HPV16 E2 mutant Y131A is a consequence of mutation at that specific site affecting an alternative process rather than being a direct consequence of the loss of ChIR1 interaction. If the amino acid Y131 is a phosphorylation site within E2 which is required for signalling proteasomal degradation, mutation at this site resulting in a loss of phosphorylation could impact on the protein degradation pathway resulting in a longer protein half-life. There is some evidence to suggest that proteasomal degradation of E2 is controlled through post-translational modification of sites within the TAD [185, 191], and therefore it is possible that mutation of Y131 could be affecting E2 protein stability through these pathways.

To investigate this, open-access bioinformatics tools were used to assess the HPV16 E2 protein sequence for potential phosphorylation sites. The results of these investigations were inconclusive, with Y131 being highlighted as a potential phosphorylation site using the online tools GPS 2.1 [192] and NetPhos 2.0 [193] but not being identified using the alternative tools NetPhosK 1.0 [194] and Scansite [195]. These contradictory results suggest that Y131 is not strongly predicted to be phosphorylated, but that this residue could exist within an alternative, less common consensus motif. Further analysis of this residue as a potential phosphorylation site would require phospho-mapping using mass spectrometry experiments [196]. Due to time constraints this complex process was not carried out but it remains a subject of interest for further work.

### *E2 protein sub-cellular localisation*

An alternative explanation for the dramatic increase in protein stability observed with the E2 mutant Y131A is that the loss of interaction with ChIR1 results in a change in E2 localisation within the cell. If E2 is no longer localised to the cellular chromatin through an interaction with ChIR1, it is possible that the free E2 protein could form interactions with other complexes and be relocalised throughout the cell; these interactions with other cellular protein complexes could then be stabilising the protein within the cell and indirectly affecting protein degradation. Mutation of the large, polar tyrosine residue to the smaller, non-polar alanine could also affect the spatial proximity of the adjacent amino acid residues, disrupting intra-molecular binding and resulting in changes to the binding surface which could increase the affinity of E2 for other proteins within the cell. The difference in stability observed between the ChIR1 binding mutant Y131A, and WT E2 in cells depleted of ChIR1 could therefore be explained by the relative difference in protein available to interact within the cell. The stoichiometry of the E2-ChIR1 interaction is not known, and although it has been shown that E2 TAD dimers are required for association with mitotic chromosomes [76], it is possible that multiple E2 dimers are required to interact with one ChIR1 protein. This would suggest that only a small amount of ChIR1 is required to localise E2 to the cellular chromatin. Although the ChIR1 knockdown efficiency was good with an average 90% reduction in ChIR1 protein levels (Figure 5.7), the remaining ChIR1 protein could be sufficient to interact with E2 and localise the protein correctly. The E2 mutant Y131A does not completely abrogate the ChIR1 binding affinity, however it has been demonstrated to result

in a 90% reduction in binding (see Chapter 4). High levels of E2 binding with only 10% affinity to the protein may not be sufficient to localise the protein correctly, resulting in this discrepancy in the observed protein stability.

The subcellular localisation of WT and Y131A E2 was then investigated within transfected cells to ascertain if any changes in protein localisation were observed following loss of ChR1 binding. Initially the proteins were examined by IF (Figure 5.8). Cells transfected with HPV16 E2 WT exhibited nuclear E2 staining with some cytoplasmic protein observed in some cells, consistent with published data [95, 133]. Y131A E2 transfected cells showed a different staining profile however. While nuclear staining was retained, in many cells there was an increase in punctate, cytoplasmic staining indicative of protein localisation with the cytoskeleton or other cellular structures. Attempts were then made to visualise the localisation of WT and Y131A E2 in mitotic cells, as it was hypothesised that the loss of ChR1 interaction with the ChR1-binding mutant E2 Y131A would result in a loss of E2 localisation to the mitotic chromatin and a concomitant reduction in viral genome tethering. IF had previously shown that BPV-1 E2 forms discrete foci on mitotic chromosomes, while the BPV-1 E2 mutant W130R did not form visible foci [95]. While multiple PV E2s have been shown to associate with mitotic chromosomes resulting in slightly different staining profiles, the  $\alpha$ -PVs including HPV16, -31, -11 and -57 were found to behave differently, and similar chromosome-associated foci were difficult to observe [93]. The  $\alpha$ -PV E2 proteins were observed to be poorly expressed in mitotic cells and to localise with the chromatin during prophase and telophase, but no chromatin-associated E2 was detected during metaphase or anaphase. Pre-extraction of soluble proteins prior to fixation has been shown to result in a change in the staining profile of the  $\alpha$ -PV E2s however, leading to E2 foci detected in the

pericentromeric region of mitotic chromosomes, similar to the pattern observed with  $\beta$ -PV HPV8 E2. However, it is not certain if the pre-extraction step is required to induce a change in protein localisation, or if it merely unmasks a previously hidden epitope on E2 to allow for staining [93].

Consistent with published data, cells transfected with HPV16 E2 WT or Y131A showed minimal E2 localisation to mitotic chromosomes. Low protein expression coupled with relatively low populations of mitotic cells within each phase of mitosis, even following synchronisation by double thymidine block and release to enrich the mitotic populations, resulted in insufficient E2 positive mitotic cells to allow any conclusions about E2 localisation during mitosis to be drawn. Attempts to use stable, inducible cell lines expressing HPV16 E2 WT or Y131A were similarly unsuccessful. The inducible pMEP vector used to generate stable cell lines contains a metallothionein promoter. Protein expression can be induced by addition of a suitable metal salt, typically cadmium sulphate. However, due to health and safety concerns over handling the highly toxic  $\text{CdSO}_4$ , a safer alternative was utilised in this case, zinc chloride ( $\text{ZnCl}_2$ ). Although considerably safer for regular use than  $\text{CdSO}_4$ ,  $\text{ZnCl}_2$  is a less effective activator of the metallothionein promoter and is required to be used at higher concentrations (100-200 mM compared to 1-2  $\mu\text{M}$   $\text{CdSO}_4$ ). pMEP-HPV E2 cell lines had been previously shown to provide adequate numbers of E2 positive mitotic cells for IF when protein expression was induced with  $\text{CdSO}_4$  [93], however attempts to use similar cell lines with the higher concentrations of  $\text{ZnCl}_2$  coupled with double thymidine block to enrich for mitotic cell populations resulted in a large amount of cell death. As such no data could be obtained about the mitotic localisation of Y131A E2 compared to WT E2 using IF.

As an alternative approach, sub-cellular fractionation techniques were utilised to investigate the localisation of WT and Y131A E2. Transfected cells were fractionated using a commercial sub-cellular fractionation kit. In each experiment, good separation of the protein fractions was achieved, with some small overlap observed between the cytoplasmic and membrane fractions. No marker of the membrane fraction is included, as repeated attempts with available antibodies to different membrane-associated proteins failed to identify a marker which was consistently detected across experiments. As expected, HPV16 E2 WT was localised to both the cytoplasmic and nuclear soluble fractions, and to a lesser extent to the chromatin associated fraction (Figure 5.11). The lower protein levels in the chromatin fraction are assumed to be due to the relatively low proportion of mitotic cells included in the asynchronous cell population. WT E2 was also shown to localise to the cytoskeletal and nuclear matrix associated fraction (CNM), and it is presumed that this is due to interactions with the lamin proteins and intermediate filaments which make up the nuclear matrix, as it is known that E2 is associated with these structures through regions in the TAD and hinge [197].

HPV16 E2 Y131A displayed a consistent shift in the protein localisation compared to WT E2, with a significant increase in CNM associated protein and a corresponding decrease in nuclear soluble protein (Figure 5.11). Repetition of this experiment in cells synchronised at the G<sub>1</sub>/S boundary or mid-S phase by double thymidine block and release showed that this effect is cell cycle dependent, occurring in mid-S phase cells. This correlates with published data showing that the BPV-1 E2-ChIR1 interaction is observed predominantly during mid-S phase [133]; if loss of the ChIR1 interaction resulted in a shift in protein localisation it would be expected to also occur predominantly during S-phase. This suggests that the change in

protein localisation is a specific result of the loss of E2 association with ChR1. It is hypothesised that loss of the E2-ChR1 would prevent E2 association with the as yet undefined chromatin-tethering complex, and that the free E2 can therefore associate with other protein complexes, particularly within the cytoskeleton and nuclear matrix. This idea is also supported by the increased stability of the Y131A E2 protein, together suggesting that mutant E2 protein is interacting as part of an alternative protein complex which stabilises E2.

To determine if the observed differences between WT and Y131A E2 were due to the loss of ChR1 interaction or a direct effect of mutation at Y131, experiments were performed to study the sub-cellular localisation of HPV16 E2 WT in cells depleted of ChR1. In asynchronous cells, an average ChR1 knockdown efficiency of approximately 90% was achieved, similar to that shown in the previous cycloheximide stability assays. Sub-cellular fractionation of the transfected cells and the subsequent analysis by western blotting revealed a similar trend in localisation in control siRNA treated cells as was previously demonstrated with WT E2 alone. In cells depleted of ChR1 however there was a decrease in nuclear soluble E2 protein and a similar increase in CNM associated protein. The same trend as was observed with Y131A E2 although in this case the difference in protein localisation did not reach statistical significance. These data suggest that the change in protein localisation to the CNM is indeed due to the loss of ChR1 interaction resulting in E2 relocalisation to alternative complexes. The change in the magnitude of this localisation shift can again be ascribed to the difference between the mutant protein abrogating the ChR1 interaction and the siRNA depletion resulting in a reduction in ChR1 expression. E2 protein binding to ChR1

with only 10% efficiency relative to WT could result in relocalisation of E2, while the 10% ChIR1 protein remaining in cells following siRNA-mediated ChIR1 knockdown could be sufficient to interact with WT E2 with high affinity and prevent E2 relocalisation.

As the effect of the HPV16 E2 mutation Y131A on protein localisation was shown to be cell cycle dependent, the localisation of WT E2 in cells depleted of ChIR1 was also investigated in cells synchronised to mid-S phase by double thymidine block and release. In these experiments however, the ChIR1 knockdown was less efficient than in previous experiments with an average of approximately 30% ChIR1 remaining in cells compared to treatment with control siRNA. Given the higher levels of remaining ChIR1 in these experiments compared to the previous sub-cellular fractionations, it is therefore not surprising that there was no change in the localisation of WT E2 observed between control and ChIR1 siRNA treated cells.

Together, the data show that HPV16 E2 Y131A displays an increased localisation to the cytoskeletal and nuclear matrix sub-cellular fraction compared to WT E2, and that a similar shift in localisation can be observed when WT E2 is studied in cells depleted of ChIR1. This change in localisation was also shown to be S-phase specific cells, consistent with published data suggesting that the E2-ChIR1 interaction is specific to mid-S phase [133] and that therefore any change in localisation due to disruption of this interaction would be observed at this point in the cell cycle.

To look further at the increase in Y131A localisation to the CNM fraction, an alternative sub-cellular fractionation technique coupled with IF was utilised. The results showed that the number of WT E2 positive cells decreased following each subsequent extraction step. Removal of the cytoplasmic and nuclear soluble proteins reduced the number of E2 positive

cells by approximately 50% compared to the whole cell, while removal of the tightly held nuclear, and chromatin associated proteins in turn reduced the number of E2 positive cells by a further 17% and 15% respectively. In comparison, the sub-cellular distribution of Y131A E2 was altered. Removal of the cytoplasmic and nuclear soluble proteins resulted in a 29% reduction in the number of E2 positive cells compared to whole cell, a smaller reduction than that observed with WT E2. Removal of the tightly held nuclear proteins reduced the number of E2 positive cells by a further 25%, while the removal of the chromatin associated proteins only led to an additional 2% reduction in E2 positive cells. This provides evidence that there was a significant increase in nuclear matrix associated Y131A E2 and an associated decrease in the chromatin associated protein compared to WT E2. This again supports the suggestion that loss of ChIR1 interaction results in a change in E2 protein localisation, and that Y131A E2 has a reduced affinity for the chromatin fraction and subsequent increased association with the nuclear matrix fraction. This data also supports the hypothesis that the HPV16 E2-ChIR1 interaction is required for the localisation of E2 to chromatin and the E2-mediated tethering of viral genomes.

Following on from these experiments, the *in situ* fractionation and IF experiments were repeated to determine the sub-cellular localisation of WT E2 in cells depleted of ChIR1. A similar trend in E2 protein localisation was observed; removal of the chromatin associated proteins resulted in only a 10% reduction in the number of E2 positive cells relative to whole cells, compared to the 26% reduction observed in cells treated with control siRNA. As with the previous sub-cellular fractionation experiments, this trend is the same as that observed with Y131A E2 but did not reach statistical significance. Again, this is thought to be due to

the difference between remaining ChIR1 following siRNA depletion, and the reduced binding affinity exhibited by the E2 mutant Y131A.

Taken together, all of the fractionation experiments reveal a change in the localisation of E2 following disruption of the interaction with ChIR1, resulting in association of E2 protein with the nuclear matrix rather than the chromatin. This supports the hypothesis that the ChIR1-E2 interaction is required for the localisation of HPV16 E2 with mitotic chromosomes, and that abrogation of this interaction could disrupt E2-dependent maintenance of viral genomes in infected cells.

#### *HPV life cycle model*

To determine the effect of the abrogation of the ChIR1-E2 interaction on the HPV life cycle, the E2 Y131A mutation was cloned into the HPV16 genome. Primary human foreskin keratinocytes were then transfected with WT or Y131A E2 mutant genomes to establish stable cell lines following selection. This model system allows the HPV life cycle to be studied within the lab, with cells being grown in monolayer culture to investigate the early stages of the viral life cycle, or induced to differentiate to allow completion of the viral life cycle to be investigated.

Southern blot analysis of HPV16 WT and Y131A genomes extracted from HFK lines revealed a reduction in the episome copy number for mutant genomes compared to WT when quantified relative to a 50 copies per cell control. HPV-genome containing HFK lines from two independent donors were shown to establish genomes at different numbers (approximately 35 copies per cell in Nigel WT, 150 copies per cell in Jerry WT). While the number of genomes observed in Jerry-WT line is within the expected, published range [28],

the numbers in the Nigel-WT line are slightly lower than expected. In both lines however, a similar fold reduction in the mutant genome copy number was observed, reducing by approximately 7 fold in passage 1 cells (approximately 5 copies per cell in Nigel-Y131A and 20 copies per cell in Jerry-Y131A). Further reduction in mutant genome copy number was observed in passage 5 cells, with an approximately 18-fold reduction in Nigel-Y131A genome copy number compared to WT, and a 30-fold reduction in Jerry-Y131A genome copy number compared to Jerry-WT. This demonstrates that loss of the E2-ChIR1 interaction induced by the E2 mutant Y131A results in a loss of genome establishment and maintenance in HFKs. This correlates with the loss of episomal genome establishment observed with the analogous BPV-1 E2 mutant W130R [95], and with E2 mutants which abrogate binding to other cellular chromatin associated proteins Brd4 [75] and TopBP1 [109].

To find further evidence to support the hypothesis that loss of E2-ChIR1 interaction inhibits the establishment of episomal HPV genomes in primary cell culture, HPV16 WT and mutant genome containing HFK lines were grown in organotypic raft culture to emulate the viral life cycle. Unfortunately, the Nigel-WT HFK line did not fully differentiate in this raft system, preventing complete analysis of the cell morphology and proliferation, but preliminary results from the Jerry HFK lines could be assessed. H&E staining of fixed raft sections revealed that the Jerry-WT HFKs showed a hyperproliferative phenotype typified by thickening of the suprabasal layers compared to normal, HPV negative HFKs from the same donor, as expected for HPV positive rafts [1]. The raft sections of Jerry-Y131A HFKs however did not show this hyperproliferation, reverting to a phenotype similar to that of the untransfected HFKs. This supports the hypothesis that loss of the E2-ChIR1 interaction and

the subsequent loss of E2-mediated genome tethering prevents viral episome establishment and therefore prevents the hyperproliferation in the raft mid-layers caused by HPV.

To further investigate the effect of the Y131A E2 mutation on the HPV life cycle, raft cultures were incubated with BrdU prior to fixation. Raft sections were then stained with anti-BrdU antibodies to investigate the proliferation of cells within the rafts. Jerry-WT rafts showed BrdU positive cells within the lower layers of the raft, indicating ongoing DNA replication above the basal layer, in agreement with published data [1]. In contrast, Jerry-Y131A rafts showed a reduction in the number of BrdU positive cells, adding further evidence to the suggestion that loss of the E2-ChIR1 interaction prevents establishment of viral genomes and inhibits the HPV-mediated hyperproliferative phenotype observed in organotypic rafts.

Together, these data provide evidence that the E2-ChIR1 interaction is required for HPV episomal genome tethering, while not being necessary for the transcription activation and replication functions of E2. Loss of interaction with ChIR1 results in a change in localisation of E2 to the nuclear matrix, resulting in an increase in protein stability. In a model primary cell culture system, E2 mutant-containing genomes were demonstrated to not establish episomal genomes to WT levels, and further analysis of organotypic raft cultures showed that E2 mutant rafts did not retain the hyperproliferative phenotype associated with WT HPV genomes.

## Chapter 6. Conclusions and future work

The HPV E2 protein is essential for the virus life cycle, regulating viral genome replication, transcription and viral genome maintenance. E2 acts to tether viral genomes to cellular chromosomes during mitosis to ensure even segregation of the viral genomes into daughter cells. Multiple cellular proteins have been proposed to act as the chromatin receptor of E2, including the DNA helicase ChIR1. Published data demonstrated that BPV-1 E2 and ChIR1 interact during mid-S phase, and that loss of this interaction *via* a BPV-1 E2 mutant W130R was demonstrated to result in a loss of E2 localisation to mitotic chromosomes and a subsequent failure to establish mutant episomal genomes in cell culture. As such, the E2-ChIR1 interaction is proposed to be required for loading E2-tethered viral genomes onto the cellular chromatin-protein complex to ensure viral genome persistence. The E2-ChIR1 interaction is therefore an interesting novel target for the development of therapeutics to treat HPV infections by disrupting the E2-ChIR1 interaction and consequently preventing the establishment of a persistent infection.

The data presented here show work carried out to optimise conditions for using HPV11 E2 TAD as a target in FBDD screening using biophysical DSF- and NMR-based assays, and an in-depth analysis of the HPV16 E2-ChIR1 interaction.

Fluorescence-based thermal shift assays were used to optimise buffer conditions to stabilise HPV11 E2 TAD for downstream binding assays. Screening a small molecule fragment library for binding to HPV11 E2 TAD using DSF and NMR assays identified a series of fragments which are structurally dissimilar to known E1-E2 interaction inhibitors, and bind to HPV11 E2 TAD. This suggests that fragment binding can occur through alternative binding sites,

supporting the hypothesis that small molecules which bind to and disrupt E2 protein-protein interactions can be identified using this methodology. Extensive structural analysis of the fragment binding sites on E2 by NMR could not be completed, despite attempts to optimise the buffer conditions to ensure maximum HPV11 E2 TAD protein stability during the required long data acquisition periods.

The interaction between HPV16 E2 TAD and ChIR1 was mapped using a series of truncated proteins and *in vitro* binding assays. These experiments mapped the binding site of HPV16 E2 within ChIR1 to an 84 amino acid region (130-214) that is within a unique domain of the N-terminus of ChIR1. Additional data provides evidence that the E2 binding site within ChIR1 could be further mapped to a 19 amino acid region (194-213). *In vitro* binding assays performed with purified E2 and ChIR1 proteins also provided evidence of a direct, protein-protein interaction. Site-directed mutagenesis of HPV16 E2 then led to the identification of an E2 mutant, Y131A, which is deficient in binding to ChIR1 63-222. The loss of ChIR1 binding affinity conferred by Y131A mutation was confirmed by co-IP.

Identification of a HPV16 E2 mutant with reduced affinity for ChIR1 (Y131A) then allowed investigation of the role of the E2-ChIR1 interaction in the virus life cycle. Mutant E2 protein was determined to be functionally active using luciferase-based transcription activation and replication assays, however a change in the protein stability and localisation was observed. Mutation of HPV16 E2 Y131 resulted in a 10-fold increase in protein stability compared to WT E2, however this same stabilisation was not observed with WT E2 in cells depleted of ChIR1, suggesting that the mutation of Y131 could induce effects separate from the loss of

ChIR1 binding. A proposed suggestion is that phosphorylation of residue Y131 within HPV16 E2 is important for regulating proteasomal degradation of E2.

Loss of ChIR1 binding induced by HPV16 E2 mutant Y131A resulted in a cell cycle dependent change in sub-cellular localisation of E2 protein, with increased localisation to the cytoskeleton and nuclear matrix, and a concurrent decrease in nuclear soluble protein. The same increase in HPV16 E2 Y131A localisation to the nuclear matrix and a corresponding decrease in chromatin-associated protein was also demonstrated using *in situ* fractionation and IF. Similar trends in protein localisation were also demonstrated with HPV16 E2 WT in cells depleted of ChIR1, supporting the hypothesis that loss of the E2-ChIR1 interaction and the subsequent localisation of E2 to the cellular chromatin results in relocalisation of E2 to alternative protein complexes which stabilise the protein.

In primary cell culture, HFKs containing HPV16 E2 Y131A genomes failed to establish episomal genomes at levels comparable to WT genome-containing HFKs. When cellular differentiation was induced by organotypic raft culture of the HPV16 WT- and Y131A-genome containing HFKs, HPV16 Y131A HFK rafts were shown to lack the hyperproliferative phenotype observed in HPV16 WT-HFK rafts. Analysis of H&E-stained raft sections showed HPV16 WT rafts displayed a thicker midsection typical of increased HPV-induced cell proliferation compared to untransfected HFKs, while HPV16 Y131A rafts retained the normal, untransfected HFK phenotype. Staining raft sections for incorporated BrdU supported this conclusion, with fewer BrdU-positive, actively replicating cells observed in HPV16 Y131A raft sections compared to WT. However this data is preliminary and confirmation of this phenotype is required.

Together, these data demonstrate that the E2-ChIR1 interaction influences the sub-cellular localisation of E2 protein, and is required for the maintenance of viral genomes in cells. This supports the hypothesis that ChIR1 is required to load E2 onto cellular chromosomes and facilitates E2-dependent tethering of viral genomes. Loss of this interaction prevents episomal genome establishment in primary cells and subsequently a loss of the hyperproliferative phenotype in differentiated cells which typifies infection with HR-HPVs.

To continue this work, repetition of experiments to investigate the role of E2-ChIR1 interaction in organotypic raft culture is required. Although data from two independent donor lines supports the hypothesis that loss of the E2-ChIR1 interaction results in failure to establish viral episomes, the subsequent loss of the expected hyperproliferative phenotype when these lines are grown in organotypic raft culture needs to be confirmed in the second donor line to ensure that the observed effects are not specific to the cell lines shown.

Additionally, phospho-mapping of HPV16 E2 is required to identify if the amino acid Y131 is phosphorylated. This would allow the discrepancy in protein stability observed between HPV16 E2 Y131A, and HPV16 WT in cells depleted of ChIR1 to be explored, and provide valuable information about the regulation of E2 protein stability.

To confirm the binding site of E2 on ChIR1, mapping of the binding site with ChIR1 peptides could be continued. Alongside the *in vitro* binding assays discussed, biophysical screens such as SPR could be used to quantify the binding affinity of E2 to ChIR1.

Finally, optimisation of the expression and purification of HPV16 E2 WT and Y131A would allow the proteins to be used in biophysical screens to identify small molecule fragments with differential binding to the E2 proteins, using the pre-optimised assays discussed. This

could identify small molecule fragments binding at the E2-CHR1 interaction site and provide a start for development of drug-like molecules that inhibit this interaction. Further optimisation of conditions for 3D NMR and X-ray crystallography analysis could then provide information about the binding site of fragments to E2.

