Methylation dependent interactions of viral transcription factor Zta with DNA

By Kirsty Flower

A Thesis submitted for the degree of Doctor of Philosophy

School of Life Sciences

University of Sussex

September 2011

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree.

Signature:.....

Acknowledgments

The most important person that I need to thank is my supervisor, Dr Alison Sinclair. Without her unending patience, advice, encouragement and wisdom, in both an academic and personal capacity, I would not be where or who I am today. Working in her laboratory has provided me with the experiences which have shaped my career ambitions, in a supportive and enriching environment. I thank her from the bottom of my heart. I also extend my gratitude to my co-supervisor, Dr Michelle West, who has been an approachable and supportive individual that helped me throughout my DPhil. I look up to both these women as role models for the career in science I hope to achieve.

I also need to thank (soon to be Dr) Questa Karlsson, Dr Sarah Bailey and Dr Lizzie Verrall, who gave me the confidence to pursue a DPhil, and the support I needed when I started. Without their encouragement and friendship, I may never have made it through my first year! A huge amount of gratitude goes to another soon to be Dr, Natalie Braithwaite; thank you for demanding my friendship on our very first day, hula hooping distractions, lunchtime rants and summertime ice creams. I don't know what I would have done without you. I also want to thank Dr Sharada Ramasubramanyan, for her amazing friendship, support, and technical lab expertise. Thank you for always knowing the right thing to say. Thanks also to Helen Webb, Dr Andrea Gunnell, Dr Kay Osborn, Jamie Heather, Dr David Wood, Wei Duan, Lina Chen, Nick Balan, Angelica Edge and Michael McClellan for advice, conference fun, lunchtimes, pub visits, and lab pranks. Thank you to Dr Beth Hellen, David Thomas and Dr Sue Jones, for their knowledge and advice regarding bioinformatics, as well as their patience!

Thank you to the University of Sussex for funding me, and to all members of faculty who inspired me as lecturers during my undergraduate years, and became more like colleagues and friends during my postgraduate years. A special thanks goes to Dr David Whitehouse, a supportive and inspirational individual, who always had teaching hours for me, as well as time for a chat. Thank you to all the undergraduate and masters students whom I have taught in tutorials, lab demonstrations or during their final year project; you reminded me of how far I have come, as well as how far I have yet to go.

I am deeply indebted to my family; Mum, Dad, Granny, my brother Graham and last but definitely not least, my partner Sian; I got there in the end! I love you all very much, and your support has been invaluable to me throughout my time studying. Sian, thank you for dealing with my mood swings and for telling me all the things I needed to hear. Thank you for never giving up on me. My friends also deserve thanks for their support, despite not quite knowing what I did all day. Finally, thanks to Benson the dog (and of course his owners, Laura and Natty) for the time and space to write my thesis in peace whilst supposedly dogsitting.

UNIVERSITY OF SUSSEX

KIRSTY FLOWER PhD BIOCHEMISTRY

Methylation dependent interactions of viral transcription factor Zta with DNA

SUMMARY

Epstein Barr Virus, a human herpes virus associated with infectious mononucleosis, Burkitt's lymphoma, Nasopharyngeal Carcinoma and Hodgkin's disease, can infect B cells and establish latency. Zta, a member of the bZIP transcription factor family, is a viral transcription and replication factor required for activation of lytic cycle.

Zta is able to bind DNA, through specific Zta Response Elements (ZREs). Interestingly, Zta binds in a methylation dependent manner to specific CpG-containing ZREs, known as Class III ZREs. RpZRE3 is one of the first examples of such a site, and can be found in the promoter of a key viral lytic gene, *BRLF1*. Through computational analysis and *in vitro* binding assays, it was found that the core 7mer sequence of RpZRE3 is sufficient for complex formation *in vitro,* and that the core sequence can be found in a variety of human gene promoter regions.

As more methylation dependent ZREs emerged, a method to predict CpG-containing ZREs was devised using position frequency matrices, coupled with *in vitro* binding verification, leading to the discovery of 12 novel CpG containing methylation dependent ZREs. The regulatory regions of both the EBV genome (type 1) and the human genome were mapped for a list of experimentally verified ZRE core sequences (both methylation dependent and independent). This analysis allowed identification of novel viral and host genes, potentially under the control of Zta activation, which were further analysed using existing transcriptome and methylome data.

An investigation into potential host factors using the same methylation dependent mechanism was started, and an unknown protein was seen to bind in a methylation dependent manner to RpZRE3 in the absence of Zta *in vitro*.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AP-1	Activator Protein 1
ATP	Adenosine Triphosphate
AZT	3'-azido-3’-deoxythymidine
BART	BamH1 A transcripts
BCR	B Cell Receptor
BL	Burkitt's Lymphoma
bZIP	basic-leucine zipper
C/EBP	CCAAT/enhancer binding protein
ChIP	Chromatin Immunoprecipitation
Cp	latency promoter within BamH1 C digestion fragment
CpG	Cytosine Guanine dinucleotide
CRE	cyclic adenosine monophosphate (cAMP)-responsive Element
CREB	cAMP Response Element binding
CT	C Terminal
CTCF	CCTC-binding Factor
CTP	Cytidine Triphosphate
DAG	diacylglycerol
DMEM	Dulbecco/Vogt Modified Eagle's Minimal Essential Medium
DMSO	Dimethyl sulphoxide
DNA	deoxyribonucleic acid
DNMT	DNA methyl transferase
DPBS	phosphate buffered saline
DSL	Duplicated Sequence left
DSR	Duplicated Sequence right
DTT	Dithiothreitol

EBER	Epstein Barr encoded RNA
EBNA	Epstein Barr Nuclear Antigen
EBNA-LP	Epstein Barr Nuclear Antigen Leader Protein
EDTA	Ethylene Diamine Tetraacetic Acid
Egr1	Early Growth Response Protein 1
EGTA	Ethylene Glycol Tetraacetic Acid
EMSA	Electrophoretic Mobility Shift Assay
FBS	Foetal Bovine Serum
G:C	Guanine cytidine content in a specific molecule of DNA
GC	Germinal Centre
GCN4	General Control Protein
Gp	Glycoprotein
GST	Glutathione S Transferase
GTP	Guanosine Triphosphate
GTPase	Guanosine triphosphate hydrolase enzyme
HCl	Hydrochloric acid
HDAC	Histone deacetylase
HEK293	human embryonic kidney 293 cell line
HEPES	4-(2-hydroxyethyl) piperazine-1-ethanesulphonic acid
HHV-1	Human Herpes Virus 1, aka Herpes Simplex Virus 1 (HSV1)
HHV-2	Human Herpes Virus 2, aka Herpes Simplex Virus 2 (HSV2)
HHV-3	Human Herpes Virus 3, aka Varicella Zoster Virus (VZV)
HHV-4	Human Herpes Virus 4, aka Epstein Barr Virus (EBV)
HHV-5	Human Herpes Virus 5, aka Human Cytomeglovirus (HCMV)
HHV-8	Human Herpes Virus 8, aka Kaposi's Sarcoma-associated Herpesvirus (KSHV)
HL	Hodgkin's Lymphoma

HLA	Human Leukocyte Antigen
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IL	Interleukin
IM	Infectious Mononucleosis
IP3	inositol- 1,4,5-triphosphate
IR1-4	Internal repeat region 1-4
ITAM	Immunoreceptor Tyrosine-based Activator Motifs
JAK	Janus Activated Kinase
JNK	c-Jun N terminal Kinase
KCl	Potassium Chloride
LCL	Lymphoblastoid Cell Lines
LCV	Lymphocryptoviridae
LD	Lymphocyte Depleted
LMP	Latent Membrane Protein
MAGE4	Melanoma Antigen Family A4
MAPK	Mitogen Activated Protein Kinase
MBD	Methyl Binding Domain
MC	Mixed Cellularity
MeCP2	Methyl CpG binding protein 2
MEF-2D	Myocyte-specific Enhancer Factor 2D
$MgCl_2$	Magnesium Chloride
MHC	Major Histocompatibility Complex
MIRA-CHIP	Methylated CpG Island Recovery Assay Chromatin ImmunoPrecipitation
MLTF	Membrane-bound lytic murein transglycosylase
MNT	Max binding protein

mRNA	messenger Ribonucleic Acid
MS	Multiple Sclerosis
NaCl	Sodium Chloride
NCBI	National Center for Biotechnology Information
NHDL	non-Hodgkin's Lymphoma
NF1	Nuclear Factor 1
NFκB	nuclear factor kappa-light-chain-enhancer of activated B cells
NPC	Nasopharyngeal Carcinoma
NS	Nodular Sclerosing
ORF	Open Reading Frame
oriLyt	origin of lytic replication
oriP	origin of plasmid replication
PC2	polycomb protein complex 2
PENK	proenkephalin
PFM	Position Frequency Matrix
PHD	Pleckstrin Homology Domain
PI3K	Phosphatidylinositol 3-kinase
PKC	Protein Kinase C
PLC-γ2	Phospholipase C gamma 2
PMSF	phenyl methane sulfoyl fluoride
POU	family of proteins with conserved homeodomains, named for Pituitary specific transcription factor Pit-1, Octamer transcription factors, Unc-86 transcription factor
PSG	Penicillin Streptomycin Glutamine
pSS	primary Sjögren's Syndrome
PTLD	Post Transplantation Lymphoproliferative Disease
Qp	latency promoter within BamH1 Q digestion fragment

QPCR	Quantitative polymerase chain reaction
RA	Rheumatoid Arthritis
RAZ	the product of a spliced mRNA initiated from Rp consisting of part of BRLF1 and BZLF1
RNA	Ribonucleic Acid
Rp	promoter of *BRLF1*
Rpm	Rotation per minute
RPMI	Roswell Park Memorial Institute medium
RV	Rhadinoviridae
SCID	Severe Combined Immunodeficient
SDS	Sodium dodecyl sulphate
SLE	Systemic Lupus Erythematosis
SM	Protein product of mRNA spanning genes BSLF1 BMRF1
Smad	homologous to drosophila protein Mothers Against Decapentaplegic (MAD) and C. Elegans protein SMA
Sp	Specificity protein
STAT	Signal Transducer and Activator of Transcription
TBE	Tris-Borate-EDTA
TEMED	Tetramethylethylenediamine
TNF	Tumour Necrosis Factor
TPA	12-0-tetradecanoyl phorbol-13-acetate
TRAF	Tumour-necrosis-factor Receptor Associated Factor
TrkA	High affinity nerve growth factor receptor
U	Units
U1-5	Unique region 1-5
UBF	Upstream Binding Factor
UTP	Uridine Triphosphate

v/v	volume to volume
VAHS	Virus Associated Hemophagocytic Syndrome
w/v	weight to volume
WHO	World Health Organisation
Wp	latency promoter within BamH1 W digestion fragment
wt	wild type
XPC	Xeroderma pigmentosum, complementation group C
Zp	promoter of *BZLF1*
ZRE	Zta Response Element

Chapter 1: Introduction

1.1 Oncogenic virus discovery

It is currently thought that viruses account for 10-20% of all cancers. However, this association was not easily accepted for many years, due to the perception that cancer was not contagious. This view itself was derived from the inability to show a bacterial or parasitical association, which were the main infectious agents known in the late 19th century, and a failure to look on the sub-microscopic level, when microscopic organisms could not be found. In 1911, Peyton Rous showed that transmission of malignant tumours, specifically a spindle cell sarcoma in chickens, could be achieved using a cell free filtered tumour extract, but was met with opposition from those who questioned whether the filtrate was completely cell free, or whether it contained particles of cells. Those who eventually accepted the virus:chicken tumour association did not accept that such a mechanism was possible in mammals. Evidence of a viral cause of breast cancer in mice, mouse mammary tumour virus, in the 1940s, and later, a mouse leukaemia virus in the early 1950s were met with derision and not accepted until the late 1950s and early 1960s. It was not until 1966 that Rous received the Nobel Prize for his work and the recognition it deserved (reviewed by (zur Hausen, 1999, Javier and Butel, 2008, Epstein, 2001)).

Epstein Barr Virus (EBV, HHV-4) was the first human tumour causing virus identified, via the discovery of a new lymphoma. Burkitt's lymphoma (BL) was first described in 1958, after being observed as a common childhood cancer of the jaw in East Africa by Denis Burkitt (Burkitt, 1958). When he investigated the distribution of the cancer, it was noted that it appeared to be dependent on climate and geographical conditions. This presented the possibility that a virus

might be responsible, and sparked the curiosity of Antony Epstein. The virus was first isolated from a Burkitt's lymphoma tumour cell line in 1964 by Antony Epstein, Yvonne Barr and Bert Achong, and identified as a herpes virus by electron microscopy (Epstein et al., 1964). BL patients were observed to exhibit higher levels of EBV antibodies than healthy control patients. The epidemiological conditions observed were later attributed to the co-factor required for tumour formation (malarial infection) not the causative virus itself. Other conditions such as infectious mononucleosis, and cancers, such as nasopharyngeal carcinoma, are linked to EBV. The transformative properties of Epstein-Barr Virus were clearly shown by the ability to transform resting B cells *in vitro,* and to induce tumours in non-human primates, confirming the oncogenic potential of this virus (reviewed by (Rickinson and Kieff, 2001a, Rowe et al., 2009, Young and Murray, 2003)).

EBV is a prevalent human herpes virus, carried by 95% of the world adult population (WHO). Despite the high growth transforming efficiency of EBV, the majority of infected individuals do not develop EBV-associated cancers. EBV exhibits a biphasic life cycle; upon infection, the virus establishes a latent phase, during which a minimum number of proteins are expressed and the genome is replicated once per cell cycle by the host machinery. This allows the virus to remain undetected by the host immune system. Upon reactivation, the virus switches to the lytic phase, whereupon a large number of lytic genes are expressed, replicating the DNA and packaging it up in virions before release into the blood stream or saliva (Rickinson and Kieff, 2001b).

1.2 Herpesviridae family

The structure of a herpesvirus particle or virion consists of a linear double stranded DNA genome, encapsulated by an icosahedral capsid, made up of 162 hexagonal capsomeres, which in turn is surrounded by tegument, a mixture of different viral and cellular proteins, and all enclosed within an envelope studded with viral glycoproteins. The virions are 180-200nm in diameter. The expression of genes encoding a variety of enzymes for processing both proteins and nucleic acids is shared or highly homologous across the family (Roizman and Baines, 1991).

Herpesviruses are further classed into the sub-families alpha, beta and gamma on the basis of biological properties including the host range, length of reproductive cycle, latency mechanism and the individual properties of the viral genomes (Roizman et al., 1981). Alpha herpesviruses establish latency primarily in sensory ganglia, and include the simplexviruses (e.g. HSV1, also known as HHV-1 and HSV2, HHV-2) and the varicelloviruses (e.g. Varicella Zoster Virus (HHV-3), Pseudorabies Virus). Beta herpesviruses establish latency in leukocytes, replicate slowly and cause the infected cells to enlarge, and include cytomegalovirus (HCMV, HHV-5) and muromeglovirus (murine cytomegalovirus) (Roizman and Baines, 1991).

EBV (HHV-4) is a member of the gamma sub family of herpes viruses, which includes Kaposi's sarcoma-associated virus (KSHV, also known as HHV-8). This sub family is split into 2 further genera, the lymphocryptoviridae (LCV) or gamma-1 herpes virus (EBV), and the rhadinoviridae (RV) or gamma-2 herpes virus (HHV-8) (Rickinson and Kieff, 2001b). A major factor that differentiates the

behaviour of a gamma herpes virus from that of an alpha or beta herpesvirus is their ability to infect lymphocytes and induce lymphoproliferation (Barozzi et al., 2007). EBV can also successfully infect and replicate in epithelial cells (Rickinson and Kieff, 2001b).

1.3 EBV genome organisation

The EBV genome consists of a double stranded DNA molecule, approximate 172kbp in length, with 0.6 G:C content (Rickinson and Kieff, 2001b). For comparison, the human genome has an average G:C content of approximately 0.4, with the average G:C content for a 100kbp fragment ranging from 0.35 to 0.6 (Romiguier et al., 2010). There are around 86 genes, including non-coding RNAs, depending on the source of the DNA, as specific cell lines have particular deletions. There are 2 main types, known as Type 1 (de Jesus et al., 2003), and Type 2 (Dolan et al., 2006). Type 1 was the first to be sequenced (Baer et al., 1984), and is the best annotated in the NCBI genome library. Type 2 is more prevalent in people in equatorial Africa and New Guinea, whilst in developed countries, Type 1 is more dominant. However, both types can be found in immunocompromised patients. The most biologically relevant difference in sequence between Type 1 and Type 2 virus resides within the coding sequence of *EBNA-2* (Epstein-Barr viral nuclear antigen 2, a latency protein); there is only 55% homology between the 2 types in this gene, and this difference affects the ability of the virus to transform B cells; Type 2 is less efficient than Type 1. (Rickinson and Kieff, 2001b).

The genome was first sequenced using a fragment library created using BamHI digests (Baer et al., 1984). Each BamHI fragment was named from A to Z in

size from largest to smallest, and most of the ORFs are named according to their position and direction within a particular fragment, for example *BARF1* is the first rightward gene in the largest BamHI fragment (A) (Rickinson and Kieff, 2001b). There is a terminal repeat region, consisting of 0.5kb direct repeats, which mediates circularisation of the genome (Baer et al., 1984) . During lytic genome replication, long concatamers of newly synthesised genome sequence are cut at variable positions in the terminal repeat region, and upon infection, the genome will circularise, which sets the number of terminal repeats. This number is then constant during latent infection. Therefore, the number of repeats can be used to determine whether a group of latently infected cells arose from a single infected cell or the infection of many cells; this is referred to as clonality (Rickinson and Kieff, 2001b). The genome is made up of unique regions (U1-5) divided up by internal repeats (IR1-4). IR1 makes up the W repeats, consisting of multiple W promoters from which latent proteins can be transcribed, and the *BWRF1* gene, for which the function and whether a protein is even expressed is debatable (de Jesus et al., 2003). Two duplicate sequences are identified as origins of lytic replication (known as OriLyt L or DSL and OriLyt R or DSR).OriLyt L is adjacent to *BHRF1* and *BHLF1*, and OriLyt R is adjacent to *LF3*. *BHLF1* and *LF3* contain tandem repeat regions, IR1 and IR4. Two origins of lytic replication are often found in herpesviruses, however some EBV genomes have been shown to lack one or the other of the orilyts (Xue and Griffin, 2007). IR3 forms part of *EBNA-1*, and is known as the *EBNA-1* triplet (Wendelburg and Vos, 1998) (Figure 1.1).

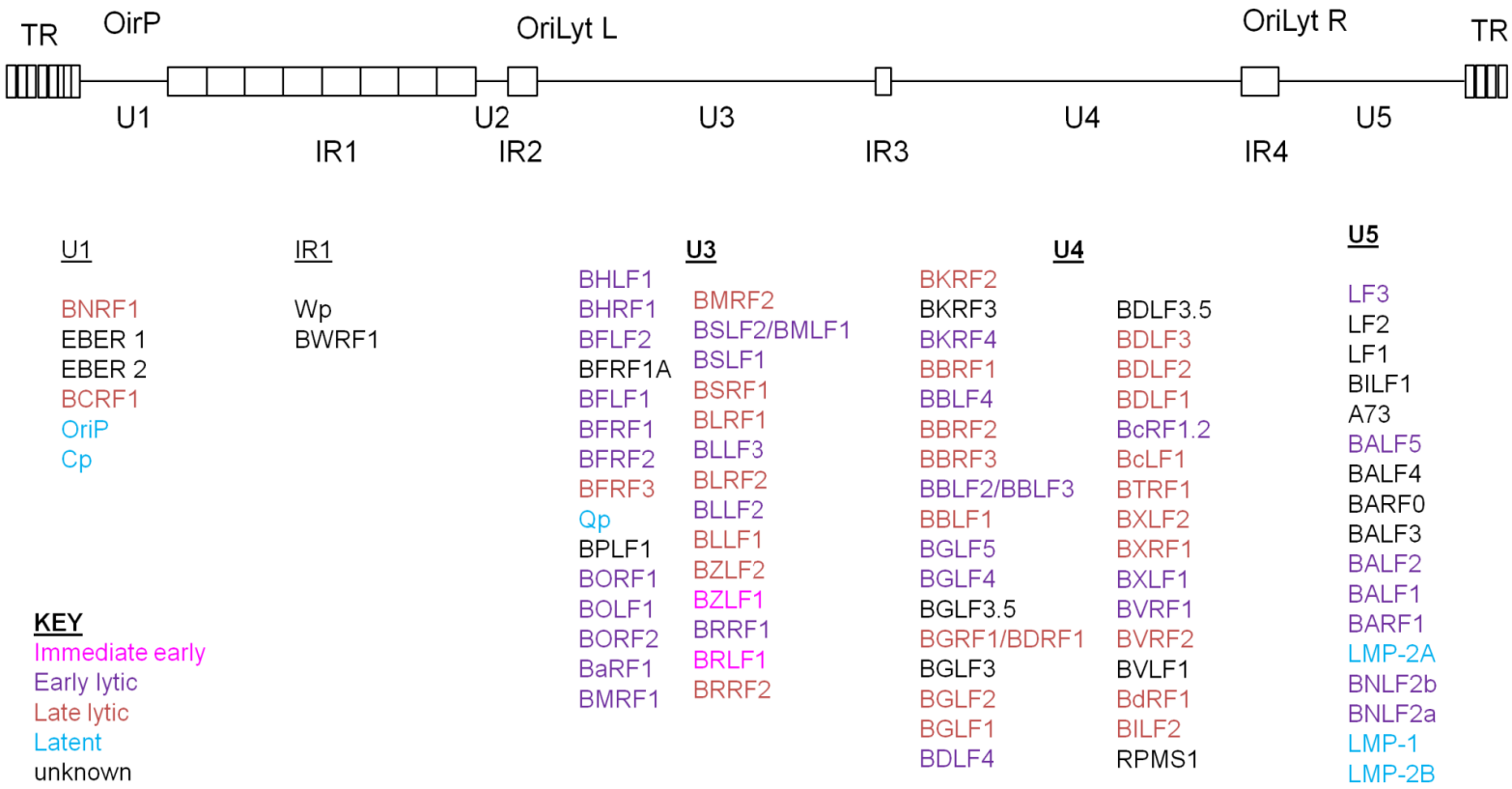


Figure 1.1 The EBV genome.

A schematic diagram to represent the features of the EBV genome. Unique regions of sequence are annotated U1-U5, and the internal repeat regions are annotated IR1 – IR4. OriP, the origin of plasmid replication, required for latent replication of the genome, is indicated within U1. The two Origins of lytic replication, OriLyt L and OriLyt R, are indicated in close proximity to IR2 and IR4. Terminal repeats (TR) at either end facilitate circularisation of the viral genome upon infection. A list of viral genes within each region is colour coded, according to their classification as a latent, immediate early lytic, early lytic or late lytic gene/promoter. Data from (Rickinson and Kieff, 2001a, Farrell, 2005, Baer et al., 1984).

The latency genes *EBNA-1, EBNA-2, EBNA-3A, EBNA-3B, EBNA-3C*, and *EBNA-LP* are spliced versions of a long transcript that spreads over 90kb of the genome. Expression can be initiated from either the Cp or Wp, and it is then differentially spliced to form the mRNA of each specific gene.. *EBNA-1* expression can also be initiated from another promoter, Qp (Rickinson and Kieff, 2001b) (Figure 1.2).

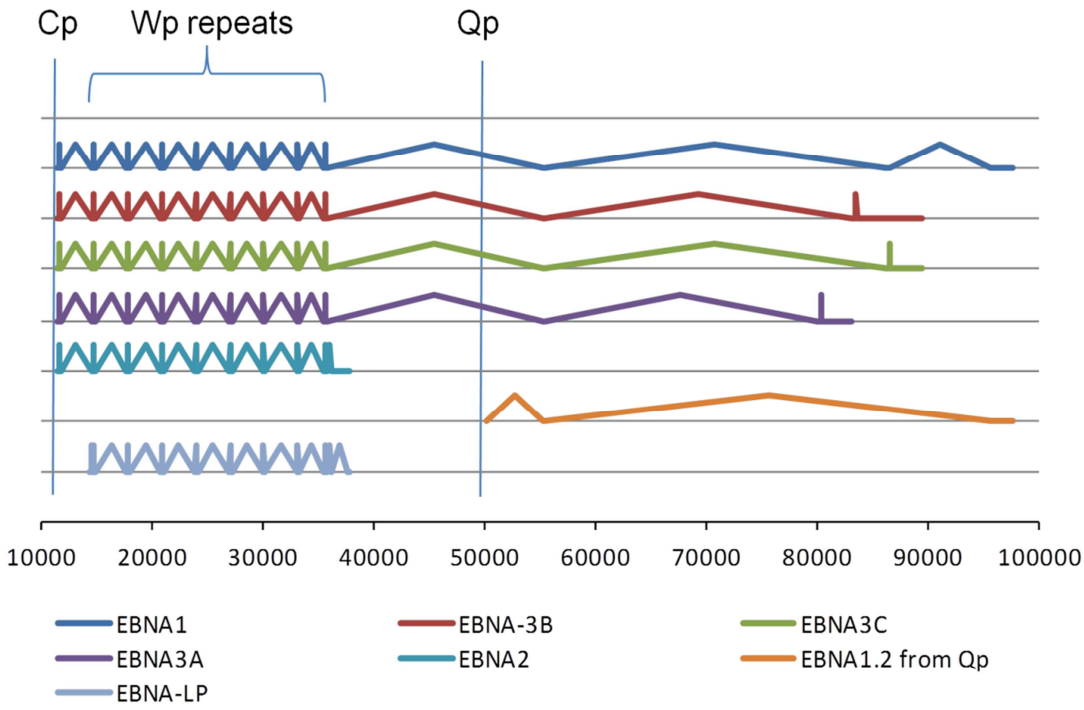


Figure 1.2 Latent gene organisation.

The differential splicing of the latency genes is represented here. Splices are indicated by deviation from the grey gridline, intended to represent the EBV genome DNA. Exons are positioned where the splice line touches the grey gridline. All the latency genes can be expressed from Cp or any Wp, but only EBNA1 can be expressed from Qp. Data from Req_Seq entry NC_007605.1 (Farrell, 2005).

Viral genes are categorised with regard to what stage during the EBV life cycle they are expressed. This is practically tested according to when stable RNA can be detected by northern blot. Latent genes are expressed constitutively. Immediate early genes are defined as those expressed after lytic activation,

without the requirement of any new protein synthesis, as the mRNA transcripts can be detected in the presence of protein synthesis inhibitors. Early lytic genes cannot be detected in the presence of protein synthesis inhibitors, suggesting the requirement of the immediate early genes expression for their regulation. They are expressed in the presence of replication inhibitors, such as acyclovir. Late lytic genes can only be detected when replication is allowed to occur. (Rickinson and Kieff, 2001b). Immediate early genes are required for successful lytic activation and replication of the viral genome, activating a cascade of early and late lytic gene expression. The products of early genes have a variety of roles, arguably the most important being replication. Late genes encode structural proteins and those important for immune evasion (Odumade et al., 2011). Table 1.1 summarises the information available for each ORF from the NCBI entry Ref_Seq NC_007605, as well as the gene counterparts that have been identified in Herpes Simplex Virus (HHV-1) and Rhadinovirus (HHV-8).

gene start (bp)	**cycle phase**	**gene name**	**Information extracted from NCBI entry, Ref_Seq NC_007605**	**Core gene family HHV-1 counter-parts**	**Rhadino-virus genus counter-parts**
1691	late lytic	*BNRF1*	tegument protein, related to formylglycinamide		
6629	unknown	*EBER 1*	small non-coding RNA		
6956	unknown	*EBER 2*	small non-coding RNA		
9631	late lytic	*BCRF1*	vIL-10		
11305	latent	*EBNA-1*	latency protein, maintenance of episomal genome		
11305	latent	*EBNA-2*	latency protein, transcription factor		
11305	latent	*EBNA-3A*	latency protein		
11305	latent	*EBNA-3B/3C*	latency protein		

12541	unknown	*BWRF1*	originally BCRF2, questionable whether a protein function is encoded		
14352	latent	*EBNA-LP*	latency protein		
40529	early lytic	*BHLF1*	basic (arginine/histidine-rich) protein product, bind single-stranded DNA preferentially		
41471	early lytic	*BHRF1*	ncRNA ebv-miR-BHRF1-1, vbcl-2a, ncRNA ebv-miR-BHRF1-2, ncRNA ebv-miR-BHRF1-3		
44793	early lytic	*BFLF2*	role in egress of capsids from nucleus	UL31	ORF69
46236	unknown	*BFRF1A*	role in DNA packaging	UL33	ORF67A
46280	early lytic	*BFLF1*	role in nuclear localization of capsids and DNA	UL32	ORF68
46544	early lytic	*BFRF1*	capsid docking protein on nuclear lamina, role in egress of capsids from nucleus	UL34	ORF67
47520	early lytic	*BFRF2*	related to HHv-5 UL49; ORF66		
49219	late lytic	*BFRF3*	small capsid protein on hexon tips	UL35	ORF65
50134	latent	*EBNA1.2*	QUK transcript, starts from the Q promoter (Qp) in latency I		
59904	unknown	*BPLF1*	large tegument protein	UL36	ORF64
62729	late lytic	*BORF1*	capsid triplex protein	UL38	ORF62
63034	unknown	*BOLF1*	tegument protein	UL37	ORF63
63881	early lytic	*BORF2*	ribonucleotide reductase large subunit	UL39	ORF61
66516	early lytic	*BaRF1*	ribonucleotide reductase small subunit	UL40	ORF60
67552	early lytic	*BMRF1*	DNA polymerase processivity subunit, early antigen D	UL42	ORF59
68491	late lytic	*BMRF2*	multiple transmembrane protein	probable counter-part of HHV-1 UL43	ORF58
72068	early lytic	*BSLF2/BMLF1*	post-transcriptional regulator, SM; EB2; Mta	UL54	ORF57
74593	early lytic	*BSLF1*	primase subunit of helicase-primase complex	UL52	ORF56
74594	late lytic	*BSRF1*	tegument protein	UL51	ORF55
76219	late lytic	*BLRF1*	envelope glycoprotein gN	UL49A	ORF53
76231	early lytic	*BLLF3*	deoxyuridine triphosphatase	UL50	ORF54

76575	late lytic	*BLRF2*	p23, LR2, tegument protein		ORF52
77763	early lytic	*BLLF2*			
79904	late lytic	*BLLF1*	envelope glycoprotein gp350, binds to CD21, role in entry		K8.1, ORF51
89828	late lytic	*BZLF2*	envelope glycoprotein gp42, binds MHCII, role in entry, contains C-type lectin domain		
90943	early lytic	*BZLF1*	bZIP transcription factor, induces switch from latent to lytic infection, EB1, Zta		K8
92728	early lytic	*BRRF1*	Na, transcriptional activator		ORF49
93925	early lytic	*BRLF1*	transactivator, role in reactivation from latency, Rta		ORF50
93955	late lytic	*BRRF2*	probable tegument protein		ORF48
97617	late lytic	*BKRF2*	virion glycoprotein gL	UL1	ORF47
98065	unknown	*BKRF3*	uracil-DNA glycosylase	UL2	ORF46
98846	early lytic	*BKRF4*	probable tegument protein		ORF45
101588	late lytic	*BBRF1*	portal protein	UL6	ORF43
101971	early lytic	*BBLF4*	helicase subunit of helicase-primase complex	UL5	ORF44
103660	late lytic	*BBRF2*	tegument protein	UL7	ORF42
106693	late lytic	*BBRF3*	envelope glycoprotein gM	UL10	ORF39
106750	early lytic	*BBLF2/BBLF3*	subunit of helicase-primase complex	UL8	ORF40
109043	late lytic	*BBLF1*	myristylated tegument protein anchored in envelope	UL11	ORF38
110053	early lytic	*BGLF5*	deoxyribonuclease	UL12	ORF37
111326	early lytic	*BGLF4*	serine-threonine protein kinase	UL13	ORF36
111829	unknown	*BGLF3.5*	tegument protein	UL14	ORF35
112650	late lytic	*BGRF1/BDRF1*	putative ATPase subunit of terminase	UL15	ORF29
112825	unknown	*BGLF3*	related to HHV-5 UL95; ORF34		ORF34
114641	late lytic	*BGLF2*	tegument protein	UL16	ORF33
116144	late lytic	*BGLF1*	tegument protein, role in DNA packaging	UL17	ORF32
116766	early lytic	*BDLF4*	related to HHV-5 UL92		ORF31
117016	unknown	*BDLF3.5*	related to HHV-5 UL91		ORF30
118816	late lytic	*BDLF3*	membrane glycoprotein gp150		ORF28
120188	late lytic	*BDLF2*	membrane protein, interacts with BMRF2		ORF27

121098	late lytic	*BDLF1*	capsid triplex protein	UL18	ORF26
125073	early lytic	*BcRF1.2*	related to HHV-5 UL87		ORF24
125422	late lytic	*BcLF1*	major capsid protein	UL19	ORF25
127416	late lytic	*BTRF1*	tegument protein	UL21	ORF23
131021	late lytic	*BXLF2*	envelope glycoprotein gH, gp85	UL22	ORF22
132571	late lytic	*BXRF1*	nuclear protein	UL24	ORF20
132846	early lytic	*BXLF1*	thymidine kinase	UL23	ORF21
133013	early lytic	*BVRF1*	putative portal-capping protein	UL25	ORF19
135432	late lytic	*BVRF2*	capside maturational protease, minor capside scaffold protein	UL26	ORF17
135627	unknown	*BVLF1*	related to HHV-5 UL79		ORF18
136331	late lytic	*BdRF1*	major capside scaffold protein	UL26.5	ORF17.5
138282	late lytic	*BILF2*	contains Ig domain, membrane glycoprotein		
138352	unknown	*RPMS1*	BART mRNAs, of which RMPS1 and part of A73 are included as representatives, are complex spliced transcripts whose significance is not understood		
143711	early lytic	*LF3*	questionable whether a protein function is encoded		
150417	unknown	*LF2*	derived from herpes virus dUTPase		ORF11
151694	unknown	*LF1*	derived from herpes virus dUTPase		ORF10
152641	unknown	*BILF1*	G protein-coupled receptor		
155549	unknown	*A73*	BART mRNAs, of which RMPS1 and part of A73 are included as representatives, are complex spliced transcripts whose significance is not understood		
156288	early lytic	*BALF5*	DNA polymerase catalytic subunit	UL30	ORF9
158864	late lytic	*BALF4*	envelope glycoprotein gB	UL27	ORF8
159121	unknown	*BARF0*	it is questionable whether a protein function is encoded		
160908	unknown	*BALF3*	subunit of terminase	UL28	ORF7
164356	early lytic	*BALF2*	single-stranded DNA-binding protein	UL29	ORF6
164984	early lytic	*BALF1*	vbcl-2b, similarity to Bcl-2 family proteins		
165008	early lytic	*BARF1*	similarity to CSF-1 receptor, secreted protein, contains Ig domain		
166011	latent	*LMP-2A*	multiple transmembrane protein, role in latency and		

			B cell survival		
166836	early lytic	*BNLF2b*			
167067	early lytic	*BNLF2a*	contains hydrophobic domain		
169088	latent	*LMP-1*	transmembrane protein expressed in latent infection, functional similarity to CD40, role in transformation		
169294	latent	*LMP-2B*	multiple transmembrane protein, role in latency and B cell survival, N-terminally truncated form of LMP-2A		

Table 1.1: The ORFs of EBV.

Annotated with the position in the Type 1 Ref_Seq NC_007605, information extracted from this NCBI entry, counterparts of genes in HHV-1 and HHV-8 (de Jesus et al., 2003, Baer et al., 1984), and gene expression classification (Farrell, 2005, Yuan et al., 2006).

1.4 EBV life cycle

The virus has a biphasic life cycle, with both latent and lytic phases. During latency, a restricted number of latent genes are expressed. These include *EBNA-1*, *EBNA-2, EBNA-3A, EBNA-3B, EBNA-3C, EBNA LP, LMP-1* (Latent Membrane Protein), *LMP-2A, LMP-2B, EBER1* (Epstein Barr viral small RNA), *EBER2* and BamHI A transcripts (BART). They are expressed at different times during the process of primary infection and the establishment of latency, to drive transformation and differentiation of the memory B cell. During the lytic phase, there is a dramatic change in expression, with new virus particles synthesised and shed into the saliva, where they can be transmitted to other individuals through close personal contact (Rickinson and Kieff, 2001b, Kutok and Wang, 2006). In developing countries, EBV antibodies are acquired early, due to exposure to the virus at a young age; however in developed countries only 50-70% of adolescents and young adults are seropositive (WHO, 2008). High standards of living can cause the individual to avoid contact with the virus until

adolescence or later life, and approximately 30 - 50% of the seronegative individuals will experience Infectious Mononucleosis (glandular fever) during primary infection (Steven, 1996). When primary infection occurs as an infant, the dose is relatively small and does not evoke the same aggressive response seen when exposure occurs as an adolescent or young adult, when the virus dose is believed to be higher (Crawford, 2001). The ability to establish latency allows the virus to evade immune responses and effectively remain hidden from the host immune system, resulting in a lifelong persistent infection.

1.4.1 Primary Infection

The infection and persistence of EBV is not yet fully understood, and controversy surrounds the initial mechanism of infection. Infection is summarised in Figure 1.3. EBV has been shown to replicate effectively in epithelial cells in culture and in oral hairy leukoplakia lesions, leading to the hypothesis that primary infection occurs via epithelial cells in the throat, where lytic replication amplifies the viral load, before circulating B lymphocytes in the bloodstream are infected. However, this has never been observed in a patient, perhaps due to the fleeting nature of initial infection and the delayed onset of symptoms. Primary infection may occur directly into circulating B cells in the bloodstream. The viral envelope protein gp350/220 interacts with B cell surface molecule CD21 (Fingeroth et al., 1984) to gain access to the B cell. Gp350 and gp220 are the unspliced and spliced products of the viral gene *BLLF1*. However, viruses that have been engineered not to express this protein can also infect B and epithelial cells, but at a reduced level, which indicates that other molecules mediate virus entry (Janz et al., 2000). The virus is then internalised via cytoplasmic vesicles (Nemerow and Cooper, 1984) and enters

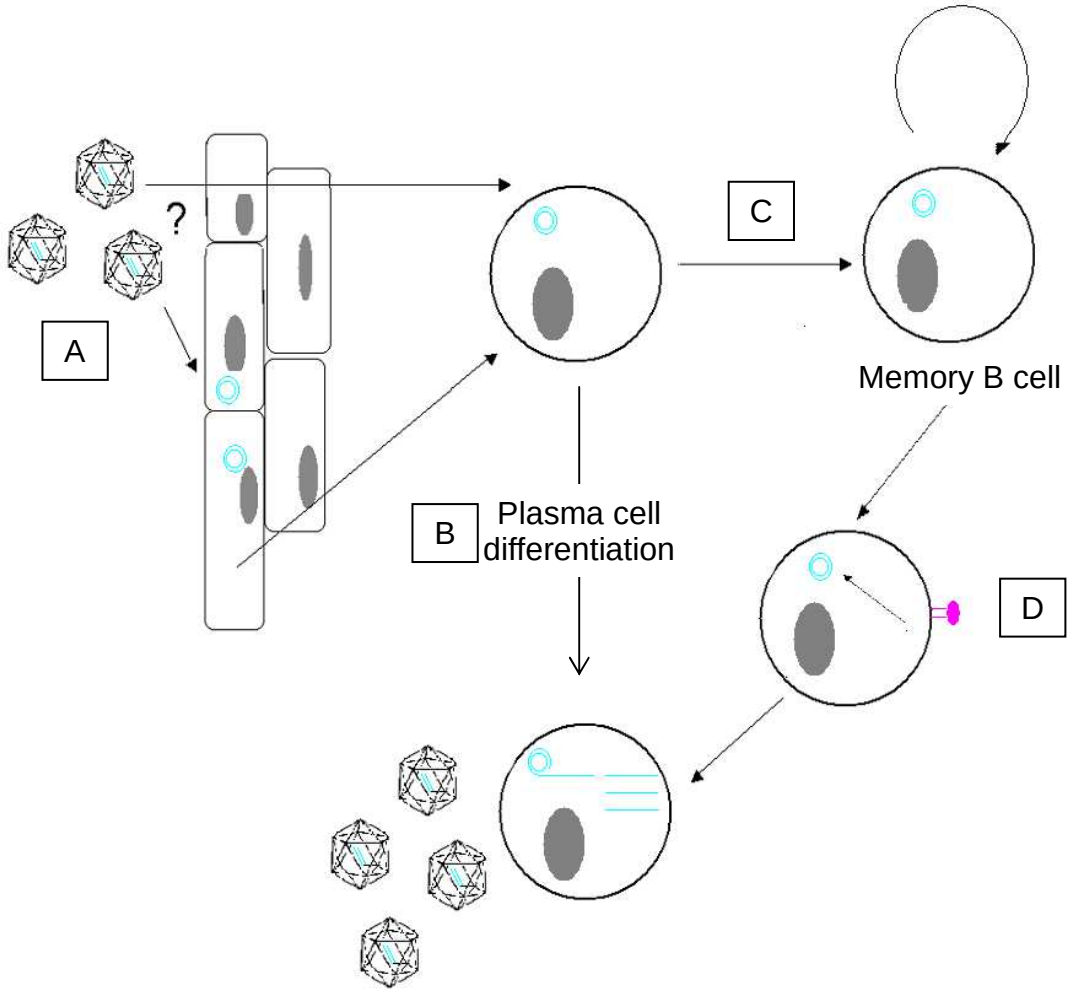


Figure 1.3 Epstein-Barr Virus life cycle.

Initial infection occurs via particles transferred from an infected individual, usually via saliva. The virus is thought to first infect epithelial cells, where the virus lytically produces viral particles, before infecting circulating B cells and establishing latency, however this is still subject to debate (A). EBV infected B cells behave in the same manner as antigen activated B cells, and differentiate in 2 different ways. Plasma cell differentiation (B) results in short lived plasmablasts with high antibody expresssion. This can activate EBV lytic cyle, and production of new virus particles. Alternatively, the B cell may migrate to germinal centres and become a memory B cell (C). When an antigen activates a memory B cell (D), lytic replication occurs, via Zta. The virus particles will then be released to infect other B cells or shed into the saliva to transmit to other people.

the latent phase, where there is no production of infectious virus particles, and a highly restricted number of latent genes are expressed. To achieve longevity within the host, as well as virus replication,, EBV exploits the normal B cell mechanisms. After activation by an antigen, B cells differentiate into plasmablasts, short-lived cells with a high expression of antibodies. This process has been shown to activate lytic cycle (Laichalk and Thorley Lawson, 2004). Alternatively, infected cells gain access to the long term memory B cell compartment, where the virus can remain undetected by the host immune system. Normally, a specific antigen activates a B cell, causing it to enter a follicle of a lymph node and proliferate to form germinal centres. These cells undergo somatic hypermutation, a process which causes alterations to the immunoglobulin genes and isotype switching. The cells are then selected for those whose new immunoglobulin genes produce higher affinity antibodies to the antigen in question. To fully mature into either short lived, antibody secreting plasma cells, or re-circulating memory B cells, two survival signals must be received, via the B Cell Receptor (BCR) and B cell surface co-receptor molecules, such as CD40, which engages with CD154 on T helper cells. The overall effect of these signals causes a cascade of cellular signalling pathways which drive proliferation of the B cell. CD40 achieves this by interacting with members of the tumour-necrosis-factor-receptor-associated factor (TRAF) family, causing activation of nuclear –factor κB (NFκB), and also with Janus activated kinase 3 (JAK3), to activate signal transducers and activators of transcription (STATs). The c-jun N-terminal kinase pathway is also engaged to activate the activator protein 1 (AP-1) transcriptional complex. The alpha and beta chains of the BCR contain immunoreceptor tyrosine-based activator motifs

(ITAMs) which are phosphorylated by Lyn, a member of the Src family of tyrosine kinases, which leads to recruitment of Syk tyrosine kinase, activating PLC-γ2, increasing intracellular levels of inositol- 1,4,5-triphosphate (IP_3) and diacylglycerol (DAG), which leads to an increase in calcium ions and activation of Protein Kinase C (PKC), resulting in survival signals. Without these signals, the cell will undergo apoptosis, and the surviving cells, primed for the specific antigen, re-circulate as memory B cells. It is suggested that EBV infected B cells also reach the germinal centre in the same manner, the required proliferation driven by the latency III growth program of latent gene expression where 9 viral genes and 2 viral RNAs, *EBER*s, are expressed. These proteins provide all the required signals without the need for any external activation. Sometime during this process, the expression of the latency genes becomes restricted to a latency II program, expressing *EBNA-1, LMP-1, LMP-2A* and the *EBER*s, turning off *EBNA-2* expression, which is known to block differentiation. *LMP-1* and *LMP-2A* are operationally similar to CD40 and BCR (Caldwell et al, 1998, Kilger et al, 1998), providing the survival signals required to save these cells from apoptosis, by driving immunoglobulin gene mutation and isotype switching. The pathways are not identical to normal B cell survival signals, but the end result is the same. *LMP-1* interacts with different members of the TRAF family to CD40, but achieves the same activation of NFκB, STATs and AP1. *LMP-2A* contains ITAMs in the amino terminal, which are phosphorylated in the same manner as BCR signalling, leading to the same activation of PKC via Syk and PLC-γ2, and survival signals (Caldwell et al, 1998). *LMP-1* downregulates the normal B cell signal for exit from the germinal centre, bcl-6, and so therefore must be differentially expressed to complete the germinal centre formation

(Kilger et al, 1998). All these factors allow the infected cells to mature into memory B cells, where all latent gene expression is switched off. (Reviewed by (Crawford, 2001, Amon and Farrell, 2005, Laichalk and Thorley-Lawson, 2005, Rickinson and Kieff, 2001b, Thorley-Lawson, 2005, Thorley-Lawson, 2001))

1.4.2 Latent cycle

At the point of latent persistence, all viral expression is shut down, with the possible exception of *LMP-2A*. This is termed latency 0. In this latent form of infection, it is only when the memory B cells divide that the viral *EBNA-1* protein is expressed, and in conjunction with the host cell DNA polymerase, will maintain the viral episome, via replication from a region called oriP, at the same time as the host chromosomes are replicated. *EBNA-1* is poorly recognised by the host immune system, ensuring that EBV remains undetected (Yin et al, 2003). In addition to *EBNA-1* and *LMP-2A, EBER-1, EBER-2* and the BamH1 A rightward transcripts (BART) are expressed. This is known as latency I.

As the growth promoting viral genes are not expressed, these cells are non-pathogenic and are also not recognised by the immune system, so therefore will not be eliminated. In this way, EBV achieves a balance between requiring proliferating activity to gain entry to the memory B cell compartment, but controlling this mechanism to achieve latent, undetected infection of the otherwise healthy host to ensure ongoing infection and propagation of the virus. This explains why EBV-associated tumours do not arise in all infected individuals, and that something must go wrong with the normal B cell biology to allow uncontrolled proliferation (Thorley-Lawson, 2005, Thorley-Lawson, 2001).

Resting B cells can be infected with EBV *in vitro* and form lymphoblastoid cell lines or LCLs. This ability exhibits the transforming potential of the virus, specifically the latent proteins, and results in a latency III gene expression profile, in which all latency genes are expressed. This resembles EBV associated tumours in immunocompromised hosts, such as post transplantation lymphoproliferative disease (PTLD), as it is the lack of an immune response which allows the full cohort of latent genes to be expressed without detection. It is more usual to observe a more restricted latency profile in EBV associated tumours. Cell lines isolated from different tumours exhibit different latency profiles, for example, Burkitt's lymphoma (such as the Akata cell line) expresses only *EBNA-1* from Qp, known as latency I, whilst B-lymphoproliferative disease expresses *EBNA-1, EBNA-2, EBNA-3A, EBNA-3B, EBNA-3C, LMP-1, LMP-2A* and *LMP-2B* initially from Wp, before switching to Cp after 24-48 hours, denoted latency III. A viral Bcl-2 homologue, *BHRF1*, is also expressed in latency III type infection, and is associated with anti-apoptotic pathways (Kelly et al, 2009). The latency II profile of gene expression is found in undifferentiated nasopharyngeal carcinoma, gastric carcinomas, Hodgkin's lymphomas, and T cell lymphomas, and consists of *LMP-1, LMP-2A* and *2B*, and Qp driven *EBNA-1*. The level of *LMP-2* expression varies between these conditions. The differential expression is thought to determine the behaviour of the cancerous cell. Non-dividing cells do not require *EBNA-1*, therefore they are classified as latency 0. This is seen in asymptomatic carriers of EBV persistence. (Reviewed by (Thorley-Lawson, 2001, Amon and Farrell, 2005, Young and Murray, 2003, Young and Rickinson, 2004, Rickinson and Kieff, 2001b, Rowe et al., 2009, Israel and Kenney, 2003, Crawford, 2001, Thorley-Lawson and Babcock, 1999, Thorley-Lawson, 2005)).

1.4.3 Lytic cycle

1.4.3.1 Activation of lytic cycle *in vivo*

The differentiation of B cells into plasma cells, activated by an antigen binding to the B cell receptor, and also the differentiation of epithelial cells, is thought to be the main mechanism for lytic activation in the host environment (Laichalk and Thorley-Lawson, 2005). Antigen activation of the B cell receptor of normal resting memory B-cells causes them to migrate to the lymph nodes and proliferate, to form a small secondary germinal centre, with production of antibody secreting plasma cells and the replenishment of the memory pool. Therefore, from time to time, EBV-carrying memory B cells will be activated, migrate and proliferate. This is required for the transmission of virus to new individuals.

The activation of the B cell receptor causes a cascade of processes within the cell that lead to immediate early viral gene expression, specifically *BZLF1* and *BRLF1*, the activation of phosphatidylinositol 3-kinase (PI3K), Ras family GTPases, and phospholipase C gamma 2 (PLC) which leads to the activation of PKC and calcium dependent factors (Kenney, 2007). All of this contributes to the drastic changes that occur to the transcription pattern of the EBV genome, where up to 62 lytic promoters are activated. The EBV genome is silenced during latency by repressive chromatin structure, caused by specific histone modification and DNA methylation at CpG dinucleotides (Kenney, 2007). The silencing of Zp and Rp is thought to be controlled in the same manner. However, DNA reporter constructs of these promoters lacking any chromatin structure are still inactive when transfected into EBV negative cell lines,

suggesting the existence of a positive factor, absent in unstimulated cells, which is required for activation (Kenney, 2007).

1.4.3.2 The immediate early genes

BZLF1 and *BRLF1* are classified as immediate early genes, and are the first viral genes expressed upon lytic cycle activation. They are required for complete viral reactivation and replication, and act as transcription factors, activating a cascade of lytic gene expression. *BZLF1* (the protein product of which is known as Zta, BZLF1, EB1, ZEBRA or Z) encodes a DNA binding protein that exists as a dimer, and binds to the promoters of both viral and host genes, as well as to the viral origin of lytic replication, directly via sequence-specific motifs within the DNA known as Zta Response Elements (ZREs). *BRLF1* (the protein product of which is known as Rta, BRLF1 or R) can also dimerise and bind to DNA. Both also interact with proteins to modulate their function, activate the transcription of the other, and cooperate to induce approximately 25 early genes. Zta can also autoactivate its own promoter. (reviewed in (Sinclair, 2003, Kenney, 2007, Amon and Farrell, 2005)).

1.4.3.3 Immediate early gene organisation

The genes coding for the two immediate early genes, *BZLF1* and *BRLF1*, are found close together in a region spanning approximately 5kb in the EBV genome (Figure 1.4). Another early lytic gene, *BRRF1,* is also found in this region, but is transcribed from the opposite strand of DNA to *BZLF1* and *BRLF1*. The mRNA message from Rp is bicistronic, containing the regions required for the expression of both Rta and Zta (Manet et al., 1989), and translation of the BZLF1 region of this transcript has been seen *in vivo.* The

majority of BZLF1 expression is derived from the Zp (Kenney, 2007). RAZ, the product of a spliced mRNA initiated from Rp consisting of part of *BRLF1* and *BZLF1*, is expressed later in infection (Furnari et al., 1994). It is thought to play a role in negative regulation of the transcriptional function of Zta; however the manner in which this is achieved is not clear. Small amounts of RAZ inhibit Zta despite not reaching a level required for effective dimerisation, and a mutated version which was incapable of dimerisation resulted in equal repression of Zta mediated transcription (Segouffin et al., 1996).

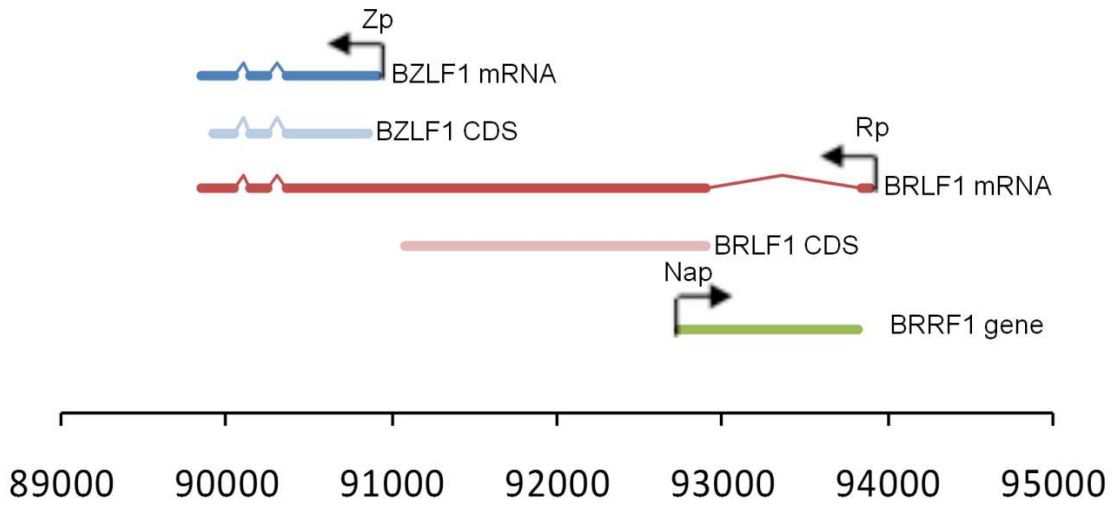


Figure 1.4 Immediate early gene organisation.

The region of the EBV genome encompassing the early immediate genes is represented here, to illustrate the mRNA splicing that *BZLF1* and *BRLF1* undergo. Coding sequence (CDS) is also shown below the appropriate mRNA, where data was available. *BRRF1* is expressed from the opposite strand to *BRLF1.*

1.4.3.4 *BRLF1* and *BZLF1* promoter control

The promoter regions of the immediate early viral genes provide response elements for a plethora of host DNA binding factors and transcriptional effectors (1.5). Through these, the host cell suppresses expression, but that suppression is also overturned upon lytic activation to allow expression of these gene

products, Zta and Rta. Five types of cis acting elements have been identified in the Zp, ZI – ZV; ZI and ZII-like elements have been found in the Rp, but have not yet been fully characterised (Kenney, 2007).

ZI elements are AT rich and have been shown in Zp to have both positive and negative regulatory roles, due to the binding of MEF-2D. In the phosphorylated form, MEF-2D interacts with histone deacetylating complexes, which cause transcriptional silencing of the surrounding area. Upon B cell receptor activation, MEF-2D becomes dephosphorylated via the activation of calcium dependent factors via PLC, which allows it to recruit histone acetyl transferase, and relieve any repressive chromatin environment, allowing expression (Gruffat et al., 2002). MEF-2D has been shown to bind to ZIA, ZIB and ZID elements. Sp1 and Sp3, transcription factors, have been shown to bind to ZIA, C and D weakly, and are not thought to bind simultaneously with MEF-2D (Speck et al., 1997). Sp1 and Sp3 are required for Rta autoactivation of Rp (Ragoczy and Miller, 2001). Another transcription factor, Smubp-2, can bind to ZI elements and has been shown to negatively regulate the associated gene expression, by disrupting stable transcription initiation machinery complex formation (Zhang et al., 1999).

ZII elements are atypical cAMP response element binding factor (CREB) response elements (CRE) that allow binding of transcription factors including CREB, ATF-1, ATF-2, c-jun, and c-fos. These transcription factors are constitutively expressed in most cells, but are activated by phosphorylation. This can occur as a result of B cell receptor activation, activating the MAPK pathway, which causes the phosphorylation of CREB/ATF transcription factors,

allowing them to bind to CRE sequences in the promoter and activate expression (Westphal et al, 1999). ZIIR overlaps ZII in Zp; this sequence has been shown to be utilised in the repression of Zp and maintaining latency, but the specific factor has not been discovered (Yu et al., 2011).

The ZIII elements (A and B) are Zta Response Elements (ZREs), and allow autoactivation of Zp. Activation of lytic cycle via anti-IgG treatment requires ZIIIA, indicating the requirement of Zta autoactivation in this process, or alternatively the presence of other cellular proteins binding this element before Zta. The importance of autoactivation in the context of the entire genome is debatable; exogenous expression of Zta has not been observed to lead to detectable levels of Zp initiated transcript (Speck et al., 1997, McDonald et al., 2010). There are 3 ZREs in Rp, and all 3 are important for activation of the gene (Chiu et al., 2007).

ZIV elements bind YY-1, a zinc finger protein associated with both positive and negative regulation of transcription, and are found in both Zp and Rp. A negative effect on expression is observed in Zp (Montalvo et al., 1995) and Rp (Zalani et al., 1997).

ZV and ZV' elements also act as negative regulators, allowing the binding of ZEB1 and ZEB2, cellular proteins which cause repression of BZLF1 expression, and lytic activation (Kraus et al., 2003). A mutated version of this sequence causes a viral genome that is defective in maintaining latency (Yu et al., 2007).

Other elements are also found in the promoter regions. Binding sites for Smad3/Smad4 allow activation of Zp and lytic cycle via the TGFβ cytokine

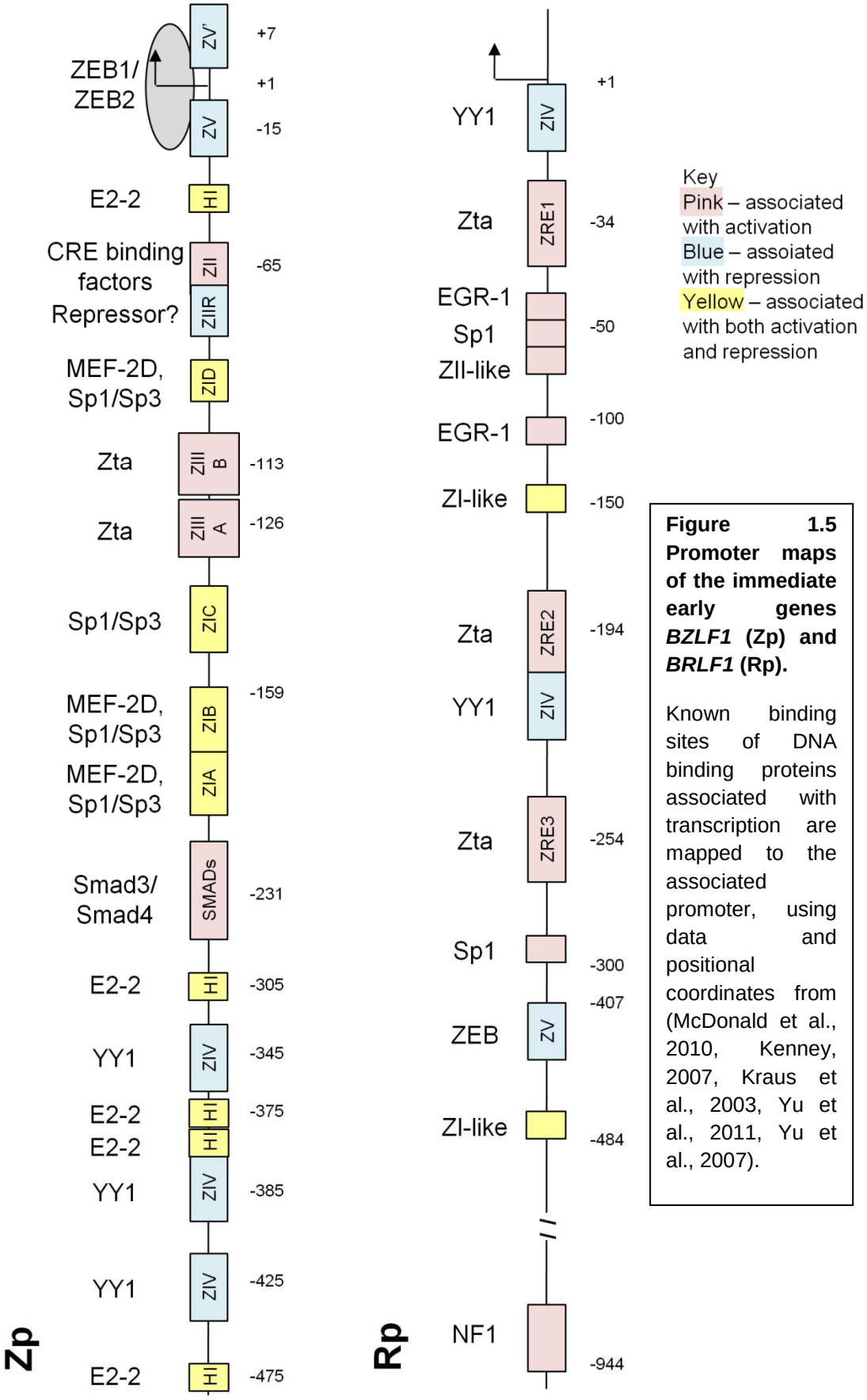


Figure 1.5 Promoter maps of the immediate early genes *BZLF1* (Zp) and *BRLF1* (Rp).

Known binding sites of DNA binding proteins associated with transcription are mapped to the associated promoter, using data and positional coordinates from (McDonald et al., 2010, Kenney, 2007, Kraus et al., 2003, Yu et al., 2011, Yu et al., 2007).

activation pathway (Kenney, 2007). H box elements in Zp bind transcription factor E2-2, which plays a repressive role in B cells but contributes to increased lytic activation in epithelial cells (Thomas et al., 2003). Rp also contains 2 Egr1 binding motifs, and the activation of lytic cycle via 12-0-tetradecanoyl phorbol-13-acetate (TPA) treatment requires these sites (Kenney, 2007). An element further upstream in Rp for Nuclear Factor 1 (NF1) was discovered to act as a positive regulator in HeLa cells, but not in lymphoid cells (Glaser et al., 1998).

1.4.3.5 Activation of lytic cycle *in vitro*

B cell receptor activation can be mimicked in cell culture *in vitro* systems by treating the cells with an antibody against human IgG or IgM, causing crosslinking of the B cell receptors leading to successful activation of lytic cycle, particularly in Akata Burkitt's lymphoma cells (Takada, 1984).

Other treatments have also been shown to activate lytic cycle. 12-0-tetradecanoyl phorbol-13-acetate (TPA) is a phorbol ester and activates lytic cycle via activation of Protein Kinase C (PKC). PKC is usually activated by diacylglycerol. Phorbol esters bind to PKC with a greater affinity, and also increase affinity for Ca^{2+}. The activation of PKC is usually controlled by the degradation of the diacylglycerol, a process which is absent with phorbol esters, leading to extended periods of activation (Castagna et al., 1982). The PKC pathway activates the expression of transcription factors such as Egr1, MEF2D and CREB, which bind to response elements in the promoters of the early immediate gene *BZLF1* and lead to the expression of Zta (Flemington and Speck, 1990b).

Sodium butyrate is a histone deacetylase (HDAC) inhibitor, and is thought to activate expression by allowing the accumulation of acetylation on lysine residues of histone tails, causing a decrease in the affinity for DNA and creating an open chromatin structure, permissive for transcription. Whether this alone is sufficient for full lytic activation is debatable (Countryman et al., 2008).

Not all cell lines are activated by all of the mechanisms detailed above. For example, Raji cells require both TPA and sodium butyrate, and the B-95.8 cell line is not responsive to sodium butyrate (Gradoville et al., 2002).

There is evidence that severe host cell stress response, such as that experienced during irradiation (Westphal et al., 2000) or chemotherapy (Feng et al., 2002) can induce EBV lytic cycle, which in conjunction with nucleoside analogues ganciclovir or 3'-azido-3'-deoxythymidine (AZT) promotes cell death.

5-azacytidine is a nucleoside analogue that is incorporated into DNA but cannot be methylated. Treatment of EBV positive human lymphoid cell lines with 5-azacytidine has been shown to activate lytic cycle. This led to the hypothesis that DNA methylation plays a role in maintaining viral latency (Ben-Sasson and Klein, 1981). An attempt to treat Nasopharyngeal Carcinoma (NPC) patients with 5-azacytidine, was not successful in activating lytic cycle, despite demonstrating demethylation of the Cp (Chan, Tao et al. 2004). This suggests that the removal of methylation marks is not sufficient for lytic activation NPC, or that demethylation did not occur at the appropriate region of the genome for lytic activation. Another explanation, suggested by recent data indicating no change in the methylation status of the genome by 5-azacytidine treatment, is

that the activation of lytic cycle is progressing in a novel mechanism (Miller et al., 2007), which may require cell specific factors.

The transfection and expression of *BZLF1* is also a successful method of inducing lytic cycle in some cell types, and has been delivered via an adenovirus vector in Jijoye cells, and severe combined immunodeficient (SCID) mice harbouring Jijoye cell tumours (Westphal et al., 1999).

Lytic induction *in vitro* is rarely successful in higher than 50% of cells, despite the use of multiple inducing agents (Kenney, 2007). This varies between cell line types, and could be due to viral or cellular factors. Expression of LMP2A, which inhibits B cell receptor activation (Miller et al., 1994) could play a role, and changes to the chromatin structure of Rp and Zp could also be responsible (Paulson and Speck, 1999). Complete resistance can indicate an integrated viral genome (Kenney, 2007). Deletion of essential viral replication components allow activation from latency into an abortive lytic reactivation, which would exhibit the expression of early lytic genes but not late, and therefore no new virus particles are produced, for example Raji cells (Taylor et al., 1989). Cellular factors that affect reactivation include NF-κB (Morrison et al., 2004) and retinoic acid (Sista et al., 1993), the receptor of which interacts with Zta and inhibits transcription. Nitric oxide has been shown to inhibit activation via inhibition of p38 phosphorylation and c-myc activation in epithelial cells (Gao et al., 2004).

1.4.3.6 Lytic Gene expression

Immediate early genes (and latent genes) contain introns, which allow them to be processed by the host cell machinery, causing splicing and associated transport out of the nucleus to allow expression in the cytoplasm. However,

most early lytic and late lytic genes contain no introns, so are not exported out of the nucleus in the usual way. The early lytic gene *BSLF2/BMLF1* encodes a protein known as SM, a nuclear RNA binding phosphoprotein. SM increases the stability of intronless mRNA, and exports it from the nucleus into the cytoplasm for translation. This function is required for lytic cycle (Swaminathan, 2005).

1.4.4 Viral genome Replication

In addition to Zta, there are six early lytic viral proteins required for replication of the EBV genome: *BALF2*, the single stranded DNA binding protein, *BALF5*, the catalytic subunit of DNA polymerase, *BBLF4*, helicase, *BSLF1*, primase, *BBLF2/BBLF3*, the primase associated protein, and *BMRF1*, the DNA polymerase processivity factor (Fixman et al., 1992). Homologues of the core six replication proteins were first identified in Herpes Simplex Virus 1 (HSV1) (Challberg, 1986), before the same techniques were implemented to discover the orthologues in EBV, and others herpesviruses, such as CMV (Pari and Anders, 1993, Pari et al., 1993) and KSHV (Wu et al., 2001). Many components of the core viral replication machinery can be swapped between members of the herpesvirus family; for example, EBV core replication proteins have been shown to replicate CMV, a beta-herpesvirus (Sarisky and Hayward, 1996), and herpesvirus papio DNA, a B lymphotrophic baboon virus with 40% overall homology to EBV (Ryon et al., 1993), and the KSHV core replication proteins have been shown to replicate the EBV genome (Wu et al., 2001). This ability highlights the similarities between these viruses in their mode of replication.

Replication initiation of herpesvirus genomes is proposed to adhere to a theta model, in which replication occurs from one or more orilyts, in a bidirectional

manner, due to the circular nature of the genome, and therefore has similarities to prokaryotic DNA replication. However, the DNA replication intermediates associated with a theta model have not been detected, nor has anyone managed to replicate this *in vitro*, casting doubt on this theory (Boehmer and Nimonkar, 2003). The early initiation events of replication, leading to strand unwinding of the lytic origins, have yet to be fully elucidated. HSV1 has the best characterised model, in which the lytic origin-binding protein of the family is the helicase, which co-ordinates with the single stranded binding protein to separate the strands, and allow initiation of replication. Zta is structurally different to the HSV1 helicase; however it is suggested that there is sequence homology in the region that confers the ability to bind single stranded DNA in the HSV1 helicase (Rennekamp and Liberman, 2010).

After replication has been initiated, the amplification of herpesvirus genomes is thought to progress via a rolling circle or sigma mechanism, generating long concatemers of the viral genome, which are then cleaved at terminal repeats and packaged into capsids (Boehmer and Nimonkar, 2003). However, viral DNA has been shown to accumulate exponentially, whereas the rolling circle model causes only linear amplification which suggests more than one mechanism of replication is involved (Rennekamp and Liberman, 2010).

1.4.4.1 Inhibition of replication

Acyclovir is a common drug used to inhibit lytic replication of EBV (and other herpesviruses), in both clinical settings and research. It is a nucleoside analogue that specifically inhibits viral DNA polymerase as a competitive inhibitor and a chain terminator of DNA replication. However, it must first be

activated by phosphorylation by the viral kinase, although it is unclear in EBV whether this is the protein kinase encoded by *BGLF4* or the tyrosine kinase encoded by *BXLF1* (Gershburg and Pagano, 2005). Ganciclovir and 3'-azido-3'-deoxythymidine (AZT) are additional nucleoside analogues that specifically target lytically active EBV infected cells.

1.5 EBV Associated Disease

EBV is associated with a large number of different diseases, including malignancies in B cells, epithelial cells and T cells linked to the proliferative properties of latency, disorders caused by an over-enthusiastic immune response to primary infection, and even associations with autoimmune diseases. Understanding the role EBV plays in each of these diseases is necessary to aid effective management and treatment. By assessing whether a tumour is mono-, oligo- or polyclonal with regards to the EBV genome (which can be detected by analysing the number of terminal repeat regions), conclusions can be drawn as to whether EBV infection is an early (mono- or oligoclonal) or late (polyclonal) event, and thus aid discovery of pathogenic mechanisms.

1.5.1 EBV and cancer

1.5.1.1 Burkitt's lymphoma

Burkitt's lymphoma is the prototypic EBV associated tumour, the most common childhood cancer in equatorial Africa. Fast growing tumours occur in the jaw, ovary, mammary gland, liver, intestine and kidneys, and are fatal if untreated. The disease is concentrated in this geographic area, classified as endemic BL, and occurs at only a low incidence worldwide, known as sporadic BL (Crawford, 2001). EBV DNA is present in tumour cells of 98% of endemic BL and in the

majority of cases exhibits a latency I profile, expressing only *EBNA-1* and the non-coding *EBER* RNAs (Allday, 2009). Some BL cases harbour a viral genome with a deletion over the *EBNA-2* gene, resulting in an expression profile consisting of all of the *EBNA*s except *EBNA-2* and the *LMP*s, but with the additional expression of *BHRF1*, a homologue of host Bcl-2, an anti-apoptotic factor. This is known as Wp-restricted latency, as all expression is activated from the Wp (Rowe et al., 2009). In contrast, EBV positive tumours account for only 15% of sporadic BL. This rises to 50% in individuals co-infected with HIV (Crawford, 2001). Both sporadic and endemic BL share a deregulation of the *Myc* oncogene on chromosome 8, as a result of a chromosomal translocation with one of three immunoglobulin genes (chromosomes 14, 22 or 2), which results in the active Ig locus activating constitutive expression of *Myc*. This causes an accumulation of Myc protein, a transcription factor with many targets including those that influence cell proliferation and apoptosis, and therefore contributes to the tumourogenic processes (Li et al., 2003). Infection with malaria appears to be an important cofactor for development of these tumours. Malaria suppresses the EBV specific T cells, causing high EBV viral loads in peripheral blood. It is also thought to increase germinal centre activity. This increase in activity increases the likelihood of the oncogenic translocation, due to the somatic hypermutation the cell is undergoing to rearrange the immunoglobulins creating an antigen specific response. As discussed in section 1.4.1, the B cells are usually selected for those that generate higher affinity antibodies, and must differentiate to survive, by receiving specific survival signals. However, due to the increased viral load, it is more likely that the B cell will be infected with EBV, which provides alternative survival signals (*LMP-1*

and *LMP-2A*) to evade cell death, and allow proliferation of a B cell with oncogenic potential (Rowe et al., 2009, Chene et al., 2009).

1.5.1.2 Hodgkin's Disease

Hodgkin's Disease (HD) or Hodgkin's lymphoma (HL) is a tumour of the lymph nodes that disrupts the recognised lymph node architecture (Kapatai and Murray, 2007). It originates from B lymphocytes. The tumour material is made up of mononuclear Hodgkin's cells and multinuclear Reed-Sternberg cells, with a reactive, non malignant cell infiltrate that dictates the exact type of disease; either nodular sclerosing (NS), mixed cellularity (MC) or the rarer lymphocyte depleted (LD) (Rickinson and Kieff, 2001a, Kutok and Wang, 2006). MC and LD have the highest rate of EBV association (60% to 90%) whilst NS has a lower association rate (20% to 40%). Monoclonal viral genomes were detected in tumour biopsies, indicating EBV infection is an early event in disease development (Rickinson and Kieff, 2001a, Young and Murray, 2003). Previous IM diagnosis increases the chance of suffering from EBV positive Hodgkin's Disease (Young and Murray, 2003). The cells exhibit a latency II profile of gene expression, with high expression of LMP1, under control of cellular transcription factor STAT3 (Rickinson and Kieff, 2001a, Young and Murray, 2003). *LMP-1* mimics normal CD40 B cell signalling, which is usually stimulated by interaction with T cells and promotes the differentiation of GC cells into memory B cells. This signalling causes constitutive NF-kB, JNK and p38 pathways activation, causing proliferation and aiding lymphomagenesis. *LMP-2* is also expressed and can also enhance survival of GC cells by recruiting src tyrosine kinases to mimic the normal antigen activation of the B cell receptor, via the PI3 pathway (Rickinson and Kieff, 2001a, Young and Murray, 2003).

1.5.1.3 Post Transplant lymphoproliferative Disorders (PTLD)

Patients that have undergone organ transplantation require heavy immunosuppression to ensure that the body does not reject the new organ (Munksgaard,2004). However, this comes with the risk of developing lymphomas, and these tumours are often associated with EBV (Maeda et al., 2009). Indeed, EBV seronegative individuals have a 20 times higher incidence of developing post transplant lymphoproliferative disorders (PTLD) than seropositive recipients. In the absence of appropriate levels of immunosurveillance in a post transplant patient, the balance between normal B cell differentiation and EBV-driven proliferation may be disrupted, leading to malignancies (Rickinson and Kieff, 2001a).

The PTLD tumours are of B cell origin and range from polyclonal B cell proliferations to aggressive monoclonal non-Hodgkin's lymphomas (NDL). Hence they have been categorised into 3 broad types: (i) diffuse B cell hyperplasias of poly or oligoclonal B cell origin which occur within 1 year of transplantation, without disturbing the normal lymph node structures; these are usually EBV positive, (ii) polymorphic lesions, exhibiting necrosis, disruption of lymphoid structure, and atypical nuclei, oligo or monoclonal in origin, which are almost always EBV positive and can appear early or late after transplantation, and (iii) monomorphic lesions, which arise several years after transplantation as diffuse, large B cell lymphomas which can be further classified into 2 subsets, those of immunoblastic origin, which the majority of which are EBV positive, or centroblastic, expressing a germinal centre B cell marker, of which a subset are EBV positive (Rickinson and Kieff, 2001a, Kutok and Wang, 2006). Overall, early onset lymphomas are 100% EBV positive, and late onset lymphomas are

80% EBV positive, and usually express a latency III profile, but other forms of latency have been detected (Young and Murray, 2003). It appears that the early onset tumours arise from EBV transformed B cells in the absence of a full T cell response, and in this way, *in vitro* transformed LCLs strongly resemble the expression profile (Rickinson and Kieff, 2001a).

