



HAL
open science

Étude de la régulation de l'expression des gènes tardifs du virus d'Epstein-Barr

Valentin Aubry

► **To cite this version:**

Valentin Aubry. Étude de la régulation de l'expression des gènes tardifs du virus d'Epstein-Barr. Virologie. Université de Lyon, 2016. Français. NNT : 2016LYSEN029 . tel-01699060

HAL Id: tel-01699060

<https://theses.hal.science/tel-01699060v1>

Submitted on 2 Feb 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Numéro National de Thèse : 2016LYSEN029

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée par
l'Ecole Normale Supérieure de Lyon

Ecole Doctorale N°340
Biologie Moléculaire, Intégrative et Cellulaire

Spécialité de doctorat : Sciences Exactes et Expérimentales
Discipline : Sciences de la vie

Soutenue publiquement le 16/11/2016, par :
Valentin AUBRY

**Etude de la régulation de l'expression
des gènes tardifs du virus
d'Epstein-Barr**

Devant le jury composé de :

Mme. Hans Fabienne, Maître de conférences à l'Université Joseph Fournier de Grenoble	Rapporteure
M. Maréchal Vincent, Professeur à l'université Pierre et Marie Curie de Paris	Rapporteur
Mme Horvat Branka, Directrice de recherche à l'INSERM	Examinatrice
M. Lomonte Patrick, Directeur de recherche au CNRS	Examineur
M. Mariamé Bernard, Chargé de Recherche à l'INSERM	Examineur
M. Gruffat Henri, Chargé de recherche à l'INSERM	Directeur de thèse

Résumé :

Le virus d'Epstein-Barr (EBV) est un Herpesvirus humain appartenant à la sous-famille des *γ-Herpesvirinae*. L'expression des gènes d'EBV est régulée très finement, de manière séquentielle et débute par l'expression des gènes immédiats précoces dont les produits contrôlent l'expression des gènes précoces. Les produits de ces derniers permettent la réplication de l'ADN viral et la synthèse des gènes tardifs. Les mécanismes contrôlant l'expression des gènes immédiats précoces et précoces sont assez bien compris, à contrario de ceux qui contrôlent l'expression des gènes tardifs. En effet, il est connu que ces derniers ne sont exprimés qu'après réplication de l'ADN viral. Par ailleurs, les promoteurs des gènes tardifs présentent une structure différente des promoteurs de gènes immédiats précoces, précoces ou des gènes cellulaires. Ils sont essentiellement caractérisés par la présence d'une boîte TATT à la place de la TATA-box canonique. EBV a la particularité de coder pour une TBP-like, la protéine BcRF1 qui est essentielle mais pas suffisante pour l'expression des gènes viraux tardifs et qui se fixe spécifiquement sur la séquence TATT des promoteurs viraux tardifs.

Notre objectif était principalement d'identifier l'ensemble des protéines virales nécessaires à l'expression des gènes tardifs d'EBV.

Nous avons ainsi démontré qu'en plus de la protéine BcRF1, EBV code pour cinq autres protéines nécessaires à l'expression des gènes viraux tardifs : BFRF2, BGLF3, BVLF1, BDLF4 et BDLF3.5. Ces six protéines virales forment un complexe qui permet le recrutement de l'ARN polymérase II sur les promoteurs des gènes viraux tardifs. Nous avons nommé ce complexe vPIC (viral Pre-Initiation Complex), par analogie avec le complexe d'initiation de la transcription cellulaire formé autour de la TATA-Binding Protein (TBP). Le vPIC est conservé chez les herpes virus des sous familles β et γ , mais est absent chez les herpes virus de la sous famille des α . Par ailleurs, nos résultats suggèrent que le complexe vPIC interagit avec le complexe de réplication viral ce qui peut expliquer la liaison entre réplication de l'ADN viral et transcription des gènes viraux tardifs. Ces résultats permettent ainsi de mieux comprendre les mécanismes de régulation de l'expression des gènes tardifs d'EBV et serviront de base pour la recherche de nouveaux antiviraux.

Mots clés : virus d'Epstein-Bar, β - et γ -Herpesvirinae, gènes tardifs, transcription, vPIC, réplication de l'ADN viral.

