INTERACTION STUDIES BETWEEN TRUNCATED HPV16/18 E7 ONCOPROTEIN WITH CTCF/YB-1 TRANSCRIPTION FACTORS COMPLEX

FATIN ATHIRAH BINTI A. WAHAB

UNIVERSITI SAINS MALAYSIA
2016

INTERACTION STUDIES BETWEEN TRUNCATED HPV16/18 E7 ONCOPROTEIN WITH CTCF/YB-1 TRANSCRIPTION FACTORS COMPLEX

by

FATIN ATHIRAH BINTI A. WAHAB

Thesis submitted in fulfilment of the requirements for the degree of Master of Science

May 2016

ACKNOWLEDGEMENT

In the name of Allah s.w.t the most gracious and the most merciful, I would like to thank all the people who contributed in various ways to the work described in this thesis. Next, I would like to express my appreciation and gratitude to my supervisor, Dr. Venugopal Balakrishnan for giving me the opportunity to pursue my postgraduate studies. He also provides invaluable suggestion, guidance and support in both my academic and mental formation here in INFORMM. A huge thankyou also goes to my co-supervisor, Prof. Shaharum Shamsuddin, who has shown great support, fruitful discussion, for always having time to listen, advice and guide to be on the right track.

I would also like to specially thank my lab-mates, Malarveli Suppan, Kalaivani Muniandi and Sanggetha Periya for their constant supports, advices, ideas, inspirations and helps throughout my project. My gratitude also goes out to Nik Nur Syazana Binti Nik Mohamed Kamal and Nur Adila Fadzil for all the things they have helped me with along the way.

High pleasure is due to my mother, Fadzilah Md. Salleh and family members for their loves, encouragements and supports during my studies. Finally, I would like to express my appreciation to my husband, Mohamad Erwan bin Hairuddin and my son, Muhammad Ruhi Mikail for their utmost loves, patience and supports. This is as much success for them as it is for me. This thesis is also dedicated to my late father, Dr. Abdul Wahab bin Abdul Hamid for being my greatest inspiration.

To INFORMM, thank you for providing a warm and conducive environment. My appreciation also goes to all lecturers, admin staffs and lab staffs for helping me

either directly or indirectly in improving my studies here. Thank you for all the supports given.

This research was financially funded by research grant from Exploratory Research Grant Scheme (ERGS), Ministry of Higher Education (MoHE). Thank you also goes to MyBrain15, MoHE for supported in funding my studies.

Fatin Athirah Binti Abdul Wahab

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

%	percentage
µg	microgram
µL	microliter
3D	three dimensional
∞	infinity
A	absorbance
Aa	amino acid
Ad	adenovirus
APP	Amyloid Protein Precursor
APS	ammonium persulfate
bp	base pair
BP	blotting paper
BSA	bovine serum albumin
C-	carboxyl
C	Celsius
$CaCl_2$	calcium chloride
CFTR	Cystic Fibrosis Transmembrane Conductance Regulator
Ci	curie
CK	casein kinase
CnBr	cyanogen bromide
Co^{2+}	cobalt ion
Co-IP	co-immunoprecipitation assay
CR	conserved region
CSD	cold shock domain
Cu^{2+}	copper ion
dbp	DNA binding protein
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTPs	deoxynucleotides
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid

EMSA	electro mobility shift assay
EtBr	ethidium bromide
FL	full length
g	gram
HCl	hydrochloric acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HF	high fidelity
HPV	human papilloma virus
hr	hour
IgG	immunoglobulin G
IMAC	immobilized metal ion affinity chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
kb	kilo base
kDa	kilo Dalton
L	litre
LB	Luria Bertani
LCR	long control region
M	molarity/molar
mCi	milicurie
mg	milligram
MHC	major histocompatibility complex
min	minute
mL	millilitre
mM	milimolar
mRNA	messenger ribonucleic acid
mRNPs	mRNA-protein complexes
N-	amino
N	normality
Na_2HPO_4	disodium hydrogen phosphate
Na_2PO_4	disodium phosphate
NaCl	sodium chloride
$NaHCO_3$	sodium hydrogen carbonate
NaOH	sodium hydroxide
NCBI	National Centre for Biotechnology Information

NeP1	negative protein 1
ng	nanogram
Ni^{2+}	nickel ion
$NiSO_4.6H_2O$	nickel sulphate
nm	nanometer
OD	optical density
ORF	open reading frame
PBS	phosphate buffered saline
PBS-T	phosphate buffered saline- Tween 20
PCR	polymerase chain reaction
pRB	retinoblastoma protein
PVDF	polyvinylidene difluoride
RFX	regulatory factor X
RNA	ribonucleic acid
s	second
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulphate – polyacrylamide gel electrophoresis
SV40	simian vacuolating virus 40
T_A	annealing temperature
TBE	tris-borate EDTA
TE	tris-EDTA
T_M	melting temperature
TS	thiol-activated sepharose
U	unit
UV	ultra-violet
V	voltage
X	times
x g	gravity
Yb-1	Y-box binding protein 1
YY1	yin and yang1
ZF	zinc finger
Zn	zinc
Zn^{2+}	zinc ion