1.5.1.4 T cell lymphoma

EBV rarely infects T cells, despite the presence of receptors on the cell surface which would allow virus entry (Kutok and Wang, 2006). However, virus is detected in some T cell lymphomas, which usually occur outside the lymph nodes (“extranodal”), and are known as NK/T cell lymphomas. The clonality of the EBV genome within the tumour has been demonstrated, confirming that EBV infection is an early event of tumourogenesis (Kutok and Wang, 2006). Virus is generally detected in between 5 and 50% of tumour cells, suggesting that infection may occur as a secondary event in oncogenesis. As the tumour progresses, the number of EBV positive cells increases, supporting the idea that EBV may confer growth or survival advantages to the transformed T cells (Young and Murray, 2003). Full details of the role of EBV in the pathogenesis of these tumours have not yet been elucidated (Kutok and Wang, 2006). T cell lymphomas are also associated with virus-associated hemophagocytic syndrome (VAHS), in which rare, monoclonal proliferations occur from mature T cells in patients suffering from chronic active EBV infection (Kutok and Wang, 2006). A latency I/II profile is expressed, and activation of tumour necrosis factor and other cytokines causes macrophage activation, leading to dangerous hemophagocytosis (Young and Murray, 2003, Kutok and Wang, 2006).

1.5.1.5 Nasopharyngeal carcinoma

Nasopharyngeal carcinoma (NPC) has been categorised by the World Health Organisation (WHO) into 3 types: Type I, Keratinising squamous cell carcinoma; Type II, Non-keratinising squamous cell carcinoma and Type III, Undifferentiated carcinoma. EBV is linked to type III, which is characterised by undifferentiated carcinoma cells surrounded by lymphocytic infiltrate, which is required for the growth of tumour cells (Young and Murray, 2003). The link was first suggested, based upon serological studies, in 1966, and was later confirmed by detection of EBV DNA in tumour cells in 1970 (Zur Hausen 1970). Monoclonality of EBV has been demonstrated in tumour cells, an indication of EBV infection before malignant cell proliferation (Shah and Young, 2009, Rickinson and Kieff, 2001a). The disease is most common in China and South East Asia, and both genetic and environmental factors have been implicated. There is a high familial risk of disease, which has not yet been fully mapped. It is suggested to reside in the HLA antigen locus on chromosome 6, a component of the major histocompatibility complex that aids the differentiation in the body of self and non-self and therefore plays an important role in immune recognition (Rickinson and Kieff, 2001a). Nitrosamine, found in salted fish, has been suggested as a dietary co-factor (Young and Murray, 2003). A latency I/II intermediate profile has been described for NPC, with Qp driven *EBNA-1* expression, *EBER*s, variable levels of *LMP-1* and high levels of spliced BARTs. Interestingly, a lytic gene product, BARF1, has also been detected, but no other lytic products can be reliably observed. However, antibody titers to lytic antigens are elevated in NPC patients, and this apparent inconsistency has not yet been explained (Rickinson and Kieff, 2001a). BARF1 protein is not detected in NPC

biopsies, due to secretion of BARF1 protein into the surrounding cell media (Seto et al., 2005).

1.5.1.6 Gastric carcinoma

The association of gastric carcinoma and EBV was first suggested due to similarities between gastric carcinoma and NPC (Deyrup, 2008). EBV is found in only a small proportion of diffuse or intestinal type gastric carcinomas (Young and Murray, 2003). The viral expression profile is more restricted than NPC, with *EBER*s, Qp *EBNA-1*, BARTs and *BARF1* expressed in all EBV associated gastric carcinomas, LMP2A expressed in approximately half, but no LMP1 expression (Rickinson and Kieff, 2001a). EBV associated gastric carcinomas are associated with a better prognosis when compared to other gastric carcinomas (Maeda et al., 2009). This may be due to a tumour-specific response directed against EBV antigens (Rickinson and Kieff, 2001a).

1.5.1.7 Breast cancer

The ubiquitous nature of EBV infection in the population can lead to controversy concerning the definition of an EBV associated cancer. Presence of the EBV genome or virus gene products in the tumour cell population should be unambiguously detected to allow such a conclusion to be drawn (Young and Murray, 2003). Unfortunately, this is not always easy to prove, and the debate over the aetiological status of EBV and breast cancer is a good example of this. EBV DNA (Labrecque et al., 1995, Luqmani and Shousha, 1995) and later *EBER-2* and *LMP-2* DNA (Bonnet et al., 1999) was first detected in whole tumour material in approximately 50% of cases. However, the presence of rare infiltrating lymphocytes into the tumour mass could account for this detection

(Murray, 2006), and therefore the evidence for an association required more specific evidence. Whole tumour tissues which were positive for EBV by PCR analysis were subjected to microdissection to isolate specific tumour cells from the surrounding non-cancerous tissue. The dissected tumour cells in this study were all negative for the presence of EBV (Murray et al., 2003). It has been demonstrated that EBV can infect breast cancer cell lines, exhibiting a latency II-like expression profile, with the additional detection of BZLF1, indicating evidence of viral replication (Huang et al 2003). However, this expression of *BZLF1* may be evidence of the abortive lytic cycle activation, or the burst of *BZLF1* expression seen upon primary infection (Halder et al., 2009).

Difficulty in detecting EBER expression, present in all states of latency, in the cancer tissue led to immunohistochemistry attempts to detect EBNA-1 (Grinstein et al., 2002). Unfortunately the antibody used (2B4) was shown to stain tissue where EBV DNA could not be detected (Murray et al., 2003), and later to cross react with another protein known to express in a wide variety of tumour types, MAGE4 (Hennard et al., 2006). Opinion remains divided on the EBV association with breast cancer, but it appears that even if an association is proven, it may be involved in only a small proportion of breast cancer.

1.5.2 Other EBV associated diseases

1.5.2.1 Infectious Mononucleosis

Infectious Mononucleosis (IM), also known as glandular fever, results from an aggressive CD8+ T cell response against EBV antigens and manifests as enlarged lymph glands, fever and malaise (Long et al., 2011). These symptoms can last between 1 and 6 months (Rea et al., 2001). The association with EBV was discovered when an EBV-seronegative technician in the Henle research

lab contracted IM, and was found to be EBV-seropositive after recovery, which led to further investigation (Rickinson and Kieff, 2001b). There is an incubation period of around 30 days between infection and onset of symptoms, during which time the EBV infected B cells circulate in the blood and infiltrate tissues (Crawford, 2001). IM is described as immunopathologic due to the cause of the symptoms residing with the activation of proinflamatory cytokines by the host immune system to primary infection of EBV. However, not all primary infection results in IM; indeed, primary infection in infants is usually asymptomatic, and it is usually EBV seronegative young adults who suffer from IM. A possible explanation for this is the difference in dose; young adults tend to contract EBV from kissing (hence the more familiar name, the kissing disease), with a potentially extended contact with the EBV carrier's saliva (Crawford, 2001). Other explanations for the aggressive immune response include changes related to age in the immune system. Existing memory B cells may be specific to other antigens that are similar to those detected during infection, causing cross reactivities, which may not be present in younger individuals (Rickinson and Kieff, 2001a).

1.5.2.2 Oral hairy leukoplakia

Oral hairy leukoplakia is an epithelial lesion on the tongue caused by uncontrolled EBV lytic infection in immunocompromised patients (Slots et al., 2006). It is seen as an early clinical indication that the immune system is failing, as it was initially observed in HIV patients, but has also been seen in post transplant patients and sometimes in healthy individuals (Rickinson and Kieff, 2001b). It appears to be as a result of repeated primary infection of the epithelial cells rather than reactivation of latent infection from memory B cells

(Sandvej et al., 1992). It is self-limiting but can be treated with viral replication inhibitors such as acyclovir, although patients are prone to recurrence when treatment is stopped (Walling et al., 2003).

1.5.2.3 Autoimmune diseases

Some autoimmune diseases are associated with chronic stimulation of the B cell system, which suggests possible interactions or association with EBV (Rickinson and Kieff, 2001a). Systemic lupus erythematosus (SLE) produces autoantibodies, which attack antigens within the body such as DNA and histones. Rheumatoid arthritis (RA) is a joint disease characterised by chronic inflammation, involving a number of different factors including genetics, hormonal, psychological, immune deregulation and environmental. Primary Sjögren's syndrome (pSS) consists of autoimmunity against the endocrine glands. Multiple sclerosis (MS) occurs when the myelin sheaths surrounding axons are damaged, causing a wide range of neurological symptoms (Toussirot and Roudier, 2008). Serological evidence has linked all these diseases to EBV, based upon co-occurrence, and an abnormally high EBV load has also been detected in autoimmune sufferers. It is difficult to say that the two are definitely connected; when a virus is as ubiquitous as EBV, the difference in sero-status between sufferers and non-sufferers is very small. For example, it was found that MS and IM had a similar distribution in the population; whether this indicates that late EBV infection causes a predisposition to MS, or if there is a common factor that increase susceptibility, is unclear (Pohl, 2009). Autoimmune patients have been documented to struggle to control EBV infection, and there is clinical evidence to suggest that EBV infection can aggravate the development of lupus (Toussirot and Roudier, 2008). Molecular mimicry may

also play a role in EBV associated autoimmune diseases; there are similarities between viral antigens and host proteins, which may cause cross-reactivities when an immune response is activated against the virus, causing harm to the host (Pohl, 2009, Toussirot and Roudier, 2008). EBV appears to possess all the qualities required of a candidate viral link for autoimmune disease; however the evidence is not yet completely convincing.

1.5.3 Lytic cycle activation as a treatment option

Activation of EBV lytic cycle provokes an immune response from the host. Where a cancer contains latent EBV, this mechanism could be exploited as a potential treatment. The transfection of a *BZLF1* expression vector into cells via an adenovirus was shown to activate lytic cycle and decrease tumour growth when it was trialled in nude mice (Westphal et al., 1999). However, the possibility remains that the increase in virus production may activate a disproportionate immune response, or cause increased infection of additional tissues. Simultaneous treatment with acyclovir is potential solution to this problem (Amon and Farrell, 2005). Efficient gene delivery methods also need to be developed before this can be considered as a legitimate treatment option (Israel and Kenney, 2003).

Other substances used to activate lytic cycle *in vitro (*as discussed in section 1.4.3.5), such as treatment with histone deacetylase inhibitors (for example sodium butyrate) are potential options, but associated side effects would need to be considered (Amon and Farrell, 2005). Despite the effectiveness of 5-azacytidine and sodium butyrate in Burkitt's lymphoma cell lines *in vitro,* these substances have not been demonstrated to effectively induce lytic cycle in *in*

vivo mouse models with tumours derived from lymphoblastoid cell lines (Israel and Kenney, 2003). However, irradiation and chemotherapy, treatments already used in patients with EBV-positive tumours, have been shown to induce lytic viral gene expression (Westphal et al., 2000, Feng et al., 2002). Augmenting this treatment with a nucleoside analogue such as ganciclovir, which requires activation by lytic-specific gene products, could promote tumour cell death (Israel and Kenney, 2003).

1.6 Epigenetics of EBV

Upon infection, the EBV genome is packaged in chromatin in a similar manner to cellular chromosome, and the establishment of latency correlates with these epigenetic marks (Tempera and Lieberman, 2010). Chromatin has 2 elements; histone octamers that make up nucleosomes, around which DNA is wrapped, and the methylation state of the DNA.

1.6.1 Histones

Histones are an important element of the chromatin structure, with 2 of each of the core histones, H2A, H2B, H3 and H4 creating an octamer known as a nucleosome, around which 147 base pairs of DNA wraps. This condenses the DNA into a chromatin fibre of 11nm diameter, which then further condenses with the addition of histone H1 to a 30nm diameter structure, which can condense still further during cell replication. Each of the histones has an unstructured N terminal tail, which can be modified by acetylation, methylation, phosphorylation and ubiquitination; each of these modifications affects the histone/DNA interaction and changes the chromatin state, either towards a more condensed state (heterochromatin) or a more open state (euchromatin).

The modifications also provide binding sites to recruit other proteins that affect the transcriptional activity of DNA, for example, proteins with bromodomains recognise histone acetylation, and those with chromo-like and PHD domains recognise histone methylation. Enzymes have been identified that add these modifications and others that remove them, indicating the dynamic nature of the structure (Kouzarides, 2007). Histone marks such as methylated H3K9, H3K27 and H4K20 are associated with silenced heterochromatin. H2K27 methylation recruits polycomb protein complex PC2, which associates with other transcription repressive proteins such as histone deacetylases and methyltransferases. Actively transcribed regions tend to exhibit high levels of acetylation, together with trimethylation of H3K4, H3K36 and H3K79 (Kouzarides, 2007). However, the association with a particular mark does not always mean a specific transcriptional outcome for a region, and some histone modifications have been shown to be strongly dependent upon context. For example, when H3K36 methylation was found in the coding region of a gene, the effect was activation, but when it was found in the promoter region, it was repressive (Vakoc et al., 2005). CCCTC-binding factor (CTCF), a transcription factor, has been shown to aid the formation of boundaries between differently transcribed regions of chromatin by recruiting specific chromatin remodelling enzymes (Kouzarides, 2007).

1.6.2 DNA methylation

Methylation of DNA in mammals occurs primarily on carbon 5 of the pyrimidine ring of cytosine residues in the context of a CpG dinucleotide. Methylation of other residues has been identified in plants and in early development in stem cells (Lister et al., 2009). Methylation patterns, established by *de novo*

methyltransferases, are heritable mitotically, maintained by methyltransferases that copy the methylation pattern to the complementary strand. Methylation patterns have been shown to be inherited across generations (through meiosis) in mice and plants, but this has not yet been shown in humans (Flanagan et al., 2010). CpG dinucleotides tend to cluster in CpG islands, first defined arbitrarily as regions approximately 200 basepairs in length with a G+C content of 50% and an observed/expected CpG ratio of 0.6 (Gardiner-Garden and Frommer, 1987). This was recently revised so as to exclude repeat regions that are not necessarily involved in direct control of expression, increasing the region length to greater than 500 basepairs, with a G+C content of 55% and an observed/expected CpG ratio of 0.65. This definition returned regions that were more likely to be associated with the 5' region of genes (Takai and Jones, 2002). CpG islands are found at the 5' end of approximately 60% of human genes (Wang and Leung, 2004) and usually remain unmethylated, whereas isolated CpG dinucleotides are more likely to be methylated. This is thought to be due to the high level of mutability of a methylated cytosine residue; spontaneous deamination of the methylated cytosine results in recognition as a T, leading to a reduction in the overall content of CpG during evolution. Methyl-CpGs mutations occur at 10-50 times the rate of a C in any other context (Walser and Furano, 2010). Those residues that are methylated will be selected against in this way, unless they are required for a vital role, such as transcriptional control.

Methylation causes transcriptional repression by attracting proteins with methyl binding sites, such as histone deacetylases and methyl transferases, which impose repressive chromatin structure upon the DNA (Jones and Baylin, 2007).

These proteins contain methyl binding domains (MBD), and crystal structures of the binding of these proteins to methylated CpG dinucleotides shows that they make multiple contacts with the methyl group, some of which are facilitated by water molecules in the major groove of DNA. There are also arginine residues that contact the G residue, confirming the CpG context required for interaction. Interaction only occurs on a double stranded symmetrically methylated CpG residue (Burkitt, 1985). The proteins containing MBDs recruit proteins that cause histone modifications associated with repression, for example Md2 is part of the Mi2/NuRD HDAC repressor complex, and plays a role in the silencing of *Xist* on the active X chromosome (Barr et al., 2007). Another MBD protein, MeCP2, associates with HDAC complexes and acts as a transcriptional repressor (Hutt and Burkitt, 1973).

Another way in which methylation contributes to transcriptional silencing is by directly interfering with the binding of some transcription factors (Jones and Baylin, 2007). C-Myc, a helix-loop-helix (HLH) DNA binding transcription factor, binds to sequences in the DNA known as Enhancer Box sequences, (E boxes) through dimerisation with another HLH protein, Max. When the central CpG within the E box was methylated, the basic region of Myc was unable to bind (Prendergast and Ziff, 1991). As c-myc is thought to regulate 15% of human promoters (Li et al., 2003), the effect of aberrant methylation could have wide-reaching effects. AP2 binding was inhibited by methylation in the promoter of the proenkaphalin gene, *PENK* (Comb and Goodman, 1990). The methylation of other transcription factor binding sites such as cAMP-responsive element (CRE) (Iguchi-Ariga and Schaffner, 1989), the binding sites for the E2F family of transcription factors (Kafuko and Burkitt, 1970) and those specific to Upstream

Binding Factor 1 (UBF1) (Burkitt, 1969) inhibit protein binding and activation of transcription. The effect of DNA methylation on the binding ability of the transcription factor Sp1, which binds to DNA through a zinc finger motif, displays a mixed phenotype. In some cases methylation appeared to inhibit binding (Douet et al., 2007, Clark et al., 1997), whilst in other cases, methylation had no effect on binding capabilities at all (Holler et al., 1988). The methylation status of surrounding CpG motifs was also shown to inhibit binding, leading to the suggestion that the effect was in fact due to the context in which the binding site was situated (Zhu et al., 2003). A similar scenario was observed in the promoter of TrkA, a nerve growth factor receptor, where specific methylation of the CpG motif within an AP1-like site did not disrupt c-jun binding (a bZIP transcription factor), but methylation of the entire oligonucleotide inhibited binding (Fujimoto et al., 2005).

1.6.3 EBV Epigenetics

H3 and H4 acetylation is increased around Cp in cell lines exhibiting a latency III profile, whereas in latency I, Qp is enriched for these marks and Cp is depleted. H3K9me3 is also detected at the W repeat and Wp during latency I. However it has not been conclusively shown that the switch between Cp and Qp driven *EBNA-1* expression is driven by the changes in chromatin structure (Alazard et al., 2003, Tempera and Lieberman, 2010).The majority of the EBV genome is populated with histone marks associated with the euchromatic state, with a relatively low amount of heterochromatic, H3K9 methylation. The OriP falls within a region of euchromatin (Tempera and Lieberman, 2010).

CTCF, an 11 Zn finger nuclear phosphoprotein, acts as an insulator of transcriptionally active regions of the genome (Phillips and Corces, 2009). It binds to unmethylated cytosines, and has been observed upstream of Qp, Cp and *EBER*, as well as downstream of Wp (Chau et al., 2006). This is thought to protect the neighbouring lytic genes from undue activation during latency. It may also regulate enhancer promoter interactions, as it has been show to cause DNA loop formation, and therefore play a part in long range regulation (Phillips and Corces, 2009). CTCF binding is increased in latency I compared to latency III, which is thought to inhibit EBNA-1 binding to the region between OriP and Cp, and hence repress expression of *EBNA-2* from Cp. CTCF binding was similar at Qp and the *EBER* promoter in latency I and III (Chau et al., 2006).

The EBV genome is extensively methylated in latently infected cells. Interestingly, from the promoter regions of 77 transcription start sites analysed using bisulphite sequencing to create a DNA methylome for the EBV genome, only five were not methylated; *EBER-1, EBER-2, Qp, LMP-2B/LMP-1* and *BZLF1* (Fernandez et al., 2009). The rest of the promoters analysed, including *BRLF1* promoter (Rp), was found to be methylated. The genome appeared to be almost completely unmethylated after viral replication has occurred, apart from *BBRF1* and *BBLF4* (Fernandez et al., 2009). Inhibition of DNA methylation relieves the repression of Cp in latency I, causing a shift to latency III expression (Tempera and Lieberman, 2010). As previously mentioned in section 1.5.3, 5'-azacitidine has been observed to activate lytic cycle, however no change in viral DNA methylation was detected, suggesting this reactivation occurred via a different, as yet unknown mechanism (Miller et al., 2007).

1.7 Zta

Crucial to the reactivation of the virus is a gene known as *BZLF1* that encodes a protein called Zta (also known as Z, BZLF1 and ZEBRA), a replication and transcription factor. This gene is silent during latency, but expressed within 30 minutes of lytic activation, and the presence of Zta is able to disrupt EBV latency (Sinclair, 2003). As a transcription factor, it binds to Zta Response Elements (ZREs) that resemble AP1 sites, in both viral and host (human) DNA. Through these ZREs, Zta activates transcription of both viral and human genes. Zta binds to ZREs in the *BZLF1* promoter (Zp) to create a positive feedback loop, as well as ZREs in lytic gene promoters. Rta activates expression from the Zp indirectly through ATF-2 and c-jun, cellular transcription factors that bind to a CRE motif (Adamson et al., 2000) as well as expression of other lytic genes. When neither *BZLF1* nor *BRLF1* are present, few lytic genes are subsequently transcribed. Zta belongs to the bZIP family of transcription factors, which includes other proteins such as yeast GCN4, c-Fos and c-Jun. As a replication factor, it interacts with the lytic origin replication, again via ZREs, to activate replication of the EBV genome itself (Burkitt, 1968). Zta can also influence cell cycle progression; enforced expression of Zta can induce cell cycle arrest in some cell types, via upregulation of C/EBPα (Wu et al., 2003).

1.7.1 Structure

Zta is 245 residues long, and exists as a dimer, containing a transactivation domain at the N terminus, a basic region that mediates DNA contact, the coiled-coil dimerisation domain, where the leucine zipper forms, also known as a bZIP domain, and a C terminal (CT) region. The bZIP region of Zta is different to the rest of the family; most bZIP proteins dimerise and bind DNA through an

approximately 60 amino acid bZIP domain, whereas Zta has a bZIP domain of 175 -220 amino acids long, and lacks a characteristic heptad repeat of leucine residues that are normally required for stable dimerisation and high affinity DNA binding (Farrell et al., 1989). Zta exhibits a variant of the bZIP fold, with a unique dimer interface and a substantial hydrophobic pocket, which allows Zta to recognise a wider range of DNA binding sites (Petosa et al., 2006). The crystal structure illustrates the dimerisation and DNA binding domain of Zta bound to DNA (Figure 1.7). The structure of the C terminal end is unknown, as it did not fully crystallise, and the N terminal transactivation domains are not shown, as they were not part of the protein used during crystallisation. It shows the protein contacting the DNA via two long bZIP helices, the basic region of each contacting the major groove of the DNA. At the C terminal end, an additional motif is formed; coiled-coil, hairpin turn, one turn helix (helix αC), then an extended stretch of residues, running antiparallel to the coiled-coil. This creates a small four helix bundle at the distal end of the homodimer. The C terminal region seems to stabilise the coiled-coil and enlarges the dimer interface, providing the extra stability required despite lack of leucine motif (Petosa et al., 2006). The CT region is required for lytic activation (Bailey et al., 2009). A cysteine at residue 189 has been shown to be required for the redox sensitive control of lytic cycle reactivation (Wang et al., 2005), and a serine residue at position 186 appears to be required for DNA binding (Bhende et al., 2005).

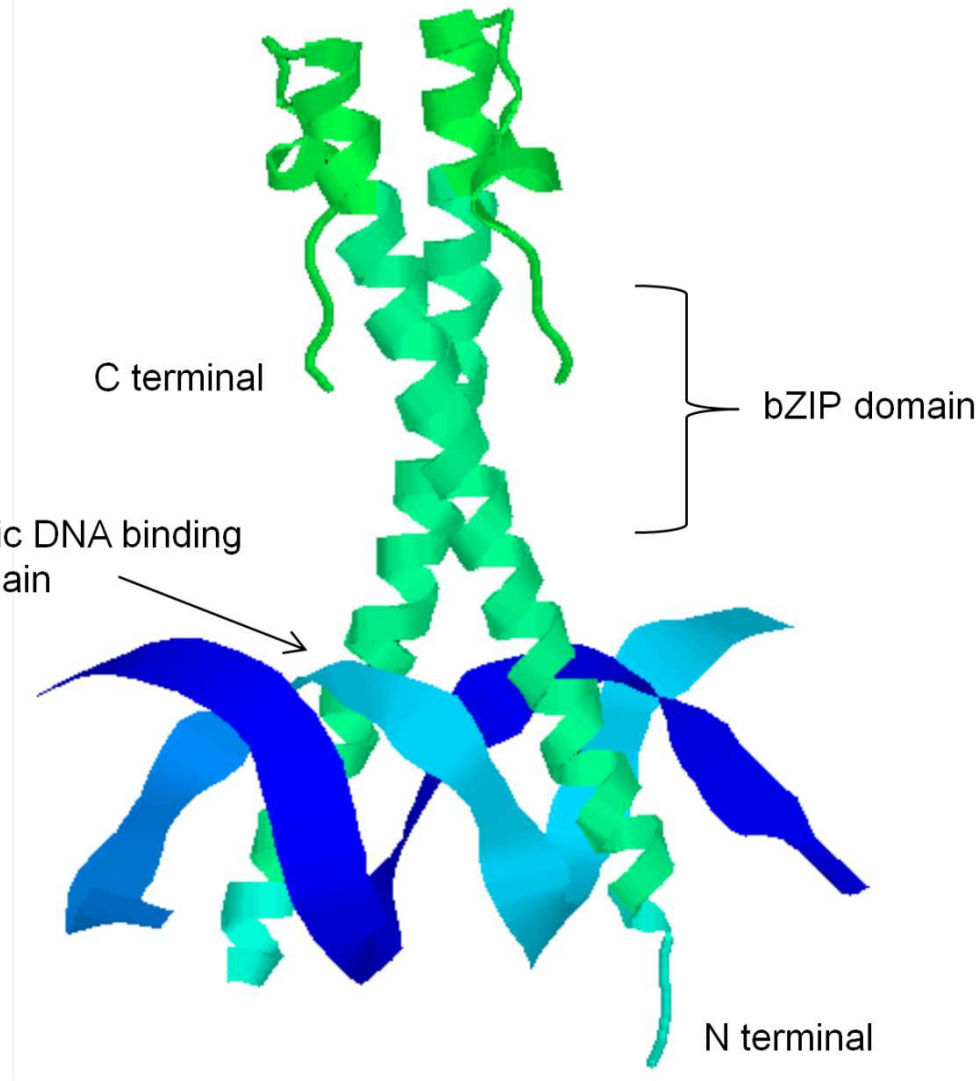


Figure 1.6 The crystal structure of Zta dimerisation and DNA binding domains.

This image was created using RASMOL and the published crystal structure data from Petosa et al, 2006. The light and dark blue represent DNA, the pale green shows the basic DNA contact regions, the bZIP helices, and the bright green shows the 4 helix bundle at the distal end of the dimer, including the start of the CT region. The very end of the CT region did not crystallise, therefore this part of the structure is not complete, and the globular transactivation domains are not included.

1.7.2 DNA interactions

Zta binds to ZREs in both the viral and host genome, and acts as a transcription factor to stabilise the complex bound to the DNA including TFIID and TFIIA (Chi et al., 1995). Binding has been established in the viral and human gene promoters detailed in table 1.2.

ZRE Core Sequence		Class	Names	References
Forward	**Reverse**			
TGAGCCA	TGGCTCA	I	Zp ZREIIIA	(Flemington and Speck, 1990a, Lieberman and Berk, 1990) (Urier et al., 1989)
			Rp ZRE1	(Sinclair et al., 1991)
TTAGCAA	TTGCTAA	I	Zp ZREIIIB	(Flemington and Speck, 1990a, Lieberman and Berk, 1990) 16
TGTGTAA	TTACACA	I	DSL ZRE1	(Lieberman and Berk, 1990)
TGAGCAA	TTGCTCA	I	DSL ZRE2	(Lieberman and Berk, 1990)
			DSL ZRE7	(Lieberman and Berk, 1990)
			BMRF1ZRE(-44)	(Taylor et al., 1991)
			BMRF1ZRE(-107)	(Quinlivan et al., 1993)
TGTGTCA	TGACACA	I	DSL ZRE3	(Lieberman and Berk, 1990)
			DSL ZRE4	(Lieberman and Berk, 1990)
			DSL ZRE6	(Lieberman and Berk, 1990)
			DHRS9 ZRE1	(Jones et al., 2007)
TGTGCAA	TTGCACA	I	DSL ZRE5	(Lieberman and Berk, 1990)
			CIITA (221)	(Li et al., 2009)

TGAGTCA	TGACTCA	I	BSLF2+BMLF1	(Taylor et al., 1991)
			BMRF1 AP-1	(Kenney et al., 1992) (Quinlivan et al., 1993)
			IL8 AP1	(Hsu et al., 2008)
			DHRS9 (ZRE2)	(Jones et al., 2007)
TGACTAA	TTAGTCA	I	Fp AP-1-Like Site	(Zetterberg et al., 2002)
TGAGTAA	TTACTCA	I	*CIITA*	(Li et al., 2009)
			IL13	(Tsai et al., 2009)
GTTGCAA	TTGCAAC	I	*IL-8 ZRE*	(Hsu et al., 2008)
GGAG**CG**A	T**CG**CTCC	III	Egr1	(Heather et al., 2009)
TGAG**CG**A	T**CG**CTCA	II	Rp ZRE2	(Sinclair et al., 1991) (Bhende et al., 2004)
TT**CGCG**A	T**CGCG**AA	III	Rp ZRE3	(Bhende et al., 2004)
TGAG**CG**T	A**CG**CTCA	III	Nap ZRE2	(Dickerson et al., 2009)
CGGG**CG**A	T**CG**CCCG	III	Nap ZRE1	(Dickerson et al., 2009)

Table 1.2 : ZRE sequences in the viral and host genome. CpG motifs are indicated in the ZREs in bold.

ZREs have been characterised in the viral promoter of *BRLF1* (Rp) (Bhende et al., 2004, Sinclair et al., 1991), *BMRF1* (the viral polymerase processivity factor) (Taylor et al., 1991, Kenney et al., 1992, Quinlivan et al., 1993), and the promoter of *BSLF2/BMLF1* (SM) (Taylor et al., 1991). These viral proteins are all required for successful lytic cycle activation. 2 interleukins have been shown to be upregulated by Zta; *IL-8* (Hsu et al., 2008) and *IL-13* (Tsai et al., 2009). Interleukins are cytokines which may play a role in tumour progression, due to their contribution to cell proliferation. Zta has also been observed to upregulate

DHRS9, an enzyme which plays a role in the conversion of retinol into the active form, retinoic acid, by directly binding to ZREs in the promoter (Jones et al., 2007). The majority of the ZREs discovered thus far have been shown to cause activation of expression, however the expression of one of these genes, *CIITA*, a component of the MHC Class II antigen presentation pathway, was actually downregulated in the presence of Zta binding. In this manner, Zta contributes to virus immune evasion (Li et al., 2009). Interestingly, both of the ZREs found in this promoter are also found in other promoters, one which is a match to a ZRE found in the viral OriLyt region, and the other found in the promoter of IL13, where it activates expression. This suggests that the final outcome of Zta binding to promoters is affected by the cellular context, perhaps the relative abundance of other proteins, or the modifications on any of these proteins.

Unlike other transcription factors, Zta has been shown to preferentially bind to methylated response elements and activate transcription of the associated gene (Bhende et al., 2004). This has major implications in understanding how Zta may reverse the epigenetic silencing imposed upon the EBV genome by the host cell. The ability of Zta to interact with methylated DNA is severely compromised if C189 or S186 are mutated. This has further implications, as C189 is highly conserved amongst bZIP proteins, suggesting that other transcription factors may have the same epigenetic reversal potential (Karlsson et al., 2008a). There are 3 ZREs in Rp, numbered 1 to 3 (figure 1.7A), and all 3 are important for activation of the gene (Chiu et al., 2007), although RpZRE1 appears to be unnecessary on a fully methylated template (Bhende et al., 2004).

The different behaviour of Zta with each of these ZREs led to the classification of the sites into three classes. RpZRE1 does not contain a CpG dinucleotide, so is therefore unaffected by methylation, and is classified as a Class I ZRE. RpZRE2 contains one CpG, and Zta has been shown to bind to it in the unmethylated state but to possess enhanced binding ability when the CpG is methylated, leading to this becoming classed as a Class II ZRE. RpZRE3 contains two CpGs, and Zta does not bind detectably to the unmethylated form, only the methylated form (Bhende et al., 2004). This is referred to as a Class III ZRE (Karlsson et al., 2008b). Each of the ZREs in table 1.2 are classified in this way.

Another example of a lytic gene promoter under the same mechanism of methylation dependent control was recently discovered (Dickerson et al., 2009). Nap, the promoter of the *BRRF1* gene, has been shown to contain two methylation dependent ZREs (figure 1.7B). Furthermore, Zta only activates Nap efficiently when it is in the methylated form (Dickerson et al., 2009). *BRRF1* encodes the protein Na, an early viral protein which induces c-jun phosphorylation and also plays a part in *BZLF1* transcription activation (Adamson et al., 2000). The Nap ZRE1 site requires both CpGs within the sequence to be methylated for effective binding. The Nap ZRE2 site contains only one CpG, and although it differs by only 1 base from the RpZRE2 site, Zta can only bind when it is methylated. Therefore both sites could be defined as class 3 ZREs. Both sites are required for activation of Nap *in vivo*. S186 is required to bind methylated Nap; however C189 is not (Dickerson et al., 2009), contrary to the behaviour of the C189 mutants in RpZRE3 (Karlsson et al., 2008a).

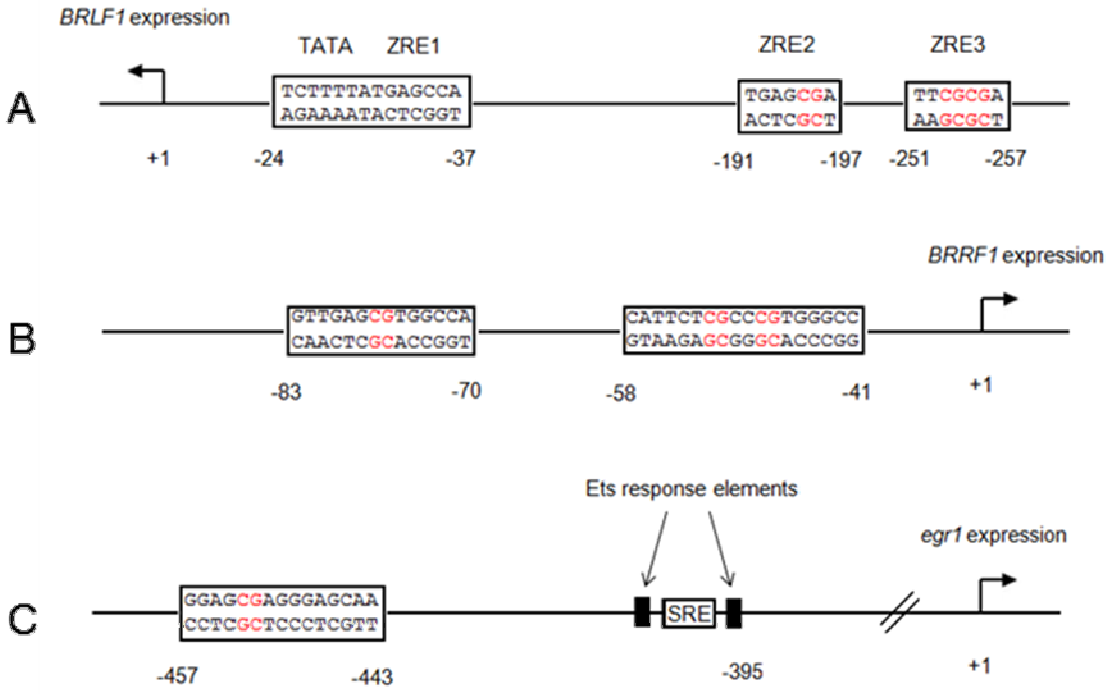


Figure 1.7 Gene promoters with CpG containing ZREs in the promoter region.

A: The *BRLF1* viral gene promoter. B: The *BRRF1* viral gene promoter. C: The *egr1* human gene promoter. The CpG dinucleotides that can be methylated are highlighted in red.

An example of methylation affecting Zta binding in human promoters has also been observed. *Egr1* encodes a human transcription factor which has been shown to be induced by Zta (Chang et al., 2006). 2 potential binding sites were identified (Chang et al., 2006) and mapped (Heather et al., 2009) to elucidate the mechanism of binding (1.7C). Interestingly, the distal site contains a CpG motif. According to one study (Seyfert et al., 1990) the promoter of *egr1* is methylated in B cells, however according to a more recent human B cell methylome (Rauch et al., 2009) is it not, suggesting that perhaps the methylation status of this area is unstable or subject to cell-specific mechanisms. Methylation of this distal site CpG enhanced Zta binding by at least 10 fold, compared to unmethylated binding, and can also activate expression *in vivo*, however this latter effect appears to be cell line specific (Heather et al., 2009).

1.8 Aims

To investigate the value of existing prediction tools for transcription factor (specifically Zta) binding sites.

To develop better tools for predicting ZREs.

To investigate the prevalence and position of ZREs within both the viral and the host genome.

To identify new Zta targets in both viral and human genes using existing methylome and transcriptome data.

Question the relevance of ZREs in the human genome and investigate host proteins that may share regulatory mechanisms with Zta.

Chapter 2. Materials and Methods

2.1 Materials

Reagent	Use	Supplier
Dimethyl sulfoxide (DMSO)	Cell culture	GIBCO
Penicillin Strep Glutamine	Cell culture	GIBCO
Fetal Calf Serum	Cell Culture	GIBCO
DMEM/RPMI	Cell culture	GIBCO
Anti IgG	EBV lytic reactivation in cell culture	Dako
Acyclovir	Inhibition of EBV replication	Sigma
EcoR1 10u/µl	restriction enzyme digest	Roche
Buffer H (10X)	restriction enzyme digest	Roche
Ethanol	Ethanol extraction of DNA	University Stores
Phenol:chloroform:isoamyl alcohol (125:24:1)	phenol-chloroform extraction of RNA	Promega
Isopropanol	phenol-chloroform extraction of RNA	Fisher
[^{35}S] Methionine	Radiolabelling in vitro translated proteins	Perkin Elmer

Stacking Buffer	protein gels	National Diagnostics
Protogel (30%) 37.5:1 Acrylamide:Bisacrylamide	protein gels	National Diagnostics
Protogel Buffer Stacking: 0.5M TrisHCl 0.4% SDS pH 6.8 Resolving: 1.5M Tris HCl 0.384% SDS pH 8.8	protein gels	National Diagnostics
10% w/v Ammonium Persulfate	protein and EMSA gels	Sigma
Tetramethylethylenediamine (TEMED)	protein and EMSA gels	Sigma
Sample buffer, Laemmli 2X concentrate	loading protein gels	Sigma
AccuGel (40%) 29:1 Acrylamide:Bisacrylamide	EMSA gels	National Diagnostics
Polynucleotide Kinase (PNK) 10u/µl	radiolabelling oligonucleotides	Roche
10X Polynucleotide Kinase	radiolabelling	Roche

(PNK) buffer	oligonucleotides	
[^{32}P] γ ATP 3000mCi/mmol	radiolabelling oligonucleotides	Perkin Elmer
G25 columns	purifying radiolabelled oligonucleotides	GE Healthcare
5X EMSA binding buffer: 50mM Tris HCl 2.5mM $MgCl_2$ 2.5mM EDTA 2.5mM DTT 250mM NaCl 0.25µg/µl PolydI:dC 20% glycerol pH 7.5	EMSAs	Promega
M.SssI Methyltransferase	*In vitro* DNA methylation	New England Biolabs
S-adenosylmethionine (SAM)	*In vitro* DNA methylation	New England Biolabs
NEBuffer 2 (10X) 50 mM NaCl 10 mM Tris-HCl 10 mM $MgCl_2$ 1 mM Dithiothreitol pH 7.9 @ 25°C	*In vitro* DNA methylation	New England Biolabs
0.1M DTT	EMSAs	Sigma

Gel loading solution, Type 1	EMSAs	Sigma
Protein A sepharose beads	Immunoprecipitation	Sigma
Protein G sepharose beads	Immunoprecipitation	Sigma
anti-Zta antibody sc-17503	Immunoprecipitation	Santa Cruz Biotechnology
SensiMix SYBR	QPCR	Bioline

Table 2.1: Reagents used and supplier

Solutions	Use	Composition
10X Tris-Glycine-SDS buffer	Running buffer for SDS PAGE protein gels	0.25M Tris base (Fisher) 1.92M Glycine (Fisher) 1% w/v SDS (Sigma) In water
Fixing solution	Fixing protein and EMSA gels	10% v/v acetic acid (Fisher) 20% v/v methanol (Fisher) In water
10X TBE buffer	Running buffer for DNA gels (1X) and EMSA gels (0.5X)	0.89M Tris base (Fisher) 0.89M Boric Acid (Acros Organics) 0.025M EDTA (BDH) In water

		pH 8.3
TE	Diluting and maintaining EMSA probes	0.01M Tris-HCl 0.001M EDTA (BDH) In water pH 7.5
Nuclear Extract buffers		
Solution A		10mM Hepes pH 7.9 10mM KCl 1.5mM $MgCl_2$ 0.1mM EGTA (chelating agent) 0.5mM DTT (reducing agent) 0.5mM PMSF (phenyl methane sulfoyl fluoride) (serine protease inhibitor)
Solution C		10mM Hepes pH 7.9 400mM NaCl 1.5mM $MgCl_2$ 0.1mM EGTA (chelating agent) 0.5mM PMSF (phenyl methane sulfoyl fluoride) (serine protease inhibitor) 5% Glycerol

ChIP buffers		
Cell lysis buffer		85mM KCl 0.5% NP-40 5mM PIPES pH 8.0 1mM PMSF 1 large protease inhibitor tablet per 50ml (Roche) In water
SDS lysis buffer		1% SDS 10 mM EDTA 50 mM Tris pH 8.0 1mM PMSF 1 large protease inhibitor tablet per 50ml (Roche) In water
IP Dilution Buffer		0.01% SDS 1.1% Triton X-100 1.2 mM EDTA 16.7 mM Tris pH 8.0 167 mM NaCl 1mM PMSF 1 large protease inhibitor tablet per 50ml (Roche)

		In water
Low salt wash buffer		0.1% SDS 1% Triton X-100 2 mM EDTA 20 mM Tris pH 8.0 150 mM NaCl In water
High salt wash buffer		0.1% SDS 1% Triton X-100 2 mM EDTA 20 mM Tris pH 8.0 500 mM NaCl In water
LiCl wash buffer		250 mM LiCl 1% NP-40 1% Na-deoxycholate 1 mM EDTA 10 mM Tris pH 8.0 In water
Elution buffer		0.01M Tris-HCl 0.005M EDTA (BDH) pH 7.5 1% SDS

		In water

Table 2.2: Solutions used, their compositions and suppliers.

Kit	**Use**	**Supplier**	**Components**
RiboMAX Large Scale RNA Production System SP6	*In vitro* transcription	Promega	Enzyme mix (RNA Polymerase, Recombinant Rnasin Ribonuclease Inhibitor and Recombinant Inorganic Pyrophosphatase) Transcription 5X buffer 100mM ATP, CTP, GTP, UTP RQ1 RNAse-Free DNAse 1u/µl 3M Sodium Acetate (pH 5.2) Nuclease-Free water
Wheat Germ Extract	*In vitro* translation	Promega	Wheat Germ Extract 1M Potassium Acetate 1mM Amino Acid Mix minus Methionine 1mM Amino Acid Mix minus Leucine 1mM Amino Acid Mix minus Cysteine

Rabbit Reticulocyte Extract	*In Vitro* translation	Promega	Rabbit Reticulocyte Extract Ribonuclease inhibitor (40U/µl) Nuclease-free water
PCR purification Kit	Purification of DNA from enzymatic reactions (post ChIP or methylation protocol)	Qiagen	PB, a high salt buffer PE, an ethanol containing wash buffer EB, a low salt elution buffer

Table 2.3: Kits used, and their components.

Vector	**Use**	**Reference**
pSP64-Zta	*In Vitro* transcription	(Hicks et al., 2003)
pSP64-Zta C189S	*In Vitro* transcription	(Schelcher et al., 2005)
pSP64-Zta C189A	*In Vitro* transcription	(Karlsson et al., 2008a)
pSP64-Zta S186A	*In Vitro* transcription	(Karlsson et al., 2008a)

Table 2.4: Expression vectors used and source.

Oligonucleotides were ordered from Invitrogen or Sigma. When indicated in the results text, specific CpG dinucleotides were methylated during synthesis by

Sigma. Alternatively, the oligonucleotides were methylated at all possible CpG dinucleotides using M.SssI methyltransferase, detailed in section 2.2.2.7.

Name	Sequence (5' - 3')	Chapter
RpZRE3 F	GTTTATAGCATCGCGAATTTTGAGTGC	4,5,6
RpZRE3 R	GCACTCAAAATTCGCGATGCTATAAAC	4,5,6
Z3CH8 F	GCTTCCCGGCTCGCGAAAGGGAGGACC	4
Z3CH8 R	GGTCCTCCCTTTCGCGAGCCGGGAAGC	4
HDAC2 F	TCCCCCACTGTCGCGAAGCTCCCGCCC	4,6
HDAC2 R	GGGCGGGAGCTTCGCGACAGTGGGGGA	4,6
XPC F	GGTGCGTCACTCGCGAAGTGGAATTTG	4
XPC R	CAAATTCCACTTCGCGAGTGACGCACC	4
MNT F	CCGCGGCGTCTCGCGAAGGGAGGGGCG	4
MNT R	CGCCCCTCCCTTCGCGAGACGCCGCGG	4
CyclinL2 F	GGGCGGCTCCTCGCGAAGCTCCACGGC	4,6
CyclinL2 R	GCCGTGGAGCTTCGCGAGGAGCCGCCC	4,6
CAPN2 F	CCGGGGAGGCTCGCGAATCGCGGTCCA	4
CAPN2 R	TGGACCGCGATTCGCGAGCCTCCCCGG	4

CDO1 F	CGTCCCAGCGTCGCGAACCACAGCGGC	4
CDO1 R	GCCGCTGTGGTTCGCGACGCTGGGACG	4
FALZ F	GGCGCGCAGCTCGCGAAATGCCCGGCG	4
FALZ R	CGCCGGGCATTTCGCGAGCTGCGCGCC	4
KIF1B F	GCTTCGGCCCTCGCGAAACTCCGCCCG	4
KIF1B R	CGGGCGGAGTTTCGCGAGGGCCGAAGC	4
LLGL1 F	TCGGCCGGGCTCGCGAAGGGACGCCCG	4
LLGL1 R	CGGGCGTCCCTTCGCGAGCCCGGCCGA	4
LMO4 F	GGATCCCGGGTCGCGAAGGGCAGCCCA	4
LMO4 R	TGGGCTGCCCTTCGCGACCCGGGATCC	4
MBD4 F	CCTCCTGCTCTTCGCGAACCGCCCCGC	4
MBD4 R	GCGGGGCGGTTCGCGAAGAGCAGGAGG	4
PLEKHJ1 F	AGCCGCTCCCTCGCGAAAGTTGGCCCC	4
PLEKHJ1 R	GGGGCCAACTTTCGCGAGGGAGCGGCT	4
PRKD1 F	CTTCCTGGGGTCGCGAACTTCCCGGGC	4
PRKD1 R	GCCCGGGAAGTTCGCGACCCCAGGAAG	4
SEC14L F	CGCCCGCTACTCGCGAAGCCCAGCCCG	4

SEC14L R	CGGGCTGGGCTTCGCGAGTAGCGGGCG	4
TADA3L F	GCTGCGCTTCTCGCGAAAGGGCAGGCA	4
TADA3L R	TGCCTGCCCTTTCGCGAGAAGCGCAGC	4
TOP2B F	CCGCGCCCCATCGCGAAGATCCGGAGC	4
TOP2B R	GCTCCGGATCTTCGCGATGGGGCGCGG	4
XPCLMNTR F	GGTGCGTCACTCGCGAAGGGAGGGGCG	4
XPCLMNTR R	CGCCCCTCCCTTCGCGAGTGACGCACC	4
MNTLXPCR F	CCGCGGCGTCTCGCGAAGTGGAATTTG	4
MNTLXPCR R	CAAATTCCACTTCGCGAGACGCCGCGG	4
XPC mut core F	GGTGCGTCACCCCCTTAGTGGAATTTG	4
XPC mut core R	CAAATTCCACTAAGGGGGTGACGCACC	4
XPC LM3 F	GGTCCGCCTCTCGCGAAGTGGAATTTG	4
XPC LM3 R	CAAATTCCACTTCGCGAGAGGCGGACC	4
XPC LM3 mut core F	GGTCCGCCTCCCCCTTAGTGGAATTTG	4
XPC LM3 mut core R	CAAATTCCACTAAGGGGGAGGCGGACC	4
AP1 mut F	GATCCACCCCTTAGAGGAAAACATACG	4

AP1 mut R	CGTATGTTTTCCTCTAAGGGGTGGATC	4
RpZRE3LCyclinL2R F	GTTTATAGCATCGCGAAGCTCCACGGC	4
RpZRE3LCyclinL2R R	GCCGTGGAGCTTCGCGATGCTATAAAC	4
CyclinL2LRpZRE3R F	GGGCGGCTCCTCGCGAATTTTGAGTGC	4
CyclinL2LRpZRE3R R	GCACTCAAAATTCGCGATGAGCCGCCC	4
ATF3 F	ATTAATAGCATTACGTCAGCCTGGGACT	4
ATF3 R	AGTCCCAGGCTGACGTAATGCTATTAAT	4
PENK F	GCGTAGGGCCTGCGTCAGCTGCAGCCC	4
PENK R	GGGCTGCAGCTGACGCAGGCCCTACGC	4
FN F	ACAGTCCCCCGTGACGTCACCCGGAGCCCG	4
FN R	CGGGCTCCGGGTGACGTCACGGGGGACTGT	4
CpG5 1F	CTCATAGGTCAGAGCGACATAGAGGCG	5
CpG5 1R	CGCCTCTATGTCGCTCTGACCTATGAG	5
CpG5 2F	AAGCGATGGCCGAGCGATGACTCGTGT	5
CpG5 2R	ACACGAGTCATCGCTCGGCCATCGCTT	5

CpG5 3F	GCCGCCGCACTCAGCGAGGAGGCCTGC	5
CpG5 3R	GCAGGCCTCCTCGCTGAGTGCGGCGGC	5
CpG5 4F	TCCAGATGACTGAGCGCACGGCCTCAA	5
CpG5 4R	TTGAGGCCGTGCGCTCAGTCATCTGGA	5
CpG5 5F	CTTTGCGCTCTGCGCGAGGACGAGCTC	5
CpG5 5R	GAGCTCGTCCTCGCGCAGAGCGCAAAG	5
CpG5 6F	TTTCAAGTCGTGGGCGAATTAACTGAG	5
CpG5 6R	CTCAGTTAATTCGCCCACGACTTGAAA	5
CpG5 7F	TTGGAAAACATTAGCGACATTTACCTG	5
CpG5 7R	CAGGTAAATGTCGCTAATGTTTTCCAA	5
CpG5 8F	GCAGGGCCCCCGCGCGATCTAGGTAGG	5
CpG5 8R	CCTACCTAGATCGCGCGGGGGCCCTGC	5
CpG5 9F	CCAATGTCTGCGTGCGAGCCGGGCTTG	5
CpG5 9R	CAAGCCCGGCTCGCACGCAGACATTGG	5
CpG5 10F	CGAGGAGGCCTGCGCGTGTTCCTCAAC	5
CpG5 10R	GTTGAGGAACACGCGCAGGCCTCCTCG	5
CpG5 11F	AGCAAGGTGCTGGGCGTGGACCGCGCG	5

CpG5 11R	CGCGCGGTCCACGCCCAGCACCTTGCT	5
CpG5 12F	TTTGGCGTCATGTGCGTCTGGATGACA	5
CpG5 12R	TGTCATCCAGACGCACATGACGCCAAA	5
CpG5 13F	CAGACTCTGGTTTGCGAGGCTGGGCGG	5
CpG5 13R	CCGCCCAGCCTCGCAAACCAGAGTCTG	5
CpG5 14F	AAAGATCTGGTTGGCGATCCGGTACAC	5
CpG5 14R	GTGTACCGGATCGCCAACCAGATCTTT	5
Zp (-96) F	TAAATTTAGGTGTGTCTATGAGGTACA	5
Zp (-96) R	TGTACCTCATAGACACACCTAAATTTA	5
Rp (-204) & RpZRE2 F	CATCTTGTCCTGTGATAAAATCGCTCATAAGCTTAGT	5
Rp (-204) & RpZRE2 R	ACTAAGCTTATGAGCGATTTTATCACAGGACAAGATG	5
RpZRE2 F	TGTGATAAAATCGCTCATAAGCTTAGT	5
RpZRE2 R	ACTAAGCTTATGAGCGATTTTATCACA	5
Rp (-439 & -447) F	ATGACTCGGGTGTGTCCTTGTGTGAGGTCTCACCTG	5
Rp (-439 & -447) R	CAGGTGAGACCTCACACAAGGACACACCCGAGTCAT	5

Rp (-114) F	CACTCATACTTAAGCGATGCTGATGCA	5
Rp (-114) R	TGCATCAGCATCGCTTAAGTATGAGTG	5
BMRF1p (-173) F	TGGGGGGTGGTGTGCCATACAAGGGAGC	5
BMRF1p (-173) R	GCTCCCTTGTATGGCACACCACCCCCCA	5
BMRF1p (-248) F	CCTTGGTGGATGTGCGAGCCATAAAGCA	5
BMRF1p (-248) R	TGCTTTATGGCTCGCACATCCACCAAGG	5
BMRF1p (-148) F	GTCATGTAGGTGAGCGGGCAGTCCTTG	5
BMRF1p (-148) R	CAAGGACTGCCCGCTCACCTACATGAC	5
RpZRE3 m2 F	GTTTATAGCATCGCGCTTTTTGAGTGC	6
RpZRE3 m2 R	GCACTCAAAAAGCGCGATGCTATAAAC	6
RpZRE3mutcore F	GTTTATAGCACCTTTTCTTTTGAGTGC	6
RpZRE3mutcore R	GCACTCAAAAGAAAAGGTGCTATAAAC	6

Table 2.5: Oligonucleotides used in EMSAs. The core ZRE binding sequence is underlined.

Name	Sequence (5' - 3')	Code	Absolute genomic position (Type 1)
BKRF4 F	CATTGCTCTCTGAGCGGTTA	SR 109	98591-98611
BKRF4 R	ACCAGATGCTTCTTGGAGTTG	SR 110	98688-98667
BGLF4 F	ACCGAGGCTCTTAGTTGCTG	SR 111	111611-111631
BGLF4 R	GTTGCGGACATGGTGACTTA	SR 112	111686-111666
BTRF1 F	AGCTACGCAATCGGAGTCA	SR 113	126981-127000
BTRF1 R	GGAGGCGCAGTCTAGCAG	SR 114	127052-127034
Region with no ZREs F	CCGCATGTCCAACCACCACG	SR 7	139875-139895
Region with no ZREs R	ATGCTACCTAGGCCTGCGTCC	SR 8	139997-139976

Table 2.6: QPCR primers.

2.2 Methods

2.2.1 Cell Culture

2.2.1.1 Maintenance of cell lines

Akata cells, an EBV positive Burkitt's lymphoma tumour cell line (B cells), are suspension cells and are maintained in RPMI (Invitrogen). HEK293 cells, a human kidney epithelial cell line, are adherent cells and are maintained in DMEM (Invitrogen). The media is supplemented with 100U/ml of penicillin, 100µg/ml of streptomycin and 2mM of L-glutamine, which was added from a pre-mixed solution PSG (Invitrogen). 10% (v/v) final volume Foetal Bovine Serum (FBS) (Invitrogen) was also added.

The EBV within Akata cells can be induced into lytic cycle by cross-linking the B cell receptors using anti-IgG at a final concentration of 0.125% (v/v).

Replication of viral genome can be inhibited by treatment with acyclovir. A 100mM stock of acyclovir was made in DMSO. Akata cells were concentrated to $2x10^6$ cells per ml of media, and acyclovir was added to a final concentration of 100µM. After 24 hours the cells were diluted to between $2\text{-}6x10^5$ cells per ml, and an appropriate amount of acyclovir was added to maintain the 100µM concentration. After 48 hours, the cells were harvested.

2.2.2 Nucleic Acids Methods

2.2.2.1 Vector linearisation

SP64 transcription vectors were prepared for use in the *in vitro* transcription assay by linearising with EcoRI restriction enzyme in the following manner:

Plasmid DNA	20µg

10X Buffer H	5µl
EcoR1	10µl
Water	up to 50µl

Incubate for 2 hours at 37°C.

2.2.2.2 Ethanol Precipitation

The linear DNA was purified by ethanol precipitation in the following manner.

0.1X the total volume of 3M Sodium Acetate (pH 5.2) and 2.5X the total volume of cold ethanol were added, before centrifugation at 13,000rpm for 20 minutes at 4°C.

The ethanol was replaced with 100µl 70% (v/v) ethanol. The solution was centrifuged at 13,000rpm for a further 10 minutes at 4°C, before the ethanol was removed, and the purified DNA resuspended in water to generate a final DNA concentration of approximately 1µg/µl.

DNA is a polar molecule due to the charged phosphate backbone, and is dissolved in water, a polar molecule. Ethanol is less polar than water, and therefore allows the DNA to form stable ionic bonds and precipitate out of solution. The presence of positive ions, such as sodium, aids this process. Centrifugation allows collection of the precipitated DNA. The pellet is washed to remove leftover salt from the precipitation procedure, and then centrifuged to collect the pellet before resuspending in pure water.

2.2.2.3 *In Vitro* Transcription

10µg linear DNA was used for RNA synthesis using a Ribomax Large Scale RNA Production System Kit.

Linear DNA	10µl (1µg/µl)
rNTPs (25mM ATP,CTP,GTP,UTP)	20µl
5X SP6 transcription buffer	20µl
Nuclease free water	40µl
SP6 RNA polymerase mix	10µl

Incubate for 2 to 4 hours at 37°C, to allow transcription of RNA from the linear DNA template.

10µl of RQ1 RNase free DNase (1u/µl) was added to the transcription mixture and incubated at 37°C for 15 minutes to digest the DNA template.

2.2.2.4 Phenol Chloroform Extraction

An equal volume of phenol:chloroform:isoamyl alcohol (125:24:1) was added to the mixture and vortexed for 1 minute, before centrifugation at 13,000rpm at room temperature for 2 minutes. Proteins are denatured by phenol and chloroform, whilst the nucleic acid remains in the aqueous phase. The upper aqueous phase was transferred to a fresh tube, and an equal volume of chloroform: isoamyl alcohol (24:1) was added and vortexed for 1 minute. This additional extraction is performed to reduce the loss of nucleic acids due to insoluble RNA:protein complexes that form at the interface between the organic and aqueous phases. This was centrifuged at 13,000rpm at room temperature

for 2 minutes. The upper aqueous phase was transferred to a fresh tube and 0.1X the total volume of 3M Sodium Acetate and 1 volume isopropanol was added. The sample was placed on ice for 5 minutes, to precipitate the nucleic acid, before centrifugation at 13,000rpm at room temperature for 10 minutes. The supernatant was removed and the RNA pellet was washed carefully with 1ml 70% (v/v) ethanol. The pellet was allowed to air dry and then resuspended in 100µl of water. The concentration was determined using UV absorbance, and stock RNA was diluted to 1µg/µl to be stored at -80°C.

2.2.2.5 DNA purification using the Qiaquick PCR purification kit

A Qiagen PCR purification kit was used to purify DNA sample after enzymatic reactions, such as methylation of oligonucleotides or proteinase K digestion. PB, a high salt buffer, was added to each sample, before applying the sample to the QIAQuick spin column. In the high salt conditions, DNA binds to the silica membrane of the column. The DNA was then washed with PE, an ethanol containing buffer, before the DNA was eluted using EB or water, where the low salt conditions allows the DNA to leave the membrane.

2.2.2.6 QPCR

Quantitative PCR is a technique used to assay relative amounts of specific regions of DNA in a sample. This was used to analysis the ChIP samples.

Each reaction contained:

SensiMix SYBR	12.5µl
Forward primer	2.5µl of 3µM primer, final concentration in reaction 0.3µM
Reverse primer	2.5µl of 3µM primer, final concentration in reaction 0.3µM

Water	5.0µl
Template DNA	2.5µl

The samples were analysed using an Applied Biosystems 7500 real time PCR machine, and subjected to the following program:

95°C 10 minutes

95°C 15 seconds
60°C 1 minute
} 40 cycles

95°C 15 seconds
60°C 1 minute
95°C 15 seconds
} Disassociation curve

A series of dilutions of input control DNA (1/4, 1/16, 1/64 and 1/256) from Akata cells were used to create input standard curves for each primer set. Results were expressed as percentage input control.

2.2.2.7 DNA methylation

The enzyme M.SssI was used to methylate CpG dinucleotides in double stranded oligonucleotides for use in EMSAs or competition EMSAs. The following reagents were combined for each methylation reaction:

33.3nM double stranded oligonucleotide/probe	10µl
10X NEBuffer 2	2µl
1600µM S-adenosylmethionine (SAM)	2µl

Water	5µl
M.SssI Methyltransferase	1µl

An equivalent mock reaction, replacing the enzyme with water, was performed as a control.

The samples were mixed by pipetting and incubated at 37°C for at least 2 hours to allow the methylation reaction to take place. The samples were then incubated at 60°C for 20 minutes to inactivate the enzyme. To reanneal the oligonucleotides, the samples were incubated at 95°C for 2 minutes, 65°C for 10 minutes, and 37°C for 30 minutes. Oligonucleotides/probes were then purified using the Qiaquick PCR purification kit.

Methylation status can be checked using BstUI digestion in oligonucleotides/probes where the motif CGCG is present.

2.2.3 Protein Methods

2.2.3.1 *In vitro* translation

Wheatgerm extract was used to express Zta and binding mutants C189S Zta, C189A Zta and S186A Zta.

Wheatgerm extract	100µl
Amino acid mix without methionine	16µl
[^{35}S]-methionine	1µl
Potassium acetate	8µl
RNA (1µg/µl)	10µl

This was incubated at 25°C for 2 hours.

To allow comparable quantification of the different Zta proteins, [^{35}S]-Methionine was added to the amino acid mix, radioactively labelling the proteins.

2.2.3.2 Protein electrophoresis

The *in vitro* translated proteins were resolved on a 15% protein gel, before quantitation of proteins using phosphor imaging and ImageQuant software.

15% polyacrylamide protein gels

Stacking gel

- 0.6ml stacking buffer
- 0.6ml Protogel (30%) 37.5:1 acrylamide:bisacrylamide
- 1.3ml water
- 20µl 10% w/v Ammonium Persulfate
- 2µl Tetramethylethylenediamine (TEMED)

Resolving gel

- 1.3ml Protogel buffer
- 2.5ml Protogel (30%) 37.5:1 acrylamide:bisacrylamide
- 1.2ml water
- 20µl 10% w/v Ammonium Persulfate
- 2µl Tetramethylethylenediamine (TEMED)

The gels were made using the BioRad Mini-PROTEAN 3 gel casting system. The resolving gel was poured into the mould to 1cm below where the bottom of the wells would be. 1ml of water was layered on top to ensure an even interface between the stacking and resolving gel. Once the resolving gel was set, the water was removed and the stacking gel added, into which the comb was inserted and allowed to set.

2µl in vitro translated protein was added to 10µl Sample buffer, Laemmli 2X concentrate and heated to 95°C for 10 minutes to denature the proteins. These samples were loaded onto the gel and run for 50 minutes at 200 volts in 1X Tris-Glycine-SDS running buffer, in BioRad Mini-PROTEAN 3 apparatus. The gel was incubated at room temperature in fixing solution for at least 1 hour, before drying at 80°C for at least 1 hour under vacuum.

2.2.3.3 Cell Extracts for EMSAs

Cells were grown to 80% confluence before harvesting in PBS and subjected to centrifugation at 13,000rpm at 4°C for 5 minutes. The supernatant was removed, and the cell pellet was resuspended in 1 packed cell volume of solution A. The solution was allowed to swell on ice for 15 minutes. The cells were then lysed using a 25 gauge hypodermic needle, but preparing the chamber with solution A, before slowly drawing the cell suspension up into the chamber and quickly expelling it a total of five times. The cells were then subjected to centrifugation for at 13,000rpm at 4°C For 20 seconds. The supernatant is discarded at this point, and the cell pellet resuspended in 2 thirds of one packed cell volume. This was incubated at 4°C for 30 minutes with agitation. A final centrifugation step was performed at 13,000rpm at 4°C for 5

minutes. The supernatant was then collected to use as nuclear extract. (Protocol adapted from (Lee et al., 1988))

2.2.4 Interaction Studies

2.2.4.1 *In Vitro*: Electrophoretic Mobility Shift Assay (EMSAs)

2.2.4.1.1 EMSA probe labelling

To generate a binding site probe to use in the assay, the oligonucleotide required was labelled with [γ-^{32}P]-ATP via a using polynucleotide kinase reaction on one strand before both strands were annealed together.

Forward single strand oligonucleotide (10µM)	0.6µl
10X PNK buffer	0.6µl
[γ- ^{32}P]- ATP 3000mCi/mmol (2X molar excess)	3.0µl
PNK enzyme (u/µl)	0.3µl
Water	1.5 µl (total volume 6 µl)

The reagents above were combined and incubated at 37°C for 30 minutes to allow phosphorylation of the 5' end of the forward oligonucleotide. To inactivate the enzyme, the reaction was incubated at 65°C for 10 minutes. A 2X excess (1.2µl of 10µM) of the 10µM reverse strand oligonucleotide was added, and the mixture was then incubated at 95°C for 2 minutes to denature any secondary structure that may have formed within the single stranded forward strand. The temperature was then decreased to 65°C for 10 minutes and 37°C for 30 minutes to allow annealing between the forward and the reverse strand. The sample was added to a G25 column (GE Healthcare), to remove any

unincorporated ^{32}P γ ATP. TE was added to make a final volume of 180µl, and G25 columns (GE Healthcare) were used to purify to probe. The final concentration of the double stranded DNA probe was 33.3nM, and was stable for approximately 2 weeks. Where EMSAs are presented, they are representative of at least 2 experiments.

2.2.4.1.2 EMSA reaction

A master mix was made so that per 10µl reaction the following components were present:

5X EMSA binding buffer	2µl
0.1M DTT	0.4µl
Water	4.6µl

To each reaction 2µl of Zta protein and 1µl of radioactively labelled probe were added, making a total reaction volume of 10µl. The mixture was incubated at room temperature for 30 minutes, and the 8% EMSA gel was pre-run for 30 minutes at 100 volts in 0.5X TBE. 2µl gel loading solution (Type 1) (Sigma) was added to each reaction, and 10µl of this was loaded onto the gel, which was electrophoresed at 100 volts for 1 hour in 0.5X TBE, to separate free probe and probe complexed with protein. The gel was incubated for 1 hour in fixing solution and then dried under vacuum for 1 hour. The dried gel was then subjected to phosphor imaging for complex quantification.

2.2.4.1.3 EMSA gels

8% acrylamide EMSA gels were made using the BioRad Mini-PROTEAN 3 gel casting system. Stock EMSA gel mix was made up in the following manner:

AccuGel (40%) 29:1 Acrylamide:Bisacrylamide 75ml	
10X TBE	25ml (0.5X final concentration)
Water	300ml (final volume 500ml)

200µl 10% (w/v) Ammonium Persulphate and 10µl Tetramethylethylenediamine (TEMED) was added to 6ml of the stock EMSA mix, mixed and immediately poured into the gel moulds. A comb was inserted and the gel was allowed to polymerise at room temperature.

2.2.4.1.4 Preparation of double stranded competitors

Equal quantities of each of the complementary oligonucleotides were added to an appropriate volume of TE and diluted to produce a solution of concentration 3.33µM. This was incubated at 95°C for 2 minutes, 65°C for 10 minutes and 37°C for 30 minutes. The labelled probe has a concentration of 33.3nM; therefore this competitor is 100 times the concentration of the probe.

2.2.4.1.5 Competition EMSA

A master mix was made in the same manner as previously detailed in 1.5. To each reaction 1µl of Zta protein, 1µl of the radioactively labelled probe and 1µl of a non-labelled double stranded DNA competitor, as described in 1.6, were added. This mixture was subjected to the same method as detailed in 1.5.

2.2.4.2 *In Vivo* Interaction studies: Chromatin Immunoprecipitation (ChIP)

2.2.4.2.1 Chromatin Preparation

After IgG treatment or an equivalent mock treatment, Akata cells were harvested at 48 hours post lytic induction. Formaldehyde was added, at a final concentration of 1%(v/v), and incubated at room temperature on a rocker for 10

minutes, to crosslink the proteins to the DNA within the cells. The reaction was then quenched with glycine, at a final concentration of 0.125M, and the cells pelleted by centrifugation at 13,000rpm at 4°C for 5 minutes. The cells were washed with DPBS twice, before a final centrifugation at 2000rpm at room temperature for 5 minutes.

Chromatin was extracted by using 300µl of ChIP cell lysis buffer per $1x10^7$ cells. This was incubated on ice 15 minutes, before centrifugation at 8,000 rpm at 4°C for 5 minutes. The supernatant was removed and the pellet was resuspended in 200µl ChIP SDS lysis buffer per $1x10^7$ cells. The sample is then sonicated to fragment the chromatin to between 200 and 600 basepairs in length. This was done on ice for 10 cycles of 10 second pulses, 30% output.

2.2.4.2.2 Immunoprecipitation

100µl of 50% bead slurry is required per immunoprecipitation. Protein A and Protein G beads were combined in a 1:1 ratio and washed 3 times with ChIP dilution buffer and centrifugation at 3000 rpm at 4°C for 5 minutes, and resuspended in ChIP dilution buffer. The beads were then preblocked by incubation with 55µl of 10mg/ml salmon sperm DNA per ml of 50% bead slurry, at 4°C with rotation for 30 minutes.

The chromatin was diluted with ChIP dilution buffer 1:10. 1ml was taken for each immunoprecipitation and precleared by adding 40µl bead slurry and incubated at 4°C with rotation for 30 minutes. The samples were then centrifuged at 3000 rpm at 4°C for 5 minutes, and the supernatant was transferred to fresh tubes. 40µl was removed from each sample at this point (4% total volume) to use as input controls.

10µg of antibody were added to each sample and incubated at 4°C with rotation for at least 30 minutes. 55µl of beads to each sample and incubated at 4°C with rotation overnight.

The beads were recovered by centrifugation at 3000 rpm at 4°C for 5 minutes. The beads were then washed in 4 different buffers at 4°C with rotation for 15 minutes in the following order:

ChIP low salt buffer

ChIP high salt buffer

ChIP LiCl buffer

TE buffer

150µl ChIP elution buffer was added to each sample and incubated at 65°C for 20 minutes, agitating once to mix. This elutes the antibody-protein-DNA from the beads. The beads were pelleted by centrifugation at 3000 rpm at 4°C for 5 minutes and the supernatant was transferred to new tubes. 150µl of elution buffer was also added to the input controls. All samples were then incubated at 65°C overnight to reverse cross-linking.

The samples were briefly centrifuged, before the addition of 45µg of proteinase K and 150µl of TE buffer to each tube and incubation at 50°C for 3 hours to degrade the protein.

A Qiagen PCR purification kit was then used to purify the DNA sample. PB, a high salt buffer, was added to each sample, before applying the sample to the QIAQuick spin column. In the high salt conditions, DNA binds to the silica

membrane of the column. The DNA was then washed with PE, an ethanol containing buffer, before the DNA was eluted using EB or water, where the low salt conditions allows the DNA to leave the membrane.

Serial 4 fold dilutions were then made using water in preparation for QPCR, as described in section.

2.2.5 *In silico* methods

2.2.5.1 Pattern Match

E. Hellen (BSMS) wrote a perl program designed to search large amounts of genomic data for specific DNA sequences. The program was used on a Linux operating system to search 500 base pairs upstream of every human gene in ensembl build 49 for the characterised RpZRE3 sequence TCGCGAA. The output contains the ensembl accession code and the position offset of the ZRE.

The pattern match search was developed and extended to include more sequences and rewritten in Java by D. Thomas (BSMS). This program searched 1000 basepairs upstream of every Havana annotated (protein coding) human gene in ensembl build 57, and also the entirety of the EBV genome (type 1). The Human herpesvirus 4 (Epstein-Barr virus) Genome NC_007605 was extracted from the NCBI in GenBank format. Features were extracted to create a table of gene locations within the genome. An experimentally verified list of 32 ZRE sequences, in forward and reverse complement form, were compared to the sequence and a table of matches produced. This data can be accessed through a website created in collaboration with D. Thomas at http://bioinf.biochem.sussex.ac.uk/EBV/.

2.2.5.2 Position Weight Matrices and Position Frequency Matrices

Position weight matrices use position frequency data collated from a collection of a specific protein DNA binding sites to calculate log likelihoods and attempt to identify conserved residues within the binding site. These position weight matrices can be used to predict other sites where the protein of interest may bind. E.Hellen (BSMS) wrote a perl program, using TFBS perl modules, designed to use position weight matrix data to search gene promoters for likely binding sites. This was used to filter our data from the initial search, after the first 5 sites had been tested via EMSA.

Position Weight Matrices use log likelihood values, whereas Position Frequency Matrices use absolute frequency values. Whilst using MatScan for ZRE prediction, the difference between PFM and PWM predictions was found to be negligible, and therefore PFMs were used.

2.2.5.3 Sequence logo generation

Weblogo, found at http://weblogo.berkeley.edu/, was used to generate sequence logos to represent position frequency matrices.

2.2.5.4 GO term over-representation and CpG island analysis

The GO terms associated with each gene were retrieved using BioMART and those annotations which were manually curated or experimentally verified were retained. The number of genes retrieved for each GO term was compared to the total number of genes in Ensembl with the same annotation. Terms which occurred with a higher frequency were retrieved, and genes that were not annotated with these over-represented terms were removed to reduce the number of candidate genes (E.Hellen, BSMS).

DAVID analysis (Huang da et al., 2009b, Huang da et al., 2009a) was also used for GO term overrepresentation.

CpG island locations were predicted in the promoter sequences, using the EMBOSS program CpGPlot with the default settings: a window size of 100; a step of 1; a minimum average observed to expected ratio of C plus G to CpG in a set of 10 windows of 0.6; a minimum average percent of G plus C in a set of 10 windows of 50% and a minimum length of 200 nucleotides. The positions of the predicted CpG islands were compared to the positions of the ZRE3 motifs to identify those ZRE3 motifs that fell within a CpG island (E.Hellen, BSMS).

2.2.5.5 PROMO and MatScan prediction

PROMO uses position frequency matrix data from version 8.3 of TRANSFAC, a database of transcription factor binding sites, to search the input DNA sequence for putative transcription factor binding sites. The settings can be adjusted to allow different extents of variability. The search options allow you to choose to search only human factors, selected species factors or all factors. The sites to be considered can also be specified; only human sites, selected species sites or all sites can be considered.

MatScan is a linux-run program, similar to PROMO, with the ability to input your own position frequency matrix for your transcription factor of choice (Blanco et al., 2006).

2.2.5.6 Comparing and aligning lists

To identify genes which appeared in more than one list, the “match” function in Windows Excel 2007 was used. To align the lists, a macro was adapted from code written by Cy Bones at

http://www.neowin.net/forum/index.php?s=d3abcb574ea5c0136be1f5d321b32a6f&showtopic=652034&pid=589603274&st=0&#entry589603274 (Accessed May 2009)

2.2.5.7 DAVID analysis

The online software DAVID (Huang da et al., 2009b, Huang da et al., 2009a) was used to annotate gene lists, convert gene names from one nomenclature to others and to search for overrepresented annotations.

Chapter 3: Methylation Sensitive ZREs in the human genome

3.1. Introduction

DNA methylation of the promoters of human genes is associated with the formation of heterochromatin, a structure of proteins and DNA that is repressive to transcription. Methylation occurs at CpG motifs, and is preserved throughout cell divisions, catalysed by DNA methyltransferases (DNMTs) (Sharma et al., 2010). It creates a repressive chromatin environment by interfering with the protein:DNA interactions of transcription factors, such as c-myc (Prendergast and Ziff, 1991) and MLTF (Watt and Molloy, 1988), and by recruiting chromatin remodelling factors, for example HDACs, via methyl binding proteins like MeCP2 (Jones et al., 1998, Nan et al., 1998). The very presence of these methyl-binding proteins could also prevent transcription factor binding. Zta interacts with DNA directly through sequence-specific sites known as Zta Response Elements (ZREs), some of which contain CpG motifs. Zta has the unusual ability of interacting with methylated ZREs (Bhende et al., 2004). The different effects methylation has upon Zta:DNA binding, depending on the specific site, has led to the classification of ZREs (Karlsson et al., 2008b). Class III ZREs contain a CpG motif, and are recognised only when methylated. At the start of this project, three genes containing this class of ZREs were studied: EBV *BRLF1* and *BRRF1,* and human *Egr1.* The methylation dependent regulation of *BRLF1* is best characterised, and the interaction of Zta with a methylated ZRE was crucial to the reactivation of EBV into lytic cycle in B cells (Bhende et al., 2004, Bhende et al., 2005, Karlsson et al., 2008a). The *BRLF1* promoter, Rp, contains three ZREs, RpZRE1,2 and 3, two of which contain a

CpG motif. A single point mutation of Zta, C189S, was able to destroy the binding to methylated RpZRE3, but not to RpZRE1 or 2, allowing the specific activation of transcription through this methylated site to be confirmed (Karlsson et al., 2008a). The viral promoter of the early lytic viral gene *BRRF1* contains two Class III ZREs, both allowing Zta to bind in their methylated form only (Dickerson et al., 2009). The discovery of ZREs in the promoter region of the human gene *Egr1* which allowed Zta to successfully activate expression (Chang et al., 2006) quickly led to further investigation into the effect of CpG methylation on this binding, and confirmed methylation dependence (Heather et al., 2009). This suggested that Zta may play a further role in overturning the epigenetic silencing of host genes, in an attempt to modulate the host cell expression.