## References

1. Kovanda, A., et al., *Characterisation of Novel Cutaneous Human Papillomavirus Genotypes HPV-150 and HPV-151*. PLoS One, 2011. **6**(7): p. e22529.
2. Bzhalava, D., et al., *Deep sequencing extends the diversity of human papillomaviruses in human skin*. Scientific Reports, 2014.
3. Van Doorslaer, K., et al., *The Papillomavirus Episteme: a central resource for papillomavirus sequence data and analysis*. Nucleic Acids Res., 2013. **41**: p. D571-8.
4. de Villiers, E.-M., et al., *Classification of papillomaviruses*. Virology, 2004. **324**: p. 17-27.
5. Campo, M.S., *Animal models of papillomavirus pathogenesis*. Virus Research, 2002. **89**: p. 249-261.
6. zur Hausen, H., *Papillomaviruses and cancer: from basic studies to clinical application*. Nat. Rev. Cancer, 2002. **2**(5): p. 342-350.
7. Desai, S., et al., *Genital warts and cost of care in England*. Sex. Transm. Infect., 2011. **87**(464-468).
8. Woodhall, S.C., et al., *Cost of treatment and QALYs lost due to genital warts: data for the economic evaluation of HPV vaccines in the United Kingdom*. Sex Transm. Dis., 2009. **36**(8): p. 515-521.
9. Lacey, C.J.N., C.M. Lowndes, and K.V. Shah, *Chapter 4: Burden and management of non-cancerous HPV-related conditions: HPV-6/11 disease*. Vaccine, 2006. **24S3**: p. S3/35-S3/41.
10. Maw, R., *Critical appraisal of commonly used treatment for genital warts*. Int. J. STD AIDS, 2004. **15**(6): p. 357-364.
11. Stanley, M.A., *Genital human papillomavirus infections - current and prospective therapies*. J. Gen. Virol., 2012. **93**: p. 681-691.
12. Viera, M.H., et al., *Herpes simplex virus and human papillomavirus genital infections: new and investigational therapeutic options*. Int. J. Dermatol., 2010. **49**(7): p. 733-749.
13. zur Hausen, H., *Human papillomaviruses and their possible role in squamous cell carcinomas*. Curr. Top. Microbiol. Immunol., 1977. **78**: p. 1-30.
14. WHO, *Human papillomavirus vaccines: WHO position paper, October 2014*. Wkly Epidemiol. Rec., 2014. **89**(43): p. 465-492.
15. Kimple, A.J., et al., *HPV-Associated head and neck cancer: Molecular and nano-scale markers for prognosis and therapeutic stratification*. Sensors, 2012. **12**: p. 5159-5169.
16. Wu, X., et al., *Human papillomavirus-associated cancers - United States, 2004-2008*. Morbidity and Mortality Weekly Report, 2012. **61**(15): p. 258-261.
17. WHO, *Human papillomavirus and related cancers in the world. Summary report 2010*. WHO/ICO Information Centre on HPV and Cervical Cancer (HPV Information Centre), 2010.
18. Ang, K.K., et al., *Human papillomavirus and survival of patients with oropharyngeal cancer*. N. Engl. J. Med., 2010. **363**(1): p. 24-35.
19. Lassen, P., *The role of human papillomavirus in head and neck cancer and the impact on radiotherapy outcome*. Radiother. Oncol., 2010. **95**(3): p. 371-380.
20. Mehanna, H., et al., *Prevalence of human papillomavirus in oropharyngeal and nonoropharyngeal head and neck cancer - systematic review and meta-analysis of trends by time and region*. Head Neck, 2013. **35**(5): p. 747-755.
21. Mehanna, H., O. Olaleye, and L. Licitra, *Oropharyngeal cancer - is it time to change management according to human papillomavirus status?* Curr. Opin. Otolaryngol. Head Neck Surg., 2012. **20**(2): p. 120-124.
22. Stanley, M., D.R. Lowy, and I. Frazer, *Prophylactic HPV vaccines: Underlying mechanisms*. Vaccine, 2006. **24S3**: p. S3/106-S3/113.

23. Stanley, M.A., *Human papillomavirus (HPV) vaccines: prospects for eradicating cervical cancer*. J. Fam. Plann. Reprod. Health Care, 2004. **30**(4): p. 213-215.
24. Paavonen, J., et al., *Efficacy of human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by oncogenic HPV types (PATRICIA): final analysis of a double-blind, randomised study in young women*. The Lancet, 2009. **374**(9686): p. 301-314.
25. Sheridan, A. and J. White, *Annual HPV vaccine coverage in England in 2009/2010*. 2010.
26. Kavanagh, K., et al., *Introduction and sustained high coverage of the HPV bivalent vaccine leads to a reduction in prevalence of HPV 16/18 and closely related HPV types*. B. J. Cancer, 2014. **110**: p. 2804-2811.
27. Chaturvedi, A.K., et al., *Human papillomavirus and rising oropharyngeal cancer incidence in the United States*. J. Clin. Oncol., 2011. **29**(32): p. 4294-4301.
28. Doorbar, J., et al., *The biology and life-cycle of human papillomaviruses*. Vaccine, 2012. **30**(S5): p. 55-70.
29. Chow, L.T., T.R. Broker, and B.M. Steinberg, *The natural history of human papillomavirus infections of the mucosal epithelia*. APMIS, 2010. **118**: p. 422-449.
30. Doorbar, J., *Molecular biology of human papillomavirus infection and cervical cancer*. Clin. Sci., 2006. **110**: p. 525-541.
31. Johnson, K.M., et al., *Role of heparan sulphate in attachment to and infection of the murine female genital tract by human papillomavirus*. J. Virol., 2009. **83**(5): p. 2067-2074.
32. Schiller, J.T., P.M. Day, and R.C. Kines, *Current understanding of the mechanism of HPV infection*. Gynecol. Oncol., 2010. **118**(1S): p. S12-7.
33. Evander, M., et al., *Identification of the alpha 6 integrin as a candidate receptor for papillomaviruses*. J. Virol., 1997. **71**(3): p. 2449-2456.
34. Bergant, M.M., et al., *Human papillomavirus L2 facilitates viral escape from late endosomes viral sorting nexin 17*. Traffic, 2012. **13**(3): p. 455-67.
35. Maglennon, G.A., P. McIntosh, and J. Doorbar, *Persistence of viral DNA in the epithelial basal layer suggests a model for papillomavirus latency following immune regression*. Virology, 2011. **414**(2): p. 153-163.
36. Lehman, C.W. and M.R. Botchan, *Segregation of viral plasmids depends on tethering to chromosomes and is regulated by phosphorylation*. Proc. Natl. Acad. Sci. U.S.A., 1998. **95**(8): p. 4338-4343.
37. McBride, A.A., *Replication and partitioning of papillomavirus genomes*. Adv. Virus. Rev., 2008. **72**: p. 155-205.
38. Flores, E.R. and P.F. Lambert, *Evidence for a switch in the mode of human papillomavirus type 16 DNA replication during the viral life cycle*. J. Virol., 1997. **71**(10): p. 7167-7179.
39. Kusumoto-Matsuo, R., T. Kanda, and I. Kukimoto, *Rolling circle replication of human papillomavirus type 16 DNA in epithelial cell extracts*. Genes Cells, 2011. **16**(1): p. 23-33.
40. Dyson, N., et al., *The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product*. Science, 1989. **243**(4893): p. 934-937.
41. Münger, K., et al., *Biological activities and molecular targets of the human papillomavirus E7 oncoprotein*. Oncogene, 2001. **20**(54): p. 7888-7898.
42. Roman, A., *The human papillomavirus E7 protein shines a spot light on the pRb family member p130*. Cell Cycle, 2006. **5**(6): p. 567-568.
43. Liu, X., et al., *Structure of the human papillomavirus E7 oncoprotein and its mechanism for inactivation of the pRb tumour suppressor*. J. Biol. Chem., 2006. **281**: p. 578-586.
44. Narisawa-Saito, M. and T. Kiyono, *Basic mechanisms of high-risk human papillomavirus-induced carcinogenesis: Roles of E6 and E7 proteins*. Cancer Sci., 2007. **98**(10): p. 1505-1511.
45. Benchimol, S., *p53-dependent pathways of apoptosis*. Cell Death Differ., 2001. **8**(11): p. 1049-1051.

46. Tomaic, V., D. Pim, and L. Banks, *The stability of the human papillomavirus E6 oncoprotein is E6AP dependent*. *Virology*, 2009. **393**(1): p. 7-10.
47. Pim, D. and L. Banks, *Interaction of viral oncoproteins with cellular target molecules: infection with high-risk vs low risk human papillomaviruses*. *APMIS*, 2010. **118**(6-7): p. 471-493.
48. Beaudenon, S. and J.M. Huibregtse, *HPV E6, E6AP and cervical cancer*. *BMC Biochem.*, 2008. **9**(S1): p. S4.
49. Klingelhutz, A.J., S.A. Foster, and J.K. McDougall, *Telomerase activation by the E6 gene product of human papillomavirus type 16*. *Nature*, 1996. **380**: p. 79-82.
50. Galloway, D.A., et al., *Regulation of telomerase by human papillomaviruses*. *Cold Spring Harb. Symp. Quant. Biol.*, 2005. **70**: p. 209-215.
51. Duensing, S. and K. Münger, *Mechanisms of genomic instability in human cancer: insights from studies with human papillomavirus oncoproteins*. *Int. J. Cancer*, 2003. **109**: p. 157-162.
52. Johansson, E. and P.F. Lambert, *Epigenetics of human papillomaviruses*. *Virology*, 2013. **445**: p. 205-212.
53. Johansson, C. and S. Schwartz, *Regulation of human papillomavirus gene expression by splicing and polyadenylation*. *Nat. Rev. Microbiol.*, 2013. **11**(4): p. 239-251.
54. Kadaja, M., et al., *Genomic instability of the host cell induced by the human papillomavirus replication machinery*. *EMBO J*, 2007. **26**(8): p. 2180-2191.
55. Pfüller, R. and W. Hammerschmidt, *Plasmid-like replicative intermediates of the Epstein-Barr virus lytic origin of DNA replication*. *J. Virol.*, 1996. **70**(6): p. 3423-3431.
56. Johansson, C., et al., *HPV-16 E2 contributes to induction of HPV-16 late gene expression by inhibiting early polyadenylation*. *EMBO J*, 2012. **31**: p. 3212-3227.
57. Hegde, R.S. and E.J. Androphy, *Crystal structure of the E2 DNA-binding domain from human papillomavirus type 16: implications for its DNA binding-site selection mechanism*. *J Mol Biol*, 1998. **284**(5): p. 1479-89.
58. Bussiere, D.E., et al., *Structure of the E2 DNA-Binding Domain from Human Papillomavirus Serotype 31 at 2.4 Å*. *Acta Crystallogr Sect D* 1998. **54**: p. 1367-1376.
59. Dell, G., et al., *Comparison of the structure and DNA-binding properties of the E2 proteins from an oncogenic and a non-oncogenic human papillomavirus*. *J. Mol. Biol.*, 2003. **334**(5): p. 979-991.
60. Thain, A., et al., *DNA binding and bending by the human papillomavirus type 16 E2 protein*. *J. Biol. Chem.*, 1997. **272**(13): p. 8236-8242.
61. Hegde, R.S., et al., *Subunit rearrangement accompanies sequence-specific DNA binding by the bovine papillomavirus-1 E2 protein*. *J Mol Biol*, 1998. **276**(4): p. 797-808.
62. Hegde, R., *The papillomavirus E2 proteins: structure, function and biology*. *Annu. Rev. Biophys. Biomol. Struct.*, 2002. **31**: p. 343-60.
63. Lewis, H., et al., *Cellular transcription factors regulate human papillomavirus type 16 gene expression by binding to a subset of the DNA sequences recognized by the viral E2 protein*. *J. Gen. Virol.*, 1999. **80**: p. 2087-2096.
64. Thierry, F. and M. Yaniv, *The BPV1-E2 trans-acting protein can be either an activator or repressor of the HPV18 regulatory region*. *EMBO J*, 1987. **6**(11): p. 3391-3397.
65. Antson, A.A., et al., *Structure of the intact transactivation domain of the human papillomavirus E2 protein*. *Nature*, 2000. **403**: p. 805-809.
66. Wang, Y., et al., *Crystal Structure of the E2 Transactivation Domain of Human Papillomavirus Type 11 Bound to a Protein Interaction Inhibitor*. *J. Biol. Chem.*, 2004. **278**(8): p. 6976-6985.
67. Hughes, F.J. and M.A. Romanos, *E1 protein of human papillomavirus is a DNA helicase/ATPase*. *Nucleic Acids Res.*, 1993. **21**(25): p. 5817-5823.
68. Sanders, C.M. and A. Stenlund, *Recruitment and loading of the E1 initiator protein: an ATP-dependent process catalysed by a transcription factor*. *EMBO J*, 1998. **17**(23): p. 7044-7055.

69. Sedman, J. and A. Stenlund, *The papillomavirus E1 protein forms a DNA-dependent hexameric complex with ATPase and DNA helicase activities*. J. Virol., 1998. **72**(8): p. 6893-6897.
70. Frattini, M.G. and L.A. Laimins, *Binding of the human papillomavirus E1 origin-recognition protein is regulated through complex formation with the E2 enhancer-binding protein*. Proc Natl Acad Sci U S A, 1994. **92**: p. 12398-12402.
71. Abbate, E.A., J.M. Berger, and M.R. Botchan, *The X-ray structure of the papillomavirus helicase in complex with its molecular matchmaker E2*. Genes Dev., 2004. **18**: p. 1981-1996.
72. Kadaja, M., et al., *Mechanism of Genomic Instability in Cells Infected with the High-Risk Human Papillomaviruses*. PLoS Pathog, 2009. **5**(4): p. e1000397.
73. Lee, A.Y. and C.M. Chiang, *Chromatin adaptor Brd4 modulates E2 transcription activity and protein stability*. J. Biol. Chem., 2009. **284**(5): p. 2778-2786.
74. Dey, A., et al., *The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis*. Proc Natl Acad Sci U S A, 2003. **100**: p. 8758-8763.
75. You, J., et al., *Interaction of the Bovine Papillomavirus E2 protein with Brd4 tethers the viral DNA to host mitotic chromosomes*. Cell, 2004. **117**: p. 349-360.
76. Baxter, M.K., et al., *The mitotic chromosome binding activity of the papillomavirus E2 protein correlates with interaction with the cellular chromosomal protein, Brd4* J. Virol., 2004. **79**(8): p. 4806-4818.
77. McPhillips, M.G., et al., *Brd4 is required for E2-mediated transcriptional activation but not genome partitioning of all papillomaviruses*. J. Virol., 2006. **80**(19): p. 9530-9543.
78. Schweiger, M.-R., J. You, and P.M. Howley, *Bromodomain protein 4 mediates the papillomavirus E2 transcriptional activation function*. J. Virol., 2006. **80**(9): p. 4276-4285.
79. Schweiger, M.-R., et al., *Brd4-independent transcriptional repression function of the papillomavirus E2 proteins*. J. Virol., 2007. **81**(18): p. 9612-9622.
80. Lambert, P.F., B.A. Spalholz, and P.M. Howley, *A transcriptional repressor encoded by BPV-1 shares a common carboxy-terminal domain with the E2 transactivator*. Cell, 1987. **50**: p. 69-78.
81. Lambert, P.F., et al., *Genetic assignment of multiple E2 gene products in bovine papillomavirus-transformed cells*. J. Virol., 1989. **63**: p. 3151-3154.
82. Lace, M.J., et al., *The truncated C-terminal E2 (E2-TR) protein of bovine papillomavirus (BPV) type-1 is a transactivator that modulates transcription in vivo and in vitro in a manner distinct from the E2-TA and E8<sup>E2</sup> gene products*. Virology, 2012. **429**: p. 99-111.
83. Lace, M.J., et al., *The E8<sup>E2</sup> gene product of human papillomavirus type 16 represses early transcription and replication but is dispensible for viral plasmid persistence in keratinocytes*. J. Virol., 2008. **82**(21): p. 10841-10853.
84. Skiadopoulos, M.H. and A.A. McBride, *Bovine papillomavirus type 1 genomes and the E2 transactivator protein are closely associated with mitotic chromatin*. J. Virol., 1998. **72**(3): p. 2079-2088.
85. Barsoum, J., et al., *Mechanism of action of the papillomavirus E2 repressor: Repression in the absence of DNA binding*. J. Virol., 1992. **66**(6): p. 3941-3945.
86. McBride, A.A., H. Romanczuk, and P.M. Howley, *The papillomavirus E2 regulatory proteins*. J. Biol. Chem., 1991. **266**(28): p. 18411-18414.
87. Nishimura, A., et al., *Mechanisms of human papillomavirus E2-mediated repression of viral oncogene expression and cervical cancer growth inhibition*. J. Virol., 2000. **74**(8): p. 3752-3760.
88. Dong, G., T.R. Broker, and L.T. Chow, *Human papillomavirus type 11 E2 proteins repress the homologous E6 promoter by interfering with the binding of host transcription factors to adjacent elements*. J. Virol., 1994. **68**(2): p. 1115-1127.

89. Tan, S.H., et al., *The human papillomavirus type 16 E2 transcription factor binds with low cooperativity to two flanking sites and represses the E6 promoter through displacement of Sp1 and TFIID*. J. Virol., 1994. **68**(10): p. 6411-6420.
90. Demeret, C., et al., *Different mechanisms contribute to the E2-mediated transcriptional repression of human papillomavirus type 18 viral oncogenes*. J. Virol., 1997. **71**(12): p. 9343-9349.
91. Pett, M. and N. Coleman, *Integration of high-risk human papillomavirus: a key event in cervical carcinogenesis?* J Pathol, 2007. **212**: p. 356-367.
92. Groves, I.J. and N. Coleman, *Pathogenesis of human papillomavirus-associated mucosal disease*. J Pathol, 2014.
93. Oliveira, J.G., L.A. Colf, and A.A. McBride, *Variations in the association of papillomavirus E2 proteins with mitotic chromosomes*. Proc. Natl. Acad. Sci. U.S.A., 2006. **103**(4): p. 1047-1052.
94. Bastien, N. and A.A. McBride, *Interaction of the papillomavirus E2 protein with mitotic chromosomes*. Virology, 2000. **270**: p. 124-134.
95. Parish, J.L., et al., *ChIR1 is required for loading papillomavirus E2 onto mitotic chromosomes and viral genome maintenance*. Mol. Cell, 2006. **24**: p. 867-876.
96. Feeney, K.M. and J.L. Parish, *Targeting mitotic chromosomes: a conserved mechanism to ensure viral persistence*. Proc. R. Soc. B., 2009. **276**: p. 1535-1544.
97. Shire, K., et al., *EBP2, a human protein that interacts with sequences of the Epstein-Barr virus nuclear antigen 1 important for plasmid maintenance*. J. Virol., 1999. **73**(4): p. 2587-2595.
98. Kapoor, P., B.D. Lavoie, and L. Frappier, *EBP2 plays a key role in Epstein-Barr virus mitotic segregation and is regulated by Aurora family kinases*. Mol. Cell. Biol., 2005. **25**(12): p. 4934-4945.
99. Ballestas, M.E. and K.M. Kaye, *Kaposi's sarcoma-associated herpesvirus latency-associated nuclear antigen 1 mediates episome persistence through cis-acting terminal repeat (TR) sequence and specifically binds TR DNA*. J. Virol., 2001. **75**(7): p. 3250-3258.
100. Ballestas, M.E., P.A. Chatis, and K.M. Kaye, *Efficient persistence of extrachromosomal KSHV DNA mediated by latency-associated nuclear antigen*. Science, 1999. **284**: p. 641-644.
101. You, J., et al., *Kaposi's sarcoma-associated herpesvirus latency-associated nuclear antigen interacts with bromodomain protein Brd4 on host mitotic chromosomes*. J. Virol., 2006. **80**(18): p. 8909-8919.
102. Viejo-Borbolla, A., et al., *Brd2/RING3 interacts with a chromatin-binding domain in the Kaposi's sarcoma-associated herpesvirus latency-associated nuclear antigen 1 (LANA-1) that is required for multiple functions of LANA-1*. J. Virol., 2005. **79**(21): p. 13618-13629.
103. Barbera, A.J., et al., *Kaposi's sarcoma-associated herpesvirus LANA hitchhikes a ride on the chromosome*. Cell Cycle, 2006. **5**(10): p. 1048-1052.
104. Barbera, A.J., et al., *The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA*. Science, 2006. **311**: p. 856-861.
105. Piirsoo, M., et al., *Cis and trans requirements for stable episomal maintenance of the BPV-1 replicator*. EMBO J, 1996. **15**(1): p. 1-11.
106. Dey, A., et al., *A bromodomain protein, MCAP, associates with mitotic chromosomes and affects G(2)-to-M transition*. Mol. Cell. Biol., 2000. **20**: p. 6537-6549.
107. Boner, W., et al., *A functional interaction between the human papillomavirus 16 transcription/replication factor E2 and the DNA damage response protein TopBP1*. J. Biol. Chem., 2002. **277**(25): p. 22297-22303.
108. Donaldson, M.M., W. Boner, and I.M. Morgan, *TopBP1 regulates human papillomavirus type 16 E2 interaction with chromatin*. J. Virol., 2007. **81**(8): p. 4338-4342.
109. Donaldson, M.M., et al., *An interaction between human papillomavirus 16 E2 and TopBP1 is required for optimum viral DNA replication and episomal genome establishment*. J. Virol., 2012. **86**(23): p. 12806-12815.

110. Liras, P., et al., *Characterization of a mutation in yeast causing nonrandom chromosome loss during mitosis*. Genetics, 1977. **88**: p. 651-671.
111. Gerring, S.L., F. Spencer, and P. Hieter, *The CHL1(CTF1) gene product of Saccharomyces cerevisiae is important for chromosome transmission and normal cell cycle progression in G<sub>2</sub>/M*. EMBO J, 1990. **9**(13): p. 4347-4358.
112. Frank, S. and S. Werner, *The human homologue of the yeast CHL1 gene is a novel keratinocyte growth factor-regulated gene*. J. Biol. Chem., 1996. **271**: p. 24337-24340.
113. Amann, J., V.J. Kidd, and J.M. Lahti, *Characterization of putative human homologues of the yeast chromosome transmission fidelity gene, CHL1*. J. Biol. Chem., 1997. **272**: p. 3823-3832.
114. Hirota, Y. and J.M. Lahti, *Characterization of the enzymatic activity of hChR1, a novel DNA helicase*. Nucleic Acids Res., 2000. **28**: p. 917-924.
115. Bharti, S.K., et al., *Molecular Functions and Cellular Roles of the ChR1 (DDX11) Helicase Defective in the Rare Cohesinopathy Warsaw Breakage Syndrome*. Cell. Mol. Life Sci., 2014. **71**: p. 2625-2639.
116. Wu, Y., A.N. Suhasini, and R.M. Brosh Jr., *Welcome to the family of FANC-J-like helicases to the block of genome stability maintenance proteins*. Cell Mol. Life Sci., 2009. **66**(7): p. 1209-1222.
117. Wu, Y., et al., *Biochemical characterisation of Warsaw Breakage syndrome helicase*. J. Biol. Chem., 2012. **287**(2): p. 1007-1021.
118. Feeney, K.M., C.W. Wasson, and J.L. Parish, *Cohesin: a regulator of genome integrity and gene expression*. Biochem. J., 2010. **428**(2): p. 147-161.
119. Haering, C.H., et al., *Molecular architecture of SMC proteins and the yeast cohesin complex*. Mol. Cell, 2002. **9**: p. 773-788.
120. Uhlmann, F., et al., *Cleavage of cohesin by the CD clan protease separin triggers anaphase in yeast*. Cell, 2000. **103**: p. 375-386.
121. Uhlmann, F., *The mechanism of sister chromatid cohesion*. Exp. Cell. Res., 2004. **296**: p. 80-85.
122. Skibbens, R.V., *Chl1p, a DNA helicase-like protein in budding yeast, functions in sister-chromatid cohesion*. Genetics, 2004. **166**: p. 33-42.
123. Ivanov, D., et al., *Eco1 is a novel acetyltransferase that can acetylate proteins involved in cohesion*. Curr. Biol., 2002. **12**: p. 323-328.
124. Parish, J.L., et al., *The DNA helicase ChR1 is required for sister chromatid cohesion in mammalian cells*. J. Cell Sci., 2006. **119**: p. 4857-4865.
125. Holloway, S.L., *CHL1 is a nuclear protein with an essential ATP binding site that exhibits a size-dependent effect on chromosome segregation*. Nucleic Acids Res., 2000. **28**(16): p. 3056-3064.
126. Das, S.P. and P. Sinha, *The budding yeast protein Chl1p has a role in transcriptional silencing, rDNA recombination, and aging*. Biochem Biophys Res Commun, 2005. **337**: p. 167-172.
127. Inoue, A., et al., *Mammalian ChR1 has a Role in Heterochromatin Organization*. Exp. Cell. Res., 2011. **317**(17): p. 2522-2535.
128. Shah, N., et al., *Roles of ChR1 DNA Helicase in Replication Recovery from DNA Damage*. Exp. Cell. Res., 2013. **319**(14): p. 2244-2253.
129. Farina, A., et al., *Studies with the human cohesin establishment factor, ChR1: association of ChR1 with Ctf18-RFC and Fen1*. J. Biol. Chem., 2008. **283**: p. 20925-20936.
130. Leman, A.R. and E. Noguchi, *Local and global functions of Timeless and Tipin in replication fork protection*. Cell Cycle, 2012. **11**(21): p. 3945-3955.
131. Lelij, P.v.d., et al., *Warsaw Breakage Syndrome, a cohesinopathy associated with mutations in the XPD helicase family member DDX11/ChR1*. American Journal of Human Genetics, 2012. **86**: p. 262-266.