Abstract :

The Epstein-Barr virus (EBV) is a human Herpesvirus belonging to the subfamily of *γ-Herpesvirinae*. EBV genes expression is tightly regulated, in a sequential manner and begins with the expression of immediate early genes whose products control the expression of early genes. Products of these allow the viral DNA synthesis and late genes expression. Mechanisms controlling immediate early and early genes expression are well documented, in contrast to those controlling late genes expression. Indeed, it is known that late genes are expressed after viral DNA replication. Furthermore, late genes promoters have a different structure of those of immediate early, early and cellular genes. They are essentially characterized by the presence of a TATT box instead of the canonical TATA-box. EBV has the feature to encode for a TBP-like, BcRF1 protein that is essential but not sufficient for late viral genes expression and that binds specifically to the TATT sequence of the late viral promoters.

Our goal was primarily to identify the set of viral proteins necessary for late EBV genes expression.

Thus, we have demonstrated that in addition to the BcRF1 protein, EBV encode for five other proteins necessary for late viral genes expression: BFRF2, BGLF3, BVLF1, BDLF4 and BDLF3.5. These six viral proteins form a complex enables to recruit the RNA polymerase II on the late viral promoters. We have named this complex vPIC (viral Pre-Initiation Complex) by analogy with the cellular complex of transcription initiation, formed around the TATA-Binding Protein (TBP). The vPIC is conserved in β - and γ -Herpesvirus subfamilies, but not in the α -Herpesvirus subfamily. Furthermore, our results suggest that vPIC interacts with the viral DNA replication complex, which can explain the link between viral DNA replication and late viral genes transcription. These results bring new insights to the mechanisms regulating late EBV genes expression and provide a basis for the search of new antiviral drugs.

Key words: Epstein-Barr virus, β - and γ -Herpesvirinae, late genes, transcription, vPIC, viral DNA replication.

Abréviations

A

ADN : Acide DésoxyriboNucléique
AcMNPV : Autographa californica NucleoPolyhedrovirus
ADP : Adénosine DiPhosphate
AP : transcription factor Activator Protein
ARN : Acide RiboNucléique
ARNm : ARN messenger
ARNr : ARN ribosomaux
ARNt : ARN de transfert
ARN pol II : ARN polymérase II cellulaire
ATF : Activating Transcription Factor ou CREB
ATP : Adénosine Tri-Phosphate
A/T : dinucléotides Adénosine/Thymidine
AU : Arbitral Unit

B

BA : Butyrate de Sodium
BKV : Virus BK
BRE : TFIIB Responsive Elements
bZIP : basic ZIPper

C

C/EBP : CCAAT/enhancer-binding protein
Calm : Calmoduline
ChIP : Chromatine Immuno-Précipitation
CpG : séquences d'ADN riches en Cytidine et Guanosine
CRE : CREB Responsive Element
CREB : Cyclic AMP-Responsive Element Binding protein
C-Ter : Carboxyl-Terminal
CTD : Carboxyl-Terminal Domain

D

DPE : Downstream Promoter Element

E

EB1 : Epstein-Barr protein 1
EBNA : Epstein-Barr Nuclear Antigen
EBV : Epstein-Barr Virus
eIF : eukaryotic Initiator Factors of translation

G

GST : Glutathion-S-transferase

H

H2A : Histone 2A
H2B : Histone 2B
H3 : Histone 3
H4 : Histone 4
HCMV : Human Cytomegalovirus
HEK293 : Human Embryonic Kindey 293 cell line
HEK293_{EBV} : Human Embryonic Kindey 293 EBV positive cell line
HEK293_{EBVΔX} : Human Embryonic Kindey 293 EBV recombinant cell line
HF : Histone Fold
HHV : Human HerpesVirus
hnRNP : heterogeneous nuclear RiboNucleoProtein
HPV : Human PapillomaVirus
Hr : Homologous Region
HSV : human Herpes Simplex Virus

I

INR : Elément Initiateur
IP : Immédiat Précoce
IR : Internal Repeat
ITR : Internal Terminal Repeat

J

JCV : Virus JC

K

K : Lysine
KDa : Kilo Dalton
Kbp : Kilo paires de bases
KSHV : Kaposi Sarcoma associated HerpesVirus

L

L4P : Promoteur L4
LCL : Lymphoblastoïd Cell Line
LCR : Long Control Region
LEF : Late Expression Factor

M

me3 : triméthylation
MHV68 : Murid HerpesVirus 68
miRNA : micro-ARN
MLP : Major Late Promoter
MS : Spectrométrie de Masse