In order to investigate the potential epigenetic overturn of silenced human genes by Zta, the RpZRE3 core sequence (TCGCGAA), which has most clearly been shown to be required for methylation dependent activation, was chosen to identify all human genes that contain an exact match of this site. This allowed identification of potential methylation dependent Zta responsive human genes. The effect of flanking sequence on methylation-dependent binding was questioned, and the utilisation of bioinformatic tools and techniques was required for candidate gene selection. Existing binding mutants were also tested, in an attempt to confirm previous findings in new flanking regions.

3.2. Results

3.2.1 Identification of Human Promoters Containing RpZRE3-Core Elements: Initial Sampling

Promoter regions (defined as −500 to +1 base pairs from the gene start site) of a sample of human protein-coding genes in Ensembl build 49 (Hubbard et al., 2007) were extracted using the Biomart data management system (Haider et al., 2009). The sample represented 40% of the human genome. A search was made for all occurrences of an exact (forwards and reverse) match to identify those promoters with the TCGCGAA RpZRE3 core element (performed by E. Hellen, BSMS)

The initial search for RpZRE3-core elements in a sample of human promoters revealed 67 genes with an RpZRE3-core element. Five of these that are involved in gene regulation were selected for DNA binding analysis. These five genes were *Z3CH8* (ENSG00000144161), *HDAC2* (ENSG00000196591), *XPC* (ENSG00000154767), *MNT* (ENSG00000070444) and *CyclinL2* (ENSG00000116148) and together with the RpZRE3 element from the viral *BRLF1* promoter (Rp), formed the RpZRE3-promoter-6 data-set.

DNA-binding assays were undertaken with the methylated forms of each of the five human promoters using 27mer double strand oligonucleotides encompassing the 7-nucleotide core element and 10-nucleotide flanking region on each side together with RpZRE3 from Rp. Electrophoretic mobility shift assays (EMSA) were undertaken with *in vitro* translated Zta protein. Zta protein/DNA complexes formed readily with the oligonucleotides from all six promoters (Figure 3.1). The specificity of the assay is shown by the lack of complex formation with non-programmed negative control protein (wheatgerm

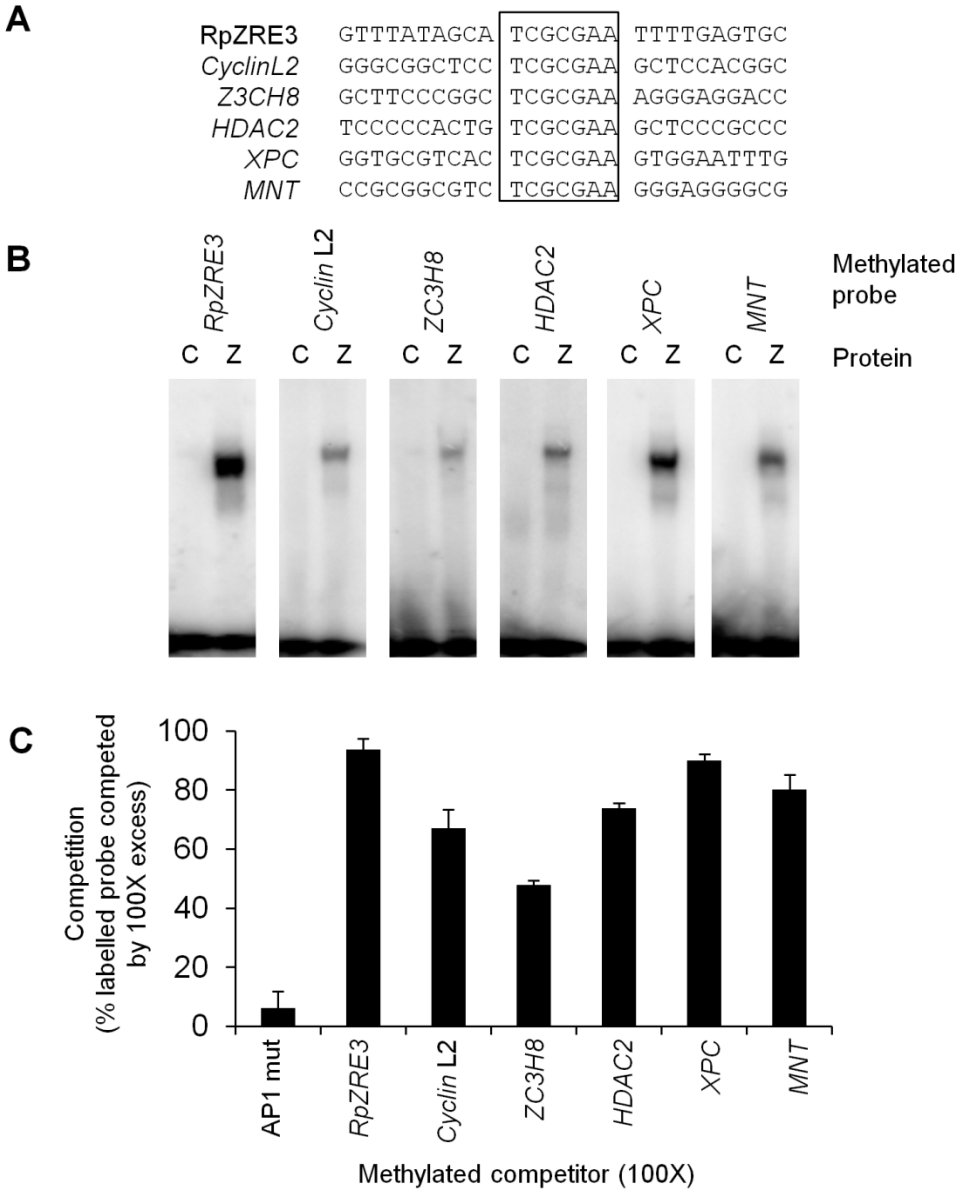


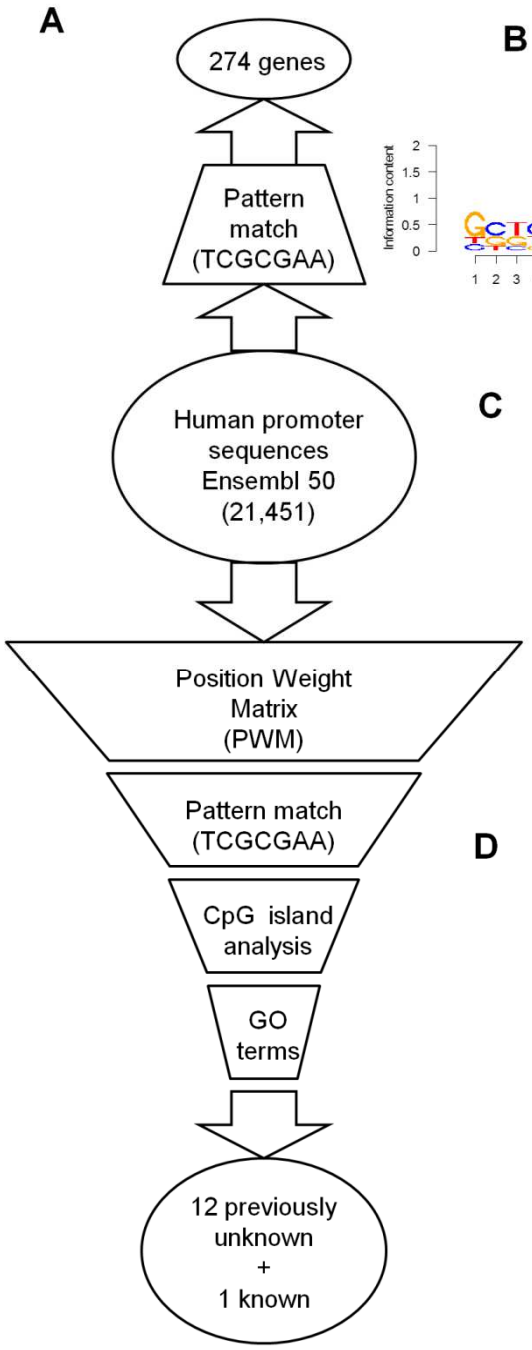
Figure 3.1 Interaction of Zta with methylated RpZRE3 core element from five human promoters.

A The nucleotide sequence of one strand of each oligonucleotide spanning the indicated ZREs is shown, with the conserved RpZRE3 core element (TCGCGAA) aligned. Double strand versions of these sequences were used as probes in EMSA, with the 2 cytosines within the CpG core motif methylated. The RpZRE3 from EBV BRLF1 promoter is included. This comprises the RpZRE3-promoter-6 data-set. **B** Double strand oligonucleotides were generated with the core ZRE sequence and at least 10 nucleotides of cognate sequence on either side. Following radiolabeling, these were incubated with *in vitro* translated Zta and subject to EMSA. The reactions contained no protein, 0, control lysate, C or Zta, Z. **C** Competition of each ZRE sequence (at 100X excess) against labelled RpZRE3 was determined by competition EMSAs. Data from at least 2 experiments was taken to calculate competition i.e. the percentage of labelled probe displaced by unlabelled competitor. Methylated ZREs were used for both probe and competition. AP1 mut, an oligonucleotide previously shown not to interact with Zta was used as a negative control to define the level of non-specific binding. Error bars indicate standard error.

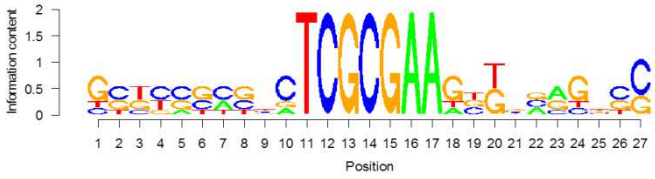
lysate without the addition of BZLF1 RNA). In addition, competition experiments were undertaken with an excess of unlabelled oligonucleotides and included a version with a mutant ZRE to further probe the specificity of the interaction. This confirms that Zta interacts with all six RpZRE3 core elements specifically. It also appears that flank may affect the strength of binding, but not specificity.

3.2.2 Identification of Human Promoters Containing RpZRE3-Core Elements: Whole Genome Scanning

An identical screen was undertaken on the entire human genome from Ensembl (50) (Flicek et al., 2008) resulting in the identification of 274 promoters which were designated the RpZRE3-promoter-274 data-set (See Appendix I). To assess whether there was an influence of flanking sequence on the interaction of Zta with these ZREs, selection of a sub-set of sites was required, so a systematic filtering process was undertaken with assistance from bioinformatician E. Hellen (BSMS), as outlined in Figure 3.2 A. The 7 nucleotides of the RpZRE3 core element, along with 10 flanking nucleotides on each side, were extracted from each of the five genes from the initial sampling. The Transcription Factor Binding Site (TFBS) Perl modules (Lenhard and Wasserman, 2002) were used to create a Position Weight Matrix (PWM) based on the entire length of these 27-nucleotide sequences. TFBS Perl modules were also used to search the promoter regions (−500 to +1) of all human protein coding genes in Ensembl build 50 (Flicek et al., 2008) which were extracted using Biomart (Haider et al., 2009). Those promoters that matched the PWM with a threshold value >80% and contained an exact match to the RpZRE3-core element were additionally filtered using two criteria (a) the presence of CpG



B



C

Overrepresented GO terms
Chromatin binding
Chromatin remodelling
Mitosis
Mitotic cell cycle checkpoint
Oxidation reduction
Satellite DNA binding
Transcription coactivator activity
Transcription corepressor activity
Transcription regulator activity

D

ENSG00000162909	*CAPN2*
ENSG00000129596	*CDO1*
ENSG00000171634	*FALZ*
ENSG00000054523	*KIF1B*
ENSG00000131899	*LLGL1*
ENSG00000143013	*LMO4*
ENSG00000129071	*MBD4*
ENSG00000104886	*PLEKHJ1*
ENSG00000184304	*PRKD1*
ENSG00000129657	*SEC14L1*
ENSG00000171148	*TADA3L*
ENSG00000077097	*TOP2B*
ENSG00000070444	*MNT*

Figure 3.2 Identification of human promoters containing RpZRE3 core elements in the human genome.

A The bioinformatics analysis undertaken for the genome wide scan is represented as a flow diagram. Genetic input or output is represented by ovals, and filters by trapeziums. **B** A Position Weight Matrix (PWM) was created using the aligned sequences of the RpZRE3-promoter-6 data-set. These sequences are shown in figure 4.1A. This was used to filter the genomic data from the human promoter sequences (−500 to +1). **C** Overrepresented GO terms found for the 274 genes were identified. **D** The 13 genes identified as candidate genes, 12 previously unknown genes and 1 from the initial screen, and their associated Ensembl accession codes.

islands and (b) function (based on the over-representation of Gene Ontology (GO) terms (Ashburner et al., 2000)). The location of CpG islands were predicted using the EMBOSS program CpGPlot (Rice et al., 2000) with default options. Only genes with a CpG island present in the promoter were retained. The GO term annotations for molecular function and biological processes were extracted from Ensembl for each gene (Flicek et al., 2008). The number of genes in the RpZRE3 promoter dataset with each GO term annotation was compared with the total number of genes in Ensembl with the same GO term. The GO terms that occurred with a significantly higher frequency ($p<0.05$) in the RpZRE3 dataset, compared to entire genome, were defined as over-represented. This resulted in 12 previously unidentified promoters and a further one that had been identified in the initial screen. Together with the promoters from the initial screen and the viral *BRLF1* promoter (RpZRE3-promoter-6 data-set), these form the RpZRE3-promoter-18 data-set. Oligonucleotides were designed using the same principles with 10 nucleotides of flanking sequence on each side of the RpZRE3-core element, and the ability of Zta to interact with each site in its methylated form was assessed. The EMSA analysis revealed that all twelve of the newly identified methylated ZREs were recognised by Zta (Figure 3.3). This analysis demonstrated that for the RpZRE3-promoter-18 data-set, the flanking sequence surrounding the methylated core 7-mer element did not have a profound effect on the ability of Zta to interact with promoters and therefore the core element was sufficient to facilitate binding. This was illustrated by the generation of a PWM using sequence from the RpZRE3-promoter-18 data-set (Figure 3.4), which revealed negligible sequence conservation outside of the core element.

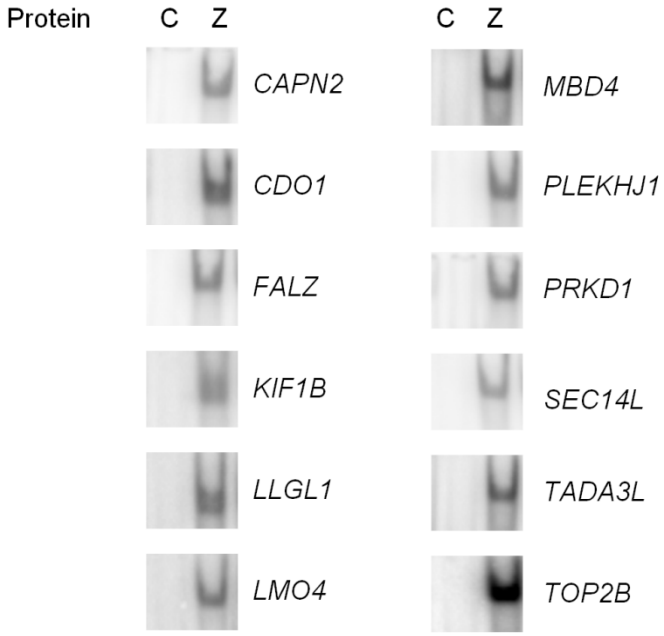


Figure 3.3 Interaction of Zta with methylated RpZRE3s found in human promoters

EMSA analysis with *in vitro* translated Zta protein (Z) or an unprogrammed translation lysate (C) with the probes indicated to the right of each gel was carried out as described in figure 3.1.

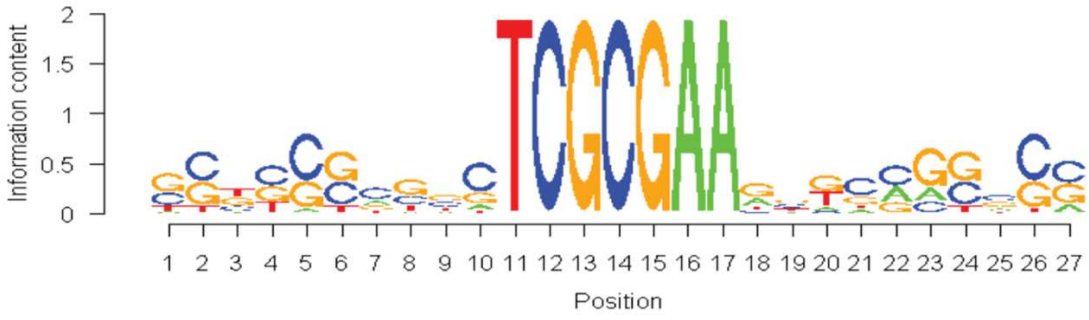


Figure 3.4 Analysis of the RpZRE3 core element flanking sequence.

A Position Weight Matrix (PWM), created using the probe sequences of the dataset RpZRE3-promoter-18.

3.2.3 Recognition of the Non-Methylated RpZRE3-Core Elements

Zta is able to recognise many response elements which do not contain a CpG motif and therefore do not have a methylated core element (Class I ZREs) (Sinclair, 2003) (Karlsson et al., 2008b). In addition, Zta recognises one CpG containing ZRE, RpZRE2, even in the absence of methylation (Class II ZREs) (Bhende et al., 2004). The ability of Zta to recognise the promoters in their non-methylated forms may impact on the ability of EBV to alter their gene expression, so the classification of ZRE for each of the 17 human promoters was investigated.

A series of DNA binding experiments with the oligonucleotides representing the human promoters in the RpZRE3-promoter-18 data-set, were undertaken comparing non-methylated with methylated RpZRE3-core elements. The interaction of Zta with the sites was quantitative, as demonstrated by the reduction in complex formation as the Zta protein concentration was titrated on each of the methylated promoters (Figure 3.5). All bar one display negligible binding to the non-methylated oligonucleotides (at least 10-fold lower than to the methylated sites) and can therefore be classified as class III ZREs.

Zta displayed a reproducible interaction with the non-methylated *XPC* oligonucleotide; the interaction reached 50% of the binding observed with the methylated oligonucleotide. This raises the possibility that the *XPC* core RpZRE3 element is influenced by the flanking sequence to behave as a class II ZRE.

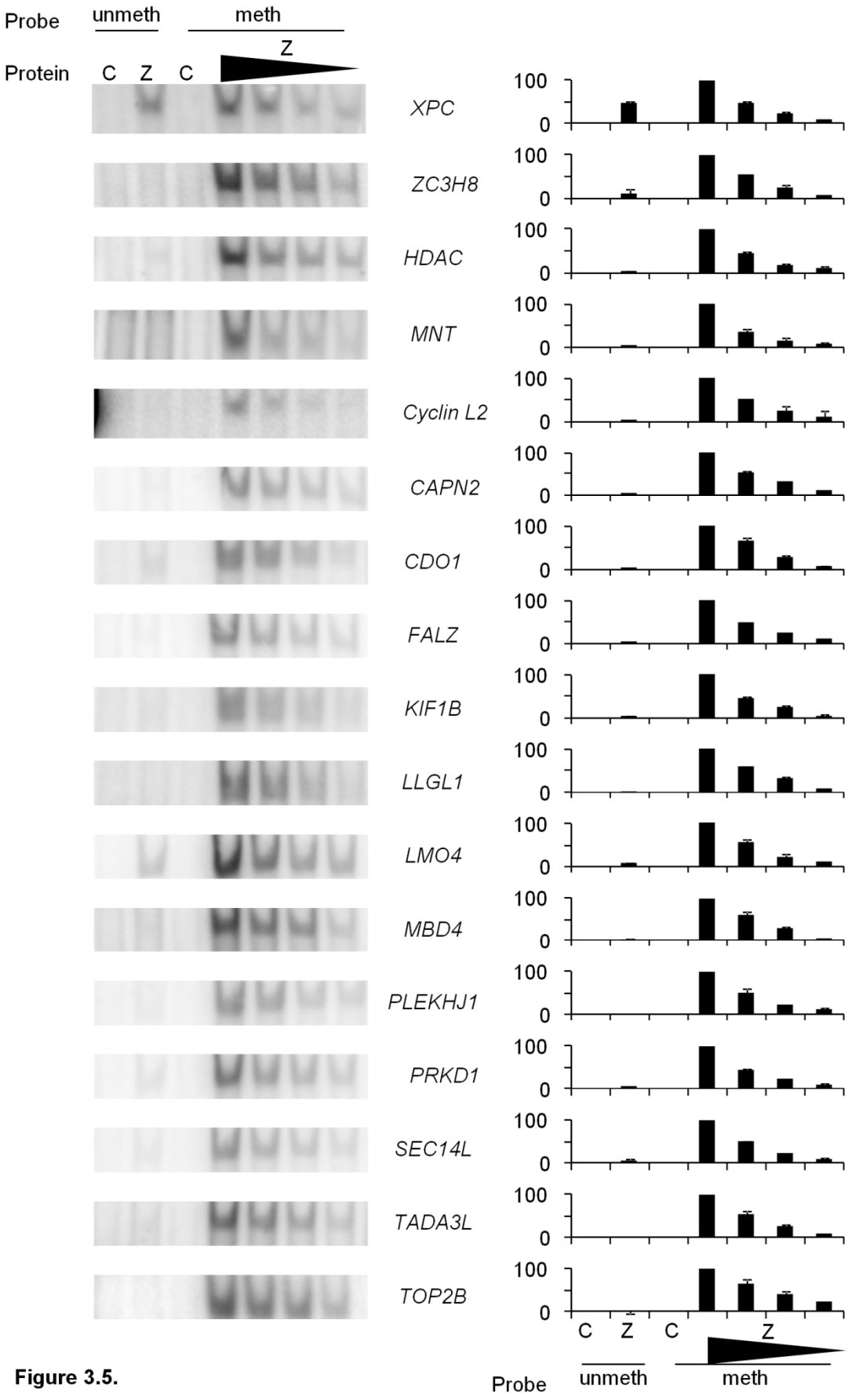

Probe
unmeth
meth
Protein
C Z C
Z
XPC
ZC3H8
HDAC
MNT
Cyclin L2
CAPN2
CDO1
FALZ
KIF1B
LLGL1
LMO4
MBD4
PLEKHJ1
PRKD1
SEC14L
TADA3L
TOP2B
100
0
C Z C Z
Probe
unmeth
meth

Figure 3.5.

Figure 3.5 Comparison of Zta interaction with non-methylated and methylated RpZRE3s from human promoters.

DNA binding analysis of *in vitro* translated Zta protein (Z) or an unprogrammed translation lysate (C) with both non-methylated and methylated versions of the probes indicated to the right of each gel was carried out by EMSA as previously described. A titration of Zta protein (1, 1:2, 1:4, 1:8) was used to compare non-methylated binding to that of the methylated equivalent. Quantitation of the complexes formed between Zta and the non-methylated and methylated probes, from at least two experiments, is represented as a histogram to the right of the corresponding gel. Complex formation is represented in the histograms relative to the maximum binding for each probe. Error bars indicate standard error.

3.2.4 Analysis of Flank Contribution to ZRE classification

To identify whether sequences conferring methylation independent recognition was confined to a specific flank of *XPC*, a series of mutant oligonucleotide probes that exchanged flanking sequences between *XPC* and a class III site (*MNT*), that is not recognised when non-methylated, were designed. Analysis of the interaction with Zta by EMSA revealed that the ability to confer Zta binding resided within the 5' sequence of *XPC*; the hybrid *XPC*L*MNT*R was able to bind but *MNT*L*XPC*R was not (Figure 3.6 A). This suggested that the 5' *XPC* flanking sequence of the RpZRE3-core element influenced binding. However, it was surprising to discover that mutation of the RpZRE3 core element within *XPC* did not prevent interaction with Zta (Figure 3.6 B).

A further explanation for the interaction of Zta with the non-methylated XPC promoter is that an obscure ZRE is present in the 5' flank. To address this further we attempted to identify putative ZREs in the XPC promoter using the transcription factor binding site prediction program PROMO (Farre et al., 2003, Messeguer et al., 2002). Although this program contains a PFM for Zta binding sites, none were predicted in this sequence. However, PROMO predicted the presence of AP1, c-fos and c-jun binding sites within the 5' flank of XPC (5'TGCGTCA) (Figure 3.7). c-fos and c-jun proteins are both members of the bZIP transcription factor family; they form AP1 DNA binding activity as either homodimers or heterodimers. It is relevant that fos/jun dimers share some DNA recognition motifs with Zta (Sinclair, 2003, Farrell et al., 1989, Urier et al., 1989, Lieberman et al., 1990, Chang et al., 1990).

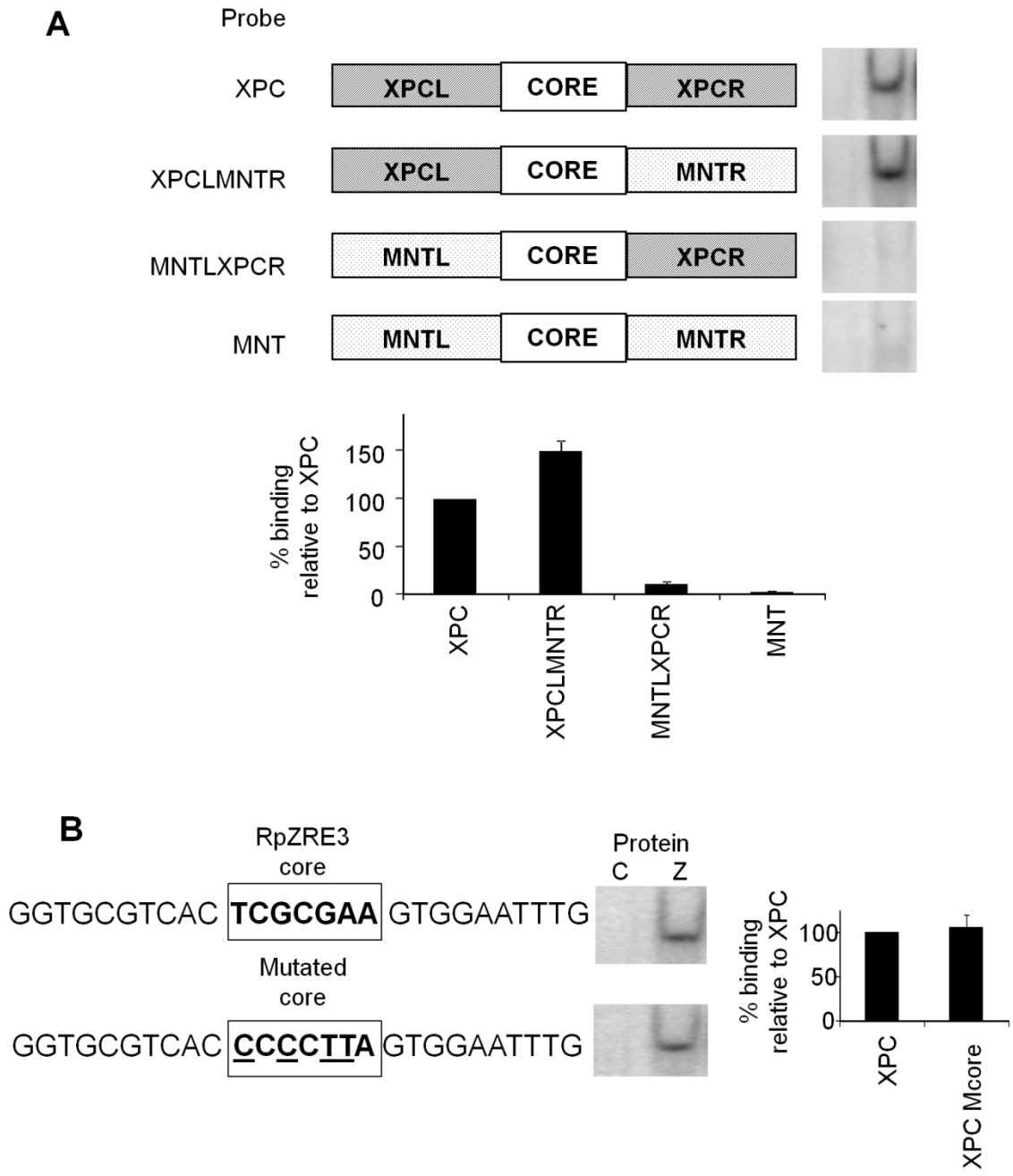


Figure 3.6 The effect of flanking region upon Zta binding to non-methylated XPC probe.

A Schematic representation of the probes, illustrating the XPC and MNT probes, and the flank swap probes XPCLMNTR and MNTLXPCR, adjacent to the corresponding EMSA analysis, carried out in the same manner as described previously. Quantification of the complexes formed with each probe, from 2 experiments was undertaken. Complex formation is represented as a percentage of the Zta complex with the non-methylated XPC probe. Error bars indicate standard error. B Schematic representation of probes, indicating mutated core sequence. EMSA analysis was carried out as described previously. Quantitation of the complexes formed with each probe, from two experiments was undertaken. Complex formation is represented as a percentage of the Zta complex formed with the non-methylated XPC probe. Error bars indicate standard error.

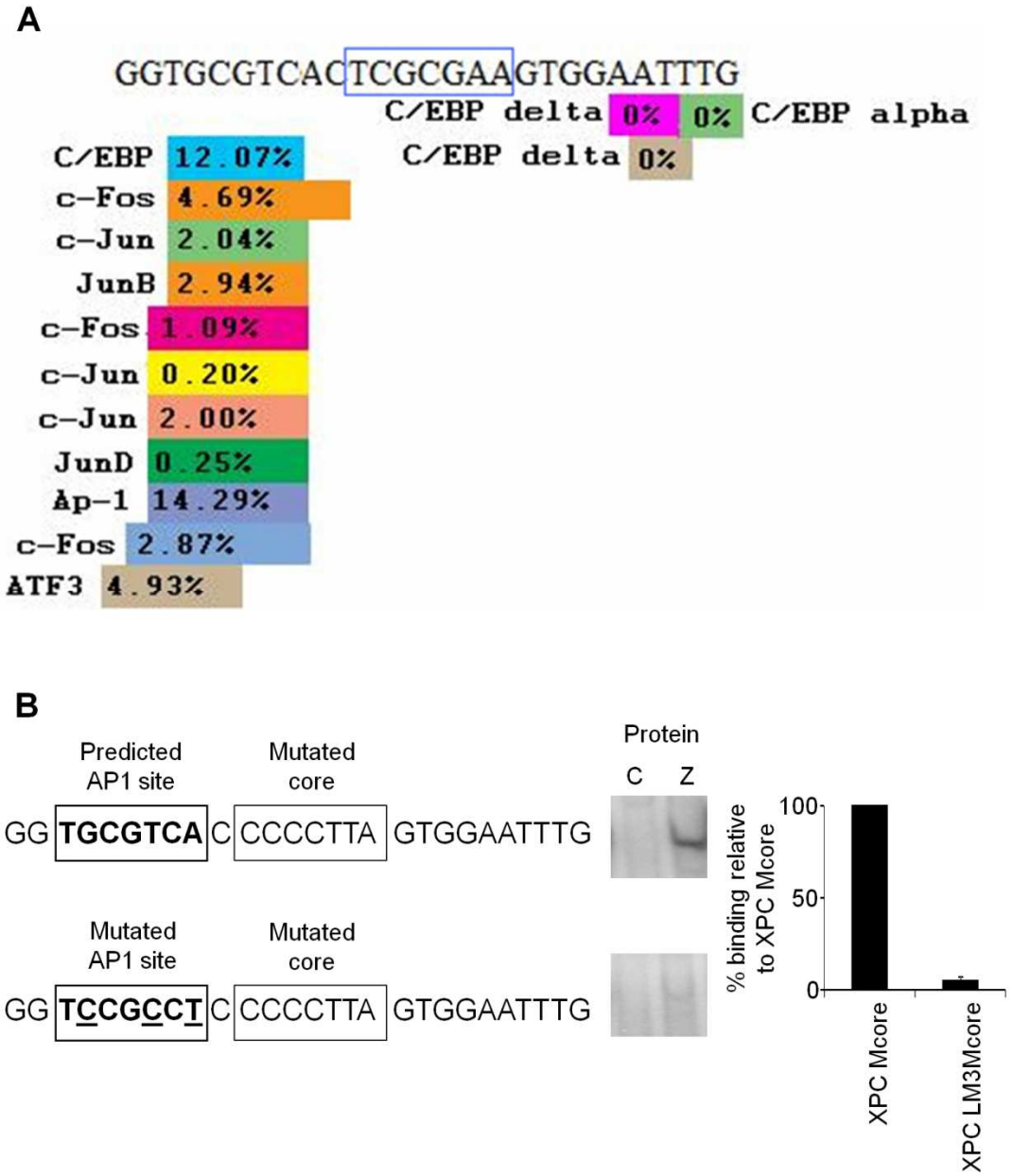


Figure 3.7 Uncovering an additional ZRE in the 5′ flank of the XPC probe.

A PROMO transcription factor binding site prediction identified an AP1 site which is highlighted in the 5′ flank sequence of XPC. Percentage values represent dissimilarity score from prediction PWM for each respective site/transcription factor. B The specific nucleotide mutations are underlined. The complexes formed by EMSA are shown to the right of the corresponding probe sequence. Quantitation of the complexes formed with each probe, from 2 experiments, was undertaken. Complex formation is represented as a percentage of the Zta complex formed with the XPC mutated core probe. Error bars indicate standard error.

This led to the hypothesis that the predicted AP1 site in the 5' flanking sequence is an additional ZRE that is responsible for binding to the non-methylated XPC oligonucleotide. To test the hypothesis, further mutations were introduced into the 5' XPC flanking sequence in the predicted AP1 site (5'TCCGCCT). The ability of Zta to interact with the double AP1/core mutant site (XPCLM3Mcore) was compared with the core-only mutant. Dual mutation of the predicted AP1 site and the core abrogated the ability of Zta to interact with the XPC oligonucleotide, demonstrating that binding of Zta to the non-methylated XPC site resided with this AP1 site (Figure 3.7).

To confirm the methylation dependent binding of Zta to the core RpZRE3 core in the *XPC* flanking environment, the novel ZRE (predicted AP1 site) was mutated to abolish binding through this site (Figure 3.8A). The oligonucleotide underwent an *in vitro* methylation using M.SssI, or an equivalent mock reaction (Figure 3.8B) and was subjected to EMSA. Significant complex formation occurred with the methylated oligonucleotide only (Figure 3.8C).

This analysis showed that the RpZRE3 core element is not recognised in its non-methylated state in the context of any of the viral or cellular promoters in the RpZRE3-promoter-18 data-set, and that this ZRE is specifically recognised when methylated. Further investigation into this additional site, a novel Zta binding sequence, will be returned to later in this chapter, in sections 3.2.6 and 3.2.7.

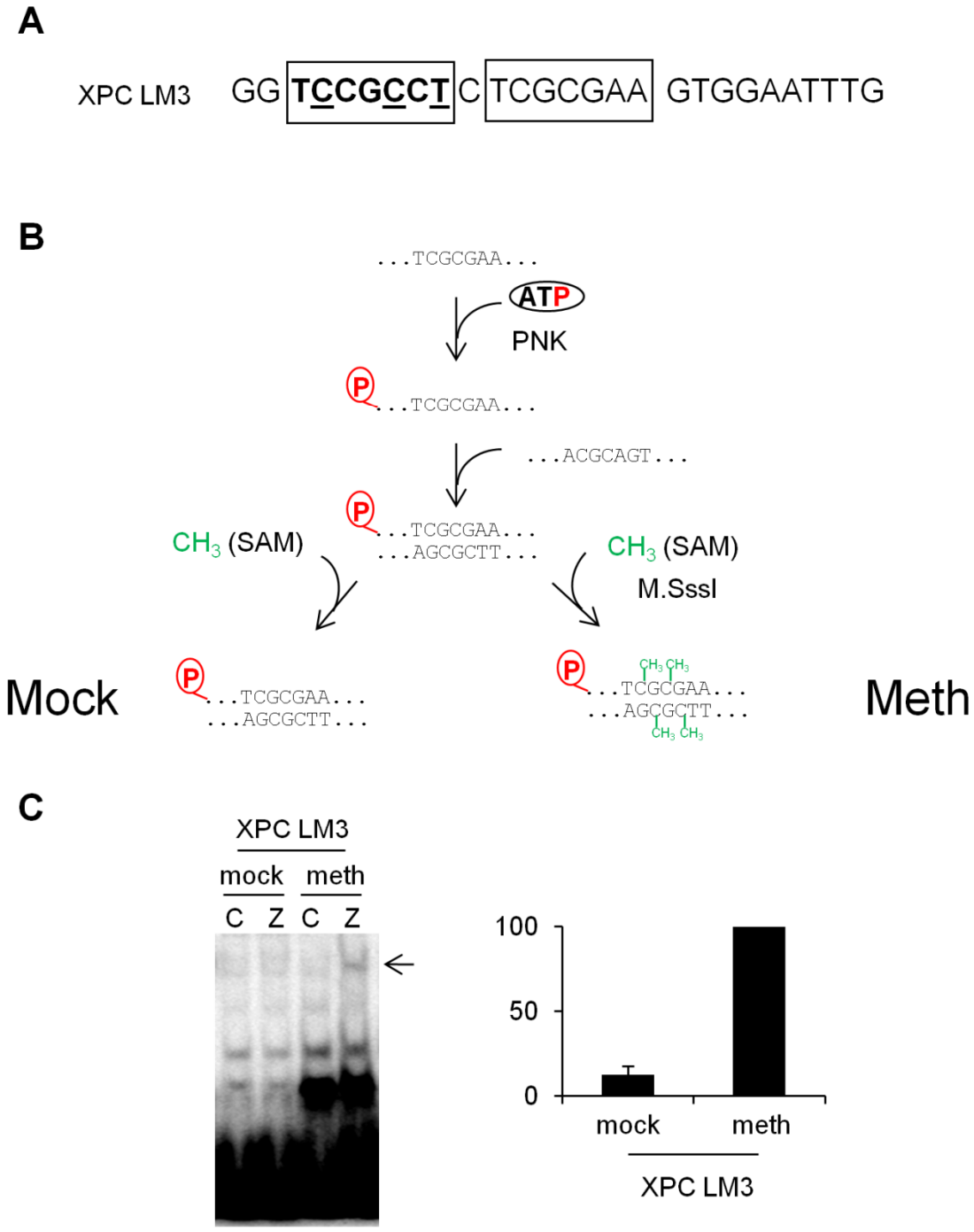


Figure 3.8 Confirmation of methylation dependent binding of core RpZRE3 element.

A Flow diagram of *in vitro* methylation method. Single stranded forward oligonucleotide was labelled with P^{32} by polynucleotide kinase (PNK), before annealing to the corresponding reverse oligonucleotide. Radioactive probes were then subjected to a methylation reaction using M.SssI and S-Adenosyl methionine (SAM), modifying every cytosine in a CpG motif. An equivalent mock reaction, performed without the M.SssI, was performed as a control for each oligonucleotide. B Oligonucleotide XPC LM3. Contains the core RpZRE3 element in the XPC cognate sequence, with the novel ZRE discovered in the left flank mutated to a non-binding sequence. C In vitro translated Zta was incubated with XPC LM3 oligonucleotide which has undergone either a methylation reaction or an equivalent mock reaction.

3.2.5 Data comparison

RpZRE3-promoter-274 data-set provides a list of human genes which contain an RpZRE3-core element in their immediate promoter region, defined as -500 to +1 of each gene start. The question remains whether these sites provide functional response elements for Zta, and under what conditions. A microarray performed in 2006 (Yuan et al., 2006) in Akata cells, an EBV positive B cell line which can be activated into lytic cycle by cross-linking with IgG (Takada, 1984), provided a list of 122 human genes affected at least 2 fold by lytic cycle activation compared to EBV negative cells cross-linked in the same manner. This list was compared with the RpZRE3-promoter-274 data-set, and only 2 genes were found to match: *LMO4* and *CEP250*. *LMO4* is a transcriptional activator over expressed in approximately half of primary breast cancer cases, and high expression is associated with poor prognosis (Sum et al., 2005). *CEP250* (also known as *CEP2*) is a core centrosomal protein important in cell cycle progression, as it is required for centriole-centriole cohesion during interphase (Mack et al., 2000).

Yuan states that *LMO4* is significantly downregulated 24-48 hours post lytic induction. However, upon further investigation of the heatmap representation of the expression data (Figure 3.9), it appears that upregulation occurs at 8-12 hours, and that downregulation is prevented at 4 hours.

No statement is made about *CEP250* in the text. However it can be seen from the heatmap that downregulation at 4-12 hours appears to be prevented by the presence of EBV, and at 8-12 hours it is upregulated compared to expression in EBV negative cells upon anti-IgG treatment.

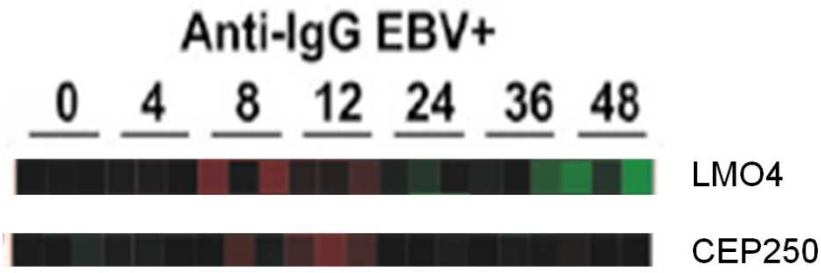


Figure 3.9 Expression profile of LMO4 and CEP250.

Cropped from Yuan et al, 2006 heatmap, expression patterns of human genes containing core RpZRE3 elements in their promoter region. The colour scale represents expression levels relative to time zero. Green indicates relative down-regulation, red indicates relative up-regulation, black indicates no change. mRNA expression was measured in Yuan et al from 3 independent experiments.

The RpZRE3-promoter-274 data-set was also compared to smaller subset of genes from another microarray (Broderick et al., 2009a); in this study, the genes were identified as being differentially regulated 6 hours after the induction of lytic cycle in Akata cells; however no matches were found.

Zta could therefore potentially play a role in the regulation of these 2 genes. However, it is important to consider whether these genes are methylated, as the presence of an RpZRE3-core element alone is not enough for Zta to bind; the CpG motif must be methylated. 3 datasets of methylome data were obtained. The first was published in 2009 (Rauch et al., 2009) from which a list of methylated promoters in circulating CD19+ B cells was produced. This was produced using a technique known as MIRA-ChIP, (Methylated CpG Island Recovery Assay Chromatin Immunoprecipitation). This uses GST-tagged MBD2b protein and His-tagged MBD3L1 protein to form a high affinity methylated DNA binding complex in order to enrich for methylated DNA, which can then be eluted, and subjected to standard microarray hybridisation techniques. The tiling arrays contained DNA probes 50-75 bps long and median probe spacing was 100bp.

In addition, methylome data was explored from germinal centre B cells, and LCLs which were generated by infecting the germinal centre B cells with EBV (Leonard et al., 2011). This data was generated using a MeIP approach, in which antibodies for methylated CpGs are used to enrich for the methylated DNA, which are then hybridised to tiling microarrays. From this, methylated promoter lists were produced, and compared to the RpZRE3-promoter-274 dataset (which had to first be converted from Ensembl code to HGDC gene name), using macro-enabled Excel. (Table 3.1) A total of 90 of the promoter regions containing RpZRE3s (RpZRE3-promoter-274 data-set) were found to be methylated in at least one cell type, 44 in at least two cell types, and eight in all three cell types (Figure 3.10). *CEP250* was found to be methylated in all three B cell types, whereas *LMO4* was found to be methylated in circulating CD19+ B cells. Therefore both genes are regulated in EBV lytic cycle, contain a methylation dependent ZRE in their promoter region, and these promoters are known to be methylated in at least one B cell type.

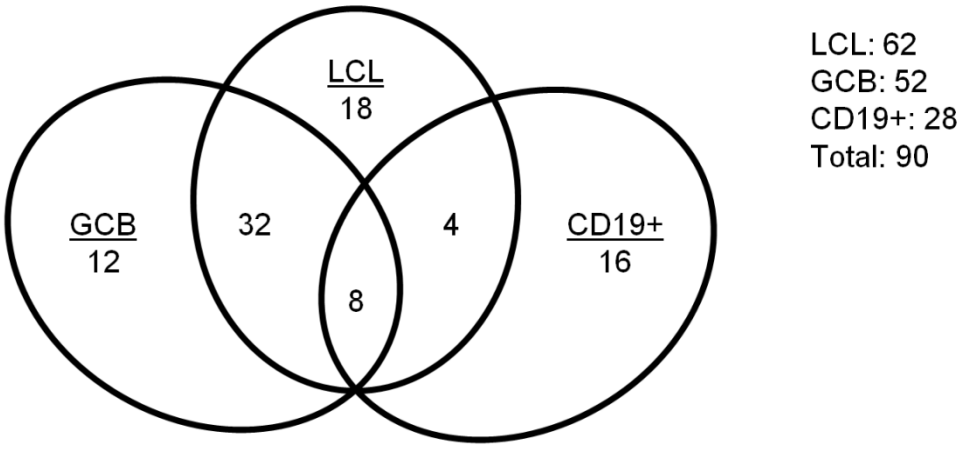


Figure 3.10 Gene promoters containing RpZRE3 core sequences and methylated in at least one of three cell types.

A Euler diagram was constructed from comparison of gene lists containing RpZRE3 core elements and methylated genes in differing B cell types. Full data is shown in table 3.1.

Gene name	LCL	GCB	CD19+
C6orf145	METH	METH	METH
CEP250	METH	METH	METH
EXTL3	METH	METH	METH
IRF2BP1	METH	METH	METH
RBBP4	METH	METH	METH
SPRY4	METH	METH	METH
STAU2	METH	METH	METH
STK10	METH	METH	METH
BAG5	METH	METH	
BAIAP2	METH	METH	
BTBD6	METH	METH	
C14orf153	METH	METH	
CCT6B	METH	METH	
CDC2L2	METH	METH	
CDCP1	METH	METH	
CSTF3	METH	METH	
CYB5D1	METH	METH	
DIP2A	METH	METH	
EIF4G2	METH	METH	
GAK	METH	METH	
GDF9	METH	METH	
GINS2	METH	METH	
AC005512.1	METH	METH	
HEATR2	METH	METH	
JAK1	METH	METH	
LINS1	METH	METH	
MTMR3	METH	METH	
NOL6	METH	METH	
OSBPL2	METH	METH	
RCSD1	METH	METH	
RNF121	METH	METH	
SEC14L1	METH	METH	
SERBP1	METH	METH	
SF3A2	METH	METH	
STK25	METH	METH	
SYNE2	METH	METH	
TM9SF1	METH	METH	
UBE2F	METH	METH	
UQCRH	METH	METH	
ZNF3	METH	METH	
CAMK2G	METH		METH
PYCR2	METH		METH
SNX17	METH		METH
ZNF155	METH		METH
C12orf57	METH		

Gene name	LCL	GCB	CD19+
CCDC77	METH		
DHX30	METH		
HSPBAP1	METH		
JMJD2A	METH		
LCMT1	METH		
LRRC41	METH		
MDN1	METH		
MMAA	METH		
NME2	METH		
PPP1R3D	METH		
RAD9B	METH		
RBM4	METH		
TAF6L	METH		
TMUB1	METH		
ZFP30	METH		
ZNF511	METH		
ZNF768	METH		
ABHD11		METH	
BAMBI		METH	
CORO7		METH	
DBF4B		METH	
DGKQ		METH	
GTF2A2		METH	
GZF1		METH	
HIST2H2AA4		METH	
HPGD		METH	
PAGE5		METH	
PCBP3		METH	
RPLP0		METH	
BCL2L11			METH
COL25A1			METH
DTNBP1			METH
DUSP6			METH
E2F5			METH
F2R			METH
IGFBP2			METH
LMO4			METH
MCHR2			METH
METT10D			METH
NT5E			METH
NUDT3			METH
PIGT			METH
PXMP3			METH
UPK3A			METH
ZNF141			METH

Table 3.1 Methylation status of human genes containing RpZRE3 core sequences in the promoter region in three different B Cell backgrounds.

3.2.6 Novel CpG-containing ZREs

The novel ZRE discovered in the flank of *XPC* RpZRE3 oligonucleotide in section 3.2.4 contained a CpG motif within the core sequence, which suggests the potential of methylation dependent regulation of binding. Further research into this site, using the TRANSFAC database (Matys et al., 2006) revealed that it matched a confirmed binding site of c-jun, in the promoter region of *PENK* known as ENKCRE-2 (Mar et al., 1992). c-jun is an important human transcription factor, and like Zta, a member of the bZIP family of transcription factors due to the characteristic leucine zipper region. c-jun can form homodimers, and, more often, heterodimers with other members of the bZIP family such as c-fos, ATF2 etc. (Shaulian and Karin, 2001). The methylation of the *PENK* promoter region by HpaII (CmCGG) was shown to inhibit AP2 binding to another site in the promoter, however the other CpG containing sites (ENKCRE1 and 2) were not investigated specifically (Comb and Goodman, 1990). Two further sequences were identified as c-jun binding sites that contained a CpG motif in the same position of the core sequence; these were found in the promoter regions of *ATF3* (Liang et al., 1996) and *FN* (Shino et al., 1997).

Oligonucleotide probes were generated for use in EMSAs to question binding of Zta to these AP1 response elements, alongside the *XPC* mut core probe from section 3.2.4. All new sites formed complexes at a comparable affinity to *XPC* mut core probe, with the exception of *FN*, which although forming a distinct band, was much weaker (Figure 3.11).

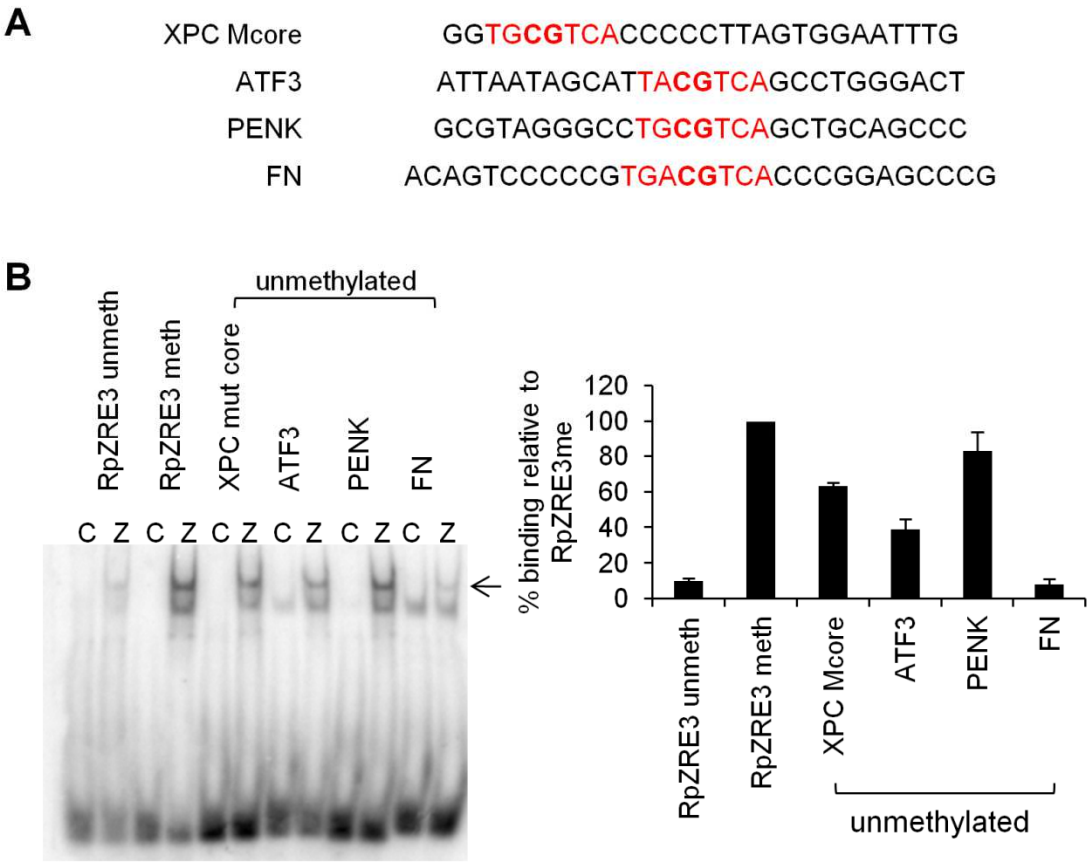


Figure 3.11 AP1-like CpG containing core sequences.

A Oligonucleotides with AP1-like CpG containing sites and natural flanking sequence. B *In vitro* translated Zta was incubated in a standard EMSA reaction with each of the oligonucleotides listed in A, compared to RpZRE3 unmethylated and methylated. Quantitation was performed relative to Zta:RpZRE3meth complex.

3.2.7 Methylation Independent Binding of Novel CpG-Containing ZREs

The 4 new AP1-like ZREs were subjected to the same methylation treatment as described in Figure 3.8, and subjected to EMSA (Figure 3.12). No significant increase in complex formation can be observed in the methylated form for any of the sequences. It appears that the methylation of this CpG-motif within the AP1-like ZREs has minimal effect on the binding affinity of Zta. This classifies the novel sites as Class II sites.

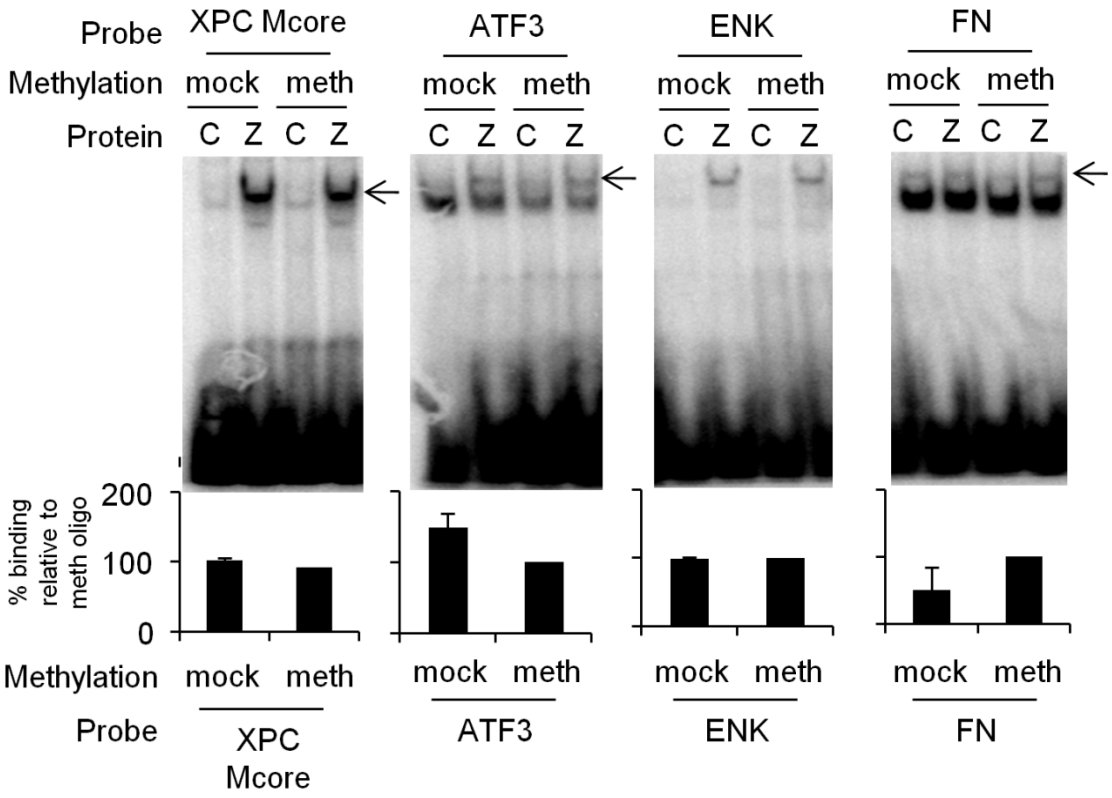


Figure 3.12 Methylation of AP1 like CpG containing ZREs appears to have a nominal effect.

In vitro translated Zta was incubated with oligonucleotides which had either undergone a methylation reaction using M.SssI or an equivalent mock reaction (as previously described in figure 4.8), and binding was assessed. Oligonucleotides used previously in Figure 4.7 and new oligonucleotides containing the AP1 like ZREs were assessed to question the methylation independence of the XPC flank site, and quantitation verifies this. Each was quantified separately, relative to each oligonucleotide which underwent the methylation reaction.

3.2.8 Unexpected effect of known binding mutants

The C189 residue of Zta has been shown to be required for binding to methylated RpZRE3, and the mutant C189S was shown to bind less than 10% the efficiency of wild type Zta. Combined C189A and S186A mutations ablated complex formation completely (Karlsson et al., 2008a). However, it was found that Zta C189S was not deficient in binding to methylated sites in the Nap (Dickerson et al., 2009). The S186A mutant has been shown to bind to some ZREs, such as those found in the promoter of *BMRF1* (Adamson et al., 2000, Francis et al., 1999), but not all, for example ZREs found in the promoter of

BRLF1 (Bhende et al., 2005). To question whether the effect of each of these binding mutants was context independent, EMSAs were performed on the RpZRE3-promoter-6 data-set, identified at the beginning of this chapter. All the Zta mutants had compromised binding ability to methylated RpZRE3, matching previous observations. However, the impaired ability of C189S to bind was not as pronounced as previously published, showing a 50% reduction in binding (Figure 3.13C). C189A was incapable of binding to the RpZRE3-promoter-6 data-set, confirming that C189 or S189 is absolutely required for DNA binding. S186A was also incapable of binding to all but one oligonucleotide, *XPC*. This is most likely explained by the presence of the AP1-like site in the flank, as the Zta mutant S186A can interact with some ZREs, and other bZIP proteins with alanine in the equivalent position of S186 in their DNA binding domain can still interact with their respective sites. C189S however, has a greater affinity than wild type Zta for each of the new sites.

There are several possibilities as to the increased binding affinity of C189S to these sites. Perhaps there is some flanking effect in these oligonucleotides that is not present in the RpZRE3 oligonucleotide, or vice versa, whether it is a separate site, a general environment effect, or a specific residue. Apart from *XPC*, none contain an alternative ZRE-like sequence in the flank when analysed by PROMO, ruling out the possibility of a separate site. The sequences of several of the sites that Zta binds to were aligned in an attempt to discover if there was an affect of the flank on C189S binding, and a sequence logo was created to represent these sequences, to allow comparison with the RpZRE3 sequence (Figure 3.14). RpZRE3 is particularly A:T rich in the 3’ (right) flank

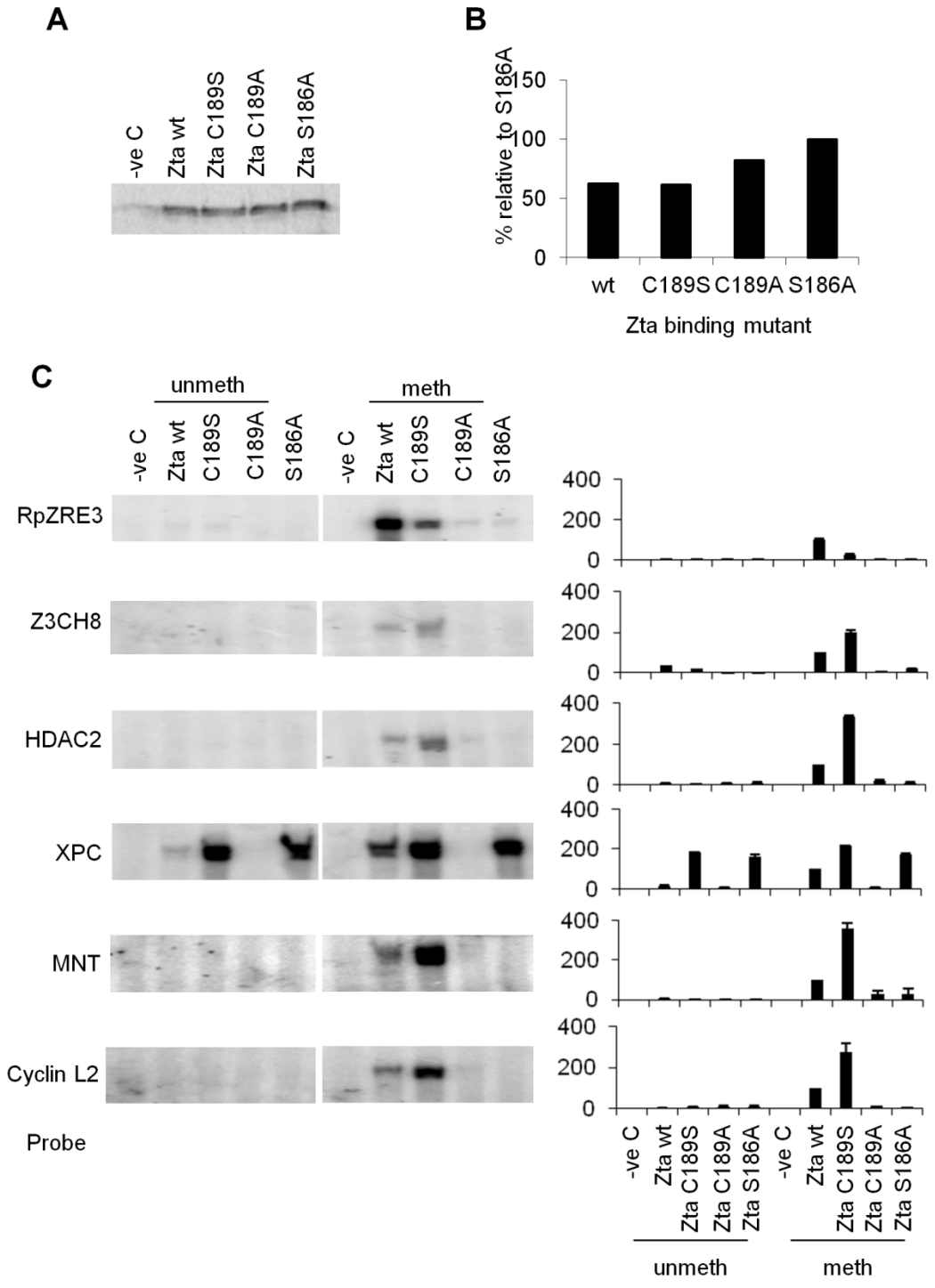
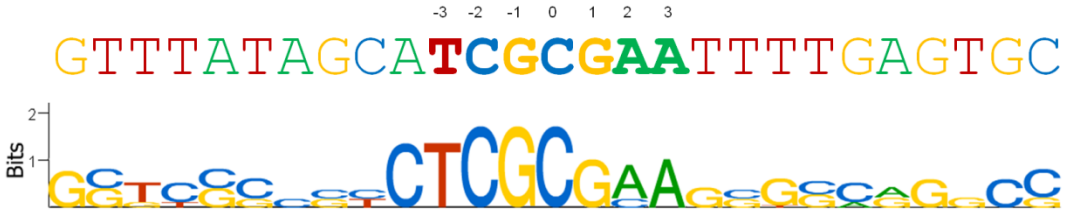


Figure 3.13 Unexpected binding with known DNA binding mutants of Zta.

A. Zta wild type, plus 3 binding mutants (C189S, C189A, S186A) were in vitro translated and run on a protein gel to ensure they were of comparable concentration. This was quantitated in B. C. In vitro translated protein was incubated with pre-methylated oligonucleotides (Sigma) or equivalent unmethylated probes, encompassing the core RpZRE3 sequence and the flanking region. Quantitation is shown adjacent from each EMSA. Wild type Zta with each site in the methylated state is set as 100%, and all other quantitation in each separate EMSA is relative to this.

A



A	0	0	0	0	0	0	2	0	1	0	1	0	0	0	0	5	6	2	0	0	1	2	2	0	1	0	0
C	1	3	1	3	3	4	3	3	3	7	0	8	0	8	1	3	1	1	3	1	3	3	1	1	1	3	3
G	4	3	2	1	3	3	1	3	1	1	0	0	8	0	6	0	1	4	3	4	3	1	2	3	2	1	2
T	1	0	3	2	0	1	2	2	3	0	7	0	0	0	1	0	0	1	2	2	0	0	0	1	1	1	0

B

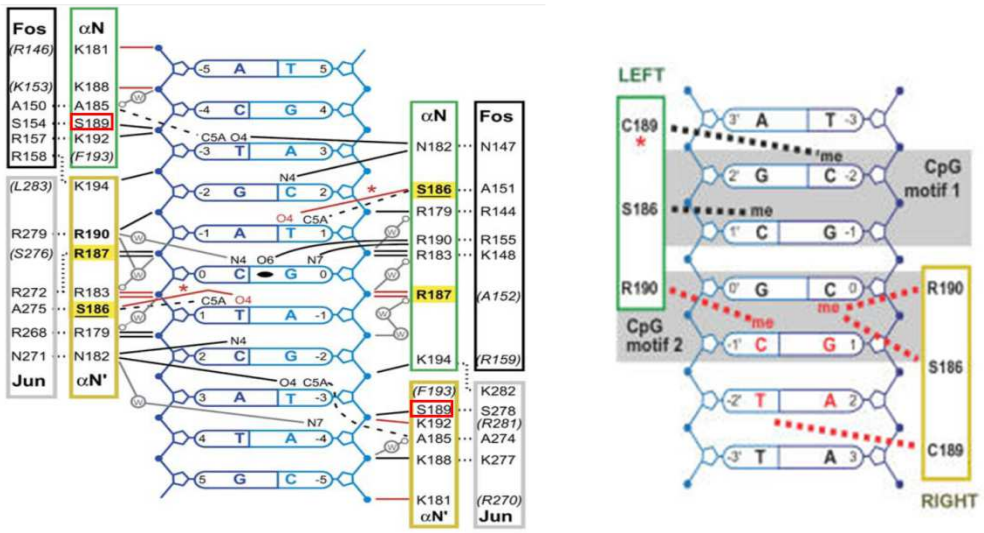


Figure 3.14 Flanking sequence may affect Zta mutant binding to core site.

A The RpZRE3 oligonucleotide sequence, a sequence logo to represent other methylation dependent binding sites and the flanking regions, and the frequency table used to create the sequence logo. B From Petosa et al, 2006, the proposed crystal model of Zta binding to an AP1 site, with mutation C189S in place, and from Karlsson et al 2008, the proposed model for RpZRE3 core methylated binding.

directly adjacent to the core site, and also in the left flank further away, compared to the sequence logo composite of the other sites in which C189S has been observed not to serve to ablate binding. The crystal structure of Zta bound to an AP1 site, solved using the C189S mutant, suggests a contact between S189 and the phosphate backbone between -3 and -4 positions of the centre of the ZRE. This is between the outermost residue of the ZRE, and the

first flanking nucleotide. A change in this nucleotide may have the potential to affect binding ability, based upon the proximity. However, the model produced for Zta binding to methylated RpZRE3 suggests that C189 is required for binding to the methyl-group on CpG motif 1 in the methylated state. Perhaps C189 binds to this methylated group if present, but binds to the phosphate backbone if not. Replacing cysteine with serine would be expected to cause a reduced binding ability, due to the lack of a methyl group, and therefore less binding overall, however some flanking residues may provide a more attractive binding position, stabilising the interaction without direct interaction with the methylated residue. However, C189S did not bind significantly to the unmethylated version of these oligonucleotides either.

In an attempt to elucidate the flank affect on binding, oligonucleotides were designed in which the left and right flanks of RpZRE3 and *CyclinL2* were swapped.

Newly synthesised Zta protein was tested on the pre-methylated probes of RpZRE3 and *CyclinL2* (Figure 3.15). The previously observed difference in binding between wild type and C189S was not as clear in these EMSAs, however it still appeared that C189S bound with a greater affinity to *CyclinL2* (Figure 3.15C). The new oligonucleotides testing the effect of flank on Zta binding required *in vitro* methylation; the C189S effect so far had only been seen using pre-methylated probes direct from Sigma, therefore RpZRE3 and *CyclinL2* were methylated *in vitro* alongside the flank swaps as a control (Figure 3.16). Surprisingly, the binding to C189S was not more than wild type. The flank swap EMSAs appeared to show a small affect of flank on the binding ability,

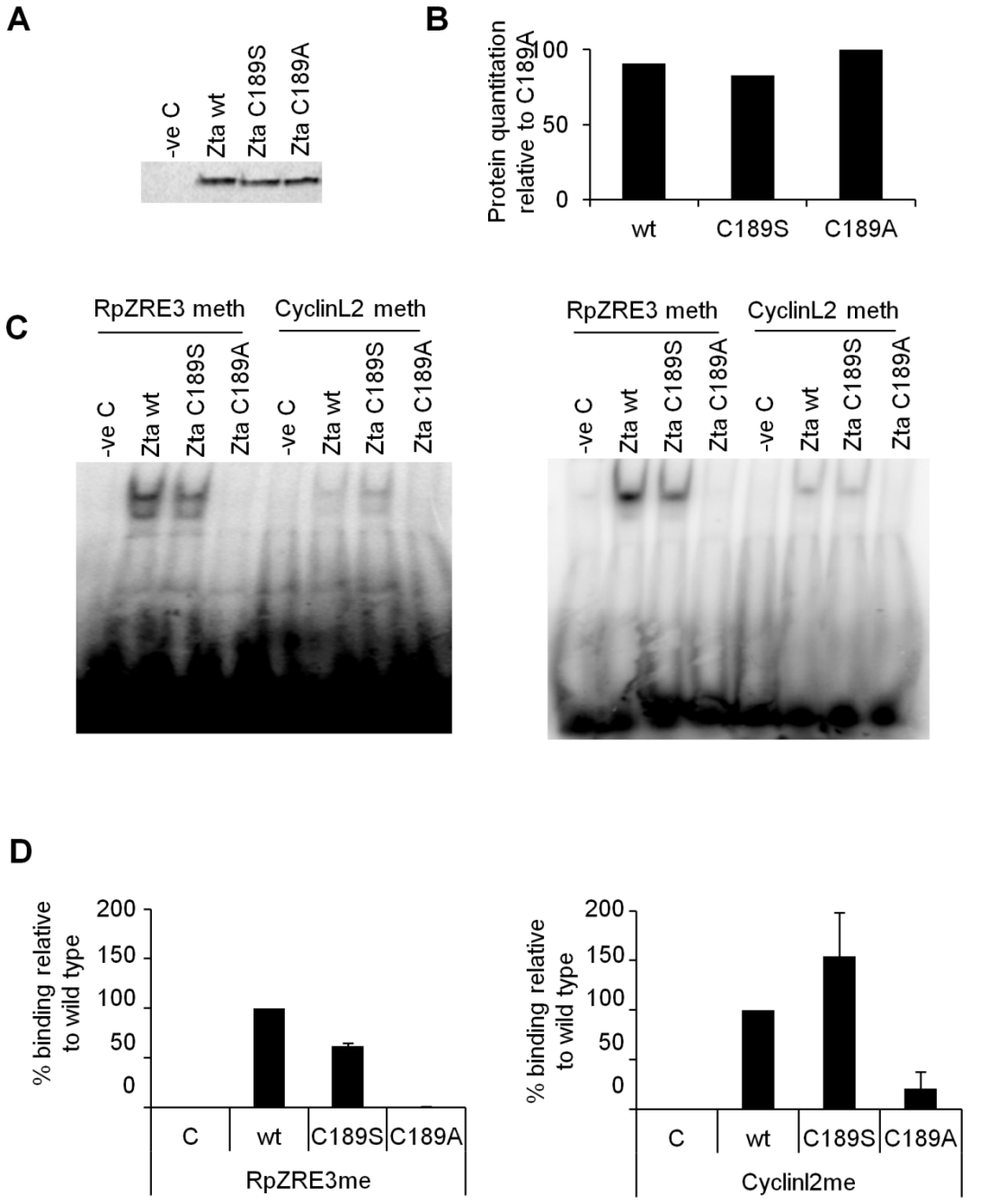


Figure 3.15 Reproduction of C189S increased binding ability is difficult

A new batch of protein synthesised in the presence of S^{35} methionine and quantified in B relative to C189A.C 2 representative EMSAs, to show the reproducibility issues. EMSAs were carried out as previously described with Sigma methylated probes. D Quantitation relative to wild type Zta complex with oligonucleotide.

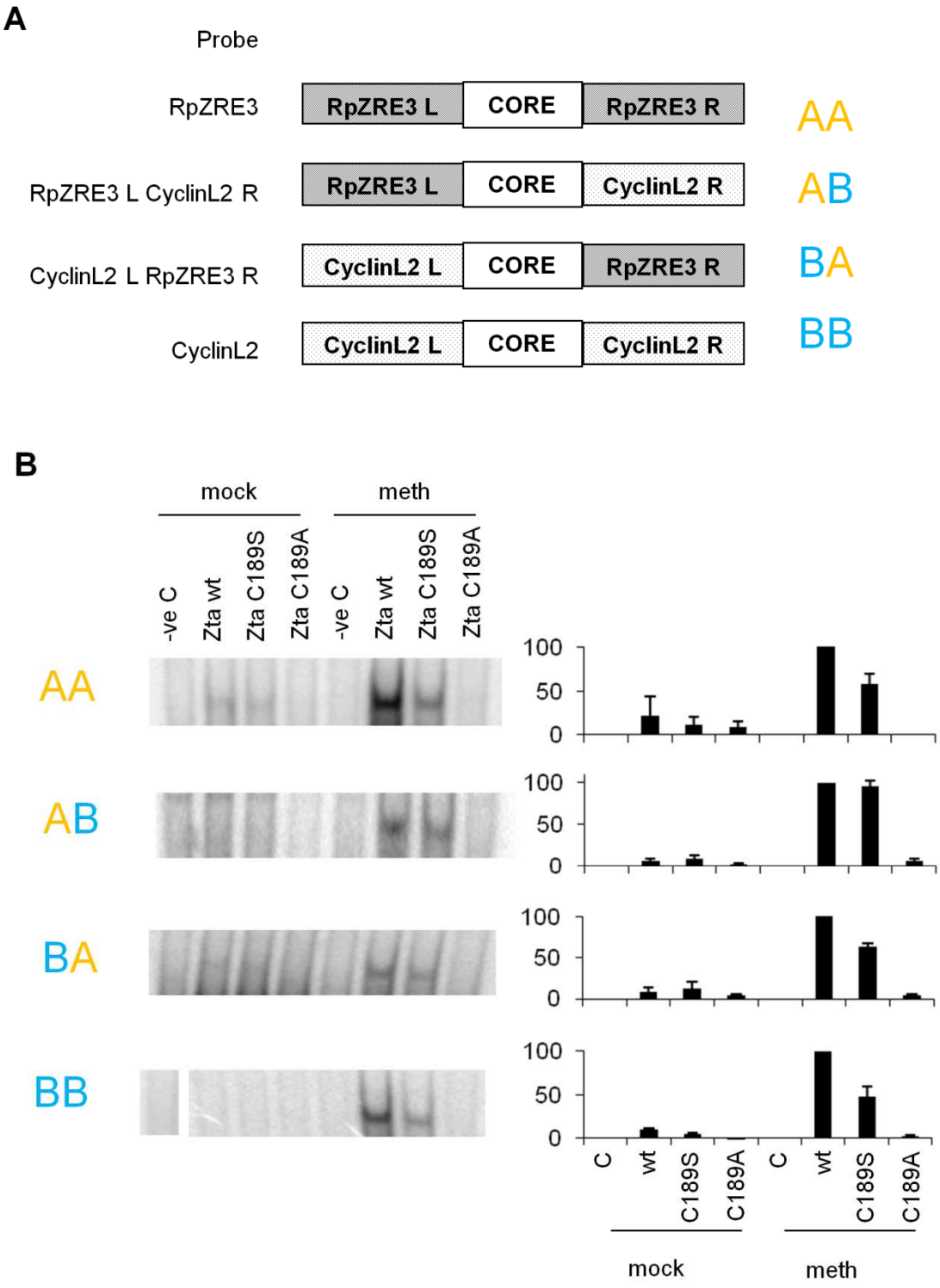


Figure 3.16 Flank affect on C189S binding deficiency in RpZRE3 oligonucleotide appears to reside in the right flank

A Schematic representation of the probes, illustrating the RpZRE3 and CyclinL2 probes, and the flank swaps. B EMSA reactions were carried out as previously described with *in vitro* methylated probes or probes that had undergone an equivalent mock reaction, as previously described. Adjacent is quantitation of complex formation, relative to wild type Zta complexed with methylated oligonucleotide.

with the oligonucleotide containing the right flank of RpZRE3 returning the C189S deficient binding behaviour (Figure 3.16). In order to address the apparent inconsistencies of the binding complex, the reaction conditions were examined. An important aspect of Zta interaction with DNA is that it appears to be required to be reduced, and therefore DTT is added to the reaction (the amount of DTT found in the binding buffer is not sufficient). DTT is labile, but can be stored for long periods of time at -20°C. All EMSAs had been performed using fresh dilutions from a number of stock 1M DTT tubes, all made in 2007. This stock DTT was tested against a new stock of DTT to establish whether this would affect the apparent difference in binding between Zta wild type and C189S (Figure 3.17).