KAJIAN INTERAKSI ANTARA TRUNKAT ONKOPROTIN HPV16 /18 E7 DENGAN KOMPLEKS FAKTOR TRANSKRIPSI CTCF/YB-1

ABSTRAK

Human papillomavirus (HPV) adalah agen etiologi kanser serviks dimana HPV16 dan HPV18 tergolong dalam kumpulan HPV berisiko tinggi. Kajian terdahulu mendedahkan interaksi antara onkoprotin HPV16/18 E7 dengan kompleks faktor transkripsi CTCF/Yb-1. Interaksi ini menyebabkan ekspresi c-myc meningkat dengan ketara yang membawa kepada percambahan sel. Dengan membahagikan HPV16 E7 dan HPV18 E7 kepada saiz yang lebih kecil, lokasi tepat di mana interaksi terhadap CTCF/Yb-1 dapat diketahui. Hal ini akan jelas mencirikan interaksi biokimia dan memberi pengetahuan yang lebih baik terhadap interaksi tersebut. Siri trunkat HPV16 E7 dan HPV18 E7 telah dihasilkan melalui tindak balas berantai polimerase, berdasarkan analisis kehidrofobian. Semua produk trunkat HPV E7, CTCF-Zn dan Yb-1 CSD telah diklon ke dalam vektor pET16bSH3 dan diekspres di dalam *E. coli* BL21 (DE3). Interaksi protin-protin antara trunkat HPV16/18 E7 dengan kompleks CTCF/Yb-1 telah dianalisa melalui ujian Pull-down. HPV16 E7 dan HPV18 E7 jujukan penuh telah digunakan sebagai kawalan. Protin-protin trunkat HPV16 E7 dan HPV18 E7 telah berjaya diekspres dan dibersihkan. Ujian Pull-down menunjukkan interaksi antara CTCF/Yb-1 dengan HPV16 dan HPV18 E7 jujukan penuh. Walau bagaimanapun, domain-domain trunkat HPV16 E7 dan HPV18 E7 tidak menunjukkan sebarang interaksi sama ada dengan CTCF-Zn mahupun Yb-1 CSD. Pertamanya, ketidakbolehan domain trunkat untuk berinteraksi dengan protin lain mungkin kerana lipatan protin dimer yang salah. Mutasi kepada

protin asal mungkin menyingkirkan interaksi hidrofobik dan mengakibatkan degradasi proteolisis. Hal ini menyebabkan struktur asli menjadi kurang stabil. Keduanya adalah kerana struktur protin dimer E7 itu. Pemangkasan tanpa domain CR1/CR2 atau domain terminal-C akan menghapuskan ikatan zink. Tanpa zink, E7 tidak dapat membentuk dimer seterusnya mengganggu interaksi dengan protin yang berpotensi menjadi pasangan. Kesimpulannya, HPV16 E7 dan HPV18 E7 tidak boleh dipendekkan berdasarkan profil kehidrofobian sahaja dimana ia menjejaskan lipatan dan kestabilan protin itu. Trunkasi HPV16 E7 dan HPV18 E7 yang kekurangan terminal -N atau –C mungkin menyebabkan kestabilan protein asal berkurangan, seterusnya menganggu interaksi protin-protin.

INTERACTION STUDIES BETWEEN TRUNCATED HPV16/18 E7 ONCOPROTEIN WITH CTCF/YB-1 TRANSCRIPTION FACTORS COMPLEX

ABSTRACT

Human Papillomavirus (HPV) is the aetiology agent of cervical cancer where HPV16 and HPV18 are the high-risk group of HPV. Previous study had revealed the interaction between HPV16/18 E7 oncoprotein with CTCF/Yb-1 transcription factors complex. This interaction resulted in marked enhancement of c-myc expression that leads cell proliferation. By truncating HPV16 E7 and HPV18 E7 into smaller size, the exact location of interaction towards CTCF/Yb-1 can be known. This will clearly characterise the interaction biochemically and provide a better insight of the interaction. Series of truncated HPV16 E7 and HPV18 E7 had been produced by polymerase chain reaction (PCR) based on hydrophobicity analysis. The truncated HPV E7 products, CTCF-Zn domain and Yb-1 cold shock domain (CSD) had been cloned into pET16bSH3 vector and expressed in *E. coli* BL21 (DE3). The protein-protein interaction between truncated HPV16/18 E7 and CTCF/Yb-1complex had been analysed via Pull-down Assay. HPV16 E7 and HPV18 E7 full length had been used as control. The series of truncated HPV16 E7 and HPV18 E7 proteins were successfully expressed and purified. Pull-down assay displayed interaction between CTCF/Yb-1 and HPV16/18 E7 full length. However, the truncated domains of HPV16 E7 and HPV18 E7 did not form any binding with neither CTCF-Zn nor Yb-1 CSD. Firstly, the inability of truncated domain to interact with other protein might be due to false folding of the dimer protein. Mutation to a wild type protein might

remove the hydrophobic interaction and cause proteolytic degradation. Thus, the native structure becomes less stable. Secondly, it might be due to structure of E7 dimer protein. Truncation that lack of CR1/CR2 domain or C-terminal domain will abolishes zinc binding. In the absence of zinc, E7 cannot form dimer thus disrupting interaction with potential protein partner. Therefore, we concluded that HPV16 E7 and HPV18 E7 cannot be truncated based on hydrophobicity profile only due to protein folding and stability. The truncation of HPV16 E7 and HPV18 E7 that lack of N- or C-terminal might reduce the stability of the native protein thus disrupt the protein-protein interaction.

CHAPTER ONE

INTRODUCTION

1.1 Human papillomavirus and cervical cancer

Human papillomavirus (HPV) is a sexually transmitted virus that comprises of circular double stranded DNA with an approximately size of 8 kb. Almost 150 genotypes have been identified so far and about 40 of them infect genital mucosa (Schiffman *et al*., 2011). These HPVs caused more than 90% of all cervical cancer cases (Muñoz *et al*., 2003) with an estimated 100% of cervical cancer cases in United Kingdom (Cancer Research UK, 2012). They are categorized into two groups; low-risk and high-risk, based on their presence in malignant lesion of the cervix. The low-risk HPVs such as HPV6 and HPV11 formed localized benign warts and these warts do not undergo malignant progression even without treatment while high-risk HPVs such as HPV16 and HPV18 can cause cervical carcinoma. According to World Health Organization (WHO, 2015), HPV16 and 18 associated with 70% of cervical cancers and precancerous cervical lesions. In 2012, about 445,000 new cervical cancer cases are diagnosed worldwide, especially in women living in less developed regions. It remains as the second most common cancer among women worldwide and the third most common cancer among Malaysian women (MoH Malaysia, 2007, WHO, 2015). High-risk HPV infections also associated with anal cancer, a subset of head and neck cancers (Nair and Pillai, 2005) and a subset of vulvar, vaginal and penile cancers (Steenbergen *et al*., 2005).