N

NCCR : Non-Coding Control Region
ncRNA : ARN non codant
NFAT : Nuclear Factor of Activated T-cells
NF-I : Nuclear Factor I
NF- κ B : Nuclear Factor κ B
NLS : Nuclear Localization Signal
nm : nanomètre
Nrf1 : facteur de transcription Nuclear respiratory factor 1
N-Ter : Amino-Terminal

O

ORF : Open Reading Frame
Ori : Origine de réplication
OriLyt : Origine de réplication Lytique
OriP : Origine de réplication Plasmidique
Oct-11 : Octamer binding protein 11 ou POU domain class 2, transcription factor 3

P

P : Précoce
PAA : PhosphonoAcetic Acid
pAE : Early polyAdenylation signal
PAGE : PolyAcrylamide Gel Electrophoresis
pAL : Late polyAdenylation signal
pb : paires de bases
PCR : Réaction de Polymérisation en Chaîne
PIC : Pre-Initiation Complex of transcription
PKC : Protéine Kinase C
PP2A : Protéine Phosphatase 2A
pTATT-Luc : plasmide rapporteur luciférase dont le gène est sous contrôle d'un promoteur de gène tardif TATT

Q

qPCR : PCR quantitative

R

RNAP : ARN polymérase
Rpb1 : RNA polymérase subunit 1
RT-qPCR : Transcription Inverse qPCR.

S

SA : Site Accepteur d'épissage
SD : Site Donneur d'épissage
SDS : Sodium DodécylSulfate
siRNA : small interfering RNA

Sp1 : Facteur de transcription Sp1

SR : Serine/Arginine rich protein

Strep : Streptavidine

SV40 : Simian Virus 40

T

T : Tardif
t-Ag : Antigène petit t
T-Ag : Antigène grand T
TAF : TATA-box Associated Factor
TAP : Tandem Affinity Purification
TBP : TATA-box Binding Protein
TGF- β : transforming Growth Factor β
TEF-1 : Transcription Enhancer Factor 1
TFIIA : Transcription Factor II A
TFIIB : Transcription Factor II B
TFIID : Transcription Factor II D
TFIIE : Transcription Factor II E
TFIIF : Transcription Factor II F
TFIIH : Transcription Factor II H
TFII-I : Transcription Factor II Initiator
TPA : 12-O-Tétradécanoyl-Phorbol 13-Acétate
TR : Terminal Repeat
TRE : TPA Responsive Element
TRF1 : TBP-Related Factor 1
TRF2 : TBP-like Factors
TRF3 : TBP-Related Factor 3
TSS : Transcriptional Start Site (Site d'initiation de la transcription)
Tst-1 : Oct-6 : POU domain class 3, transcription factor 1
T4 : Bactériophage T4

V

VLF1 : Very Late Factor 1
vPIC : viral Pre-Initiation Complex of transcription
VZV : Varicella-Zoster Virus

W

WB : Western Blot

Y

YY-1 : Transcriptional repressor protein
Ying Yang

Z

ZRE : BZLF1 Responsive Element

Table des matières

PARTIE I : INTRODUCTION.....	10
I) La transcription cellulaire.	11
A) Les séquences cis-régulatrices de la transcription.	12
B) L'assemblage du complexe général d'initiation de la transcription cellulaire.....	14
C) Une diversification des facteurs et de complexes de transcription.	16
D) L'analyse tridimensionnelle du PIC, une nouvelle ère de compréhension.	18
E) La chromatine et son rôle dans la régulation du recrutement du PIC.	22
II) Les virus à ADN.	23
A) Les <i>Caudovirales</i> (bactériophages).....	24
1) Les promoteurs des gènes viraux précoces.	25
2) Les promoteurs des gènes intermédiaires.	26
3) Les promoteurs des gènes viraux tardifs et leur régulation spécifique.	27
B) Les <i>Baculoviridae</i> (baculovirus).	32
1) L'expression des gènes viraux précoces.	33
2) L'expression des gènes viraux tardifs.	36
C) Les <i>Adenoviridae</i> (Adénovirus).	42
1) L'expression des gènes viraux précoces.	43
2) L'expression de gènes couplée à la réplication de l'ADN viral.....	45
3) La transcription des gènes tardifs.....	47
4) L'épissage des ARNm tardifs, une étape clé dans leur expression.....	50
5) La traduction des ARNm viraux est favorisée lors de l'étape tardive de l'infection.....	53
D) Les <i>Polyomaviridae</i> (Polyomavirus).	56
1) Généralités.	56
2) La région précoce.....	59
3) La région tardive.	60
E) Les <i>Papillomaviridae</i> (Papillomavirus).	66
1) Généralités.	67
2) La transcription virale.	71
a) Structure du promoteur précoce.....	71
b) Structure du promoteur tardif.....	73
3) La régulation de l'expression des gènes viraux tardifs par l'épissage et la polyadénylation.	75
a) Les mécanismes d'inhibition de l'expression des gènes viraux tardifs dans la région précoce.	75
b) La régulation de l'expression des gènes tardifs dans la région tardive.....	78
F) Synthèse sur les virus à ADN.	82
G) Les <i>Herpesviridae</i> (Herpesvirus).	83
1) Le virus d'Epstein-Barr.....	84
2) Nomenclature.....	85
3) Infection et pathologies associées.....	85
4) Le cycle viral.....	86