132. Capo-Chichi, J.-M., et al., *Identification and Biochemical Characterization of a Novel Mutation in DDX11 Causing Warsaw Breakage Syndrome*. Human Mutations, 2013. **34**(1): p. 103-107.
133. Feeney, K.M., et al., *In vivo analysis of the cell cycle dependent association of the bovine papillomavirus E2 protein and ChIR1*. Virology, 2011. **414**: p. 1-9.
134. Vedadi, M., et al., *Chemical screening methods to identify ligands that promote protein stability, protein crystallisation and structure determination*. Proc Natl Acad Sci U S A, 2006. **103**(43): p. 15835-15840.
135. Niesen, F., H. Berglund, and M. Vedadi, *The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability*. Nature Protocols, 2007. **2**(9): p. 2212-2221.
136. Sawasdichai, A., et al., *In situ Subcellular Fractionation of Adherent and Non-adherent Mammalian Cells*. JoVE, 2010. **41**.
137. White, P.W., et al., *Biphenylsulfonacetic acid inhibitors of the human papillomavirus type 6 E1 helicase inhibit ATP hydrolysis by an allosteric mechanism involving tyrosine 486*. Antimicrob Agents Chemother, 2005. **2005**(49): p. 12.
138. Faucher, A.-M., et al., *Discovery of small-molecule inhibitors of the ATPase activity of human papillomavirus E1 helicase*. J. Med. Chem, 2004. **47**: p. 18-21.
139. Davidson, W., et al., *Characterization of the binding site for inhibitors of the HPV11 E1-E2 protein interaction on the E2 transactivation domain by photoaffinity labeling and mass spectrometry*. Anal. Chem., 2004. **76**: p. 2095-2102.
140. White, P.W., A.-M. Faucher, and N. Goudreau, *Small Molecule Inhibitors of the Human Papillomavirus E1-E2 Interaction*
- Small-Molecule Inhibitors of Protein-Protein Interactions*, L. Vassilev and D. Fry, Editors. 2011, Springer Berlin Heidelberg. p. 61-88.
141. Goudreau, N., et al., *Optimization and determination of the absolute configuration of a series of potent inhibitors of human papillomavirus type-11 E1-E2 protein-protein interaction: A combined medicinal chemistry, NMR and computational chemistry approach*. Bioorg. Med. Chem., 2007. **15**: p. 2690-2700.
142. Lipinski, C.A., et al., *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings*. Adv. Drug Deliv. Rev., 2001. **46**: p. 3-26.
143. Hann, M.M., A.R. Leach, and G. Harper, *Molecular complexity and its impact on the propability of finding leads for drug discovery*. J. Chem. Inf. Comput. Sci., 2001. **41**: p. 856-864.
144. Scott, D.E., et al., *Fragment-based approaches in drug discovery and chemical biology*. Biochemistry, 2012. **51**: p. 4990-5003.
145. Congreve, M., et al., *Recent developments in Fragment-based drug discovery*. J. Med. Chem, 2008. **51**(13): p. 3661-3680.
146. Congreve, M., et al., *A 'rule of three' for fragment-based lead discovery*. DDT, 2003. **8**(19): p. 876-877.
147. Bohacek, R.S., C. McMartin, and W.C. Guida, *The art and practice of structure-based drug design: a molecular modeling perspective*. Med. Res. Rev., 1996. **16**: p. 3-50.
148. Fink, T. and J.-L. Reymond, *Virtual exploration of the chemical universe up to 11 atoms of C, N, O, F: Assembly of 26.4 million structures (110.9 million stereoisomers) and analysis for new ring systems, stereochemistry, physiochemical properties, compound classes and drug discovery*. J. Chem. Inf. Model., 2007. **47**: p. 342-353.
149. Carr, R. and H. Jhoti, *Structure-based screening of low-affinity compounds*. DDT, 2002. **7**(9): p. 522-527.
150. Shuker, S.B., et al., *Discovering high-affinity ligands for proteins: SAR by NMR*. Science, 1996. **274**: p. 1531-1534.

151. Tsai, J., et al., *Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity*. Proc. Natl. Acad. Sci. U.S.A., 2008. **105**: p. 3041-3046.
152. Boivin, S., S. Kozak, and R. Meijers, *Optimization of protein purification and characterisation using Thermofluor screens*. Protein Expr. Purif., 2013. **91**: p. 192-206.
153. Leibly, D.J., et al., *Stabilizing additives added during cell lysis aid in the solubilization of recombinant proteins*. PLoS One, 2012. **7**(12).
154. Mao, Y.J., X.R. Sheng, and X.M. Pan, *The effects of NaCl concentration and pH on the stability of hyperthermophilic protein Ssh10b*. BMC Biochem., 2007. **8**.
155. Baldwin, R.L., *How Hofmeister ion interactions affect protein stability*. Biophys. J., 1996. **71**(4): p. 2056-2063.
156. Erlanson, D.A., R.S. McDowell, and T. O'Brien, *Fragment-based drug discovery*. J. Med. Chem, 2004. **47**(14): p. 3463-3482.
157. Ericsson, U.B., et al., *Thermofluor-based high-throughput stability optimisation of proteins for structural studies*. Anal. Biochem., 2006. **357**: p. 289-298.
158. Major, L.L. and T.K. Smith, *Screening the MayBridge Rule of 3 Fragment Library for Compounds That Interact with the Trypanosoma brucei myo-Inositol-3-Phosphate Synthase and/or Show Trypanocidal Activity*. Molecular Biology International, 2011. **2011**.
159. Kelly, A.E., et al., *Low-conductivity buffers for high-sensitivity NMR measurements*. J. Am. Chem. Soc., 2002. **124**(40): p. 12013-12019.
160. white, P.W., et al., *Inhibition of human papillomavirus DNA replication by small molecule antagonists of the E1-E2 protein interaction*. J. Biol. Chem., 2003. **278**(29): p. 26765-26772.
161. D'Abramo, C.M. and J. Archambault, *Small molecule inhibitors of human papillomavirus protein-protein interactions*. The Open Virology Journal, 2011. **5**: p. 80-95.
162. Pantoliano, M.W., et al., *High-density miniaturized thermal shift assays as a general strategy for drug discovery*. J. Biomol. Screen., 2001. **6**(6): p. 429-440.
163. Lo, M.-C., et al., *Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery*. Anal. Biochem., 2004. **332**(1): p. 153-159.
164. Date, M.S. and B.N. Dominy, *Modeling the influence of salt on the hydrophobic effect and protein fold stability*. Commun. Comput. Phys., 2013. **13**(1): p. 90-106.
165. Basse, N., et al., *Towards the rational design of p53-stabilizing drugs: probing the surface of the oncogenic Y220C mutant*. Chem. Biol., 2010. **17**(1): p. 46-56.
166. Davis, B.J. and D.A. Erlanson, *Learning from our mistakes: the "unknown knows" in fragment screening*. Bioorg Med Chem Lett, 2013. **23**: p. 2844-2852.
167. Meyer, B. and T. Peters, *NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors*. Angew. Chem. Int. Ed., 2003. **42**(8): p. 864-890.
168. Dalvit, C., et al., *WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability*. J. Biomol. NMR, 2001. **21**: p. 349-359.
169. Zartler, E.R. and H. Mo, *Practical aspects of NMR-based fragment discovery*. Current topics in Medicinal Chemistry, 2007. **7**: p. 1592-1599.
170. Ludwig, C. and U.L. Guenther, *Ligand based NMR methods for drug discovery*. Front. Biosci., 2009. **14**: p. 4565-4574.
171. Androphy, E.J., D.R. Lowy, and J.T. Schiller, *Bovine papillomavirus E2 trans-activating gene product binds to specific sites in papillomavirus DNA*. Nature, 1987. **325**: p. 70-73.
172. Kasukawa, H., P.M. Howley, and J.D. Benson, *A fifteen-amino-acid peptide inhibits human papillomavirus E1-E2 interaction and human papillomavirus DNA replication in vitro*. J. Virol., 1998. **72**(10): p. 8166-8173.
173. Sakai, H., et al., *Targeted mutagenesis of the human papillomavirus type 16 E2 transactivation domain reveals separable transcriptional activation and DNA replication functions*. J. Virol., 1996. **70**(3): p. 1602-1611.

174. McPhillips, M.G., et al., *Brd4 is required for E2-mediated transcriptional activation but not genome partitioning of all papillomaviruses*. J. Virol., 2006. **80**(19): p. 9530-9543.
175. de Bernardez Clark, E., *Refolding of recombinant proteins*. Curr. Op. Biotech., 1998. **9**: p. 157-163.
176. Bennion, B.J. and V. Daggett, *The molecular basis for the chemical denaturation of proteins by urea*. Proc. Natl. Acad. Sci. U.S.A., 2003. **100**(9): p. 5142-5147.
177. Yang, Z., et al., *Highly efficient production of soluble proteins from insoluble inclusion bodies by a two-step denaturing and refolding method*. PLoS One, 2011. **6**(7).
178. Wilson, R. and L.A. Laimins, *Differentiation of HPV-containing cells using organotypic "raft" culture or methylcellulose*. Methods Mol. Med., 2005. **119**: p. 157-169.
179. Fradet-Turcotte, A., et al., *Development of quantitative and high-throughput assays of polyomavirus and papillomavirus DNA replication*. Virology, 2010. **399**(1): p. 65-76.
180. Garreau de Loubresse, N., et al., *Structural basis for the inhibition of the eukaryotic ribosome*. Nature, 2014. **513**: p. 517-522.
181. Schneider-Poetsch, T., et al., *Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin*. Nat. Chem. Biol., 2010. **6**(3): p. 209-217.
182. King, L.E., et al., *Human papillomavirus 16 E2 stability and transcriptional activation is enhanced by E1 via a direct protein-protein interaction*. Virology, 2011. **414**(1): p. 26-33.
183. Li, J., et al., *Brd4-mediated nuclear retention of the papillomavirus E2 protein contributes to its stabilization in host cells*. Viruses, 2014. **6**: p. 319-335.
184. Sekhar, V. and A.A. McBride, *Phosphorylation regulates binding of the human papillomavirus type 8 E2 protein to host chromosomes*. J. Virol., 2012. **86**(18): p. 10047-10058.
185. Bellanger, S., et al., *Stability of the human papillomavirus type 18 E2 protein is regulated by a proteasome degradation pathway through its amino-terminal transactivation domain*. J. Virol., 2001. **75**(16): p. 7244-7251.
186. Sakakibara, N., R. Mitra, and A.A. McBride, *The papillomavirus E1 helicase activates a cellular DNA damage response in viral replication foci*. J. Virol., 2011. **85**(17): p. 8981-8995.
187. He, D., J.A. Nickerson, and S. Penman, *Core filaments of the nuclear matrix*. J. Cell. Biol., 1990. **110**: p. 569-580.
188. Fey, E.G., K.M. Wan, and S. Penman, *Epithelial cytoskeletal framework and nuclear matrix-intermediate filament scaffold: three-dimensional organization and protein composition*. J. Cell. Biol., 1984. **98**: p. 1973-1984.
189. Kirnbauer, R., et al., *Efficient self-assembly of human papillomavirus type 16 L1 and L1-L2 into virus-like particles*. J. Virol., 1993. **67**(12): p. 6929-6936.
190. Vance, K.W., M.S. Campo, and I.M. Morgan, *An Enhanced Epithelial Response of a Papillomavirus Promotor to Transcriptional Activators*. J. Biol. Chem., 1999. **274**: p. 27839-27844.
191. Penrose, K.J., et al., *Casein kinase II phosphorylation-induced conformational switch triggers degradation of the papillomavirus E2 protein*. J. Biol. Chem., 2004. **279**: p. 22430-22439.
192. Xue, Y., et al., *GPS 2.1: enhanced prediction of kinase-specific phosphorylation sites with an algorithm of motif length selection*. Protein Eng. Des. Sel., 2011. **24**(3): p. 255-260.
193. Blom, N., S. Gammeltoft, and S. Brunak, *Sequence and structure-based prediction of eukaryotic protein phosphorylation sites*. J. Mol. Biol., 1999. **294**: p. 1351-1362.
194. Blom, N., et al., *Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence*. Proteomics, 2004. **4**: p. 1633-1649.
195. Obenauer, J.C., L.C. Cantley, and M.B. Yaffe, *Scansite 2.0: proteome-wide prediction of cell signalling interactions using short sequence motifs*. Nucleic Acids Res., 2003. **31**(13): p. 3635-3641.
196. Dephoure, N., et al., *Mapping and analysis of phosphorylation sites: a quick guide for cell biologists*. Mol. Biol. Cell, 2013. **24**(5): p. 535-542.

197. Zou, N., et al., *The hinge of the human papillomavirus type 11 E2 protein contains major determinants for nuclear localisation and nuclear matrix association*. J. Virol., 2000. **74**(8): p. 3761-3770.

## Appendix 1. Buffers and solutions

Details of all buffers and solutions used throughout are included here. See Chapter 2, materials and methods for details of usage.

### *SDS PAGE*

#### Acrylamide gels

	8%	10%	12%	15%	Stacking (6%)
ddH <sub>2</sub> O	4.65 mL	3.95 mL	3.3 mL	2.3 mL	2.05 mL
1.5 M Tris pH 8.8	2.5 mL	2.5 mL	2.5 mL	2.5 mL	-
0.5 M Tris pH 6.8	-	-	-	-	375 µL
30% Bis-Acrylamide	2.65 mL	3.35 mL	4 mL	5 mL	0.5 mL
10% SDS	100 µL	100 µL	100 µL	100 µL	30 µL
10% APS	100 µL	100 µL	100 µL	100 µL	30 µL
TEMED	6 µL	4 µL	4 µL	4 µL	3 µL

**Table 2.2 Solutions for preparing acrylamide gels for SDS-PAGE**

All solution volumes are for 1 gel (10 mL resolving, 3 mL stack). TEMED was added immediately prior to pouring. The resolving gel was poured and topped with isopropanol until set. The isopropanol was removed before the stack was poured. Gels were poured and used on the same day.

#### 6x SDS-gel loading buffer

300 mM Tris pH 6.8

12% w/v SDS

60% v/v glycerol

600 mM DTT

Bromophenol blue

Aliquot and store at -20 °C.

#### 10x SDS running buffer

250 mM Tris

250 mM glycine  
1% w/v SDS  
Make up to 1 L with ddH<sub>2</sub>O  
Dilute 1:10 in ddH<sub>2</sub>O for use.

### *Protein purification*

Protease inhibitor cocktail (Sigma) (1% by volume) and 0.1 mM DTT were added to each buffer immediately prior to use. Buffers were stored at 4 °C.

#### Lysis buffer A (E2 TAD)

50 mM Tris.HCl pH 7.9  
500 mM NaCl  
10 mM imidazole  
10% (v/v) glycerol

#### Lysis buffer B (ChIR1 1-130)

25 mM Tris pH 8  
500 mM NaCl  
10% (v/v) glycerol  
0.1% (v/v) Triton-X-100

#### Lysis buffer C (ChIR1 63-214, 63-222)

50 mM sodium phosphate pH 6.5  
150 mM NaCl

#### Lysis buffer D (MBP-His-ChIR1 63-222)

25 mM Tris pH 7.4  
250 mM NaCl  
2 mM EDTA  
1% lysozyme

Lysis buffer E (GST-ChIR1 peptides)

25 mM Tris pH 7.5

150 mM NaCl

2 mM EDTA

Lysis buffer F

25 mM Tris pH 8

250 mM NaCl

2 mM EDTA

10% glycerol

0.5% Triton-X-100

1% lysozyme

5 mM DTT

Triton lysis buffer

50 mM Tris pH 8

50 mM NaCl

2% Triton-X-100

1.6 M urea

Urea lysis buffer 1

50 mM Tris pH 8

50 mM NaCl

8 M urea

10 mM DTT

Refolding buffer

25 mM Tris pH 8

250 mM NaCl

2 mM EDTA

5 mM DTT

SE buffer (size exclusion chromatography)

25 mM Tris pH 7.5

500mM NaCl

NMR buffer

50 mM sodium phosphate pH 7.5 + 10% d<sub>2</sub>O

*Cell lysis buffers*

All IP buffers were supplemented with the specified concentration DTT and 1% protease inhibitor cocktail immediately prior to use.

IP lysis buffer

50 mM Tris HCl pH 7.4

100 mM NaCl

20 mM NaF

10 mM KH<sub>2</sub>PO<sub>4</sub>

1% Triton-x-100

0.1 mM DTT

IP binding buffer

50 mM Tris HCl pH 7.4

100 mM KCl

0.1 mM EDTA

0.2% NP40

0.1% BSA

2.5% glycerol

2 mM DTT

IP wash buffer

100 mM Tris HCl pH 7.4  
100 mM NaCl  
0.5% NP40  
2 mM DTT

#### Urea Lysis buffer 2

50 mM Tris HCl pH 7.4  
8 M urea  
1 mM DTT + 1% protease inhibitors were added immediately prior to use

#### *DNA buffers*

##### DNA lysis buffer

10 mM Tris-HCl pH 7.4  
400 mM NaCl  
10 mM EDTA  
Autoclaved to sterilise before use, stored at RT

#### *Southern Blotting*

##### 20 x SSC buffer

3 M sodium chloride  
0.34 M tri-sodium citrate  
800 mL ddH<sub>2</sub>O  
pH 7.0; made up to 1L with ddH<sub>2</sub>O

##### 50 x Denhardts

1 g Ficoll 400 (1% w/v)  
1 g Polyvinylpyrrolidone  
1 g BSA  
Made up to 100 mL with ddH<sub>2</sub>O

### 2 x hybridisation buffer

10 mL 50x Denhardt's

20 mL 20x SSC buffer

10 g dextran sulphate

Made up to 50 mL with ddH<sub>2</sub>O and filtered through an 8.0 µM syringe filter

For 1x hybridisation buffer, dilute 1:1 with deionised formamide and add 2 mL 10% SDS. Mix well before use.

### Stringency wash buffers

1. 2 x SSC + 0.1% w/v SDS
2. 0.5 x SSC + 0.1% w/v SDS
3. 0.1 x SSC + 0.1% w/v SDS
4. 0.1 x SSC + 1% w/v SDS

### *Immunofluorescence*

#### Pre-extraction buffer

80 mM PIPES pH 6.8

5 mM EGTA

1 mM MgCl<sub>2</sub>

0.1% Triton-x-100

#### HINGS block solution

20 % heat inactivated goat serum (HINGS)

0.1% BSA, sterile filtered before use

Made up to 50 mL with PBS

Sterile filtered and stored at 4 °C

#### BSA block solution

3% BSA in PBS

Sterile filtered before use and stored at 4 °C

*In-situ fractionation buffers*

CSK buffer

10 mM PIPES pH 6.8

300 mM sucrose

100 mM NaCl

3 mM MgCl<sub>2</sub>

1 mM EGTA

Buffer was prepared fresh and sterile filtered (0.2 µm syringe filter) prior to use

CSK-1: CSK + 0.1% Triton-X-100

CSK-2: CSK + 0.5% Triton-X-100

CSK-3: CSK + 100 µg/mL DNase I

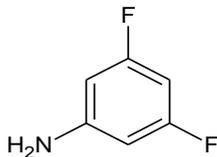
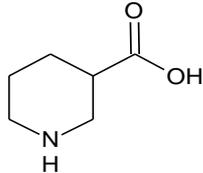
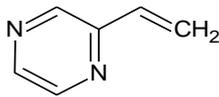
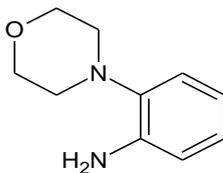
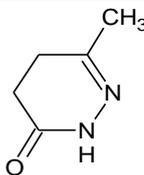
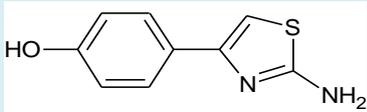
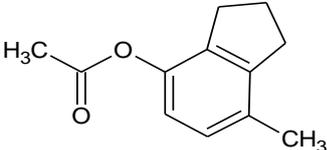
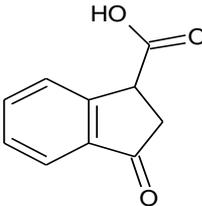
TES buffer

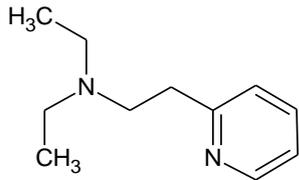
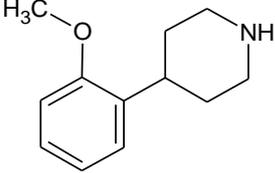
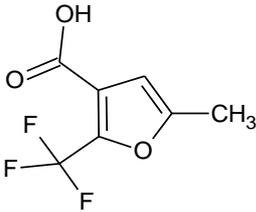
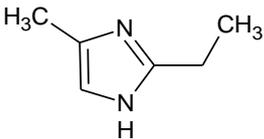
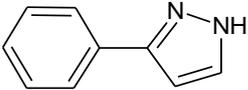
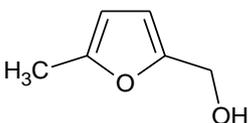
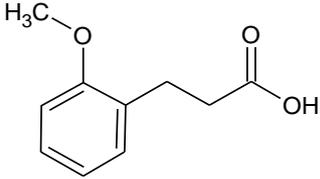
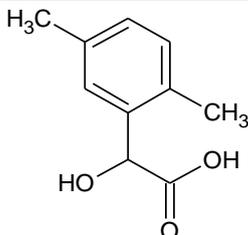
20 mM Tris-HCl pH 7.4

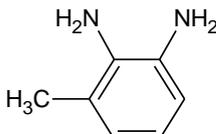
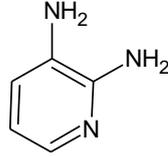
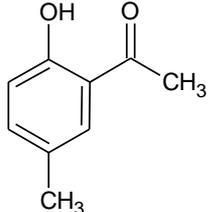
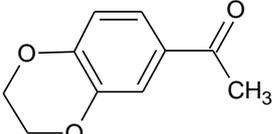
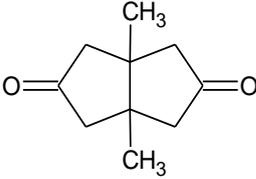
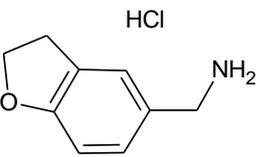
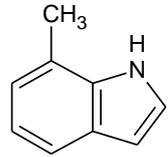
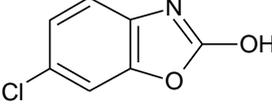
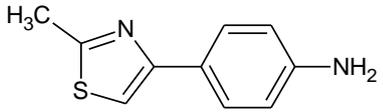
1mM EDTA

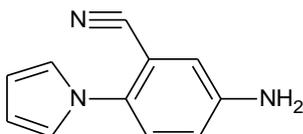
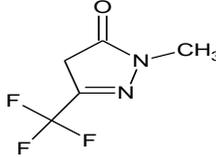
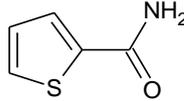
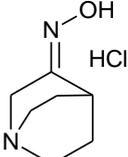
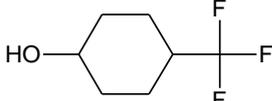
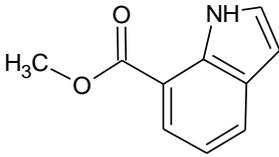
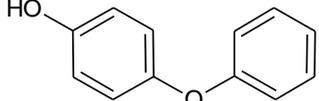
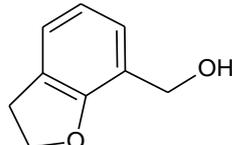
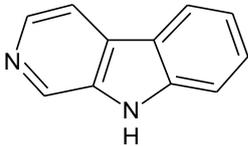
1% SDS

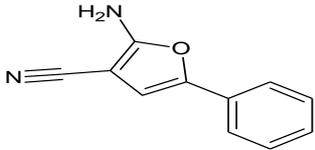
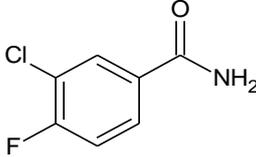
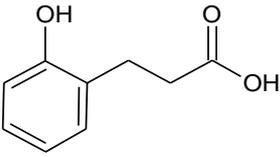
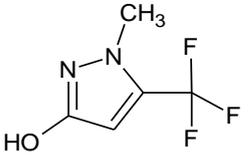
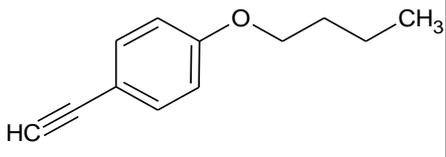
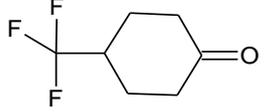
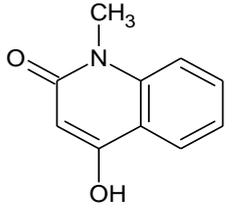
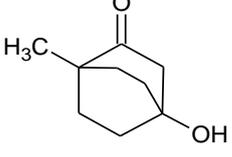
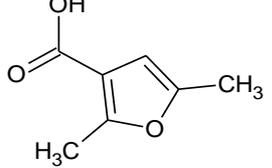
## Appendix 2. Details of NMR fragment library