1.2 HPV genome organization and life cycle

The HPV genome can be separated into three regions; an early (E) region, a late (L) region and a non-coding long control region (LCR). The early region encodes non-structural proteins; the late region encodes structural proteins while LCR controls viral replication and gene expression. They are labelled as E or L based on their expression specified in early or late stage of epithelium differentiation, and numbered based on size. As the number increase, the corresponding open reading frame (ORF) becomes smaller. There are six ORFs expressed early in the epithelial differentiation marked as E1, E2, E4, E5, E6 and E7 and two late ORFs; L1 and L2. The HPV infection was established by maintaining the expression of the viral genome at the basal layer of the epithelium. The viral life cycle goes through success stages of genome amplification, virus assembly and virus release throughout the basal epithelial cells differentiation. The virus released simultaneously cause changes in gene expression pattern of early to late genes including L1 and L2 that formed the viral capsid where L1 is the major capsid protein while L2 functions as a link to the plasmid DNA (Schiffman *et al.*, 2007) . The HPV genome organization and the general function of their respective proteins are illustrated in Figure 1.1.

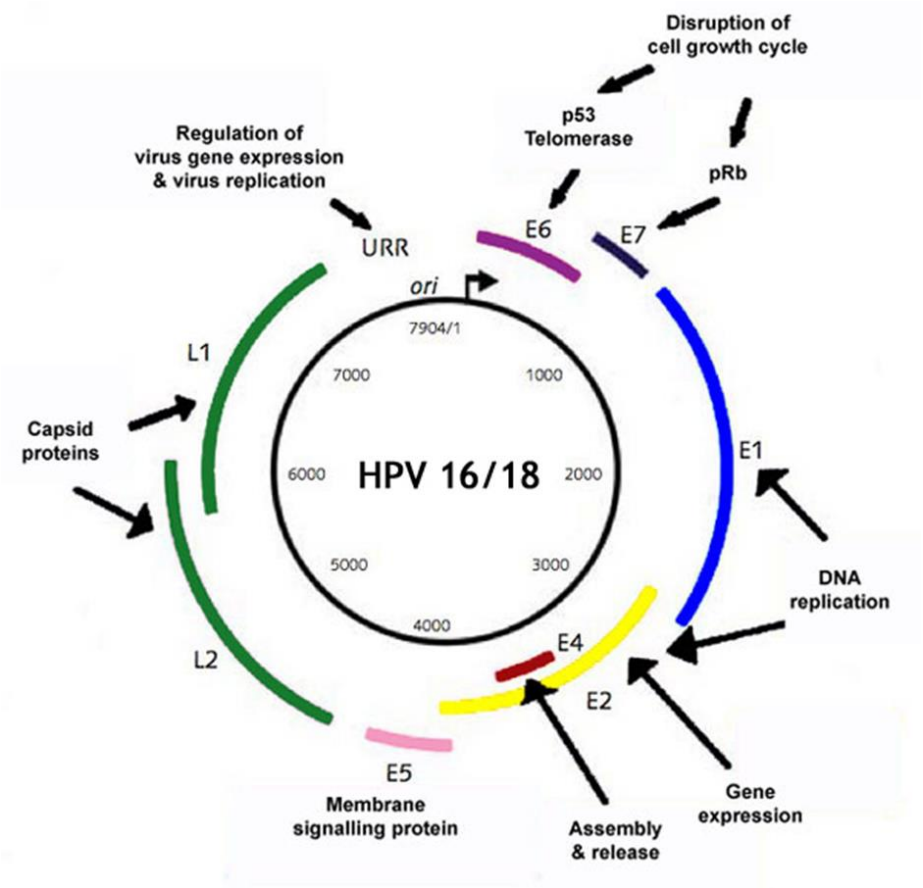

Figure 1.1: HPV16/18 genome organization and the general function of their respective proteins. The HPV genome can be divided into three parts; early (E) region, a late (L) region and non-coding long control region (LCR). Adapted from Schiffman *et al*. (2007).

The E6 and E7 oncogenes are associated with almost all cervical carcinomas cases. These viral proteins are constantly expressed in the tumours and they are necessary for both induction and maintenance of the transformed phenotype. Mostly, the carcinogenesis of human cells is caused by inactivation of tumour suppressor p53 and/or retinoblastoma protein (pRB) where E6 associates with p53 and E7 associates with pRB (Narisawa-Saito and Kiyono, 2007). Among these two viral proteins, the E7 protein was marked as the main transforming activity of high-risk HPVs (McLaughlin-Drubin and Münger, 2009) where E7 alone is sufficient to immortalize human epithelial cell. Therefore E7 is considered to be a target of immunotherapy for cervical cancer (Fernando *et al.*, 1995).