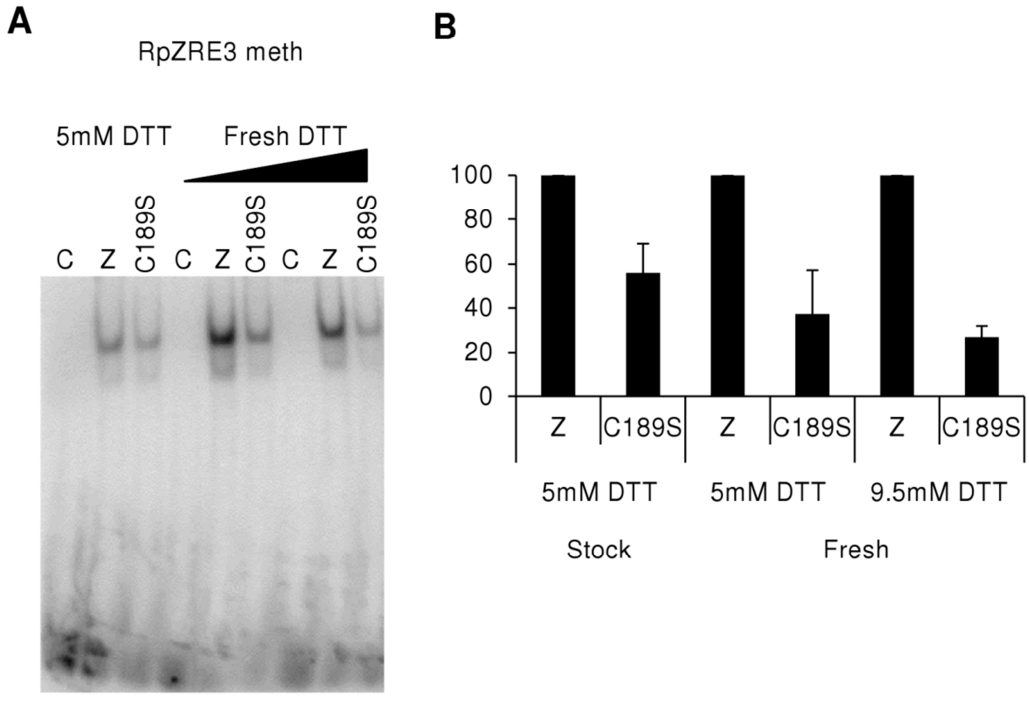


Figure 3.17 Concentration of DTT in the reaction affects relative binding of C189S Zta compared to wild type.

A *In vitro* translated Zta was incubated with labelled RpZRE3 probe with differing stocks or concentrations of DTT. B Quantitation of complex formed is relative to wild type:probe complex for each different DTT condition, to allow C189S comparison.

An increase in DTT, whether due to the fresh stock or the increased amount, increased the binding ability of the wild type protein. This increases the difference between wild type Zta DNA binding and Zta C189S DNA binding. This can be explained by the fact that this mutant has lost the cysteine at which disulphide bonds can form, therefore the increase in DTT affects the complex formation less than the wild type. However, the observable difference in complex formation between wild type Zta and the C189S mutant was not as considerable as previously observed (Karlsson et al., 2008a). The error bars are large, indicating inconsistency, and reproduction of the effect was very difficult.

All EMSAs so far have been performed with Zta protein in vitro translated using wheatgerm extract. Where C189S mutation was shown to have the greatest affect on binding the protein was expressed using rabbit reticulate lysate (Karlsson et al., 2008a). Different translation conditions could affect the oxidation state of the protein overall, so the EMSAs were repeated with proteins expressed using the rabbit reticulate lysate system (Figure 3.18). The binding of the probes tested to C189S appeared to almost completely ablated, and the binding appeared more consistent. Therefore when the conditions are right for wild type Zta to bind maximally, C189S has a reduced not enhanced binding ability. The phenotype *in vivo* is still unclear and which translation system best represents *in vivo* conditions is unknown.

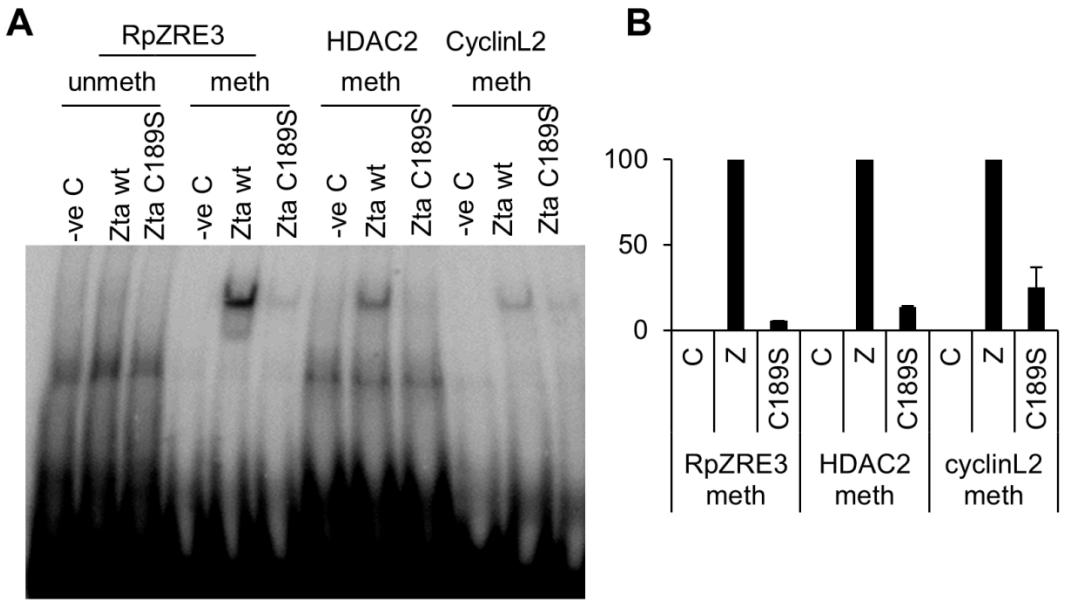


Figure 3.18 Translation system affects relative binding of C189S Zta compared to wild type, in a similar manner to DTT concentration.

A Zta *in vitro* translated in the rabbit reticulate lysate system was incubated with labelled oligonucleotide probes in standard EMSA conditions as previously described. B Quantitation of complex formed is relative to wild type Zta:probe complex for each separate probe used.

3.3. Discussion

A computational search revealed a set of 274 human genes with cellular promoters containing the 7-mer RpZRE3 core element. Flanking sequence can influence the interaction of some transcription factors with DNA; for example, the interaction of the E2F family with DNA is influenced by a region of at least 8 nucleotides 5’ and 11 nucleotides 3’ to the central nucleotide (Tao et al., 1997). Before assigning all 274 genes as candidates for regulation by Zta, it was important to assess whether the cognate flanking sequence had a profound effect on binding by Zta. Consideration of the sequences represented in the flanking sequence of this set of genes revealed it to be diverse; all four nucleotides were represented in 55% of positions and three of the four nucleotides were represented at the remaining positions. Indeed, the immediate flank of the RpZRE3 core element, consisting of four nucleotides both 5’ and 3’, had 100% representation of all four nucleotides. It can therefore be concluded that the flanking sequence does not prevent binding to the methylated RpZRE3 core, and a sequence logo generated from a PWM of the RpZRE3-promoter-18 data-set showed little conservation outside the core motif. In contrast, the interaction with the non-methylated core element initially appeared to be influenced by flanking sequence in only one of the 18 promoters. However, further analysis revealed this to be due to the presence of a second ZRE in the flanking sequence.

These results show that the approach of undertaking a computational pattern match search for the 7-mer RpZRE3 core element is a fast and reliable method to identify promoters containing ZREs that are recognized by Zta only when they are methylated (class III ZREs).

The identification of two adjacent juxtaposed ZREs in the *XPC* promoter presents an interesting problem; can both sites be occupied at once? The natural juxtaposition of ZREs has been previously described for a viral promoter, *BZLF1* and a cellular promoter, *Egr1*. In the *BZLF1* promoter, the elements are situated with 12 base pairs between the central nucleotides; both can be occupied simultaneously and both are functionally relevant (Flemington and Speck, 1990a, Urier et al., 1989, Lieberman et al., 1990). For the *Egr1* promoter, the elements are immediately adjacent with just 8 nucleotides between the central nucleotides (Chang et al., 2006, Heather et al., 2009). There is no evidence for simultaneous occupation by Zta and only one element appears to be functional *in vivo* (Heather et al., 2009). The arrangement of the *XPC* promoter places the centre of the elements 8 nucleotides apart, identical to *Egr1*, and there is no evidence from the DNA-binding experiments for simultaneous occupation of the sites. As one ZRE is recognised in a methylation-dependent manner and the other is recognised when non-methylated, it is possible that the arrangement may provide a fail-safe mechanism to ensure that the gene is regulated in both its methylated and non-methylated states, more likely by a host regulatory factor, the mechanism of which could be taken advantage of by Zta.

The simple computational approach taken to identify the location of this methylation-dependent and novel DNA-binding motif in human gene promoters identified 274 genes potentially regulated by overcoming epigenetic silencing during the viral replicative cycle. Given the known functions of Zta in reprogramming viral gene expression, disrupting cell cycle control and replicating viral DNA, it is interesting to note that the Gene Ontology terms that

are over-represented in the RpZRE3-promoter-274 gene-set are largely involved in transcription, chromatin re-modelling and mitosis. This strongly suggests that Zta may activate this set of cellular genes in order to accomplish these functions. Testing whether these genes are activated during latency disruption in memory B-lymphocytes *in vivo* is technically challenging given both the scarcity of memory B-lymphocytes in peripheral circulation and the infrequency of latency disruption *in vivo*.

The co-location of the RpZRE3 core element in 274 human promoters is unlikely to have been driven by an evolutionary advantage to the virus, but may reflect the involvement of a cellular transcription factor interacting with the same element. This would suggest that this set of genes share a common mode of regulation during human development or differentiation. Furthermore, it would be interesting to question whether co-occurrence of the RpZRE3-core element with other transcription factor binding sites forming a co-operative *cis*-regulator module which may illuminate the regulation of these host and viral genes.

The identification of a novel ZRE in the flank of one of the RpZRE3 core sequence sites led to further investigation regarding the nature of this site, and others like it. Through the TRANSFAC database, it was found to match exactly that of an AP1 site found in the promoter of *PENK*, and the presence of a CpG motif opened the possibility of an effect by methylation status. Other AP1 sites were identified with a CpG motif in the same position of the site, and these also became binding candidates for Zta. All sites were found to bind *in vitro* in a methylation independent manner to varying degrees. It is interesting to consider how methylation of a CpG motif in one sequence can have such a profound

effect, whereas the methylation of a CpG motif in a displaced by one nucleotide position within the site has almost none.

C189S is an interesting Zta mutant, as it appears to play a role in methylation dependent binding, and has been successfully used to show the importance of methylation dependent activation of the *BRLF1* promoter through RpZRE3 (Karlsson et al., 2008a). C189 regulates the redox-sensitivity of DNA-binding activity (Wang et al., 2005). Therefore it was valuable to check how it interacted with the new pattern match sites, to ascertain whether flank has any role to play in binding behaviour of this mutant. The result was surprising, and ultimately unclear. Initially it suggested that the effect of the mutation was reversed; that the replacement of cysteine with serine actually made complex formation stronger with the new RpZRE3 core sequences. However, this only applied to the new sites, as RpZRE3 probe had reduced binding with C189S, although the extent to which the binding was reduced (approx. 50%) was not as much as previously observed (<10%). There was also discrepancy between probes that were methylated in different ways; initially all probes were synthesised by Sigma in the methylated state, however when more probes were required with very slight differences, *in vitro* methylation with M.SssI was used. When the binding of wild type and C189S mutant Zta protein was investigated to these different sets of probes, more conflicting evidence emerged. The DTT concentration seems to play an important role in consistent EMSA complex formation, and this could go a long way to explaining some of the inconsistencies observed. Another element to consider is the different translation systems, and the possibility of Zta protein being more reduced when made in rabbit reticulate lysate compared to wheatgerm. C189S is less affected

by the reduction state due to the loss of cysteine, therefore it is hypothesised that this binding would remain more constant, whereas the increase in DTT would cause wild type Zta increased reduction, and therefore greater binding affinity. The poor reproducibility of these results indicates that more factors are at work than were controlled for. To fully investigate this issue, the assays would need to be controlled in every possible way, something that would be easier using a reliable non-radioactive labelling technique, which has not yet been optimised for Zta:DNA interactions.

Chapter 4: ZRE prediction and mapping in the viral and human genome

4.1. Introduction

The RpZRE_274_dataset from the previous chapter identifies all those human genes with the potential to have their epigenetic silencing reversed by Zta binding to the RpZRE3 core sequence found in the promoter region. However, it only takes into account one methylation dependent site, whereas there are currently four published methylation dependent unique sequence ZREs found in viral promoters, and another found in a human promoter. There is also no evidence to suggest that these are the only methylation dependent sites. Additionally, it would not be possible to say without doubt that a particular gene can only be activated by Zta when methylated; it may contain other non-methylation dependent binding sites in the same region. This chapter aims to provide a fuller picture of the potential of Zta-controlled regulation in both viral and human promoter regions, using and extending techniques developed in the previous chapter.

The infection of human B-lymphocytes by Epstein-Barr virus results in the establishment of a latent state in which a highly restricted set of viral genes are expressed (Rowe et al., 2009). This is accompanied by extensive methylation of CpG motifs in non-expressed viral gene promoter regions (Fernandez et al., 2009, Kalla et al., 2010). In response to physiological stimuli, such as engagement of the B-cell receptor, epigenetic silencing of the viral genome is evaded, resulting in widespread activation of viral gene expression and lytic replication (Bhende et al., 2005, Fernandez et al., 2009). The expression of a subset of host genes is also altered during this period (Chang et al., 2006, Yuan

et al., 2006, Broderick et al., 2009a). As previously discussed, Zta is the key lytic activator protein which switches the virus expression from latency to lytic cycle, and methylation dependent binding is critical to this. Interestingly, methylation of the EBV genome has been shown to be required for lytic reactivation (Kalla et al., 2010). The genome enters the cell during infection in the unmethylated state and gradually becomes methylated over time. The ability of Zta to bind methylated ZREs suggests that it might have a direct role in evading the epigenetic silencing of the viral genome to activate expression from viral promoters required for lytic replication.

An additional function of the methylation requirement for Zta binding could contribute to the establishment of latency during the immortalisation of infected cells. It is known that Zta is expressed transiently during early infection (Halder et al., 2009). However, it does not activate the complete lytic cycle cohort of genes at this early stage. A plausible explanation for this is that key lytic cycle genes are controlled by class III ZREs that do not function unless the CpG ZREs are methylated. Zta has also been shown to be required for immortalisation (Kalla et al., 2010). Thus far the only direct targets of Zta are lytic genes, however perhaps Zta has a further role to play during early infection by activating other genes though different ZREs.

4.2. Results

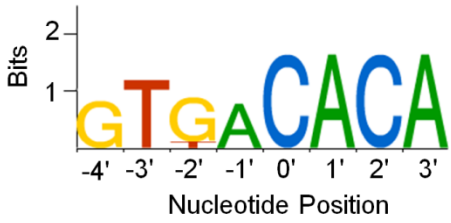
4.2.1 Computational prediction of ZREs core sequences bound by Zta using PROMO

In order to test the accuracy of existing prediction tools and to find novel ZRE core sequences, three well-characterized Zta-responsive promoters from the EBV genome were analysed by James Heather (*BZLF1* promoter [Zp] [Flemington and Speck, 1990a, Lieberman et al., 1990, Urier et al., 1989, Packham et al., 1990], *BRLF1* promoter [Rp] [Bhende et al., 2004, Packham et al., 1990, Lieberman et al., 1990, Sinclair et al., 1991] and *BMRF1* promoter [Quinlivan et al., 1993, Taylor et al., 1991, Kenney et al., 1992]) using PROMO, an online transcription factor binding site (TFBS) prediction tool (Farre et al., 2003, Messeguer et al., 2002). PROMO uses a position frequency matrix (PFM) created from the Transfac 8.3 (Matys et al., 2006) Zta transcription factor entry T00923 (Figure 4.1A). This is made up from six existing ZRE sequences, which match sequences found in the orilyt region, ZRE1, ZRE3, ZRE4, ZRE6 and ZRE5. ZRE3, ZRE4 and ZRE6 are the same core sequences, and are repeated 3 times in this matrix. The identity of the 6^{th} site was not available from the public Transfac database (7.0) and does not match any known ZREs. The search was conducted 500bp upstream of the published transcription start site. PROMO provides dissimilarity rates to score matches, and a positive match was taken as one with a dissimilarity rate of <= 15%.

A

ZRE core sequence	Name
TTACACA	ZRE1
TGACACA	ZRE3
TGACACA	ZRE4
TGACACA	ZRE6
TTGCACA	ZRE5
GAGCCACA	?

A	0	1	0	4	0	6	0	6
C	0	0	0	1	6	0	6	0
G	1	0	4	1	0	0	0	0
T	0	5	2	0	0	0	0	0



B

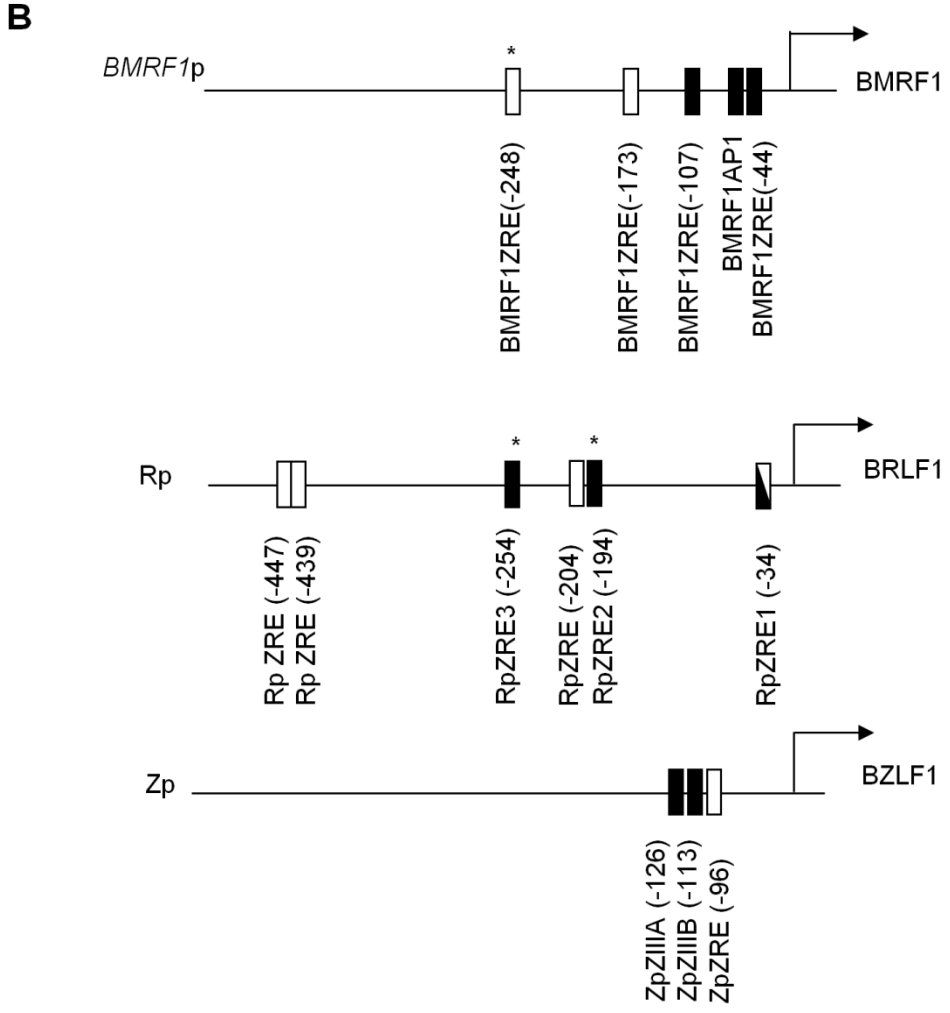


Figure 4.1 ZREs in essential early lytic promoters.

A The ZRE sequences used by PROMO to create a Position Frequency Matrix (PFM), the PFM itself and a sequence logo representation of the prediction matrix. B Previously known ZREs are represented by black boxes, ZREs predicted using PROMO at default settings are represented by black boxes. An asterix (*) indicates the presence of a CpG motif within the ZRE. RpZRE1 is the only existing ZRE that was also predicted by PROMO and is indicated in both black and white.

These three promoters contain eight previously published sites (Zp (2) (Sinclair et al., 1991, Bhende et al., 2004), Rp (3) (Urier et al., 1989, Flemington and Speck, 1990a, Lieberman et al., 1990, Sinclair et al., 1991) and BMRF1p (3) (Kenney et al., 1992, Taylor et al., 1991, Quinlivan et al., 1993). However, the PROMO algorithm only predicted one of these sites (RpZRE1 (Sinclair et al., 1991)) but in addition predicted six novel sites (Figure 4.1B and Table 4.1). The ability of Zta to interact with each predicted site was assessed using electrophoretic mobility shift assays (EMSA), in collaboration with James Heather and final year BSc project students. Conditions were chosen in which *in vitro* synthesized Zta forms readily detectable complexes with previously characterized ZREs, as shown for *BMRF1*ZRE (-44), *BMRF1*ZRE (AP1) and RpZRE2 (-194) (Figure 4.2A). Zta interacted with two of the predicted sites (*BMRF1*ZRE (-173) and ZpZRE (-96)) but not with a DNA probe containing the predicted RpZRE (-439), RpZRE (-447) or RpZRE (-204) sites.

Promoter	Position	ZRE core sequence	EMSA Binding	Methylation-dependent
Zp	-96	TGTGTCT	+	N/A
Rp	-204	TGTGATA	-	N/A
Rp	-439	TGTGTCC	-	N/A
Rp	-447	TGTGTGA	-	N/A
BMRF1	-173	TGGCACA	+	N/A
BMRF1	-248	TGTG**CG**A	+	+

Table 4.1 PROMO predicted sites

The ZRE core sequence predicted is shown, alongside the promoter name, the offset from the published transcription start site, the binding status *in vitro* and the methylation dependence.

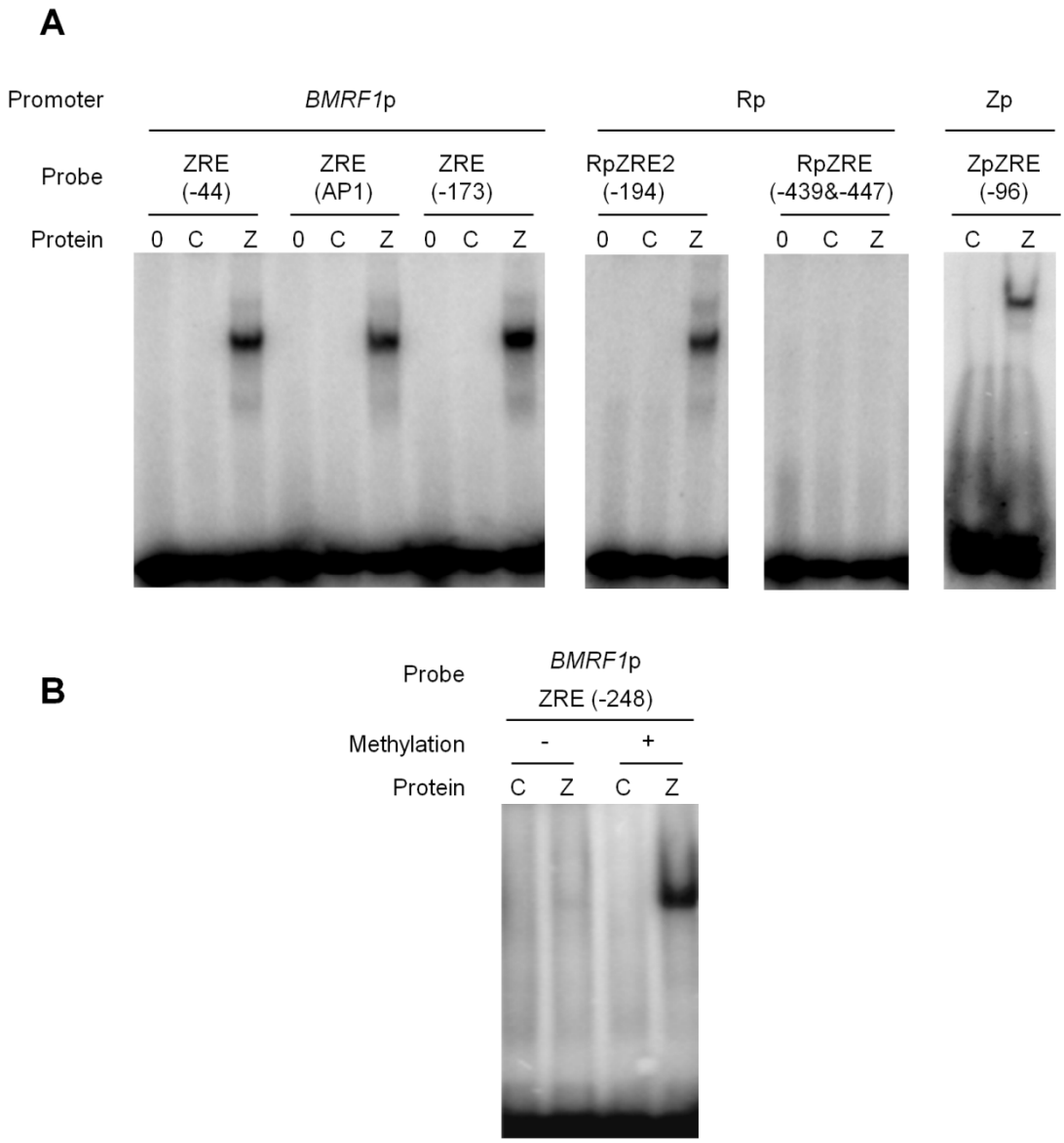


Figure 4.2 Interrogating ZRE PROMO predictions by EMSA.

A Double stranded oligonucleotides were generated with the core ZRE sequence and at least 10 nucleotides of cognate sequence on either side. Following radiolabelling, these were incubated with *in vitro* translated Zta and subject to EMSA. The reactions contained no protein, 0, control lysate, C, or Zta, Z. B Double stranded oligonucleotides were generated with the core ZRE sequence and 10 nucleotides of cognate sequence on either side. Following radiolabeling, these were subject to *in vitro* methylation with M.SssI methyl transferase, or a mock reaction. Subsequently, they were incubated with *in vitro* translated Zta and subjected to EMSA. The reactions contained control lysate, C, or Zta, Z. EMSAs in this figure were performed by undergraduate students, with the exception of ZpZRE(-96).

One predicted site in *BMRF1*p contains a CpG motif (BMRF1ZRE (-248)). Zta only displays a marginal interaction with the site (6-10 fold less than another ZRE from the same promoter) in its non-methylated form (Figure 4.2B). In contrast, complex formation was dramatically enhanced following *in vitro* methylation of the CpG motif, identifying it as a methylation-dependent class III ZRE.

Of the PROMO predicted sites, four were found to bind (one was a previously known ZRE, one of these in a methylation dependent manner), but the prediction tool missed six known ZREs.

4.2.2 Testing a novel ZRE PFM to predict CpG containing ZREs

A position frequency matrix (PFM) was generated using the CpG-containing ZRE core binding sites (denoted PFM_{CpG5}) (Figure 4.3). The accuracy of the matrix was evaluated by searching for ZREs in the three well-characterised viral promoters, using MatScan (Blanco et al., 2006) in collaboration with James Heather (University of Sussex). The PFM_{CpG5} identified all five CpG containing sites (used to create the PFM) and predicted two novel sites; one located in Rp, centred on -114 (RpZRE (-114)), and one located in the *BMRF1* promoter, centred on -148 (BMRF1ZRE (-148)). DNA binding experiments demonstrate that Zta interacts with both sites in a methylation-dependent manner, characteristic of class III ZREs (Figure 4.4).

4.2.3 Optimisation of the prediction matrix

The discovery of these novel ZREs invites the question as to whether inclusion of the new sites would improve the prediction capabilities. A new PFM including

the two new CpG containing ZREs discovered in Rp and BMRF1p, as well as an

A

ZRE core Sequence	Name
TGAGCGA	RpZRE2
TTCGCGA	RpZRE3
TGAGCGT	NapZRE2
CGGGCGA	NapZRE1
TGTGCGA	BMRF1p (-248)

B

	-3'	-2'	-1'	0	1'	2'	3'
A	0	0	2	0	0	0	4
C	1	0	1	0	5	0	0
G	0	4	1	5	0	5	0
T	4	1	1	0	0	0	1

C

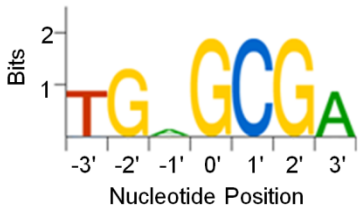


Figure 4.3 Generation of PFM_{CpG5}

A 5 CpG containing ZREs, found in viral promoters B A Position Frequency Matrix (PFM) showing the frequency of each base at a given position. This was supplied to MatScan to predict new CpG-containing sites. C A sequence logo, graphical representation of the motif, generated using WebLogo.

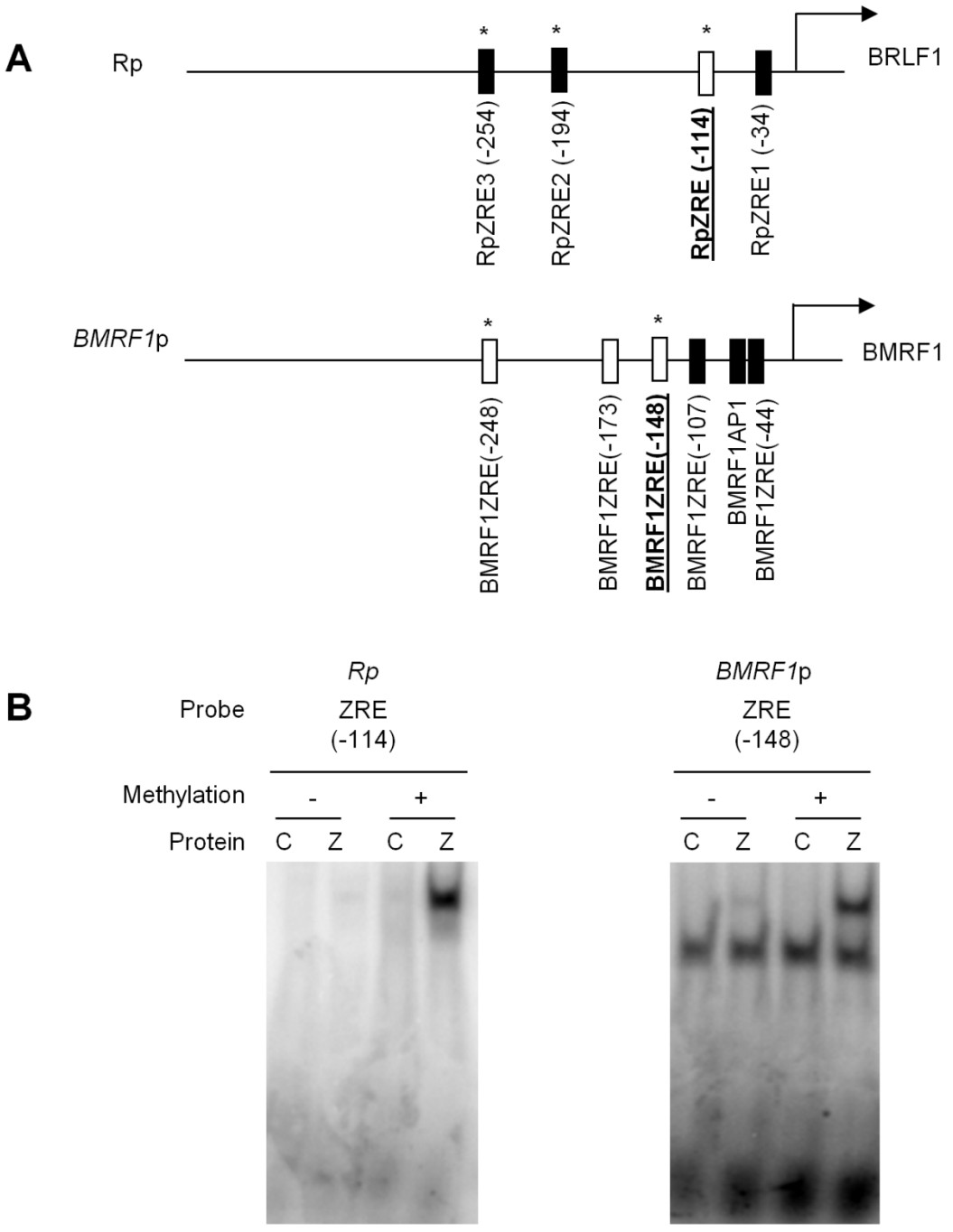


Figure 4.4 Newly predicted CpG-containing ZREs in the promoters of BRLF1 and BMRF1.

A The PFM_{CpG5} was used in MatScan to predict further CpG containing ZREs. These newly predicted sites are indicated in the promoter maps by white boxes, with the name in bold and underlined. B Double stranded oligonucleotides were generated with the core predicted ZRE sequence and 10 nucleotides of cognate sequence on either side. Following radiolabeling, these were subject to *in vitro* methylation with M.SssI methyl transferase, or a mock reaction. Subsequently they were incubated with *in vitro* translated Zta and subjected to EMSA. The reactions contained control lysate, C or Zta, Z.

existing methylation dependent site found in the promoter of Egr1 (Chang, Lee et al. 2006), was created and denoted PFM_{CpG8}.

In order to compare these matrices, both PFM_{CpG5} and PFM_{CpG8} were used in MatScan to search the entirety of the Type 1 EBV genome (Ref Seq NC_007605) (Figure 4.5). Core sequences were identified and duplicates due to multiple instances were removed, creating two lists of unique core sequence predictions. This was compared to the eight previously identified CpG containing ZREs. PFM_{CpG5} predicted a total of 22 unique core sequences, eight of which matched the previously identified CpG containing ZREs, and 14 novel sequence predictions. PFM_{CpG8} predicted a total of 13 unique sequences, six of which matched the previously identified CpG containing ZREs, and seven novel sequence predictions, all of which were predicted by PFM_{CpG5}. I questioned which list was more accurate. PFM_{CpG8} missed the RpZRE3 and NapZRE1 sequences in the prediction, despite the fact these are contained within the PFM. It was therefore decided to continue the investigation using the predictions from PFM_{CpG5}, as this matrix did not miss any of the known sites.

4.2.4 *In vitro* confirmation of predicted binding site sequences

The 14 novel discrete ZRE core sequences, denoted CpG5A-N (Table 4.2), predicted by PFM_{CpG5} in the Type 1 EBV genome were tested by EMSA in both the unmethylated and methylated state (Figure 4.6). All predicted core ZRE sequences bound in the methylated form, except CpG5 H and N. No sites bound significantly in the unmethylated form, except CpG B. Therefore 11 out of 14 predictions can be classified as Class III ZREs, and one can be classified as Class II. Overall, the prediction matrix PFM_{CpG5} has predicted a total of 22

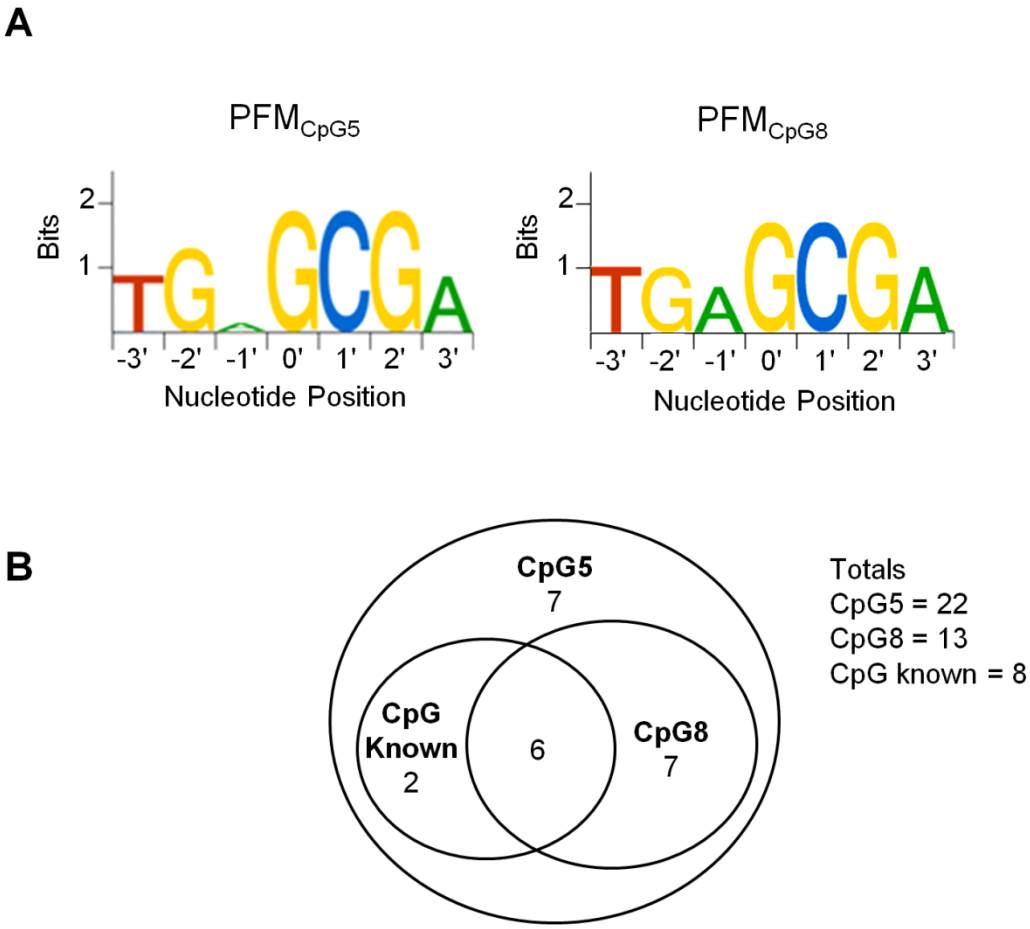


Figure 4.5 Attempt to improve CpG containing ZRE prediction

A Sequence logo motif representations of the PFM_{CpG5} and PFM_{CpG8} to illustrate the effect of three additional sites on the motif appearance. B A Euler diagram comparing the predictions made using PFM_{CpG5} and PFM_{CpG8} .

Name	Predicted ZRE core sequence	score
A	CGAGCGA	0.89
B	TGAGCGC	0.86
C	TTGGCGA	0.86
D	TTAGCGA	0.89
E	TGTGCGT	0.86
F	TTTGCGA	0.86
G	TCAGCGA	0.86
H	TGCGCGT	0.86
I	CGTGCGA	0.86
J	TGCGCGA	0.96
K	CGCGCGA	0.86
L	AGAGCGA	0.86
M	TGGGCGA	0.96
N	TGGGCGT	0.86

Table 4.2 Predicted ZRE core sequences

ZRE predictions from PFM_{CpG5} used in MatScan on the entire EBV genome, together with their respective score, a numerical representation of the differences between the predicted site and the max score possible from PFM_{CpG5} .

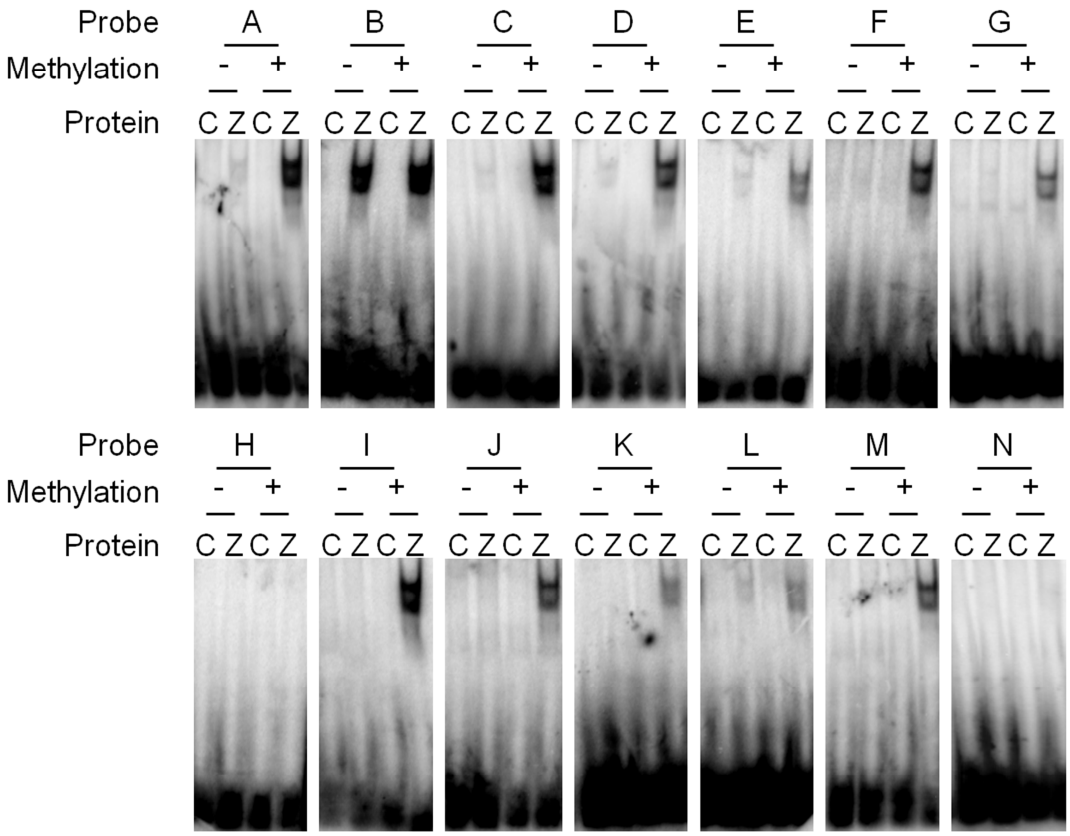


Figure 4.6. Confirmation of Zta recognition and methylation dependence of PFM_{CpG5} predicted CpG containing core ZREs.

Double strand oligonucleotides were generated with the core predicted ZRE sequence and 10 nucleotides of cognate sequence on either side. Following radiolabeling, these were subject to *in vitro* methylation with SssI methyl transferase (+), or a mock reaction (-). Subsequently, they were incubated with *in vitro* translated Zta and subject to EMSA. The reactions contained control lysate, C; or Zta, Z.

unique sequences, 20 of which can be seen to form complexes *in vitro*, a 91% successful prediction tool.

4.2.5 Genomic position prediction of ZRE core binding sequences in the EBV genome

Global predictions for the 171,823 nucleotide type 1 Epstein-Barr virus genome (Ref Seq NC_007605) was undertaken by bioinformatician D. Thomas (BSMS), using an exact pattern match with 32 validated ZRE sequences (Table 4.3, Figure 4.7). This revealed 469 locations within the genome that matched one of the 32 ZRE core sequences (Appendix II).

The occurrence of ZREs in each promoter regulatory region was analysed. The promoters of the EBV genes was taken as between 500 base pairs upstream and 200 base pairs downstream because the annotated gene start sites are known not to correlate fully with known mRNA start sites. This revealed that 79 out of 86 EBV regulatory regions contained at least one ZRE core sequence (Table 4.3). The occurrence of ZREs in 92% of EBV regulatory regions suggests that Zta has the potential to regulate the expression of most EBV genes. Furthermore, 58 EBV regulatory regions contained at least one CpG containing ZRE (Table 4.3), potentially controlling the expression of 41 lytic genes. These regions are all methylated during latency (Fernandez et al., 2009, Kalla et al., 2010), allowing methylation-dependent Zta interaction with ZREs that could influence the expression of a broad range of EBV genes once Zta is expressed during lytic cycle.

ZRE Core Sequence		Class	Names	References
Forward	**Reverse**			
TGAGCCA	TGGCTCA	I	Zp ZREIIIA	(Flemington and Speck, 1990a, Lieberman and Berk, 1990) (Urier et al., 1989)
			Rp ZRE1	(Sinclair et al., 1991)
			Zp (-365)	This chapter
TTAGCAA	TTGCTAA	I	Zp ZREIIIB	(Flemington and Speck, 1990a, Lieberman and Berk, 1990) 16
TGTGTAA	TTACACA	I	DSL ZRE1	(Lieberman and Berk, 1990)
TGAGCAA	TTGCTCA	I	DSL ZRE2	(Lieberman and Berk, 1990)
			DSL ZRE7	(Lieberman and Berk, 1990)
			*BMRF1*ZRE(-44)	(Taylor et al., 1991)
			*BMRF1*ZRE(-107)	(Quinlivan et al., 1993)
TGTGTCA	TGACACA	I	DSL ZRE3	(Lieberman and Berk, 1990)
			DSL ZRE4	(Lieberman and Berk, 1990)
			DSL ZRE6	(Lieberman and Berk, 1990)
			DHRS9 ZRE1	(Jones et al., 2007)
TGTGCAA	TTGCACA	I	DSL ZRE5	(Lieberman and Berk, 1990)
			CIITA (221)	(Li et al., 2009)
TGAGTCA	TGACTCA	I	*BSLF2+BMLF1*	(Taylor et al., 1991)
			BMRF1 AP-1	(Kenney et al., 1992) (Quinlivan et al., 1993)
			IL8 AP1	(Hsu et al., 2008)
			DHRS9 (ZRE2)	(Jones et al., 2007)
TGACTAA	TTAGTCA	I	Fp AP-1-Like Site	(Zetterberg et al., 2002)
TGTGTCT	AGACACA	I	Zp (-96)	This chapter
TGGCACA	TGTGCCA	I	*BMRF1* (-173)	This chapter
TGAGTAA	TTACTCA	I	*CIITA*	(Li et al., 2009)
			IL13	(Tsai et al., 2009)
GTTGCAA	TTGCAAC	I	*IL-8 ZRE*	(Hsu et al., 2008)
GGAG**CG**A	T**CG**CTCC	III	*Egr1*	(Heather et al., 2009)

TGAG**CG**A	T**CG**CTCA	II	Rp ZRE2	(Sinclair et al., 1991) (Bhende et al., 2004)
TT**CGCG**A	T**CGCG**AA	III	Rp ZRE3	(Bhende et al., 2004)
TAAG**CG**A	T**CG**CTTA	III	*BRLF1* (-114)	This chapter
TGAG**CG**G	C**CG**CTCA	III	*BMRF1* (-148)	This chapter
TGTG**CG**A	T**CG**CACA	III	*BMRF1* (-248)	This chapter
TGAG**CG**T	A**CG**CTCA	III	Nap ZRE2	(Dickerson et al., 2009)
CGGG**CG**A	T**CG**CCCG	III	Nap ZRE1	(Dickerson et al., 2009)
CGAGCGA	TCGCTCG	III	CpG5 A	This chapter
TGAGCGC	GCGCTCA	II	CpG5 B	This chapter
TTGGCGA	TCGCCAA	III	CpG5 C	This chapter
TTAGCGA	TCGCTAA	III	CpG5 D	This chapter
TGTGCGT	ACGCACA	III	CpG5 E	This chapter
TTTGCGA	TCGCAAA	III	CpG5 F	This chapter
TCAGCGA	TCGCTGA	III	CpG5 G	This chapter
CGTGCGA	TCGCACG	III	CpG5 I	This chapter
TGCGCGA	TCGCGCA	III	CpG5 J	This chapter
CGCGCGA	TCGCGCG	III	CpG5 K	This chapter
AGAGCGA	TCGCTCT	III	CpG5 L	This chapter
TGGGCGA	TCGCCCA	III	CpG5 M	This chapter

Table 4.3 ZRE_{32}

The 32 ZREs characterised by EMSA or DNase footprinting, including the novel ZREs discovered in this work.

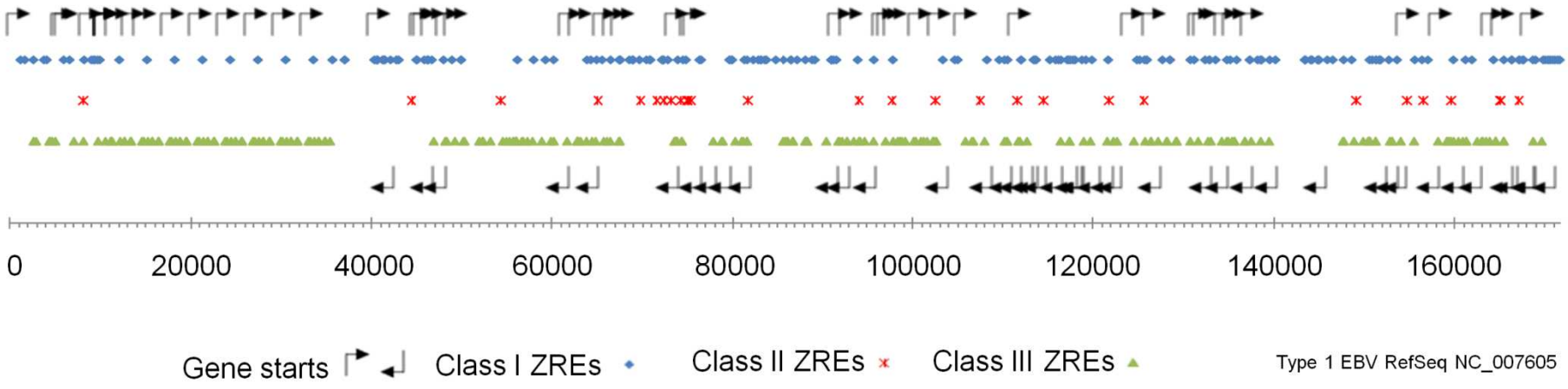


Figure 4.7 Genome-wide map of core ZREs

The entire EBV genome was subjected to an exact pattern match search, using the ZRE_{32} set of sequences. Each site was classed by binding behaviour, and plotted by the first nucleotide of the site to form a genome wide map of core ZREs. Class I sites are indicated by blue diamonds, Class II sites are indicated by red stars, Class III sites are indicated by green triangles, and gene starts and direction are indicated by arrows.

Interestingly 22 EBV gene regulatory regions contain only CpG ZRE binding sequences (highlighted in Table 4.4). Those genes are prime contenders to be regulated in a strictly methylation-dependent manner by Zta, as binding will primarily occur when the promoter region is methylated. These genes were originally classified as displaying early lytic, late lytic and latent patterns of gene expression, but importantly, a more recent genome wide expression study shows that all are upregulated during lytic cycle in Akata BL cells, with the majority reaching a peak expression value approximately 24 hours after lytic expression (Yuan et al., 2006).

cycle phase	Promoter name	Class I	Class II	Class III	Total
late lytic	*BNRF1*	1	0	0	1
unknown	*EBER 1*	1	0	0	1
unknown	*EBER 2*	1	0	1	2
late lytic	*BCRF1*	3	0	1	4
latent	*Cp*	0	0	3	3
unknown	*BWRF1*	0	0	3	3
latent	*Wp*	0	0	1	1
early lytic	*BHLF1*	4	0	0	4
early lytic	*BHRF1*	4	0	0	4
early lytic	*BFLF2*	2	0	0	2
unknown	*BFRF1A*	4	0	0	4
early lytic	*BFLF1*	4	0	0	4
early lytic	*BFRF1*	4	0	0	4
early lytic	*BFRF2*	0	0	0	0
late lytic	*BFRF3*	1	0	1	2
latent	*Qp*	2	0	2	4
unknown	*BPLF1*	1	0	2	3
late lytic	*BORF1*	0	0	3	3
unknown	*BOLF1*	0	0	2	2
early lytic	*BORF2*	1	0	1	2
early lytic	*BaRF1.1*	1	0	2	3
early lytic	*BMRF1*	4	0	2	6
late lytic	*BMRF2*	1	0	0	1
early lytic	*BSLF2/BMLF1*	4	1	0	5
early lytic	*BSLF1*	2	1	0	3
late lytic	*BSRF1*	1	2	2	5

late lytic	*BLRF1*	1	0	0	1
early lytic	*BLLF3*	2	0	0	2
late lytic	*BLRF2*	2	0	0	2
early lytic	*BLLF2*	0	0	1	1
late lytic	*BLLF1*	2	0	1	3
late lytic	*BZLF2*	0	0	0	0
early lytic	*BZLF1*	4	0	0	4
early lytic	*BRRF1*	0	0	2	2
early lytic	*BRLF1*	1	1	2	4
late lytic	*BRRF2*	1	1	2	4
late lytic	*BKRF2*	0	1	1	2
unknown	*BKRF3*	1	1	1	3
early lytic	*BKRF4*	0	0	3	3
late lytic	*BBRF1*	0	0	2	2
early lytic	*BBLF4*	0	0	3	3
late lytic	*BBRF2*	1	0	0	1
late lytic	*BBRF3*	0	0	2	2
early lytic	*BBLF2/BBLF3*	0	0	1	1
late lytic	*BBLF1*	0	0	0	0
early lytic	*BGLF5*	2	0	3	5
early lytic	*BGLF4*	0	1	1	2
unknown	*BGLF3.5*	2	0	2	4
late lytic	*BGRF1/BDRF1*	0	0	1	1
unknown	*BGLF3*	0	0	0	0
late lytic	*BGLF2*	0	1	0	1
late lytic	*BGLF1*	2	0	1	3
early lytic	*BDLF4*	1	0	0	1
unknown	*BDLF3.5*	3	0	2	5
late lytic	*BDLF3*	2	0	1	3
late lytic	*BDLF2*	0	0	0	0
late lytic	*BDLF1*	0	0	1	1
early lytic	*BcRF1.2*	2	0	0	2
late lytic	*BcLF1*	1	1	1	3
late lytic	*BTRF1*	0	0	1	1
late lytic	*BXLF2*	1	0	0	1
late lytic	*BXRF1*	1	0	1	2
early lytic	*BXLF1*	2	0	2	4
early lytic	*BVRF1*	2	0	2	4
late lytic	*BVRF2*	1	0	3	4
unknown	*BVLF1*	1	0	2	3
late lytic	*BdRF1*	1	0	0	1
late lytic	*BILF2*	4	0	1	5
unknown	*RPMS1*	2	0	1	3
early lytic	*LF3*	2	0	0	2
unknown	*LF2*	1	0	4	5

unknown	*LF1*	3	0	0	3
unknown	*BILF1*	1	0	3	4
unknown	*A73*	1	0	1	2
early lytic	*BALF5*	1	1	0	2
late lytic	*BALF4*	0	0	3	3
unknown	*BARF0*	0	0	2	2
unknown	*BALF3*	1	0	3	4
early lytic	*BALF2*	1	0	3	4
early lytic	*BALF1*	0	2	2	4
early lytic	*BARF1.2*	0	2	1	3
latent	*LMP-2A*	1	0	0	1
early lytic	*BNLF2b*	2	1	0	3
early lytic	*BNLF2a*	2	1	0	3
latent	*LMP-1*	1	0	0	1
latent	*LMP-2B*	1	0	0	1

Table 4.4 ZRE predictions in all EBV genes. Those genes with only CpG-containing ZREs in the promoter region are highlighted.

The three AP1 like-sites discussed in Chapter 4 were omitted from the ZRE list, due to unconvincing binding data regarding their methylation dependency at the time of analysis. However, when they are included in retrospect, these three core sequences are found a total of 28 times within the EBV genome, increasing the total number of ZRE matches found to 497. Two genes which were previously found to contain no ZREs, *BBLF1* and *BGLF2*, contain at least one of these AP1 like sites. Their inclusion does not significantly affect the list of genes which are contenders for methylation dependent regulation, as these sites can be bound in their unmethylated state, therefore Zta is not prevented from binding. (Full table can be found in appendix III).

The 22 predicted methylation dependent Zta responsive regions were examined to identify ChIP candidates, and primers were designed for the CpG-ZREs in three gene promoters: *BKRF4, BGLF4* and *BTRF1*. The position of the ZREs and the region amplified by the qPCR primers is shown in figure 4.8A. Lytic

cycle was activated by anti-IgG treatment in Akata cells and acyclovir treatment was used to inhibit replication, to retain the methylation pattern of the unreplicated genome. Chromatin was harvested at the 48 hour time point and subjected to Zta immunoprecipitation in conjunction with S. Ramasubramanyan (University of Sussex). QPCR was undertaken and the enrichment relative to percent input is shown in figure 4.8B, successfully demonstrating the ability of Zta to bind to these novel ZREs at the onset of replication. Zta binds to these regions to a comparable amount to that seen in the Zp (Ramasubramanyan, S., unpublished data).

4.2.6 Integrating ChIP Seq data with positional ZRE predictions

Bergbauer et al performed a ChIP Seq experiment using a mutant version of Zta, a GFP-tagged protein lacking the transactivation domain (residues 149-245) (Bergbauer et al., 2010), referred to hereafter as Zta^{MUT}. The reasoning behind this was to identify only regions of the genome where Zta bound directly to the DNA through the binding domain, and not indirectly through interactions with other proteins through the transactivation domain. However this approach does not recognise the potential for the transactivation domain to affect the DNA binding ability of Zta. The basepair resolution data was kindly shared with us for the 4 conditions they used; *in vitro* unmethylated, the 165 kbp genomic DNA of EBV B95.8 strain cloned in E.coli and incubated with purified Zta^{MUT}, *in vitro* methylated, the same cloned DNA but incubated with purified Zta^{MUT} after it had undergone CpG methylation by de novo methyltransferase M.SssI, *in vivo* Raji, a cell line which is strictly latent, and *in vivo* B95.8, a cell line semi-permissive for lytic cycle. Both *in vivo* cell lines stably expressed the Zta^{MUT}.

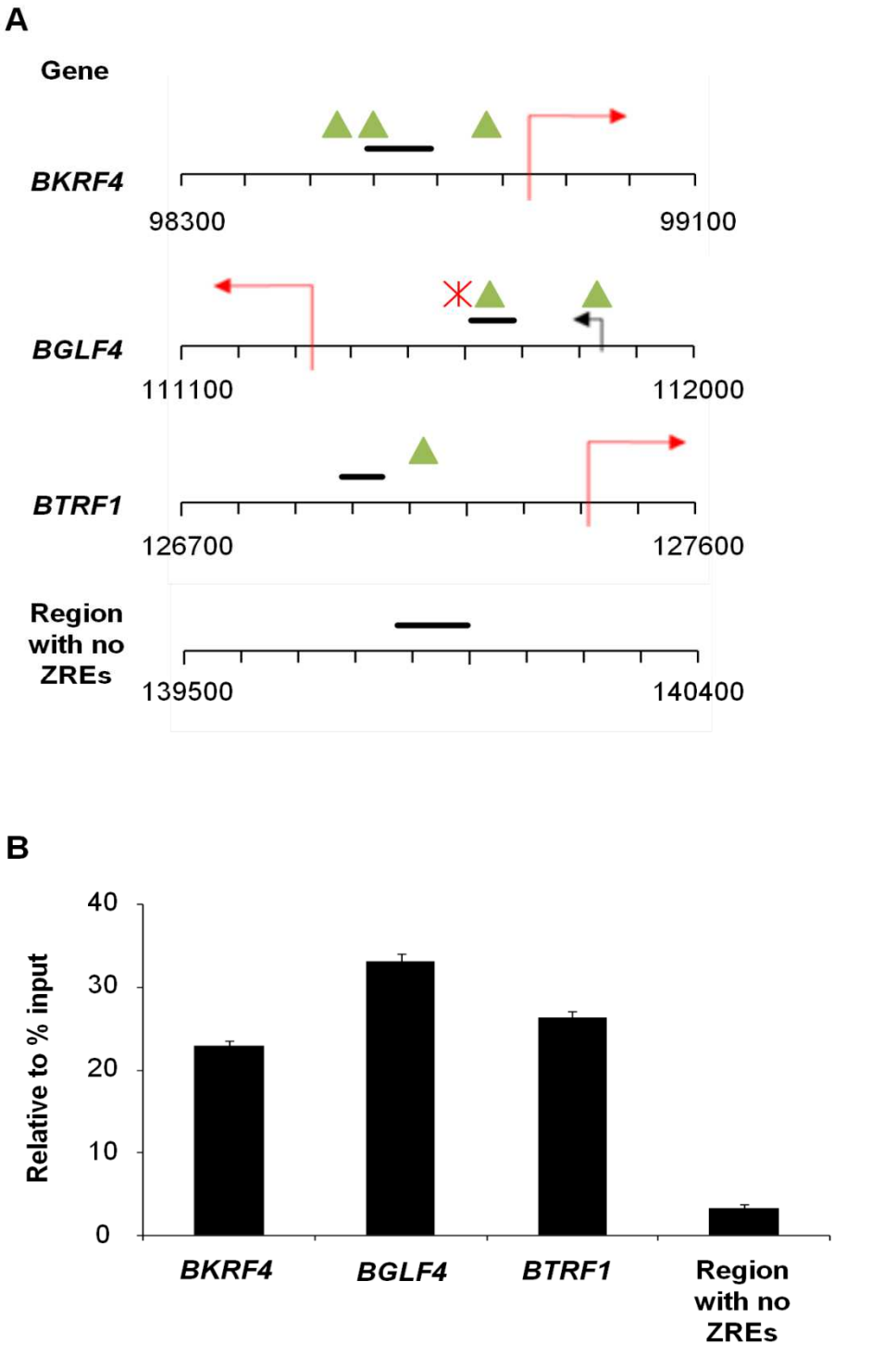


Figure 4.8 ChIP Analysis of candidate genes.

A Regions of EBV selected for Chromatin Immunoprecipitation (ChIP). Green triangles indicate Class III ZREs, red stars indicate Class II ZREs. B Quantitation by QPCR shown relative to % input.

The *in vitro* unmethylated and methylated binding profiles give us an idea of the potential of Zta binding in the absence of other viral or cellular proteins, whereas the *in vivo* data provides a snapshot of binding in a cellular environment.

The ChIP Seq data was aligned to the 172kbp B95.8 EBV genome (Baer et al., 1984) which has some differences to the Type 1 genome sequence used so far, containing a different number of W repeats and a deleted region over the LF1, 2 and 3 genes, and orilyt right. An exact pattern match of the ZRE_{32} sequences was performed on this sequence, revealing a total of 443 ZRE positional predictions. To ensure gene mapping consistency between the genomes, each regulatory region of interest was aligned with the B95.8 sequence using BLAST. The ChIP Seq data and the ZRE positional predictions were then plotted for the specified promoter regions. Initially, the three well characterised promoters, previously used in the chapter to initially predict ZREs (*BRLF1, BMRF1 and BZLF1*) were plotted to question these predictions (Figure 4.9). Rp shows peaks both *in vitro* and *in vivo* over RpZRE1 and RpZRE2, and less obvious “shoulders” of binding peaks over RpZRE (-114) and RpZRE3. There is a considerable increase in binding in the methylated *in vitro* experiment, compared to the unmethylated *in vitro* experiment.. *BMRF1*p shows a broad peak of binding across all the sites both *in vitro,* where it appears to be methylation dependent, and *in vivo.* Zp exhibits a small binding peak over ZpZREIII A and B, which are slightly more pronounced *in vitro*, but none over a ZpZREIIIA pattern match, denoted RpZRE(-365). To question why there was no peak, I undertook EMSAs to assess *in vitro* binding of Zta to a double stranded

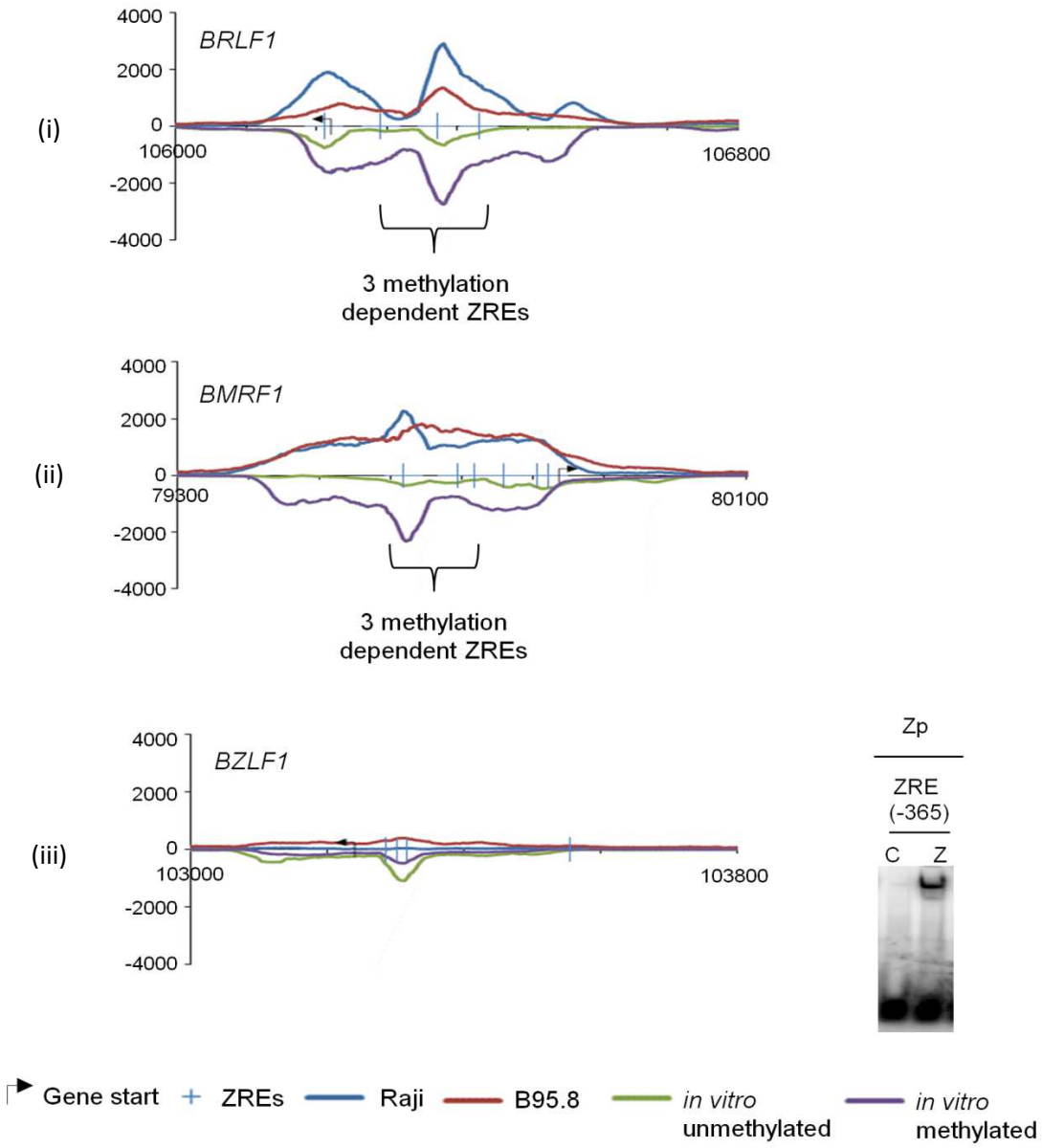


Figure 4.9 Integration of Bergbauer et al ChIP Seq data with ZRE position predictions: the three well characterised promoters.

The Zta^{MUT} ChIP Seq data and the ZRE position predictions were extracted for each previously defined promoter region of the 3 well defined promoters initially investigated, and plotted. Blue lines on the x axis indicate ZRE position. The *in vitro* data is plotted as negative for the sole purpose to separate the data, and should be read as equivalent binding as above the x axis. The ZREs from left to right are (i) Rp: RpZRE1, RpZRE(-114), RpZRE2, RpZRE3 (ii) *BMRF1*p: BMRF1ZRE(-248), BMRF1(-173), BMRF1(-148), BMRF1p(-107), BMRF1 AP-1, BMRF1(-44) (iii) Zp: ZpZRE(-96), ZpZREIIIB, ZpZREIIIA, ZpZRE(-365). ZpZRE(-365) is an exact match to the core sequence of ZpZREIIIA, and the ability of Zta to bind in this context was confirmed by EMSA, as shown adjacent to the Zp graph. Approximately 500bp upstream and 200bp downstream is shown.

oligonucleotide consisting of the core sequence, with 10 basepairs of cognate sequence on both flanks. Binding was confirmed in this context, suggesting that the inability of Zta^{mut} to bind in the ChIP experiments is not due to the immediate surrounding sequence.

The 22 regulatory regions containing only CpG ZREs were also plotted. BALF1, BARF1.2, BBLF4 and BGLF2 all showed specific methylation dependent binding *in vitro* that aligned with at least 1 of the predicted CpG-containing ZREs, and also exhibited strong *in vivo* binding (Figure 4.10). BBLF2/BBLF3, BBRF3, BKRF2 and BKRF4 showed a small but specific methylation dependent binding *in vitro*, and some *in vivo* binding (Figure 4.11). BALF4, BARF0, BGLF4, BOLF1, BORF1, BTRF1 and Cp showed some specific methylation dependent binding *in vitro* but none *in vivo* (Figure 4.12). BBRF1, BDLF1, BGRF1/BDRF1, BLLF2, BRRF1, BWRF1 repeats and Wp showed no increase in Zta binding (not shown).

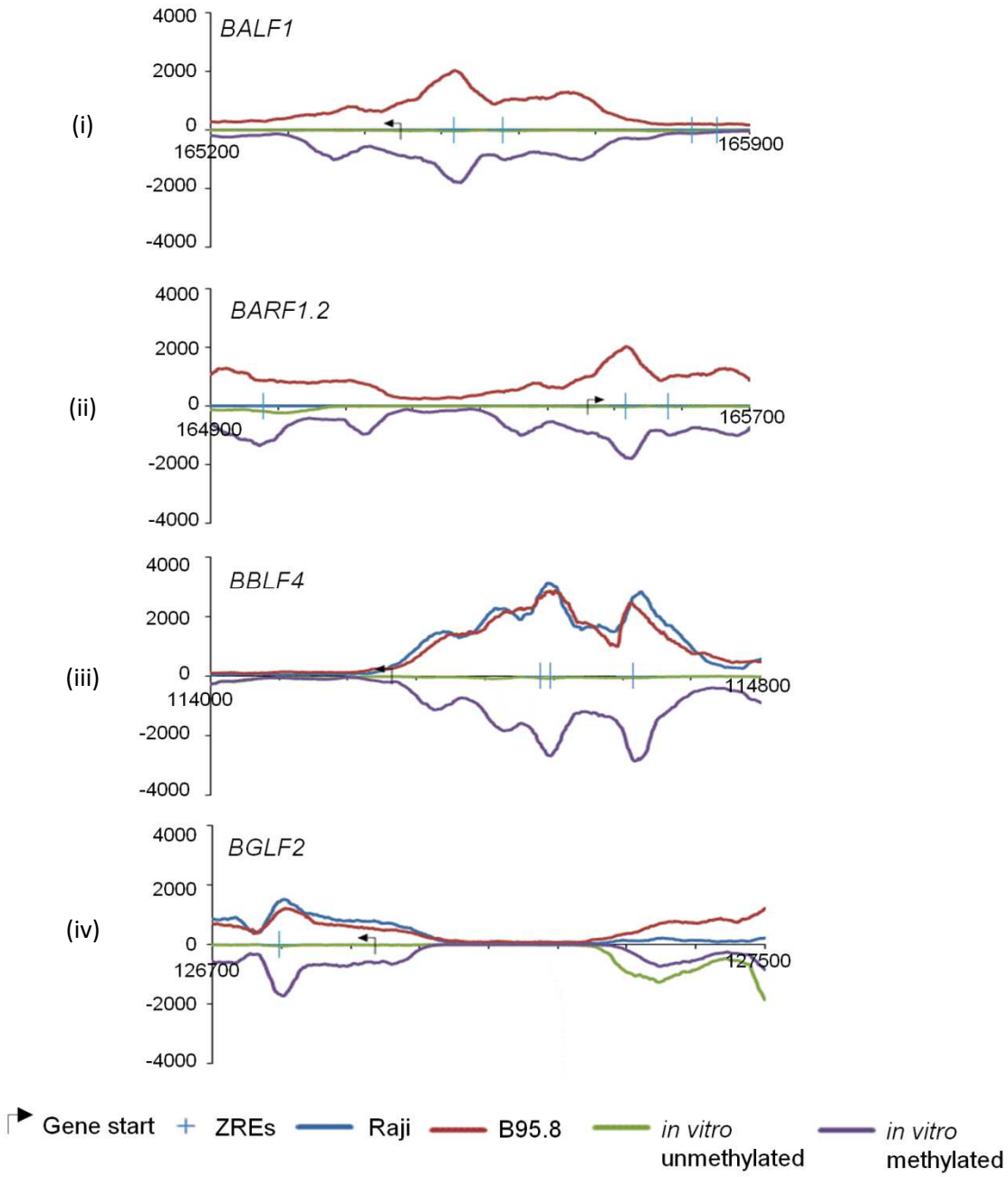


Figure 4.10 Integration of Bergbauer et al ChIP Seq data with ZRE position predictions: significant Zta^{MUT} binding peaks.

The ChIP Seq data and the ZRE position predictions were extracted for each of the defined promoter region and plotted. Blue lines on the x axis indicate ZRE position. The *in vitro* data is plotted as negative for the sole purpose to separate the data, and should be read as equivalent binding as above the x axis. The ZRE prediction matches from left to right are (i) *BALF1*p: RpZRE2, RpZRE2, PFM_{CpG5} M, NapZRE2 (ii) *BARF1.2*p: PFM_{CpG5} M, RpZRE2, RpZRE2 (iii) *BBLF4*p: PFM_{CpG5} A, *BMRF1*(-248), NapZRE2 (iv) *BGLF2*p: RpZRE2. . Approximately 500bp upstream and 200bp downstream is shown.

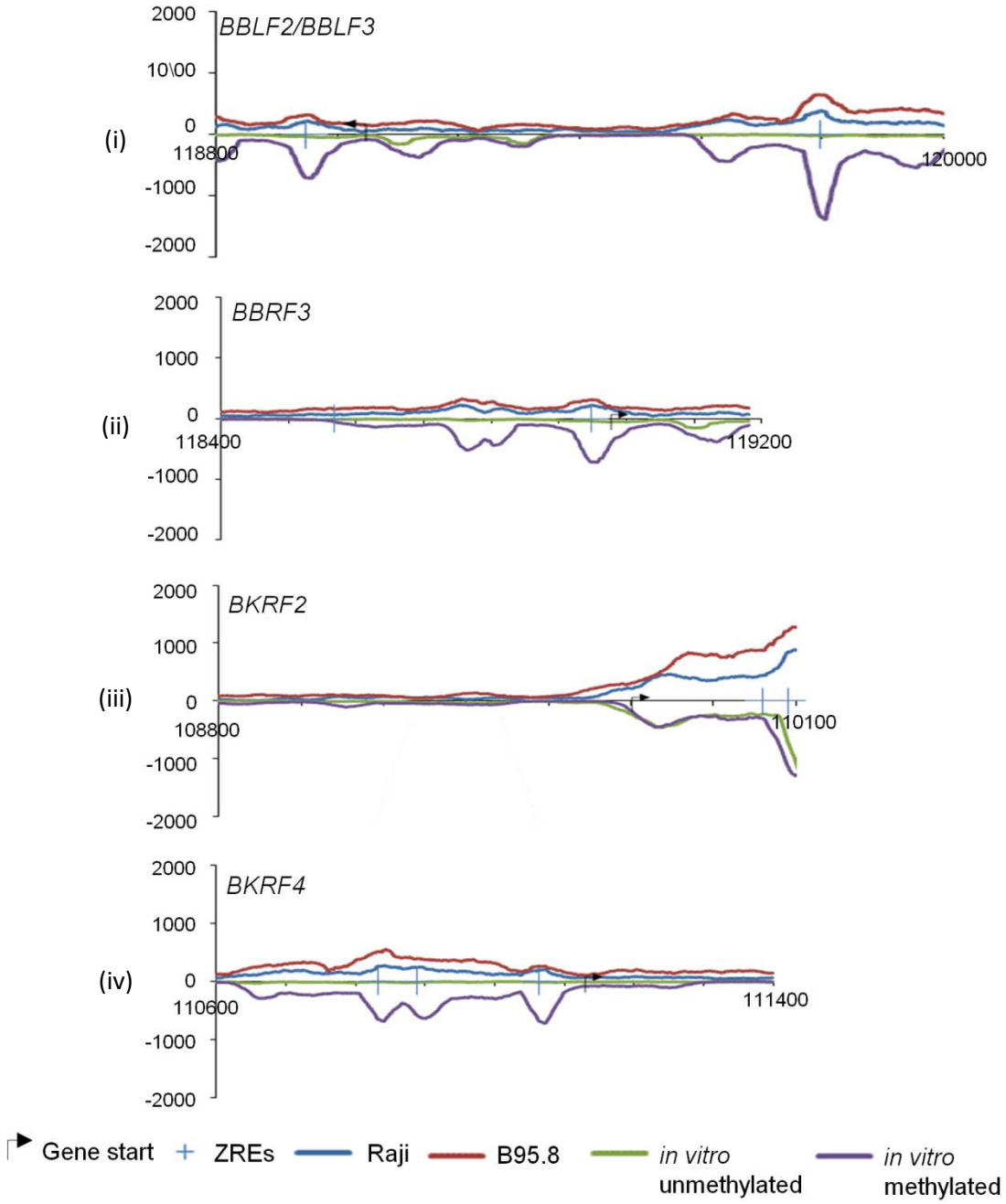


Figure 4.11 Integration of Bergbauer et al ChIP Seq data with ZRE position predictions: specific Zta^{MUT} binding peaks at a lower magnitude.