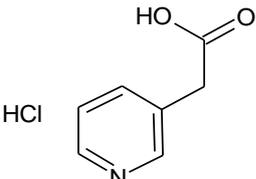
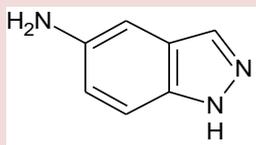
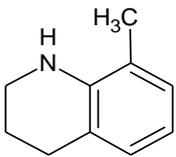
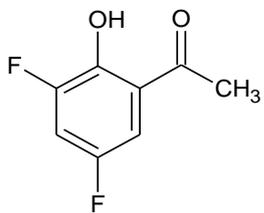
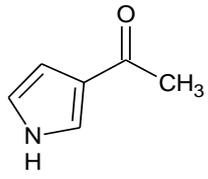
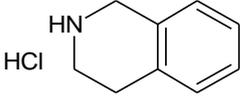
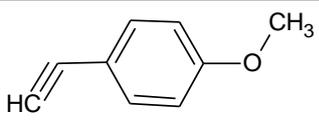
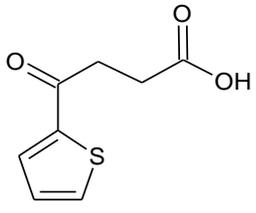
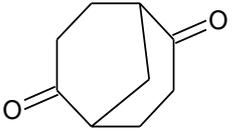
ID	Name	CAS	Structure	Tm shift/°C	NMR
1A1	3,5-difluoroaniline	372-39-4		1.48	N/A
1A2	3-piperidinecarboxylic acid	498-95-3		1.29	N/A
1A3	2-vinylpyrazine	4177-16-6		0.86	N/A
1A4	2-morpholinoaniline	5585-33-1		1.47	-
1A5	6-methyl-2,3,4,5-tetrahydropyridazin-3-one	5157-08-4		1.11	-
1A6	4-(2-amino-1,3-thiazol-4-yl)phenol	57634-55-6		7.37	-
1A7	7-methyl-2,3-dihydro-1H-inden-4-yl acetate	175136-12-6		1.35	-
1A8	3-oxoindane-1-carboxylic acid	29427-69-8		1.18	-

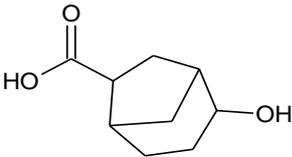
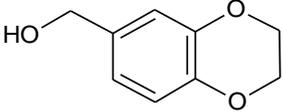
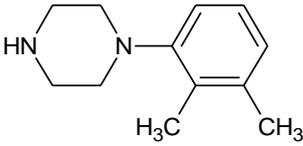
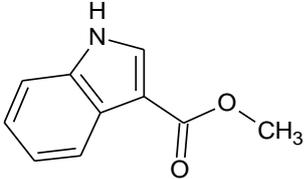
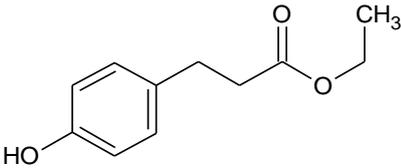
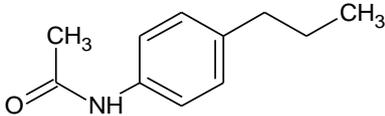
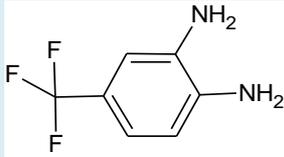
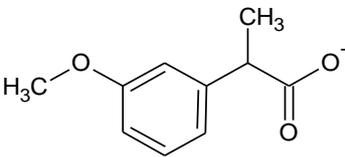
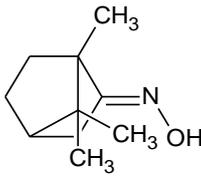
1A9	N,N-diethyl-N-[2-(2-pyridyl)ethyl]amine	25877-30-9		1.00	-
1A10	4-(2-methoxyphenyl)piperidine	58333-75-8		1.08	-
1A11	5-methyl-2-(trifluoromethyl)-3-furoic acid	17515-74-1		1.60	-
1B1	2-ethyl-4-methyl-1H-imidazole	931-36-2		0.93	-
1B2	3-phenyl-1H-pyrazole	2458-26-6		0.46	-
1B3	(5-methyl-2-furyl)methanol	3857-25-8		* -2.22	N/A
1B4	3-(2-methoxyphenyl)propanoic acid	6342-77-4		1.37	-
1B5	2-(2,5-dimethylphenyl)-2-hydroxyacetic acid	5766-40-5		1.49	-

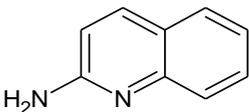
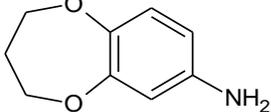
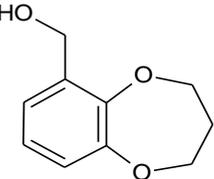
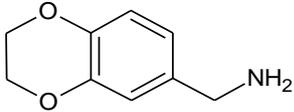
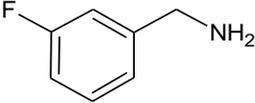
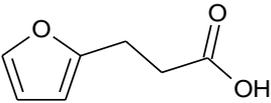
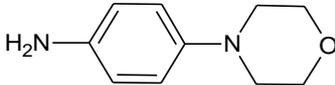
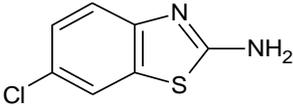
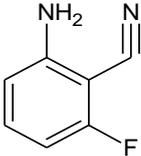
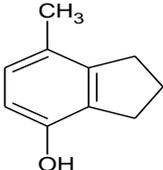
1B6	3-methylbenzene-1,2-diamine	2687-25-4		* 7.75	-
1B7	pyridine-2,3-diamine	452-58-4		* -2.2	-
1B8	1-(2-hydroxy-5-methylphenyl)ethan-1-one	1450-72-2		1.11	-
1B9	1-(2,3-dihydro-1,4-benzodioxin-6-yl)ethan-1-one	2879-20-1		0.84	-
1B10	3a,6a-dimethylperhydropentalene-2,5-dione	21170-10-5		1.39	-
1B11	2,3-dihydrobenzo[b]furan-5-ylmethylamine hydrochloride	55745-74-9		1.51	-
1C1	7-methyl-1H-indole	933-67-5		0.19	*
1C2	6-chloro-1,3-benzoxazol-2-ol	19932-84-4		0.85	-
1C3	4-(2-methyl-1,3-thiazol-4-yl)phenylamine	25021-49-2		0.63	-

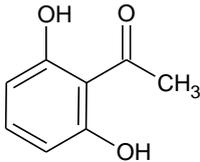
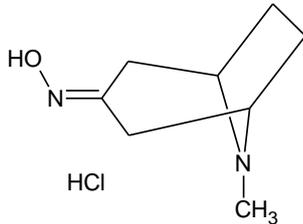
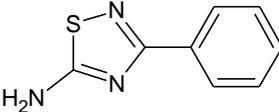
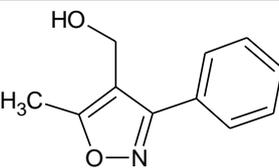
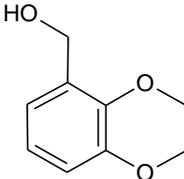
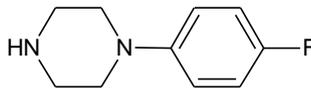
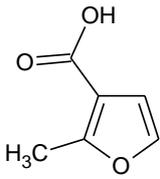
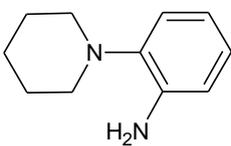
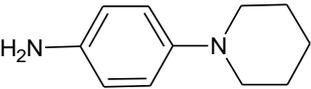
1C4	5-amino-2-(1H-pyrrol-1-yl)benzonitrile	106981-51-5		0.89	-
1C5	1-methyl-3-(trifluoromethyl)-4,5-dihydro-1H-pyrazol-5-one	1481-02-3		0.52	N/A
1C6	thiophene-2-carboxamide	5813-89-8		0.79	-
1C7	quinuclidine-3-one oxime hydrochloride	76883-37-9		1.03	N/A
1C8	4-trifluoromethyl-cyclohexan-1-ol	30129-18-1		1.11	N/A
1C9	methyl 1H-indole-7-carboxylate	93247-78-0		0.77	-
1C10	4-phenoxyphenol	831-82-3		1.16	*
1C11	2,3-Dihydrobenzo[b]furan-7-methanol	151155-53-2		0.75	-
1D1	9H-beta-carboline	244-63-3		1.38	-
1D2	3-methyl-1,3-benzoxazole-2(3H)-thione	13673-63-7		1.37	-

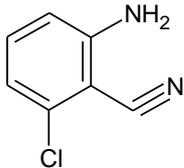
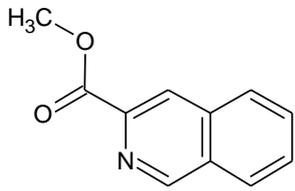
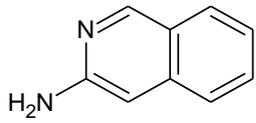
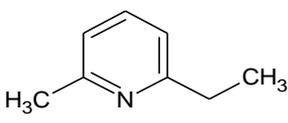
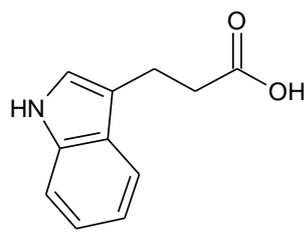
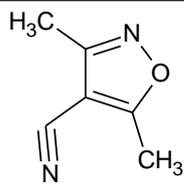
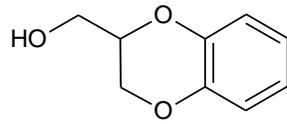
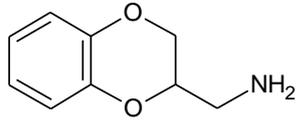
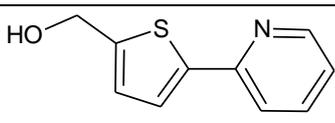
1D3	2-amino-5-phenyl-3-furonitrile	14742-32-6		1.63	N/A
1D4	3-chloro-4-fluorobenzamide	701-43-9		1.06	N/A
1D5	3-(2-hydroxyphenyl)propanoic acid	495-78-3		1.01	-
1D6	1-methyl-5-(trifluoromethyl)-1H-pyrazol-3-ol	119022-51-4		1.27	-
1D7	1-butoxy-4-ethynylbenzene	79887-15-3		0.97	N/A
1D8	4-(trifluoromethyl)cyclohexan-1-one	75091-99-5		1.08	-
1D9	4-hydroxy-1-methyl-1,2-dihydroquinolin-2-one	1677-46-9		0.27	-
1D10	4-hydroxy-1-methylbicyclo[2.2.2]octan-2-one	5122-77-0		1.35	*
1D11	2,5-dimethyl-3-furoic acid	636-44-2		* -2.18	-

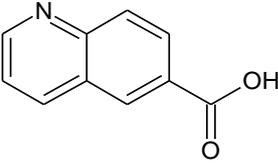
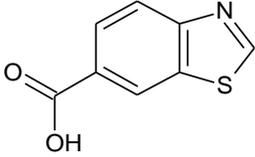
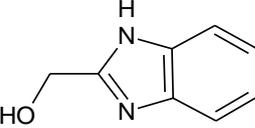
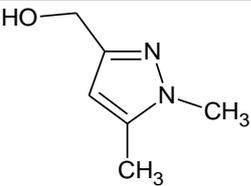
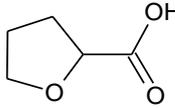
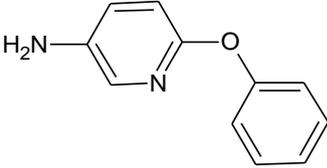
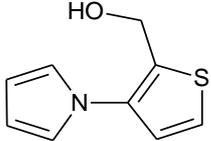
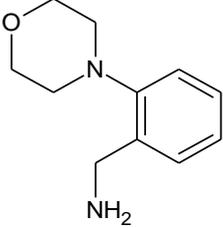
1E1	pyridin-3-ylacetic acid hydrochloride	6419-36-9	HCl 	1.31	-
1E2	1H-indazol-5-amine	19335-11-6		-3.36	-
1E3	8-methyl-1,2,3,4-tetrahydroquinoline	52601-70-4		0.82	-
1E4	1-(3,5-difluoro-2-hydroxyphenyl)ethan-1-one	140675-42-9		1.03	-
1E5	1-(1H-pyrrol-3-yl)ethan-1-one	1072-82-8		0.35	-
1E6	1,2,3,4-tetrahydroisoquinoline hydrochloride	14099-81-1	HCl 	0.87	-
1E7	1-eth-1-ynyl-4-methoxybenzene	768-60-5		0.83	-
1E8	4-oxo-4-(2-thienyl)butanoic acid	4653-08-1		0.85	-
1E9	bicyclo[3.3.1]nonane-2,6-dione	16473-11-3		1.02	N/A

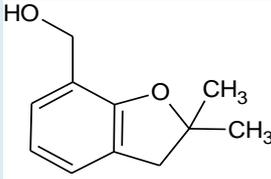
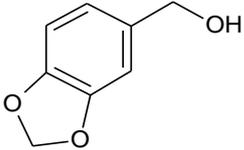
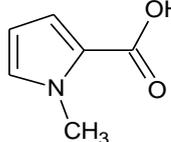
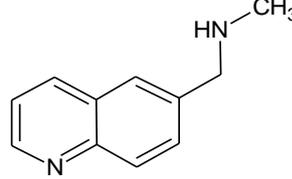
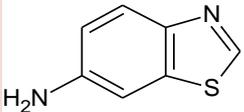
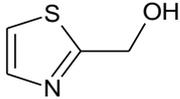
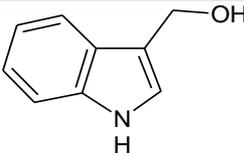
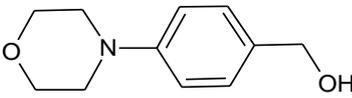
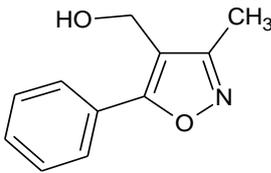
1E10	2-hydroxybicyclo[3.2.1]octane-6-carboxylic acid	257932-29-9		1.37	N/A
1E11	2,3-dihydro-1,4-benzodioxin-6-ylmethanol	39270-39-8		0.85	-
1F1	1-(2,3-dimethylphenyl)piperazine	1013-22-5		0.54	-
1F2	methyl 1H-indole-3-carboxylate	942-24-5		1.03	-
1F3	ethyl 3-(4-hydroxyphenyl)propanoate	23795-02-0		0.74	-
1F4	N1-(4-propylphenyl)acetamide	20330-99-8		0.64	N/A
1F5	4-(trifluoromethyl)benzene-1,2-diamine	368-71-8		2.90	-
1F6	methyl-2-(3-methoxyphenyl)acetate	18927-05-4		0.88	N/A
1F7	1,7,7-trimethylbicyclo[2.2.1]heptan-2-one oxime	13559-66-5		0.72	-

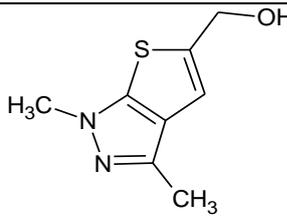
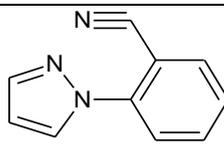
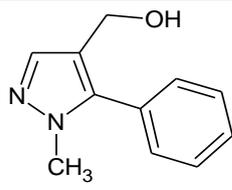
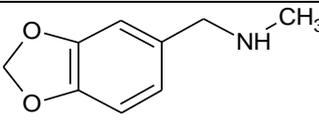
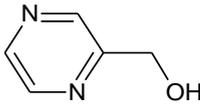
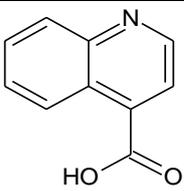
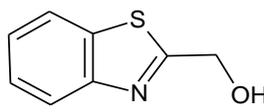
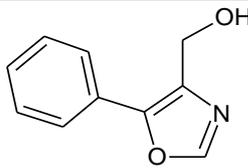
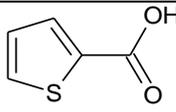
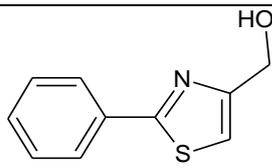
1F8	quinolin-2-amine	580-22-3		0.58	-
1F9	3,4-dihydro-2H-1,5-benzodioxepin-7-amine	175136-34-2		1.19	-
1F10	3,4-dihydro-2H-1,5-benzodioxepin-6-ylmethanol	499770-81-9		1.13	-
1F11	2,3-dihydro-1,4-benzodioxin-6-ylmethanamine	17413-10-4		-1.36	-
1G1	3-fluorobenzylamine	100-82-3		-1.59	-
1G2	3-(2-furyl)propionic acid	935-13-7		-0.17	N/A
1G3	4-morpholinoaniline	2524-67-6		* -3.14	-
1G4	6-chloro-1,3-benzothiazol-2-amine	95-24-9		-0.29	-
1G5	2-amino-6-fluorobenzonitrile	77326-36-4		0.03	-
1G6	7-methylindan-4-ol	16400-13-8		1.40	-

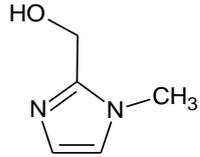
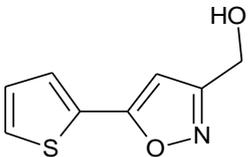
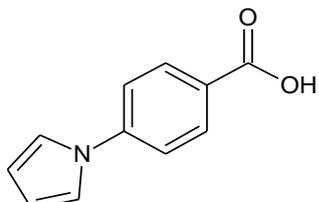
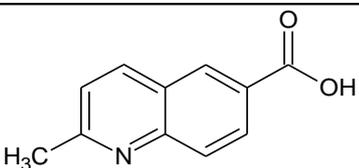
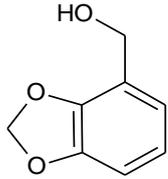
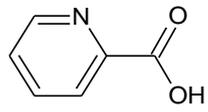
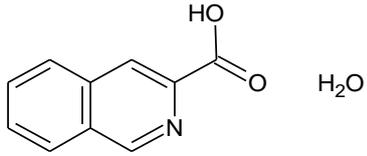
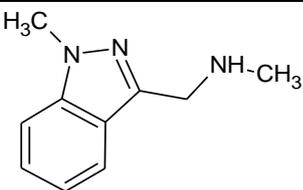
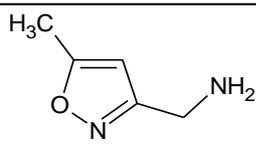
1G7	1-(2,6-dihydroxyphenyl)ethan-1-one	699-83-2		0.92	-
1G8	8-methyl-8-azabicyclo[3.2.1]octane-3-one oxime hydrochloride	212079-30-6		0.68	N/A
1G9	3-phenyl-1,2,4-thiadiazol-5-amine	17467-15-1		0.06	-
1G10	(5-methyl-3-phenyl-4-isoxazolyl)methanol	18718-79-1		0.06	-
1G11	2,3-dihydro-1,4-benzodioxin-5-ylmethanol	274910-19-9		0.94	-
1H1	1-(4-fluorophenyl)piperazine	2252-63-3		-0.34	-
1H2	2-methyl-3-furoic acid	6947-94-0		-0.77	-
1H3	2-piperidinoaniline	39643-31-7		0.03	-
1H4	4-piperidinoaniline	2359-60-6		N/A	-

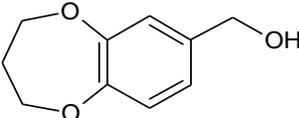
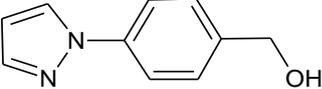
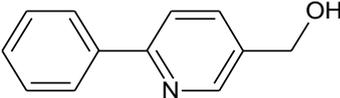
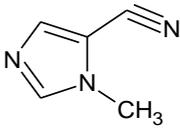
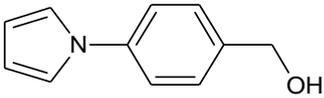
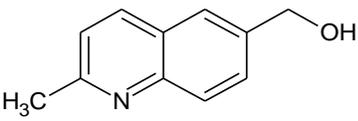
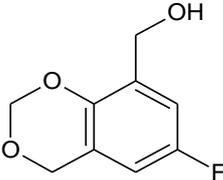
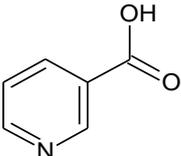
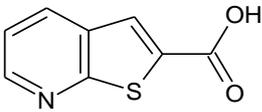
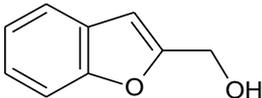
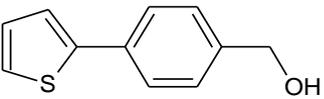
1H5	2-amino-6-chlorobenzonitrile	6575-11-7		0.93	-
1H6	methyl isoquinoline-3-carboxylate	27104-73-0		0.08	-
1H7	isoquinolin-3-amine	25475-67-6		1.13	-
1H8	2-ethyl-6-methylpyridine	1122-69-6		0.52	N/A
1H9	3-(1H-indol-3-yl)propanoic acid	830-96-6		1.17	-
1H10	3,5-dimethyl-4-isoxazolecarbonitrile	31301-46-9		0.17	-
1H11	2,3-dihydro-1,4-benzodioxin-2-ylmethanol	3663-82-9		0.69	-
2A1	2,3-dihydro-1,4-benzodioxin-2-ylmethanamine	4442-59-5		* -2.54	-
2A2	[5-(2-pyridinyl)-2-thienyl]methanol	197899-76-6		-0.40	-

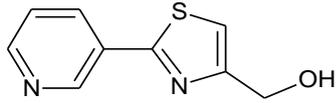
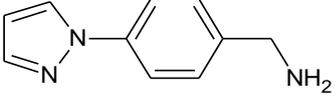
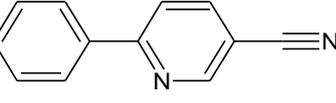
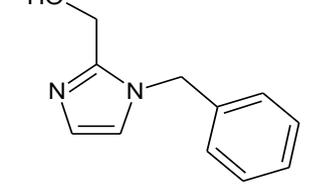
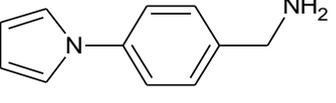
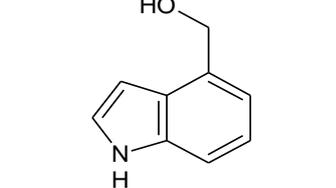
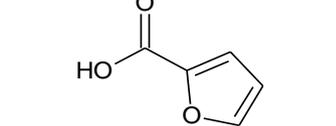
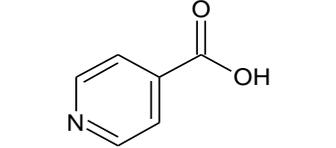
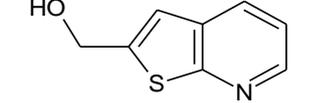
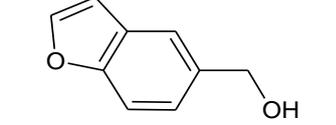
2A3	quinoline-6-carboxylic acid	10349-57-2		0.40	-
2A4	1,3-benzothiazole-6-carboxylic acid	3622-35-3		* 9.34	-
2A5	1H-benzimidazol-2-ylmethanol	4856-97-7		0.38	-
2A6	(1,5-dimethyl-1H-pyrazol-3-yl)methanol	153912-60-8		0.10	-
2A7	tetrahydro-2-furancarboxylic acid	16874-33-2		0.78	-
2A8	6-phenoxy-3-pyridinamine	25194-67-6		-0.35	-
2A9	[3-(1H-pyrrol-1-yl)-2-thienyl]methanol	107073-27-8		0.09	-
2A10	2-morpholinobenzylamine	204078-48-8		-0.15	N/A