1.3 HPV E7 oncoprotein

1.3.1 Biochemical characterization of E7 oncoprotein

The E7 oncoprotein was identified as the first oncogene of high-risk HPV (Vousden *et al.*, 1988). E7 is a small polypeptide composed of nearly 100 amino acid residues. The intrinsic enzymatic activity of this protein is not known but it has phosphoprotein characteristics; a consensus casein kinase II (CK II) phosphorylation site in their amino terminal. Their amino terminal also have a region of sequence homology with some conserved region (CR) 1 and entire CR2 of adenovirus (Ad) E1A, and related sequences in simian vacuolating virus 40 T antigen (SV40 T) (Figge *et al.*, 1988, Phelps *et al.*, 1988). These two conserve regions play important roles in transforming activities of cancerous cells by HPV E7 (Watanabe *et al.*, 1990, Phelps *et al.*, 1992, Jewers *et al.*, 1992). Besides that, CR2 domain consists of a LXCXE motif (Leucine, Cysteine and E is glutamic acid) that is important for

association of E7 protein with pRB (Münger *et al*., 1989). However, some studies claim that C-terminal E7 domain also contains pRB binding site but with low affinity (Patrick *et al*., 1994, Liu *et al*., 2006). Adjacent to Adjacent to the LXCXE motif is a consensus casein kinase II (CK II) that involve in phosphorylation of E7 protein (Firzlaff *et al*., 1989). The carboxyl terminal of E7 contains a zinc-binding domain; comprise of two copies CXXC motif separated by 29 amino acid (Barbosa *et al.*, 1989). This domain associates in dimerization of E7 in which it found to be involved in metal binding and mediate direct interaction with several proteins partner (Clemens *et al*., 1995, Münger *et al*., 2001). These important regions of E7 were illustrated in Figure 1.2. Moreover, a study by Alonso *et al.* (2002) had revealed the oligomerization properties of E7 where monomers, dimers and tetramers were detected.

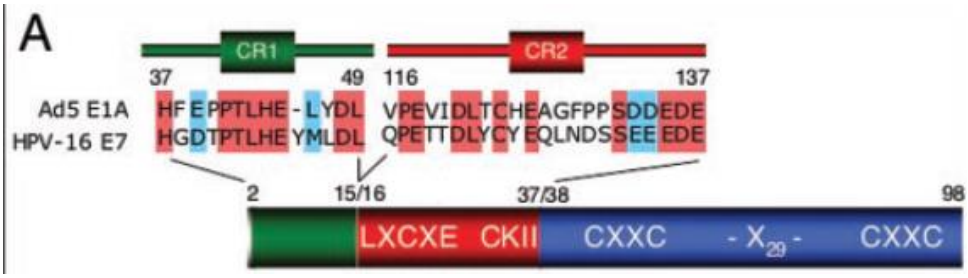

Figure 1.2: Illustration of important regions in HPV16 E7. The N-terminal consists of CR1 (green) and CR2 (red) where CR2 contains the core pRB-binding site (LXCXE motif), while the C-terminal (blue) comprises of a metal binding motif (two copies of CXXC motifs spaced by 29 amino acids). Adapted from Münger *et al*. (2004).

1.3.2 HPV E7 life cycle and cellular target

The HPV E7 protein is consistently expressed in HPV positive cervical carcinomas. During carcinogenic progression, the HPV genome frequently integrates into a host cell chromosome. These oncogenic activities of high-risk HPV E7 reflect their function during the viral life cycle where it intimately associated with the differentiation process of the infected epithelial cell. In order to establish a persistent infection, the HPV genomes were maintained at low copy number where it exclusively occurs in the terminally differentiated layers of epithelium (reviewed in Lee and Laimins, 2007).

The most important cellular target for HPV E7 in cell transformation is pRB tumour suppressor. The pRB is a family of closely linked protein including p107 and p130. It was found bind to E2F and inhibit transcriptional activation function. E2F is a heterodimer transcription factor that involved in regulation of G1/S phase of cell cycle progression. In normal cells, the hypophosphorylated pRB protein can bind to E2F-family transcription factor. Then, the pRB-E2F complex down-regulated the transcription of certain genes necessary for DNA synthesis and cell cycle progression (Dyson, 1998). When G1 cyclin-dependent kinases phosphorylate pRb, E2F will be released and cause the cell cycle to progress into S phase. In cancer cells, E7 bind to unphosphorylated pRB via LXCXE motif and it prematurely induces cells to enter S phase by disrupting pRB-E2F complexes (Yim and Park, 2005). Thus, lead to abnormal cell progression. The high-risk HPVs bound to pRB with higher affinities compared to low-risk HPVs (Münger *et al.*, 1989). Furthermore, A. Suhrbier has found that cleavage of pRB by calcium activated cysteine protease calpain was promoted by E7 protein. This cleavage is necessary before E7 can stimulate the proteasomal degradation of pRB (Narisawa-Saito and Kiyono, 2007). Figure 1.3

summarized how HPV E7 binds with pRb interfering with the normal functions of the tumour suppressors.

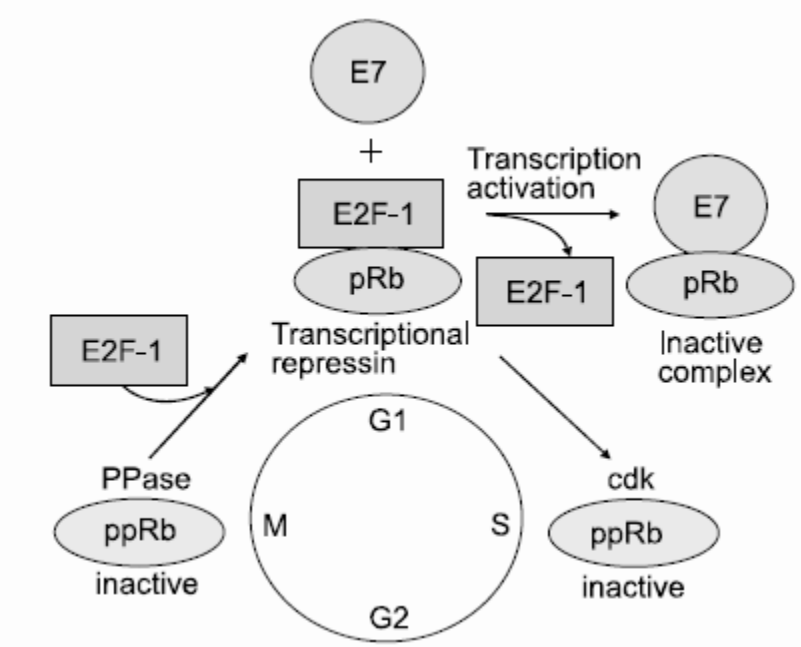

Figure 1.3: Life cycle of pRB-E2F complex disrupted by HPV E7. E7 interfering with the normal functions of the pRB tumour suppressors. Adapted from Yim and Park (2005).