The ChIP Seq data and the ZRE position predictions were extracted for each defined promoter and plotted. Blue lines on the x axis indicate ZRE position. The *in vitro* data is plotted as negative for the sole purpose to separate the data, and should be read as equivalent binding as above the x axis. The ZRE prediction matches from left to right are (i) *BBLF2/BBLF3*p: PFM_{CpG5} M (ii) *BBRF3*p: PFM_{CpG5} F, PFM_{CpG5} M (iii) *BKRF2*p: PFM_{CpG5} B, PFM_{CpG5} D (iv) *BKRF4*p: PFM_{CpG5} I, BMRF1 (-148), PFM_{CpG5} D. . Approximately 500bp upstream and 200bp downstream is shown.

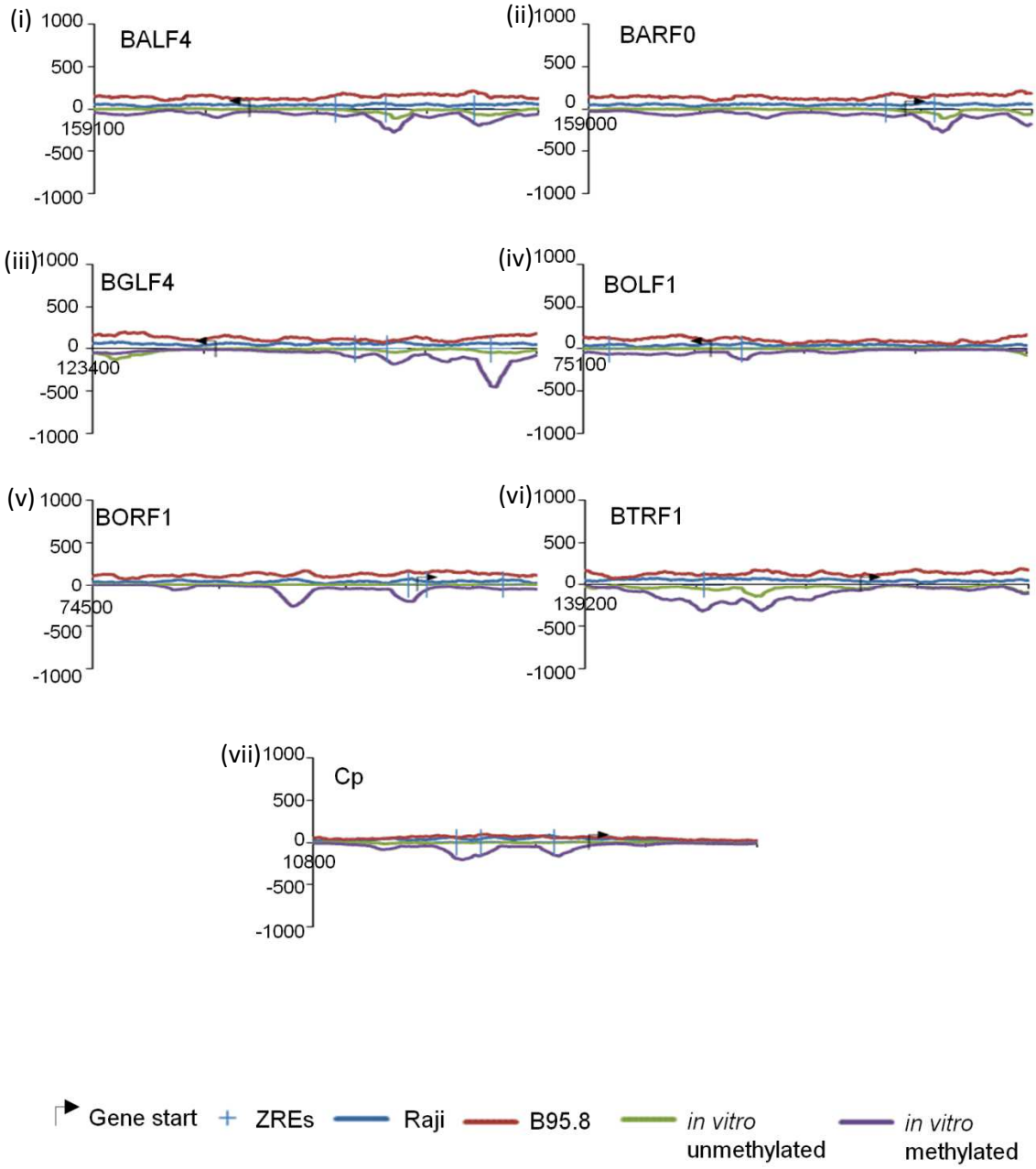


Figure 4.12 Integration of Bergbauer et al ChIP Seq data with ZRE position predictions: in vitro binding only.

The ChIP Seq data and the ZRE position predictions were extracted for each of the defined promoter region and plotted. Blue lines on the x axis indicate ZRE position. The *in vitro* data is plotted as negative for the sole purpose to separate the data, and should be read as equivalent binding as above the x axis. The ZRE prediction matches from left to right are (i) *BALF4*p: PFM_{CpG5} J, PFM_{CpG5} A, PFM_{CpG5} E (ii) *BARF0*p: RpZRE2, PFM_{CpG5} A (iii) *BGLF4*p: PFM_{CpG5} B, BRLF1 (-114) (iv) *BOLF1*p: NapZRE1, PFM_{CpG5} G (v) *BORF1*p: BMRF1(-148), PFM_{CpG5} J, NapZRE1 (vi) *BTRF1*p: PFM_{CpG5} M (vii) Cp: Egr1, PFM_{CpG5} F, PFM_{CpG5} M. . Approximately 500bp upstream and 200bp downstream is shown.

The previous analysis suggests that Zta^{MUT} binds to some of these predicted sites but not to others. To obtain a genome-wide view, and further assess the newly predicted sites, the ChIP Seq data corresponding to the 500 basepairs up- and down-stream of each unique predicted ZRE position was extracted, and averaged before plotting on a graph. The profile for each region was also plotted separately. 0 corresponds to the first nucleotide of the ZRE. RpZRE1, 2 and 3 were first analysed in this manner to assess the confidence with which the data can be analysed in this way (Figure 4.13). RpZRE2 shows significant specific methylation dependent binding of Zta^{MUT} in the *in vitro* data, and this is reflected by the *in vivo* data. The individual plots of each of the 17 occurrences shows that all exhibit a peak of binding in the *in vitro* methylated data, between 487 depth and 3129 depth, and similar peaks can be seen in the *in vivo* data, except for 2 Raji occurrences, where there is no data due to deletions. RpZRE1 shows small peaks at the sites in all cell lines, and as RpZRE1 does not contain a CpG motif this is reflected by the similarity of the *in vitro* unmethylated and methylated binding. The peaks are however very small compared to RpZRE2. When each separate site is plotted individually, six out of the 24 occurrences of the RpZRE1 sequence show peaks of at least 500 depth, whereas the other binding profiles are relatively flat, accounting for the smaller average peaks. In the Raji data there is only one corresponding peak, which matches the occurrence in Rp, and in the B95.8 data there are six peaks, but the peaks are much lower than *in vitro*. RpZRE3 shows no specific binding peaks, and this is confirmed by the separate plots as well as the average. However this is ambiguous, as in some of the separate plots there are shoulders of binding, as seen figure 4.9 for *BRLF1*p, which RpZRE3 appears to contribute to.

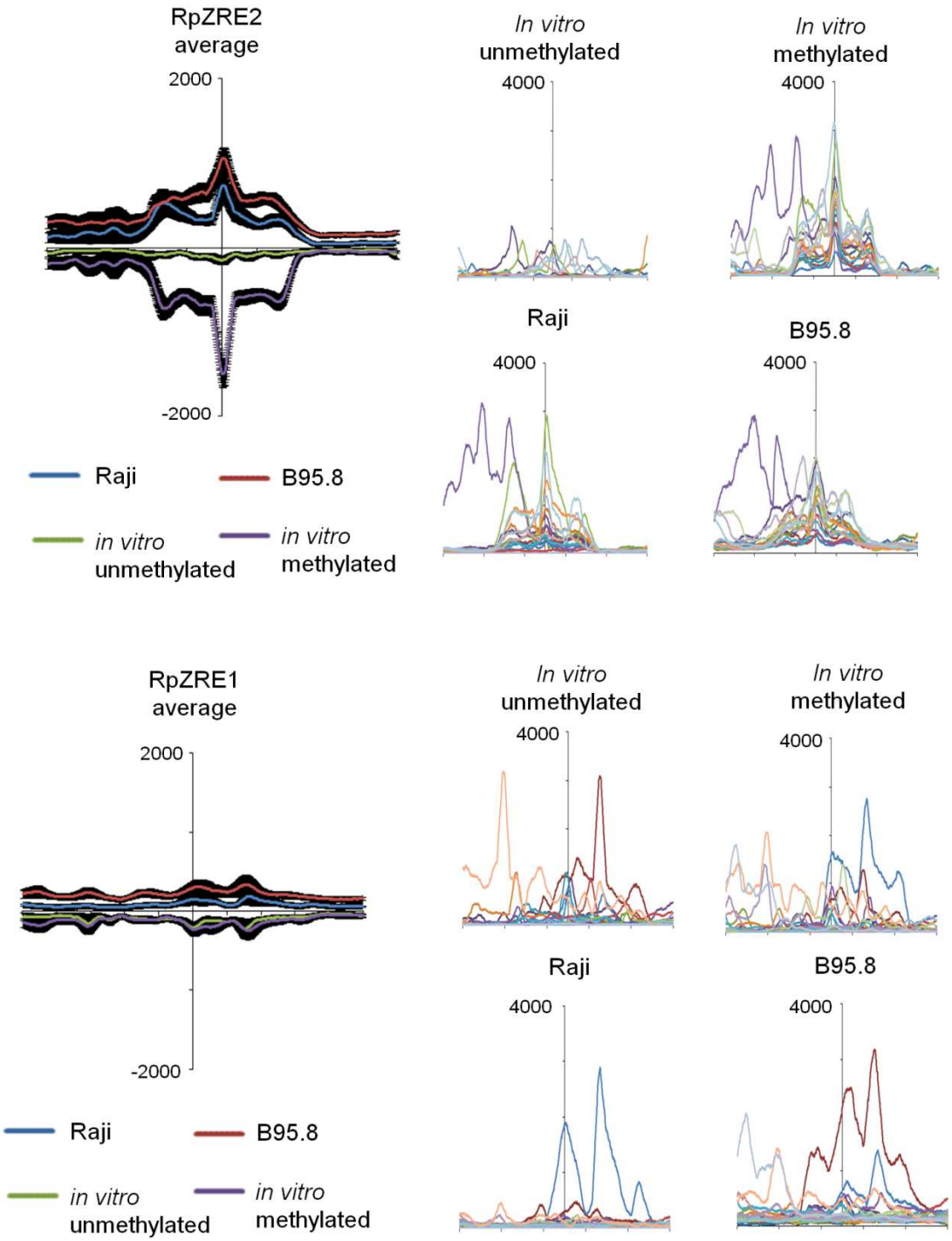


Figure 4.13 Individual site analysis of pattern matches to the RpZREs.

ChIP Seq data surrounding each occurrence of the predicted ZRE match positions of RpZRE2 (n=17), RpZRE1 (n=25) and RpZRE3 (n=6) was extracted and plotted either as an average with standard error as the error bars, or as a separate plot.

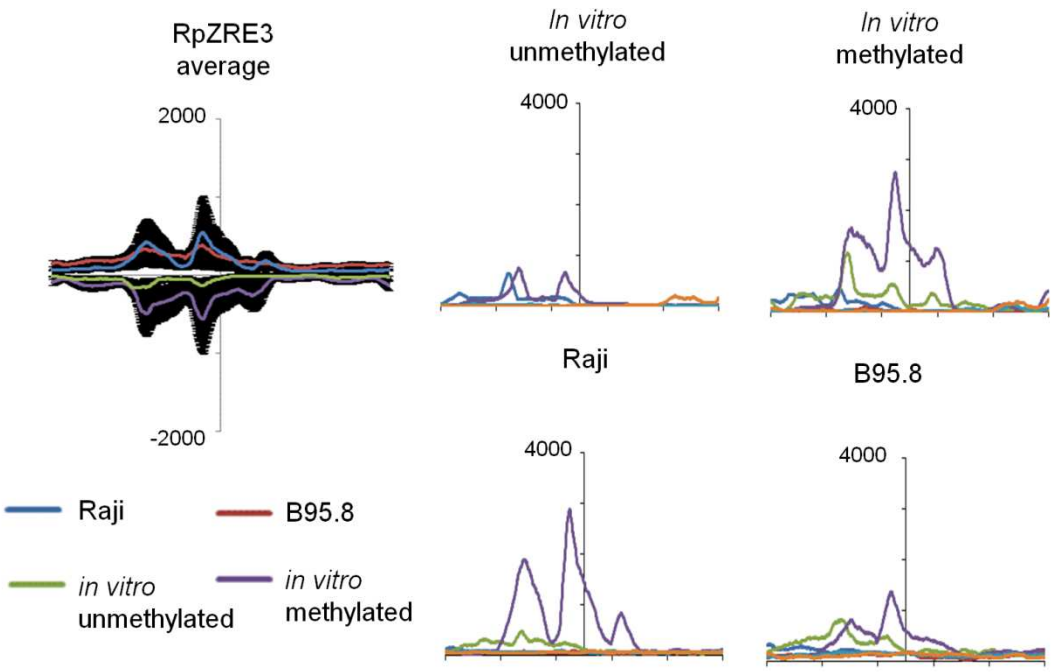


Figure 4.13 Continued.

The two PFM_{CpG5} predictions that did not bind in EMSAs, H and N, were analysed in the same manner (Figure 4.14). This shows that H has no binding peaks in both the average analysis and the 9 separate plots. N does not appear to show any specific peaks on average, however the 16 separate plots are slightly harder to interpret. Whilst no specific peaks are seen, there are 3 broader peaks to which N could be contributing, depending upon the surround sites. Overall however, it appears that the inability of Zta to bind these two sequences in EMSAs is matched by the ChIP Seq data.

The same analysis was performed on all the predicted sites made using PFM_{CpG5}. BMRF1 -248 and PFM_{CpG5} A both show distinct, methylation specific binding peaks, with large error bars (Figure 4.15). The 3 separate plots reveal two significant peaks for BMRF1 -248 and one smaller peak, both *in vitro*

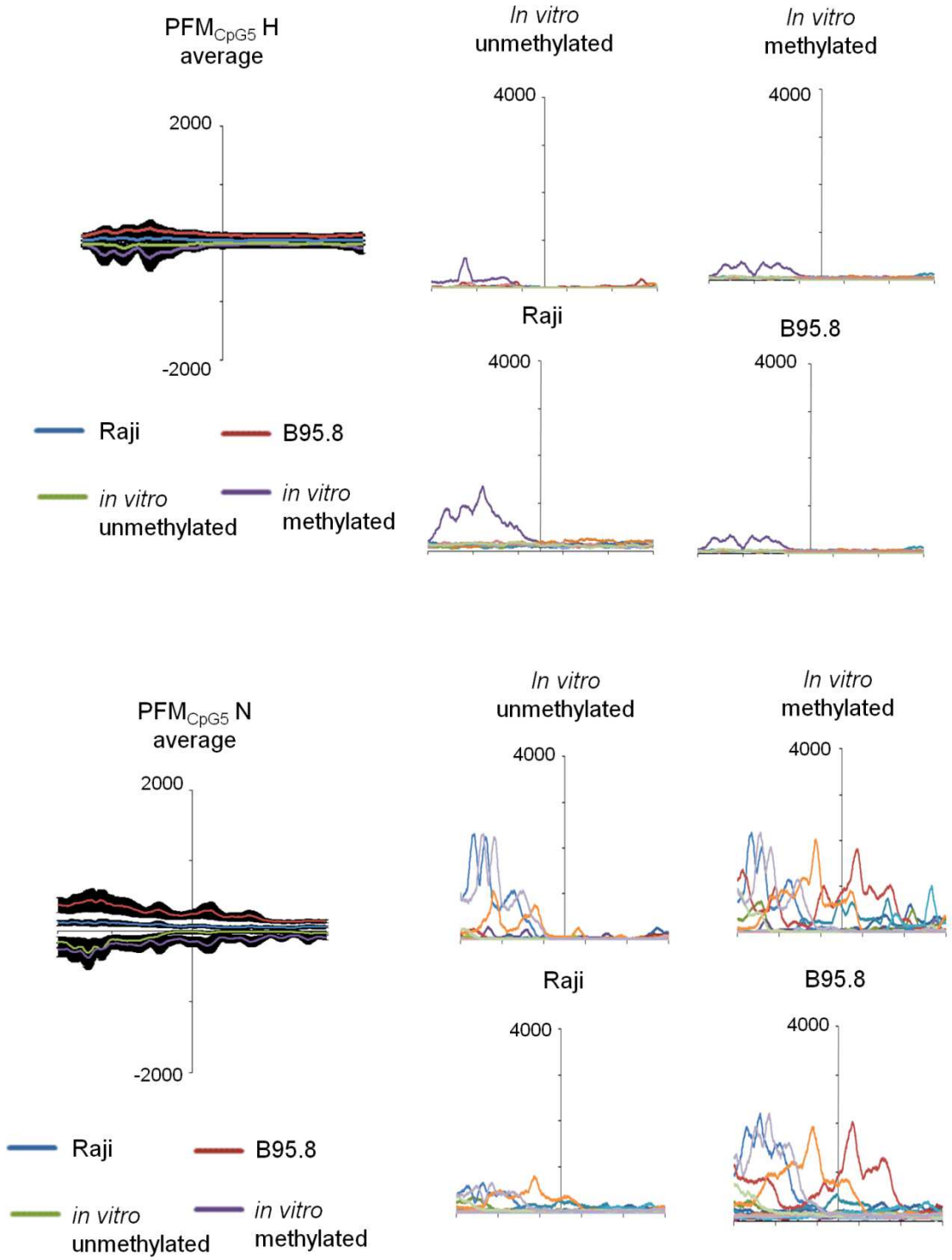


Figure 4.14 Individual site analysis of pattern matches to the PFM$_{CpG5}$ Predictions that did not bind by EMSA

ChIP Seq data surrounding each occurrence of the PFM$_{CpG5}$ H (n=9) and N (n=16) predictions was extracted and plotted either as an average with standard error as the error bars, or as a separate plot.

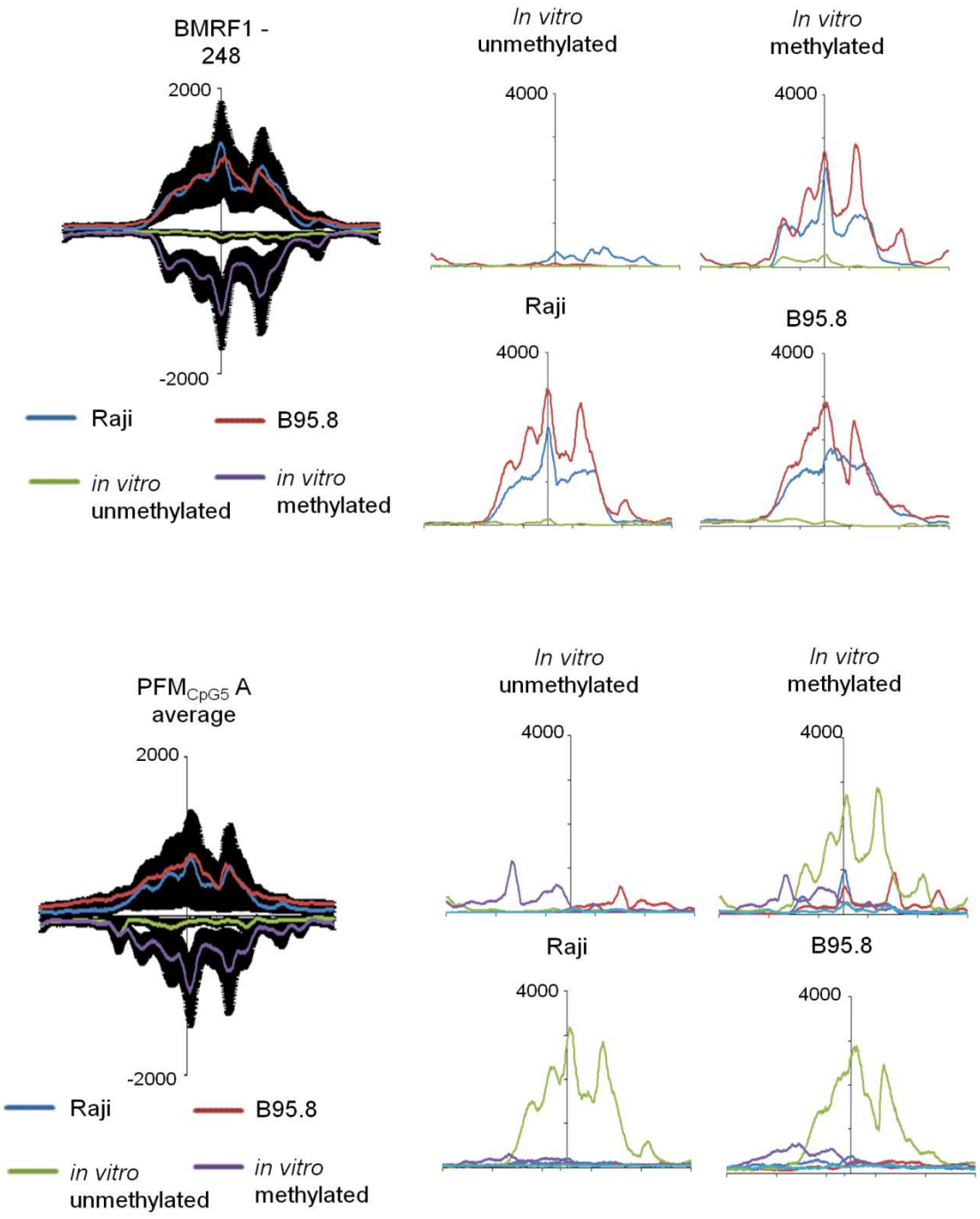


Figure 4.15 Individual site analysis of pattern matches to the PFM_{CpG5} Predictions that appear to bind significantly in the ChIP Seq data.

ChIP Seq data surrounding each occurrence of the BMRF1(-248) (n=3) and PFM_{CpG5} A (n=5) predictions was extracted and plotted either as an average with standard error as the error bars, or as a separate plot.

methylated and *in vivo.* PFM_{CpG5} A has varying peaks *in vitro* methylated, but only one out of five *in vivo*, which corresponds to the promoter region of *BBLF4*, identified previously. PFM_{CpG5} K and D show some binding on average but at a lower level to the previous site predictions (Figure 4.16). All five PFM_{CpG5} D sites show some binding *in vitro,* but only 1 appears to bind *in vivo.* One out of the three PFM_{CpG5} K binds both *in vitro* and *in vivo*. PFM_{CpG5} F and I have very low average peaks, but specific methylation dependent peaks can still be seen in the ten separate plots *in vitro* but only on two occasions for PFM_{CpG5} F and once for PFM_{CpG5} I in the *in vivo* B95.8 data (Figure 4.17). PFM_{CpG5} G, J, E, M, B, BMRF1-148 and RpZRE (-114) show almost no detectable binding on average, however in at least one of the separate plots the site looks to contribute some binding to the peak profile (Figure 4.18). PFM_{CpG5} C and L appear to show no binding peaks in any condition (Figure 4.19).

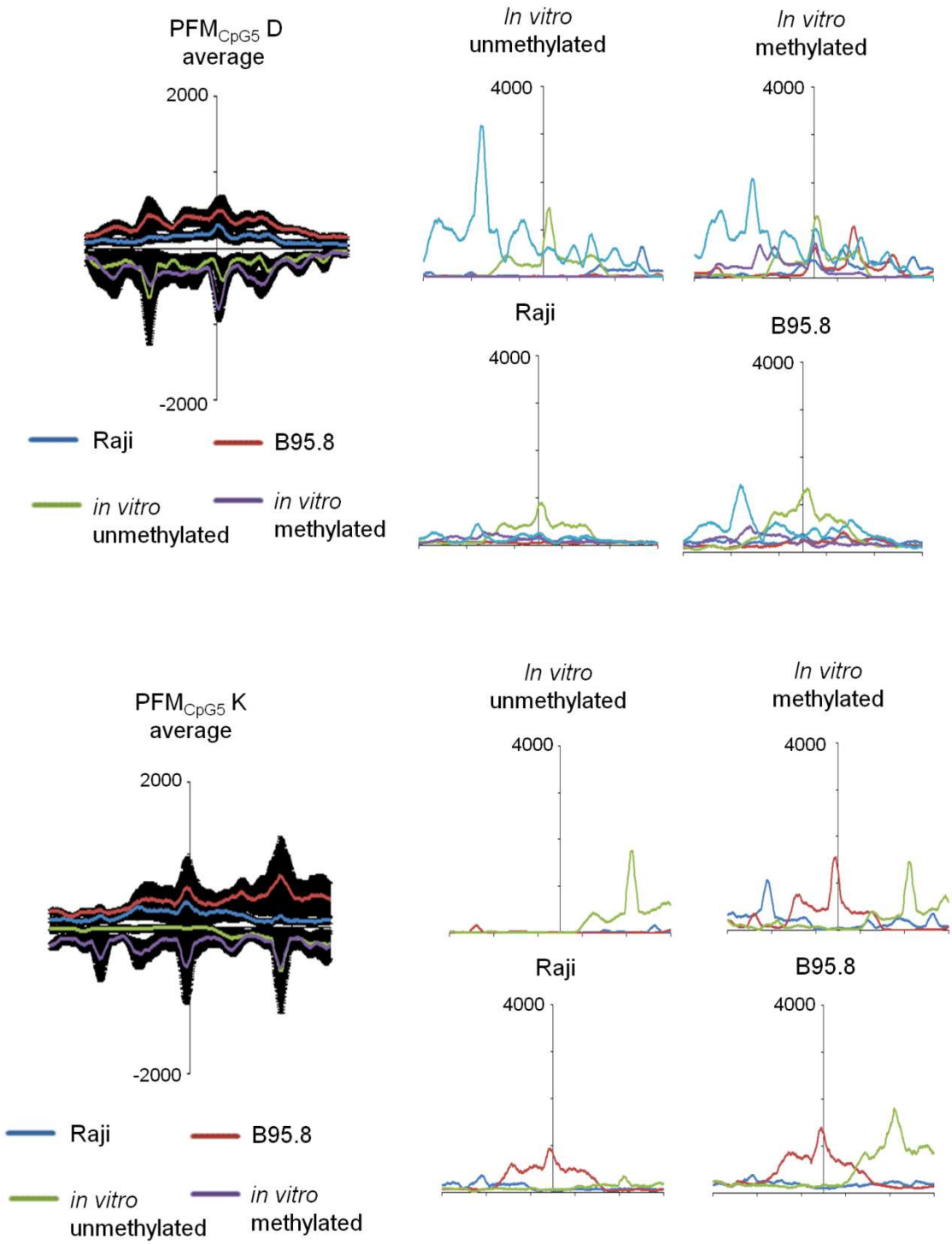


Figure 4.16 Individual site analysis of pattern matches to the PFM$_{CpG5}$ Predictions that appear to show weak binding in the ChIP Seq data.

ChIP Seq data surrounding each occurrence of the PFM$_{CpG5}$ D (n=5) and K (n=3) predictions was extracted and plotted either as an average with standard error as the error bars, or as a separate plot.

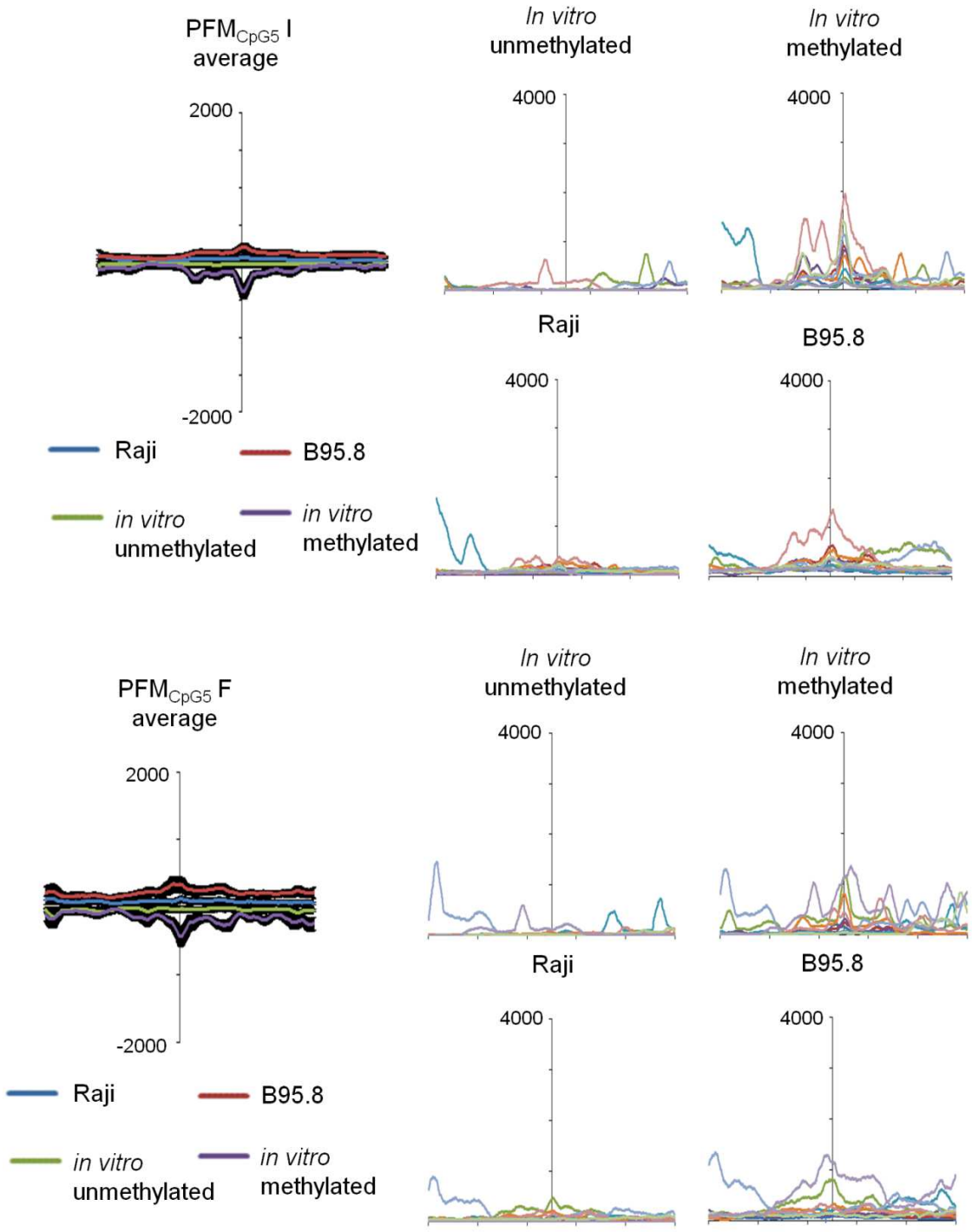


Figure 4.17 Individual site analysis of pattern matches to the PFM$_{CpG5}$ Predictions that appear to show weak binding in the *in vitro* ChIP Seq data.

ChIP Seq data surrounding each occurrence of the PFM$_{CpG5}$ I (n=10) and F(n=10) predictions was extracted and plotted either as an average with standard error as the error bars, or as a separate plot.

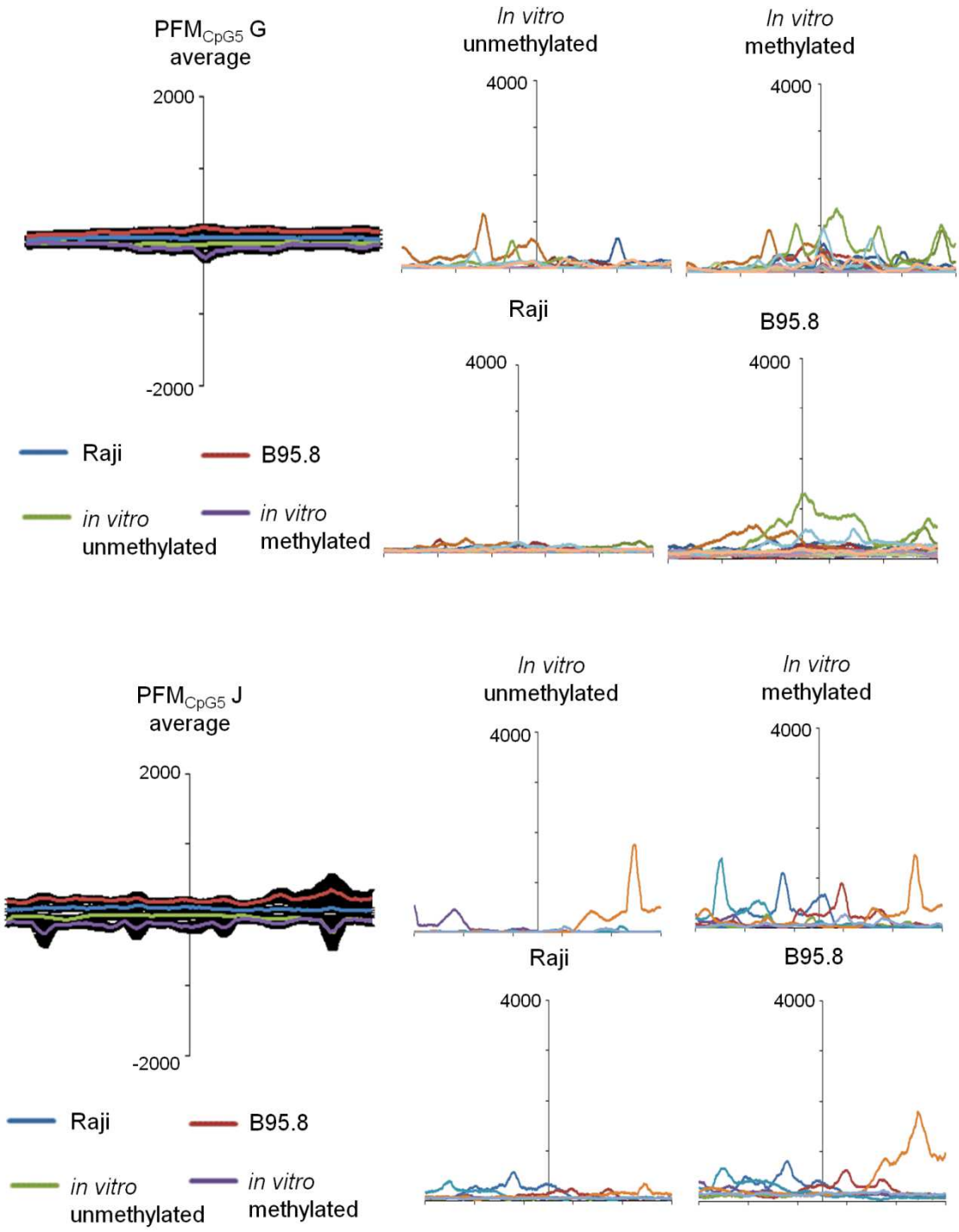


Figure 4.18 Individual site analysis of pattern matches to the PFM$_{CpG5}$ Predictions that appear to bind at least 1 site in the *in vitro* ChIP Seq data.

ChIP Seq data surrounding each occurrence of the PFM$_{CpG5}$ G (n=24), J (n=7), E (n=8), M (n=31), B, BMRF1-148 (n=17) and RpZRE (-114) (n=5) predictions was extracted and plotted either as an average with standard error as the error bars, or as a separate plot.

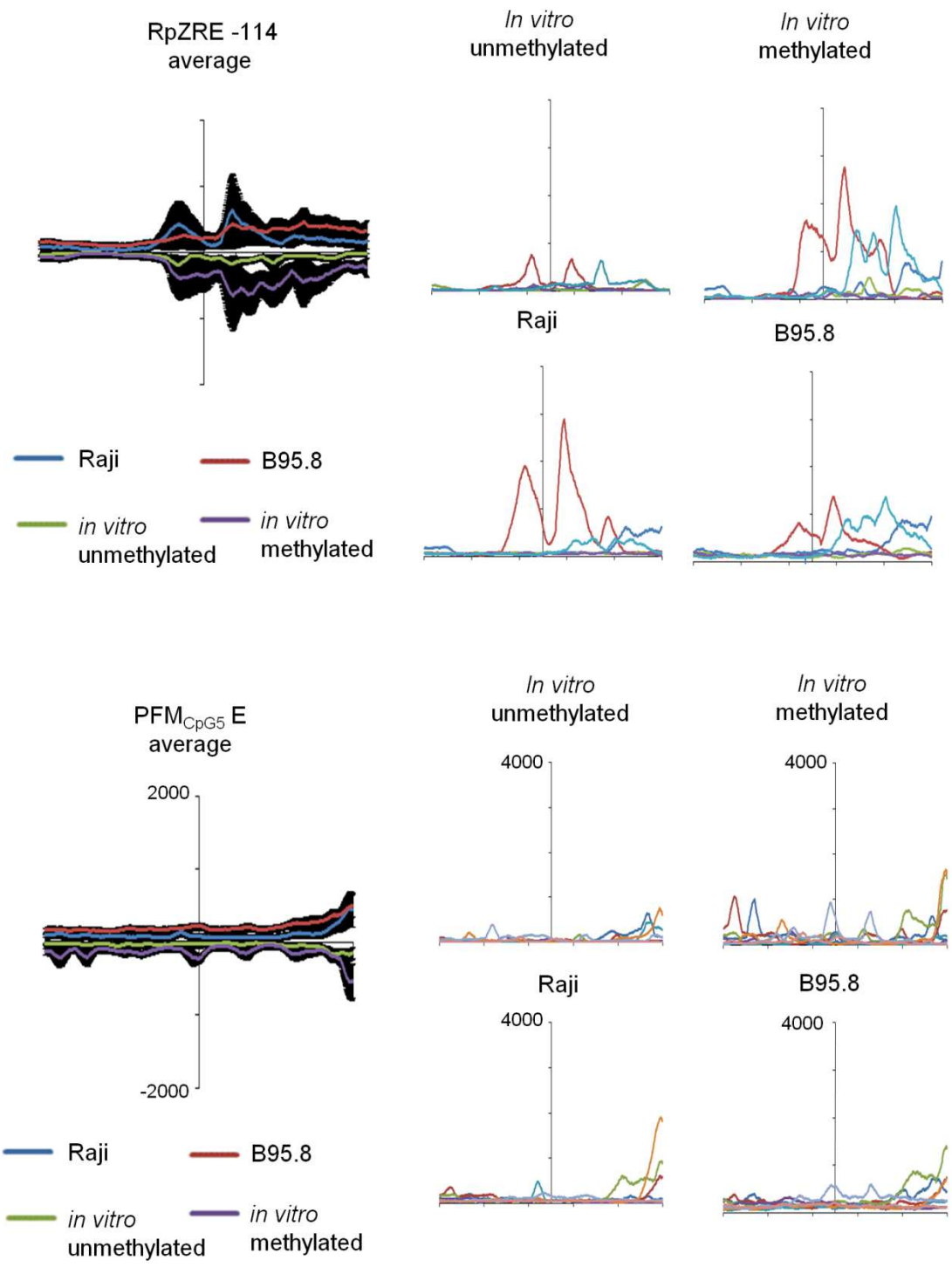


Figure 4.18 Continued

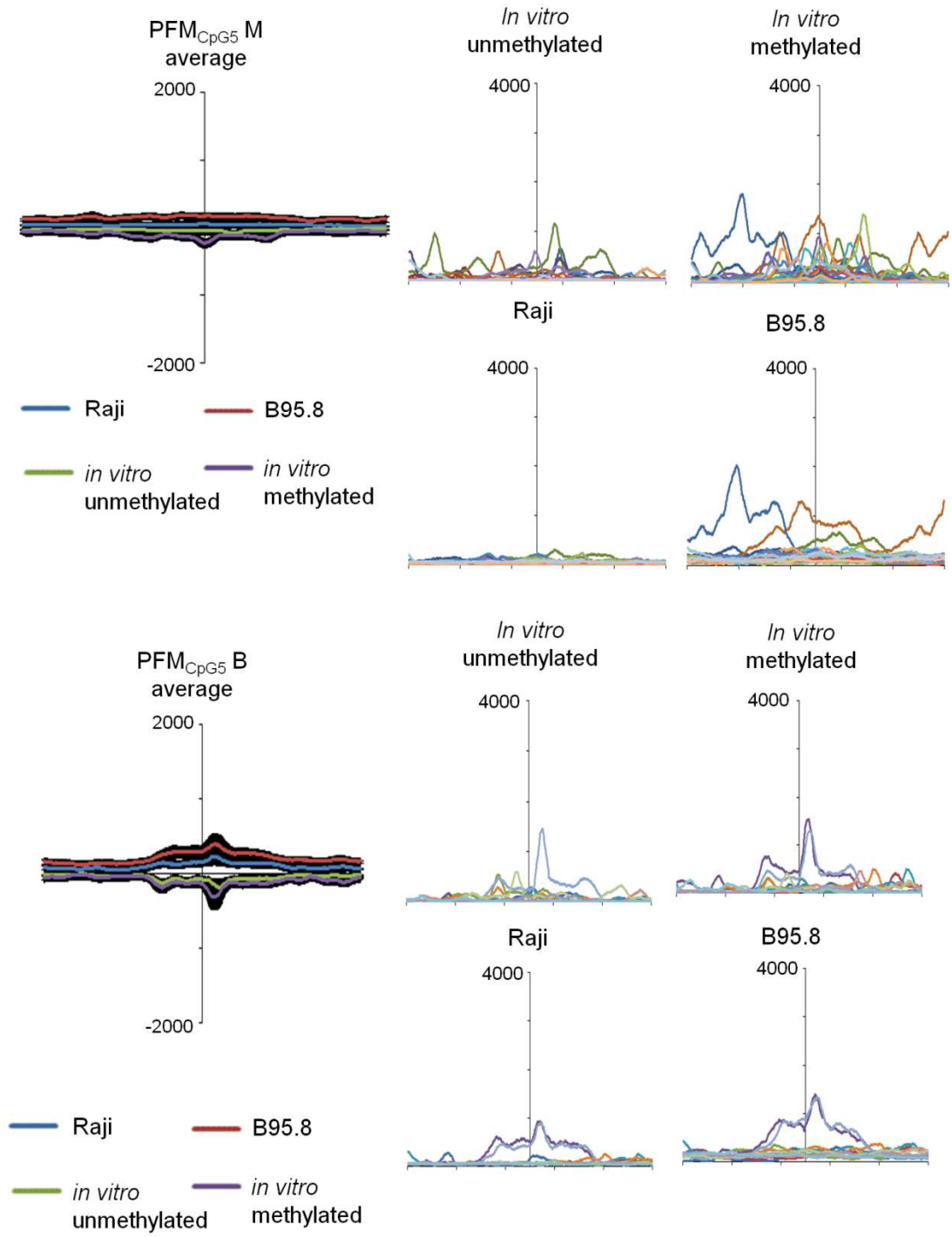


Figure 4.18 Continued

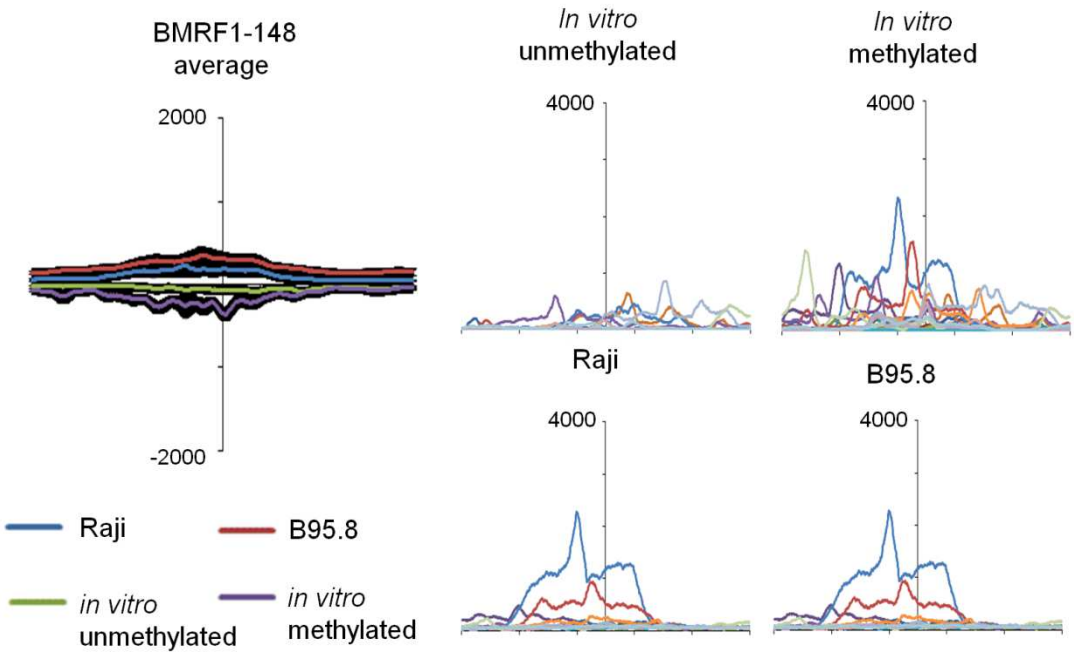


Figure 4.18 Continued

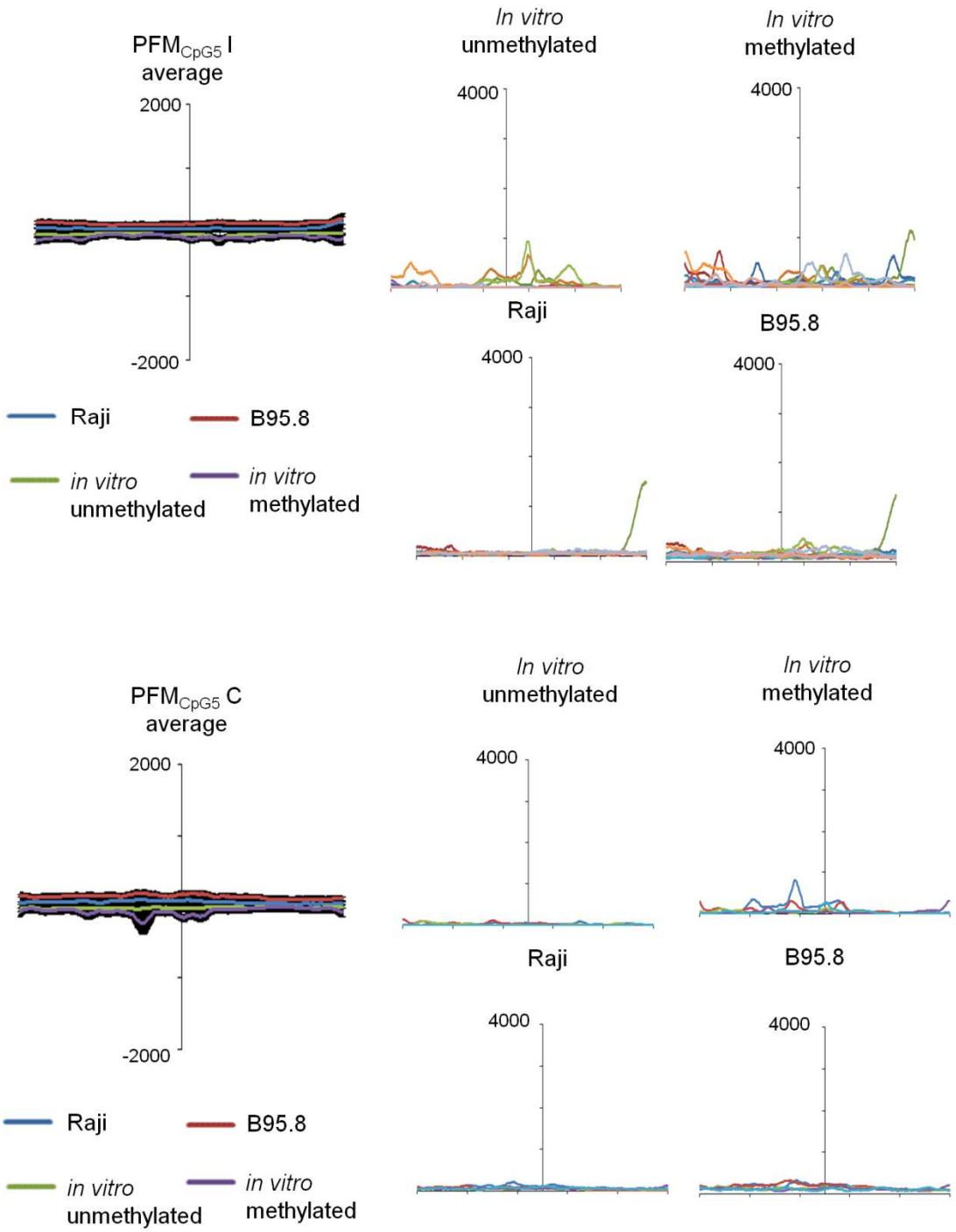


Figure 4.19 Individual site analysis of pattern matches to the PFM$_{CpG5}$ Predictions that do not appear to bind in the ChIP Seq data.

ChIP Seq data surrounding each occurrence of the PFM$_{CpG5}$ L (n=14) and C(n=5) predictions was extracted and plotted either as an average with standard error as the error bars, or as a separate plot.

4.2.7 Genomic position prediction of ZRE core binding sequences in the Human genome

Global predictions for the 3.2Gb Human Genome (NC GRCh37 extracted from Ensembl 57 (Flicek et al., 2009)) was undertaken by bioinformatician D. Thomas (BSMS), using an exact pattern match with the 32 validated ZRE core sequences. Of the 22,224 regulatory regions searched, 7,538 (33.9%) genes contained at least one CpG ZRE, whereas 3,390 (15.3%) genes contained CpG ZREs but have no methylation independent ZREs in their regulatory regions (Figure 4.20A). Methylation-dependent ZRE core sequences are therefore fairly common in human promoters.

To explore whether Zta may overturn epigenetic silencing of the host genes that contain methylation-dependent ZRE core sequences in their regulatory regions, we first questioned whether these regulatory regions are methylated in B-cells. EBV infects and establishes latency in B-cells and epithelial cells. Data-sets containing methylome information from circulating B cells (CD19+) (Rauch et al., 2009), germinal centre (GC) B cells and EBV-infected lymphoblastoid cell lines (LCL) derived from GC B cells (Leonard et al., 2011), were interrogated to identify those CpG ZRE containing promoters that are methylated in a representative B-cell, as was done in the previous chapter. This identified about 7% of human genes that are both methylated and contain a methylation-dependent ZRE in their regulatory regions (Figure 4.20B). The differences in methylation status between the methylomes in this gene list are shown in Figure 4.20 C and D.

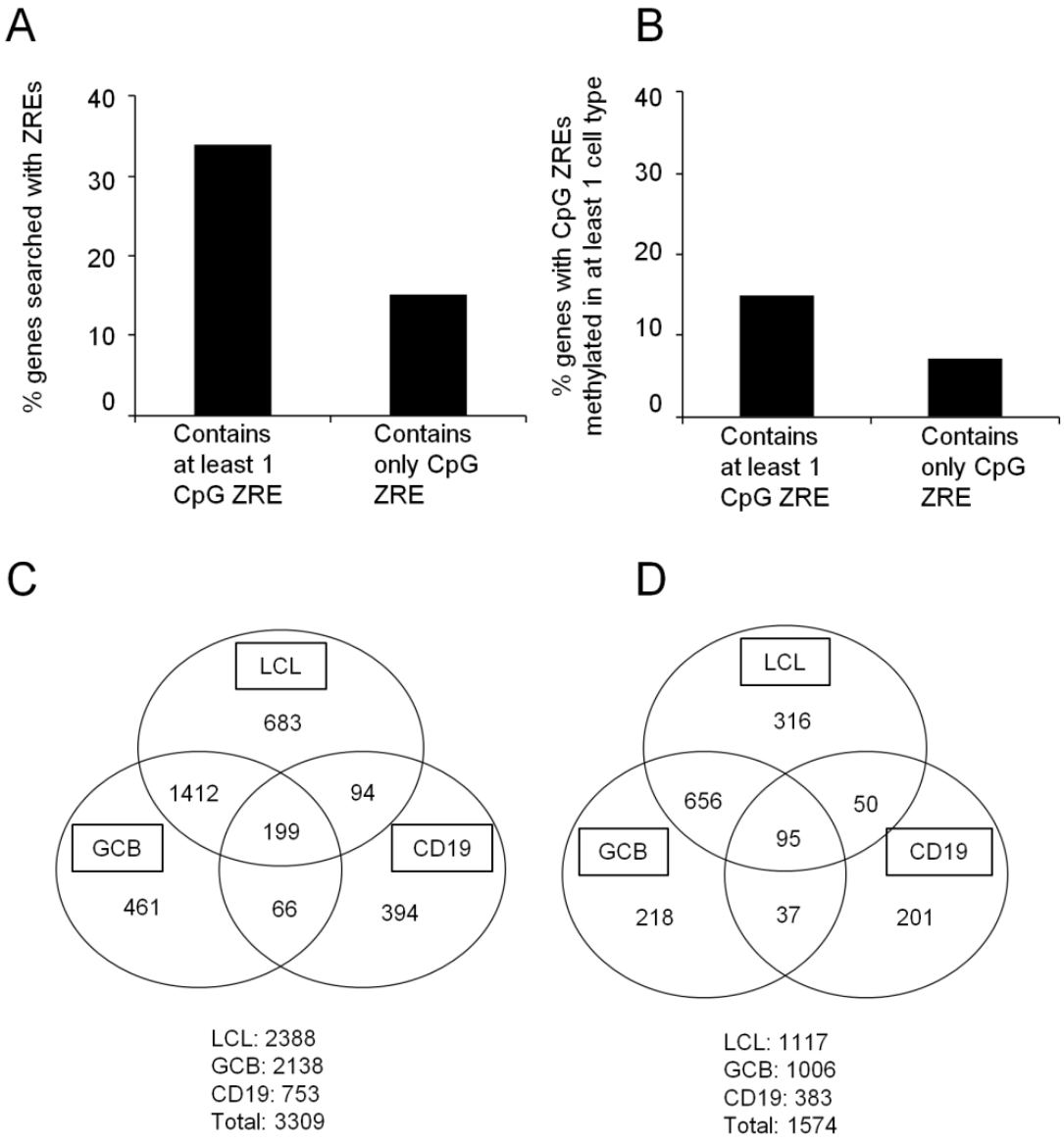


Figure 4.20 Interrogating methylome data to identify methylated promoter regions with CpG-containing ZREs

A A graph to show percentage of total gene promoters searched with (i) at least one CpG-containing ZRE (ii) only CpG-containing ZREs. B A graph to show percentage of total gene promoters searched and found to be methylated in at least one cell type with (i) at least one CpG-containing ZRE (ii) only CpG containing ZREs. C A Venn diagram to represent the methylome data compared to the human gene lists of those promoters with at least one CpG-containing ZRE.D A Venn diagram to represent the methylome data compared to the human gene lists of those promoters with only CpG-containing ZREs.

Genes that are regulated when Zta is expressed during EBV lytic cycle from two published transcriptome studies (Broderick et al., 2009b, Yuan et al., 2006) were compared with the set of methylated promoters that contain methylation dependent ZREs. This revealed 30 genes that contain the features required to be regulated by Zta through a methylated ZRE in their promoter regions (Table 4.5). The up- or down-regulation of the each of the genes was assessed from either the classification in the paper, if one was made, or the immediate behaviour following lytic induction (between 0 and 12 hours) by interrogating the heat map. 12 genes appear to be down-regulated, and the remaining 18 up-regulated.

These 30 genes were analysed using the online bioinformatic program DAVID (Huang da et al., 2009b, Huang da et al., 2009a). This clusters similar terms together to create clusters of similar function genes. In this gene set the enrichment scores were fairly low, suggesting many different functions are regulated.

Microarray	Ensembl ID	Gene name	Class I	Class II	Class III	-/+
Yuan et al, 2006	ENSG00000167393	PR48	0	0	7	-
	ENSG00000084652	DKFZp451J0118	0	0	1	-
	ENSG00000106236	NPTX2	0	0	1	-
	ENSG00000170889	RPS9	3	0	3	-
	ENSG00000149418	ST14	3	0	1	-
	ENSG00000110324	IL10RA	1	0	2	-
	ENSG00000130193	LOC51337	1	0	2	-
	ENSG00000143603	KCNN3	1	1	1	-
	ENSG00000204237	LOC339229	1	0	2	-
	ENSG00000143013	LMO4	1	0	1	-
	ENSG00000148154	UGCG	0	0	3	+
	ENSG00000138166	DUSP5	0	0	1	+
	ENSG00000161011	SQSTM1	0	0	1	+
	ENSG00000167460	TPM4	0	0	1	+
	ENSG00000170876	MGC3222	0	0	1	+
	ENSG00000173334	C8FW	3	2	2	+
	ENSG00000120910	PPP3CC	1	0	3	+
	ENSG00000007168	PAFAH1B1	2	0	1	+
	ENSG00000115738	ID2	1	0	2	+
	ENSG00000021355	SERPINB1	1	0	1	+
	ENSG00000139718	KIAA1076	0	1	1	+
	ENSG00000140650	PMM2	1	0	1	+
	ENSG00000153113	CAST	1	0	1	+
	ENSG00000167720	SRR	1	0	1	+
	ENSG00000170542	SERPINB9	1	0	1	+
	ENSG00000186480	INSIG1	1	0	1	+
Broderick et al, 2009	ENSG00000163110	PDLIM5	2	0	1	-
	ENSG00000125354	SEPT6	0	0	2	-
	ENSG00000160957	RECQL4	0	0	4	+
	ENSG00000104856	RELB	0	0	2	+

Table 4.5 Human gene promoters with at least 1 CpG-containing ZRE, and associated gene expression shown by microarray to be affected by lytic induction in Akata cells.

The Ensembl code and associated gene name are shown, and the microarray source of data is indicated by the publication. The sites are broken down into the appropriate classes, and the affect of lytic cycle upon expression is signified by either “-” corresponding to downregulation, or “+” upregulation

4.3. Discussion

Following several iterations of a predictive and evaluative approach, we identified a set of 32 distinct sequence variants in the core 7 nucleotide sequence to which Zta can bind. This includes 20 sites containing a CpG motif, the majority of which are only recognized by Zta when they are methylated.

The consensus binding site identified for non-CpG ZREs is similar to the binding sites originally described for Zta. In contrast, the consensus binding site for CpG containing ZREs is remarkably different (Figure 4.21). This sequence is dominated by an almost invariant G 5' to the absolute prerequisite for me-CpG at positions 1' and 2' in the right-half of the core sequence. Beyond that, there is some resemblance to the non-CpG consensus only at the extreme 5' and 3' ends.

The identification of 58 EBV genes that contain methylation dependent CpG ZREs in their proximal regulatory regions, combined with the knowledge that the EBV genome is heavily methylated during latency (Fernandez et al., 2009, Kalla et al., 2010), suggests that Zta plays an important role in overturning epigenetic silencing of over half of the EBV genes during lytic replication. Zta^{MUT}-bound sequences resulted in a similar consensus site with an almost invariant GCG (Bergbauer et al, 2011) to that shown in Figure 4.21. It has previously shown that wild-type Zta interacts with both Rp (Karlsson et al., 2008b) and the *BRRF1* promoter (Dickerson et al., 2009) *in vivo.* In contrast, Zta^{MUT} was unable to interact with the full set of ZREs in Rp, nor did it recognize the *BRRF1* promoter (Bergbauer et al., 2010). These conflicting data can be rationalised if it is considered that the regions of Zta excluded from the mutant form influence

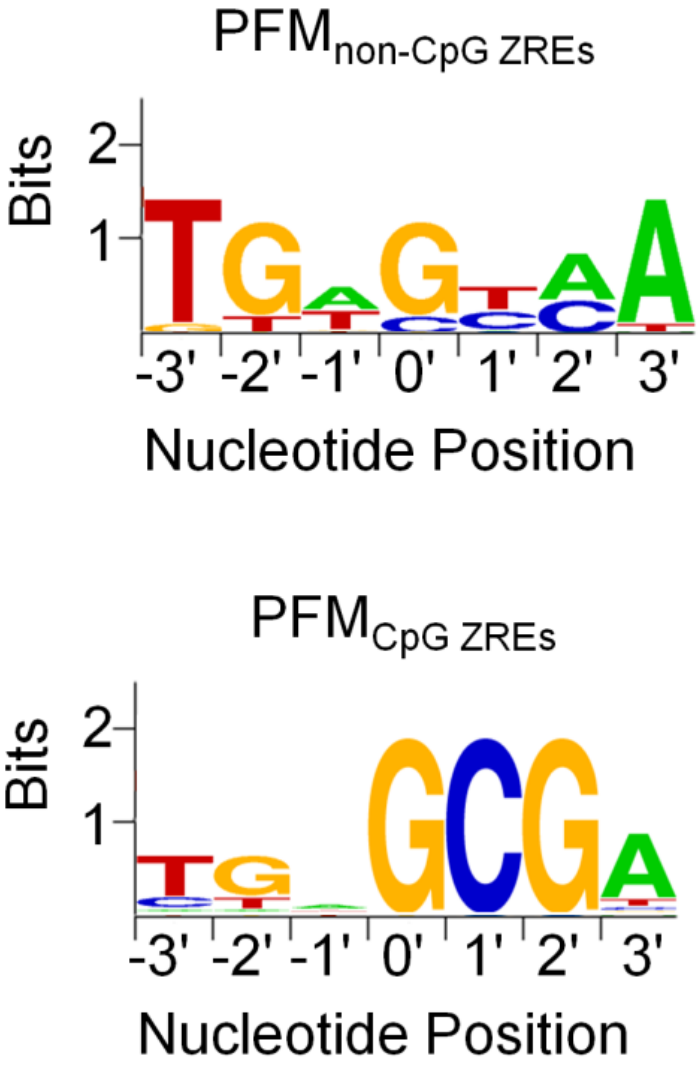


Figure 4.21 Sequence logo representations of methylation dependent and independent ZREs.

The ZRE_{32} list was split into those containing CpG motifs and those without. Position Frequency Matrices (PFMs) were created and Sequence Logos were produced to represent the consensus sequence of each.

the stability of DNA binding *in vivo*. This therefore casts some doubt over the completeness of this ChIP Seq study, and repetition of this with full length Zta would be beneficial.

The 22 EBV genes that contain methylation dependent but no methylation independent ZREs are particularly interesting and informative. Several of these genes are required for EBV replication and include components of the helicase/primase complex (*BBLF4, BBLF2/BBLF3*), the viral protein kinase (*BGLF4)*, and glycoproteins gL (*BKRF2*) and gB (*BALF4*). All of these regulatory regions are heavily methylated during viral latency yet unmethylated following replication and immediately after infection (Fernandez et al., 2009, Kalla et al., 2010). The promoter for *BBLF4* has been validated as being absolutely dependent on methylation for Zta activation, and the ZREs in BBLF2/3 were shown to be dependent on methylation for binding in EMSAs (Bergbauer et al., 2010). The binding of Zta to some of the novel ZREs was confirmed in this chapter in replication stalled IgG treated Akata cells by ChIP, including those found in viral protein kinase BGLF4 and 2 other regions, BKRF4 and BTRF1, tegument proteins.

The discovery that around 20% of EBV regulatory regions contain CpG ZREs but have no methylation independent ZREs in their regulatory regions strongly supports the hypothesis that the unmethylated status of the EBV genome guards against the expression of the full range of lytic genes and therefore lytic replication during the establishment of latency. However, a further five gene promoter regions were shown in the same paper to require methylation for expression (*BRLF1, BALF2, BMRF1, BBLF1* and *BALF5*); these regions were

overlooked in this analysis due to the fact that they contain methylation independent Class I ZREs in their promoter as well as the CpG-containing ZREs. This suggests that even in the presence of Class I sites, a methylation dependent mode of activation could still be required, increasing the number of genes potentially controlled in this manner to 58.

It is interesting to compare the results of our ChIP with that of the Zta^{mut} ChIP. Demonstration of full length Zta binding to the promoter region of *BKRF4* was also shown in the ChIP seq data, to a lesser extent. However, *BGLF4* and *BTRF1* promoter binding was not reflected in an equivalent manner, with the ChIP Seq only displaying some binding *in vitro.* This could be due to the differences in the protein (full length versus truncated version) or cell background (Akata versus Raji/B95.8).

Each unique ZRE sequence was analysed by averaging the ChIP Seq values for the positions where each occurrence was found. This allowed the behaviour of the site with Zta^{MUT} both *in vitro* and *in vivo* to be assessed. First the well-described ZREs within Rp were analysed, to assess the suitability of this technique for evaluating the novel ZRE sequences. The sequence logo resulting from the ChIP Seq (Bergbauer et al, 2010) very closely matched RpZRE2, and by averaging the 17 occurrences of this ZRE, there is a clear binding peak in both *in vivo* samples, and in the methylated *in vitro* experiment, compared to almost non-existent binding in the unmethylated *in vitro* condition. However, the same could not be shown to a comparable extent for RpZRE1 and RpZRE3. There are several possibilities to explain the differences. The ChIP Seq used Zta^{MUT} lacking the transactivation domain, suggesting that this

region could influence sequence-specific DNA binding. It would be interesting to investigate the ability of the Zta^{MUT} to bind to the predicted ZREs in EMSAs. Equally, comparison with ChIP Seq data performed using wild type Zta would reveal the full binding possibilities within the EBV genome.

The continuous nature of the ChIP Seq binding data can make it difficult to specifically interpret the binding ability of each novel PFM_{CpG5} predicted site. The proximity of other ZREs also makes analysis difficult. However, the predictions can be placed upon a scale of increasing binding ability, starting with those that do no appear to bind at all (PFM_{CpG5} H and N), those which can only be seen to bind in EMSAs (PFM_{CpG5} C and L), or not able to bind in all contexts (PFM_{CpG5} B, E, G, J, M, BMRF1(-148) and RpZRE(-114)), those which bind to varying extents, perhaps performing better *in vitro* than *in vivo* (PFM_{CpG5} F and I, PFM_{CpG5} K and D) and those which show fairly well defined binding profiles (BMRF1 (-248) and PFM_{CpG5} A). It is interesting to note that the main CpG containing sequences picked out by the analysis in the Bergbauer et al paper were matches to RpZRE2 and BMRF1(-248); this data is therefore likely to confirm these as competent ZREs.

The *in vitro* binding gives a good indication of the potential of Zta to bind to a specific sequence, which is more similar to the effects seen in EMSAs. However, in cell lines where normal chromatin formation will occur, the actual binding positions of Zta may be obscured by these structures, or even other binding proteins, including methyl-binding proteins. Perhaps in the latter case, the relative binding affinities of the proteins will affect overall expression. Another factor which has not been taken into account is whether the

occurrences of each novel site fall within promoter regions; the data was not filtered in this way previous to analysis, and therefore does not take this into account. A hypothesis would be that Zta is more likely to bind in promoter regions, therefore the averaged binding data of specifically those found there may form more specific averaged peaks, when analysed without sites that occur outside the defined promoter regions. However, due to the compact nature of the EBV genome, it is difficult to assess whether many of these positional predictions do occur outside a potential promoter region.

Analysis of the host genome exposed a set of genes with regulatory regions resembling those of EBV genes. They contain methylation dependent ZREs, are methylated in cells targeted by EBV and are regulated following expression of Zta. These promoters therefore share the characteristics that would allow epigenetic silencing to be overturned during EBV replication through the interaction of Zta with methylated ZREs. Zta has been shown to both upregulate (*Egr1, Il-8, Il-13*) and downregulate (CIITA) human gene expression, and these opposite effects can occur through interaction with the same core ZRE; the ZRE found in the promoter of *CIITA*, causing down-regulation, is an exact match to that found in the promoter IL13, where interaction of Zta with this site has been shown to cause up-regulation. The potential for gene expression control is divergent, and not reliant on a specific ZRE sequence. How can one protein interacting with the same DNA sequence have two such differing roles? Perhaps the inhibitive effect of Zta on this promoter is due to overlapping of the ZRE with another response element, usually required for a different protein to interact with and activate gene expression. Another possible mechanism would involve cis-acting modules of other transcription control proteins in the same

promoter; for activation, one set of proteins is required and for inhibition, a separate set is involved. The overall effect of Zta binding to the promoter would be a context and environmental dependent outcome.

Zta^{MUT}, used by Bergbauer et al in the ChIP Seq data interrogated in this chapter, lacks the transactivation domain. The behaviour of this mutant in some specific regions of the EBV genome is different to previously published data, and ChIP PCR performed in this chapter. This discrepancy suggests that this region could influence sequence-specific DNA binding, perhaps due to any effect these domains may have on overall structure or dimerisation ability, or the interaction of Zta with other proteins. It would be interesting to investigate the ability of the Zta^{MUT} to bind to the predicted ZREs in EMSAs. Equally, comparison with ChIP Seq data performed using wild type Zta would reveal the full binding possibilities within the EBV genome. The transactivation domain is known to interact with other proteins, and this may lead to co-operative binding to some DNA sequences. Therefore without this region in Zta^{MUT} may not bind to all the sites shown in a wild type Zta ChIP. The EMSAs were performed with *in vitro* translated Zta, and without all the cellular proteins that would be present in a nuclear extract, or in a ChIP experiment. However, some other proteins are present in the translation systems, which sometimes show as non-specific bands on an EMSA gel. These could contribute to complex formation.

In a ChIP experiment, the binding of Zta to specific DNA sequences is affected by the surrounding chromatin structure; if the sequence is hidden within a condensed section of chromatin, Zta will be unable to interact with that sequence. It would not be surprising if Zta was prevented from binding in some

contexts for this reason, and this would explain some of the graphs with large error bars showing the averaged binding to specific sequences.

Chromatin structure will also affect the sonication step; silenced chromatin does not shear to the same extent as euchromatin, leading to an overrepresentation of long fragments in heterochromatic regions. These long fragments would not pass the size-selection of the protocol, and also have less fragment ends within the region, reducing the number of reads (Teytelman et al., 2009). All these reasons would lead to underrepresentation of the silenced regions (Liu et al., 2010).

Regions with a higher GC content have been shown to have a high correlation to an increased number of reads (Dohm et al., 2008), which again will lead to overrepresentation of specific regions. This could be due to the higher temperature required to melt GC DNA, therefore AT rich DNA may be denatured and unusable at a lower temperature prior to PCR, allowing GC rich DNA to perpetuate due to their increased stability (Dohm et al., 2008, Margulies et al., 2001). However, GC rich regions that possess strong secondary structure with a higher melting point can be difficult to amplify in standard PCR reactions (Hube et al., 2005), such as that which would be used as to amplify the DNA before sequencing. This would lead to an underrepresentation of GC rich regions, making the overall effect of GC content on reliability of ChIP Seq data difficult to predict. Sequencing biases would be controlled for by using an appropriate input sequencing run concurrently.

Chapter 5: Host methylation dependent DNA binding factor

5.1. Introduction

The occurrence of methylation dependent ZREs in the human genome does not necessarily imply they are there for the benefit of EBV infection; rather it suggests that Zta shares a common mechanism with a host DNA binding protein. This led to the question of whether a host factor could bind to the RpZRE3 sequence in a methylation dependent manner.

DNA methylation status strongly correlates with the associated gene expression, and aberrant methylation is linked to cancer (Esteller, 2008). Hypomethylation of specific genes was first shown to differentiate between cancerous and normal cells (Feinberg and Vogelstein, 1983). Hypermethylation of tumour suppressor genes such as *VHL* (Herman et al., 1994) and *p16* (Herman et al., 1995, Merlo et al., 1995) provided further evidence linking methylation status with expression and hence cell behaviour in cancer.

Whether this is merely an association, or a causative mechanism has long been a source of controversy. A recent paper appeared to confirm a causative link; methylation of 2 specific tumour suppressor genes (*HIC1* and *RassF1A)* was shown to transform mesenchymal stem cells into cancer stem cells (Teng et al., 2011).

The methylation of CpGs is thought to contribute to silencing in 3 main ways. The methylated CpG motif directly interferes with the interaction between transcription factors and DNA; this has been confirmed for a number of transcription factors, including CREB (Iguchi-Ariga and Schaffner, 1989), MLTF

(Watt and Molloy, 1988) and c-myc (Prendergast and Ziff, 1991). It also provides a binding site for methyl-binding proteins such as chromatin remodelling factors, and in this way it contributes to the packaging of the DNA region (Esteller, 2008). These binding proteins may also play a role in obstructing the access of transcription activating proteins. CpG methylation can also affect nucleosome positions on the DNA by affecting the rigidity of the double helix (Segal and Widom, 2009). In the same way as the methyl binding proteins, this may prevent transcription machinery loading onto promoters.

Although long associated with the silencing of genes, DNA methylation is increasingly being shown to be involved in more complicated mechanisms. The requirement of CpG methylation was first shown in the Rp promoter, allowing Zta to bind and activate transcription (Karlsson et al., 2008a, Bhende et al., 2004, Bhende et al., 2005). In this way, CpG methylation is creating a binding site for Zta, similar to the way it destroys binding sites for other transcription factors.

5.2. Results

5.2.1 Host methylation dependent binding protein in HEK293 extract

Nuclear extract from HEK293 cells was titrated in an EMSA with a double stranded oligonucleotide encompassing the central RpZRE3 core sequence, and 10 basepairs of cognate flank sequence on either side, in both the unmethylated and methylated state. A large complex is only visible with the methylated form of the probe, and increases in intensity with increasing amount of nuclear extract, denoted “M”. Conversely, a lower molecular weight complex can be seen only with the unmethylated probe, denoted “U”, but this complex could not be reliably reproduced. A cluster of 3 bands beneath the “M” complex are hard to differentiate, however there appears to be a complex formed only with the methylated oligonucleotide, denoted as “A” (Figure 5.1).

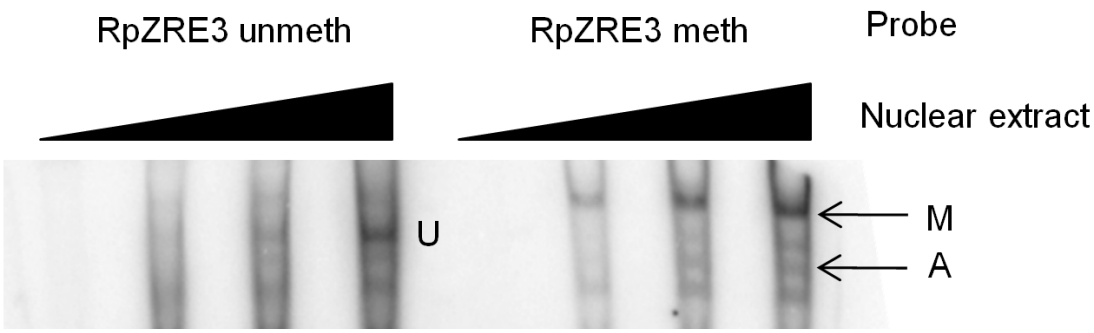


Figure 5.1 Host protein binding to ZRE affected by methylation.