5) Le cycle viral productif.....	88
6) Structure des promoteurs viraux immédiats précoces et précoces.....	91
7) Les gènes viraux tardifs.....	93
8) Caractérisation de BcRF1, une avancée majeure dans la compréhension de la régulation de l'expression des gènes tardifs.....	95
9) Homologies entre facteurs protéiques de Beta et Gamma-Herpesvirus.....	96
10) Mécanismes de régulation de l'expression des gènes tardifs d'Alpha-Herpesvirus.....	98
11) Contexte et hypothèse.....	99
PARTIE II : RESULTATS	101
I) Matériel et Méthode :.....	102
II) Résultats :	107
A) L'expression des gènes viraux tardifs est dépendante d'un complexe d'initiation viral de la transcription : résumé de l'article scientifique publié en 2014 (Aubry et al., 2014).	107
B) BVLF1, BDLF4 et BDLF3.5 sont nécessaires à l'expression des gènes viraux tardifs.	109
1) BVLF1, BDLF4 et BDLF3.5 sont nécessaires à l'expression des gènes viraux tardifs <i>in vitro</i>	110
2) BVLF1, BDLF4 et BDLF3.5 ne sont pas impliquées dans l'expression des protéines virales précoces, <i>in vivo</i>	111
3) BVLF1, BDLF4 et BDLF3.5 n'ont pas de rôle <i>in vivo</i> , dans la réplication de l'ADN viral.	113
4) BVLF1, BDLF4 et BDLF3.5 sont nécessaires à l'expression des ARNm tardifs du virus d'Epstein-Barr.	114
5) Conclusion sur le rôle de BVLF1, BDLF4 et BDLF3.5.....	116
C) L'interactome du vPIC.....	116
1) Le vPIC et l'ARN polymérase II cellulaire.....	117
2) Interactions individuelles entre protéines du vPIC.....	118
a) GST-BcRF1	119
b) Synthèse des interactions individuelles des protéines du vPIC.....	120
3) L'interactome du vPIC : toutes les protéines du vPIC moins une.....	122
a) GST-BcRF1 + vPIC-1.....	122
b) Synthèse des interactions des protéines du vPIC.....	125
4) L'association avec BDLF4 dans le complexe viral.....	127
5) Conclusions sur l'interactome vPIC.....	128
D) Recherche des partenaires éventuels du vPIC.....	129
1) BTRF1 et BGLF4 ont-elles un rôle dans l'expression des gènes tardifs d'EBV ?.....	130
2) Identification des partenaires potentiels du vPIC en condition virale.....	134
PARTIE III : Conclusions & Discussions	143
I) La découverte d'un complexe de pré-initiation de la transcription des gènes viraux tardifs, conservé entre β- et γ-Herpesvirus.....	144
II) Caractérisation du rôle de BVLF1, BDLF4 et BDLF3.5 dans l'expression des gènes viraux tardifs.	147
III) Interactome du vPIC.....	149

IV) Identification des partenaires potentiels du vPIC.....	151
1) BTRF1 et BGLF4.....	151
2) Analyse par spectrométrie de masse.....	152
V) Le complexe de réplication de l'ADN viral, partenaire du vPIC ?	153
1) L'hélicase/primase servirait-elle de cargo pour le transport du vPIC dans le noyau ?	154
2) L'hélicase/primase recruteraient-elles le vPIC sur l'ADN ?	155
3) L'expression des gènes viraux tardifs est-elle dépendante du complexe de réplication de l'ADN viral ou bien nécessite-t-elle uniquement un changement de conformation de la chromatine ?	156
VI) Modèle.....	158
VII) Perspectives du projet.....	160
PARTIE IV : Références Bibliographiques.....	162
PARTIE V : Annexes Résultats	199

PARTIE I :

INTRODUCTION

I) La transcription cellulaire.