In addition, E7 also has different targets that associated with cellular transformation as it can override the growth inhibitory activities of the cyclin dependent kinase inhibitors, including $p21^{CIP1}$ and $p27^{KIP1}$ (Münger *et al*., 2004). Besides that, cervical carcinoma cells that were infected with HPV E7 showed a significant increase in the level of c-myc expression (Gewin and Galloway, 2001, Oh *et al*., 2001, DeFilippis *et al.*, 2003). This finding showed that cervical carcinoma was tightly associated with *c-myc* gene amplification especially when infected with HPV16 where *c-myc* gene increased according to the grade lesion of cervical carcinoma (Abba *et al*., 2004). Furthermore, the cells transfected with E7 protein also significantly increases the level of *c-myc* gene expression (Liu *et al*., 2007). These findings have confirmed that the E7 oncoprotein plays an important role in the activation of *c-myc* gene, but the mechanism of action is still unknown.

1.4 Transcription factors

1.4.1 CTCF protein

CTCF is also known as CCCTC-binding factor, is highly expressed and conserved in eukaryotes. It is well-known for its 11 zinc fingers (ZFs) that perform various regulatory functions in the cell. It was first isolated by Lobanenkov (1990) as a transcriptional repressor that bind to 50-60 bp sequence within the promoter region of chicken *c-myc* gene (Klenova *et al*., 1993). Following that, CTCF was identified bind to numerous different sequences in the mouse, avian and human c-myc promoter (Filippova *et al*., 1996, Klenova *et al*., 2001). The structure of CTCF can be divided into three parts; an amino (N) terminal domain, a central zinc (Zn) domain (with 11 zinc fingers) and a carboxyl (C) terminal domain (Vostrov *et al*., 2002). The Zn domain displays almost 100% homology between mouse, chicken and human. The ZFs can be classified into two types; the first ten ZFs are 30 amino acid residues comprised of a pair of cysteine and a pair of histidine (C_2H_2-type) that is separated by 12 amino acids while the 11^{th} ZF is the C_2HC-type (Ohlsson *et al*., 2001). The structural features of CTCF were illustrated in Figure 1.4.

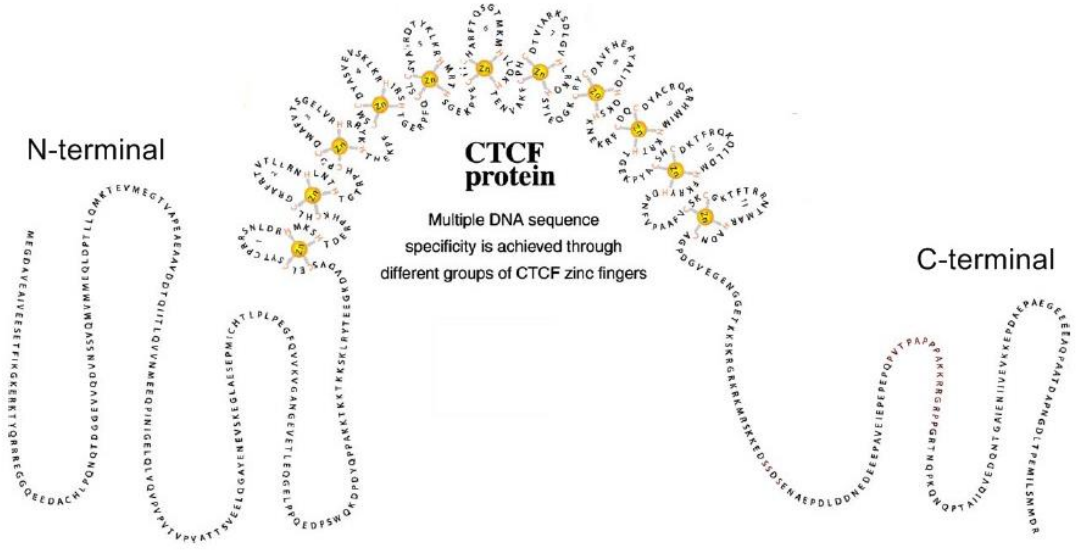

Figure 1.4: Schematic illustration of CTCF structural features adapted from Klenova *et al*. (2002). Each ZF contains a pair of cysteine residues and a pair of histidine, 12 amino acids apart except for the 11th ZF (C_2HC-type). The Zn domain displays almost 100% homology between mouse, chicken and human. Certain sets of ZFs is necessary for the CTCF binding to a target sequence but unnecessary to form binding with others.

CTCF found in nearly 65,000 sites in the genome of mammals in which 15% of them bind on the promoter proximal. Therefore, it was concluded that CTCF is capable of managing the architecture of the chromosome by binding to different sites. In addition, CTCF also has functions in the regulation of gene expression through repression or activation promoter, the regulation of distant chromatin interactions and the insulation of enhancers or silencer (Phillips and Corces, 2009).