Double stranded oligonucleotides were generated with the core RpZRE3 sequence and at least 10 nucleotides of cognate sequence on either side, with the two CpG motifs within the ZRE either unmethylated or methylated. Following radiolabelling, these were incubated with HEK293 nuclear extract.

To confirm the specificity of these unknown complexes for methylated RpZRE3, oligonucleotides were synthesised with a mutated core sequence that Zta has previously been shown not to bind, known as RpZRE3 mut core. Another mutant oligonucleotide (replacing the 3' AA with CT), Mut2, was synthesised to question whether the complexes were forming due to the double CpG nucleotide in the centre of RpZRE3 wild type sequence (Figure 5.2A).

Complex M was confirmed to require the central RpZRE3 motif in the methylated state, including the 3' AA. Complex A requires methylation but does not require the 3' AA (Figure 5.2B).

This is summarised by the model in Figure 5.3, and highlights the following question; is the binding confined to our definition of the site boundaries, or does it overlap with the flanking sequence specific to the *BRLF1* promoter? If the binding of M is specific to the core sequence only, then the protein that binds here would also bind in all of the host sites identified in chapter 4 that contain this sequence, and would suggest an equivalent host mechanism for methylated DNA dependent binding.

5.2.2 Host methylation dependent binding protein is specific to the RpZRE3 context

To question whether this core sequence alone is sufficient for host factor binding, previously used RpZRE3-site oligonucleotides CyclinL2 and HDAC2 were incubated with nuclear extract, in both the unmethylated and methylated state. These oligonucleotides were chosen after analysis by PROMO showed that these sequences had the least number of predicted binding sites for other host proteins. This EMSA suggests that complex M requires a specific

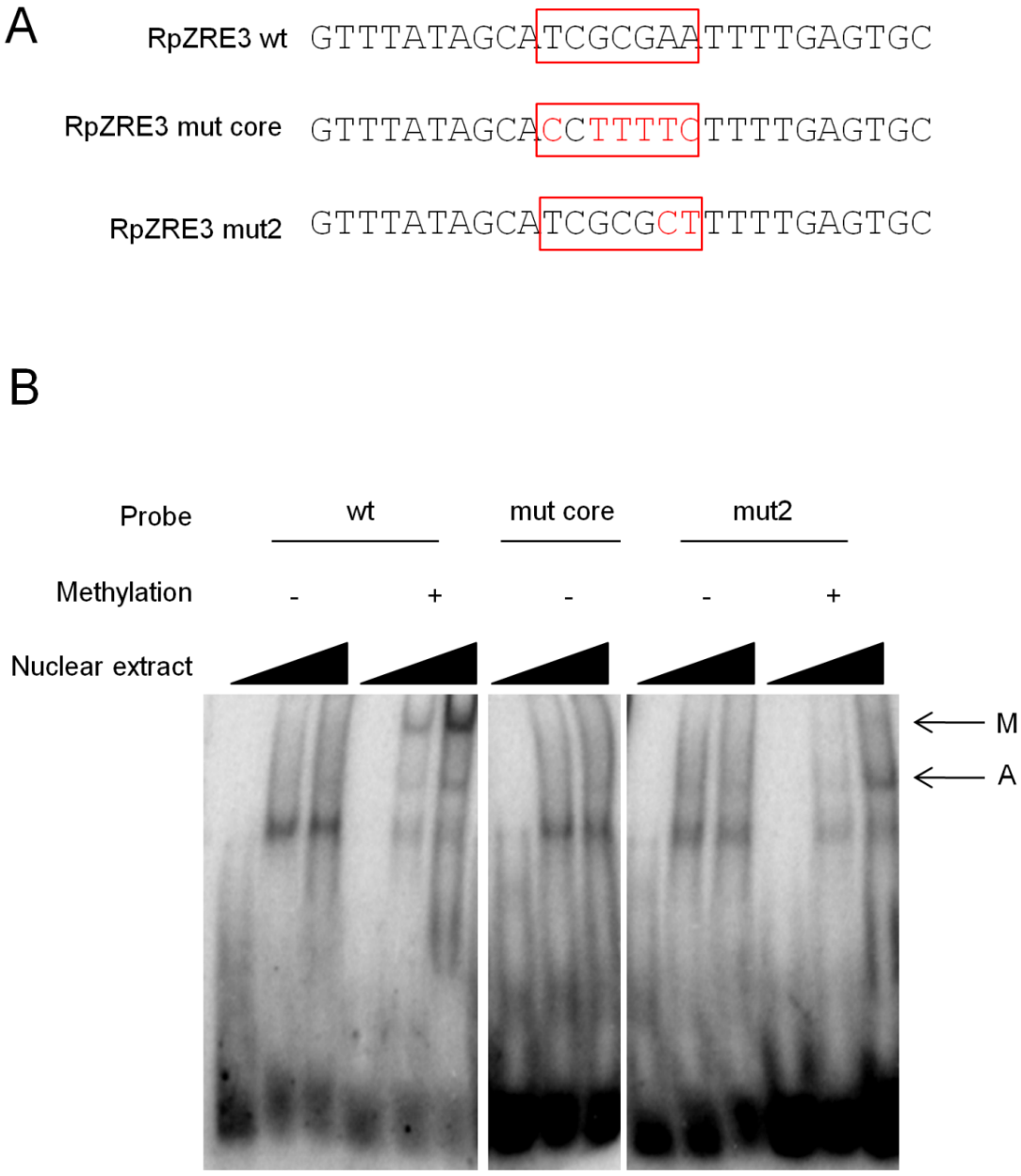


Figure 5.2 Mapping methylation dependent host protein binding.

A Double stranded oligonucleotides were generated with the core RpZRE3 sequence, or a mutated version of the sequence, and at least 10 nucleotides of cognate sequence on either side, with the 2 CpG motifs of RpZRE3 or RpZRE3 mut core CGCG either unmethylated or methylated. B Following radiolabelling, these were incubated with HEK293 nuclear extract, and subjected to EMSA analysis.

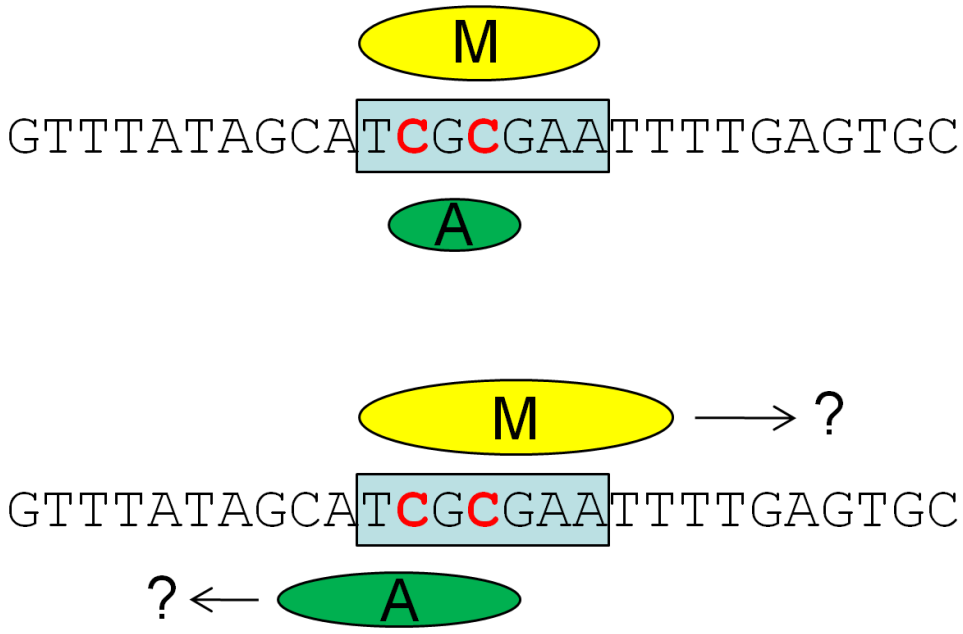


Figure 5.3 Host protein binding to RpZRE3 model.

Both complex A and M form in the presence of methylated CpG dinucleotides, indicated by red Cs. Complex M requires the 3' AA of the core RpZRE3 site, whereas complex A does not. The extent of the binding site of each protein has been not fully revealed.

environment that is provided by the flank of the RpZRE3 oligonucleotide, but that complex A is less dependent on a specific flanking sequence (Figure 5.4). This implies that complex M is specific to the RpZRE3 in the *BRLF1* promoter, requiring part of the 3' flank to the core RpZRE3 site, and may not bind to all instances where the sequence can be found in the host genome. In contrast, complex A is unaffected by the change in flank. This, combined with the effect of the oligonucleotide mut2, suggests that A is specific to the sequence TCGCG.

5.2.3 C/EBPalpha as a candidate for complex M

At this stage of the investigation, the ability of C/EBPalpha to bind to the methylated version of a CpG containing CRE sequence (TGACGTCA) was revealed (Rishi et al., 2010). C/EBPalpha is a bZIP protein, like Zta, and dimerises to bind to DNA (Ramji and Foka, 2002). The consensus sites of Zta and C/EBPalpha are very similar. A comparison using protein BLAST (NCBI) was carried out (Figure 5.5). Zta possesses C189 within the DNA binding domain, a residue shown to be required for interaction and activation through RpZRE3 (Karlsson et al., 2008a) and therefore postulated to be required for methylation dependent binding. However, C/EBPalpha contains a serine residue in the corresponding position.

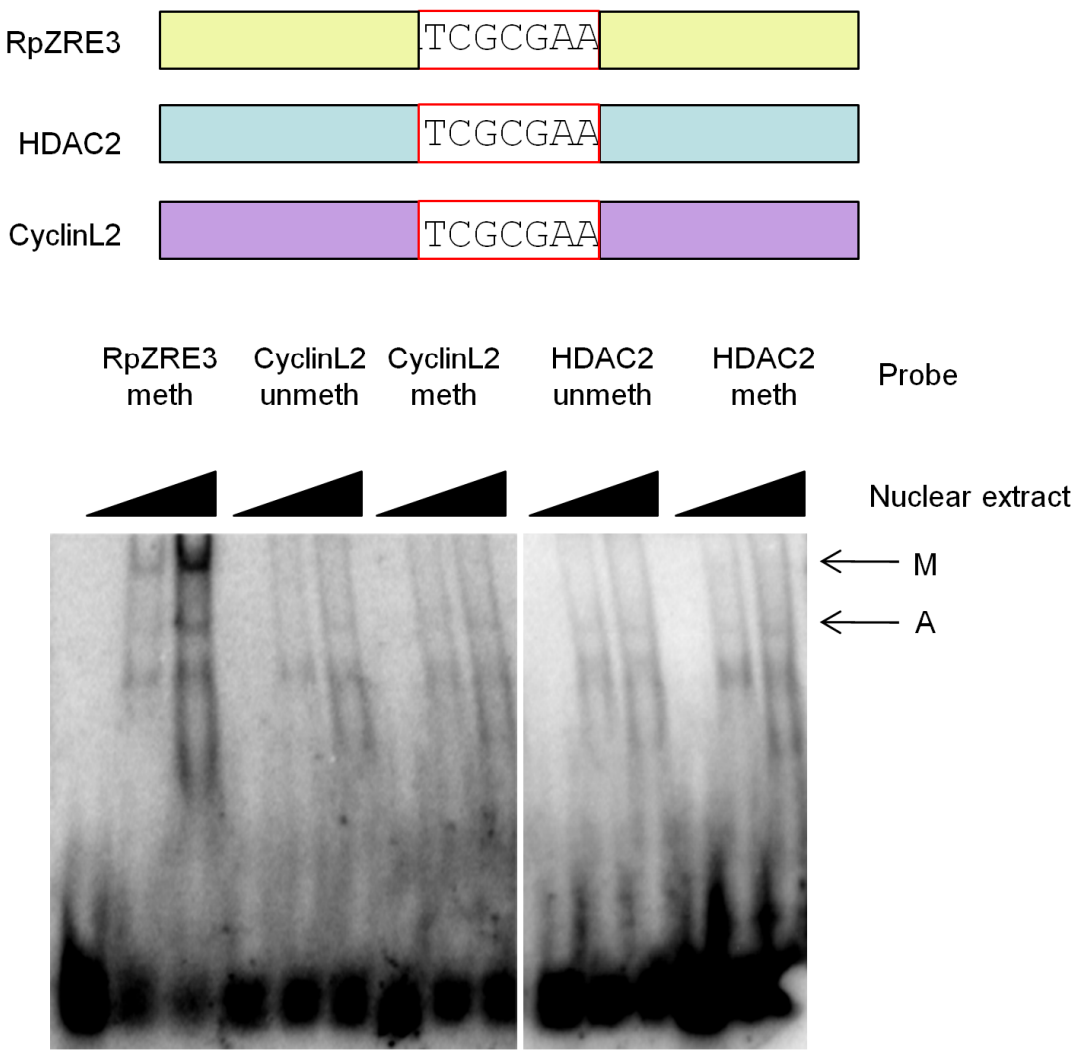


Figure 5.4 Methylation dependent host protein binding affected by flanking sequence.

Oligonucleotides with the same core RpZRE3 sequence but differing flank regions were synthesised in both the unmethylated and methylated state. After radiolabelling, oligonucleotides were incubated with nuclear extract from HEK293 cells and subjected to EMSA.

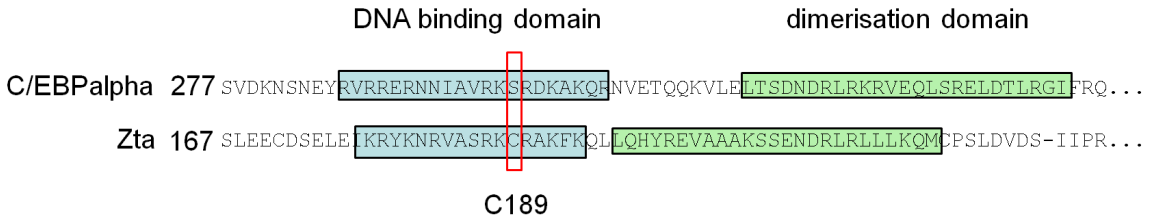


Figure 5.5 Protein BLAST alignment of C/EBPalpha and Zta

The DNA binding domains and bZIP dimerisation domains of both proteins are highlighted. The important C189 residue for the ability of Zta to bind to methylated RpZRE3 sequence is also highlighted, and the equivalent residue in C/EBPalpha.

5.2.4 PROMO analysis of RpZRE3 oligonucleotide

In an attempt to discover the identity of the unknown host binding protein, PROMO (Messeguer et al., 2002, Farre et al., 2003) was used to predict transcription factor binding sites in the RpZRE3 oligonucleotide, and both the mutated core versions (Figure 5.5). Several C/EBP isoforms had predicted sites in all the oligonucleotides (data not shown), however these were either confined to the flanking regions, or were very short recognition sequences, such as CGC, which had already been controlled for in the EMSA mutations. However, these C/EBP isoforms could be candidates for complex A. This short consensus site does not match the most recent publications on C/EBPalpha sites. Transcription factors with sites predicted only in the wild type RpZRE3 were identified (and not the HDAC2 and CyclinL2 oligonucleotides), and then filtered for only those containing a CpG dinucleotide. This identified 5 transcription factors, all with predicted binding sites that overlapped the core site and the 3' flank, making them candidates for complex M identification (Figure 5.6).

TRANSFAC ID	Factor name	Start position	End position	Dissimilarity	String
[T02042]	Cutl1	12	17	9.54%	GCGAAT
[T00691]	POU1F1a	13	17	7.51%	CGAAT
[T02690]	Dof2	13	22	5.57%	CGAATTTTGA
[T00630]	POU3F2	13	19	5.96%	CGAATTT
[T01042] and [T02104]	HSF1 (long) and (short)	10	17	8.96%	TCGCGAAT

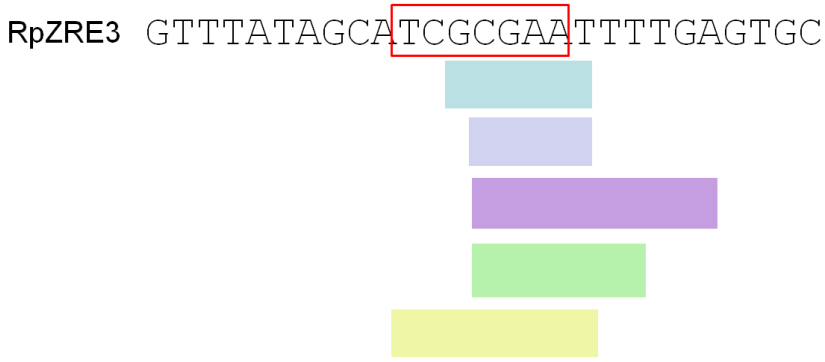


Figure 5.6 PROMO predictions of CpG containing transcription factor binding sites found in RpZRE3 oligonucleotide.

The RpZRE3 oligonucleotide sequence was input into the PROMO prediction tool, and the transcription factor predictions were compared with those found in the mutated oligonucleotides. Those found only in the wild type oligonucleotide and containing a CpG dinucleotide are shown here. The coloured blocks represent the respective transcription factor binding site position in the RpZRE3 oligonucleotide.

5.3. Discussion

The potential for a host binding protein with the same methylation dependent binding mechanism as Zta introduces complications to the existing silencing theories of DNA methylation, and also provides a potential evolutionary ancestor to the viral protein. A methylated promoter is not necessarily silenced, depending on the expression of methylation dependent binding factors.

In this chapter, a complex (labelled M) is shown to form with the methylated version of the RpZRE3 oligonucleotide, and that this complex formation requires the presence of the core site, not merely the methylated double CpG dinucleotide. However, the core RpZRE3 sequence is not sufficient for complex formation, implicated the flank of the oligonucleotide in the binding ability. A second complex, (labelled A) is also shown to form with the methylated RpZRE3 oligonucleotide, but only appears to require the first (5') 5 nucleotides of the site. The approximate binding positions of these complexes, with the potential binding dynamics with regards to Zta complex formation, is depicted in figure 5.7.

After these EMSAs were performed, C/EBPalpha was shown to bind in a methylation dependent manner to the CRE sequence TGACGTCA, and that this activity was important for the activation of methylated promoters during keratinocyte differentiation (Rishi et al., 2010). C/EBPbeta was also shown to bind in a similar manner in the initial investigation; however this was not pursued further in this paper. This CRE sequence matches the site found in the FN oligonucleotide. Zta has been shown in chapter 4 to bind to this site relatively weakly, in both the unmethylated and the methylated state. The

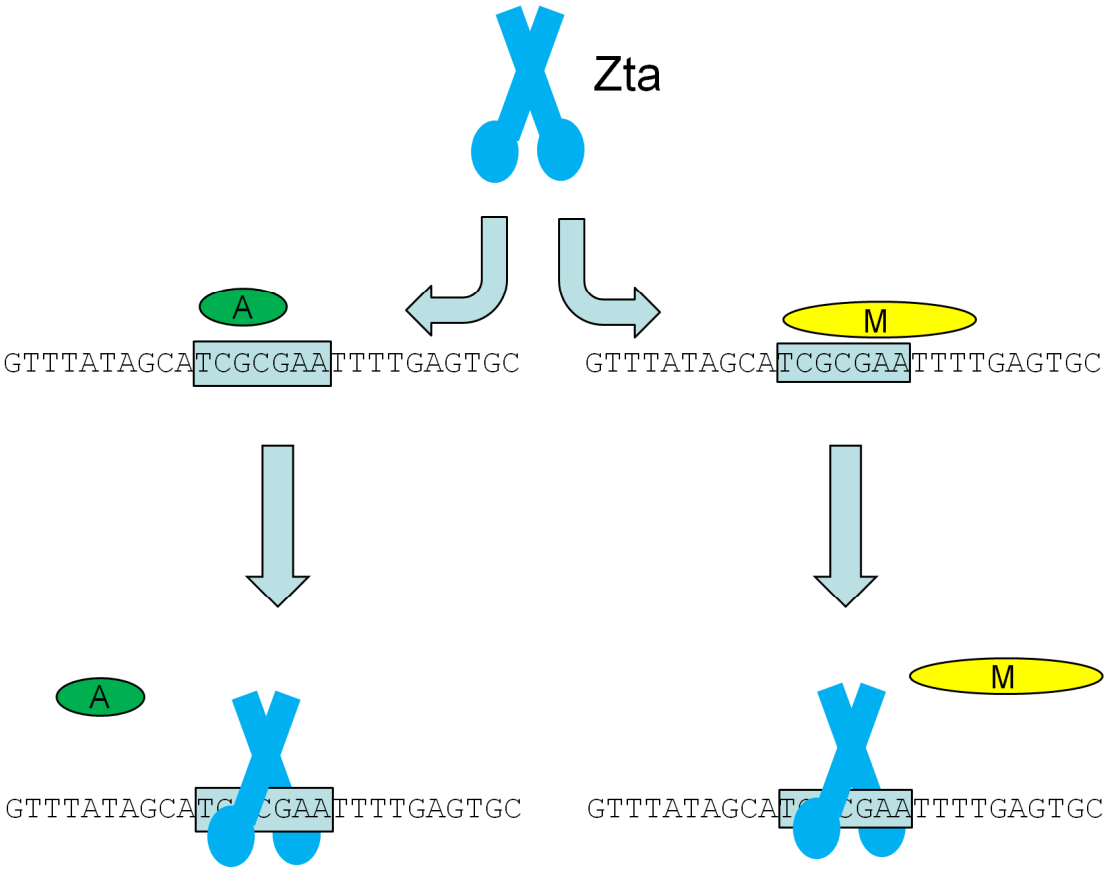


Figure 5.7 Host protein/Zta dynamics of binding to the methylated RpZRE3 oligonucleotide.

Potential displacement of these human factors by Zta upon *BZLF1* expression.

possibility of C/EBPalpha binding to the RpZRE3 oligonucleotide in a methylation dependent manner is considered here.

C/EBPalpha could have the potential to bind to RpZRE3 and other CpG containing sequences in the same methylation dependent manner as Zta. TTGCGCAA is the consensus C/EBPalpha binding site, to which it can bind in unmethylated form, but preferentially binds when it is methylated (Rishi et al., 2010). It contains the invariant GCG shown in the position frequency matrix; however it is difficult to know how to align this site with ZREs, due the fact it is an octamer compared to the heptamer ZREs. The site to which it was shown to bind in a methylation dependent manner contains a single central CpG dinucleotide, which again is difficult to align with known ZREs; the CpG dinucleotide is seen to be offset from the centre in all the specifically methylation dependent sites. The CpG motif is found centrally in the AP-1 like sites, investigated in chapter 4. These sites were bound by Zta in both the unmethylated and methylated state, suggesting that both the unmethylated and the methylated residue can be tolerated at this position.

The binding abilities of C/EBPalpha could be classified in the same way as Zta. C/EBPalpha binds to non-CpG containing, and therefore methylation independent, sites, like the Class I ZREs. The consensus site of C/EBPalpha contains a CpG, and can be bound in both the unmethylated and methylated states, Class II ZRE behaviour. Finally, the CREB consensus site is bound in a completely methylation dependent manner, like the Class III ZREs.

The C189 residue is known to be required for Zta interaction with RpZRE3 sequence (Karlsson et al., 2008a) however this has not been confirmed for

other ZREs (such as NaZREs (Dickerson et al., 2009)). The fact that C/EBPalpha does not contain this residue in an equivalent position after aligning the protein sequences using BLAST may suggest that this residue is only required to stabilise the interaction of Zta with RpZRE3, whereas other bZIP transcription factors may not require this for interaction with the same site.

The consensus sequences for the PROMO predicted transcription factors were examined. Cutl1, POU1F1A, POU3F2 and HSF1 do not contain a CpG dinucleotide, however the predicted binding site does. PROMO allows a maximum of 15% dissimilarity from the consensus sequence, which obviously can lead to nucleotide differences in the predicted sequence. This does not necessarily mean that these proteins will not bind to this sequence, but it has not yet been confirmed and the effect of methylation is unknown. Cutl1, POU1F1A and POU3F2 are members of the homeodomain family of DNA binding proteins. Other members of this family of DNA binding proteins, Oct1 and Oct2, have been shown to be involved in control of *BRLF1* expression, but this is not via DNA binding; Oct1 activates lytic cycle, via interaction with Rta through Rta Response Elements (RREs) and Oct2 is thought to inhibit lytic cycle. Oct2 is restricted to B cells, suggesting a mechanism for latency establishment, whereas Oct1 is found in a wide variety of cell types (Robinson et al., 2011). HSF1 (Heat Shock Factor 1) contains a leucine zipper, and functions as a homotrimer when a cell undergoes heat shock or stress (Batista-Nascimento et al., 2011).

The predicted binding site of Dof2, a transcription factor currently found only in maize with an unusual zinc finger-like structure (Tapia-Ramirez et al., 1997),

contains a CpG dinucleotide at the very start of the consensus site. It is unknown whether the methylation of this CpG dinucleotide would affect the ability of Dof2 to bind.

Unfortunately all these predicted proteins appear to be unlikely candidates, due to differences between the consensus and predicted sites, or the unlikelihood of their occurrence in human epithelial cells. Based upon the optimisation carried out to improve the position frequency matrix for ZRE prediction in chapter 5, it is perhaps not surprising that PFMs in PROMO for the binding sites of other transcription factors may be found deficient. Alternatively, the protein that binds in these nuclear extract EMSAs may not be very well characterised, and therefore the binding sites not well predicted by PROMO.

Chapter 6: Discussion

The prediction of methylation dependent ZREs in a number of lytic promoters, supports the model that methylation is required for full lytic cycle activation, and the lack of methylation works to prevent lytic gene expression during primary infection.

Upon infection, the genome circularises and expresses some latency genes which help drive the virus into the latent phase of the life cycle. Zta has been shown to be present at this early stage just after infection (Halder et al., 2009) so therefore has the potential to bind to promoters and activate expression. However, as the genome is unmethylated, Zta cannot bind to all of the required response elements. As identified in chapter 5, the promoters of some viral genes required for full lytic replication, including components of the replication machinery, appear to only contain ZREs that require methylation for interaction with Zta. The unmethylated state of these promoters at this early stage prevents their expression. This in turn prevents the full cohort of lytic gene expression and replication, ensuring that gene products that induce an immune response are not expressed, allowing the viral genome to establish latency and remain hidden from the host. The genome becomes progressively methylated over the coming weeks, becoming almost fully methylated after 3 weeks (Kalla et al., 2010). At this stage an external stimulus, such as activation of the B cell receptor, could activate expression of Zta. Those genes containing only methylation dependent ZREs will now be successfully activated, as they are now methylated. Full lytic cycle will then ensue.

As discussed in chapter 4, the zta^{MUT} ChIP, performed by Bergbauer et al, has its limitations, and it was prudent to attempt a ChIP using full length Zta. This has been undertaken in the lab after I left. They have used conditions in which lytic cycle is activated by cross-linking of Akata cells with IgG, but replication is stalled with the addition of acyclovir. This allows a “snapshot” of the binding profile of Zta to be taken. The results so far have been interesting, and one key observation that verifies the ZRE position predictions has been that all the early lytic genes exhibit Zta ChIP signal greater than 1.75 reads per million. For comparison, OriLyt, a region we know relevant Zta binding is occurring, has a ChIP signal of approximately 2 reads per million.

An important factor to consider in ChIP-seq profiling of Zta is the different template environments at different points of the life cycle. Without a fully synchronised homogenous population of cells, a mixture of these templates will be present. Upon infection, the genome is unmethylated and naked with regards to histone and chromatin formation. The expression of Zta has been shown to be important in the immortalisation of the cell at this point (Kalla et al., 2010). During latency, the genome is methylated and bound by histones to form a chromatin structure, and this will be the state the template is in upon induction of lytic activation. After the lytic activation, the replicated genome will be naked and unmethylated. These 3 different situations could account for the ability of Zta to bind to some sites and not others.

Replication of the genome potentially provides a mechanism to shift the binding of Zta from these methylated response elements, some of which Zta has been shown to have a higher affinity for (Bergbauer et al., 2010), to those response

elements that do not require methylation. As the double stranded DNA genome is unzipped for replication to occur, that DNA will now become hemi-methylated. It has previously been shown that a hemi-methylated version of a methylation dependent response element causes a reduction in complex formation with Zta (Karlsson et al., 2008a). It is also likely that any DNA binding proteins will be displaced by the replication machinery during replication, including Zta. When the double stranded genome is reformed, Zta is now free to bind to unmethylated, potentially lower affinity binding sites. These response elements may control late lytic genes, those which require replication to occur before they are expressed.

Whether Zta binds to newly replicated genome, and whether it activates any expression, is an interesting question, which does not appear to have been addressed in the literature yet. Using BrdU, the newly replicated genome could be labelled, and then immunoprecipitated for. A Zta re-ChIP could then be performed to address this question. I would predict that the binding profile would change, with those regions containing methylation dependent ZREs losing signal, and those regions with non-methylation dependent ZREs gaining signal, shifting the distribution of Zta. However, the large number of unmethylated genomes may mean that the relative binding, and therefore signal, at each region is diluted, as only a finite amount of Zta is available to bind, unless new Zta is being expressed.

Some of the gene promoters identified with only methylated dependent sites in their promoter regions have previously been classified as late lytic genes; why would these genes be activated in a methylation dependent context, and how

would the stable RNA products only be detected when replication was allowed to take place? Some reasons for this could include a requirement of the expression of a true late lytic gene, only expressed post-replication, which is needed to stabilise the RNA product. Another explanation could be that the chromatin structure of the particular region is such that even though the DNA is methylated, the response element is hidden from Zta, and therefore unavailable for binding, until after replication has taken place to displace the chromatin structure.

Zta is the prototypic member of a family of transcription factors that interact with DNA in a methylation dependent manner. C/EBP alpha has recently been shown to share the same characteristics (Rishi et al., 2010). It has been suggested that the interaction between C/EBP alpha and it is methylated sequence elements are needed to activate tissue specific genes during differentiation. Although both proteins have the ability to recognize some DNA sequence elements only when the CpG motifs are methylated, and are both bZIP transcription factors, the consensus sites for methylation dependent binding are not the same. As shown in chapter 4 however, Zta does show some affinity for the C/EBPalpha site, although this interaction requires further investigation for full characterization. In chapter 5, it can be observed that a human factor has affinity for the methylated RpZRE3 site; perhaps this is C/EBP alpha, or another, as yet unknown, methylation dependent binding factor. To question this, I was interested to perform a DNA affinity pulldown, using biotinylated versions of the RpZRE3 oligonucleotide used in the EMSAs. The sample would then be prepared to send to mass spectroscopy to attempt to identify the proteins binding to this sequence. The conditions for this needed to

be optimised and a high quality sample collected to ensure a good mass spectroscopy reading, and therefore I did not have time to attempt this experiment.

The biphasic methylation cycle is observed for several different classes of viruses that establish latency (Fernandez et al., 2009). Yet even KSHV, which is closely related to EBV, does not contain a *BZLF1* homologue. Perhaps the recent discovery that a cellular transcription factor also has methylation dependent DNA binding properties (Rishi et al., 2010) implies that other viruses may rely on host methylation dependent transcription factors to differentially control the expression of their genomes during the establishment of latency or replication.

The C189S mutation has been observed to be methylation dependent for some ZREs, but not others. However, the ability to reactivate lytic cycle was shown to be diminished when the C189S mutation was used (Schelcher et al., 2005). Perhaps this suggests an additional role for Zta in lytic reactivation than specifically direct DNA binding. Zta is known to interact with other proteins, both viral and of the host; the C189 residue may be required for one of these key interactions. Alternatively, the C189 residue may only be required in binding to the methylated CpG in RpZRE3, and is involved in competition with other proteins for this particular site during reactivation. EMSAs performed with rabbit reticulocyte lysate (RRL) and wheatgerm would contain other proteins, which may be involved in the observed binding status. Bacterially expressed Zta was used in Wang et al. (Wang et al., 2005), in which they saw very little difference between wildtype and C189S mutation in the methylation dependent binding to

RpZRE2. This concurs with the finding in Schelcher et al (Schelcher et al., 2005) for RpZRE2. However, RpZRE3 binding was not addressed by Wang et al. It would be interesting to see if the reduction observed in chapter 4 in ZREs binding with C189S Zta *in vitro* translated in RRL or wheatgerm was replicated with bacterially expressed Zta.

Bibliography

ADAMSON, A. L., DARR, D., HOLLEY-GUTHRIE, E., JOHNSON, R. A., MAUSER, A., SWENSON, J. & KENNEY, S. 2000. Epstein-Barr virus immediate-early proteins BZLF1 and BRLF1 activate the ATF2 transcription factor by increasing the levels of phosphorylated p38 and c-Jun N-terminal kinases. *J Virol,* 74**,** 1224-33.

ALAZARD, N., GRUFFAT, H., HIRIART, E., SERGEANT, A. & MANET, E. 2003. Differential hyperacetylation of histones H3 and H4 upon promoter-specific recruitment of EBNA2 in Epstein-Barr virus chromatin. *J Virol,* 77**,** 8166-72.

ALLDAY, M. J. 2009. How does Epstein-Barr virus (EBV) complement the activation of Myc in the pathogenesis of Burkitt's lymphoma? *Semin Cancer Biol,* 19**,** 366-76.

AMON, W. & FARRELL, P. J. 2005. Reactivation of Epstein-Barr virus from latency. *Rev Med Virol,* 15**,** 149-56.

ASHBURNER, M., BALL, C. A., BLAKE, J. A., BOTSTEIN, D., BUTLER, H., CHERRY, J. M., DAVIS, A. P., DOLINSKI, K., DWIGHT, S. S., EPPIG, J. T., HARRIS, M. A., HILL, D. P., ISSEL-TARVER, L., KASARSKIS, A., LEWIS, S., MATESE, J. C., RICHARDSON, J. E., RINGWALD, M., RUBIN, G. M. & SHERLOCK, G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet,* 25**,** 25-9.

BAER, R., BANKIER, A. T., BIGGIN, M. D., DEININGER, P. L., FARRELL, P. J., GIBSON, T. J., HATFULL, G., HUDSON, G. S., SATCHWELL, S. C., SEGUIN, C. & ET AL. 1984. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature,* 310**,** 207-11.

BAILEY, S. G., VERRALL, E., SCHELCHER, C., RHIE, A., DOHERTY, A. J. & SINCLAIR, A. J. (2009) Functional Interaction Between Epstein-Barr Virus Replication Protein Zta and Host DNA-Damage Response Protein 53BP1. *J Virol*.

BAROZZI, P., POTENZA, L., RIVA, G., VALLERINI, D., QUADRELLI, C., BOSCO, R., FORGHIERI, F., TORELLI, G. & LUPPI, M. 2007. B cells and Herpesviruses: A model of lymphoproliferation. *Autoimmunity Reviews,* 7**,** 132-136.

BARR, H., HERMANN, A., BERGER, J., TSAI, H. H., ADIE, K., PROKHORTCHOUK, A., HENDRICH, B. & BIRD, A. 2007. Mbd2 contributes to DNA methylation-directed repression of the Xist gene. *Mol Cell Biol,* 27**,** 3750-7.

BATISTA-NASCIMENTO, L., NEEF, D. W., LIU, P. C., RODRIGUES-POUSADA, C. & THIELE, D. J. 2011. Deciphering human heat shock transcription factor 1 regulation via post-translational modification in yeast. *PLoS One,* 6**,** e15976.

BEN-SASSON, S. A. & KLEIN, G. 1981. Activation of the Epstein-Barr virus genome by 5-aza-cytidine in latently infected human lymphoid lines. *Int J Cancer,* 28**,** 131-5.

BERGBAUER, M., KALLA, M., SCHMEINCK, A., GOBEL, C., ROTHBAUER, U., ECK, S., BENET-PAGES, A., STROM, T. M. & HAMMERSCHMIDT, W. CpG-methylation regulates a class of Epstein-Barr virus promoters. *PLoS Pathog,* 6.

BERGBAUER, M., KALLA, M., SCHMEINCK, A., GOBEL, C., ROTHBAUER, U., ECK, S., BENET-PAGES, A., STROM, T. M. & HAMMERSCHMIDT, W. 2010. CpG-methylation regulates a class of Epstein-Barr virus promoters. *PLoS Pathog,* 6.

BHENDE, P. M., SEAMAN, W. T., DELECLUSE, H. J. & KENNEY, S. C. 2004. The EBV lytic switch protein, Z, preferentially binds to and activates the methylated viral genome. *Nat Genet,* 36**,** 1099-104.

BHENDE, P. M., SEAMAN, W. T., DELECLUSE, H. J. & KENNEY, S. C. 2005. BZLF1 activation of the methylated form of the BRLF1 immediate-early promoter is regulated by BZLF1 residue 186. *J Virol,* 79**,** 7338-48.

BLANCO, E., MESSEGUER, X., SMITH, T. F. & GUIGO, R. 2006. Transcription factor map alignment of promoter regions. *PLoS Comput Biol,* 2**,** e49.

BOEHMER, P. E. & NIMONKAR, A. V. 2003. Herpes virus replication. *IUBMB Life,* 55**,** 13-22.

BONNET, M., GUINEBRETIERE, J. M., KREMMER, E., GRUNEWALD, V., BENHAMOU, E., CONTESSO, G. & JOAB, I. 1999. Detection of Epstein-Barr virus in invasive breast cancers. *J Natl Cancer Inst,* 91**,** 1376-81.

BRODERICK, P., HUBANK, M. & SINCLAIR, A. J. 2009a. Effects of Epstein-Barr virus on host gene expression in Burkitt's lymphoma cell lines. *unpublished*.

BRODERICK, P., HUBANK, M. & SINCLAIR, A. J. 2009b. Effects of Epstein-Barr virus on host gene expression in Burkitt's lymphoma cell lines. *Chinese Journal of Cancer,* in press.

BURKITT, D. 1958. A sarcoma involving the jaws in African children. *Br J Surg,* 46**,** 218-23.

BURKITT, D. P. 1968. The epidemiology of Burkitt's lymphoma. *West Afr Med J Niger Pract,* 17**,** 258-9.

BURKITT, D. P. 1969. Etiology of Burkitt's lymphoma--an alternative hypothesis to a vectored virus. *J Natl Cancer Inst,* 42**,** 19-28.

BURKITT, D. P. 1985. The beginnings of the Burkitt's lymphoma story. *IARC Sci Publ***,** 11-5.

CALDWELL, R. G., WILSON, J. B., ANDERSON, S. J. & LONGNECKER, R. 1998. Epstein-Barr virus LMP2A drives B cell development and survival in the absence of normal B cell receptor signals. Immunity, 9, 405-11.

CASTAGNA, M., TAKAI, Y., KAIBUCHI, K., SANO, K., KIKKAWA, U. & NISHIZUKA, Y. 1982. Direct activation of calcium-activated, phospholipid-dependent protein kinase by tumor-promoting phorbol esters. *J Biol Chem,* 257**,** 7847-51.

CHALLBERG, M. D. 1986. A method for identifying the viral genes required for herpesvirus DNA replication. *Proc Natl Acad Sci U S A,* 83**,** 9094-8.

CHANG, Y., LEE, H. H., CHEN, Y. T., LU, J., WU, S. Y., CHEN, C. W., TAKADA, K. & TSAI, C. H. 2006. Induction of the early growth response 1 gene by Epstein-Barr virus lytic transactivator Zta. *J Virol,* 80**,** 7748-55.

CHANG, Y. N., DONG, D. L., HAYWARD, G. S. & HAYWARD, S. D. 1990. The Epstein-Barr virus Zta transactivator: a member of the bZIP family with unique DNA-binding specificity and a dimerization domain that lacks the characteristic heptad leucine zipper motif. *J Virol,* 64**,** 3358-69.

CHAU, C. M., ZHANG, X. Y., MCMAHON, S. B. & LIEBERMAN, P. M. 2006. Regulation of Epstein-Barr virus latency type by the chromatin boundary factor CTCF. *J Virol,* 80**,** 5723-32.

CHENE, A., DONATI, D., OREM, J., MBIDDE, E. R., KIRONDE, F., WAHLGREN, M. & BEJARANO, M. T. 2009. Endemic Burkitt's lymphoma as a polymicrobial disease: new insights on the interaction between Plasmodium falciparum and Epstein-Barr virus. *Semin Cancer Biol,* 19**,** 411-20.

CHI, T., LIEBERMAN, P., ELLWOOD, K. & CAREY, M. 1995. A general mechanism for transcriptional synergy by eukaryotic activators. *Nature,* 377**,** 254-7.

CHIU, Y. F., TUNG, C. P., LEE, Y. H., WANG, W. H., LI, C., HUNG, J. Y., WANG, C. Y., KAWAGUCHI, Y. & LIU, S. T. 2007. A comprehensive library of mutations of Epstein Barr virus. *J Gen Virol,* 88**,** 2463-72.

CLARK, S. J., HARRISON, J. & MOLLOY, P. L. 1997. Sp1 binding is inhibited by (m)Cp(m)CpG methylation. *Gene,* 195**,** 67-71.

COMB, M. & GOODMAN, H. M. 1990. CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2. *Nucleic Acids Res,* 18**,** 3975-82.

COUNTRYMAN, J. K., GRADOVILLE, L. & MILLER, G. 2008. Histone hyperacetylation occurs on promoters of lytic cycle regulatory genes in Epstein-Barr virus-infected cell lines which are refractory to disruption of latency by histone deacetylase inhibitors. *J Virol,* 82**,** 4706-19.

CRAWFORD, D. H. 2001. Biology and disease associations of Epstein-Barr virus. *Philos Trans R Soc Lond B Biol Sci,* 356**,** 461-73.

DAVID, J. & BURKITT, D. 1968. Burkitt's lymphoma: remissions following seemingly non-specific therapy. *Br Med J,* 4**,** 288-9.

DE JESUS, O., SMITH, P. R., SPENDER, L. C., ELGUETA KARSTEGL, C., NILLER, H. H., HUANG, D. & FARRELL, P. J. 2003. Updated Epstein-Barr virus (EBV) DNA sequence and analysis of a promoter for the BART (CST, BARF0) RNAs of EBV. *J Gen Virol,* 84**,** 1443-50.

DEYRUP, A. T. 2008. Epstein-Barr virus-associated epithelial and mesenchymal neoplasms. *Hum Pathol,* 39**,** 473-83.

DICKERSON, S. J., XING, Y., ROBINSON, A. R., SEAMAN, W. T., GRUFFAT, H. & KENNEY, S. C. 2009. Methylation-dependent binding of the epstein-barr virus BZLF1 protein to viral promoters. *PLoS Pathog,* 5**,** e1000356.

DOHM, J. C., LOTTAZ, C., BORODINA, T. & HIMMELBAUER, H. 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res,* 36**,** e105.

DOLAN, A., ADDISON, C., GATHERER, D., DAVISON, A. J. & MCGEOCH, D. J. 2006. The genome of Epstein-Barr virus type 2 strain AG876. *Virology,* 350**,** 164-70.

DOUET, V., HELLER, M. B. & LE SAUX, O. 2007. DNA methylation and Sp1 binding determine the tissue-specific transcriptional activity of the mouse Abcc6 promoter. *Biochem Biophys Res Commun,* 354**,** 66-71.

EPSTEIN, M. A. 2001. Historical background. *Philos Trans R Soc Lond B Biol Sci,* 356**,** 413-20.

EPSTEIN, M. A., ACHONG, B. G. & BARR, Y. M. 1964. Virus Particles in Cultured Lymphoblasts from Burkitt's Lymphoma. *Lancet,* 1**,** 702-3.

ESTELLER, M. 2008. Epigenetics in cancer. *N Engl J Med,* 358**,** 1148-59.

FARRE, D., ROSET, R., HUERTA, M., ADSUARA, J. E., ROSELLO, L., ALBA, M. M. & MESSEGUER, X. 2003. Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic Acids Res,* 31**,** 3651-3.

FARRELL, P. J. 2005. *Epstein-Barr Virus Genome*, In ROBERTSON, E.S, (ed.) Epstein-Barr Virus, Caister Academic Press.

FARRELL, P. J., ROWE, D. T., ROONEY, C. M. & KOUZARIDES, T. 1989. Epstein-Barr virus BZLF1 trans-activator specifically binds to a consensus AP-1 site and is related to c-fos. *Embo J,* 8**,** 127-32.

FEINBERG, A. P. & VOGELSTEIN, B. 1983. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature,* 301**,** 89-92.

FENG, W. H., ISRAEL, B., RAAB-TRAUB, N., BUSSON, P. & KENNEY, S. C. 2002. Chemotherapy induces lytic EBV replication and confers ganciclovir susceptibility to EBV-positive epithelial cell tumors. *Cancer Res,* 62**,** 1920-6.

FERNANDEZ, A. F., ROSALES, C., LOPEZ-NIEVA, P., GRANA, O., BALLESTAR, E., ROPERO, S., ESPADA, J., MELO, S. A., LUJAMBIO, A., FRAGA, M. F., PINO, I., JAVIERRE, B., CARMONA, F. J., ACQUADRO, F., STEENBERGEN, R. D., SNIJDERS, P. J., MEIJER, C. J., PINEAU, P., DEJEAN, A., LLOVERAS, B., CAPELLA, G., QUER, J., BUTI, M., ESTEBAN, J. I., ALLENDE, H., RODRIGUEZ-FRIAS, F., CASTELLSAGUE, X., MINAROVITS, J., PONCE, J., CAPELLO, D., GAIDANO, G., CIGUDOSA, J. C., GOMEZ-LOPEZ, G., PISANO, D. G., VALENCIA, A., PIRIS, M. A., BOSCH, F. X., CAHIR-MCFARLAND, E., KIEFF, E. & ESTELLER, M. 2009. The dynamic DNA methylomes of double-stranded DNA viruses associated with human cancer. *Genome Res,* 19**,** 438-51.

FINGEROTH, J. D., WEIS, J. J., TEDDER, T. F., STROMINGER, J. L., BIRO, P. A. & FEARON, D. T. 1984. Epstein-Barr virus receptor of human B lymphocytes is the C3d receptor CR2. *Proc Natl Acad Sci U S A,* 81**,** 4510-4.

FIXMAN, E. D., HAYWARD, G. S. & HAYWARD, S. D. 1992. trans-acting requirements for replication of Epstein-Barr virus ori-Lyt. *J Virol,* 66**,** 5030-9.

FLANAGAN, J. M., COCCIARDI, S., WADDELL, N., JOHNSTONE, C. N., MARSH, A., HENDERSON, S., SIMPSON, P., DA SILVA, L., KHANNA, K., LAKHANI, S., BOSHOFF, C. & CHENEVIX-TRENCH, G. 2010. DNA methylome of familial breast cancer identifies distinct profiles defined by mutation status. *Am J Hum Genet,* 86**,** 420-33.

FLEMINGTON, E. & SPECK, S. H. 1990a. Autoregulation of Epstein-Barr virus putative lytic switch gene BZLF1. *J Virol,* 64**,** 1227-32.

FLEMINGTON, E. & SPECK, S. H. 1990b. Identification of phorbol ester response elements in the promoter of Epstein-Barr virus putative lytic switch gene BZLF1. *J Virol,* 64**,** 1217-26.

FLICEK, P., AKEN, B. L., BALLESTER, B., BEAL, K., BRAGIN, E., BRENT, S., CHEN, Y., CLAPHAM, P., COATES, G., FAIRLEY, S., FITZGERALD, S., FERNANDEZ-BANET, J., GORDON, L., GRAF, S., HAIDER, S., HAMMOND, M., HOWE, K., JENKINSON, A., JOHNSON, N., KAHARI, A., KEEFE, D., KEENAN, S., KINSELLA, R., KOKOCINSKI, F., KOSCIELNY, G., KULESHA, E., LAWSON, D., LONGDEN, I., MASSINGHAM, T., MCLAREN, W., MEGY, K., OVERDUIN, B., PRITCHARD, B., RIOS, D., RUFFIER, M., SCHUSTER, M., SLATER, G., SMEDLEY, D., SPUDICH, G., TANG, Y. A., TREVANION, S., VILELLA, A., VOGEL, J., WHITE, S., WILDER, S. P., ZADISSA, A., BIRNEY, E., CUNNINGHAM, F., DUNHAM, I., DURBIN, R., FERNANDEZ-SUAREZ, X. M., HERRERO, J., HUBBARD, T. J., PARKER, A., PROCTOR, G., SMITH, J. & SEARLE, S. M. 2009. Ensembl's 10th year. *Nucleic Acids Res,* 38**,** D557-62.

FLICEK, P., AKEN, B. L., BEAL, K., BALLESTER, B., CACCAMO, M., CHEN, Y., CLARKE, L., COATES, G., CUNNINGHAM, F., CUTTS, T., DOWN, T., DYER, S. C., EYRE, T., FITZGERALD, S., FERNANDEZ-BANET, J., GRAF, S., HAIDER, S., HAMMOND, M., HOLLAND, R., HOWE, K. L., HOWE, K., JOHNSON, N., JENKINSON, A., KAHARI, A., KEEFE, D., KOKOCINSKI, F., KULESHA, E., LAWSON, D., LONGDEN, I., MEGY, K., MEIDL, P., OVERDUIN, B., PARKER, A., PRITCHARD, B., PRLIC, A., RICE, S., RIOS, D., SCHUSTER, M., SEALY, I., SLATER, G., SMEDLEY, D., SPUDICH, G., TREVANION, S., VILELLA, A. J., VOGEL, J., WHITE, S., WOOD, M., BIRNEY, E., COX, T., CURWEN, V., DURBIN, R., FERNANDEZ-SUAREZ, X. M., HERRERO, J., HUBBARD, T. J., KASPRZYK, A., PROCTOR, G., SMITH, J., URETA-VIDAL, A. & SEARLE, S. 2008. Ensembl 2008. *Nucleic Acids Res,* 36**,** D707-14.

FRANCIS, A., RAGOCZY, T., GRADOVILLE, L., HESTON, L., EL-GUINDY, A., ENDO, Y. & MILLER, G. 1999. Amino acid substitutions reveal distinct functions of serine 186 of the ZEBRA protein in activation of early lytic cycle genes and synergy with the Epstein-Barr virus R transactivator. *J Virol,* 73**,** 4543-51.

FUJIMOTO, M., KITAZAWA, R., MAEDA, S. & KITAZAWA, S. 2005. Methylation adjacent to negatively regulating AP-1 site reactivates TrkA gene expression during cancer progression. *Oncogene,* 24**,** 5108-18.

FURNARI, F. B., ZACNY, V., QUINLIVAN, E. B., KENNEY, S. & PAGANO, J. S. 1994. RAZ, an Epstein-Barr virus transdominant repressor that modulates the viral reactivation mechanism. *J Virol,* 68**,** 1827-36.

GAO, X., WANG, H. & SAIRENJI, T. 2004. Inhibition of Epstein-Barr virus (EBV) reactivation by short interfering RNAs targeting p38 mitogen-activated protein kinase or c-myc in EBV-positive epithelial cells. *J Virol,* 78**,** 11798-806.

GARDINER-GARDEN, M. & FROMMER, M. 1987. CpG islands in vertebrate genomes. *J Mol Biol,* 196**,** 261-82.

GERSHBURG, E. & PAGANO, J. S. 2005. Epstein-Barr virus infections: prospects for treatment. *J Antimicrob Chemother,* 56**,** 277-81.

GLASER, G., VOGEL, M., WOLF, H. & NILLER, H. H. 1998. Regulation of the Epstein-Barr viral immediate early BRLF1 promoter through a distal NF1 site. *Arch Virol,* 143**,** 1967-83.

GRADOVILLE, L., KWA, D., EL-GUINDY, A. & MILLER, G. 2002. Protein kinase C-independent activation of the Epstein-Barr virus lytic cycle. *J Virol,* 76**,** 5612-26.

GRINSTEIN, S., PRECIADO, M. V., GATTUSO, P., CHABAY, P. A., WARREN, W. H., DE MATTEO, E. & GOULD, V. E. 2002. Demonstration of Epstein-Barr virus in carcinomas of various sites. *Cancer Res,* 62**,** 4876-8.

GRUFFAT, H., MANET, E. & SERGEANT, A. 2002. MEF2-mediated recruitment of class II HDAC at the EBV immediate early gene BZLF1 links latency and chromatin remodeling. *EMBO Rep,* 3**,** 141-6.

HAIDER, S., BALLESTER, B., SMEDLEY, D., ZHANG, J., RICE, P. & KASPRZYK, A. 2009. BioMart Central Portal--unified access to biological data. *Nucleic Acids Res,* 37**,** W23-7.

HALDER, S., MURAKAMI, M., VERMA, S. C., KUMAR, P., YI, F. & ROBERTSON, E. S. 2009. Early events associated with infection of Epstein-Barr virus infection of primary B-cells. *PLoS One,* 4**,** e7214.

HEATHER, J., FLOWER, K., ISAAC, S. & SINCLAIR, A. J. 2009. The Epstein-Barr virus lytic cycle activator Zta interacts with methylated ZRE in the promoter of host target gene egr1. *J Gen Virol,* 90**,** 1450-4.

HENNARD, C., PFUHL, T., BUETTNER, M., BECKER, K. F., KNOFEL, T., MIDDELDORP, J., KREMMER, E., NIEDOBITEK, G. & GRASSER, F. 2006. The antibody 2B4 directed against the Epstein-Barr virus (EBV)-encoded nuclear antigen 1 (EBNA1) detects MAGE-4: implications for studies on the EBV association of human cancers. *J Pathol,* 209**,** 430-5.

HERMAN, J. G., LATIF, F., WENG, Y., LERMAN, M. I., ZBAR, B., LIU, S., SAMID, D., DUAN, D. S., GNARRA, J. R., LINEHAN, W. M. & ET AL. 1994. Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc Natl Acad Sci U S A,* 91**,** 9700-4.

HERMAN, J. G., MERLO, A., MAO, L., LAPIDUS, R. G., ISSA, J. P., DAVIDSON, N. E., SIDRANSKY, D. & BAYLIN, S. B. 1995. Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. *Cancer Res,* 55**,** 4525-30.

HICKS, M. R., AL-MEHAIRI, S. S. & SINCLAIR, A. J. 2003. The zipper region of Epstein-Barr virus bZIP transcription factor Zta is necessary but not sufficient to direct DNA binding. *J Virol,* 77**,** 8173-7.

HOLLER, M., WESTIN, G., JIRICNY, J. & SCHAFFNER, W. 1988. Sp1 transcription factor binds DNA and activates transcription even when the binding site is CpG methylated. *Genes Dev,* 2**,** 1127-35.

HSU, M., WU, S. Y., CHANG, S. S., SU, I. J., TSAI, C. H., LAI, S. J., SHIAU, A. L., TAKADA, K. & CHANG, Y. 2008. Epstein-Barr virus lytic transactivator Zta enhances chemotactic activity through induction of interleukin-8 in nasopharyngeal carcinoma cells. *J Virol,* 82**,** 3679-88.

HUANG DA, W., SHERMAN, B. T. & LEMPICKI, R. A. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res,* 37**,** 1-13.

HUANG DA, W., SHERMAN, B. T. & LEMPICKI, R. A. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc,* 4**,** 44-57.

HUBBARD, T. J., AKEN, B. L., BEAL, K., BALLESTER, B., CACCAMO, M., CHEN, Y., CLARKE, L., COATES, G., CUNNINGHAM, F., CUTTS, T., DOWN, T., DYER, S. C., FITZGERALD, S., FERNANDEZ-BANET, J., GRAF, S., HAIDER, S., HAMMOND, M., HERRERO, J., HOLLAND, R., HOWE, K., JOHNSON, N., KAHARI, A., KEEFE, D., KOKOCINSKI, F., KULESHA, E., LAWSON, D., LONGDEN, I., MELSOPP, C., MEGY, K., MEIDL, P., OUVERDIN, B.,

PARKER, A., PRLIC, A., RICE, S., RIOS, D., SCHUSTER, M., SEALY, I., SEVERIN, J., SLATER, G., SMEDLEY, D., SPUDICH, G., TREVANION, S., VILELLA, A., VOGEL, J., WHITE, S., WOOD, M., COX, T., CURWEN, V., DURBIN, R., FERNANDEZ-SUAREZ, X. M., FLICEK, P., KASPRZYK, A., PROCTOR, G., SEARLE, S., SMITH, J., URETA-VIDAL, A. & BIRNEY, E. 2007. Ensembl 2007. *Nucleic Acids Res,* 35**,** D610-7.

HUBE, F., REVERDIAU, P., IOCHMANN, S. & GRUEL, Y. 2005. Improved PCR method for amplification of GC-rich DNA sequences. *Mol Biotechnol,* 31**,** 81-4.

HUTT, M. S. & BURKITT, D. P. 1973. Aetiology of Burkitt's lymphoma. *Lancet,* 1**,** 439.

IGUCHI-ARIGA, S. M. & SCHAFFNER, W. 1989. CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes Dev,* 3**,** 612-9.

ISRAEL, B. F. & KENNEY, S. C. 2003. Virally targeted therapies for EBV-associated malignancies. *Oncogene, 22***,** 5122-30.

JANZ, A., OEZEL, M., KURZEDER, C., MAUTNER, J., PICH, D., KOST, M., HAMMERSCHMIDT, W. & DELECLUSE, H. J. 2000. Infectious Epstein-Barr virus lacking major glycoprotein BLLF1 (gp350/220) demonstrates the existence of additional viral ligands. *J Virol,* 74**,** 10142-52.

JAVIER, R. T. & BUTEL, J. S. 2008. The history of tumor virology. *Cancer Res,* 68**,** 7693-706.

JONES, P. A. & BAYLIN, S. B. 2007. The epigenomics of cancer. *Cell,* 128**,** 683-92.

JONES, P. L., VEENSTRA, G. J., WADE, P. A., VERMAAK, D., KASS, S. U., LANDSBERGER, N., STROUBOULIS, J. & WOLFFE, A. P. 1998. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet,* 19**,** 187-91.

JONES, R. J., DICKERSON, S., BHENDE, P. M., DELECLUSE, H. J. & KENNEY, S. C. 2007. Epstein-Barr virus lytic infection induces retinoic acid-responsive genes through induction of a retinol-metabolizing enzyme, DHRS9. *J Biol Chem,* 282**,** 8317-24.

KAFUKO, G. W. & BURKITT, D. P. 1970. Burkitt's lymphoma and malaria. *Int J Cancer,* 6**,** 1-9.

KALLA, M., SCHMEINCK, A., BERGBAUER, M., PICH, D. & HAMMERSCHMIDT, W. 2010. AP-1 homolog BZLF1 of Epstein-Barr virus has two essential functions dependent on the epigenetic state of the viral genome. *Proc Natl Acad Sci U S A,* 107**,** 850-5.

KAPATAI, G. & MURRAY, P. 2007. Contribution of the Epstein Barr virus to the molecular pathogenesis of Hodgkin lymphoma. *J Clin Pathol,* 60**,** 1342-9.

KARLSSON, Q. H., SCHELCHER, C., VERRALL, E., PETOSA, C. & SINCLAIR, A. J. 2008a. Methylated DNA recognition during the reversal of epigenetic silencing is regulated by cysteine and serine residues in the Epstein-Barr virus lytic switch protein. *PLoS Pathog,* 4**,** e1000005.

KARLSSON, Q. H., SCHELCHER, C., VERRALL, E., PETOSA, C. & SINCLAIR, A. J. 2008b. The reversal of epigenetic silencing of the EBV genome is regulated by viral bZIP protein. *Biochem Soc Trans,* 36**,** 637-9.

KELLY, G. L., LONG, H. M., STYLIANOU, J., THOMAS, W. A., LEESE, A., BELL, A. I., BORNKAMM, G. W., MAUTNER, J., RICKINSON, A. B. & ROWE, M. 2009. An Epstein-Barr virus anti-apoptotic protein constitutively expressed in transformed cells and implicated in burkitt lymphomagenesis: the Wp/BHRF1 link. PLoS Pathog, 5, e1000341.

KENNEY, S. C. 2007. Reactivation and lytic replication of EBV. In: ARVIN A, C.-F. G., MOCARSKI E, MOORE PS, ROIZMAN B, WHITLEY R, YAMANISHI K, (ed.) Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. 2011/02/25 ed. Cambridge: Cambridge University Press.

KENNEY, S. C., HOLLEY-GUTHRIE, E., QUINLIVAN, E. B., GUTSCH, D., ZHANG, Q., BENDER, T., GIOT, J. F. & SERGEANT, A. 1992. The cellular oncogene c-myb can interact synergistically with the Epstein-Barr virus BZLF1 transactivator in lymphoid cells. *Mol Cell Biol,* 12**,** 136-46.

KILGER, E., PICHLER, B., GOETZ, A. E., RANK, N., WELTE, M., MORSTEDT, K., VETTER, H. O., GODJE, O., SCHMITZ, C., LAMM, P., ENGELSCHALK, E., MUEHLBEYER, D. & FREY, L. 1998. Procalcitonin as a marker of systemic inflammation after conventional or minimally invasive coronary artery bypass grafting. Thorac Cardiovasc Surg, 46, 130-3.KOUZARIDES, T. 2007. Chromatin modifications and their function. *Cell,* 128**,** 693-705.

KRAUS, R. J., PERRIGOUE, J. G. & MERTZ, J. E. 2003. ZEB negatively regulates the lytic-switch BZLF1 gene promoter of Epstein-Barr virus. *J Virol,* 77**,** 199-207.

KUTOK, J. L. & WANG, F. 2006. Spectrum of Epstein-Barr virus-associated diseases. *Annu Rev Pathol,* 1**,** 375-404.

LABRECQUE, L. G., BARNES, D. M., FENTIMAN, I. S. & GRIFFIN, B. E. 1995. Epstein-Barr virus in epithelial cell tumors: a breast cancer study. *Cancer Res,* 55**,** 39-45.

LAICHALK, L. L. & THORLEY-LAWSON, D. A. 2005. Terminal differentiation into plasma cells initiates the replicative cycle of Epstein-Barr virus in vivo. *J Virol,* 79**,** 1296-307.

LEE, K. A., BINDEREIF, A. & GREEN, M. R. 1988. A small-scale procedure for preparation of nuclear extracts that support efficient transcription and pre-mRNA splicing. *Gene Anal Tech,* 5**,** 22-31.

LENHARD, B. & WASSERMAN, W. W. 2002. TFBS: Computational framework for transcription factor binding site analysis. *Bioinformatics,* 18**,** 1135-6.

LEONARD, S., WEI, W., ANDERTON, J., VOCKERODT, M., ROWE, M., MURRAY, P. G. & WOODMAN, C. B. 2011. Epigenetic and Transcriptional Changes Which Follow Epstein-Barr Virus Infection of Germinal Center B Cells and Their Relevance to the Pathogenesis of Hodgkin's Lymphoma. *J Virol,* 85**,** 9568-77.

LI, D., QIAN, L., CHEN, C., SHI, M., YU, M., HU, M., SONG, L., SHEN, B. & GUO, N. 2009. Down-regulation of MHC class II expression through inhibition of CIITA transcription by lytic transactivator Zta during Epstein-Barr virus reactivation. *J Immunol,* 182**,** 1799-809.

LI, Z., VAN CALCAR, S., QU, C., CAVENEE, W. K., ZHANG, M. Q. & REN, B. 2003. A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. *Proc Natl Acad Sci U S A,* 100**,** 8164-9.

LIANG, G., WOLFGANG, C. D., CHEN, B. P., CHEN, T. H. & HAI, T. 1996. ATF3 gene. Genomic organization, promoter, and regulation. *J Biol Chem,* 271**,** 1695-701.

LIEBERMAN, P. M. & BERK, A. J. 1990. In vitro transcriptional activation, dimerization, and DNA-binding specificity of the Epstein-Barr virus Zta protein. *J Virol,* 64**,** 2560-8.

LIEBERMAN, P. M., HARDWICK, J. M., SAMPLE, J., HAYWARD, G. S. & HAYWARD, S. D. 1990. The zta transactivator involved in induction of lytic cycle gene expression in Epstein-Barr virus-infected lymphocytes binds to both AP-1 and ZRE sites in target promoter and enhancer regions. *J Virol,* 64**,** 1143-55.

LISTER, R., PELIZZOLA, M., DOWEN, R. H., HAWKINS, R. D., HON, G., TONTI-FILIPPINI, J., NERY, J. R., LEE, L., YE, Z., NGO, Q. M., EDSALL, L., ANTOSIEWICZ-BOURGET, J., STEWART, R., RUOTTI, V., MILLAR, A. H., THOMSON, J. A., REN, B. & ECKER, J. R. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature,* 462**,** 315-22.

LIU, E. T., POTT, S. & HUSS, M. 2010. Q&A: ChIP-seq technologies and the study of gene regulation. *BMC Biol,* 8**,** 56.

LONG, H. M., TAYLOR, G. S. & RICKINSON, A. B. 2011. Immune defence against EBV and EBV-associated disease. *Curr Opin Immunol,* 23**,** 258-64.

LUQMANI, Y. & SHOUSHA, S. 1995. Presence of epstein-barr-virus in breast-carcinoma. *Int J Oncol,* 6**,** 899-903.

MACK, G. J., OU, Y. & RATTNER, J. B. 2000. Integrating centrosome structure with protein composition and function in animal cells. *Microsc Res Tech,* 49**,** 409-19.

MAEDA, E., AKAHANE, M., KIRYU, S., KATO, N., YOSHIKAWA, T., HAYASHI, N., AOKI, S., MINAMI, M., UOZAKI, H., FUKAYAMA, M. & OHTOMO, K. 2009. Spectrum of Epstein-Barr virus-related diseases: a pictorial review. *Jpn J Radiol,* 27**,** 4-19.

MANET, E., GRUFFAT, H., TRESCOL-BIEMONT, M. C., MORENO, N., CHAMBARD, P., GIOT, J. F. & SERGEANT, A. 1989. Epstein-Barr virus bicistronic mRNAs generated by facultative splicing code for two transcriptional trans-activators. *EMBO J,* 8**,** 1819-26.

MAR, E. C., SUH, H. H. & HONG, J. S. 1992. Regulation of proenkephalin expression in C6 rat glioma cells. *Mol Cell Neurosci,* 3**,** 518-28.

MARGULIES, E. H., KARDIA, S. L. & INNIS, J. W. 2001. Identification and prevention of a GC content bias in SAGE libraries. *Nucleic Acids Res,* 29**,** E60-0.

MATYS, V., KEL-MARGOULIS, O. V., FRICKE, E., LIEBICH, I., LAND, S., BARRE-DIRRIE, A., REUTER, I., CHEKMENEV, D., KRULL, M., HORNISCHER, K., VOSS, N., STEGMAIER, P., LEWICKI-POTAPOV, B., SAXEL, H., KEL, A. E. & WINGENDER, E. 2006. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res,* 34**,** D108-10.

MCDONALD, C., KARSTEGL, C. E., KELLAM, P. & FARRELL, P. J. 2010. Regulation of the Epstein-Barr virus Zp promoter in B lymphocytes during reactivation from latency. *J Gen Virol,* 91**,** 622-9.

MERLO, A., HERMAN, J. G., MAO, L., LEE, D. J., GABRIELSON, E., BURGER, P. C., BAYLIN, S. B. & SIDRANSKY, D. 1995. 5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. *Nat Med,* 1**,** 686-92.

MESSEGUER, X., ESCUDERO, R., FARRE, D., NUNEZ, O., MARTINEZ, J. & ALBA, M. M. 2002. PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics,* 18**,** 333-4.

MILLER, C. L., LEE, J. H., KIEFF, E., BURKHARDT, A. L., BOLEN, J. B. & LONGNECKER, R. 1994. Epstein-Barr virus protein LMP2A regulates reactivation from latency by negatively regulating tyrosine kinases involved in sIg-mediated signal transduction. *Infect Agents Dis,* 3**,** 128-36.

MILLER, G., EL-GUINDY, A., COUNTRYMAN, J., YE, J. & GRADOVILLE, L. 2007. Lytic cycle switches of oncogenic human gammaherpesviruses. *Adv Cancer Res,* 97**,** 81-109.

MONTALVO, E. A., COTTAM, M., HILL, S. & WANG, Y. J. 1995. YY1 binds to and regulates cis-acting negative elements in the Epstein-Barr virus BZLF1 promoter. *J Virol,* 69**,** 4158-65.

MORRISON, T. E., MAUSER, A., KLINGELHUTZ, A. & KENNEY, S. C. 2004. Epstein-Barr virus immediate-early protein BZLF1 inhibits tumor necrosis factor alpha-induced signaling and apoptosis by downregulating tumor necrosis factor receptor 1. *J Virol,* 78**,** 544-9.

MURRAY, P. G. 2006. Epstein-Barr virus in breast cancer: artefact or aetiological agent? *J Pathol,* 209**,** 427-9.

MURRAY, P. G., LISSAUER, D., JUNYING, J., DAVIES, G., MOORE, S., BELL, A., TIMMS, J., ROWLANDS, D., MCCONKEY, C., REYNOLDS, G. M., GHATAURA, S., ENGLAND, D., CAROLL, R. & YOUNG, L. S. 2003. Reactivity with A monoclonal antibody to Epstein-Barr virus (EBV) nuclear antigen 1 defines a subset of aggressive breast cancers in the absence of the EBV genome. *Cancer Res,* 63**,** 2338-43.

NAN, X., NG, H. H., JOHNSON, C. A., LAHERTY, C. D., TURNER, B. M., EISENMAN, R. N. & BIRD, A. 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature,* 393**,** 386-9.

NEMEROW, G. R. & COOPER, N. R. 1984. Early events in the infection of human B lymphocytes by Epstein-Barr virus: the internalization process. *Virology,* 132**,** 186-98.

ODUMADE, O. A., HOGQUIST, K. A. & BALFOUR, H. H., JR. 2011. Progress and problems in understanding and managing primary Epstein-Barr virus infections. *Clin Microbiol Rev,* 24**,** 193-209.

PACKHAM, G., ECONOMOU, A., ROONEY, C. M., ROWE, D. T. & FARRELL, P. J. 1990. Structure and function of the Epstein-Barr virus BZLF1 protein. *J Virol,* 64**,** 2110-6.

PARI, G. S. & ANDERS, D. G. 1993. Eleven loci encoding trans-acting factors are required for transient complementation of human cytomegalovirus oriLyt-dependent DNA replication. *J Virol,* 67**,** 6979-88.

PARI, G. S., KACICA, M. A. & ANDERS, D. G. 1993. Open reading frames UL44, IRS1/TRS1, and UL36-38 are required for transient complementation of human cytomegalovirus oriLyt-dependent DNA synthesis. *J Virol,* 67**,** 2575-82.

PAULSON, E. J. & SPECK, S. H. 1999. Differential methylation of Epstein-Barr virus latency promoters facilitates viral persistence in healthy seropositive individuals. *J Virol,* 73**,** 9959-68.

PETOSA, C., MORAND, P., BAUDIN, F., MOULIN, M., ARTERO, J. B. & MULLER, C. W. 2006. Structural basis of lytic cycle activation by the Epstein-Barr virus ZEBRA protein. *Mol Cell,* 21**,** 565-72.

PHILLIPS, J. E. & CORCES, V. G. 2009. CTCF: master weaver of the genome. *Cell,* 137**,** 1194-211.

POHL, D. 2009. Epstein-Barr virus and multiple sclerosis. *J Neurol Sci,* 286**,** 62-4.

PRENDERGAST, G. C. & ZIFF, E. B. 1991. Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region. *Science,* 251**,** 186-9.

QUINLIVAN, E. B., HOLLEY-GUTHRIE, E. A., NORRIS, M., GUTSCH, D., BACHENHEIMER, S. L. & KENNEY, S. C. 1993. Direct BRLF1 binding is required for cooperative BZLF1/BRLF1 activation of the Epstein-Barr virus early promoter, BMRF1. *Nucleic Acids Res,* 21**,** 1999-2007.

RAGOCZY, T. & MILLER, G. 2001. Autostimulation of the Epstein-Barr virus BRLF1 promoter is mediated through consensus Sp1 and Sp3 binding sites. *J Virol,* 75**,** 5240-51.

RAMJI, D. P. & FOKA, P. 2002. CCAAT/enhancer-binding proteins: structure, function and regulation. *Biochem J,* 365**,** 561-75.

RAUCH, T. A., WU, X., ZHONG, X., RIGGS, A. D. & PFEIFER, G. P. 2009. A human B cell methylome at 100-base pair resolution. *Proc Natl Acad Sci U S A,* 106**,** 671-8.

REA, T. D., RUSSO, J. E., KATON, W., ASHLEY, R. L. & BUCHWALD, D. S. 2001. Prospective study of the natural history of infectious mononucleosis caused by Epstein-Barr virus. *J Am Board Fam Pract,* 14**,** 234-42.

RENNEKAMP, A. J. & LIBERMAN, P. M. 2010. Initiation of lytic DNA replication in Epstein-Barr virus: search for a common family mechanism. *Future Virology,* 5**,** 65-83.

RICE, P., LONGDEN, I. & BLEASBY, A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet,* 16**,** 276-7.

RICKINSON, A. & KIEFF, E. 2001a. Epstein-Barr virus. *In:* B FIELDS, D. K., P HOWLEY (ed.) *Fields Virology.* Philadelphia: Lippincot-Raven.

RICKINSON, A. & KIEFF, E. 2001b. Epstein-Barr virus and its replication. *In:* B FIELDS, D. K., P HOWLEY (ed.) *Fields Virology.* Philadelphia: Lippincot-Raven.

RISHI, V., BHATTACHARYA, P., CHATTERJEE, R., ROZENBERG, J., ZHAO, J., GLASS, K., FITZGERALD, P. & VINSON, C. 2010. CpG methylation of half-

CRE sequences creates C/EBPalpha binding sites that activate some tissue-specific genes. *Proc Natl Acad Sci U S A,* 107**,** 20311-6.

ROBINSON, A. R., KWEK, S. S., HAGEMEIER, S. R., WILLE, C. K. & KENNEY, S. C. 2011. Cellular transcription factor Oct-1 interacts with the Epstein-Barr virus BRLF1 protein to promote disruption of viral latency. *J Virol,* 85**,** 8940-53.

ROIZMAN, B. & BAINES, J. 1991. The diversity and unity of Herpesviridae. *Comp Immunol Microbiol Infect Dis,* 14**,** 63-79.

ROIZMAN, B., CARMICHAEL, L. E., DEINHARDT, F., DE-THE, G., NAHMIAS, A. J., PLOWRIGHT, W., RAPP, F., SHELDRICK, P., TAKAHASHI, M. & WOLF, K. 1981. Herpesviridae. Definition, provisional nomenclature, and taxonomy. The Herpesvirus Study Group, the International Committee on Taxonomy of Viruses. *Intervirology,* 16**,** 201-17.

ROMIGUIER, J., RANWEZ, V., DOUZERY, E. J. & GALTIER, N. 2010. Contrasting GC-content dynamics across 33 mammalian genomes: relationship with life-history traits and chromosome sizes. *Genome Res,* 20**,** 1001-9.

ROWE, M., KELLY, G. L., BELL, A. I. & RICKINSON, A. B. 2009. Burkitt's lymphoma: the Rosetta Stone deciphering Epstein-Barr virus biology. *Semin Cancer Biol,* 19**,** 377-88.

RYON, J. J., FIXMAN, E. D., HOUCHENS, C., ZONG, J., LIEBERMAN, P. M., CHANG, Y. N., HAYWARD, G. S. & HAYWARD, S. D. 1993. The lytic origin of herpesvirus papio is highly homologous to Epstein-Barr virus ori-Lyt: evolutionary conservation of transcriptional activation and replication signals. *J Virol,* 67**,** 4006-16.

SANDVEJ, K., KRENACS, L., HAMILTON-DUTOIT, S. J., RINDUM, J. L., PINDBORG, J. J. & PALLESEN, G. 1992. Epstein-Barr virus latent and replicative gene expression in oral hairy leukoplakia. *Histopathology,* 20**,** 387-95.

SARISKY, R. T. & HAYWARD, G. S. 1996. Evidence that the UL84 gene product of human cytomegalovirus is essential for promoting oriLyt-dependent DNA replication and formation of replication compartments in cotransfection assays. *J Virol,* 70**,** 7398-413.

SCHELCHER, C., VALENCIA, S., DELECLUSE, H. J., HICKS, M. & SINCLAIR, A. J. 2005. Mutation of a single amino acid residue in the basic region of the Epstein-Barr virus (EBV) lytic cycle switch protein Zta (BZLF1) prevents reactivation of EBV from latency. *J Virol,* 79**,** 13822-8.

SEGAL, E. & WIDOM, J. 2009. What controls nucleosome positions? *Trends Genet,* 25**,** 335-43.

SEGOUFFIN, C., GRUFFAT, H. & SERGEANT, A. 1996. Repression by RAZ of Epstein-Barr virus bZIP transcription factor EB1 is dimerization independent. *J Gen Virol,* 77 (Pt 7)**,** 1529-36.