La régulation de l'expression des gènes, qu'ils soient d'origine eucaryote, bactérienne, d'archée ou virale, est une étape clé du bon fonctionnement des organismes. De nombreux dispositifs moléculaires permettent de contrôler cette expression. La première étape de régulation de l'expression des gènes est la transcription.

La transcription est un mécanisme permettant la synthèse d'ARN à partir d'un brin matrice d'ADN. Elle est assurée par des enzymes nommées ARN polymérases.

Chez les eucaryotes, il existe trois catégories d'ARN polymérases, impliquées dans la transcription des différentes classes de gènes :

- l'ARN polymérase de type I (ARN pol I) permet la transcription des ARN ribosomiaux (ARNr), à l'exception de l'ARNr 5S (Russell and Zomerdijk, 2006).

- l'ARN polymérase de type II (ARN pol II) transcrit des ARN messagers (ARNm) et de petits ARN non codants (ncRNA).

- l'ARN polymérase de type III (ARN pol III) synthétise des ARN de transfert (ARNt), l'ARNr 5S et des ncRNA (Dieci et al., 2007).

Certains facteurs impliqués dans l'initiation de la transcription de ces trois catégories de gènes sont conservés ou sont homologues (Vannini and Cramer, 2012). Malgré le fait que des éléments puissent être retrouvés dans les différents systèmes de transcription, nous nous focaliserons par la suite, uniquement sur la synthèse d'ARN messagers (ARNm) par l'ARN polymérase II.

La transcription débute au niveau du site d'initiation de la transcription ou site +1 (TSS), correspondant au premier nucléotide transcrit, et s'achève au niveau du site de terminaison de la transcription. En amont et/ou en aval du site d'initiation de la transcription, des séquences d'ADN régulatrices sont présentes. Ce sont des séquences dites cis-régulatrices sur lesquelles se fixent des facteurs protéiques (trans-régulateurs). Il existe une multitude d'éléments cis-régulateurs et de protéines cellulaires capables de s'y lier. L'ensemble de ces acteurs s'assemble de manière très précise, et sont finement régulés. Ils ont ainsi un rôle primordial dans le mécanisme de synthèse des ARNm.

A) Les séquences cis-régulatrices de la transcription.

Parmi toutes les séquences d'ADN régulatrices, deux catégories d'éléments sont classiquement définies : les éléments promoteurs et les éléments « enhancers ». Les promoteurs sont de courtes séquences nucléotidiques situées à proximité du site d'initiation de la transcription. À l'inverse, les séquences situées à plusieurs centaines voire milliers de paires de bases (pb) du site d'initiation de la transcription sont des séquences dites « enhancers ».

Il existe une multitude de promoteurs. Parmi ceux-ci, deux types peuvent être mis en lumière : les promoteurs avec TATA-box et ceux sans boîte TATA.

La première catégorie contient en général une boîte TATA, dont la séquence consensus est TATAAT ou TATATA. Elle est généralement localisée une trentaine de nucléotides en amont du site d'initiation de la transcription chez les eucaryotes (cf. figure 1). D'autres éléments de séquence, appelés éléments BRE (TFIIB Responsive Elements), sont fréquemment retrouvés à proximité de la TATA-box. Ce sont des éléments de réponse à la protéine TFIIB, situés en amont et en aval de la boîte TATA (Lagrange et al., 1998). Il existe également les éléments INR (pour éléments initiateurs), situés aux alentours ou à cheval du site d'initiation de la transcription et/ou une séquence DPE (Downstream Promoter Element), localisée vers +28 à +32 (cf. figure 1). Ces deux dernières séquences sont liées dans le sens où, un complexe protéique nommé TFIID se fixe aléatoirement dessus. Cependant, une seule mutation dans l'un ou l'autre de ces éléments peut entraîner la perte de la transcription dans le cas d'un promoteur sans TATA-box. Outre ces éléments « classiques » des promoteurs, il existe une multitude de séquences cis-régulatrices, situées en amont de la TATA-box, où se fixent de nombreux facteurs cellulaires. Il est également connu que des séquences situées parfois à plusieurs milliers de paires de base du promoteur (« enhancer ») sont la cible de protéines cellulaires (ou virales) qui vont activer ou réprimer les protéines impliquées dans la transcription cellulaire (Levine and Tjian, 2003; Levine et al., 2014).