CTCF is involved in various roles in gene regulation including context-dependent promoter activation or repression such as regulation of *c-myc* gene expression. It was found to bind the c-myc promoter and down-regulated the expression. Besides that, CTCF was found to associate with Negative protein 1 (NeP1) at a modular silencing element, 2.4 kb upstream of the chicken *lysozyme* gene (Baniahmad *et al.*, 1990, Burcin *et al.*, 1997). The CTCF was also reported as a transcriptional repressor of NeP1 based on the observation that the chicken lysozyme silencer containing a CTCF-binding site adjacent to a thyroid hormone response element, was synergistically decrease the reporter gene expression (Kohne *et al.*, 1993). However, the full composite element displayed inconsistent results dependent on the cell line used. Therefore, it had been concluded that CTCF binding site has only minimal impact on the reporter gene expression. Following analysis had identified that CTCF also involve in activation of gene expression. In other study, Vostrov and Quitschke (1997) identified CTCF was bind immediately to upstream Amyloid Protein Precursor (APP) and up-regulate the gene expression. In this work, transgene assays were used to support the conclusion that CTCF also can serve as a transcriptional activator. Besides that, CTCF also involved in activation of the *APP* gene where it binds to the promoter region (APPβ) and up-regulates the gene transcription (Chen *et al.*, 2013).

Using chromatin immunoprecipitation with microarrays (ChiP-chip), it was found that CTCF and cohesin work together to regulate a function (Wendt *et al*., 2008). CTCF co-localize with cohesin and cause changes to chromatin structure and control gene regulatory elements. Cohesin is a protein complex that regulates stabilization of chromatin loops, gene transcriptional and DNA repair (Kagey *et al*., 2010, Faure *et al*., 2012). Usually, cohesin does not form binding with DNA but it was found associated with CTCF (Stedman *et al*., 2008). CTCF leads cohesin to its binding sites and cohesin helps CTCF to perform its insulator function. Hence, sequence-specific cohesin binding is dependent on CTCF but CTCF is not dependent on cohesin for its function (Wendt *et al*., 2008). The CTCF-cohesin complexes sustain the chromatin structures by co-localizing at various sites across the genome, such as Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) locus. At CTFR locus, CTCF formed complex with cohesin and created a chromatin loop that cause enhancer to reach promoter and activate transcription. Immediately upon knock-down of CTCF, the chromatin structure of CTFR locus was influenced, resulting in enhancement of gene expression (Gosalia *et al*., 2014).

Furthermore, CTCF was found to play an important role as an insulator. An insulator is a DNA sequence that suppresses gene activation by inhibiting enhancer from activating promoter (Herold *et al*., 2012). β-globin locus is the first individual locus that found in the discovery of CTCF's function. Chung *et al*. (1993) had found that chicken β-globin locus can block the enhancer activity where a 42 bp fragment had been identified responsible for it. The fragment was found to be the CTCF-binding site thus suggesting that CTCF plays a role in the insulation activity (Bell *et al*., 1999). Moreover, CTCF assist in formation of chromatin loops that incorporate *β-globin* gene and the locus control. During chromatin loops formation, β-globin

sites are relocated and blocked the enhancer signals thus repressed the gene transcription (Splinter *et al*., 2006).

Another significant locus is the imprinted H19/Igf2 locus. The *H19* gene is expressed exclusively from the maternal chromosome and *Igf2* genes are expressed exclusively from the paternal chromosome. They are far apart from each other where the imprinting control region (ICR) lies between them (Pidsley *et al*., 2012). This is situated between 7 and 9 kb downstream of the *H19* gene and 90 kb downstream of the *Igf2* gene. This region contains four CTCF-binding sites. On the maternal chromosome, the CTCF formed binding with unmethylated ICR. This binding blocked the interaction between H19-proximal enhancer and Igf2 promoter thus inactivated Igf2. However, CTCF cannot bind the paternal ICR due to methylation, causing the H19 enhancer to activate Igf2 (Hark *et al*., 2000). CTCF binding can formed various chromatin loops integrating specific alleles with enhancers and promoters at the maternal ICR (Yoon *et al*., 2007).

Moreover, subsequent analysis revealed that CTCF helps in long-range chromosomal interaction via looping. CTCF was found helps in interchromosomal interaction formation by co-localizing the H19/Igf2 locus on chromosome 7 with the Wsb1/Nf1 locus on chromosome 11. When CTCF was knocked down, no interchromosomal interaction formed thus cause changes in *Wsb1/Nf1* gene expression (Ling *et al*., 2006). Furthermore, CTCF-binding sites are found at both active and inactive domain boundaries (Cuddapah *et al*., 2009), and some are situated at the borders of the lamina-associated domains. The enrichments of CTCF binding at the lamina-associated domain boundaries suggested that CTCF has a role in shaping the three-dimensional chromatin organization (Guelen *et al*., 2008).

Meanwhile, CTCF can bind to numerous protein partners. The DNA-binding proteins that form complex with CTCF are include Y-box binding protein 1 (Yb-1) (Chernukhin *et al*., 2000), Kaiso (Defossez *et al*., 2005), regulatory factor X (RFX) and MHC class II transactivator (CIITA) (Majumder *et al.*, 2006) and Yin and yang 1 (YY1) (Donohoe *et al*., 2007). Among these proteins, Yb-1 interacts with CTCF-Zn domain and cooperates with CTCF in transcriptional repression of *c-myc* gene (Chernukhin *et al*., 2000, Balakrishnan, 2008). The CTCF-Zn domain is capable of binding either to DNA or protein, unlike other multi-ZF proteins that bind to DNA only or protein only.

1.4.2 Yb-1 protein

The multifunctional Y-box-binding (Yb) proteins are characterized by a highly conserved nucleic acid binding domain, termed cold-shock domain (CSD). It is widely distributed among bacteria, plants, animals and human (Wolffe *et al*., 1992). These proteins bind to DNA containing Y-box or inverted CCAAT box in the promoter region of various eukaryotic genes and regulate gene expression. Previous reports showed that the Yb-1 protein was the most responsive to DNA damage, drug treatment, stress-related stimuli including ultra-violet (UV) irradiation, anti-cancer agents and heat in cancer cells in vitro (Ito *et al*., 2012, Wang *et al*., 2012).