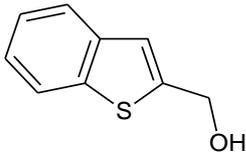
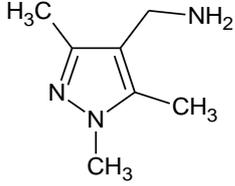
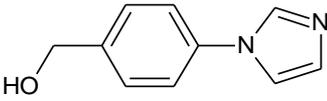
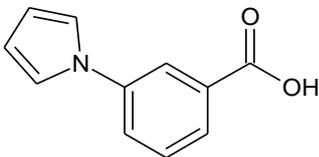
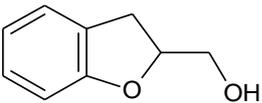
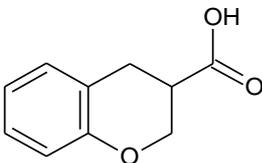
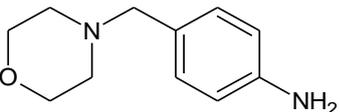
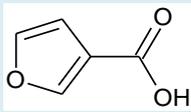
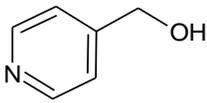
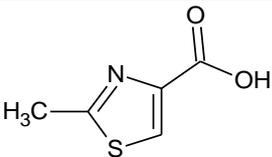
2A11	(2,2-dimethyl-2,3-dihydro-1-benzofuran-7-yl)methanol	38002-89-0		2.03	-
2B1	1,3-benzodioxol-5-ylmethanol	495-76-1		0.37	-
2B2	1-methyl-1H-pyrrole-2-carboxylic acid	6973-60-0		0.45	-
2B3	N-methyl-N-(quinolin-6-yl)methylamine	179873-36-0		14.10	-
2B4	1,3-benzothiazol-6-amine	533-30-2		-4.98	-
2B5	1,3-thiazol-2-ylmethanol	14542-12-2		-1.38	-
2B6	1H-indol-3-ylmethanol	700-06-1		3.40	-
2B7	(4-morpholin-4-yl-phenyl)-methanol	280556-71-0		0.29	-
2B8	(3-methyl-5-phenyl-4-isoxazolyl)methanol	113826-87-2		-0.14	-

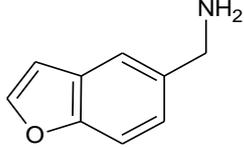
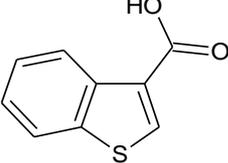
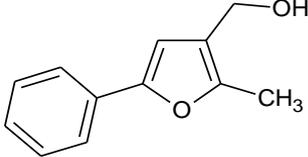
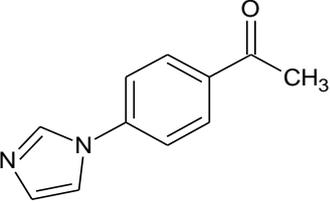
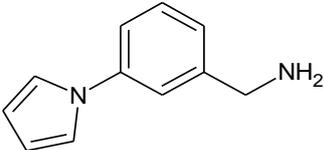
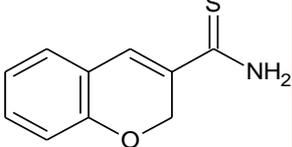
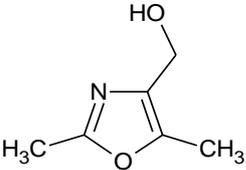
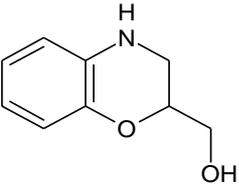
2B9	(1,3-dimethyl-1H-thieno[2,3-c]pyrazol-5-yl)methanol	423769-75-9		-1.36	-
2B10	2-(1H-pyrazol-1-yl)benzotrile	25775-03-5		-1.62	-
2B11	(1-methyl-5-phenyl-1H-pyrazol-4-yl)methanol	499785-47-6		0.02	*
2C1	N-(1,3-benzodioxol-5-ylmethyl)-N-methylamine	15205-27-3		-0.64	-
2C2	2-pyrazinylmethanol	6705-33-5		0.40	-
2C3	quinoline-4-carboxylic acid	486-74-8		0.27	-
2C4	1,3-benzothiazol-2-ylmethanol	37859-42-0		0.29	-
2C5	(5-phenyl-1,3-oxazol-4-yl)methanol	352018-88-3		0.45	-
2C6	thiophene-2-carboxylic acid	527-72-0		0.67	-
2C7	(2-phenyl-1,3-thiazol-4-yl)methanol	23780-13-4		-0.54	-

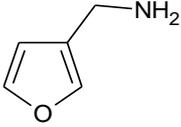
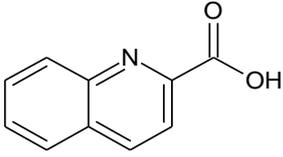
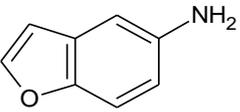
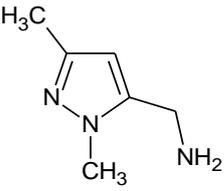
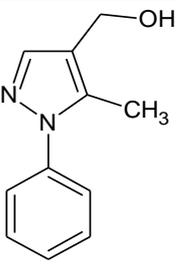
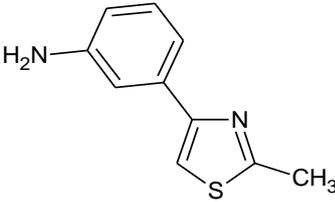
2C8	(1-methyl-1H-imidazol-2-yl)methanol	17334-08-6		-0.05	-
2C9	[5-(2-thienyl)-3-isoxazolyl]methanol	194491-44-6		-0.15	-
2C10	4-(1H-pyrrol-1-yl)benzoic acid	22106-33-8		-0.40	-
2C11	2-methyl-6-quinolinecarboxylic acid	635-80-3		0.12	-
2D1	1,3-benzodioxol-4-ylmethanol	769-30-2		0.90	w
2D2	2-pyridinecarboxylic acid	98-98-6		0.50	-
2D3	isoquinoline-3-carboxylic acid hydrate	6624-49-3		0.11	-
2D4	N-methyl-N-[(1-methyl-1H-indazol-3-yl)methyl]amine	124491-38-9		-0.46	-
2D5	(5-methyl-3-isoxazolyl)methylamine	154016-48-5		-0.50	-

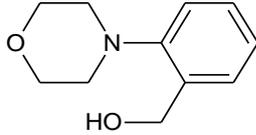
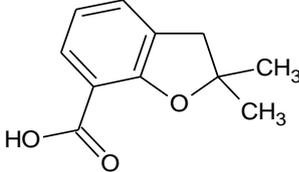
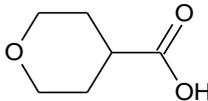
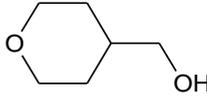
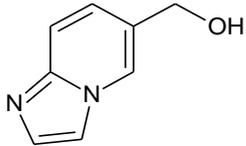
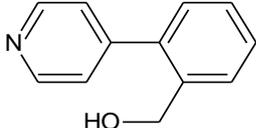
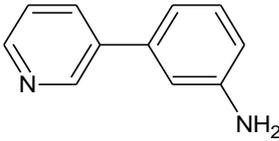
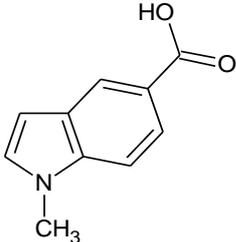
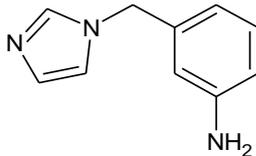
2D6	3,4-dihydro-2H-1,5-benzodioxepin-7-ylmethanol	62823-14-7		0.28	-
2D7	[4-(1H-pyrazol-1-yl)phenyl]methanol	143426-49-7		-0.08	-
2D8	(6-phenyl-3-pyridinyl)methanol	4634-09-7		-0.71	-
2D9	1-methyl-1H-imidazole-5-carbonitrile	66121-66-2		-0.35	-
2D10	[4-(1H-pyrrol-1-yl)phenyl]methanol	143426-51-1		-1.04	-
2D11	(2-methyl-6-quinolinyl)methanol	108166-02-5		0.03	-
2E1	(6-fluoro-4H-1,3-benzodioxin-8-yl)methanol	306934-89-4		0.38	-
2E2	nicotinic acid	59-67-6		0.28	-
2E3	thieno[2,3-b]pyridine-2-carboxylic acid	59944-76-2		0.11	-
2E4	1-benzofuran-2-ylmethanol	55038-01-2		0.08	-
2E5	(4-thien-2-ylphenyl)methanol	81443-44-9		0.35	-

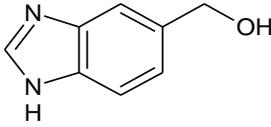
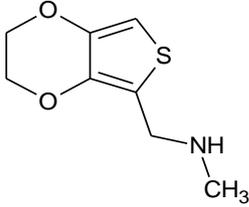
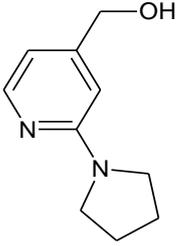
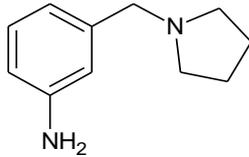
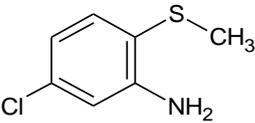
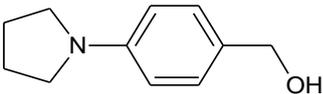
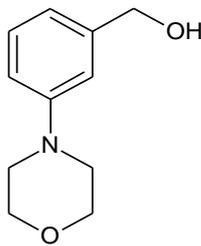
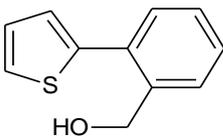
2E6	(2-pyridin-3-yl-1,3-thiazol-4-yl)methanol	138745-99-0		0.24	-
2E7	4-(1H-pyrazol-1-yl)benzylamine	368870-03-5		-0.31	-
2E8	6-phenylnicotinonitrile	39065-54-8		-0.14	-
2E9	(1-benzyl-1H-imidazol-2-yl)methanol	5376-10-3		0.14	-
2E10	4-(1H-pyrrol-1-yl)benzylamine	465514-27-6		0.48	-
2E11	1H-indol-4-ylmethanol	1074-85-7		0.78	-
2F1	2-furoic acid	88-14-2		-1.23	-
2F2	isonicotinic acid	55-22-1		1.81	-
2F3	thieno[2,3-b]pyridin-2-ylmethanol	131337-81-0		-1.91	-
2F4	1-benzofuran-5-ylmethanol	31823-05-9		0.30	-

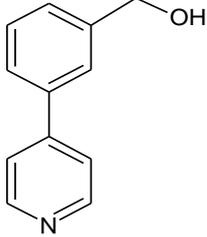
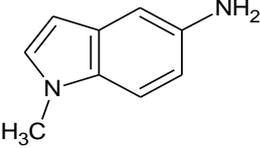
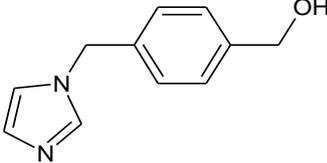
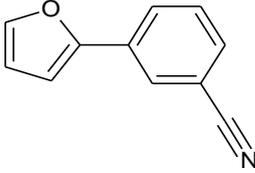
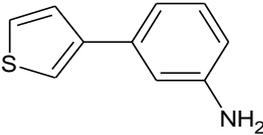
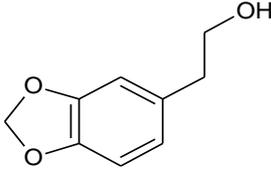
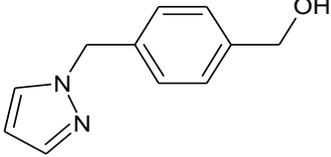
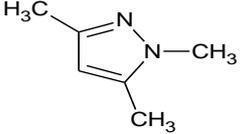
2F5	1-benzothiophen-2-ylmethanol	17890-56-1		-0.34	-
2F6	(1,3,5-trimethyl-1H-pyrazol-4-yl)methylamine	352018-93-0		-0.43	*
2F7	[4-(1H-imidazol-1-yl)phenyl]methanol	86718-08-3		0.12	-
2F8	3-(1H-pyrrol-1-yl)benzoic acid	61471-45-2		-0.50	-
2F9	2,3-dihydro-1-benzofuran-2-ylmethanol	66158-96-1		-0.15	-
2F10	3-chromanecarboxylic acid	115822-57-6		0.33	-
2F11	4-(morpholinomethyl)aniline	51013-67-3		-0.14	-
2G1	3-furoic acid	488-93-7		* 6.94	-
2G2	4-pyridinylmethanol	586-95-8		-0.47	-
2G3	2-methyl-1,3-thiazole-4-carboxylic acid	35272-15-2		0.19	-

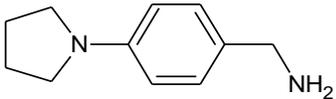
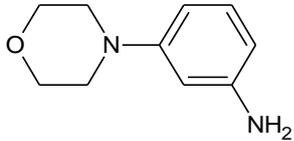
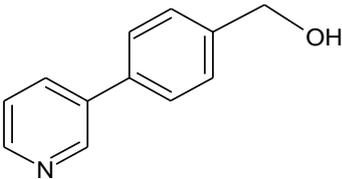
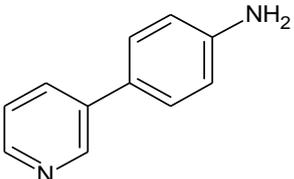
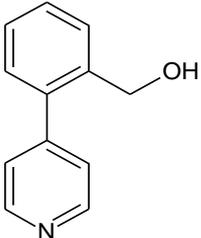
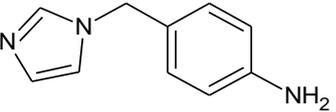
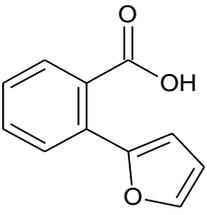
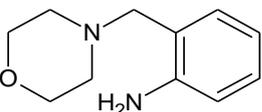
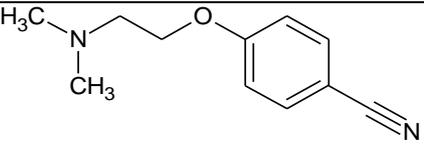
2G4	1-benzofuran-5-yl-methylamine	37798-08-6		-0.94	N/A
2G5	1-benzothiophene-3-carboxylic acid	5381-25-9		-0.13	-
2G6	(2-methyl-5-phenyl-3-furyl)methanol	111787-91-8		-4.12	-
2G7	1-[4-(1H-imidazol-1-yl)phenyl]-1-ethanone	10041-06-2		0.11	-
2G8	3-(1H-pyrol-1-yl)benzylamine	368869-95-8		-0.77	N/A
2G9	2H-chromene-3-carbothioamide	423768-57-4		-3.31	-
2G10	(2,5-dimethyl-1,3-oxazol-4-yl)methanol	92901-94-5		-0.21	-
2G11	3,4-dihydro-2H-1,4-benzoxazin-2-ylmethanol	82756-74-9		-0.70	-

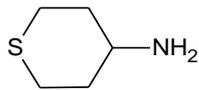
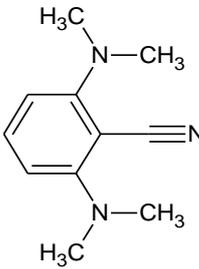
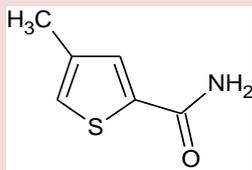
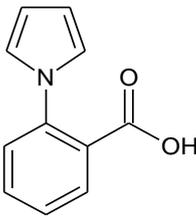
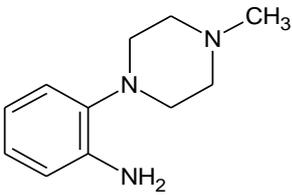
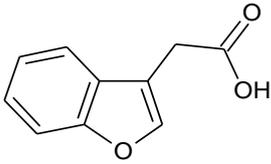
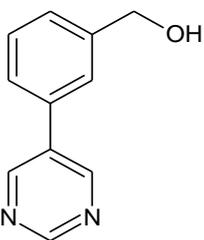
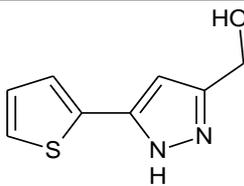
2H1	3-furylmethylamine	4543-47-9		-0.32	N/A
2H2	2-quinolinecarboxylic acid	93-10-7		-0.5	N/A
2H3	(2,4-dimethyl-1,3-thiazol-5-yl)methanol	50382-32-6		-1.31	N/A
2H4	1-benzofuran-5-amine	58546-89-7		* 46.38	*
2H5	(1,3-dimethyl-1H-pyrazol-5-yl)methylamine	499770-63-7		-1.77	-
2H6	(5-methyl-1-phenyl-1H-pyrazol-4-yl)methanol	153863-35-5		-0.26	-
2H7	3-(2-methyl-1,3-thiazol-4-yl)aniline	89250-34-0		0.13	w
2H8	3-(1H-pyrrol-1-yl)thiophene-2-carboxylic acid	74772-17-1		-5.87	-

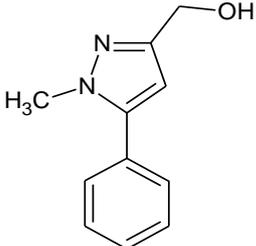
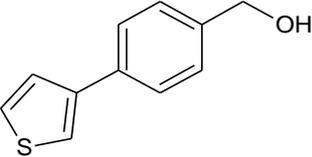
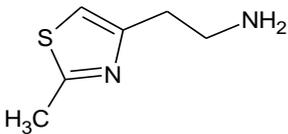
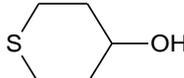
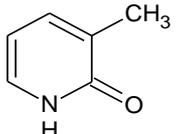
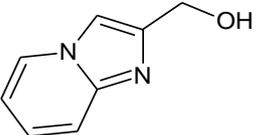
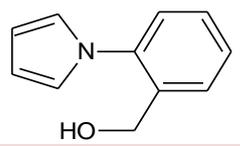
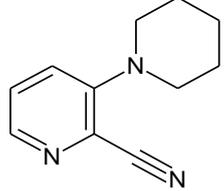
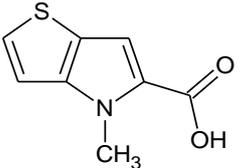
2H9	(2-morpholinophenyl)methanol	465514-33-4		0.44	-
2H10	2,2-dimethyl-2,3-dihydro-1-benzofuran-7-carboxylic acid	42327-95-7		N/A	-
2H11	tetrahydro-2H-pyran-4-carboxylic acid	5337-03-1		0.66	-
3A1	tetrahydro-2H-pyran-4-ylmethanol	14774-37-9		-0.42	-
3A2	imidazo[1,2-a]pyridin-6-ylmethanol	132213-07-1		-0.55	-
3A3	(4-pyrid-4-ylphenyl)methanol	217192-22-8		-0.61	-
3A4	3-pyridin-3-ylaniline	57976-57-5		* -7.88	-
3A5	1-methyl-1H-indole-5-carboxylic acid	186129-25-9		0.52	-
3A6	3-(1H-imidazol-1-ylmethyl)aniline	120107-85-9		0.91	-

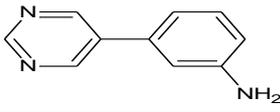
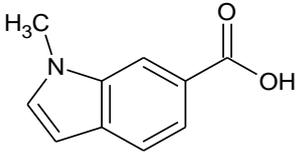
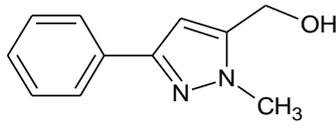
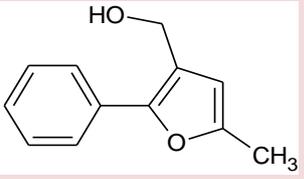
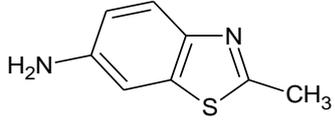
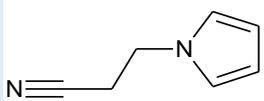
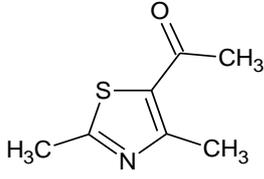
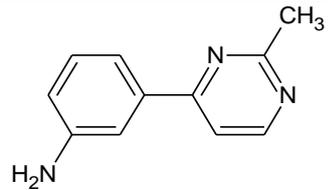
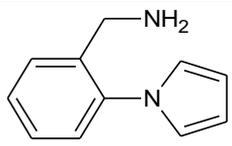
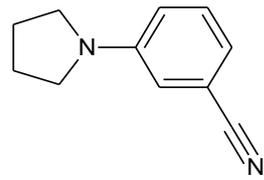
3A7	1H-benzimidazol-5-ylmethanol	106429-29-2		0.08	-
3A8	N-(2,3-dihydrothieno[3,4-b][1,4]dioxin-5-ylmethyl)-N-methylamine	859851-03-9		-1.25	W
3A9	(2-pyrrolidin-1-ylpyrid-4-yl)methanol	906352-65-6		-0.63	-
3A10	3-(pyrrolidin-1-ylmethyl)aniline	183365-31-3		-0.53	-
3A11	5-chloro-2-(methylthio)aniline	16423-54-4		-0.66	*
3B1	[4-(1-pyrrolidinyl)phenyl]methanol	676245-12-8		0.23	-
3B2	(3-morpholinophenyl)methanol	145127-38-4		0.14	-
3B3	(2-thien-2-ylphenyl)methanol	773872-97-2		-0.60	*

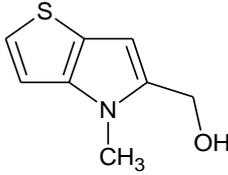
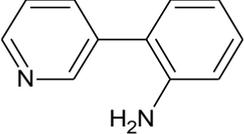
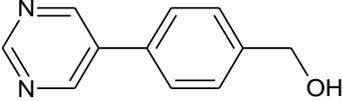
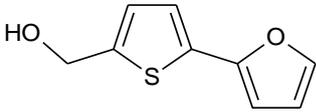
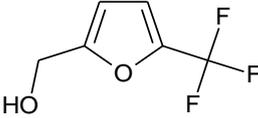
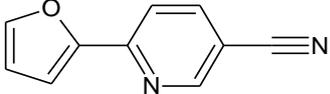
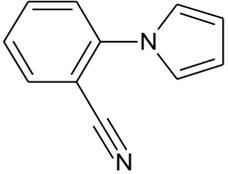
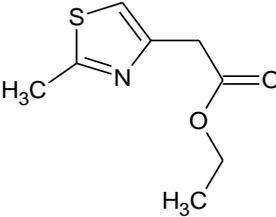
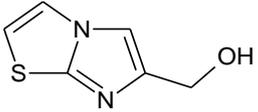
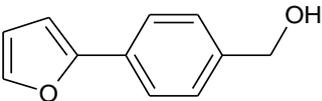
3B4	(4-pyrid-3-ylphenyl)methanol	217189-04-3		-0.15	-
3B5	1-methyl-1H-indol-5-amine	102308-97-4		* -5.17	-
3B6	[4-(1H-imidazol-1-ylmethyl)phenyl]methanol	103573-92-8		-0.19	-
3B7	3-(2-furyl)benzonitrile	112598-77-3		-1.83	-
3B8	3-thien-3-ylaniline	161886-96-0		-0.65	-
3B9	2-(1,3-benzodioxol-5-yl)ethanol	6006-82-2		-0.41	*
3B10	[4-(1H-pyrazol-1-ylmethyl)phenyl]methanol	160388-55-6		-0.78	-
3B11	1,3,5-trimethyl-1H-pyrazole	1072-91-9		0.04	N/A

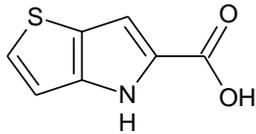
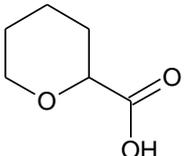
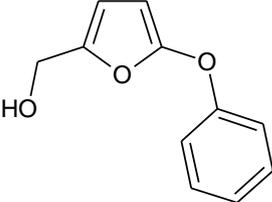
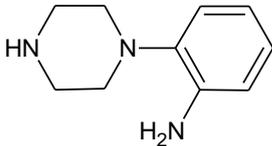
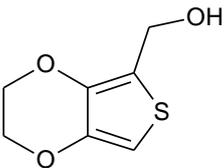
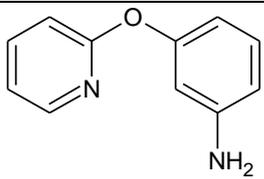
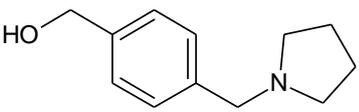
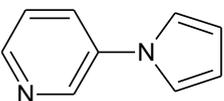
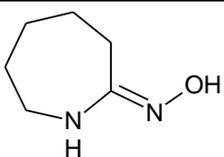
3C1	[4-(1-pyrrolidinyl)phenyl]methanamine	114365-04-7		-0.31	*
3C2	3-morpholin-4-ylaniline	159724-40-0		-0.24	-
3C3	(3-pyrid-4-ylphenyl)methanol	85553-55-5		-1.91	-
3C4	4-pyridin-3-ylaniline	82261-42-5		-0.23	-
3C5	(4-pyrid-2-ylphenyl)methanol	98061-39-3		-1.45	-
3C6	4-(1H-imidazol-1-ylmethyl)aniline	56643-85-7		-0.46	-
3C7	2-(2-furyl)benzoic acid	331942-47-3		-1.06	-
3C8	2-(morpholin-4-ylmethyl)aniline	95539-61-0		-0.38	-
3C9	4-[2-(dimethylamino)ethoxy]benzonitrile	24197-95-3		-1.23	-