L'ensemble de ces séquences promotrices est nécessaire aux mécanismes de transcription impliquant le complexe cellulaire TFIID. Ces éléments cis-régulateurs constituent la base des promoteurs majoritairement impliqués dans la transcription des ARNm (cf. figure 1). Toutefois, il est possible de retrouver ces différentes séquences en absence de

boîte TATA. En effet, les promoteurs contenant une TATA-box ne représentent qu'environ un tiers des promoteurs (Suzuki et al., 2001).

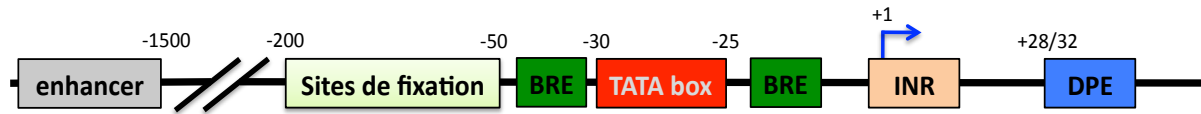


Figure 1 : Représentation schématique d'un promoteur classique de gène cellulaire.

Les séquences promotrices sont constituées de différents éléments, notamment d'une TATA-box, de séquences de réponses à TFIIB (BRE), mais aussi d'éléments INR et DPE qui permettent également le recrutement des facteurs de transcription cellulaire. En amont de ces éléments, des facteurs cellulaires se fixent à leurs sites de liaison respectifs. Des éléments situés à plusieurs milliers de paires de bases peuvent jouer le rôle d'« enhancers » ou « silencers », suivant la protéine liant cette séquence et modulant l'activité des facteurs de transcription impliqués.

De nombreuses analyses tendent à démontrer qu'une multitude d'autres promoteurs existent mais les mécanismes de régulations sont encore mal connus (Müller et al., 2007). En effet, le dernier exemple en date est l'identification de la séquence DTIE en aval du +1 d'initiation de la transcription, pouvant correspondre au promoteur d'un micro-ARN (miR-22) ainsi qu'à celui du gène codant pour la protéine p53. DTIE serait liée à deux séquences présentes une centaine de paires de bases en amont du +1. DTIE serait également présente dans 20% des promoteurs, et recruterait indirectement le complexe protéique TFIID sur cette séquence (Marbach-Bar et al., 2015). Ainsi, parmi tous les éléments cis-régulateurs étudiés à ce jour, il existe une liste de combinaisons entre séquences promotrices extrêmement variée car certains éléments cis-régulateurs peuvent être associés, couplés, et d'autres peuvent être indépendants.

La deuxième catégorie de promoteur est celui ne possédant pas de promoteur classique, c'est à dire sans TATA-INR-DPE. Ce sont des séquences généralement riches en nucléotides C et G : les îlots « CpG » vont servir de séquences initiatrices afin de réaliser la transcription. Ce type de séquence est souvent affilié à la transcription des gènes de ménage (présents dans tous les types cellulaires). Cette catégorie de promoteur est également associée à la présence de sites multiples d'initiation de transcription, engendrant une extrémité 5' variable pour des ARNm issus d'un même gène. Cependant, les mécanismes de recrutement des facteurs de transcription cellulaire sur ce type de promoteurs sont encore mal connus. Il est néanmoins admis que le facteur Sp1 est impliqué dans ce dispositif car de nombreux sites de fixation pour cette protéine sont retrouvés en amont de gènes sous le contrôle de promoteur

sans TATA-INR-DPE. (Kadonaga, 2012). D'autres protéines telles que Nrf1, E2F et ETS ont été retrouvées enrichies au niveau de ces îlots CpG (Deaton and Bird, 2011). Hormis ce système de régulation impliquant des facteurs de transcription, d'autres mécanismes, évoqués plus loin, peuvent influencer la transcription de ces promoteurs ainsi que ceux contenant des séquences cis-régulatrices.