Y-box binding protein 1 (Yb-1) was named in 1988 when transcription factors were found binding to Y-box motif of major histocompatibility complex (MHC) class II gene (Didier *et al*., 1988). Y-box proteins can be divided into three subfamilies. The first subfamily is Yb-1 that includes human Yb-1, rabbit p50, chicken-Yb-1, bovine EFI-A, mouse MSY-1 and others. It contains 100% homology

with DNA binding protein B (dbpB). These proteins have characteristic of somatic cells that perform various functions and they are the most studied ones. The second subfamily is Yb-2 subfamily, includes protein mouse MSY-2 and FRGY2 (p54/56). These proteins also known as dpbC and contrin, and they are specific for germ cells. Finally is the Yb-3 subfamily, contains mouse MSY-3 and human dpbA (csdA) protein. These proteins supposed were synthesized during embryonic development and disappeared by the moment of birth. However, Yb-3 mRNA can be detected in some tissues of adult organism (Mastrangelo and Kleene, 2000).

The human *Yb-1* gene is located in the first chromosome (1p34) where it consists of 19 kb nucleotides including eight exons. After splicing, the *Yb-1* mRNA has length about 1,500 bp only encoding 43 kDa protein (Makino *et al*., 1996). The structure of Yb-1 is comprised of three domains; N-terminal domain that contains high content of alanine and proline (A/P domain), CSD domain and elongated C-terminal domain containing alternating clusters of positively and negatively charged amino acid residues (Wolffe, 1994). The amino acid sequences of proteins from various subfamilies had demonstrated that their CSDs are identical by more than 90%, while in other parts of the molecules, no significant homology is observed. It was discovered that the amino acid sequence of dbpB is completely identical to Yb-1, and dbpA is homologous to it by 46%. The Yb-1 and dpbA have 44% sequence identical to Cold shock protein A (CspA) from *Escherichia coli* (Kudo *et al*., 1995). The three-dimensional (3D) structures of Yb-1 CSD and CspA turned out to be very close to each other and they consist of five β-strands packed antiparallel in a β-barrel. The CSD has the so-called RNA recognition motifs, RNP-1 and RNP-2 that involve in specific and non-specific interactions with DNA and RNA. The presence of a CSD is a specific feature of Y-box binding protein that allows them attribute to a wider

group of proteins containing a CSD (Eliseeva *et al*., 2011). The structural organization of Yb-1 and the difference between structure Yb-1, Yb-2 and Yb-3 were illustrated in Figure 1.5.

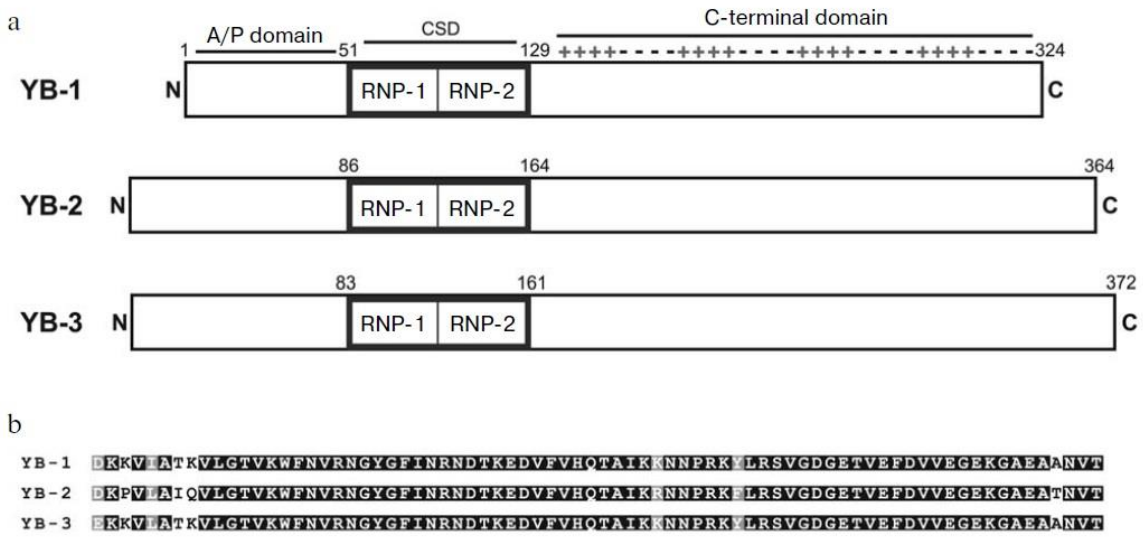

Figure 1.5: The domains of Y-box binding proteins subfamilies. Adapted from from Eliseeva *et al*. (2011).

In both somatic and germ cells, Yb-1 protein act as the major translational repressor and/or mRNA chaperones of mRNA-protein complexes (mRNPs) in the cytoplasm. It is primarily localized in the cytoplasm but in response to physical and environmental signals, Yb-1 translocate into the nucleus. Within the nucleus, the Yb-1 protein has been involved in major nucleus activities such as DNA repair, pre-mRNA splicing and transport, and transcriptional regulation. By passing from the cytoplasm to the cell nucleus, Yb-1 also activates transcription of several protective proteins including proteins that provide multidrug resistance of cells. It also enhances resistance of cells to xenobiotics and ionizing radiation when involved in DNA repair in the nucleus (Eliseeva *et al*., 2011). Therefore, Y-box proteins have been suggested as multifunctional coordinators of gene expression in both nucleus and cytoplasm (Wang *et al*., 2012).