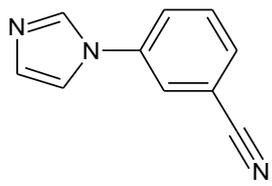
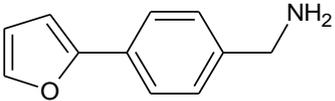
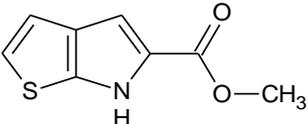
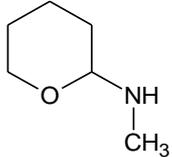
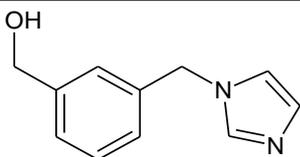
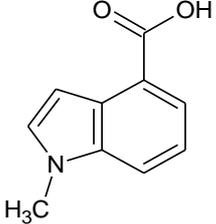
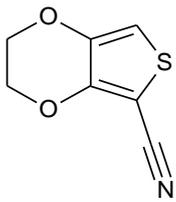
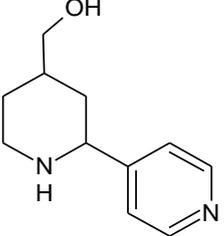
3C10	tetrahydrothiopyran-4-ylamine	21926-00-1		-1.64	w
3C11	2,6-di(dimethylamino)benzonitrile	20926-04-9		-0.31	-
3D1	4-methylthiophene-2-carboxamide	83933-16-8		-3.97	*
3D2	2-(1H-pyrrol-1-yl)benzoic acid	10333-68-3		-1.08	-
3D3	2-(4-methylpiperazin-1-yl)aniline	180605-36-1		* 2.08	-
3D4	benzo[b]furan-3-ylacetic acid	64175-51-5		-1.28	-
3D5	(3-pyrimidin-5-ylphenyl)methanol	852180-75-7		-0.24	-
3D6	(5-thien-2-yl-1H-pyrazol-3-yl)methanol	852228-02-5		-0.35	w

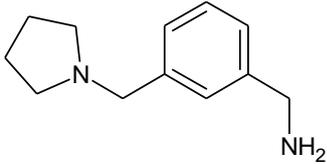
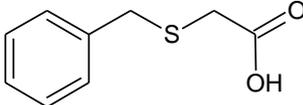
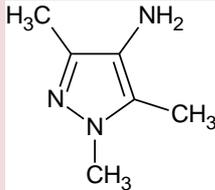
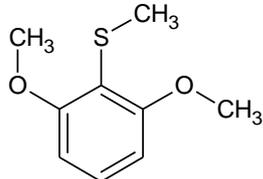
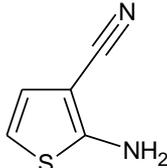
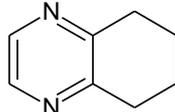
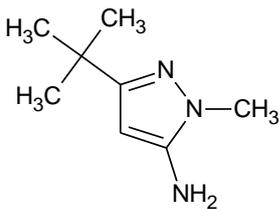
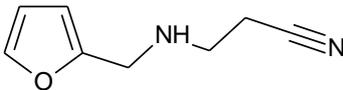
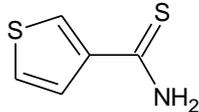
3D7	(1-methyl-5-phenyl-1H-pyrazol-3-yl)methanol	124344-98-5		-0.16	-
3D8	(4-thien-3-ylphenyl)methanol	160278-20-6		-0.64	-
3D9	2-(2-methyl-1,3-thiazol-4-yl)ethylamine	165115-15-1		-1.95	-
3D10	tetrahydro-2H-thiopyran-4-ol	29683-23-6		-0.44	-
3D11	3-methyl-1,2-dihydropyridin-2-one	1003-56-1		-0.27	*
3E1	imidazo[1,2-a]pyridin-2-ylmethanol	82090-52-6		-0.43	-
3E2	[2-(1H-pyrrol-1-yl)phenyl]methanol	61034-86-4		-2.75	*
3E3	3-piperidin-1-ylpyridine-2-carbonitrile	780802-33-7		-0.42	-
3E4	4-methyl-4H-thieno[3,2-b]pyrrole-5-carboxylic acid	841222-62-6		-0.11	-

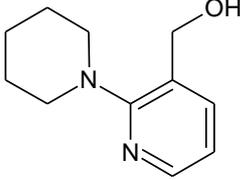
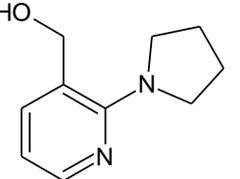
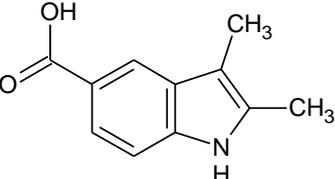
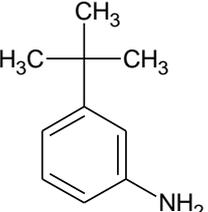
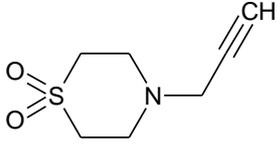
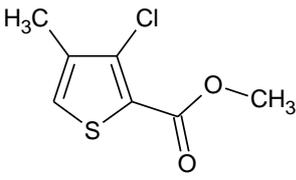
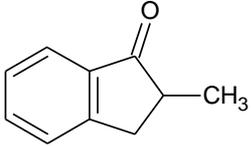
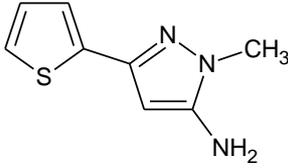
3E5	3-pyrimidin-5-ylaniline	69491-59-4		-0.1	N/A
3E6	1-methyl-1H-indole-6-carboxylic acid	202745-73-1		-0.44	-
3E7	(1-methyl-3-phenyl-1H-pyrazol-5-yl)methanol	864068-97-3		-0.63	*
3E8	(5-methyl-2-phenyl-3-furyl)methanol	183210-33-5		-3.59	-
3E9	2-methyl-1,3-benzothiazol-6-ylamine	2941-62-0		-0.42	-
3E10	3-(1H-pyrrol-1-yl)propanenitrile	43036-06-2		2.31	-
3E11	1-(2,4-dimethyl-1,3-thiazol-5-yl)ethan-1-one	38205-60-6		-0.48	-
3F1	3-(2-methylpyrimidin-4-yl)aniline	175201-90-8		-0.74	*
3F2	[2-(1H-pyrrol-1-yl)phenyl]methylamine	39243-88-4		-0.52	-
3F3	3-pyrrolidin-1-ylbenzonitrile	175696-73-8		-0.28	N/A

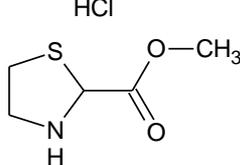
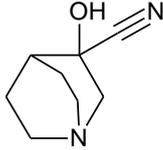
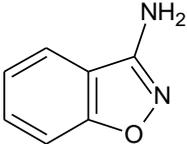
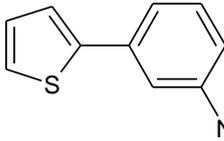
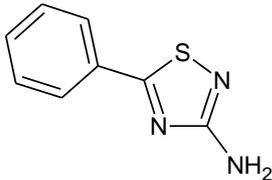
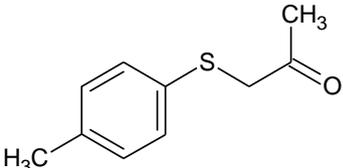
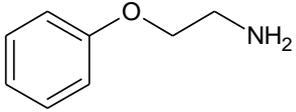
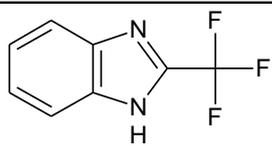
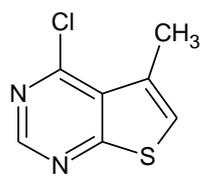
3F4	(4-methyl-4H-thieno[3,2-b]pyrrol-5-yl)methanol	121933-59-3		0.26	-
3F5	2-pyridin-3-ylaniline	177202-83-4		-0.66	*
3F6	(4-pyrimidin-5-ylphenyl)methanol	198084-13-8		0.33	-
3F7	[5-(2-furyl)thien-2-yl]methanol	868755-65-1		-1.00	-
3F8	[5-(trifluoromethyl)2-furyl]methanol	65865-28-3		-0.53	N/A
3F9	6-(2-furyl)nicotinonitrile	619334-28-0		0.30	-
3F10	2-(1H-pyrrol-1-yl)benzotrile	33265-71-3		-0.68	-
3F11	ethyl 2-(2-methyl-1,3-thiazol-4-yl)acetate	37128-24-8		-0.72	*
3G1	imidazo[2,1-b]thiazol-6-ylmethanol	349480-74-6		-0.57	-
3G2	[4-(2-furyl)phenyl]methanol	17920-85-3		* -4.5	-

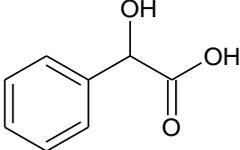
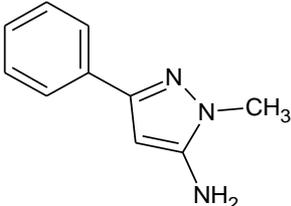
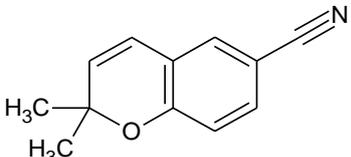
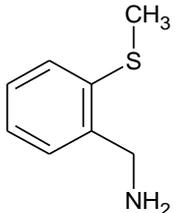
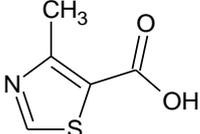
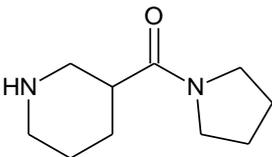
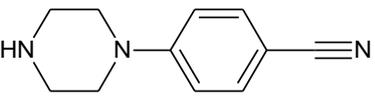
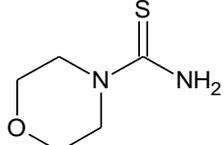
3G3	4H-thieno[3,2-b]pyrrole-5-carboxylic acid	39793-31-2		N/A	-
3G4	tetrahydropyran-2-carboxylic acid	51673-83-7		-0.05	-
3G5	(5-phenoxy-2-furyl)methanol	51551-74-7		-1.67	-
3G6	2-piperazin-1-ylaniline	13339-02-1		-0.32	-
3G7	2,3-dihydrothieno[3,4-b][1,4]dioxin-5-ylmethanol	859851-01-7		-0.81	*
3G8	3-(pyridin-2-yloxy)aniline	86556-09-4		-0.46	-
3G9	[4-(pyrrolidin-1-ylmethyl)phenyl]methanol	91271-60-2		-1.58	-
3G10	3-(1H-pyrrol-1-yl)pyridine	72692-99-0		-1.50	-
3G11	azepan-2-one oxime	19214-08-5		-0.58	-

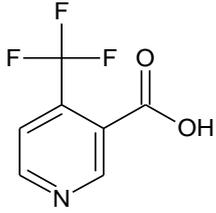
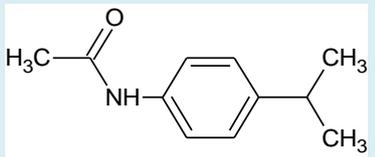
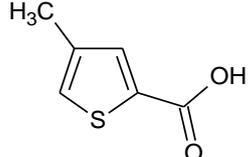
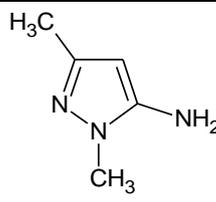
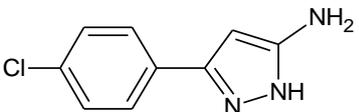
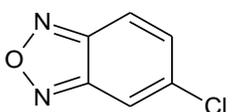
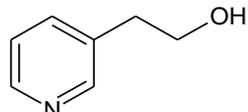
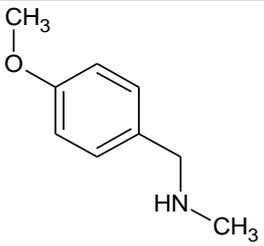
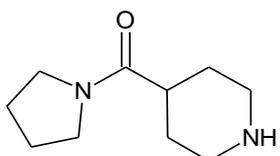
3H1	3-(1H-imidazol-1-yl)benzonitrile	25699-85-8		-0.30	-
3H2	[4-(2-furyl)phenyl]methylamine	771573-27-4		* -4.64	-
3H3	methyl 6H-thieno[2,3-b]pyrrole-5-carboxylate	118465-49-9		-0.40	-
3H4	tetrahydropyran-2-ylmethylamine	6628-83-7		-0.77	-
3H5	[3-(1H-imidazol-1-ylmethyl)phenyl]methanol	151055-79-7		-1.46	-
3H6	1-methyl-1H-indole-4-carboxylic acid	90924-06-4		-0.56	-
3H7	2,3-dihydrothieno[3,4-b][1,4]dioxine-5-carbonitrile	859851-02-8		-0.06	N/A
3H8	(2-piperidinopyrid-4-yl)methanol	888070-04-0		-0.63	-

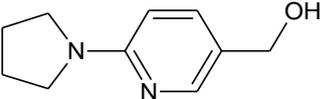
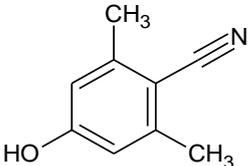
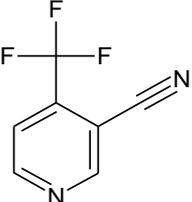
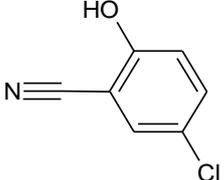
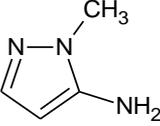
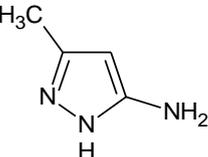
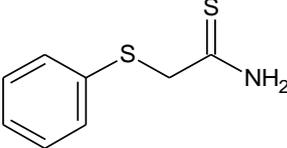
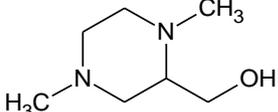
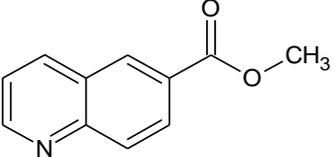
3H9	[3-(1-pyrrolidinylmethyl)phenyl]methanamine	91271-78-2		-1.92	-
3H10	2-(benzylthio)acetic acid	103-46-8		-1.03	w
3H11	1,3,5-trimethyl-1H-pyrazol-4-amine	28466-21-9		-2.57	-
4A1	1,3-dimethoxy-2-(methylthio)benzene	33617-67-3		0.43	*
4A2	2-aminothiophene-3-carbonitrile	4651-82-5		* 5.07	-
4A3	5,6,7,8-tetrahydroquinoxaline	34413-35-9		0.43	-
4A4	3-(tert-butyl)-1-methyl-1H-pyrazol-5-amine	118430-73-2		1.10	*
4A5	1,3-thiazolane-2-carboxylic acid	16310-13-7		* 11.17	N/A
4A6	3-[(2-furylmethyl)amino]propane nitrile	6788-68-7		1.89	-
4A7	thiophene-3-carbothioamide	24044-76-6		* -2.9	*

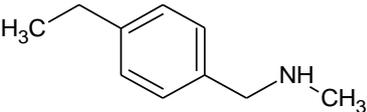
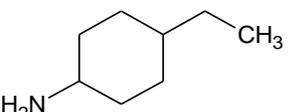
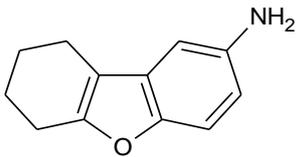
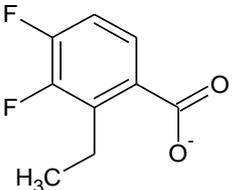
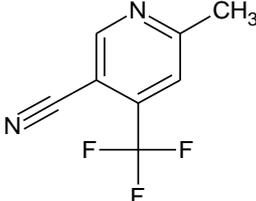
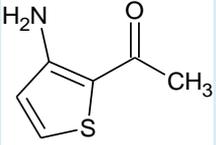
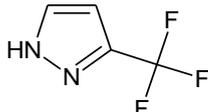
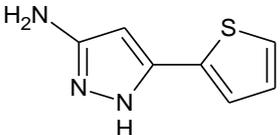
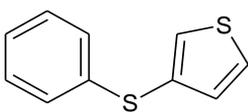
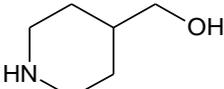
4A8	(2-piperidino-3-pyridinyl)methanol	690632-84-9		0.21	-
4A9	[2-(1-pyrrolidinyl)-3-pyridinyl]methanol	690632-85-0		* 7.14	-
4A10	2,3-dimethyl-1H-indole-5-carboxylic acid	14844-73-6		-1.92	-
4A11	3-(tert-butyl)aniline	5369-19-7		0.54	*
4B1	4-(Prop-2-ynyl)thiomorpholine 1,1-dioxide	10442-03-2		0.47	-
4B2	methyl 3-chloro-4-methylthiophene-2-carboxylate	175137-11-8		1.87	-
4B3	2-methylindan-1-one	17496-14-9		1.37	-
4B4	1-methyl-3-(2-thienyl)-1H-pyrazol-5-amine	118430-78-7		-0.63	-

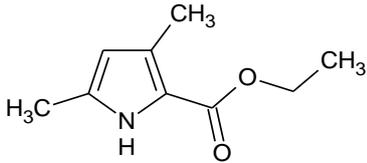
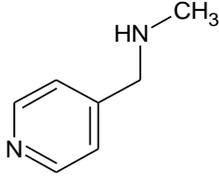
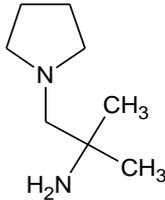
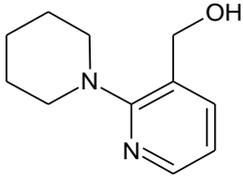
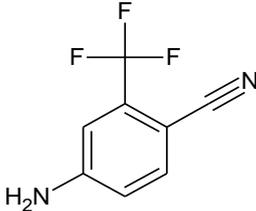
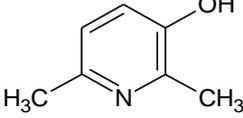
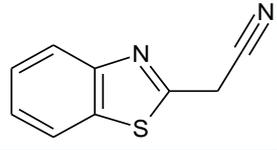
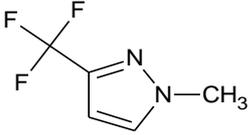
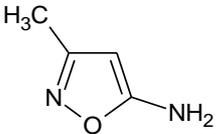
4B5	methyl 1,3-thiazolane-2-carboxylate hydrochloride	33305-08-7		1.95	-
4B6	3-hydroxyquinuclidine-3-carbonitrile	6238-30-8		-0.06	-
4B7	1,2-benzisoxazol-3-amine	36216-80-5		-0.12	*
4B8	3-(2-thienyl)aniline	92057-12-0		N/A	*
4B9	5-phenyl-1,2,4-thiadiazol-3-amine	27182-54-3		-0.20	-
4B10	1-[(4-methylphenyl)thio]acetone	1200-13-1		* 10.66	-
4B11	2-phenoxyethylamine	1758-46-9		0.07	-
4C1	2-(trifluoromethyl)-1H-benzo[d]imidazole	312-73-2		1.68	-
4C2	4-chloro-5-methylthieno[2,3-d]pyrimidine	43088-67-1		* -5.94	-

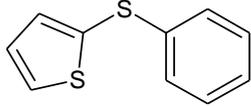
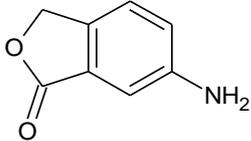
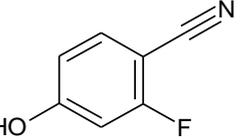
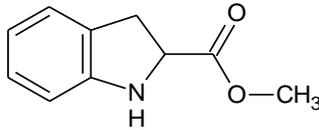
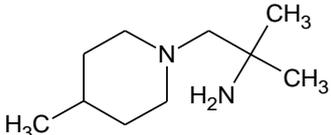
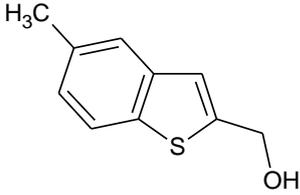
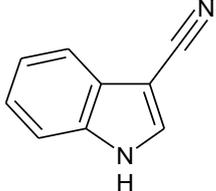
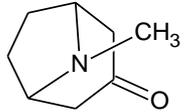
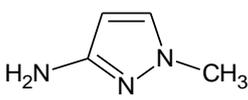
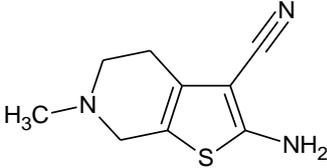
4C3	2-hydroxy-2-phenylacetic acid	90-64-2		1.63	-
4C4	1-methyl-3-phenyl-1H-pyrazol-5-amine	10199-50-5		-0.50	-
4C5	2,2-dimethyl-2H-chromene-6-carbonitrile	33143-29-2		-0.09	-
4C6	2-(methylthio)benzylamine	56004-83-2		-0.11	-
4C7	4-Methyl-thiazole-5-carboxylic acid	20485-41-0		0.28	-
4C8	3-piperidinyl(1-pyrrolidinyl)methanone	35090-94-9		-0.45	-
4C9	4-Piperazin-1-yl-benzonitrile	68104-63-2		-0.05	-
4C10	morpholine-4-carbothioamide	14294-10-1		0.77	-

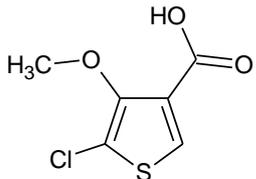
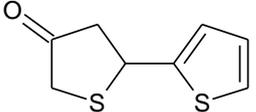
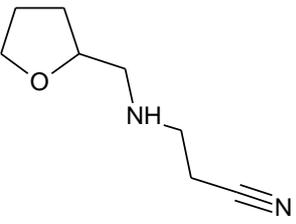
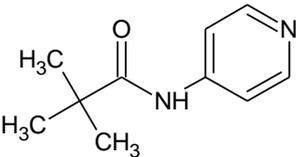
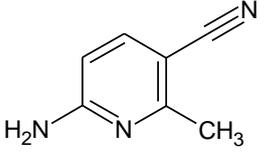
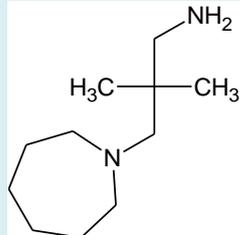
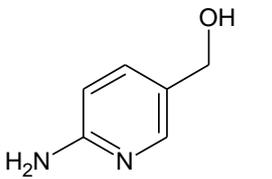
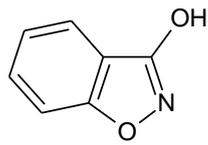
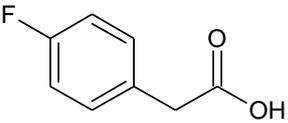
4C11	4-(trifluoromethyl)nicotinic acid	158063-66-2		0.21	-
4D1	N1-(4-isopropylphenyl)acetamide	5702-74-9		4.87	-
4D2	4-methylthiophene-2-carboxylic acid	14282-78-1		1.31	-
4D3	1,3-dimethyl-1H-pyrazol-5-amine	3524-32-1		0.23	-
4D4	3-(4-chlorophenyl)-1H-pyrazol-5-amine	78583-81-0		-0.46	-
4D5	5-chloro-2,1,3-benzoxadiazole	19155-86-3		N/A	-
4D6	2-(3-pyridyl)ethan-1-ol	6293-56-7		N/A	-
4D7	N-(4-methoxybenzyl)-N-methylamine	702-24-9		0.14	-
4D8	4-piperidiny(1-pyrrolidiny)methanone	35090-95-0		-0.40	-

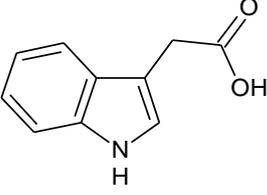
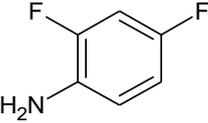
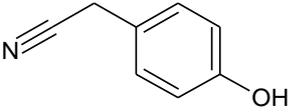
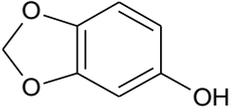
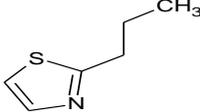
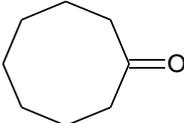
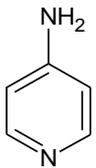
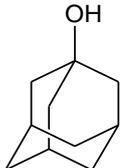
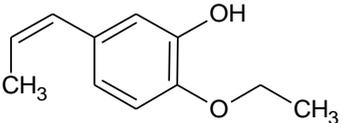
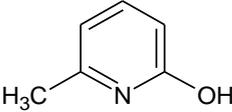
4D9	[6-(1-pyrrolidinyl)-3-pyridinyl]methanol	690632-01-0		-0.30	-
4D10	4-hydroxy-2,6-dimethylbenzonitrile	58537-99-8		0.11	-
4D11	4-(trifluoromethyl)nicotinonitrile	13600-43-6		0.17	-
4E1	5-chloro-2-hydroxybenzonitrile	13589-72-5		-1.02	*
4E2	1-methyl-1H-pyrazol-5-ylamine	1192-21-8		0.36	-
4E3	3-methyl-1H-pyrazol-5-amine	31230-17-8		0.17	-
4E4	2-(phenylthio)ethanethioamide	59865-82-6		3.35	-
4E5	(1,4-dimethyl-2-piperazinyl)methanol	14675-44-6		-0.05	-
4E6	methyl quinoline-6-carboxylate	38896-30-9		N/A	-