B) L'assemblage du complexe général d'initiation de la transcription cellulaire.

L'ARN polymérase ne reconnaît pas directement les promoteurs. En revanche, elle est recrutée sur ces sites à l'aide de certains facteurs de transcription, qui eux, sont capables de se fixer spécifiquement aux séquences d'ADN. Dans le cas des promoteurs contenant une boîte TATA, un assemblage séquentiel de protéines est observé. Une protéine nommée TBP (pour TATA-Box Binding Protein) reconnaît la TATA-box et se fixe au niveau du petit sillon de l'ADN (Kim et al., 1993). Un ensemble de treize autres protéines, nommées TAF (pour TATA-Associated Factors, conservées dans de nombreuses espèces) (Tora, 2002), va s'associer dans un ordre précis à la protéine TBP, formant ainsi le complexe protéique TFIID (cf. figure 2) (Dubrovskaya et al., 1996; Goodrich et al., 1993; Hisatake et al., 1995; Hoffmann and Roeder, 1996; Jacq et al., 1994; Mengus et al., 1995; Takada et al., 1992; Tao et al., 1997; Weinzierl et al., 1993). Toutefois, il existe en plus des variants tissu spécifique de ces TAF (Müller et al., 2010).

Les protéines TAF sont également capables de fixer des séquences cis-régulatrices, notamment en l'absence de TATA-box. En effet, l'équipe de Tjian a démontré que certaines TAF (dont TAF1-TAF2) reconnaissent la séquence INR des promoteurs (Verrijzer et al., 1995). TFII-I et YY-1 sont également capables de se fixer sur ces séquences INR afin de recruter la TBP et l'ensemble de TFIID. Il a été montré que TAF6 et TAF9 se fixent sur les séquences DPE et conféraient la reconnaissance de cette région au complexe d'initiation de la transcription (Kadonaga, 2002; Smale and Kadonaga, 2003). La poursuite de l'assemblage du complexe de pré-initiation de la transcription (couramment appelé PIC) s'effectuerait, à priori, à l'aide d'un mécanisme identique aux promoteurs contenant une boîte TATA. Dans le cas de la synthèse des ARNm, ce sont les TAF spécifiques à la transcription des ARNm qui permettront l'assemblage avec l'ARN polymérase II.

Certaines protéines TAF peuvent également se lier aux facteurs de transcription. Par exemple, plusieurs domaines de la protéine TAF6 humaine sont nécessaires pour son interaction avec TFIIE et TFIIF (Transcription Factor II E / F) (Hisatake et al., 1995). Dans un autre contexte, l'homologue de TAF9 chez la drosophile interagit avec le facteur de transcription cellulaire TFIIB (Goodrich et al., 1993) alors que TAF5 se lie à TFIIF (Dubrovskaya et al., 1996). Cependant, le complexe TFIID, précurseur de l'initiation de la transcription, est assemblé de manière variable. En effet, TBP est capable de s'associer avec différentes classes de TAF, entraînant ainsi la formation de complexes impliquant l'ARN polymérase I, II ou III. Cette association de TAF est également dépendante de la composition du promoteur (TATA+ ou TATA-) (Comai et al., 1994; Hernandez, 1993; Peter Verrijzer and Tjian, 1996).