The Yb-1 protein has broad functions in nucleic acid-binding properties. It can bind to both DNA- and RNA-binding protein that has properties of a nucleic acid chaperone, and it interacts with a great variety of other proteins. It shows a higher affinity for single stranded regions of DNA and RNA and markedly decreases the melting temperature of double helicases. Yb-1 has a higher affinity for damaged DNA structure and promotes DNA repair. During transcription, Yb-1 binds to pre-mRNA on chromosomes and accompanies mRNA molecules all the way through their life. In the cell nucleus, Yb-1 performed alternative splicing of pre-mRNA. In the cell cytoplasm, almost all Yb-1 molecules are associated with translated or non-translated mRNAs, determining their functional activity, stability and localization of translationally active mRNAs on actin skeleton. Thus, Yb-1 is involved not only in DNA- and mRNA-dependent processes, but in all steps of mRNA bio-genesis and functioning as well (Skabkin *et al*., 2006).

Besides that, Yb-1 content was found increased in cancer cells. In some studies, Yb-1 can be considered as an oncoprotein that stimulate cell proliferation and promotes metastasis, but in other studies showed that Yb-1 can be a tumour suppressor in a number of cases. Therefore it is suggested to be one of the most indicative markers for cancer therapy (Dahl *et al*., 2009, Lasham *et al*., 2013). Yb-1 can give implication in the activation of transcription of c-*myc* gene where it binds to the CT-rich sequence in the promoter region, results in docking of core transcription factors and RNA polymerase II. Yb-1 protein is also known to interact with diverse range of proteins including transcription factors, viral proteins, repair proteins and actin filaments. One of the interacting proteins is CTCF transcription factor. This Yb-1/CTCF interaction was found to regulate c-*myc* gene expression (Chernukhin *et al*., 2000, Kohno *et al*., 2003).

1.4.3 Effect of CTCF/Yb-1 complex towards c-*myc* oncogene and interaction with HPV16 E7 and HPV18 E7

The dysregulation of cell signalling and cell proliferation is commonly caused by inactivation of tumour suppressor genes and over-expression or mutation of proto-oncogenes. Activation of some oncogenes has been associated with many cancer cells including cervical carcinoma cells (Golijow *et al*., 2001, zur Hausen, 2000, Mouron *et al*., 2000). The c-*myc* gene has been found associated in many solid tumours and various human cell lines (Riou *et al*., 1990). The c-*myc* is a proto-oncogene and the cellular homologous of the v-*myc* transforming gene that encodes a helix-loop-helix leucine zipper domain transcription factor (Colby *et al*., 1983). The c-*myc* over-expression is a common event in cervical carcinoma (Iwasaka *et al*.,

1992) and low-level c-*myc* gene amplification often found in HPV infected cases. This outcome indicates that HPV may has a critical biological impact on the development and progression of carcinomas of the uterine cervix since HPV-associated cervical carcinomas can allow frequent alteration of c-*myc* gene (Zhang *et al*., 2002). Moreover, the integration of HPV sequences close to the c-myc locus has been found in some cervical cell lines, suggesting a synergistic role for HPV and c-*myc* gene in the development of cervical carcinoma (Nair *et al*., 1998).

Yb-1 multifunctional factor was found to interact with CTCF and are likely to play multiple roles in regulation of major cellular process. CTCF was known to bind promoter region of c-*myc* gene (Lobanenkov *et al*., 1990) and this interaction might affect c-myc expression. Expression of Yb-1 alone had no effect on c-myc reporter activity but co-expression of Yb-1 with CTCF resulted in a marked enhancement of CTCF-dependent transcriptional repression (Chernukhin *et al*., 2000). This binding was illustrated in Figure 1.6.

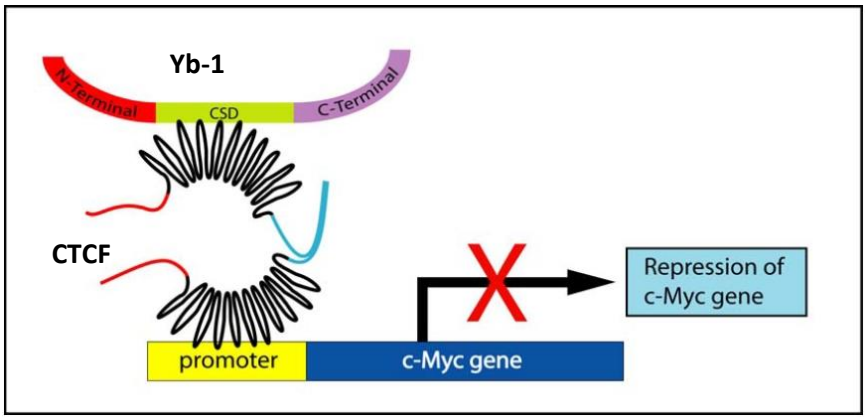

Figure 1.6: Effect of CTCF/Yb-1 complex on c-*myc* gene. Yb-1 CSD formed complex with ZFs of CTCF on promoter region of *c-myc* gene and cause down-regulation of gene expression (Chernukhin *et al*., 2000).

However, in cervical carcinoma infected with HPV16, high level of c-myc expression was observed (Abba *et al*., 2004). In addition, the presence of E7 oncoprotein was tightly associated with increased level of c-myc expression compare to normal cells (Oh *et al*., 2001; Gewin and Galloway, 2001; Lui *et al*., 2007). Then, Balakrishnan (2008) had found that the HPV16/18 E7 complex was interacting with CTCF, and dissociates Yb-1 from CTCF/Yb-1 complex on the c-myc promoter as illustrated in Figure 1.9. Before analysing the interaction between these complexes, the interaction of HPV16 and HPV18 E7 with CTCF/Yb-1 were analysed individually. When HPV16 E7 was introduced, it forms binding of CTCF-Zn domain and dissociate Yb-1 from the CTCF/Yb-1 complex as illustrated in Figure 1.7. In other hand, HPV18 E7 was found to bind Yb-1 without disrupting the CTCF/Yb-1 complex as illustrated in Figure 1.8. These interactions suggested the release of Yb-1 might induce the expression of c-*myc* gene thus contribute towards the enhancement of cell proliferation and formation of carcinoma cells.