SETO, E., YANG, L., MIDDELDORP, J., SHEEN, T. S., CHEN, J. Y., FUKAYAMA, M., EIZURU, Y., OOKA, T. & TAKADA, K. 2005. Epstein-Barr virus (EBV)-encoded BARF1 gene is expressed in nasopharyngeal carcinoma and EBV-associated gastric carcinoma tissues in the absence of lytic gene expression. *J Med Virol,* 76**,** 82-8.

SEYFERT, V. L., MCMAHON, S., GLENN, W., CAO, X. M., SUKHATME, V. P. & MONROE, J. G. 1990. Egr-1 expression in surface Ig-mediated B cell activation. Kinetics and association with protein kinase C activation. *J Immunol,* 145**,** 3647-53.

SHAH, K. M. & YOUNG, L. S. 2009. Epstein-Barr virus and carcinogenesis: beyond Burkitt's lymphoma. *Clin Microbiol Infect,* 15**,** 982-8.

SHARMA, S., KELLY, T. K. & JONES, P. A. 2010. Epigenetics in cancer. *Carcinogenesis,* 31**,** 27-36.

SHAULIAN, E. & KARIN, M. 2001. AP-1 in cell proliferation and survival. *Oncogene,* 20**,** 2390-400.

SHINO, Y., SHIRASAWA, H., KINOSHITA, T. & SIMIZU, B. 1997. Human papillomavirus type 16 E6 protein transcriptionally modulates fibronectin gene expression by induction of protein complexes binding to the cyclic AMP response element. *J Virol,* 71**,** 4310-8.

SINCLAIR, A. J. 2003. bZIP proteins of human gammaherpesviruses. *J Gen Virol,* 84**,** 1941-9.

SINCLAIR, A. J., BRIMMELL, M., SHANAHAN, F. & FARRELL, P. J. 1991. Pathways of activation of the Epstein-Barr virus productive cycle. *J Virol,* 65**,** 2237-44.

SISTA, N. D., PAGANO, J. S., LIAO, W. & KENNEY, S. 1993. Retinoic acid is a negative regulator of the Epstein-Barr virus protein (BZLF1) that mediates disruption of latent infection. *Proc Natl Acad Sci U S A,* 90**,** 3894-8.

SLOTS, J., SAYGUN, I., SABETI, M. & KUBAR, A. 2006. Epstein-Barr virus in oral diseases. *J Periodontal Res,* 41**,** 235-44.

SPECK, S. H., CHATILA, T. & FLEMINGTON, E. 1997. Reactivation of Epstein-Barr virus: regulation and function of the BZLF1 gene. *Trends Microbiol,* 5**,** 399-405.

SUM, E. Y., SEGARA, D., DUSCIO, B., BATH, M. L., FIELD, A. S., SUTHERLAND, R. L., LINDEMAN, G. J. & VISVADER, J. E. 2005. Overexpression of LMO4 induces mammary hyperplasia, promotes cell invasion, and is a predictor of poor outcome in breast cancer. *Proc Natl Acad Sci U S A,* 102**,** 7659-64.

SWAMINATHAN, S. 2005. Post-transcriptional gene regulation by gamma herpesviruses. *J Cell Biochem,* 95**,** 698-711.

TAKADA, K. 1984. Cross-linking of cell surface immunoglobulins induces Epstein-Barr virus in Burkitt lymphoma lines. *Int J Cancer,* 33**,** 27-32.

TAKAI, D. & JONES, P. A. 2002. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A,* 99**,** 3740-5.

TAO, Y., KASSATLY, R. F., CRESS, W. D. & HOROWITZ, J. M. 1997. Subunit composition determines E2F DNA-binding site specificity. *Mol Cell Biol,* 17**,** 6994-7007.

TAPIA-RAMIREZ, J., EGGEN, B. J., PERAL-RUBIO, M. J., TOLEDO-ARAL, J. J. & MANDEL, G. 1997. A single zinc finger motif in the silencing factor REST represses the neural-specific type II sodium channel promoter. *Proc Natl Acad Sci U S A,* 94**,** 1177-82.

TAYLOR, N., COUNTRYMAN, J., ROONEY, C., KATZ, D. & MILLER, G. 1989. Expression of the BZLF1 latency-disrupting gene differs in standard and defective Epstein-Barr viruses. *J Virol,* 63**,** 1721-8.

TAYLOR, N., FLEMINGTON, E., KOLMAN, J. L., BAUMANN, R. P., SPECK, S. H. & MILLER, G. 1991. ZEBRA and a Fos-GCN4 chimeric protein differ in their DNA-binding specificities for sites in the Epstein-Barr virus BZLF1 promoter. *J Virol,* 65**,** 4033-41.

TEMPERA, I. & LIEBERMAN, P. M. 2010. Chromatin organization of gammaherpesvirus latent genomes. *Biochim Biophys Acta,* 1799**,** 236-45.

TENG, I. W., HOU, P. C., LEE, K. D., CHU, P. Y., YEH, K. T., JIN, V. X., TSENG, M. J., TSAI, S. J., CHANG, Y. S. D., WU, C. S., SUN, H. S., TSAI, K. D., JENG, L. B., NEPHEW, K. P., HUANG, T. H., HSIAO, S. H. & LEU, Y. W. 2011. Targeted methylation of two tumor suppressor genes is sufficient to transform mesenchymal stem cells into cancer stem/initiating cells. *Cancer Res*.

TEYTELMAN, L., OZAYDIN, B., ZILL, O., LEFRANCOIS, P., SNYDER, M., RINE, J. & EISEN, M. B. 2009. Impact of chromatin structures on DNA processing for genomic analyses. *PLoS One,* 4**,** e6700.

THOMAS, C., DANKESREITER, A., WOLF, H. & SCHWARZMANN, F. 2003. The BZLF1 promoter of Epstein-Barr virus is controlled by E box-/HI-motif-binding factors during virus latency. *J Gen Virol,* 84**,** 959-64.

THORLEY-LAWSON, D. A. 2001. Epstein-Barr virus: exploiting the immune system. *Nat Rev Immunol,* 1**,** 75-82.

THORLEY-LAWSON, D. A. 2005. EBV the prototypical human tumor virus--just how bad is it? *J Allergy Clin Immunol,* 116**,** 251-61; quiz 262.

THORLEY-LAWSON, D. A. & BABCOCK, G. J. 1999. A model for persistent infection with Epstein-Barr virus: the stealth virus of human B cells. *Life Sci,* 65**,** 1433-53.

TOUSSIROT, E. & ROUDIER, J. 2008. Epstein-Barr virus in autoimmune diseases. *Best Pract Res Clin Rheumatol,* 22**,** 883-96.

TSAI, S. C., LIN, S. J., CHEN, P. W., LUO, W. Y., YEH, T. H., WANG, H. W., CHEN, C. J. & TSAI, C. H. 2009. EBV Zta protein induces the expression of interleukin-13, promoting the proliferation of EBV-infected B cells and lymphoblastoid cell lines. *Blood,* 114**,** 109-18.

URIER, G., BUISSON, M., CHAMBARD, P. & SERGEANT, A. 1989. The Epstein-Barr virus early protein EB1 activates transcription from different responsive elements including AP-1 binding sites. *EMBO J,* 8**,** 1447-53.

VAKOC, C. R., MANDAT, S. A., OLENCHOCK, B. A. & BLOBEL, G. A. 2005. Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. *Mol Cell,* 19**,** 381-91.

WALLING, D. M., FLAITZ, C. M. & NICHOLS, C. M. 2003. Epstein-Barr virus replication in oral hairy leukoplakia: response, persistence, and resistance to treatment with valacyclovir. *J Infect Dis,* 188**,** 883-90.

WALSER, J. C. & FURANO, A. V. 2010. The mutational spectrum of non-CpG DNA varies with CpG content. *Genome Res,* 20**,** 875-82.

WANG, P., DAY, L., DHEEKOLLU, J. & LIEBERMAN, P. M. 2005. A redox-sensitive cysteine in Zta is required for Epstein-Barr virus lytic cycle DNA replication. *J Virol,* 79**,** 13298-309.

WANG, Y. & LEUNG, F. C. 2004. An evaluation of new criteria for CpG islands in the human genome as gene markers. *Bioinformatics,* 20**,** 1170-7.

WATT, F. & MOLLOY, P. L. 1988. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev,* 2**,** 1136-43.

WENDELBURG, B. J. & VOS, J. M. 1998. An enhanced EBNA1 variant with reduced IR3 domain for long-term episomal maintenance and transgene expression of oriP-based plasmids in human cells. *Gene Ther,* 5**,** 1389-99.

WESTPHAL, E. M., BLACKSTOCK, W., FENG, W., ISRAEL, B. & KENNEY, S. C. 2000. Activation of lytic Epstein-Barr virus (EBV) infection by radiation and sodium butyrate in vitro and in vivo: a potential method for treating EBV-positive malignancies. *Cancer Res,* 60**,** 5781-8.

WESTPHAL, E. M., MAUSER, A., SWENSON, J., DAVIS, M. G., TALARICO, C. L. & KENNEY, S. C. 1999. Induction of lytic Epstein-Barr virus (EBV) infection in EBV-associated malignancies using adenovirus vectors in vitro and in vivo. *Cancer Res,* 59**,** 1485-91.

WHO. 2008. *Viral Cancers [online] Available: http://www.who.int/vaccine_research/diseases/viral_cancers/en/index1.html* [Online]. [Accessed 20 May 2009].

WU, F. Y., AHN, J. H., ALCENDOR, D. J., JANG, W. J., XIAO, J., HAYWARD, S. D. & HAYWARD, G. S. 2001. Origin-independent assembly of Kaposi's sarcoma-associated herpesvirus DNA replication compartments in transient cotransfection assays and association with the ORF-K8 protein and cellular PML. *J Virol,* 75**,** 1487-506.

WU, F. Y., CHEN, H., WANG, S. E., APRHYS, C. M., LIAO, G., FUJIMURO, M., FARRELL, C. J., HUANG, J., HAYWARD, S. D. & HAYWARD, G. S. 2003. CCAAT/enhancer binding protein alpha interacts with ZTA and mediates ZTA-induced p21(CIP-1) accumulation and G(1) cell cycle arrest during the Epstein-Barr virus lytic cycle. *J Virol,* 77**,** 1481-500.

XUE, S. A. & GRIFFIN, B. E. 2007. Complexities associated with expression of Epstein-Barr virus (EBV) lytic origins of DNA replication. *Nucleic Acids Res,* 35**,** 3391-406.

YIN, Y., MANOURY, B. & FAHRAEUS, R. 2003. Self-inhibition of synthesis and antigen presentation by Epstein-Barr virus-encoded EBNA1. Science, 301, 1371-4.

YOUNG, L. S. & MURRAY, P. G. 2003. Epstein-Barr virus and oncogenesis: from latent genes to tumours. *Oncogene,* 22**,** 5108-21.

YOUNG, L. S. & RICKINSON, A. B. 2004. Epstein-Barr virus: 40 years on. *Nat Rev Cancer,* 4**,** 757-68.

YU, X., MCCARTHY, P. J., LIM, H. J., IEMPRIDEE, T., KRAUS, R. J., GORLEN, D. A. & MERTZ, J. E. 2011. The ZIIR element of the Epstein-Barr virus BZLF1 promoter plays a central role in establishment and maintenance of viral latency. *J Virol,* 85**,** 5081-90.

YU, X., WANG, Z. & MERTZ, J. E. 2007. ZEB1 regulates the latent-lytic switch in infection by Epstein-Barr virus. *PLoS Pathog,* 3**,** e194.

YUAN, J., CAHIR-MCFARLAND, E., ZHAO, B. & KIEFF, E. 2006. Virus and cell RNAs expressed during Epstein-Barr virus replication. *J Virol,* 80**,** 2548-65.

ZALANI, S., COPPAGE, A., HOLLEY-GUTHRIE, E. & KENNEY, S. 1997. The cellular YY1 transcription factor binds a cis-acting, negatively regulating element in the Epstein-Barr virus BRLF1 promoter. *J Virol,* 71**,** 3268-74.

ZETTERBERG, H., JANSSON, A., RYMO, L., CHEN, F., KARLSSON, A., KLEIN, G. & BRODIN, B. 2002. The Epstein-Barr virus ZEBRA protein activates transcription from the early lytic F promoter by binding to a promoter-proximal AP-1-like site. *J Gen Virol,* 83**,** 2007-14.

ZHANG, Q., WANG, Y. C. & MONTALVO, E. A. 1999. Smubp-2 represses the Epstein-Barr virus lytic switch promoter. *Virology,* 255**,** 160-70.

ZHU, W. G., SRINIVASAN, K., DAI, Z., DUAN, W., DRUHAN, L. J., DING, H., YEE, L., VILLALONA-CALERO, M. A., PLASS, C. & OTTERSON, G. A. 2003. Methylation of adjacent CpG sites affects Sp1/Sp3 binding and activity in the p21(Cip1) promoter. *Mol Cell Biol,* 23**,** 4056-65.

ZUR HAUSEN, H., SCHULTE-HOLTHAUSEN, H., KLEIN, G., HENLE, W., HENLE, G., CLIFFORD, P. & SANTESSON, L. 1970. EBV DNA in biopsies of Burkitt tumours and anaplastic carcinomas of the nasopharynx. Nature, 228, 1056-8.

ZUR HAUSEN, H. 1999. Viruses in human cancers. *Eur J Cancer,* 35**,** 1878-85.

Appendix I

Ensembl Accession Code	Gene Name (HGNC)
ENSG00000106077	ABHD11
ENSG00000214450	AC005180.3
ENSG00000176533	AC005512.1
ENSG00000213050	AC005795.1
ENSG00000209593	AC007421.12
ENSG00000215498	AC007731.16
ENSG00000184771	AC008738.7
ENSG00000199405	AC018607.12
ENSG00000131351	AC020908.7
ENSG00000184536	AC022080.37
ENSG00000214194	AC073346.12
ENSG00000174558	AC087441.9
ENSG00000214944	AC093283.3
ENSG00000207167	AC099518.3
ENSG00000163074	AC104066.3
ENSG00000198468	AC104333.2
ENSG00000175901	AC127496.5
ENSG00000217812	AC134684.8
ENSG00000156709	AIFM1
ENSG00000212924	AL049872.3
ENSG00000206754	AL137783.12
ENSG00000180153	AL139281.7
ENSG00000206660	AL161932.15
ENSG00000215909	AL354712.18
ENSG00000203875	AL355615.12
ENSG00000208759	AL356957.27
ENSG00000202491	AL591670.10
ENSG00000189229	AL691432.53
ENSG00000175548	ALG10B
ENSG00000180715	AP001011.6
ENSG00000196763	AP002490.4
ENSG00000152056	AP1S3
ENSG00000183475	ASB7
ENSG00000166170	BAG5
ENSG00000175866	BAIAP2
ENSG00000095739	BAMBI
ENSG00000153094	BCL2L11
ENSG00000184887	BTBD6
ENSG00000127720	C12orf26
ENSG00000111678	C12orf57
ENSG00000100908	C14orf122
ENSG00000127150	C14orf153
ENSG00000138614	C15orf44
ENSG00000116922	C1orf109
ENSG00000117477	C1orf114
ENSG00000127423	C1orf135
ENSG00000163875	C1orf149
ENSG00000171806	C1orf156
ENSG00000183250	C21orf67.
ENSG00000198685	C3orf27
ENSG00000168994	C6orf145
ENSG00000175161	CADM2
ENSG00000148660	CAMK2G
ENSG00000162909	CAPN2
ENSG00000166734	CASC4
ENSG00000183323	CCDC125
ENSG00000120647	CCDC77
ENSG00000132141	CCT6B
ENSG00000008128	CDC2L2
ENSG00000163814	CDCP1
ENSG00000129596	CDO1
ENSG00000126001	CEP250
ENSG00000135837	CEP350
ENSG00000166037	CEP57
ENSG00000110172	CHORDC1
ENSG00000188517	COL25A1
ENSG00000103426	CORO7
ENSG00000121005	CRISPLD1
ENSG00000088766	CRLS1
ENSG00000176102	CSTF3
ENSG00000182224	CYB5D1
ENSG00000179091	CYC1
ENSG00000161692	DBF4B
ENSG00000145214	DGKQ
ENSG00000157379	DHRS1
ENSG00000132153	DHX30
ENSG00000160305	DIP2A
ENSG00000105993	DNAJB6
ENSG00000067334	DNTTIP2
ENSG00000047579	DTNBP1
ENSG00000139318	DUSP6
ENSG00000133740	E2F5
ENSG00000205609	EIF3CL

ENSG00000184110	EIF3S8
ENSG00000063046	EIF4B
ENSG00000110321	EIF4G2
ENSG00000198018	ENTPD7
ENSG00000218266	ERVWE2
ENSG00000012232	EXTL3
ENSG00000181104	F2R
ENSG00000135472	FAIM2
ENSG00000171634	FALZ (BPTF)
ENSG00000119632	FAM14A
ENSG00000165874	FAM35B.
ENSG00000128578	FAM40B
ENSG00000077458	FAM76B
ENSG00000127585	FBXL16
ENSG00000198793	FRAP1
ENSG00000068438	FTSJ1
ENSG00000162613	FUBP1
ENSG00000104290	FZD3
ENSG00000178950	GAK
ENSG00000125965	GDF5
ENSG00000164404	GDF9
ENSG00000102886	GDPD3
ENSG00000131067	GGTL3
ENSG00000131153	GINS2
ENSG00000198932	GPRASP1
ENSG00000137106	GRHPR
ENSG00000140307	GTF2A2
ENSG00000125812	GZF1
ENSG00000196591	HDAC2
ENSG00000164818	HEATR2
ENSG00000092148	HECTD1
ENSG00000183558	HIST2H2AA3
ENSG00000203812	HIST2H2AA4
ENSG00000216550	HIST2H2BC
ENSG00000218157	HIST2H2BD
ENSG00000164120	HPGD
ENSG00000207598	hsa-mir-124-3
ENSG00000199038	hsa-mir-210
ENSG00000207707	hsa-mir-639
ENSG00000169087	HSPBAP1
ENSG00000115457	IGFBP2
ENSG00000170604	IRF2BP1
ENSG00000162434	JAK1
ENSG00000096968	JAK2
ENSG00000066135	JMJD2A
ENSG00000182255	KCNA4
ENSG00000134901	KDELC1
ENSG00000137261	KIAA0319
ENSG00000163808	KIF15
ENSG00000054523	KIF1B
ENSG00000205629	LCMT1
ENSG00000140471	LINS1
ENSG00000131899	LLGL1
ENSG00000073350	LLGL2
ENSG00000169223	LMAN2
ENSG00000143013	LMO4
ENSG00000132128	LRRC41
ENSG00000170860	LSM3
ENSG00000164109	MAD2L1
ENSG00000139915	MAMDC1
ENSG00000135341	MAP3K7
ENSG00000129071	MBD4
ENSG00000076770	MBNL3
ENSG00000176274	MCART6
ENSG00000152034	MCHR2
ENSG00000112159	MDN1
ENSG00000123066	MED13L
ENSG00000145794	MEGF10
ENSG00000197889	MEIG1
ENSG00000127804	METT10D
ENSG00000151611	MMAA
ENSG00000169184	MN1
ENSG00000020426	MNAT1
ENSG00000070444	MNT
ENSG00000112651	MRPL2
ENSG00000172689	MS4A10
ENSG00000100330	MTMR3
ENSG00000125814	NAPB
ENSG00000154654	NCAM2
ENSG00000125912	NCLN
ENSG00000101882	NKAP
ENSG00000179873	NLRP11
ENSG00000011052	NME2
ENSG00000165271	NOL6
ENSG00000135318	NT5E
ENSG00000070081	NUCB2
ENSG00000112664	NUDT3
ENSG00000130703	OSBPL2
ENSG00000158639	PAGE5

ENSG00000099864	PALM
ENSG00000183570	PCBP3
ENSG00000185808	PIGP
ENSG00000124155	PIGT
ENSG00000104886	PLEKHJ1
ENSG00000182013	PNMAL1
ENSG00000137709	POU2F3
ENSG00000125686	PPARBP
ENSG00000058272	PPP1R12A
ENSG00000132825	PPP1R3D
ENSG00000170955	PRKCDBP
ENSG00000184304	PRKD1
ENSG00000132600	PRMT7
ENSG00000110958	PTGES3
ENSG00000164751	PXMP3
ENSG00000143811	PYCR2
ENSG00000151164	RAD9B
ENSG00000162437	RAVER2
ENSG00000162521	RBBP4
ENSG00000173933	RBM4
ENSG00000198771	RCSD1
ENSG00000137522	RNF121
ENSG00000219319	RP11-179H18.4
ENSG00000218456	RP11-262H14.10
ENSG00000220674	RP11-289I10.2
ENSG00000219599	RP11-348A7.2
ENSG00000218038	RP11-430G17.1
ENSG00000220544	RP11-452D2.2
ENSG00000219132	RP11-764K9.3
ENSG00000215036	RP13-140E4.1
ENSG00000220736	RP13-55H22.2
ENSG00000218048	RP3-407E4.4
ENSG00000218110	RP4-672J20.2
ENSG00000089157	RPLP0
ENSG00000102104	RS1
ENSG00000176700	SCAND2.
ENSG00000128228	SDF2L1
ENSG00000129657	SEC14L1
ENSG00000142864	SERBP1
ENSG00000197632	SERPINB2
ENSG00000104897	SF3A2
ENSG00000125454	SLC25A19
ENSG00000133460	SLC2A11
ENSG00000108604	SMARCD2
ENSG00000198952	SMG5
ENSG00000184602	SNN
ENSG00000028528	SNX1
ENSG00000115234	SNX17
ENSG00000196104	SPOCK3
ENSG00000187678	SPRY4
ENSG00000179119	SPTY2D1
ENSG00000040341	STAU2
ENSG00000123473	STIL
ENSG00000072786	STK10
ENSG00000115694	STK25
ENSG00000054654	SYNE2
ENSG00000171148	TADA3L
ENSG00000135801	TAF5L
ENSG00000162227	TAF6L
ENSG00000178913	TAF7
ENSG00000160551	TAOK1
ENSG00000092607	TBX15
ENSG00000100926	TM9SF1
ENSG00000164124	TMEM144
ENSG00000065600	TMEM206
ENSG00000112697	TMEM30A
ENSG00000156171	TMEM77
ENSG00000163472	TMEM79
ENSG00000164897	TMUB1
ENSG00000168591	TMUB2
ENSG00000185361	TNFAIP8L1
ENSG00000077097	TOP2B
ENSG00000167333	TRIM68
ENSG00000206557	TRIM71
ENSG00000144935	TRPC1
ENSG00000163728	TTC14
ENSG00000200368	U47924.1
ENSG00000150991	UBA52
ENSG00000165006	UBAP1
ENSG00000184182	UBE2F
ENSG00000174607	UGT8
ENSG00000147854	UHRF2
ENSG00000111981	ULBP1
ENSG00000100373	UPK3A

ENSG00000156467	UQCRB
ENSG00000173660	UQCRH
ENSG00000111237	VPS29
ENSG00000139719	VPS33A
ENSG00000163811	WDR43
ENSG00000206061	Z97652.9
ENSG00000213062	Z99572.1
ENSG00000176261	ZBTB8OS
ENSG00000120784	ZFP30

ENSG00000131127	ZNF141
ENSG00000204920	ZNF155
ENSG00000166526	ZNF3
ENSG00000198546	ZNF511
ENSG00000171425	ZNF581
ENSG00000173545	ZNF622
ENSG00000143373	ZNF687
ENSG00000169957	ZNF768
ENSG00000198346	ZNF813

Genes which contain a RpZRE3 core element in the promoter region, defined as 500bp upstream of the gene start.

Appendix II

Seq	Found Position	Class
TGAGCAA	1064	Class1
TGACACA	1539	Class1
TGACACA	2496	Class1
CCGCTCA	2531	Class3
ACGCTCA	2809	Class3
TGGCACA	3659	Class1
TGAGCAA	3970	Class1
TCGCTAA	4274	Class3
TCAGCGA	4385	Class3
TCGCCCA	4595	Class3
TCGCGAA	4943	Class3
TGTGCAA	5851	Class1
TGGCACA	6496	Class1
ACGCTCA	6994	Class3
TCAGCGA	8055	Class3
GCGCTCA	8097	Class2
TGTGTAA	8114	Class1
TGACTAA	8916	Class1
TTAGCAA	9158	Class1
TGTGTAA	9299	Class1
TGTGCAA	9598	Class1
GGAGCGA	9676	Class3
TGAGCAA	9845	Class1
TCGCCCA	10488	Class3
GGAGCGA	11057	Class3
TCGCAAA	11101	Class3
TCGCCCA	11235	Class3
TGGCACA	12009	Class1
GGAGCGA	12119	Class3
CGGGCGA	12587	Class3
CGGGCGA	12613	Class3
GGAGCGA	13263	Class3
CGGGCGA	13334	Class3
TCAGCGA	14532	Class3
TCGCCCA	14733	Class3
TGGCACA	15081	Class1
GGAGCGA	15191	Class3
CGGGCGA	15659	Class3
CGGGCGA	15685	Class3
GGAGCGA	16335	Class3
CGGGCGA	16406	Class3
TCAGCGA	17604	Class3
TCGCCCA	17805	Class3
TGGCACA	18153	Class1
GGAGCGA	18263	Class3
CGGGCGA	18731	Class3
CGGGCGA	18757	Class3
GGAGCGA	19407	Class3
CGGGCGA	19478	Class3
TCAGCGA	20676	Class3
TCGCCCA	20877	Class3
TGGCACA	21225	Class1
GGAGCGA	21335	Class3
CGGGCGA	21803	Class3
CGGGCGA	21829	Class3
GGAGCGA	22479	Class3
CGGGCGA	22550	Class3
TCAGCGA	23748	Class3
TCGCCCA	23949	Class3
TGGCACA	24297	Class1
GGAGCGA	24407	Class3
CGGGCGA	24875	Class3
CGGGCGA	24901	Class3
GGAGCGA	25551	Class3
CGGGCGA	25622	Class3
TCAGCGA	26820	Class3
TCGCCCA	27021	Class3
TGGCACA	27369	Class1
GGAGCGA	27479	Class3
CGGGCGA	27947	Class3
CGGGCGA	27973	Class3
GGAGCGA	28623	Class3
CGGGCGA	28694	Class3
TCAGCGA	29892	Class3
TCGCCCA	30093	Class3
TGGCACA	30441	Class1
GGAGCGA	30551	Class3
CGGGCGA	31019	Class3
CGGGCGA	31045	Class3
GGAGCGA	31695	Class3
CGGGCGA	31766	Class3
TCAGCGA	32964	Class3

TCGCCCA	33165	Class3
TGGCACA	33513	Class1
GGAGCGA	33623	Class3
CGGGCGA	34091	Class3
CGGGCGA	34117	Class3
GGAGCGA	34767	Class3
CGGGCGA	34838	Class3
TCGCCCA	35355	Class3
TGAGCCA	35627	Class1
TGGCACA	36973	Class1
TGAGTAA	36999	Class1
TGGCACA	40268	Class1
TGTGTAA	40545	Class1
TGAGCAA	40563	Class1
TGACACA	40601	Class1
TGACACA	40634	Class1
TGTGCAA	41182	Class1
TGAGCCA	41299	Class1
TGACACA	41341	Class1
TGAGCAA	41446	Class1
TGGCACA	42064	Class1
TGAGCCA	42719	Class1
TGACACA	42745	Class1
TGTGTAA	42977	Class1
AGACACA	44098	Class1
TCGCTCA	44488	Class2
TGGCACA	44931	Class1
TGGCACA	45013	Class1
TGTGTAA	45804	Class1
TGGCACA	46136	Class1
TGAGCCA	46229	Class1
TGACACA	46236	Class1
TGACACA	46649	Class1
CGGGCGA	46830	Class3
GTTGCAA	47938	Class1
ACGCTCA	48106	Class3
TCGCGAA	48276	Class3
TGAGCCA	48874	Class1
CCGCTCA	49180	Class3
TGAGCCA	49865	Class1
TGACTAA	49898	Class1
TCGCAAA	50112	Class3
CGTGCGA	50239	Class3
TCAGCGA	51867	Class3
TCGCGCA	52223	Class3
CGTGCGA	52313	Class3
CCGCTCA	53115	Class3
TCGCTCA	54329	Class2
TCGCGCA	54479	Class3
CGGGCGA	54867	Class3
ACGCTCA	55224	Class3
GGAGCGA	55612	Class3
TCGCAAA	55691	Class3
TCGCGAA	55922	Class3
GGAGCGA	55978	Class3
CCGCTCA	56000	Class3
TGACACA	56120	Class1
AGAGCGA	56585	Class3
TCGCAAA	56741	Class3
AGAGCGA	57203	Class3
GGAGCGA	57770	Class3
TGACACA	57916	Class1
TCGCAAA	58796	Class3
TGAGTAA	59231	Class1
CGTGCGA	59810	Class3
CGGGCGA	60129	Class3
TGTGTAA	60137	Class1
CGGGCGA	61602	Class3
CCGCTCA	62709	Class3
TCGCGCA	62739	Class3
CGGGCGA	62859	Class3
TCAGCGA	63099	Class3
AGAGCGA	63745	Class3
TGAGTAA	63836	Class1
TGTGTAA	64251	Class1
TCGCGCA	64356	Class3
TCAGCGA	64756	Class3
TGGCACA	64904	Class1
GCGCTCA	65095	Class2
TGGCACA	65510	Class1
CGAGCGA	65653	Class3
ACGCACA	66022	Class3
CCGCTCA	66379	Class3
TGAGCAA	66453	Class1
TCGCACA	67330	Class3
TGGCACA	67405	Class1
CCGCTCA	67430	Class3
TGAGCAA	67471	Class1
TGACTCA	67517	Class1
TGAGCAA	67534	Class1

TGAGCCA	68533	Class1
AGACACA	68865	Class1
AGACACA	68959	Class1
TGACACA	69642	Class1
GCGCTCA	69772	Class2
TGTGTAA	70425	Class1
TGGCACA	70771	Class1
TTAGCAA	70869	Class1
TCGCTCA	71579	Class2
TGACTCA	72140	Class1
TGGCACA	72260	Class1
TGAGCAA	72289	Class1
TGGCACA	72331	Class1
TCGCTCA	72357	Class2
AGACACA	72799	Class1
TCGCTCA	73151	Class2
CGCGCGA	73472	Class3
ACGCACA	73603	Class3
GGAGCGA	73674	Class3
TAAGCGA	73744	Class3
ACGCACA	73772	Class3
TGTGCAA	73895	Class1
GCGCTCA	74228	Class2
TCGCTCA	74259	Class2
CGCGCGA	74278	Class3
CCGCTCA	74313	Class3
TGGCACA	74474	Class1
TTAGCAA	74800	Class1
GCGCTCA	74961	Class2
GCGCTCA	75097	Class2
TCGCTCA	75398	Class2
TGGCACA	76293	Class1
TGACTAA	76489	Class1
CCGCTCA	77781	Class3
TCGCAAA	78712	Class3
TCGCCAA	78832	Class3
TGAGCCA	79610	Class1
TGACACA	79839	Class1
TGACTCA	79928	Class1
TCGCCCA	80217	Class3
TCAGCGA	80925	Class3
AGAGCGA	81114	Class3
TGGCACA	81283	Class1
TCGCTAA	81516	Class3
CGGGCGA	81523	Class3
TGACACA	81591	Class1
TCGCTCA	81676	Class2
TGAGCCA	82263	Class1
TGACACA	82824	Class1
TGAGCCA	83076	Class1
TGACACA	83158	Class1
TGAGCAA	83534	Class1
TGACTCA	83711	Class1
AGACACA	84759	Class1
ACGCTCA	85520	Class3
TGACACA	85672	Class1
CGAGCGA	85752	Class3
TCGCCCA	85961	Class3
CCGCTCA	86129	Class3
TTAGCAA	86512	Class1
AGACACA	86563	Class1
TCGCCAA	86656	Class3
ACGCACA	86722	Class3
TTAGCAA	87215	Class1
GGAGCGA	88057	Class3
TGTGTAA	88128	Class1
TCGCCCA	88189	Class3
TCGCCCA	88280	Class3
CGTGCGA	88298	Class3
TGACTCA	88879	Class1
TGACACA	89176	Class1
ACGCACA	90390	Class3
AGACACA	90998	Class1
TTAGCAA	91015	Class1
TGAGCCA	91028	Class1
TGAGCCA	91267	Class1
TGACACA	91564	Class1
AGAGCGA	91720	Class3
AGAGCGA	91867	Class3
CCGCTCA	92152	Class3
ACGCTCA	92675	Class3
CGGGCGA	92703	Class3
ACGCACA	93436	Class3
TGAGCCA	93923	Class1
TAAGCGA	94003	Class3
TCGCTCA	94083	Class2
TCGCGAA	94143	Class3
CGTGCGA	94635	Class3
ACGCTCA	95613	Class3
AGACACA	95714	Class1

GGAGCGA	96954	Class3
GCGCTCA	97772	Class2
TCGCTAA	97802	Class3
TGAGCAA	97820	Class1
TCGCAAA	98293	Class3
CGTGCGA	98544	Class3
CCGCTCA	98600	Class3
TCGCTAA	98776	Class3
AGAGCGA	99131	Class3
TCGCGAA	100070	Class3
TCGCCCA	100301	Class3
ACGCTCA	100875	Class3
TCGCGAA	101174	Class3
AGAGCGA	101557	Class3
CGAGCGA	102192	Class3
TCGCACA	102206	Class3
ACGCTCA	102328	Class3
TCGCTCA	102515	Class2
GGAGCGA	102573	Class3
TGACTAA	103356	Class1
AGACACA	104685	Class1
AGACACA	105007	Class1
CGGGCGA	105798	Class3
TCGCAAA	106516	Class3
TCGCCCA	106659	Class3
TCGCTCA	107508	Class2
TCGCGCA	107910	Class3
TGAGTAA	108242	Class1
TGAGCCA	109615	Class1
TCGCCCA	110120	Class3
TGACTCA	110187	Class1
TCAGCGA	110385	Class3
CGAGCGA	110422	Class3
TGAGCAA	110520	Class1
GCGCTCA	111586	Class2
TAAGCGA	111642	Class3
ACGCTCA	111830	Class3
TGGCACA	112019	Class1
TGGCACA	112041	Class1
CGGGCGA	112601	Class3
TGAGCCA	113412	Class1
TGGCACA	113561	Class1
TTAGCAA	113695	Class1
TCGCTCA	114508	Class2
TGAGCAA	115217	Class1
TGGCACA	115843	Class1
TGAGTAA	116169	Class1
TCAGCGA	116329	Class3
AGACACA	116442	Class1
TGACACA	117134	Class1
TCAGCGA	117301	Class3
TGAGCCA	117305	Class1
ACGCACA	117330	Class3
TGACTCA	117482	Class1
TGACTAA	117779	Class1
TGAGCCA	118010	Class1
TGGCACA	118770	Class1
CCGCTCA	118883	Class3
TGACTAA	119208	Class1
AGAGCGA	119628	Class3
TGAGCCA	119873	Class1
CGGGCGA	121488	Class3
TGACTAA	121678	Class1
GCGCTCA	121732	Class2
TCGCCCA	122177	Class3
ACGCTCA	122582	Class3
CGTGCGA	124443	Class3
TGAGCAA	124867	Class1
TGACACA	125177	Class1
GCGCTCA	125625	Class2
TAAGCGA	125682	Class3
TGAGCCA	125705	Class1
GTTGCAA	125960	Class1
GTTGCAA	125961	Class1
AGAGCGA	126320	Class3
ACGCTCA	126339	Class3
TCGCCCA	127124	Class3
TGACACA	127692	Class1
CGGGCGA	128065	Class3
TCGCCAA	128124	Class3
TGGCACA	128514	Class1
TGGCACA	128541	Class1
CCGCTCA	129105	Class3
GGAGCGA	129184	Class3
TCAGCGA	130682	Class3
TGAGCAA	130785	Class1
TGACTCA	131207	Class1
CGGGCGA	131551	Class3
TGACTCA	132419	Class1
TAAGCGA	132656	Class3

TGAGCAA	132865	Class1
TGGCACA	132939	Class1
CGTGCGA	132953	Class3
TGAGCCA	133577	Class1
TCGCCAA	134174	Class3
TCGCCCA	134722	Class3
TGAGCCA	134836	Class1
CGTGCGA	134890	Class3
TGAGCCA	135205	Class1
CCGCTCA	135309	Class3
TCGCGCA	135520	Class3
CGCGCGA	135588	Class3
TGTGCAA	135908	Class1
TCGCCAA	136720	Class3
ACGCTCA	136730	Class3
TGAGCCA	137354	Class1
TGACACA	137394	Class1
TCGCAAA	137420	Class3
TGACTCA	138228	Class1
TCGCTAA	138488	Class3
TGAGCCA	138537	Class1
TGAGCAA	138615	Class1
TTAGCAA	138682	Class1
CGGGCGA	139408	Class3
AGACACA	139577	Class1
TGAGCCA	140206	Class1
TGTGCAA	143463	Class1
TGAGCAA	143481	Class1
TGTGTAA	143553	Class1
TGTGCAA	144101	Class1
TGAGCCA	144218	Class1
TGACACA	144260	Class1
TGTGCAA	144875	Class1
TGGCACA	144985	Class1
TGAGCAA	145019	Class1
TGTGCAA	145048	Class1
TGAGCCA	145957	Class1
TGGCACA	146560	Class1
ACGCACA	147574	Class3
GGAGCGA	147692	Class3
AGACACA	147765	Class1
TGGCACA	148653	Class1
CGCGCGA	148804	Class3
GCGCTCA	149138	Class2
AGAGCGA	149824	Class3
CGGGCGA	150363	Class3
CGAGCGA	150440	Class3
TCGCCCA	150532	Class3
CGTGCGA	150576	Class3
TGGCACA	150639	Class1
TGACTCA	150945	Class1
AGAGCGA	151198	Class3
TCGCCCA	151207	Class3
AGACACA	151353	Class1
AGACACA	151359	Class1
TGACACA	151673	Class1
TGAGCAA	151806	Class1
TGTGCAA	151864	Class1
GTTGCAA	152777	Class1
GGAGCGA	152784	Class3
GGAGCGA	153004	Class3
TCAGCGA	153043	Class3
TGTGTAA	153679	Class1
AGAGCGA	153979	Class3
TCGCTCA	154723	Class2
TCGCCCA	155458	Class3
TGACACA	155644	Class1
TGAGCAA	156452	Class1
TCGCTCA	156537	Class2
TGGCACA	157098	Class1
CCGCTCA	158123	Class3
TCGCGCA	159076	Class3
CGAGCGA	159166	Class3
ACGCACA	159326	Class3
AGAGCGA	159547	Class3
GCGCTCA	159626	Class2
CGGGCGA	159847	Class3
TGACTCA	159919	Class1
TCGCCCA	160372	Class3
ACGCTCA	160846	Class3
TCAGCGA	161224	Class3
TCGCCCA	161241	Class3
TGTGCAA	161262	Class1
TGGCACA	161974	Class1
CCGCTCA	162449	Class3
AGAGCGA	162946	Class3
TCGCCCA	163069	Class3
CGTGCGA	163432	Class3
GGAGCGA	163832	Class3
TGAGTAA	164363	Class1

TCAGCGA	164458	Class3
TCGCAAA	164485	Class3
TCGCCCA	164520	Class3
TCGCTCA	165058	Class2
TCGCTCA	165122	Class2
TCGCCCA	165367	Class3
ACGCTCA	165400	Class3
TGAGCCA	165516	Class1
GTTGCAA	166363	Class1
TGACTCA	166416	Class1
TCGCTCA	167108	Class2
TGACACA	167192	Class1
TGGCACA	167265	Class1
TGGCACA	167602	Class1
TGACTAA	167701	Class1
AGAGCGA	168648	Class3
TGAGTAA	168744	Class1
TGACACA	169027	Class1
TCGCACA	169590	Class3
AGACACA	169878	Class1
TGACACA	170114	Class1
AGACACA	170416	Class1
TGACACA	170652	Class1
AGACACA	170939	Class1
TGACACA	171175	Class1
AGACACA	171477	Class1
TGACACA	171713	Class1

Positional predictions using ZRE_{32} in the EBV genome

Appendix III

cycle phase	Gene name	Class I	Class II	Class III	Total
LL	*BNRF1*	1	0	0	0
U	*EBER 1*	1	1	0	2
U	*EBER 2*	1	1	1	3
LL	*BCRF1*	3	0	1	4
L	*Cp*	0	0	3	3
U	*BWRF1*	0	0	3	3
L	*Wp*	0	0	1	1
EL	*BHLF1*	4	0	0	4
EL	*BHRF1*	4	0	0	4
EL	*BFLF2*	2	0	0	2
U	*BFRF1A*	4	0	0	4
EL	*BFLF1*	4	0	0	4
EL	*BFRF1*	4	0	0	4
EL	*BFRF2*	0	0	0	0
LL	*BFRF3*	1	0	1	2
L	*Qp*	2	0	2	4
U	*BPLF1*	1	0	2	3
LL	*BORF1*	0	0	3	3
U	*BOLF1*	0	0	2	2
EL	*BORF2*	1	0	1	2
EL	*BaRF1*	1	0	2	3
EL	*BMRF1*	4	0	2	6
LL	*BMRF2*	1	3	0	4
EL	*BSLF2/BMLF1*	4	1	0	5
EL	*BSLF1*	2	1	0	3
LL	*BSRF1*	1	2	2	5
LL	*BLRF1*	1	0	0	1
EL	*BLLF3*	2	0	0	2
LL	*BLRF2*	2	0	0	2
EL	*BLLF2*	0	0	1	1
LL	*BLLF1*	2	0	1	3
LL	*BZLF2*	0	0	0	0
EL	*BZLF1*	4	0	0	4
EL	*BRRF1*	0	0	2	2
EL	*BRLF1*	1	1	2	4
LL	*BRRF2*	1	1	2	4
LL	*BKRF2*	0	1	1	2
U	*BKRF3*	1	1	1	3
EL	*BKRF4*	0	0	3	3
LL	*BBRF1*	0	1	2	3
EL	*BBLF4*	0	0	3	3

LL	*BBRF2*	1	1	0	2
LL	*BBRF3*	0	0	2	2
EL	*BBLF2/BBLF3*	0	0	1	1
LL	*BBLF1*	0	2	0	2
EL	*BGLF5*	2	0	3	5
EL	*BGLF4*	0	1	1	2
U	*BGLF3.5*	2	0	2	4
LL	*BGRF1/BDRF1*	0	0	1	1
U	*BGLF3*	0	0	0	0
LL	*BGLF2*	0	1	0	1
LL	*BGLF1*	2	0	1	3
EL	*BDLF4*	1	0	0	1
U	*BDLF3.5*	3	0	2	5
LL	*BDLF3*	2	0	1	3
LL	*BDLF2*	0	0	0	0
LL	*BDLF1*	0	0	1	1
EL	*BcRF1.2*	2	0	0	2
LL	*BcLF1*	1	1	1	3
LL	*BTRF1*	0	1	1	2
LL	*BXLF2*	1	0	0	1
LL	*BXRF1*	1	0	1	2
EL	*BXLF1*	2	0	2	4
EL	*BVRF1*	2	0	2	4
LL	*BVRF2*	1	0	3	4
U	*BVLF1*	1	0	2	3
LL	*BdRF1*	1	0	0	1
LL	*BILF2*	4	0	1	5
U	*RPMS1*	2	0	1	3
EL	*LF3*	2	0	0	2
U	*LF2*	1	0	4	5
U	*LF1*	3	0	0	3
U	*BILF1*	1	0	3	4
U	*A73*	1	0	1	2
EL	*BALF5*	1	1	0	2
LL	*BALF4*	0	1	3	4
U	*BARF0*	0	0	2	2
U	*BALF3*	1	0	3	4
EL	*BALF2*	1	1	3	5
EL	*BALF1*	0	2	2	4
EL	*BARF1*	0	2	1	3
L	*LMP-2A*	1	0	0	1
EL	*BNLF2b*	2	2	0	4
EL	*BNLF2a*	2	1	0	3
L	*LMP-1*	1	1	0	2
L	*LMP-2B*	1	1	0	2

ZREs in the promoter regions of EBV genes including the positional predictions of the 3 AP1-like sites not included in the initial search. Changes in count of Class II ZREs are highlighted.

OPEN ACCESS Freely available online

PLoS one

Evaluation of a Prediction Protocol to Identify Potential Targets of Epigenetic Reprogramming by the Cancer Associated Epstein Barr Virus

Kirsty Flower[1☯], Elizabeth Hellen[2☯], Melanie J. Newport[2], Susan Jones[1], Alison J. Sinclair[1]*

1 School of Life Sciences, University of Sussex, Brighton, United Kingdom, **2** Brighton and Sussex Medical School, Brighton, United Kingdom

Abstract

Background: Epstein Barr virus (EBV) infects the majority of the human population, causing fatal diseases in a small proportion in conjunction with environmental factors. Following primary infection, EBV remains latent in the memory B cell population for life. Recurrent reactivation of the virus occurs, probably due to activation of the memory B-lymphocytes, resulting in viral replication and re-infection of B-lymphocytes. Methylation of the viral DNA at CpG motifs leads to silencing of viral gene expression during latency. Zta, the key viral protein that mediates the latency/reactivation balance, interacts with methylated DNA. Zta is a transcription factor for both viral and host genes. A sub-set of its DNA binding sites (ZREs) contains a CpG motif, which is recognised in its methylated form. Detailed analysis of the promoter of the viral gene *BRLF1* revealed that interaction with a methylated CpG ZRE (RpZRE3) is key to overturning the epigenetic silencing of the gene.

Methodology and Principal Findings: Here we question whether we can use this information to identify which host genes contain promoters with similar response elements. A computational search of human gene promoters identified 274 targets containing the 7-nucleotide RpZRE3 core element. DNA binding analysis of Zta with 17 of these targets revealed that the flanking context of the core element does not have a profound effect on the ability of Zta to interact with the methylated sites. A second juxtaposed ZRE was observed for one promoter. Zta was able to interact with this site, although co-occupancy with the RpZRE3 core element was not observed.

Conclusions/Significance: This research demonstrates 274 human promoters have the potential to be regulated by Zta to overturn epigenetic silencing of gene expression during viral reactivation from latency.

Citation: Flower K, Hellen E, Newport MJ, Jones S, Sinclair AJ (2010) Evaluation of a Prediction Protocol to Identify Potential Targets of Epigenetic Reprogramming by the Cancer Associated Epstein Barr Virus. PLoS ONE 5(2): e9443. doi:10.1371/journal.pone.0009443

Editor: Maria G. Masucci, Karolinska Institutet, Sweden

Received October 14, 2009; **Accepted** December 2, 2009; **Published** February 26, 2010

Funding: This work was funded by studentships from the University of Sussex and the Brighton and Sussex Medical School. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: a.j.sinclair@sussex.ac.uk

☯ These authors contributed equally to this work.

Introduction

Epstein Barr virus (EBV) infects and causes several diseases in humans including Burkitt's lymphoma, nasopharyngeal carcinoma, Hodgkin's disease, post-transplant lymphoproliferative disorder and infectious mononucleosis (glandular fever) [1–4]. Like other members of the gammaherpesviruses family, EBV infection persists for life following primary infection. The virus is maintained in a state of latency in memory B-lymphocytes and occasional reactivation and replication is considered to maintain the virus within individuals [5]. In EBV-induced lymphomas, EBV is also present in a latent state. In cell lines derived from these lymphomas, the viral genome is predominantly methylated [6–9]. The effect of silencing viral gene expression not only aids evasion from the immune system [10], but also prevents the destruction of tumour cells by viral replication. Indeed, reactivation of EBV from latency has been proposed as a route to treat EBV-associated lymphomas [11,12].

EBV latency is disrupted following physiological activation of B-lymphocytes, through the expression of Zta (BZLF1, ZEBRA, EB1, Z) [13–15]. Zta is a sequence-specific DNA-binding protein, which resembles the bZIP family of transcription factors and plays a critical role in the reactivation of viral gene expression and replication of the genome. Through direct interaction with Zta response elements (ZREs) in promoters, Zta regulates the expression of viral and cellular genes. Many host and viral promoters that have been evaluated to date contain ZREs within the proximal five hundred nucleotides of 5′ sequence. Thus far, eight have experimentally verified binding sites for Zta in their proximal promoter regions:

BSLF2+BMLF1 [16,17]; *BRLF1* [18]; *BZLF1* [17,19,20]; the joint promoter for *BHLF1* and *BHRF1* [19]; the lytic *EBNA1* promoter Fp [21]; *BRRF1* [22]; and *BMRF1* [23]. Furthermore, 6 host genes are directly regulated by Zta through ZREs in their promoters *DHRS9* [24]; *EGR1* [25,26]; *CIITA* [27]; *IL-8* [28]; *IL-10* [29]; and *IL-13* [30]. Zta interacts with a diverse range of ZREs

[31] and multiple sites exist in some Zta-responsive promoters, sometimes in close proximity. One example is the viral promoter for the *BZLF1* gene, Zp, which contains the two functional ZIIIA and ZIIIB ZREs within a total span of 20 nucleotides [17,19,20].

Zta has the unusual feature of interacting with a sub-set of ZREs that contain a methylated CpG motif [32]. In some cases Zta is able to interact with the non-methylated ZRE, while for other ZREs the interaction with Zta is dependent on methylation [22,26,32–34]. This has led to the classification of ZREs into three classes: class I ZREs do not contain a CpG motif; class II ZREs contain a CpG motif that is recognized in both the methylated and non-methylated states; and class III ZREs contain a CpG motif but are only recognized when methylated [35]. The ability of Zta to interact with methylated CpG-containing ZREs allows Zta to activate gene expression in the latent viral genome despite the repressive methylation status and thus overturn the epigenetic silencing of the viral genome [22,32–35].

To date, three genes with CpG-containing ZREs have been studied: EBV *BRLF1*, EBV *BRRF1* and human *EGR1* [22,26,32–35]. Of these, investigation of the regulation of the EBV gene *BRLF1* provided compelling evidence that the interaction of Zta with a methylated ZRE was instrumental in reactivating EBV into lytic cycle in B-lymphocytes [32–35]. The viral *BRLF1* gene includes three ZREs in the promoter proximal region [18,32]. Two of the ZREs contain CpG motifs; RpZRE2, is a class II ZRE and the other, RpZRE3 is a class III ZRE [35]. The ability of a single point-mutation in Zta to differentiate between the interaction of Zta with methylated and non-methylated RpZRE3 allowed the relevance of the interaction of Zta with this promoter to be established [34,36,37]. The viral *BRRF1* gene contains two CpG-containing ZREs, which Zta only interacts with in their methylated states [22]. The presence of a CpG-containing ZRE in the promoter of the human *EGR1* gene, which is recognized by Zta in its methylated form, suggests that EBV may overturn epigenetic silencing of host genes in order to modulate the host cell environment [26].

The RpZRE3 core element (TCGCGAA), which was clearly shown to be required for overturning epigenetic silencing of *BRLF1* in B-lymphocytes [34,36,37], was chosen to identify all human genes that contain an exact match to this CpG-containing ZRE within the −500 to +1 region of the promoter and evaluate whether they are recognised by Zta in their natural context.

Materials and Methods

Identification of Human Promoters Containing RpZRE3-Core Elements: Initial Sampling

Promoter regions (defined as −500 to +1 base pairs from the transcription start site) of a sample of human protein-coding genes in Ensembl (49) [38] were extracted using the Biomart data management system [39]. The sample represented 40% of the human genome. A search was made for all occurrences of an exact (forwards and reverse) match to identify those promoters with the TCGCGAA RpZRE3 core element.

Identification of Human Promoters Containing RpZRE3-Core Elements: Whole Genome Scanning

An identical screen was undertaken on the entire human genome from Ensembl (50) [40], resulting in the identification of 274 promoters which were designated the RpZRE3-promoter-274 data-set. The 7 nucleotides of the RpZRE3 core element, along with 10 flanking nucleotides on each side, were extracted from each of the 5 genes from the initial sampling. The Transcription Factor Binding Site (TFBS) Perl modules [41] were used to create a Position Weight Matrix (PWM) based on the entire length of these 27-nucleotide sequences. TFBS Perl modules were also used to search the promoter regions (−500 to +1) of all human protein coding genes in Ensembl 50 [40] which were extracted using Biomart [39]. Those promoters that matched the PWM with a threshold value >80% and contained an exact match to the RpZRE3-core element were additionally filtered using 2 criteria (a) the presence of CpG islands and (b) function (based on the over-representation of Gene Ontology (GO) terms [42]). The location of CpG islands were predicted using the EMBOSS program CpGPlot [43] with default options. Only genes with a CpG island present in the promoter were retained. The GO term annotations for molecular function and biological processes were extracted from Ensembl for each gene [40]. The number of genes in the RpZRE3 promoter dataset with each GO term annotation was compared with the total number of genes in Ensembl with the same GO term. The GO terms that occurred with a significantly higher frequency ($p<0.05$) in the RpZRE3 dataset, compared to entire genome, were defined as over-represented.

DNA Binding Analysis by EMSA on RpZRE3 Containing Promoters

Double stranded labelled DNA probes were made by labelling 6 pmol of oligonucleotide (27 nt long) at the 5′ end with [γ-^{32}P]ATP (30 μCi) using polynucleotide kinase (Roche). 12 pmol of the complementary oligonucleotide strand was added and incubated with the labelled single strand at 95°C for 2 minutes, 65°C for 10 minutes, and 37°C for 30 minutes to anneal the strands. The concentration of the probe was 33.3 nM. The oligonucleotides comprised the core 7-mer sequence surrounded by 20 nucleotides corresponding to the cognate flanking sequence for each promoter and were synthesised. Where indicated in the figure, the central CpG motif was methylated on both cytosines during synthesis (Sigma).

Zta protein was *in vitro* translated using wheatgerm extract (Promega). This was incubated with the labelled probe (at a final probe concentration of 3.3 nM) for 30 minutes at room temperature, before the sample was fractionated on an 8% native polyacrylamide gel at 100 volts for 1 hour. Following detection of the radio labelled DNA using a Storm phosphorimager, the relative signals were quantitated using ImageQuant software (GE Healthcare, UK).

Competition EMSAs were carried out with a 100X excess of a double stranded non-labelled oligonucleotide in addition to the standard EMSA reaction. Equal quantities of each of the complementary oligonucleotides were annealed in the same manner as the labelled probes and diluted to yield a solution concentration of 3.33 μM. This was added to the EMSA reaction at a final concentration of 333 nM (i.e. 100X excess).

Oligonucleotides

The oligonculeotides used were double strand versions of the following sequences (5′-3′):

XPC GGTGCGTCACTCGCGAAGTGGAATTTG
ZC3H8 GCTTCCCGGCTCGCGAAAGGGAGGACC
HDAC2 TCCCCCACTGTCGCGAAGCTCCCGCCC
MNT CCGCGGCGTCTCGCGAAGGGAGGGGCG
Cyclin L2 GGGCGGCTCCTCGCGAAGCTCCACGGC
RpZRE3 GTTTATAGCATCGCGAATTTTGAGTGC
CAPN2 CCGGGGAGGCTCGCGAATCGCGGTCCA
CDO1 CGTCCCAGCGTCGCGAACCACAGCGGC
FALZ GGCGCGCAGCTCGCGAAATGCCCGGCG
KIF1B GCTTCGGCCCTCGCGAAACTCCGCCCG
LLGL1 TCGGCCGGGCTCGCGAAGGGACGCCCG

LMO4 GGATCCCGGGTCGCGAAGGGCAGCCCA
MBD4 CCTCCTGCTCTTCGCGAACCGCCCCGC
PLEKHJ1 AGCCGCTCCCTCGCGAAAGTTGGCCCC
PRKD1 CTTCCTGGGGTCGCGAACTTCCCGGGC
SEC14L CGCCCGCTACTCGCGAAGCCCAGCCCG
TADA3L GCTGCGCTTCTCGCGAAAGGGCAGGCA
TOP2B CCGCGCCCCATCGCGAAGATCCGGAGC
XPCLMNTR GGTGCGTCACTCGCGAAGGGAGGGGCG
MNTLXPCR CCGCGGCGTCTCGCGAAGTGGAATTTG
XPC Mcore GGTGCGTCACCCCCTTAGTGGAATTTG
XPCLM3Mcore GGTCCGCCTCCCCCTTAGTGGAATTTG
AP1 mut GATCCACCCCTTAGAGGAAAACATACG

Prediction of Transcription Factor Binding Sites Using Promo

The transcription factor binding site prediction program PROMO [44] was used with the default setting of 15% dissimilarity value, to identify potential bZIP transcription factor binding sites within the XPC oligonucleotide.

Results

Identification of Human Promoters Containing the RpZRE3-Core Element: Initial Sampling

The initial search for RpZRE3-core elements in a sample of human promoters revealed 67 genes with an RpZRE3-core element. 5 of these that are involved in gene regulation were selected for DNA binding analysis. These 5 genes were ZC3H8 (ENSG00000144161), HDAC2 (ENSG00000196591), XPC (ENSG00000154767), MNT (ENSG00000070444) and CyclinL2 (ENSG00000116148) and together with the RpZRE3 element from the viral *BRLF1* promoter (Rp), formed the RpZRE3-promoter-6 data-set.

DNA-binding assays were undertaken with the methylated forms of each of the 5 human promoters using 27mer double strand oligonucleotides encompassing the 7-nucleotide core element and 10-nucleotide flanking region on each side together with RpZRE3 from Rp. Electrophoretic mobility shift assays (EMSA) were undertaken with *in vitro* translated Zta protein. Zta protein/DNA complexes formed readily with the oligonucleotides from all six promoters (Figure 1). The specificity of the assay is shown by the lack of complex formation with control protein. In addition, we undertook competition experiments with an excess of unlabelled oligonucleotides and included a version with a mutant ZRE to further probe the specificity of the interaction. This confirms that Zta interacts with all six RpZRE3 core elements specifically.

Identification of Human Promoters Containing RpZRE3-Core Elements: Whole Genome Scanning

The complete human genome was scanned for additional RpZRE3-core elements. This resulted in a data-set of 274 genes, denoted the RpZRE3-promoter-274 data-set (see Table S1). To assess whether there was an influence of flanking sequence on the interaction of Zta with these ZREs we undertook a systematic filtering process. The RpZRE3-promoter-274 dataset was first filtered by a Position Weight Matrix (PWM), which was generated from the RpZRE3-promoter-6 data-set. Matches were further filtered for genes with at least one CpG island in the promoter and an over-represented GO term. This resulted in 12 previously unidentified promoters and a further one that had been identified in the initial screen (Figure 2). Together with the promoters from the initial screen and the viral *BRLF1* promoter (RpZRE3-promoter-6 data-set), these form the RpZRE3-promoter-18 data-set.

Oligonucleotides were designed using the same principles with 10 nucleotides of flanking sequence on each side of the RpZRE3-core element, and the ability of Zta to interact with each site in its methylated form was assessed. The EMSA analysis revealed that all twelve of the newly identified methylated promoters were recognised by Zta (Figure 3).

This analysis demonstrated that for the RpZRE3-promoter-18 data-set, the flanking sequence surrounding the methylated core 7-mer element did not have a profound effect on the ability of Zta to interact with promoters and therefore the core element was sufficient to facilitate binding. This was illustrated by the generation of a PWM using sequence from the RpZRE3-promoter-18 data-set (Figure 4), which revealed negligible sequence conservation outside of the core element.

Recognition of the Non-Methylated Promoters

Zta is able to recognise many response elements which do not contain a CpG motif and therefore do not have a methylated core element (Class I ZREs) [15,35]. In addition, Zta recognises one CpG containing ZRE, RpZRE2, even in the absence of methylation (Class II ZREs) [33]. The ability of Zta to recognise the promoters in their non-methylated forms may impact on the ability of EBV to alter their gene expression, so we investigated the classification of ZRE for each of the 17 human promoters.

A series of DNA binding experiments with the oligonucleotides representing the human promoters in the RpZRE3-promoter-18 data-set, were undertaken comparing non-methylated with methylated RpZRE3-core elements. The interaction of Zta with the sites was quantitative, as demonstrated by the reduction in complex formation as the Zta protein concentration was titrated on each of the methylated promoters (Figure 5). All bar one display negligible binding to the non-methylated oligonucleotides (at least 10-fold lower than to the methylated sites) and can therefore be classified as class III ZREs.

Zta displayed a reproducible interaction with the non-methylated XPC oligonucleotide; the interaction reached 50% of the binding observed with the methylated oligonucleotide (Figure 5). This raises the possibility that the XPC core RpZRE3 element is influenced by the flanking sequence to behave as a class II ZRE.

Additional ZRE Juxtaposed with an RpZRE3-Core Element in Flanking Region

To identify whether sequences conferring methylation independent recognition was confined to a specific flank of XPC, a series of mutant oligonucleotide probes that exchanged flanking sequences between XPC and a class III site (MNT), that is not recognised when non-methylated, were designed. Analysis of the interaction with Zta by EMSA revealed that the ability to confer Zta binding resided within the 5′ sequence of XPC; the hybrid XPCLMNTR was able to bind but MNTLXPCR was not (Figure 6). This suggested that the 5′ XPC flanking sequence of the RpZRE3 element influenced binding. However, it was surprising to discover that mutation of the RpZRE3 element within XPC did not prevent interaction with Zta (Figure 6).

A further explanation for the interaction of Zta with the non-methylated XPC promoter is that an obscure ZRE is present in the 5′ flank. To address this further we attempted to identify putative ZREs in the XPC promoter using the transcription factor binding site prediction program PROMO [44,45]. Although this program contains a PWM for Zta binding sites, none were predicted in this sequence. However, PROMO predicted the

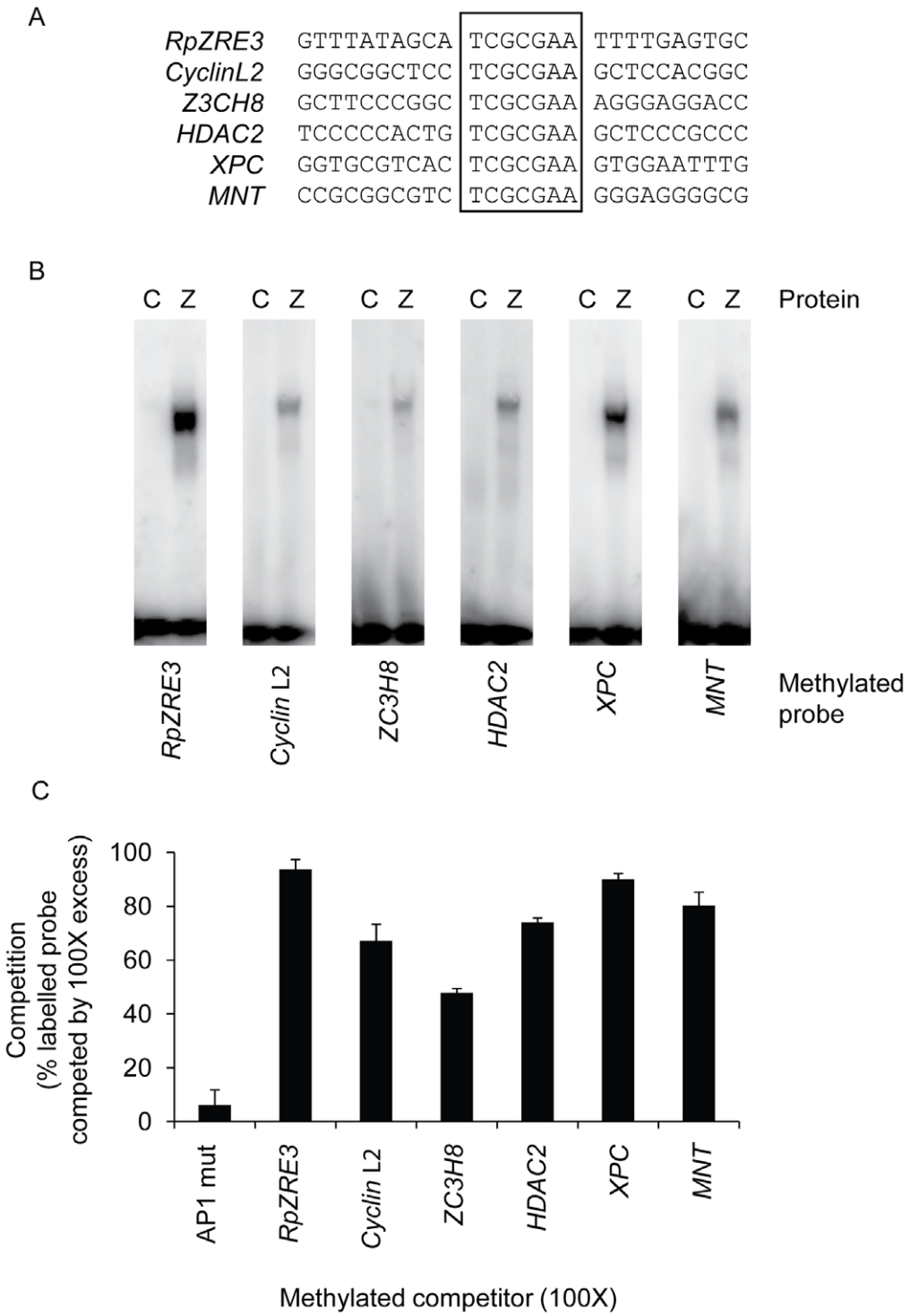
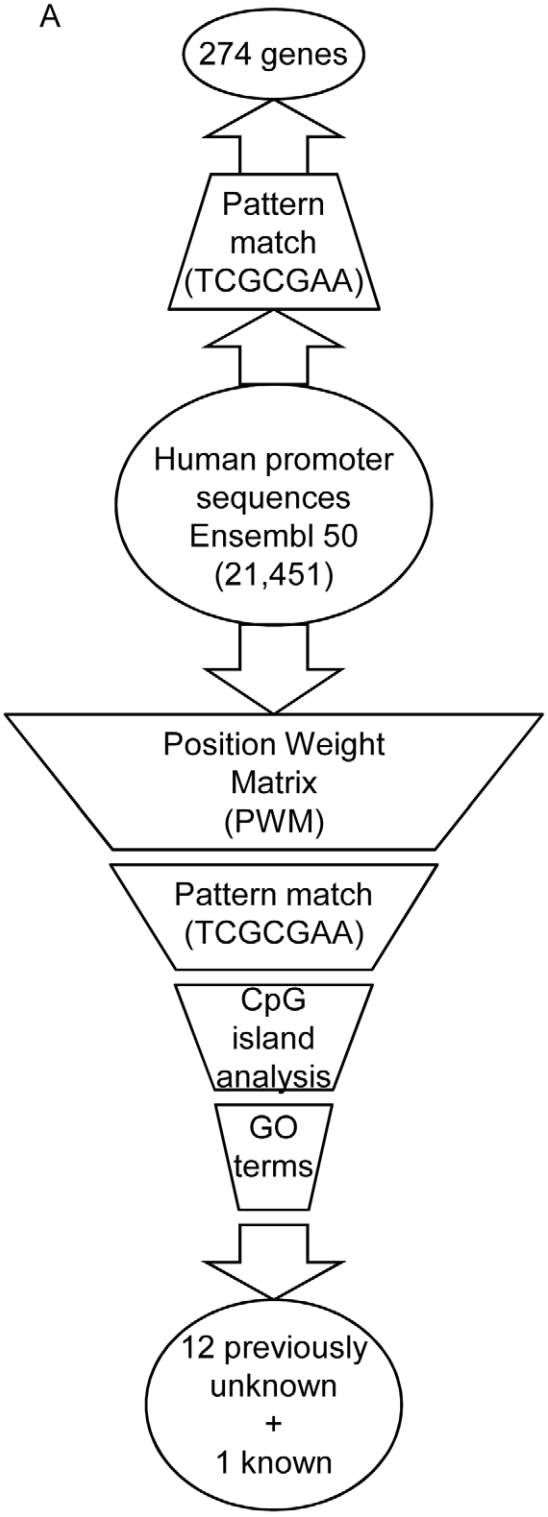
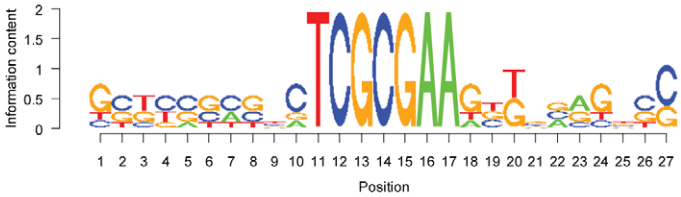


Figure 1. Interaction of Zta with methylated RpZRE3 core element from 5 human promoters. (a) The nucleotide sequence of one strand of each oligonucleotide spanning the indicated ZREs is shown, with the conserved RpZRE3 core element (TCGCGAA) aligned. Double strand versions of these sequences were used as probes in EMSA, with the 2 cytosines within the CpG core motif methylated. The RpZRE3 from EBV BRLF1 promoter is included. (b) DNA binding between *in vitro* translated Zta with each probe was undertaken by EMSA. The ability of Zta (Z) to interact with the probe was compared with an unprogrammed translation lysate (C) as a negative control. The complex was separated on an 8% gel by electrophoresis and visualised by phosphoimaging. The excess probe can be seen at the bottom of the gel. The probe used in each experiment is indicated below the gel. (c) Competition of each ZRE sequence (at 100X excess) against labelled RpZRE3 was determined by competition EMSAs. Data from at least 2 experiments was taken to calculate competition i.e. the percentage of labelled probe displaced by unlabelled competitor. Methylated ZREs were used for both probe and competition. AP1 mut, an oligonucleotide previously shown not to interact with Zta [36], was used as a negative control to define the level of non-specific binding. Error bars indicate standard error.
doi:10.1371/journal.pone.0009443.g001

A



B



C

Overrepresented GO terms
Chromatin binding
Chromatin remodelling
Mitosis
Mitotic cell cycle checkpoint
Oxidation reduction
Satellite DNA binding
Transcription coactivator activity
Transcription corepressor activity
Transcription regulator activity

D

ENSG00000162909	CAPN2
ENSG00000129596	CDO1
ENSG00000171634	FALZ
ENSG00000054523	KIF1B
ENSG00000131899	LLGL1
ENSG00000143013	LMO4
ENSG00000129071	MBD4
ENSG00000104886	PLEKHJ1
ENSG00000184304	PRKD1
ENSG00000129657	SEC14L1
ENSG00000171148	TADA3L
ENSG00000077097	TOP2B
ENSG00000070444	MNT

Figure 2. Identification of human promoters containing RpZRE3 core elements in the human genome. (a) The bioinformatics analysis undertaken for the genome wide scan is represented as a flow diagram. Genetic input or output is represented by ovals, and filters by trapeziums. (b) A Position Weight Matrix (PWM) was created using the aligned sequences of the RpZRE3-promoter-6 data-set. These sequences are shown in Fig. 1(a). This was used to filter the genomic data from the human promoter sequences (−500 to +1). (c) Overrepresented GO terms found for the 274 genes were identified. (d) The 13 genes identified as candidate genes, 12 previously unknown genes and 1 from the initial screen, and their associated Ensembl accession codes.

doi:10.1371/journal.pone.0009443.g002

Figure 3. Interaction of Zta with methylated RpZRE3s found in human promoters. EMSA analysis with *in vitro* translated Zta protein (Z) or an unprogrammed translation lysate (C) with the probes indicated to the right of each gel was carried out as described in Fig. 1.
doi:10.1371/journal.pone.0009443.g003

presence of AP1, c-fos and c-jun binding sites within the 5′ flank of XPC (5′TGCGTCA) (Figure 7). c-fos and c-jun proteins are both members of the bZIP transcription factor family; they form AP1 DNA binding activity as either homodimers or heterodimers. It is relevant that fos/jun dimers share some DNA recognition motifs with Zta [15–17,31,46–49]. This led to the hypothesis that the predicted AP1 site in the 5′ flanking sequence is an additional ZRE that is responsible for binding to the non-methylated XPC oligonucleotide. To test the hypothesis, further mutations were introduced into the 5′ XPC flanking sequence in the predicted AP1 site (5′TCCGCCT). The ability of Zta to interact with the double AP1/core mutant site (XPCLM3Mcore) was compared with the core-only mutant. Dual mutation of the predicted AP1 site and the core abrogated the ability of Zta to interact with the XPC oligonucleotide, demonstrating that binding of Zta to the non-methylated XPC site resided with this AP1 site (Figure 7).

This analysis showed that the RpZRE3 core element is not recognised in its non-methylated state in the context of any of the viral or cellular promoters in the RpZRE3-promoter-18 data-set, and that this ZRE is specifically recognised when methylated.

Discussion

A computational search revealed a set of 274 human genes with cellular promoters containing the 7-mer RpZRE3 core element. Flanking sequence can influence the interaction of some transcription factors with DNA, for example, the interaction of the E2F family with DNA is influenced by a region of at least 8 nucleotides 5′ and 11 nucleotides 3′ to the central nucleotide [50]. Before assigning all 274 genes as candidates for regulation by Zta, it was important to assess whether the cognate flanking sequence had a profound effect on binding by Zta. Consideration of the sequences represented in the flanking sequence of this set of genes revealed it to be diverse; all four nucleotides were represented in 55% of positions and three of the four nucleotides were represented at the remaining positions. Indeed, the immediate flank of the RpZRE3 core element, consisting of four nucleotides both 5′ and 3′, had 100% representation of all four nucleotides. It can therefore be concluded that the flanking sequence does not prevent binding to the methylated RpZRE3 core, and a PWM generated from the RpZRE3-promoter-18 data-set showed little conservation outside the core motif. In contrast, the interaction with the non-methylated core element initially appeared to be influenced by flanking sequence in only 1 of the 18 promoters. However, further analysis revealed this to be due to the presence of a second ZRE in the flanking sequence.

These results show that the approach of undertaking a computational pattern match search for the 7-mer RpZRE3 core element is a fast and reliable method to identify promoters containing ZREs that are recognized by Zta only when they are methylated (class III ZREs).

The identification of two adjacent juxtaposed ZREs in the XPC promoter presents an interesting problem; can both sites be occupied at once? The natural juxtaposition of ZREs has been previously described for a viral promoter; *BZLF1* and a cellular promoter *EGR1*. In the *BZLF1* promoter, the elements are situated with 12 base pairs between the central nucleotides; both can be occupied simultaneously and both are functionally relevant [17,19–20]. For the *EGR1* promoter, the elements are immediately adjacent with just 8 nucleotides between the central nucleotides [25,26]. There is no evidence for simultaneous occupation by Zta and only one element appears to be functional *in vivo* [26]. The arrangement of the *XPC* promoter places the elements 8 nucleotides apart, identical to *EGR1*, and there is no evidence from the DNA-binding experiments for simultaneous occupation of the sites. As one ZRE is recognised in a methylation-dependent manner and the other is recognised when non-methylated, it is possible that the arrangement may provide a fail-safe mechanism to ensure that the gene is regulated in both its methylated and non-methylated states.