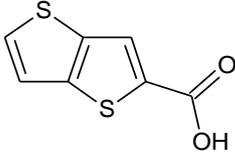
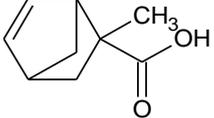
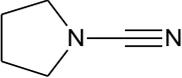
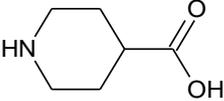
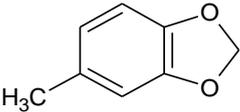
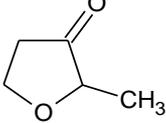
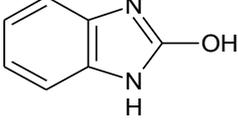
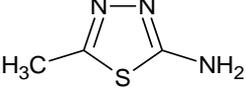
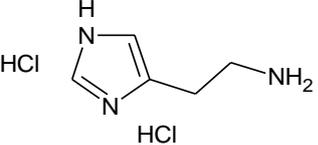
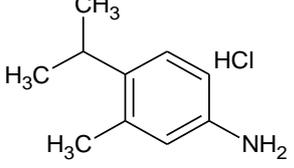
4E7	N-(4-ethylbenzyl)-N-methylamine	568577-84-4		-0.09	-
4E8	4-ethylcyclohexanamine	23775-39-5		0.25	-
4E9	6,7,8,9-tetrahydrodibenzo[b,d]furan-2-amine	38084-44-5		-0.30	-
4E10	ethyl-3,4-difluorobenzoate	144267-96-9		-0.1	N/A
4E11	6-methyl-4-(trifluoromethyl)nicotinonitrile	13600-49-2		0.3	N/A
4F1	1-(3-amino-2-thienyl)ethan-1-one	31968-33-9		2.93	*
4F2	3-(trifluoromethyl)-1H-pyrazole	20154-03-4		-0.12	-
4F3	5-thien-2-yl-1H-pyrazol-3-amine	96799-03-0		0.02	N/A
4F4	3-(phenylthio)thiophene	16718-11-9		0.65	N/A
4F5	4-piperidylmethanol	6457-49-4		0.33	-

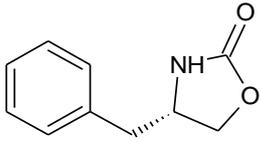
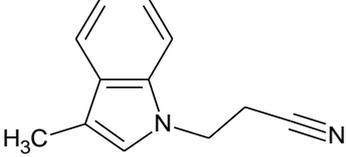
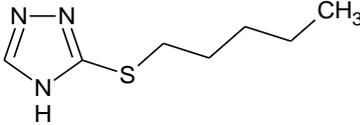
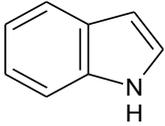
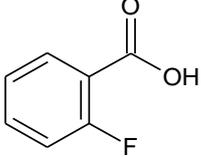
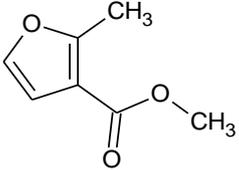
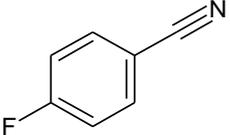
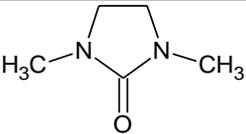
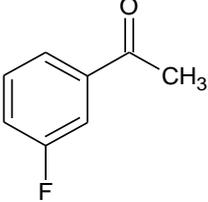
4F6	ethyl 3,5-dimethyl-1H-pyrrole-2-carboxylate	2199-44-2		-0.36	-
4F7	N-methyl-N-(4-pyridinylmethyl)amine	6971-44-4		0.47	-
4F8	2-methyl-1-(1-pyrrolidinyl)-2-propanamine	34155-39-0		-0.04	-
4F9	(Piperidino-3-pyridinyl)methanol	690631-99-3		-0.24	-
4F10	4-amino-2-(trifluoromethyl)benzonitrile	654-70-6		-1.34	-
4F11	2,6-dimethylpyridin-3-ol	1122-43-6		0.12	-
4G1	2-(1,3-benzothiazol-2-yl)acetonitrile	56278-50-3		8.93	-
4G2	1-methyl-3-(trifluoromethyl)-1H-pyrazole	154471-65-5		0.26	N/A
4G3	3-methylisoxazol-5-amine	14678-02-5		-0.42	-

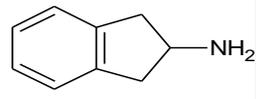
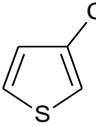
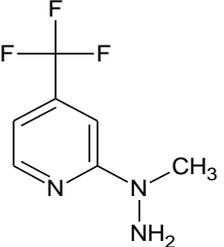
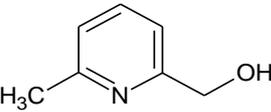
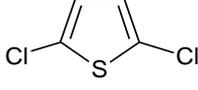
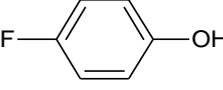
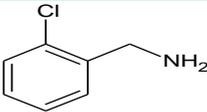
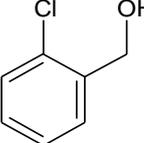
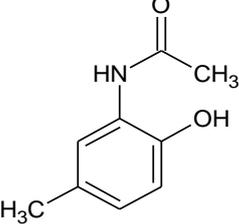
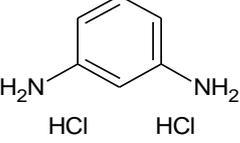
4G4	2-(phenylthio)thiophene	16718-12-0		-0.09	N/A
4G5	6-amino-1,3-dihydroisobenzofuran-1-one	57319-65-0		0.44	-
4G6	2-fluoro-4-hydroxybenzonitrile	82380-18-5		-0.11	-
4G7	methyl 2-indolinecarboxylate	96056-64-3		6.78	-
4G8	2-methyl-1-(4-methylpiperidino)-2-propanamine	690632-11-2		-0.19	N/A
4G9	5-Methylbenzo[b]thiophene-2-methanol	22962-49-8		N/A	-
4G10	1H-indole-3-carbonitrile	5457-28-3		-1.10	w
4G11	8-methyl-8-azabicyclo[3.2.1]octan-3-one	532-24-1		-0.29	-
4H1	1-methyl-1H-pyrazol-3-amine	1904-31-0		0.90	-
4H2	2-amino-6-methyl-4,5,6,7-tetrahydrothieno[2,3-c]pyridine-3-carbonitrile	37578-06-6		N/A	-

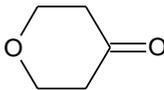
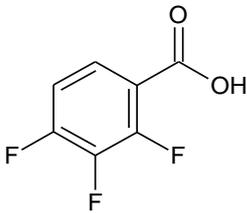
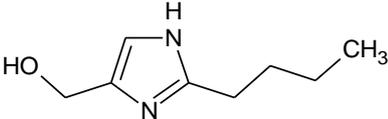
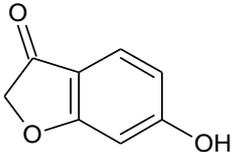
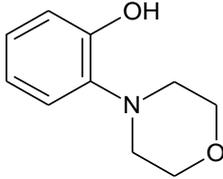
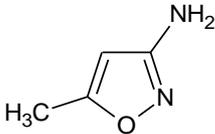
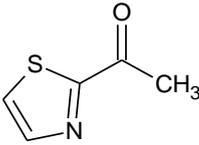
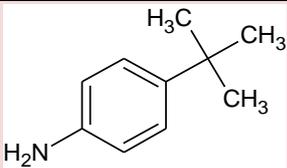
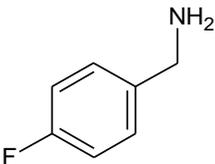
4H3	5-chloro-4-methoxythiophene-3-carboxylic acid	133659-14-0		* 5.34	-
4H4	5-(2-thienyl)tetrahydrothiophen-3-one	108372-48-1		* 5.71	-
4H5	3-[(tetrahydrofuran-2-yl)methyl]amino]propanenitrile	90322-18-2		0.19	-
4H6	2,2-dimethyl-N-(4-pyridinyl)propanamide	70298-89-4		0.01	-
4H7	6-amino-2-methylnicotinonitrile	183428-90-2		-1.07	-
4H8	3-(1-azepanyl)-2,2-dimethylpropylamine	845885-85-0		10.70	-
4H9	(6-amino-3-pyridinyl)methanol	113293-71-3		0.14	-
4H10	benzo[d]isoxazol-3-ol	21725-69-9		-0.38	-
4H11	2-(4-fluorophenyl)acetic acid	405-50-5		0.56	-

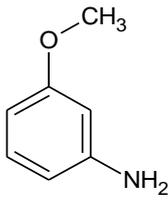
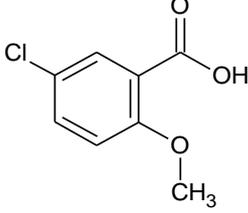
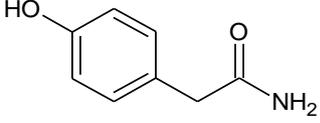
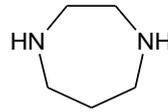
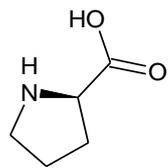
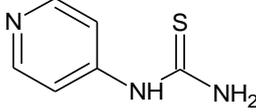
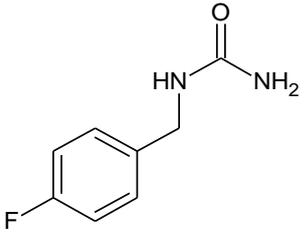
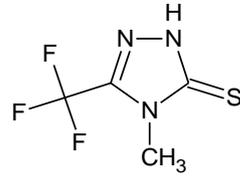
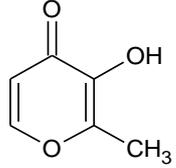
5A1	2-(1H-indol-3-yl)acetic acid	87-51-4		N/A	-
5A2	2,4-difluoroaniline	367-25-9		-0.64	N/A
5A3	2-(4-hydroxyphenyl)acetonitrile	14191-95-8		-1.14	-
5A4	1,3-benzodioxol-5-ol	533-31-3		* 8.51	w
5A5	2-propyl-1,3-thiazole	17626-75-4		-0.67	N/A
5A6	cyclooctan-1-one	502-49-8		-0.59	-
5A7	pyridin-4-amine	504-24-5		-1.09	-
5A8	adamantan-1-ol	768-95-6		-0.97	-
5A9	2-ethoxy-5-prop-1-enylphenol	94-86-0		* -6.62	-
5A10	6-methylpyridin-2-ol	3279-76-3		-1.19	-

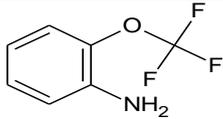
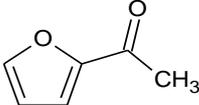
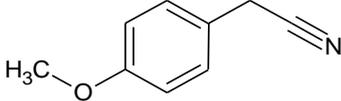
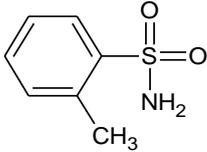
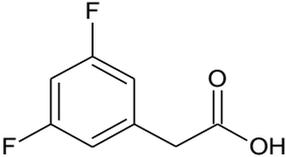
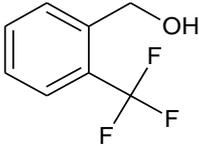
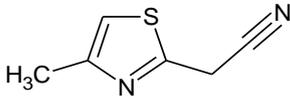
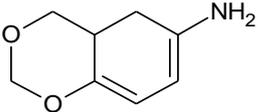
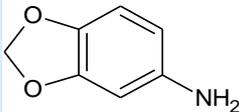
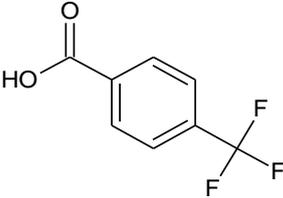
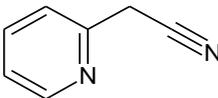
5A11	thieno[3,2-b]thiophene-2-carboxylic acid	1723-27-9		*-3.91	-
5B1	2-methylbicyclo[2.2.1]hept-5-ene-2-carboxylic acid	825-03-6		-1.17	-
5B2	pyrrolidine-1-carbonitrile	1530-88-7		-0.61	-
5B3	piperidine-4-carboxylic acid	498-94-2		-0.64	-
5B4	5-methyl-1,3-benzodioxazole	7145-99-5		N/A	-
5B5	2-methyltetrahydrofuran-3-one	3188-00-9		-0.89	-
5B6	1H-benzimidazol-2-ol	615-16-7		N/A	-
5B7	5-methyl-1,3,4-thiadiazol-2-amine	108-33-8		-0.72	-
5B8	2-(1H-imidazol-4-yl)ethylamine dihydrochloride	56-92-8		-0.39	-
5B9	4-isopropyl-3-methylaniline hydrochloride	4534-11-6		-0.95	*

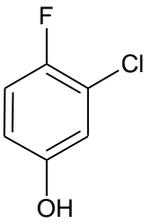
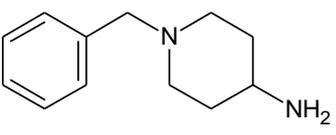
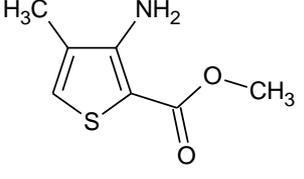
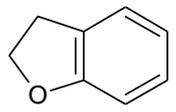
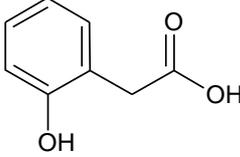
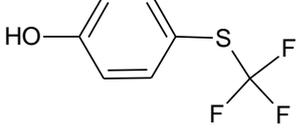
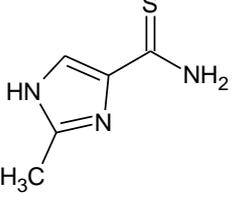
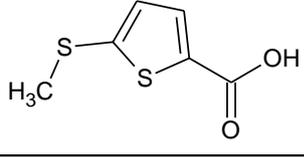
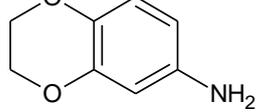
5B10	(S)-4-BENZYL-2-OXAZOLIDINONE	90719-32-7		N/A	-
5B11	3-(3-methyl-1H-indol-1-yl)propanenitrile	4414-81-7		-0.37	-
5C1	3-(pentylthio)-4H-1,2,4-triazole	71705-07-2		N/A	-
5C2	1H-indole	120-72-9		-1.25	*
5C3	2-fluorobenzoic acid	445-29-4		N/A	-
5C5	methyl 2-methyl-3-furoate	6141-58-8		N/A	N/A
5C6	4-fluorobenzonitrile	1194-02-1		N/A	-
5C7	1,3-dimethylimidazolidin-2-one	80-73-9		-0.96	w
5C8	1-(3-fluorophenyl)ethan-1-one	455-36-7		N/A	N/A

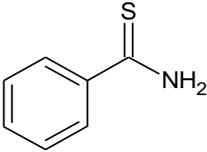
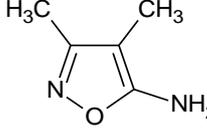
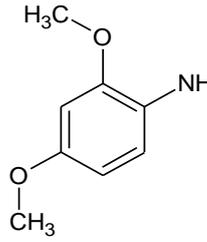
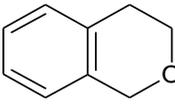
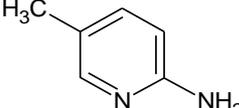
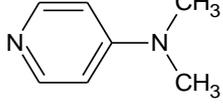
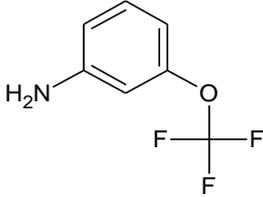
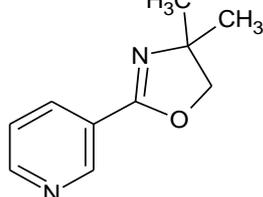
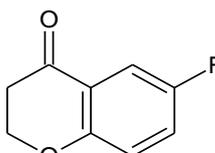
5C9	indan-2-amine	2975-41-9		-2.27	N/A
5C10	3-chlorothiophene	17249-80-8		-0.94	N/A
5C11	2-(1-methylhydrazino)-4-(trifluoromethyl)pyridine	208720-09-6		-0.97	-
5D1	(6-methylpyridin-2-yl)methanol	1122-71-0		0.04	-
5D2	2,5-dichlorothiophene	3172-52-9		-0.30	-
5D3	4-fluorophenol	371-41-5		-0.83	-
5D4	2-chlorobenzylamine	89-97-4		* 2.53	N/A
5D5	(2-chlorophenyl)methanol	17849-38-6		-1.07	-
5D6	N1-(2-hydroxy-5-methylphenyl)acetamide	6375-17-3		-1.06	-
5D7	benzene-1,3-diamine dihydrochloride	541-69-5		-0.72	-

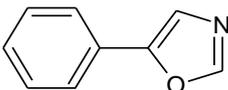
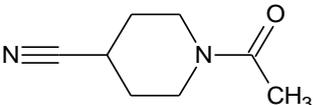
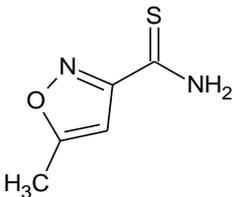
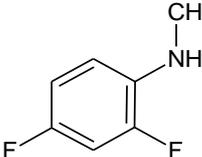
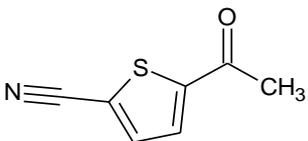
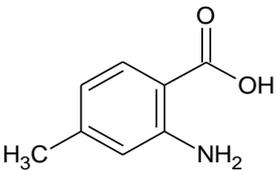
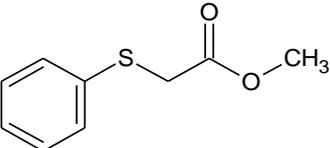
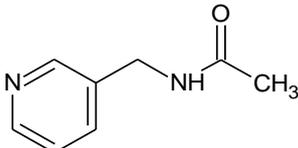
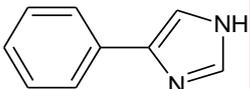
5D8	tetrahydro-4H-pyran-4-one	29943-42-8		-0.89	-
5D9	2,3,4-trifluorobenzoic acid	61079-72-9		-1.17	-
5D10	(2-butyl-1H-imidazol-4-yl)methanol	68283-19-2		-0.94	-
5D11	6-hydroxy-2,3-dihydrobenzo[b]furan-3-one	6272-26-0		-0.62	-
5E1	2-morpholinophenol	41536-44-1		-0.81	-
5E2	5-methylisoxazol-3-amine	1072-67-9		-0.65	-
5E3	1-(1,3-thiazol-2-yl)ethan-1-one	24295-03-2		-0.88	N/A
5E4	4-(tert-butyl)aniline	769-92-6		-2.96	-
5E5	4-fluorobenzylamine	140-75-0		-0.26	-

5E6	3-methoxyaniline	536-90-3		-1.45	-
5E7	5-chloro-2-methoxybenzoic acid	3438-16-2		-0.88	-
5E8	2-(4-hydroxyphenyl)acetamide	17194-82-0		N/A	-
5E9	1,4-diazepane	505-66-8		N/A	-
5E10	D-Proline	344-25-2		N/A	N/A
5E11	N-(4-pyridyl)thiourea	164670-44-4		-1.47	-
5F1	N-(4-fluorobenzyl)urea	76523-24-5		-0.14	-
5F2	4-methyl-3-(trifluoromethyl)-4,5-dihydro-1H-1,2,4-triazole-5-thione	30682-81-6		* 7.2	-
5F3	3-hydroxy-2-methyl-4H-pyran-4-one	118-71-8		-1.64	-

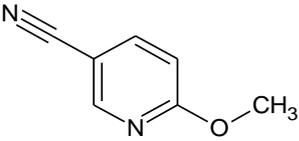
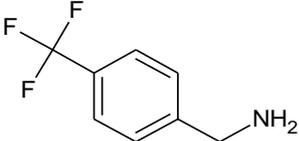
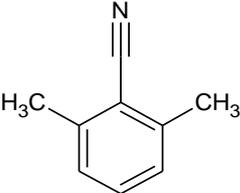
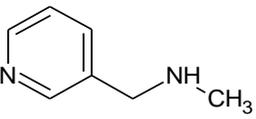
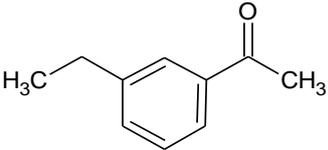
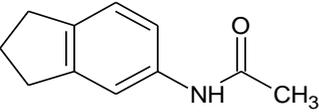
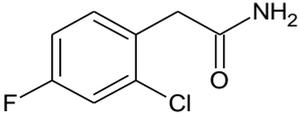
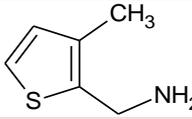
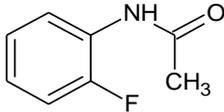
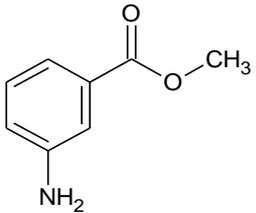
5F4	2-trifluoromethoxyaniline	1535-75-7		-0.74	N/A
5F5	1-(2-furyl)ethan-1-one	1192-62-7		-0.93	N/A
5F6	2-(4-methoxyphenyl)acetonitrile	104-47-2		-0.69	-
5F7	2-methylbenzene-1-sulfonamide	88-19-7		-0.58	-
5F8	2-(3,5-difluorophenyl)acetic acid	105184-38-1		-0.66	-
5F9	[2-(trifluoromethyl)phenyl]methanol	346-06-5		N/A	N/A
5F10	2-(4-methyl-1,3-thiazol-2-yl)acetonitrile	19785-39-8		N/A	-
5F11	4,5-dihydro-1,3-benzodioxine-6-amine	22791-64-6		* -4.09	-
5G1	1,3-benzodioxol-5-amine	14268-66-7		2.78	-
5G2	4-(trifluoromethyl)benzoic acid	455-24-3		-1.16	N/A
5G3	2-(2-pyridyl)acetonitrile	2739-97-1		-1.15	-

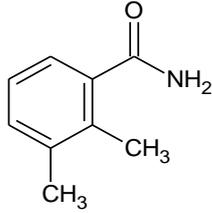
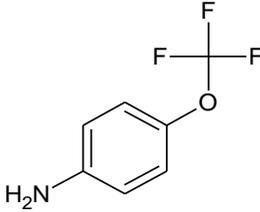
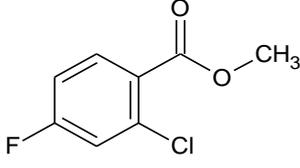
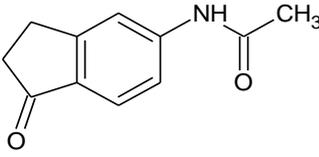
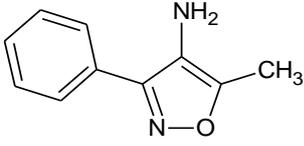
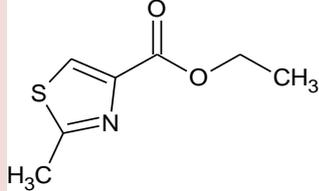
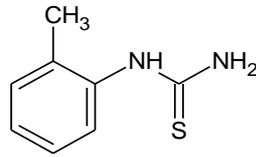
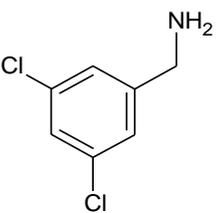
5G4	3-chloro-4-fluorophenol	2613-23-2		-0.96	-
5G5	1-benzylpiperidin-4-amine	50541-93-0		-0.47	-
5G6	methyl 3-amino-4-methylthiophene-2-carboxylate	85006-31-1		-1.90	-
5G7	2,3-dihydrobenzo[b]furan	496-16-2		-0.19	N/A
5G8	2-(2-hydroxyphenyl)acetic acid	614-75-5		-0.59	-
5G9	4-[(trifluoromethyl)thio]phenol	461-84-7		-1.06	-
5G10	2-methyl-1H-imidazole-4-carbothioamide	129486-91-5		-1.72	-
5G11	5-(methylthio)thiophene-2-carboxylic acid	20873-58-9		-0.32	-
5H1	2,3-dihydro-1,4-benzodioxin-6-amine	22013-33-8		* -3.35	-