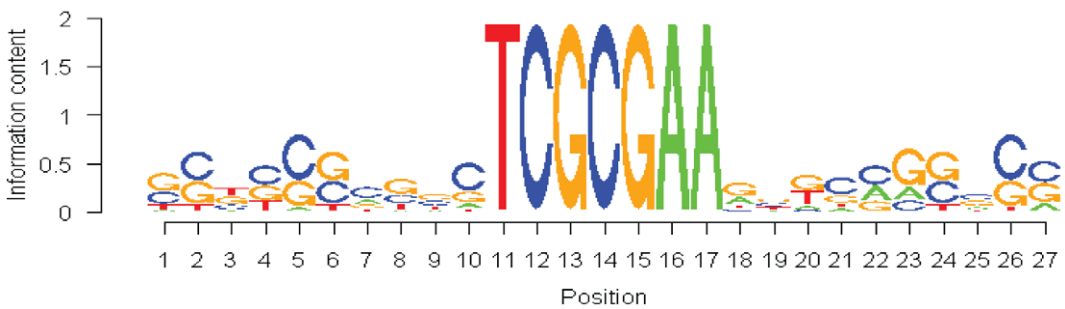


Figure 4. Analysis of the RpZRE3 core element flanking sequence. (a) A Position Weight Matrix (PWM), created using the probe sequences of the dataset RpZRE3-promoter-18.
doi:10.1371/journal.pone.0009443.g004

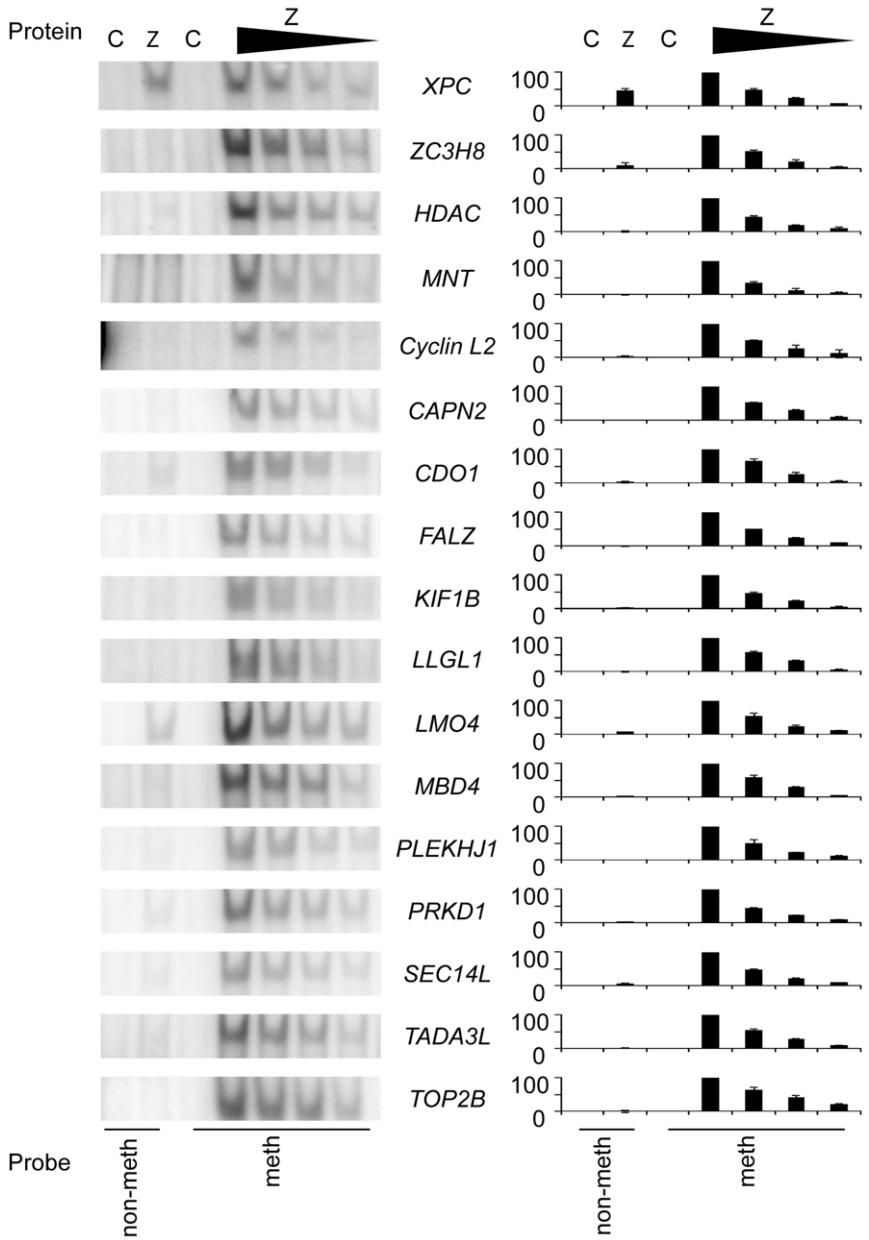


Figure 5. Comparison of Zta interaction with non-methylated and methylated RpZRE3s from human promoters. DNA binding analysis of *in vitro* translated Zta protein (Z) or an unprogrammed translation lysate (C) with both non-methylated and methylated versions of the probes indicated to the right of each gel was carried out by EMSA as described in Fig. 1. A titration of Zta protein (1:2, 1:4, 1:8) was used to compare non-methylated binding to that of the methylated equivalent. Quantitation of the complexes formed between Zta and the non-methylated and methylated probes, from at least 2 experiments, is represented as a histogram to the right of the corresponding gel. Complex formation is shown relative to the maximum binding for each probe. Error bars indicate standard error.
doi:10.1371/journal.pone.0009443.g005

The simple computational approach taken to identify the location of this methylation-dependent and novel DNA-binding motif in human gene promoters identified 274 genes potentially regulated by overcoming epigenetic silencing during the viral replicative cycle. Given the known functions of Zta in reprogramming viral gene expression, disrupting cell cycle control and replicating viral DNA, it is interesting to note that the Gene Ontology terms that are over-represented in the RpZRE3-promoter-274 gene-set are largely involved in transcription, chromatin re-modelling and mitosis. This strongly suggests that Zta may activate this set of cellular genes in order to accomplish these functions. Testing whether these genes are activated during latency disruption in memory B-lymphocytes *in vivo* is technically challenging given both the scarcity of memory B-lymphocytes in peripheral circulation and the infrequency of latency disruption *in vivo*.

The co-location of the RpZRE3 core element in 274 human promoters is unlikely to have been driven by an evolutionary advantage to the virus, but may reflect the involvement of a cellular transcription factor interacting with the same element. This would suggest that this set of genes share a common mode of regulation during human development or differentiation. Further

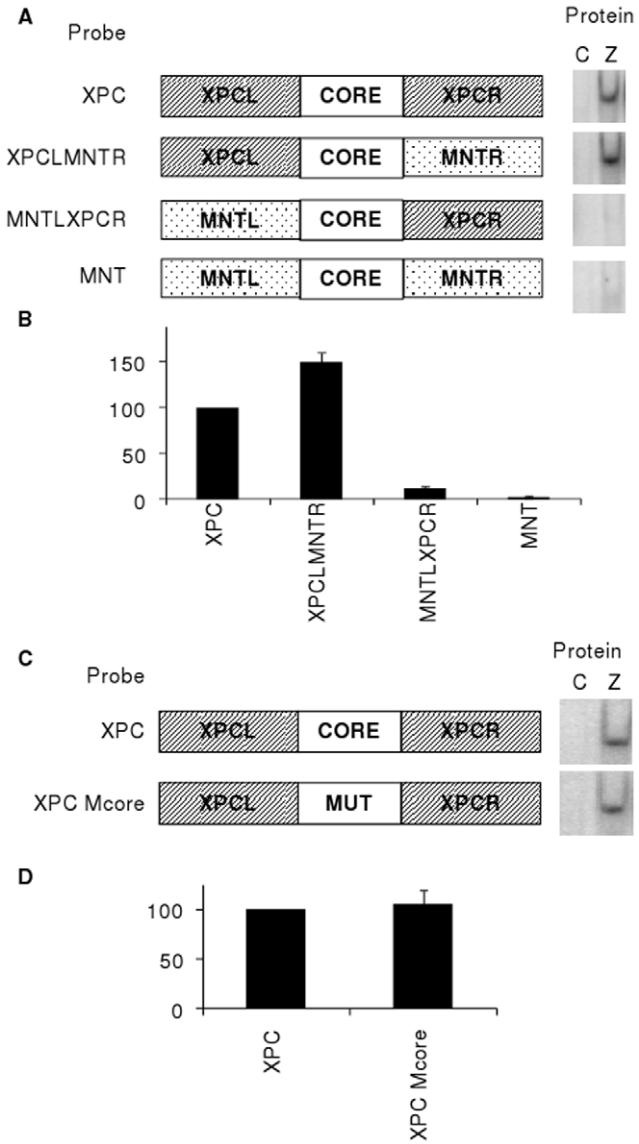


Figure 6. The effect of flanking region upon Zta binding to non-methylated XPC probe. (a) Schematic representation of the probes, illustrating the XPC and MNT probes, and the flank swap probes XPCLMNTR and MNTLXPCR, adjacent to the corresponding EMSA analysis, carried out in the same manner as described in Fig. 1. (b) Quantification of the complexes formed with each probe, from 2 experiments was undertaken. Complex formation is represented as a percentage of the Zta complex with the non-methylated XPC probe. Error bars indicate standard error. (c) Schematic representation of probes, indicating mutated core sequence. EMSA analysis was carried out as described in Fig. 1. (d) Quantitation of the complexes formed with each probe, from 2 experiments was undertaken. Complex formation is represented as a percentage of the Zta complex formed with the non-methylated XPC probe. Error bars indicate standard error.
doi:10.1371/journal.pone.0009443.g006

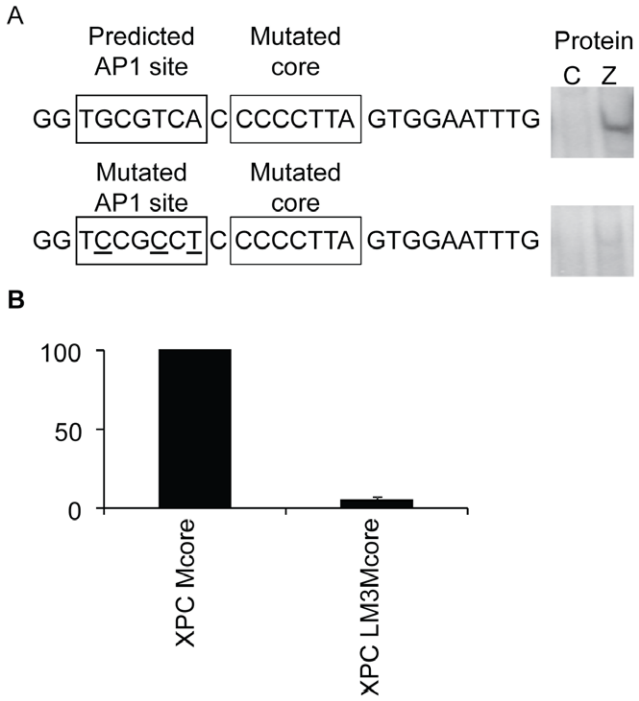


Figure 7. Uncovering an additional ZRE in the 5′ flank of the XPC probe. (a) PROMO transcription factor binding site prediction software identified an AP1 site which is highlighted in the 5′ flank sequence of XPC. The specific nucleotide mutations are underlined. The complexes formed by EMSA are shown to the right of the corresponding probe sequence. (b) Quantitation of the complexes formed with each probe, from 2 experiments, was undertaken. Complex formation is represented as a percentage of the Zta complex formed with the XPC mutated core probe. Error bars indicate standard error.
doi:10.1371/journal.pone.0009443.g007

more, it will be interesting to question whether co-occurrence of the RpZRE3-core element with other transcription factor binding sites forming a co-operative *cis*-regulator module which may illuminate the regulation of these host and viral genes.

Supporting Information

Table S1 RpZRE3-promoter-274 data-set: A table of genes which contain a RpZRE3 core element in the 500 bp promoter region.
Found at: doi:10.1371/journal.pone.0009443.s001 (0.05 MB XLS)

Acknowledgments

We thank Queensta Miller and James Heather for discussion.

Author Contributions

Conceived and designed the experiments: KF EH MJN SJ AJS. Performed the experiments: KF EH. Analyzed the data: KF EH. Contributed reagents/materials/analysis tools: KF EH. Wrote the paper: KF EH MJN SJ AJS.

References

1. Murray PG, Young LS (2001) Epstein-Barr virus infection: basis of malignancy and potential for therapy. Expert Rev Mol Med. pp 1–20.
2. Talbot SJ, Crawford DH (2004) Viruses and tumours–an update. Eur J Cancer 40: 1998–2005.
3. Young LS, Rickinson AB (2004) Epstein-Barr virus: 40 years on. Nat Rev Cancer 4: 757–768.
4. Thorley-Lawson DA, Allday MJ (2008) The curious case of the tumour virus: 50 years of Burkitt's lymphoma. Nat Rev Microbiol 6: 913–924.

5. Babcock GJ, Decker LL, Volk M, Thorley-Lawson DA (1998) EBV persistence in memory B cells in vivo. Immunity 9: 395–404.
6. Niller HH, Wolf H, Minarovits J (2009) Epigenetic dysregulation of the host cell genome in Epstein-Barr virus-associated neoplasia. Semin Cancer Biol 19: 158–164.
7. Minarovits J (2006) Epigenotypes of latent herpesvirus genomes. Curr Top Microbiol Immunol 310: 61–80.
8. Ernberg I, Falk K, Minarovits J, Busson P, Tursz T, et al. (1989) The role of methylation in the phenotype-dependent modulation of Epstein-Barr nuclear antigen 2 and latent membrane protein genes in cells latently infected with Epstein-Barr virus. J Gen Virol 70 (Pt 11): 2989–3002.
9. Masucci MG, Contreras-Salazar B, Ragnar E, Falk K, Minarovits J, et al. (1989) 5-Azacytidine up regulates the expression of Epstein-Barr virus nuclear antigen 2 (EBNA-2) through EBNA-6 and latent membrane protein in the Burkitt's lymphoma line rael. J Virol 63: 3135–3141.
10. Tao Q, Robertson KD (2003) Stealth technology: how Epstein-Barr virus utilizes DNA methylation to cloak itself from immune detection. Clin Immunol 109: 53–63.
11. Israel BF, Kenney SC (2003) Virally targeted therapies for EBV-associated malignancies. Oncogene 22: 5122–5130.
12. Feng WH, Hong G, Delecluse HJ, Kenney SC (2004) Lytic induction therapy for Epstein-Barr virus-positive B-cell lymphomas. J Virol 78: 1893–1902.
13. Miller G (1989) The switch between EBV latency and replication. Yale J Biol Med 62: 205–213.
14. Speck SH, Chatila T, Flemington E (1997) Reactivation of Epstein-Barr virus: regulation and function of the BZLF1 gene. Trends in Microbiology 5: 399–405.
15. Sinclair AJ (2003) bZIP proteins of human Gamma Herpesviruses. Journal of General Virology 84: 1941–1949.
16. Farrell P, Rowe D, Rooney CM, Kouzarides T (1989) Epstein-Barr virus BZLF1 trans-activator specifically binds to a consensus AP-1 site and is related to c-fos. EMBO J 8: 127–132.
17. Urier G, M. B, Chambard P, Sergeant A (1989) The Epstein-Barr virus early protein EB1 activates trasncription from different responsive elements including AP-1 binding sites. EMBO J 8: 1447–1453.
18. Sinclair AJ, Brimmell M, Shanahan F, Farrell PJ (1991) Pathways of activation of the Epstein-Barr virus productive cycle. J Virol 65: 2237–2244.
19. Lieberman PM, Hardwick JM, Sample J, Hayward GS, Hayward SD (1990) The zta transactivator involved in induction of lytic cycle gene expression in Epstein-Barr virus-infected lymphocytes binds to both AP-1 and ZRE sites in target promoter and enhancer regions. J Virol 64: 1143–1155.
20. Flemington E, Speck SH (1990) Autoregulation of Epstein-Barr Virus putative lytic switch gene BZLF1. Journal of Virology 64: 1227–1232.
21. Zetterberg H, Jansson A, Rymo L, Chen F, Karlsson A, et al. (2002) The Epstein-Barr virus ZEBRA protein activates transcription from the early lytic F promoter by binding to a promoter- proximal AP-1-like site. Journal of General Virology 83: 2007–2014.
22. Dickerson SJ, Xing Y, Robinson AR, Seaman WT, Gruffat H, et al. (2009) Methylation-dependent binding of the epstein-barr virus BZLF1 protein to viral promoters. PLoS Pathog 5: e1000356.
23. Quinlivan EB, Holley-Guthrie EA, Norris M, Gutsch D, Bachenheimer SL, et al. (1993) Direct BRLF1 binding is required for cooperative BZLF1/BRLF1 activation of the Epstein-Barr virus early promoter, BMRF1. Nucleic Acids Res 21: 1999–2007.
24. Jones RJ, Dickerson S, Bhende PM, Delecluse HJ, Kenney SC (2007) Epstein-Barr virus lytic infection induces retinoic acid-responsive genes through induction of a retinol-metabolizing enzyme, DHRS9. J Biol Chem 282: 8317–8324.
25. Chang Y, Lee HH, Chen YT, Lu J, Wu SY, et al. (2006) Induction of the early growth response 1 gene by Epstein-Barr virus lytic transactivator Zta. J Virol 80: 7748–7755.
26. Heather J, Flower K, Isaac S, Sinclair AJ (2009) The Epstein-Barr virus lytic cycle activator Zta interacts with methylated ZRE in the promoter of host target gene egr1. J Gen Virol 90: 1450–1454.
27. Li D, Qian L, Chen C, Shi M, Yu M, et al. (2009) Down-regulation of MHC class II expression through inhibition of CIITA transcription by lytic transactivator Zta during Epstein-Barr virus reactivation. J Immunol 182: 1799–1809.
28. Hsu M, Wu SY, Chang SS, Su IJ, Tsai CH, et al. (2008) Epstein-Barr virus lytic transactivator Zta enhances chemotactic activity through induction of interleukin-8 in nasopharyngeal carcinoma cells. J Virol 82: 3679–3688.
29. Mahot S, Sergeant A, Drouet E, Gruffat H (2003) A novel function for the Epstein-Barr virus transcription factor EB1/Zta: induction of transcription of the hIL-10 gene. J Gen Virol 84: 965–974.
30. Tsai SC, Lin SJ, Chen PW, Luo WY, Yeh TH, et al. (2009) EBV Zta protein induces the expression of interleukin-13, promoting the proliferation of EBV-infected B cells and lymphoblastoid cell lines. Blood 114: 109–118.
31. Kouzarides T, Packham G, Cook A, Farrell PJ (1991) The BZLF1 protein of EBV has a coiled coil dimerization domain without a heptad leucine repeat but with homology to the C/EBP leucine zipper. Oncogene 6: 195–204.
32. Bhende PM, Seaman WT, Delecluse HJ, Kenney SC (2005) BZLF1 activation of the methylated form of the BRLF1 immediate-early promoter is regulated by BZLF1 residue 186. J Virol 79: 7338–7348.
33. Bhende PM, Seaman WT, Delecluse HJ, Kenney SC (2004) The EBV lytic switch protein, Z, preferentially binds to and activates the methylated viral genome. Nat Genet 36: 1099–1104.
34. Karlsson QH, Schelcher C, Verrall E, Petosa C, Sinclair AJ (2008) Methylated DNA recognition during the reversal of epigenetic silencing is regulated by cysteine and serine residues in the Epstein-Barr virus lytic switch protein. PLoS Pathog 4: e1000005.
35. Karlsson QH, Schelcher C, Verrall E, Petosa C, Sinclair AJ (2008) The reversal of epigenetic silencing of the EBV genome is regulated by viral bZIP protein. Biochem Soc Trans 36: 637–639.
36. Schelcher C, Valencia S, Delecluse HJ, Hicks M, Sinclair AJ (2005) Mutation of a single amino acid residue in the basic region of the Epstein-Barr virus (EBV) lytic cycle switch protein Zta (BZLF1) prevents reactivation of EBV from latency. J Virol 79: 13822–13828.
37. Wang P, Day L, Dheekollu J, Lieberman PM (2005) A redox-sensitive cysteine in Zta is required for Epstein-Barr virus lytic cycle DNA replication. J Virol 79: 13298–13309.
38. Hubbard TJ, Aken BL, Beal K, Ballester B, Caccamo M, et al. (2007) Ensembl 2007. Nucleic Acids Res 35: D610–617.
39. Haider S, Ballester B, Smedley D, Zhang J, Rice P, et al. (2009) BioMart Central Portal–unified access to biological data. Nucleic Acids Res 37: W23–27.
40. Flicek P, Aken BL, Beal K, Ballester B, Caccamo M, et al. (2008) Ensembl 2008. Nucleic Acids Res 36: D707–714.
41. Lenhard B, Wasserman WW (2002) TFBS: Computational framework for transcription factor binding site analysis. Bioinformatics 18: 1135–1136.
42. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25: 25–29.
43. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16: 276–277.
44. Farre D, Roset R, Huerta M, Adsuara JE, Rosello L, et al. (2003) Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. Nucleic Acids Res 31: 3651–3653.
45. Messeguer X, Escudero R, Farre D, Nunez O, Martinez J, et al. (2002) PROMO: detection of known transcription regulatory elements using species-tailored searches. Bioinformatics 18: 333–334.
46. Taylor N, Flemington E, Kolman JL, Baumann RP, Speck SH, et al. (1991) ZEBRA and a Fos-GCN4 chimeric protein differ in their DNA-binding specificities for sites in the Epstein-Barr virus BZLF1 promoter. J Virol 65: 4033–4041.
47. Lieberman PM, Berk AJ (1990) In vitro transcriptional activation, dimerization, and DNA-binding specificity of the Epstein-Barr virus Zta protein. J Virol 64: 2560–2568.
48. Chang YN, Dong DLY, Hayward GS, Hayward SD (1990) The Epstein-Barr-Virus Zta transactivator - a member of the bZip family with unique DNA-binding specificity and a dimerization domain that lacks the characteristic heptad leucine zipper motif. Journal of Virology 64: 3358–3369.
49. Sinclair AJ (2006) Unexpected structure of Epstein-Barr virus lytic cycle activator Zta. Trends Microbiol 14: 289–291.
50. Tao Y, Kassatly RF, Cress WD, Horowitz JM (1997) Subunit composition determines E2F DNA-binding site specificity. Mol Cell Biol 17: 6994–7007.

OPEN ACCESS Freely available online

PLoS one

Epigenetic Control of Viral Life-Cycle by a DNA-Methylation Dependent Transcription Factor

Kirsty Flower[1,2], David Thomas[1], James Heather[1,3], Sharada Ramasubramanyan[1], Susan Jones[1,4], Alison J. Sinclair[1]*

1 School of Life Sciences, University of Sussex, Brighton, United Kingdom, **2** Epigenetics Unit, Department of Surgery and Cancer, Imperial College, London, United Kingdom, **3** Infection and Immunity Division, University College London, London, United Kingdom, **4** The James Hutton Institute, Dundee, United Kingdom

Abstract

Epstein-Barr virus (EBV) encoded transcription factor Zta (BZLF1, ZEBRA, EB1) is the prototype of a class of transcription factor (including C/EBPalpha) that interact with CpG-containing DNA response elements in a methylation-dependent manner. The EBV genome undergoes a biphasic methylation cycle; it is extensively methylated during viral latency but is reset to an unmethylated state following viral lytic replication. Zta is expressed transiently following infection and again during the switch between latency and lytic replication. The requirement for CpG-methylation at critical Zta response elements (ZREs) has been proposed to regulate EBV replication, specifically it could aid the activation of viral lytic gene expression from silenced promoters on the methylated genome during latency in addition to preventing full lytic reactivation from the non-methylated EBV genome immediately following infection. We developed a computational approach to predict the location of ZREs which we experimentally assessed using *in vitro* and *in vivo* DNA association assays. A remarkably different binding motif is apparent for the CpG and non-CpG ZREs. Computational prediction of the location of these binding motifs in EBV revealed that the majority of lytic cycle genes have at least one and many have multiple copies of methylation-dependent CpG ZREs within their promoters. This suggests that the abundance of Zta protein coupled with the methylation status of the EBV genome act together to co-ordinate the expression of lytic cycle genes at the majority of EBV promoters.

Citation: Flower K, Thomas D, Heather J, Ramasubramanyan S, Jones S, et al. (2011) Epigenetic Control of Viral Life-Cycle by a DNA-Methylation Dependent Transcription Factor. PLoS ONE 6(10): e25922. doi:10.1371/journal.pone.0025922

Editor: Fatah Kashanchi, George Mason University, United States of America

Received June 27, 2011; **Accepted** September 13, 2011; **Published** October 11, 2011

Funding: This research was supported by grants from the Wellcome Trust, United Kingdom, and University of Sussex. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: a.j.sinclair@sussex.ac.uk

Introduction

Infection of human B-lymphocytes by Epstein-Barr virus results in the establishment of a latent state in which a highly restricted set of viral genes are expressed [1]. This is accompanied by extensive methylation of CpG motifs in non-expressed viral genes [2,3,4]. In response to physiological stimuli, such as engagement of the B-cell receptor, epigenetic silencing of the viral genome is overturned, resulting in widespread activation of viral gene expression and lytic replication [4,5]. The expression of a subset of host genes is also altered during this period [6,7,8,9,10,11,12,13,14].

The switch between latency and the lytic cycle is orchestrated by the viral gene *BZLF1*, which encodes the protein Zta (also known as ZEBRA, BZLF1, EB1, or Z) [15,16,17]. Zta resembles the AP1 family of bZIP transcription factors but has a unique dimerisation domain and does not form heterodimers with cellular bZIP proteins [18]. Three classes of Zta DNA binding sites (Zta response elements (ZREs)) have been defined for Zta [19]. Class I ZREs include classical AP1-like recognition elements. However, some Zta binding sites contain a CpG motif and Zta has the unusual property of binding preferentially to these ZREs when they are methylated [20,21], defining class II ZREs [19]. Remarkably, some CpG-containing ZREs are only recognized in their methylated form (class III ZREs) [19,20,21,22,23,24,25]. Methylation of the viral genome occurs during latency and has recently been shown to be required for EBV replication [3]. The ability of Zta to bind to methylated ZREs suggests that Zta may have a direct role in overriding the epigenetic silencing of the viral genome to activate expression of viral genes required for lytic replication.

The requirement for methylation at critical ZREs may also contribute to the establishment of latency during the immortalization of infected cells. The EBV genome is not methylated when it enters cells but the genome gradually becomes methylated during immortalization and the establishment of viral latency [3,4]. Zta is transiently expressed during the early period immediately after infection and is required for efficient immortalization [3]. It is therefore essential that Zta should not activate the full lytic replication cycle at this stage. A plausible hypothesis to explain this is that expression of key lytic cycle genes are controlled by class III ZREs that do not function in their unmethylated form.

We developed a computational approach to identify candidate ZREs and applied it to a genome-wide analysis of the EBV genome that revealed many novel target loci. The implications of these data for the ability of EBV to evade epigenetic silencing of the host viral genome is discussed.

Results

Prediction of ZREs core sequences bound by Zta using PROMO

In order to predict novel ZRE core sequences, we started by searching three well-characterized Zta-responsive promoters from the EBV genome (*BZLF1* promoter (Zp) [26,27,28,29], *BRLF1* promoter (Rp) [27,29,30] and *BMRF1* promoter [31] using the PROMO algorithm [33,34] and the position frequency matrix (PFM) for Tranfac 8.3 Zta transcription factor entry T00923 [32,33]. These 3 promoters are known to contain eight previously verified sites: in Zp (ZREIIIA and ZREIIIB); in Rp (ZRE1, ZRE2 and ZRE3) and in BMRF1 promoter (AP1, ZRE(−44) and ZRE (−107)) (Figure 1 and Table S1) however, the PROMO algorithm only predicted one of these sites (RpZRE1) using the PFM T00923. In addition, 6 novel sites where predicted (Table S2).The ability of Zta to interact with each predicted site was assessed using electrophoretic mobility shift assays (EMSA) (Figure 1, Table S2), although three novel sites were identified, eight known sites were missed and three false positives were predicted indicating that the PFM used had a low sensitivity.

Application of a novel ZRE PFM to predict CpG containing ZREs

A new PFM was generated using the core sequences of five CpG-containing ZREs (denoted PFM_{CpG5}) from the promoters described above and the *BRRF1* promoter [22] (Figure 2). The accuracy of the PFM was evaluated by searching for ZREs in the well-characterized viral promoters (Rp, Zp and *BMRF1*p). PFM_{CpG5} identified all 5 verified CpG containing sites and predicted two novel sites; one located in Rp, centered on −114, and one located in the *BMRF1* promoter, centered on −148. DNA binding experiments demonstrate that Zta interacts with both sites in a methylation-dependent manner, characteristic of class III ZREs (Figure 3), thus the new PFM ($ZRECpG_5$) has a high level of sensitivity. The PFM_{CpG5} was then used to predict core ZREs in the complete EBV genome. Within the EBV genome a total of 16 novel sequence variants of CpG ZREs were predicted (A–P) (Figure 3). EMSAs were undertaken with each of the novel ZRE core sequences (both non-methylated and methylated) to evaluate Zta binding (Figure 3). All but two of the predicted ZRE sequences bound in the methylated form. Only one sequence bound significantly in the unmethylated form. Therefore 13 out of 16 predictions are classified as Class III ZREs, 1 is classified as Class II and 2 did not interact with Zta significantly. Combined with previously published ZREs, this resulted in a total set of 32 distinct sequence variants of ZREs (ZRE_{32}) (Table 1).

Identification of ZRE core binding sequences in the EBV genome

Global analysis of the EBV genome was then undertaken using an exact pattern match with the 32 validated variants of the ZRE core sequence (Figure 4). This revealed 469 locations within the EBV genome that matched one of the ZRE core sequences (Table S3 and http://bioinf.biochem.sussex.ac.uk/EBV).

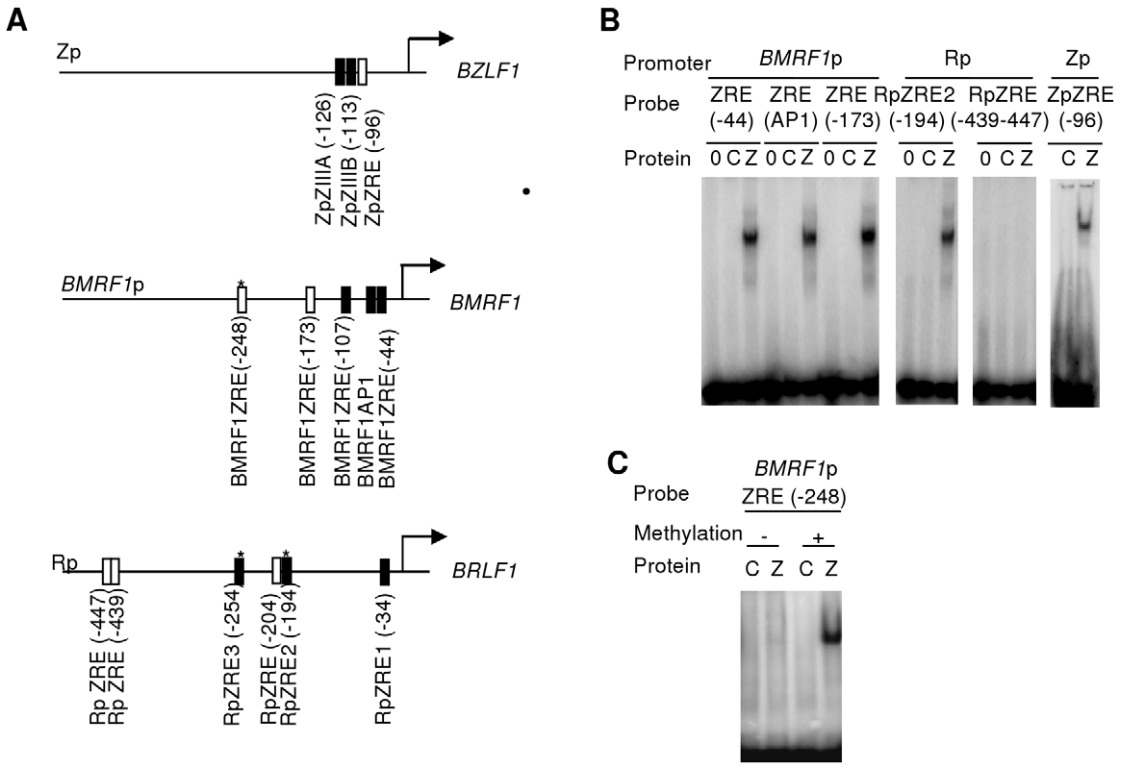


Figure 1. Evaluation of predicted ZREs in three EBV promoters. A. Summary of the known (filled box) and predicted (open box) core ZRE sequences within the proximal 500 nucleotides of the indicated *BZLF1*, *BRLF1 and BMRF1* promoters. The arrows represent the transcription start sites. Stars represent CpG ZREs. B. Double strand oligonucleotides were generated with the core ZRE sequence and at least 10 nucleotides of cognate sequence on either side. Following radio labeling, these were incubated with *in vitro* translated Zta and subject to EMSA. The reactions contained no protein, 0, control lysate, C or Zta, Z. The DNA probes are indicated above with their originating promoters. C. Double strand oligonucleotides were generated with the core ZRE sequence and 10 nucleotides of cognate sequence on either side. Following radio labeling, these were subject to *in vitro* methylation with SssI methyl transferase (+), or a mock reaction (−). Subsequently, they were incubated with *in vitro* translated Zta and subject to EMSA. The reactions contained control lysate, C; or Zta, Z.
doi:10.1371/journal.pone.0025922.g001

A

ZRE core Sequence	Name
TGAGCGA	RpZRE2
TTCGCGA	RpZRE3
TGAGCGT	NapZRE2
CGGGCGA	NapZRE1
TGTGCGA	BMRF1p (-248)

B

	-3'	-2'	-1'	0	1'	2'	3'
A	0	0	2	0	0	0	4
C	1	0	1	0	5	0	0
G	0	4	1	5	0	5	0
T	4	1	1	0	0	0	1

C

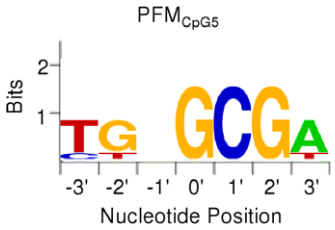


Figure 2. Position Frequency Matrix (PFM) of CpG containing ZREs (PFM$_{CpG5}$). A. The core seven nucleotide element of the CpG containing ZREs found within the *BRLF1, BRRF1, and BMRF1* (Rp, Nap and BMRF1p) promoters are shown. B. The number of incidences of each nucleotide at each position and the thus the frequency were calculated. C. The consensus sequence is represented as a sequence LOGO.
doi:10.1371/journal.pone.0025922.g002

The occurrence of ZREs throughout the EBV genome appears to be widespread; with 81 out of 86 (94%) EBV promoters containing at least 1 ZRE core sequence (Table S4). This suggests that Zta has the potential to regulate the expression of the majority of EBV genes. Furthermore, 58 EBV promoters contained at least 1 CpG containing ZRE. These regions are methylated during latency [3,4], suggesting that methylation-dependent Zta interaction with ZREs could influence the expression of a broad range of EBV genes once Zta is synthesized at the onset of lytic cycle.

Of particular relevance to the control of EBV gene expression immediately after infection are 22 EBV genes that contained CpG ZREs but have no methylation independent ZREs in their promoters (Table 2). These genes are prime contenders to be regulated in a strictly methylation-dependent manner by Zta. These were originally classified as displaying early lytic, late lytic and latent patterns of gene expression [34], but importantly, genome wide expression studies revealed that all are up regulated during lytic cycle in BL cells, with the majority reaching peak levels approximately 24 hours after lytic activation [6].

Three of these promoters were chosen to question whether Zta interacts with the novel CpG ZREs *in vivo*; *BKRF4*, *BGLF4* and *BTRF1*. The location of the CpG ZREs in each promoter is indicated in Figure 5. Lytic cycle was activated in Akata cells [35] by surface immunoglobulin ligation, undertaken in the presence of acyclovir to inhibit genome replication. Chromatin was subjected to Zta immunoprecipitation (ChIP) and the interaction of Zta with these promoters was assessed by Q-PCR. The ability of this antibody to precipitate Zta bound to chromatin is demonstrated by western blotting in Figure 6. In addition, we show that Zta as opposed to a control antibody specifically precipitates chromatin from a region of oriLyt containing multiple ZREs. Using primer sets proximal to the CpG ZREs from *BKRF4*, *BGLF4* and *BTRF1* compared to primer sets from three regions of the EBV genome devoid of ZREs, we reveal that Zta specifically binds to all three of these promoters that contain novel CpG ZREs *in vivo* (Figure 7).

Discussion

Following several iterations of a predictive and evaluative approach, we identified a set of 32 distinct sequence variants in the core 7-nucleotide sequence to which Zta can bind. This includes 20 variants containing a CpG motif, the majority of which (90%) are only recognized by Zta when they are methylated.

The consensus binding sites identified for non-CpG ZREs are similar to the binding sites originally described for Zta (Figure 8). In contrast, the binding sites for CpG containing ZREs are remarkably different. This sequence is dominated by an almost invariant G 5′ to the absolute prerequisite for me-CpG at positions 1′ and 2′ in the right-half of the core sequence.

The identification of 58 EBV promoters that harbor methylation dependent CpG ZREs, combined with the knowledge that the EBV genome is heavily methylated during latency [3,4], suggests that Zta plays an important role in overturning epigenetic silencing of over half of the EBV genes during lytic replication. Indeed, all three of the promoters tested displayed a strong interaction with Zta *in vivo* in ChIP analyses. A genome-wide DNA binding analysis was recently published identifying sequences to which a mutant form of Zta, that is replication and transactivation dead, can interact [3]. This report highlighted the strength of the interaction between Zta and methylation dependent binding of Zta to CpG ZREs in the EBV genome.

The EBV genes that contain only methylation-dependent ZREs are of particular interest. All of these genes are heavily methylated during viral latency yet unmethylated following replication and immediately after infection [3,4]. Several are required for EBV replication and include components of the helicase/primase complex (*BBLF4*, *BBLF2/BBLF3*), the viral protein kinase (*BGLF4*), and glycoproteins gL (*BKRF2*) and gB (*BALF4*). In addition, the promoters for *BBLF4* and *BBLF2/BBLF3* have been validated as being targets for Zta that are completely dependent on methylation for Zta activation [3]. Our discovery that one in five EBV promoters contain CpG ZREs but have no methylation independent ZREs strongly supports the hypothesis that the unmethylated status of the EBV genome guards against the expression of the full range of lytic genes and therefore lytic replication during the establishment of latency.

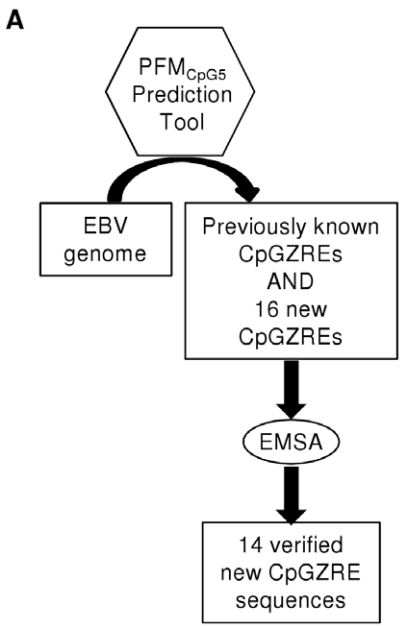
Zta is the prototypic member of a family of transcription factors that interact with DNA in a methylation-dependent manner. C/EBP alpha has recently been shown to share the same characteristics [36]. It has been suggested that the interaction between C/EBP alpha and methylated sequence elements are needed to activate tissue specific genes during differentiation [36].

The biphasic methylation cycle is observed for several different classes of viruses that establish latency [4]. Yet even KSHV, which is closely related to EBV, does not contain a functional Zta homologue. The question arises as to how the methylated genomes of these viruses can be reactivated. We suggest that the recent discovery that a cellular transcription factor also has methylation dependent DNA binding properties [36] implies that other viruses may rely on host methylation dependent transcription factors to differentially control the expression of their genomes during the establishment of latency or replication.

Methods

Computational prediction of ZREs core sequences bound by Zta

The starting point for the computational approach was the Zta transcription factor entry T00923 in Transfac 8.3 that includes 6 experimentally verified ZRE binding sites [32,33]. The Promo algorithm [37,38] generated a position weight matrix (PWM)



B

Name	Novel predicted ZRE core sequence
A	CGAGCGA
B	TGAGCGC
C	TTGGCGA
D	TTAGCGA
E	TGTGCGT
F	TTTGCGA
G	TCAGCGA
H	TGCGCGT
I	CGTGCGA
J	TGCGCGA
K	CGCGCGA
L	AGAGCGA
M	TGGGCGA
N	TGGGCGT
O	TAAGCGA
P	TGAGCGG

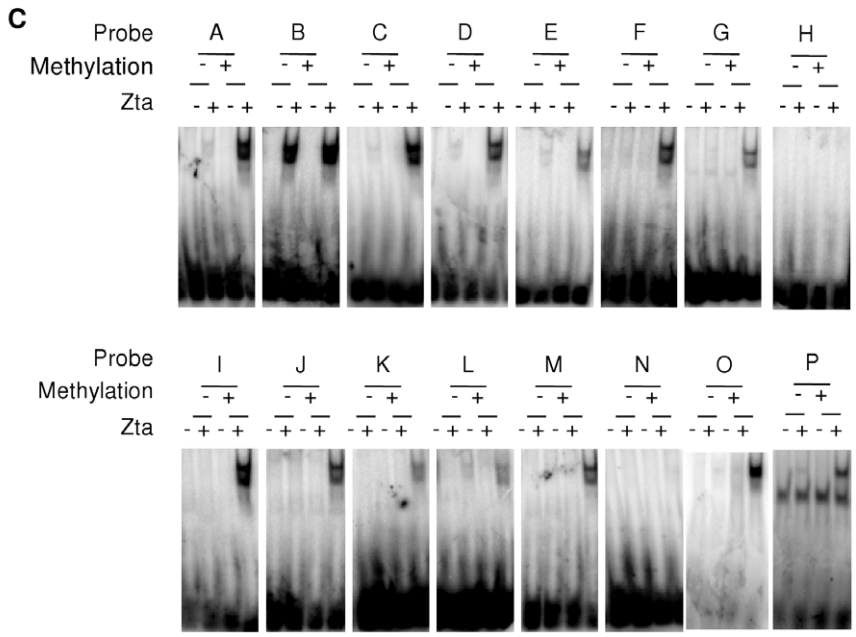


Figure 3. Zta recognition and methylation dependence of PFM$_{CpG5}$ predicted CpG containing ZREs. A. Flow diagram illustrating the information flow from the PFM to the predictions of novel ZREs in the EBV genome and their subsequent evaluation. B. Core heptamer sequences, in both forward and reverse complement, of PFM$_{CpG5}$ predicted CpG containing ZREs within the EBV genome. C. PFM$_{CpG5}$ was used to predict the potential for further ZREs in the EBV genome. Double strand oligonucleotides were generated. Following radio labeling, these were subject to *in vitro* methylation with SssI methyl transferase (+), or a mock reaction (−). Subsequently, they were incubated with *in vitro* translated Zta and subject to EMSA. The reactions contained control lysate, C; or Zta, Z.
doi:10.1371/journal.pone.0025922.g003

based on the T00923 transcription factor entry, and we used it to search 3 well-characterized Zta-responsive promoters from the EBV genome (*BZLF1* promoter (Zp) 500 bp upstream of the published transcription start sites were included. A positive match was taken as one with an 85% similarity rate.

DNA binding assays

Electrophoretic mobility shift assays (EMSA) were undertaken using Zta protein generated in a wheat germ *in vitro* translation system (Promega) and [^{32}P]-radio labeled double strand oligonucleotides, as described previously [39].

Table 1. List of core ZREs included within ZRE_{32}.

ZRE Core Sequence		Class	Names
Forward	**Reverse**		
TGAGCCA	TGGCTCA	I	Zp ZREIIIA
TGAGCCA	TGGCTCA	I	Rp ZRE1
TGAGCCA	TGGCTCA	I	Zp (−365)
TTAGCAA	TTGCTAA	I	Zp ZREIIIB
TGTGTAA	TTACACA	I	DSL ZRE1
TGAGCAA	TTGCTCA	I	DSL ZRE2
TGAGCAA	TTGCTCA	I	DSL ZRE7
TGAGCAA	TTGCTCA	I	BMRF1ZRE(−44)
TGAGCAA	TTGCTCA	I	BMRF1ZRE(−107)
TGTGTCA	TGACACA	I	DSL ZRE3
TGTGTCA	TGACACA	I	DSL ZRE4
TGTGTCA	TGACACA	I	DSL ZRE6
TGTGTCA	TGACACA	I	*DHRS9 ZRE1*
TGTGCAA	TTGCACA	I	DSL ZRE5
TGTGCAA	TTGCACA	I	*CIITA (221)*
TGAGTCA	TGACTCA	I	BSLF2+BMLF1
TGAGTCA	TGACTCA	I	BMRF1 AP-1
TGAGTCA	TGACTCA	I	*IL8 AP1*
TGAGTCA	TGACTCA	I	*DHRS9 (ZRE2)*
TGACTAA	TTAGTCA	I	Fp AP-1-Like Site
TGTGTCT	AGACACA	I	Zp (−96)
TGGCACA	TGTGCCA	I	BMRF1 (−173)
TGAGTAA	TTACTCA	I	*CIITA*
TGAGTAA	TTACTCA	I	*IL13*
GTTGCAA	TTGCAAC	I	*IL-8 ZRE*
GGAG**CG**A	T**CG**CTCC	III	Egr1
TGAG**CG**A	T**CG**CTCA	II	Rp ZRE2
TT**CGCG**A	T**CGCG**AA	III	Rp ZRE3
TGTG**CG**A	T**CG**CACA	III	BMRF1 (−248)
TGAG**CG**T	A**CG**CTCA	III	Nap ZRE2
CGGG**CG**A	T**CG**CCCG	III	Nap ZRE1
CGAGCGA	TCGCTCG	III	CpG5 A
TGAGCGC	GCGCTCA	II	CpG5 B
TTGGCGA	TCGCCAA	III	CpG5 C
TTAGCGA	TCGCTAA	III	CpG5 D
TGTGCGT	ACGCACA	III	CpG5 E
TTTGCGA	TCGCAAA	III	CpG5 F
TCAGCGA	TCGCTGA	III	CpG5 G
CGTGCGA	TCGCACG	III	CpG5 I
TGCGCGA	TCGCGCA	III	CpG5 J
CGCGCGA	TCGCGCG	III	CpG5 K
AGAGCGA	TCGCTCT	III	CpG5 L
TGGGCGA	TCGCCCA	III	CpG5 M
TAAG**CG**A	T**CG**CTTA	III	CpG5 O
TGAG**CG**G	C**CG**CTCA	III	CpG5 P

doi:10.1371/journal.pone.0025922.t001

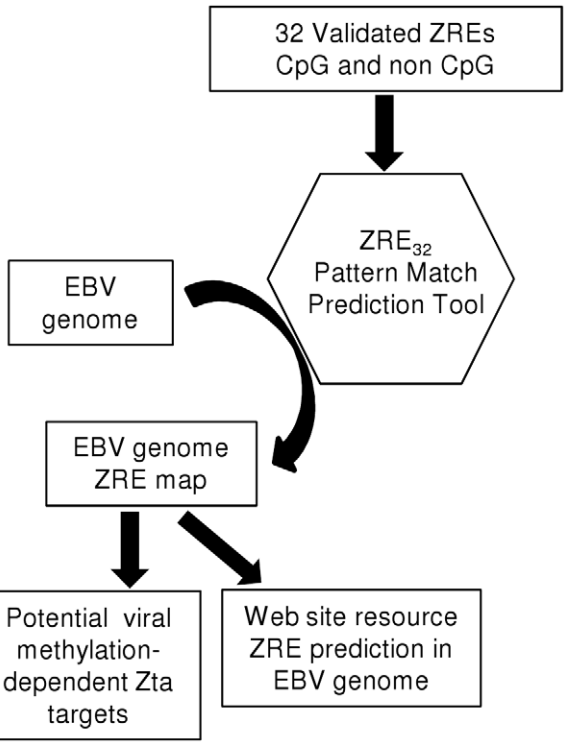


Figure 4. Information flow for the identification of ZREs in the EBV genome. Flow diagram illustrating the use of the pattern matching tool (ZRE_{32}) to identify exact sequence matches for the 32 verified ZREs in the EBV genome.
doi:10.1371/journal.pone.0025922.g004

Where indicated in the figure, the central CpG motif was methylated on both cytosine residues during synthesis (Sigma) or methylated probes were synthesized or methylated *in vitro* using the CpG methyltransferase M.sssI (NEB) [23].

Zta protein (B95-8 strain) was *in vitro* translated using wheatgerm extract (Promega).

Zp (−96): TAAATTTAGGTGTGTCTATGAGGTACA

Zp (−365): ACAGATGGACCTGAGCCACCCGCC

Zp (−662): CCTCTTTGGCTGACACACCTCTCGCCC

Rp (−204) & RpZRE2: CATCTTGTCCTGTGATAAAATCGCTCATAAGCTTAGT

Rp (−439 & −447): ATGACTCGGGTGTGTCCTTGTGTGAGGTCTCACCTG

Rp (−114): CACTCATACTTAAGCGATGCTGATGCA

*BMRF1*p (−173): TGGGGGGTGGTGTGCCATACAAGGGAGC

*BMRF1*p (−248): CCTTGGTGGATGTGCGAGCCATAAAGCA

CpG5 A: AAGCGATGGCCGAGCGATGACTCGTGT

CpG5 B: TCCAGATGACTGAGCGCACGGCCTCAA

CpG5 C: AAAGATCTGGTTGGCGATCCGGTACAC

CpG5 D: TTGGAAAACATTAGCGACATTTACCTG

Table 2. EBV genes that contain CpG ZREs but have no methylation independent ZREs in their regulatory regions, with the kinetics and extent of any change in their expression in Akata cells undergoing lytic cycle [6].

Cycle	Promoter/Gene	Class	Sequence	Offset From Gene Start	Peak expression time (hr)	fold change expression
early lytic	BALF1	II	TCGCTCA	−74	12	8.2
		II	TCGCTCA	−138		
		III	TCGCCCA	−383		
		III	ACGCTCA	−416		
late lytic	BALF4	III	TCGCGCA	−164	24	15.2
		III	TCGCTCG	−254		
		III	ACGCACA	−414		
unknown	BARF0	III	CGAGCGA	45	36	7.3
		III	TGCGCGA	−45		
early lytic	BARF1.2	II	TGAGCGA	114	24	9.2
		II	TGAGCGA	50		
		III	TGGGCGA	−488		
early lytic	BBLF2/BBLF3	III	TCGCCCA	91	12	19.5
early lytic	BBLF4	III	TCGCTCG	−221	12	7.9
		III	TCGCACA	−235		
		III	ACGCTCA	−357		
late lytic	BBRF1	III	AGAGCGA	−31	24	7
		III	TTCGCGA	−414		
late lytic	BBRF3	III	TGGGCGA	−34	24	20.5
		III	TTTGCGA	−177		
late lytic	BDLF1	III	TCGCCCG	−390	24	21.5
late lytic	BGLF2	II	TCGCTCA	133	24	14.3
early lytic	BGLF4	II	GCGCTCA	−260	8	12.1
		III	TCGCTTA	−316		
late lytic	BGRF1/BDRF1	III	CGGGCGA	−49	24	8.3
late lytic	BKRF2	II	TGAGCGC	155	24	14.1
		III	TTAGCGA	185		
early lytic	BKRF4	III	TTAGCGA	−70	24	16.3
		III	TGAGCGG	−246		
		III	CGTGCGA	−302		
early lytic	BLLF2	III	CCGCTCA	−18	24	12.4
	BOLF1	III	TCGCCCG	175	24	4.2
		III	TCGCTGA	−65		
late lytic	BORF1	III	CGGGCGA	130	24	12.4
		III	TGCGCGA	10		
		III	TGAGCGG	−20		
early lytic	BRRF1	III	CGGGCGA	−25	24	7
		III	TGAGCGT	−53		
late lytic	BTRF1	III	TGGGCGA	−292	24	8.5
unknown	BWRF1 repeats	III	CGGGCGA	72	~	~
		III	CGGGCGA	46		
		III	GGAGCGA	−422		
latent	Cp EBNAs	III	TGGGCGA	−70	24–48	6–8.2
		III	TTTGCGA	−204		
		III	GGAGCGA	−248		
latent	Wp EBNAs	III	TCAGCGA	180	48	6.1

doi:10.1371/journal.pone.0025922.t002

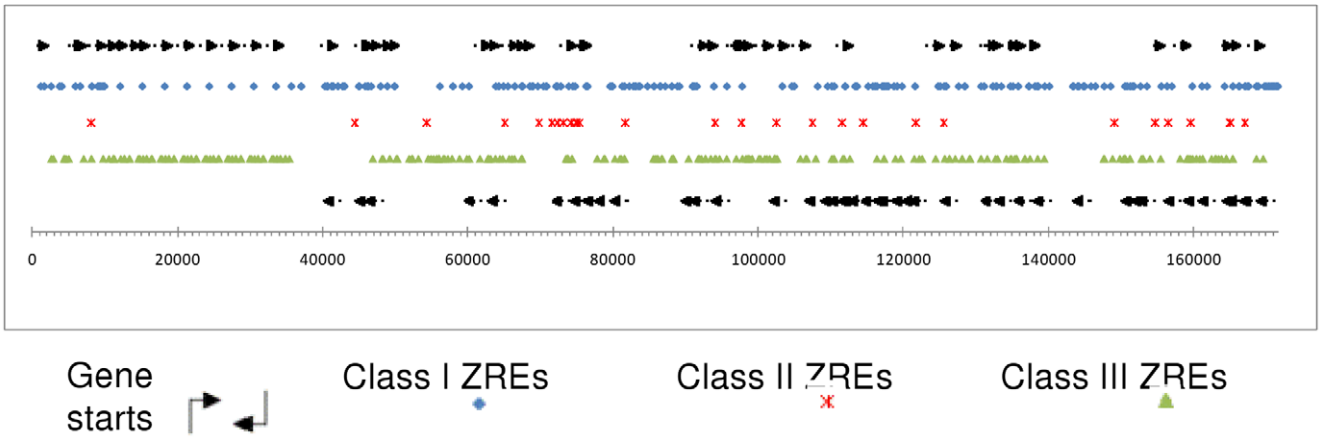


Figure 5. ZREs in the EBV genome. The entire EBV genome was subjected to an exact pattern match search, using the ZRE_{32} set of core sequences. Each site was classed by binding behaviour, and plotted by the first nucleotide of the site to form a genome wide map of ZREs. Class I sites are indicated by blue diamonds, Class II sites are indicated by red stars, Class III sites are indicated by green triangles, and gene starts and direction are indicated by arrows. The location of transcription start sites and their orientation are indicated by arrows.
doi:10.1371/journal.pone.0025922.g005

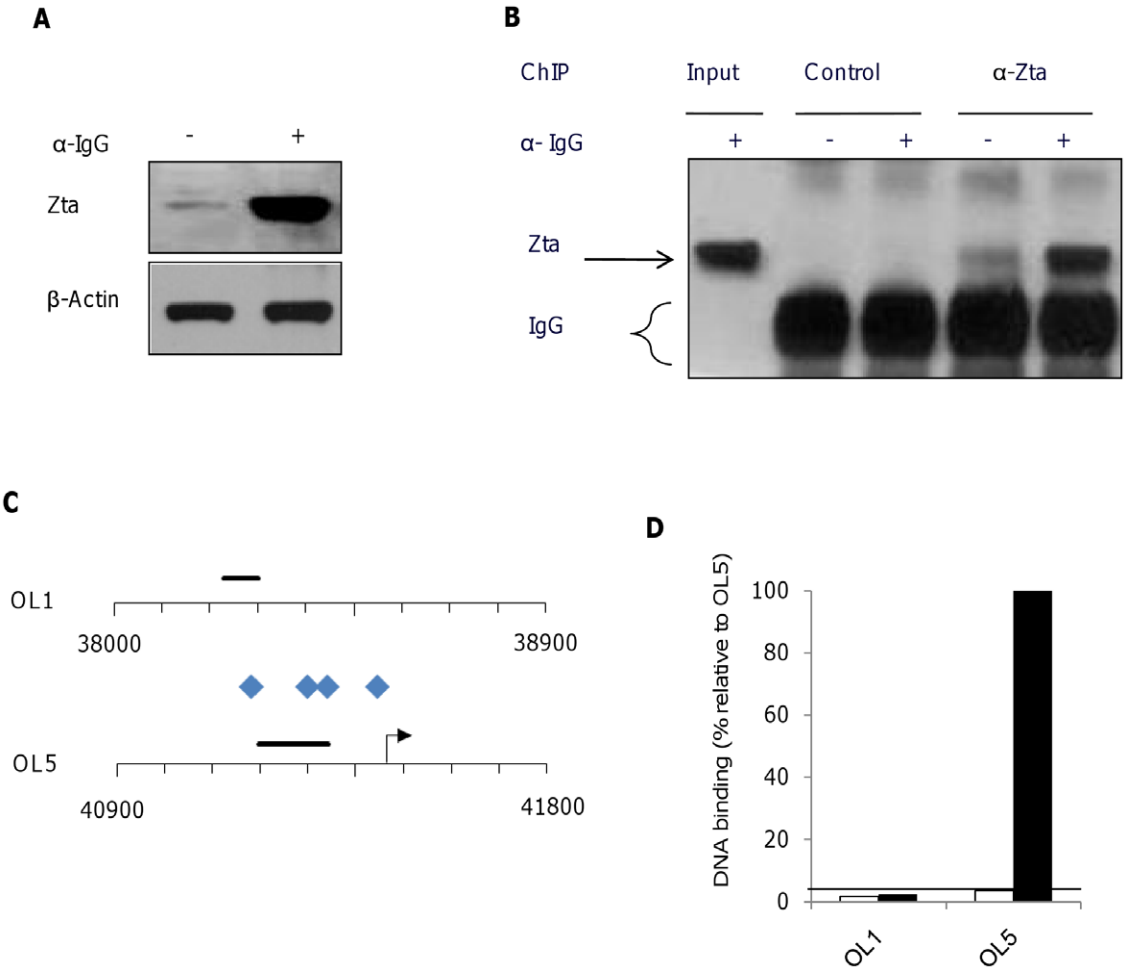
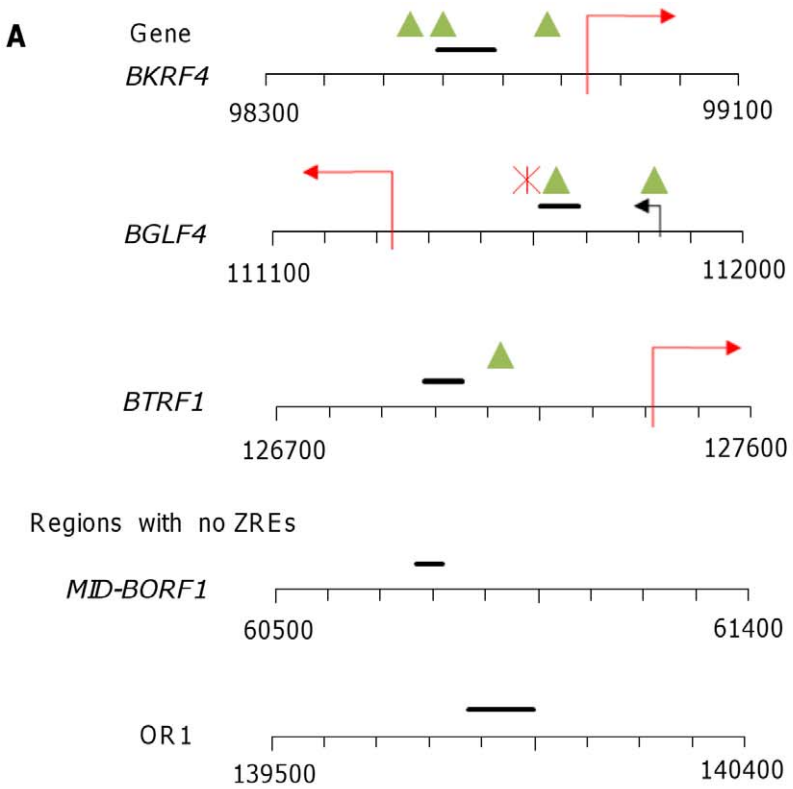


Figure 6. Chromatin precipitation (ChIP) analysis of Zta. Akata cells were induced or not to enter lytic cycle by the addition if anti-IgG as indicated. 48 hours later the cross-linking agent was added and chromatin and total proteins were harvested. A. The proteins were subject to western blot analysis for Zta expression. B. A chromatin precipitation experiment was undertaken with the Zta antibody and a control antibody. The precipitated chromatin and input chromatin were subject to western blot analysis to detect Zta protein. C. The location of ZREs and Q-PCR amplicons are illustrated for a region central to OriLyt left and for a region flanking OriLyt left. The genome co-ordinates are indicated. Transcription start sites and the direction of transcription are shown by arrows. Blue diamonds represents the class I ZREs. The amplicons used for Q-PCR are indicated as black horizontal bars. D. Association of Zta with OriLyt left was assessed using chromatin precipitation from Akata cells in lytic cycle, followed by Q-PCR. The binding is shown relative to maximal binding to Orilyt left.
doi:10.1371/journal.pone.0025922.g006



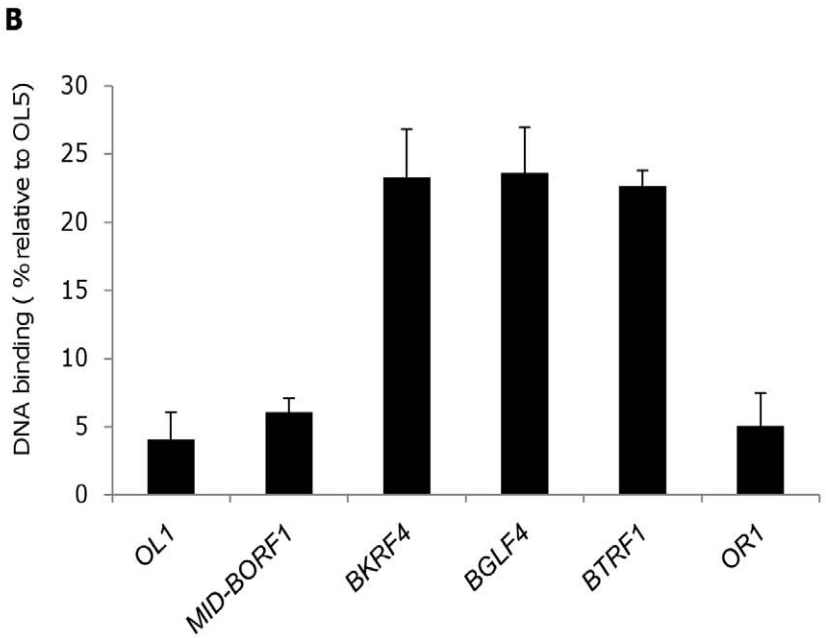


Figure 7. Chromatin precipitation (ChIP) analysis of Zta binding with the EBV genome. A. Five regions of the EBV genome are illustrated, together with gene names and sequence co-ordinates. Three contain core CpG ZREs and two do not. Transcription start sites and the direction of transcription are indicated with arrows. Green triangles represent class III CpG ZREs and the red star represents a class II ZRE. The amplicons used for Q-PCR are indicated as black horizontal bars. F. Association of Zta with the indicated regions of the EBV genome was assessed using chromatin precipitation from Akata cells in early lytic cycle (stalled prior to DNA replication with acyclovir), followed by Q-PCR. The binding is shown relative to maximal binding to OriLyt left (OL5).
doi:10.1371/journal.pone.0025922.g007

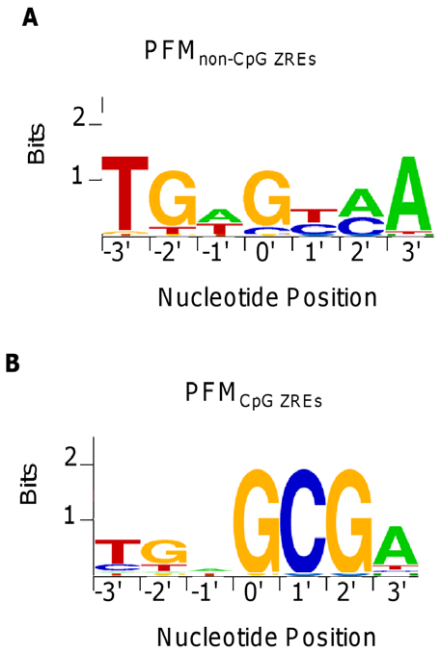


Figure 8. Sequence Logos CpG and non-CpG ZREs. The sequences comprising ZRE_{32}, were divided into those sites which do not contain a CpG motif, class I ZREs and those that do, class II and class III ZREs. A. A PFM was created from ($PFM_{non\text{-}CpG\ ZREs}$) and is displayed using relative letter height. B. A PFM was created from ($PFM_{CpG\ ZREs}$) and is displayed using relative letter height.
doi:10.1371/journal.pone.0025922.g008

CpG5 E: TTTGGCGTCATGTGCGTCTGGATGACA

CpG5 F: CAGACTCTGGTTTGCGAGGCTGGGCGG

CpG5 G: GCCGCCGCACTCAGCGAGGAGGCCTGC

CpG5 H: CGAGGAGGCCTGCGCGTGTTCCTCAAC

CpG5 I: CCAATGTCTGCGTGCGAGCCGGGCTTG

CpG5 J: CTTTGCGCTCTGCGCGAGGACGAGCTC

CpG5 K: GCAGGGCCCCCGCGCGATCTAGGTAGG

CpG5 L: CTCATAGGTCAGAGCGACATAGAGGCG

CpG5 M: TTTCAAGTCGTGGGCGAATTAACTGAG

CpG5 N: AGCAAGGTGCTGGGCGTGGACCGCGCG

CpG5 O: CACTCATACTTAAGCGATGCTGATGCA

CpG5 P: GTCATGTAGGTGAGCGGGCAGTCCTTG

Chromatin Immunoprecipitation

Chromatin was prepared from Akata cells [35], following induction with anti IgG, in the presence of 100 μM acyclovir essentially as described in [40], except that a mixture of Protein A and protein G were used to capture antibodies. Precipitation was undertaken using an amino-terminal Zta antibody from Santa Cruz.

Primers: absolute genomic position and sequence

OL5 F 41195-41216 CAGCTGACCGATGCTCGCCA

OL5 R 41342-41323 ATGGTGAGGCAGGCAAGGCG

BKRF4 F 98591-98611 CATTGCTCTCTGAGCGGTTA

BKRF4 R 98688-98667 ACCAGATGCTTCTTGGAGTTG

BGLF4 F 111611-111631 ACCGAGGCTCTTAGTTGCTG

BGLF4 R 111686-111666 GTTGCGGACATGGTGACTTA

BTRF1 F 126981-127000 AGCTACGCAATCGGAGTCA

BTRF1 R 127052-127034 GGAGGCGCAGTCTAGCAG

Region with no ZREs

OR1 F 139875-139895 CCGCATGTCCAACCACCACG

OR1 R 139997-139976 ATGCTACCTAGGCCTGCGTCC

OL1 F 38229-38249 GCGCAACAGTGCCACCAACC

OL1 R 38302-38282 CAGGACCTGGCGGTAGTGCAG

MID-BORF1 F 60768-60788 TGCCTGAGACCTCTCGGACGG

MID-BORF1 R 60817-60792 CCACGACGCAGTCCTTAGGATCATG

Generation and application of ZRE PFMs

A position frequency matrix (PFM_{CpG5}) was created using 5 CpG containing ZRE core binding sequences (Figure 3) and used to search (a) the 3 EBV promoters using the algorithm Matscan [41], and (b) the complete Human herpesvirus 4 (Epstein-Barr virus) Genome NC_007605 extracted from GenBank [42]. An 85% similarity score was used to define a positive match to the PFM_{CpG5}. PFMs for non-CpG ZREs and CpG ZREs were generated in a similar manner and displayed using WEBLOGO [43].

Exact pattern matching was employed to search for each of the 32 core ZRE sequences within the EBV genome. A rolling window of seven nucleotides was used and an exact comparison of each of the core ZREs sequences, was made. In addition, the reverse complement of the sequence was checked in the same manner.

A MySQL database of the locations of the exact matches within the EBV genome was generated and simple analyses can be conducted using a web interface which is publically available at URL: http://bioinf.biochem.sussex.ac.uk/EBV/. The database uses gene annotations extracted from the RefSeq NC_007605 entry and data from the exact match predictions.

Supporting Information

Table S1 The sequences of published ZREs used in this study are shown, together with their names and references. CpG motifs are shown in bold.
(DOCX)

Table S2 ZREs predicted using PROMO, together with the results of their evaluation by DNA binding assays (EMSA) are shown. CpG motifs are shown in bold.
(DOCX)

Table S3 All ZREs in the EBV genome are shown with the position of the central nucleotide using RefSeq NC_007605.
(DOCX)

Table S4 The number of ZREs close to each EBV promoter (between −1000 and +200 from the annotated gene start) are shown using RefSeq NC_007605.
(DOCX)

Acknowledgments

We thank Dr Kay Osborn, Carl May and Lisa Murray for addition technical help, Drs Michelle West and Gordon Peters for comments on the manuscript.

Author Contributions

Conceived and designed the experiments: KF DT JH SJ AJS. Performed the experiments: KF DT JH SR. Analyzed the data: KF DT JH SR SJ AJS. Wrote the paper: KF AJS.

References

1. Rowe M, Kelly GL, Bell AI, Rickinson AB (2009) Burkitt's lymphoma: the Rosetta Stone deciphering Epstein-Barr virus biology. Semin Cancer Biol 19: 377–388.
2. Minarovits J (2006) Epigenotypes of latent herpesvirus genomes. Curr Top Microbiol Immunol 310: 61–80.
3. Kalla M, Schmeinck A, Bergbauer M, Pich D, Hammerschmidt W (2010) AP-1 homolog BZLF1 of Epstein-Barr virus has two essential functions dependent on the epigenetic state of the viral genome. Proc Natl Acad Sci U S A 107: 850–855.
4. Fernandez AF, Rosales C, Lopez-Nieva P, Grana O, Ballestar E, et al. (2009) The dynamic DNA methylomes of double-stranded DNA viruses associated with human cancer. Genome Res 19: 438–451.
5. Miller G, El-Guindy A, Countryman J, Ye J, Gradoville L (2007) Lytic cycle switches of oncogenic human gammaherpesviruses(1). Adv Cancer Res 97: 81–109.
6. Yuan J, Cahir-McFarland E, Zhao B, Kieff E (2006) Virus and cell RNAs expressed during Epstein-Barr virus replication. J Virol 80: 2548–2565.
7. Broderick P, Hubank M, Sinclair AJ (2009) Effects of Epstein-Barr virus on host gene expression in Burkitt's lymphoma cell lines. Chinese Journal of Cancer. In press.
8. Chang Y, Lee HH, Chen YT, Lu J, Wu SY, et al. (2006) Induction of the early growth response 1 gene by Epstein-Barr virus lytic transactivator Zta. J Virol 80: 7748–7755.
9. Jones RJ, Dickerson S, Bhende PM, Delecluse HJ, Kenney SC (2007) Epstein-Barr virus lytic infection induces retinoic acid-responsive genes through induction of a retinol-metabolizing enzyme, DHRS9. J Biol Chem 282: 8317–8324.
10. Li D, Qian L, Chen C, Shi M, Yu M, et al. (2009) Down-regulation of MHC class II expression through inhibition of CIITA transcription by lytic transactivator Zta during Epstein-Barr virus reactivation. J Immunol 182: 1799–1809.
11. Morrison TE, Mauser A, Wong A, Ting JP, Kenney SC (2001) Inhibition of IFN-gamma signaling by an Epstein-Barr virus immediate-early protein. Immunity 15: 787–799.
12. Tsai SC, Lin SJ, Chen PW, Luo WY, Yeh TH, et al. (2009) EBV Zta protein induces the expression of interleukin-13, promoting the proliferation of EBV-infected B cells and lymphoblastoid cell lines. Blood 114: 109–118.
13. Hsu M, Wu SY, Chang SS, Su IJ, Tsai CH, et al. (2008) Epstein-Barr virus lytic transactivator Zta enhances chemotactic activity through induction of interleukin-8 in nasopharyngeal carcinoma cells. J Virol 82: 3679–3688.
14. Mahot S, Sergeant A, Drouet E, Gruffat H (2003) A novel function for the Epstein-Barr virus transcription factor EB1/Zta: induction of transcription of the hIL-10 gene. J Gen Virol 84: 965–974.
15. Countryman J, Miller G (1985) Activation of expression of latent Epstein-Barr herpesvirus after transfer with a small cloned subfragment of heterogeneoous viral DNA. Proc Natl Acad Sci USA 81: 7632–7636.
16. Sinclair AJ (2003) bZIP proteins of human Gamma Herpesviruses. Journal of General Virology 84: 1941–1949.
17. Sinclair AJ (2006) Unexpected structure of Epstein-Barr virus lytic cycle activator Zta. Trends Microbiol 14: 289–291.
18. Petosa C, Morand P, Baudin F, Moulin M, Artero JB, et al. (2006) Structural Basis of Lytic Cycle Activation by the Epstein-Barr Virus ZEBRA Protein. Mol Cell 21: 565–572.
19. Karlsson QH, Schelcher C, Verrall E, Petosa C, Sinclair AJ (2008) The reversal of epigenetic silencing of the EBV genome is regulated by viral bZIP protein. Biochem Soc Trans 36: 637–639.
20. Bhende PM, Seaman WT, Delecluse HJ, Kenney SC (2004) The EBV lytic switch protein, Z, preferentially binds to and activates the methylated viral genome. Nat Genet 36: 1099–1104.
21. Bhende PM, Seaman WT, Delecluse HJ, Kenney SC (2005) BZLF1 activation of the methylated form of the BRLF1 immediate-early promoter is regulated by BZLF1 residue 186. J Virol 79: 7338–7348.
22. Dickerson SJ, Xing Y, Robinson AR, Seaman WT, Gruffat H, et al. (2009) Methylation-dependent binding of the epstein-barr virus BZLF1 protein to viral promoters. PLoS Pathog 5: e1000356.
23. Heather J, Flower K, Isaac S, Sinclair AJ (2009) The Epstein-Barr virus lytic cycle activator Zta interacts with methylated ZRE in the promoter of host target gene egr1. J Gen Virol 90: 1450–1454.
24. Karlsson QH, Schelcher C, Verrall E, Petosa C, Sinclair AJ (2008) Methylated DNA recognition during the reversal of epigenetic silencing is regulated by cysteine and serine residues in the Epstein-Barr virus lytic switch protein. PLoS Pathog 4: e1000005.
25. Bergbauer M, Kalla M, Schmeinck A, Gobel C, Rothbauer U, et al. (2010) CpG-methylation regulates a class of Epstein-Barr virus promoters. PLoS Pathog 6.
26. Flemington E, Speck SH (1990) Autoregulation of Epstein-Barr Virus putative lytic switch gene BZLF1. J Virol 64: 1227–1232.
27. Lieberman PM, Hardwick JM, Sample J, Hayward GS, Hayward SD (1990) The zta transactivator involved in induction of lytic cycle gene expression in Epstein-Barr virus-infected lymphocytes binds to both AP-1 and ZRE sites in target promoter and enhancer regions. J Virol 64: 1143–1155.
28. Urier G, M. B, Chambard P, Sergeant A (1989) The Epstein-Barr virus early protein EB1 activates trasncription from different responsive elements including AP-1 binding sites. EMBO J 8: 1447–1453.
29. Packham G, Economou A, Rooney CM, Rowe DT, Farrell PJ (1990) Structure and function of the Epstein-Barr virus BZLF1 protein. J Virol 64: 2110–2116.
30. Sinclair AJ, Brimmell M, Shanahan F, Farrell PJ (1991) Pathways of activation of the Epstein-Barr virus productive cycle. J Virol 65: 2237–2244.
31. Taylor N, Flemington E, Kolman JL, Baumann RP, Speck SH, et al. (1991) ZEBRA and a Fos-GCN4 chimeric protein differ in their DNA-binding specificities for sites in the Epstein-Barr virus BZLF1 promoter. J Virol 65: 4033–4041.
32. Wingender E, Dietze P, Karas H, Knuppel R (1996) TRANSFAC: a database on transcription factors and their DNA binding sites. Nucleic Acids Res 24: 238–241.
33. Fu Y, Weng Z (2005) Improvement of TRANSFAC matrices using multiple local alignment of transcription factor binding site sequences. Genome Inform 16: 68–72.
34. Farrell PJ (2005) Epstein-Barr virus genome. In: Robertson ES, ed. Epstein-Barr virus. Wymondham: Caister. pp 263–288.
35. Takada K, Ono Y (1989) Synchronous and sequential activation of latently infected Epstein-Barr virus genomes. J Virol 63: 445–449.
36. Rishi V, Bhattacharya P, Chatterjee R, Rozenberg J, Zhao J, et al. (2010) CpG methylation of half-CRE sequences creates C/EBPalpha binding sites that activate some tissue-specific genes. Proc Natl Acad Sci U S A 107: 20311–20316.
37. Farre D, Roset R, Huerta M, Adsuara JE, Rosello L, et al. (2003) Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. Nucleic Acids Res 31: 3651–3653.
38. Messeguer X, Escudero R, Farre D, Nunez O, Martinez J, et al. (2002) PROMO: detection of known transcription regulatory elements using species-tailored searches. Bioinformatics 18: 333–334.
39. Flower K, Hellen E, Newport MJ, Jones S, Sinclair AJ (2010) Evaluation of a prediction protocol to identify potential targets of epigenetic reprogramming by the cancer associated Epstein Barr virus. PLoS One 5: e9443.
40. Bark-Jones SJ, Webb HM, West MJ (2006) EBV EBNA 2 stimulates CDK9-dependent transcription and RNA polymerase II phosphorylation on serine 5. Oncogene 25: 1775–1785.
41. Blanco E, Messeguer X, Smith TF, Guigo R (2006) Transcription factor map alignment of promoter regions. PLoS Comput Biol 2: e49.
42. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2009) GenBank. Nucleic Acids Res 37: D26–31.
43. Schneider TD, Stephens RM (1990) Sequence logos: a new way to display consensus sequences. Nucleic Acids Res 18: 6097–6100.