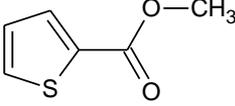
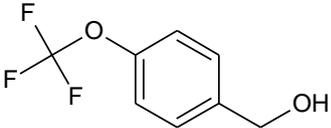
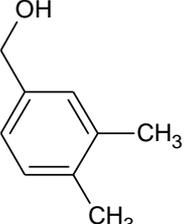
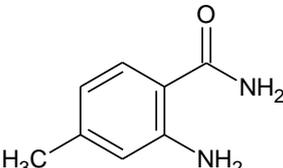
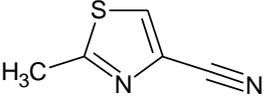
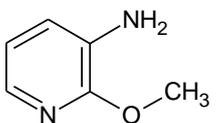
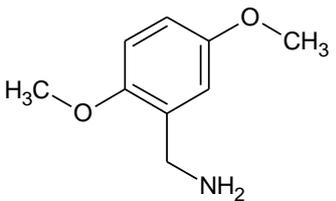
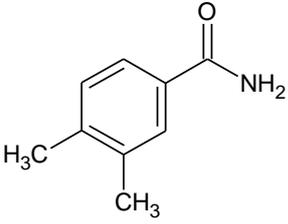
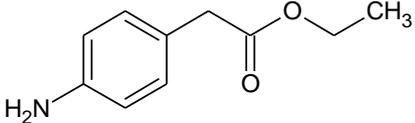
5H2	benzene-1-carbothioamide	2227-79-4		* -3.75	-
5H3	3,4-dimethylisoxazol-5-amine	19947-75-2		-0.55	-
5H4	2,4-dimethoxyaniline	2735-04-8		*-3.17	-
5H5	3,4-dihydro-1H-benzo[c]pyran	493-05-0		-0.31	-
5H6	5-methylpyridin-2-amine	1603-41-4		-0.22	-
5H7	N4,N4-dimethylpyridin-4-amine	1122-58-3		-0.62	-
5H8	3-(trifluoromethoxy)aniline	1535-75-5		-0.13	N/A
5H9	3-(4,4-dimethyl-4,5-dihydro-1,3-oxazol-2-yl)pyridine	68981-86-2		-1.11	-
5H10	6-fluorochroman-4-one	66892-34-0		-0.86	-

5H11	5-phenyl-1,3-oxazole	1006-68-4		0.00	-
6A1	1-acetylpiperidine-4-carbonitrile	25503-91-7		0.66	-
6A2	5-methylisoxazole-3-carbothioamide	77358-26-0		-0.36	-
6A3	N1-methyl-2,4-difluoroaniline	138564-16-6		-0.14	-
6A4	4-acetylbenzonitrile	1443-80-7		0.06	-
6A5	5-acetylthiophene-2-carbonitrile	88653-55-8		-1.04	-
6A6	2-amino-4-methylbenzoic acid	2305-36-4		0.87	-
6A7	methyl 2-(phenylthio)acetate	17277-58-6		-0.12	-
6A8	N1-(3-pyridylmethyl)acetamide	22977-34-0		-0.31	-
6B1	4-phenyl-1H-imidazole	670-95-1		-2.42	-

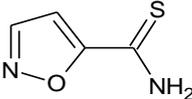
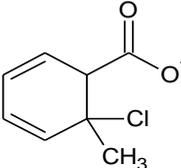
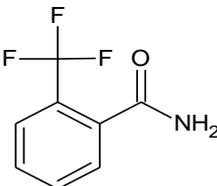
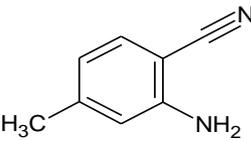
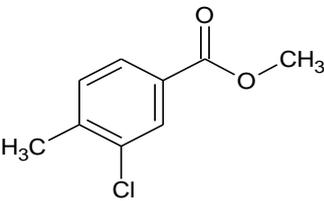
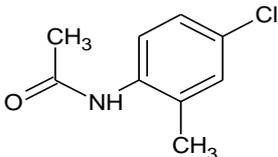
6B2	6-methyl-2-oxo-1,2-dihydropyridine-3-carbonitrile	4241-27-4		-0.30	-
6B3	2,3-dimethylbenzylamine	51586-20-0		-0.59	-
6B4	3-morpholinophenol	27292-49-5		0.74	-
6B5	2-(trifluoromethoxy)benzonitrile	63968-85-4		0.28	N/A
6B6	3,5-dimethylisoxazole	300-87-8		-0.22	-
6B7	methyl 4-amino-3-methylbenzoate	18595-14-7		0.84	-
6B8	4-chloro-2-methylbenzoic acid	7499-07-2		-0.14	-
6C1	methyl 2,5-dimethyl-1H-pyrrole-3-carboxylate	69687-80-5		-1.56	*

6C2	6-methoxynicotinonitrile	15871-85-9		-0.29	-
6C3	4-(trifluoromethyl)benzylamine	3300-51-4		-0.70	-
6C4	2,6-dimethylbenzonitrile	6575-13-9		-0.50	-
6C5	N-methyl-N-(3-pyridylmethyl)amine	20173-04-0		-0.02	-
6C6	1-(3-ethylphenyl)ethan-1-one	22699-70-3		0.39	N/A
6C7	N1-(2,3-dihydro-1H-inden-5-yl)acetamide	59856-06-3		-0.03	-
6C8	2-(2-chloro-4-fluorophenyl)acetamide	306937-35-9		-0.09	-
6D1	(3-methyl-2-thienyl)methylamine	104163-35-1		-2.39	-
6D2	N1-(2-fluorophenyl)acetamide	399-31-5		-0.45	-
6D3	methyl 3-aminobenzoate	4518-10-9		-0.35	-

6D4	2,3-dimethylbenzamide	5580-34-7		-0.28	-
6D5	4-(trifluoromethoxy)aniline	461-82-5		-0.28	N/A
6D6	methyl 2-chloro-4-fluorobenzoate	85953-29-3		-0.09	-
6D7	N1-(1-oxo-2,3-dihydro-1H-inden-5-yl)acetamide	58161-35-6		-0.47	-
6D8	5-methyl-3-phenyl-4-isoxazolamine	21169-65-3		1.18	-
6E1	ethyl 2-methyl-1,3-thiazole-4-carboxylate	6436-59-5		-3.53	-
6E2	N-(2-methylphenyl)thiourea	614-78-8		-0.34	-
6E3	3,5-dichlorobenzylamine	39989-43-0		-1.01	N/A

6E4	methyl thiophene-2-carboxylate	5380-42-7		-0.16	N/A
6E5	[4-(trifluoromethoxy)phenyl]methanol	1736-74-9		1.22	-
6E6	(3,4-dimethylphenyl)methanol	05/10/6966		0.09	N/A
6E7	2-amino-4-methylbenzamide	39549-79-6		0.62	N/A
6F1	2-methyl-1,3-thiazole-4-carbonitrile	21917-76-0		-0.98	-
6F2	2-methoxypyridin-3-amine	20265-38-7		-0.21	-
6F3	2,5-dimethoxybenzylamine	3275-95-4		0.17	-
6F4	3,4-dimethylbenzamide	5580-33-6		-0.14	-
6F5	ethyl 2-(4-aminophenyl)acetate	5438-70-0		-1.43	-

6F6	3-chloro-4-methylbenzoic acid	5162-82-3		-0.37	-
6F7	indolin-2-one	59-48-3		0.43	-
6G1	methyl-2-fluorobenzoate	394-35-4		-0.49	N/A
6G2	4-(4-methylpiperazino)aniline	16153-81-4		N/A	-
6G3	methyl 3-methylbenzoate	99-36-5		-0.04	-
6G4	1-(3-chloro-4-fluorophenyl)ethan-1-one	2923-66-2		-0.41	-
6G5	2-(2-thienyl)ethylamine	30433-91-1		-0.49	-
6G6	methyl-2-methylbenzoate	89-71-4		-0.12	N/A
6G7	N-methyl-N-phenylthiourea	4104-75-0		-1.35	-

6H1	isoxazole-5-carbothioamide	175334-72-2		-1.46	w
6H2	5-fluoro-2-(1H-imidazol-1-yl)aniline	251649-52-2		-0.67	-
6H3	methyl-2-chlorobenzoate	610-96-8		-0.07	-
6H4	2-(trifluoromethyl)benzamide	360-64-5		-0.07	-
6H5	2-amino-4-methylbenzonitrile	26830-96-6		-0.78	-
6H6	methyl 3-chloro-4-methylbenzoate	56525-63-4		*7.39	-
6H7	4-chloro-2-methylacetanilide	5202-86-8		-0.1	N/A
6H8				* 5.24	N/A

**Table A.1 Details of the Maybride rule of 3 fragment library screened for binding to HPV11 E2 TAD.**

The  $\Delta T_m$  of HPV11 E2 TAD with each fragment compared to a DMSO control is shown. Fragments resulting in an increase in protein  $T_m > 2$  °C are highlighted in blue; fragments resulting in a decrease in protein  $T_m > 2$  °C are highlighted in blue. Fragments resulting in altered fluorescence profiles are highlighted as \* $T_m$ . N/A indicates no  $T_m$  was calculated. Results of screening fragments for binding to HPV11 E2 TAD using NMR assays are included. Fragments not binding are shown as -; fragments binding in both STD and wLOGSY experiments as \*; fragments binding in wLOGSY alone as w. N/A indicates fragments which were not included in the NMR screens.

## Appendix 3. Details of NMR buffers screened with HPV11 E2 TAD

Gene ID	1	2	3	4	5	6	7	8	9	10	11	12
<b>A</b>	Na Phosphate pH 7.0	Na Phosphate pH 7.0	Na Phosphate pH 7.0 100 mM NaCl	Tris pH 8.0, 10 mM	Tris pH 8.0, 20 mM	Tris pH 8.0, 20 mM 100 mM NaCl	MOPS pH 7.0	MOPS pH 7.0	MOPS pH 7.0 100 mM NaCl	HEPES pH 7.5	HEPES pH 7.5	HEPES pH 7.5 100 mM NaCl
<b>B</b>	Na Phosphate pH 7.0 10 mM MgCl2	Na Phosphate pH 7.0 20 mM sucrose	Na Phosphate pH 7.0 20 mM sucrose 100 mM NaCl	Tris pH 8.0, 10 mM MgCl2	Tris pH 8.0, 50 mM sucrose	Tris pH 8.0, 20 mM sucrose 100 mM NaCl	MOPS pH 7.0 10 mM MgCl2	MOPS pH 7.0 20 mM sucrose	MOPS pH 7.0 20 mM sucrose 100 mM NaCl	HEPES pH 7.5 10 mM MgCl2	HEPES pH 7.5 20 mM sucrose	HEPES pH 7.5 20 mM sucrose 100 mM NaCl
<b>C</b>	Na Phosphate pH 7.0 20 mM MgCl2	Na Phosphate pH 7.0 50 mM sucrose	Na Phosphate pH 7.0 50 mM sucrose 100 mM NaCl	Tris pH 8.0, 20 mM MgCl2	Tris pH 8.0, 50 mM sucrose	Tris pH 8.0, 50 mM sucrose 100 mM NaCl	MOPS pH 7.0 20 mM MgCl2	MOPS pH 7.0 50 mM sucrose	MOPS pH 7.0 50 mM sucrose 100 mM NaCl	HEPES pH 7.5 20 mM MgCl2	HEPES pH 7.5 50 mM sucrose	HEPES pH 7.5 50 mM sucrose 100 mM NaCl
<b>D</b>	Na Phosphate pH 7.0 50 mM MgCl2	Na Phosphate pH 7.0 100 mM sucrose	Na Phosphate pH 7.0 100 mM sucrose 100 mM NaCl	Tris pH 8.0, 50 mM MgCl2	Tris pH 8.0, 150 mM sucrose	Tris pH 8.0, 100 mM sucrose 100 mM NaCl	MOPS pH 7.0 50 mM MgCl2	MOPS pH 7.0 100 mM sucrose	MOPS pH 7.0 100 mM sucrose 100 mM NaCl	HEPES pH 7.5 50 mM MgCl2	HEPES pH 7.5 100 mM sucrose	HEPES pH 7.5 100 mM sucrose 100 mM NaCl
<b>E</b>	Na Phosphate pH 7.0 100 mM MgCl2	Na Phosphate pH 7.0 150 mM sucrose	Na Phosphate pH 7.0 150 mM sucrose 100 mM NaCl	Tris pH 8.0, 100 mM MgCl2	Tris pH 8.0, 200 mM sucrose	Tris pH 8.0, 150 mM sucrose 100 mM NaCl	MOPS pH 7.0 100 mM MgCl2	MOPS pH 7.0 150 mM sucrose	MOPS pH 7.0 150 mM sucrose 100 mM NaCl	HEPES pH 7.5 100 mM MgCl2	HEPES pH 7.5 150 mM sucrose	HEPES pH 7.5 150 mM sucrose 100 mM NaCl
<b>F</b>	Na Phosphate pH 7.0 150 mM MgCl2	Na Phosphate pH 7.0 200 mM sucrose	Na Phosphate pH 7.0 200 mM sucrose 100 mM NaCl	Tris pH 8.0, 150 mM MgCl2	Tris pH 8.0, 200 mM sucrose	Tris pH 8.0, 200 mM sucrose 100 mM NaCl	MOPS pH 7.0 150 mM MgCl2	MOPS pH 7.0 200 mM sucrose	MOPS pH 7.0 200 mM sucrose 100 mM NaCl	HEPES pH 7.5 150 mM MgCl2	HEPES pH 7.5 200 mM sucrose	HEPES pH 7.5 200 mM sucrose 100 mM NaCl
<b>G</b>	Na Phosphate pH 7.0 200 mM MgCl2	Na Phosphate pH 7.0 250 mM sucrose	Na Phosphate pH 7.0 250 mM sucrose 100 mM NaCl	Tris pH 8.0, 200 mM MgCl2	Tris pH 8.0, 250 mM sucrose	Tris pH 8.0, 250 mM sucrose 100 mM NaCl	MOPS pH 7.0 200 mM MgCl2	MOPS pH 7.0 250 mM sucrose	MOPS pH 7.0 250 mM sucrose 100 mM NaCl	HEPES pH 7.5 200 mM MgCl2	HEPES pH 7.5 250 mM sucrose	HEPES pH 7.5 250 mM sucrose 100 mM NaCl
<b>H</b>	Na Phosphate pH 7.0 250 mM MgCl2	Na Phosphate pH 7.0 500 mM sucrose	Na Phosphate pH 7.0 500 mM sucrose 100 mM NaCl	Tris pH 8.0, 250 mM MgCl2	Tris pH 8.0, 500 mM sucrose	Tris pH 8.0, 500 mM sucrose 100 mM NaCl	MOPS pH 7.0 250 mM MgCl2	MOPS pH 7.0 500 mM sucrose	MOPS pH 7.0 500 mM sucrose 100 mM NaCl	HEPES pH 7.5 250 mM MgCl2	HEPES pH 7.5 500 mM sucrose	HEPES pH 7.5 500 mM sucrose 100 mM NaCl

**Table A.2 NMR buffer screen 2 buffer compositions**

Details of the NMR-buffer screen 2 buffer compositions used in thermal shift assays to determine the stability of HPV11 E2 TAD using DSF assays.

	1	2	3	4	5	6	7	8	9	10	11	12
<b>A</b>	41.25	41.25	42.23	40.33	41.23	40.32	40.28	40.28	41.22	40.28	40.28	42.22
<b>B</b>	42.23	42.23	43.23	41.22	42.25	40.23	41.22	40.23	42.23	41.22	40.22	42.23
<b>C</b>	42.23	42.23	43.23	42.23	41.22	40.23	41.22	40.23	43.23	41.22	40.23	42.23
<b>D</b>	43.2	42.22	43.2	43.2	42.2	40.23	44.22	40.23	42.22	43.2	40.23	42.22
<b>E</b>	43.22	42.22	43.22	44.23	42.22	41.27	44.22	40.3	42.22	44.2	40.3	42.22
<b>F</b>	44.22	42.43	44.22	44.22	42.22	41.27	45.25	40.28	42.22	45.25	40.28	43.22
<b>G</b>	44.22	42.23	44.22	45.3	42.23	41.27	45.28	40.32	42.23	45.25	40.32	43.3
<b>H</b>	43.23	42.27	45.22	44.2	43.25	42.27	46.23	41.23	44.2	45.23	42.25	43.23

**Table A.3 NMR buffer screen 2 DSF results**

Gene ID	1	2	3	4	5	6	7	8	9	10	11	12
<b>A</b>	Na Phosphate pH 7.0	Na Phosphate pH 7.0 100 mM NaCl	Na Phosphate pH 7.0 100 mM MgCl2	Tris pH 8.0,	Tris pH 8.0, 100 mM NaCl	Tris pH 8.0, 100 mM MgCl2	MOPS pH 7.0	MOPS pH 7.0 100 mM NaCl	MOPS pH 7.0 100 mM MgCl2	HEPES pH 7.5	HEPES pH 7.5 100 mM NaCl	HEPES pH 7.5 100 mM MgCl2
<b>B</b>	Na Phosphate pH 7.0 5 mM Arg/Glu	Na Phosphate pH 7.0 5 mM Arg/Glu 100 mM NaCl	Na Phosphate pH 7.0 5 mM Arg/Glu 100 mM MgCl2	Tris pH 8.0, 5 mM Arg/Glu	Tris pH 8.0, 5 mM Arg/Glu 100 mM NaCl	Tris pH 8.0, 5 mM Arg/Glu 100 mM MgCl2	MOPS pH 7.0 5 mM Arg/Glu	MOPS pH 7.0 5 mM Arg/Glu 100 mM NaCl	MOPS pH 7.0 5 mM Arg/Glu 100 mM MgCl2	HEPES pH 7.5 5 mM Arg/Glu	HEPES pH 7.5 5 mM Arg/Glu 100 mM NaCl	HEPES pH 7.5 5 mM Arg/Glu 100 mM MgCl2
<b>C</b>	Na Phosphate pH 7.0 10 mM Arg/Glu	Na Phosphate pH 7.0 10 mM Arg/Glu 100 mM NaCl	Na Phosphate pH 7.0 10 mM Arg/Glu 100 mM MgCl2	Tris pH 8.0, 10 mM Arg/Glu	Tris pH 8.0, 10 mM Arg/Glu 100 mM NaCl	Tris pH 8.0, 10 mM Arg/Glu 100 mM MgCl2	MOPS pH 7.0 10 mM Arg/Glu	MOPS pH 7.0 10 mM Arg/Glu 100 mM NaCl	MOPS pH 7.0 10 mM Arg/Glu 100 mM MgCl2	HEPES pH 7.5 10 mM Arg/Glu	HEPES pH 7.5 10 mM Arg/Glu 100 mM NaCl	HEPES pH 7.5 10 mM Arg/Glu 100 mM MgCl2
<b>D</b>	Na Phosphate pH 7.0 20 mM Arg/Glu	Na Phosphate pH 7.0 20 mM Arg/Glu 100 mM NaCl	Na Phosphate pH 7.0 20 mM Arg/Glu 100 mM MgCl2	Tris pH 8.0, 20 mM Arg/Glu	Tris pH 8.0, 20 mM Arg/Glu 100 mM NaCl	Tris pH 8.0, 20 mM Arg/Glu 100 mM MgCl2	MOPS pH 7.0 20 mM Arg/Glu	MOPS pH 7.0 20 mM Arg/Glu 100 mM NaCl	MOPS pH 7.0 20 mM Arg/Glu 100 mM MgCl2	HEPES pH 7.5 20 mM Arg/Glu	HEPES pH 7.5 20 mM Arg/Glu 100 mM NaCl	HEPES pH 7.5 20 mM Arg/Glu 100 mM MgCl2
<b>E</b>	Na Phosphate pH 7.0 50 mM Arg/Glu	Na Phosphate pH 7.0 50 mM Arg/Glu 100 mM NaCl	Na Phosphate pH 7.0 50 mM Arg/Glu 100 mM MgCl2	Tris pH 8.0, 50 mM Arg/Glu	Tris pH 8.0, 50 mM Arg/Glu 100 mM NaCl	Tris pH 8.0, 50 mM Arg/Glu 100 mM MgCl2	MOPS pH 7.0 50 mM Arg/Glu	MOPS pH 7.0 50 mM Arg/Glu 100 mM NaCl	MOPS pH 7.0 50 mM Arg/Glu 100 mM MgCl2	HEPES pH 7.5 50 mM Arg/Glu	HEPES pH 7.5 50 mM Arg/Glu 100 mM NaCl	HEPES pH 7.5 50 mM Arg/Glu 100 mM MgCl2
<b>F</b>	Na Phosphate pH 7.0 100 mM Arg/Glu	Na Phosphate pH 7.0 100 mM Arg/Glu 100 mM NaCl	Na Phosphate pH 7.0 100 mM Arg/Glu 100 mM MgCl2	Tris pH 8.0, 100 mM Arg/Glu	Tris pH 8.0, 100 mM Arg/Glu 100 mM NaCl	Tris pH 8.0, 100 mM Arg/Glu 100 mM MgCl2	MOPS pH 7.0 100 mM Arg/Glu	MOPS pH 7.0 100 mM Arg/Glu 100 mM NaCl	MOPS pH 7.0 100 mM Arg/Glu 100 mM MgCl2	HEPES pH 7.5 100 mM Arg/Glu	HEPES pH 7.5 100 mM Arg/Glu 100 mM NaCl	HEPES pH 7.5 100 mM Arg/Glu 100 mM MgCl2
<b>G</b>	Na Phosphate pH 7.0 150 mM Arg/Glu	Na Phosphate pH 7.0 150 mM Arg/Glu 100 mM NaCl	Na Phosphate pH 7.0 150 mM Arg/Glu 100 mM MgCl2	Tris pH 8.0, 150 mM Arg/Glu	Tris pH 8.0, 150 mM Arg/Glu 100 mM NaCl	Tris pH 8.0, 150 mM Arg/Glu 100 mM MgCl2	MOPS pH 7.0 150 mM Arg/Glu	MOPS pH 7.0 150 mM Arg/Glu 100 mM NaCl	MOPS pH 7.0 150 mM Arg/Glu 100 mM MgCl2	HEPES pH 7.5 150 mM Arg/Glu	HEPES pH 7.5 150 mM Arg/Glu 100 mM NaCl	HEPES pH 7.5 150 mM Arg/Glu 100 mM MgCl2
<b>H</b>	Na Phosphate pH 7.0 200 mM Arg/Glu	Na Phosphate pH 7.0 200 mM Arg/Glu 100 mM NaCl	Na Phosphate pH 7.0 200 mM Arg/Glu 100 mM MgCl2	Tris pH 8.0, 200 mM Arg/Glu	Tris pH 8.0, 200 mM Arg/Glu 100 mM NaCl	Tris pH 8.0, 200 mM Arg/Glu 100 mM MgCl2	MOPS pH 7.0 200 mM Arg/Glu	MOPS pH 7.0 200 mM Arg/Glu 100 mM NaCl	MOPS pH 7.0 200 mM Arg/Glu 100 mM MgCl2	HEPES pH 7.5 200 mM Arg/Glu	HEPES pH 7.5 200 mM Arg/Glu 100 mM NaCl	HEPES pH 7.5 200 mM Arg/Glu 100 mM MgCl2

**Table A.4 NMR buffer screen 3 buffer compositions**

Details of the NMR-buffer screen 2 buffer compositions used in thermal shift assays to determine the stability of HPV11 E2 TAD using DSF assays.

	1	2	3	4	5	6	7	8	9	10	11	12
A	41.23	42.25	43.23	40.23	42.25	44.25	44.23	41.25	44.23	40.23	41.25	44.23
B	41.25	42.23	43.23	40.25	41.23	43.22	42.23	41.23	44.22	40.25	41.23	44.23
C	41.23	41.23	43.23	40.27	41.25	44.25	41.27	41.27	44.23	40.27	41.27	44.23
D	41.27	42.27	42.27	40.23	41.27	44.23	41.27	41.27	44.23	40.25	41.27	43.23
E	41.25	42.27	43.25	40.25	41.23	43.25	40.25	41.23	44.28	40.25	41.23	44.28
F	41.27	42.25	43.23	40.27	41.27	43.25	41.27	41.27	43.25	40.27	41.27	44.3
G	42.27	42.27	43.25	41.27	42.27	43.25	41.27	42.27	44.28	40.27	42.27	44.28
H	42.25	43.27	43.25	41.25	42.25	44.23	41.25	42.25	44.23	41.25	42.25	44.23

**Table A.5 NMR buffer screen 3 DSF results**