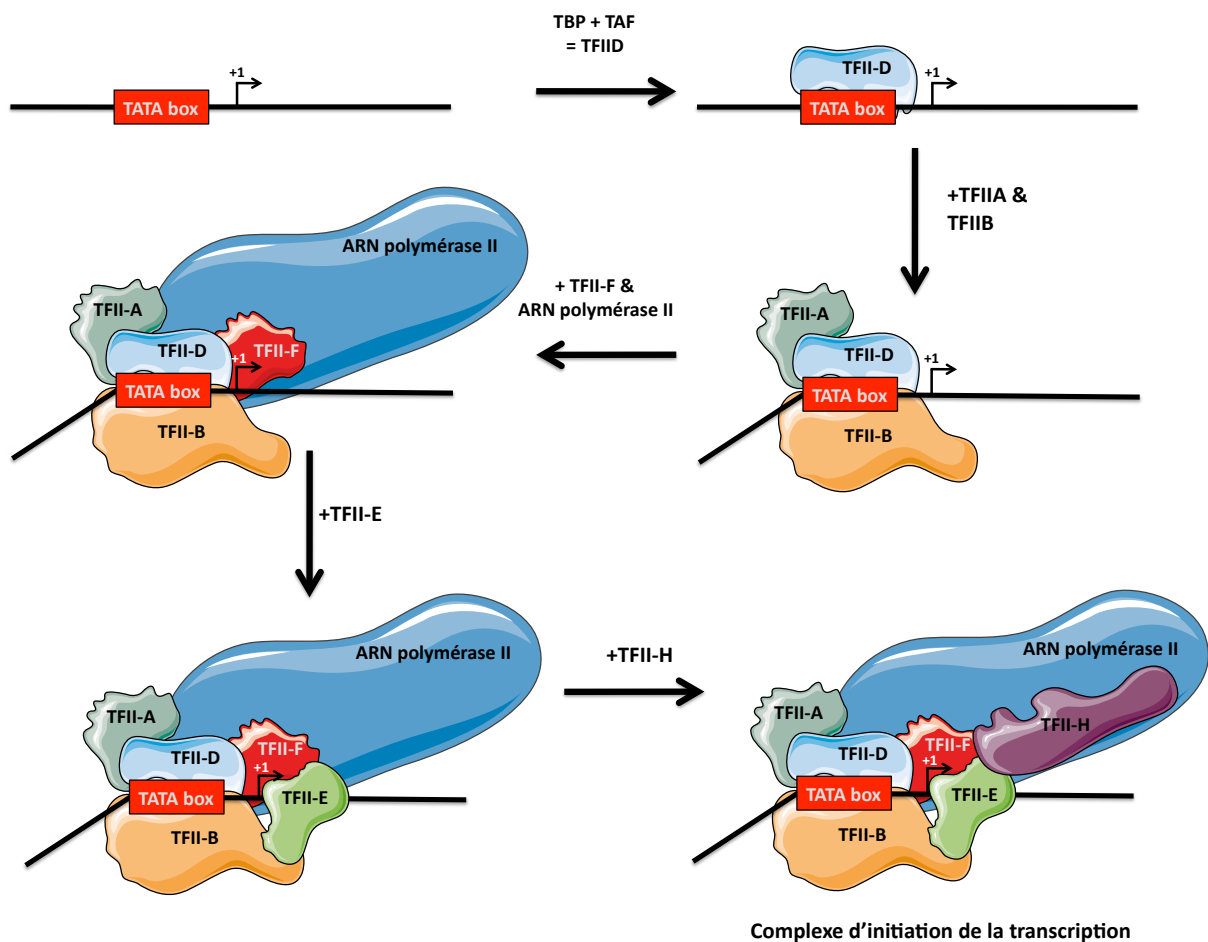


Figure 2: Représentation schématique de l'assemblage du complexe d'initiation de la transcription cellulaire.

Le complexe protéique TFIID est constitué de la protéine TBP (TATA-binding protein) ainsi que d'un ensemble de treize autres protéines associées nommées TAF (TATA-associated factors). TFIID est recruté sur les promoteurs, notamment sur la TATA-box ou d'autres séquences régulatrices. Le recrutement du complexe TFIID sur cette boîte est stabilisé par TFIIA puis TFIIB. TFIIF, associé à l'ARN polymérase, rejoint le

complexe à l'aide des TAF et de TFIIB. Le facteur TFIIF va alors permettre l'arrivée de TFIIE sur les promoteurs, renforçant ainsi la solidité du système. Le complexe de pré-initiation de la transcription est alors fermé, c'est-à-dire, inactif. L'enrôlement du facteur TFIIH va alors modifier la structure de l'ADN via son activité hélicase et phosphoryler le domaine Carboxyl-Terminal de l'ARN polymérase II afin d'initier la transcription des gènes cellulaires.

Suite à l'assemblage de ce complexe TFIID, le facteur de transcription TFIIA (Transcription Factor II A) vient s'ajouter au complexe afin de stabiliser sa liaison à l'ADN (Oelgeschläger et al., 1996). Le facteur TFIIB arrive ensuite pour se greffer au complexe et peut interagir avec TBP (Choy and Green, 1993). De plus, TFIIB est capable de se lier à des séquences d'ADN BRE. Cette fixation peut entraîner une inhibition ou une activation de la transcription, variable en fonction du contexte étudié. À ce moment, le facteur TFIIF, en association avec l'ARN pol II (composée de douze sous-unités), va se fixer aux protéines présentes sur la boîte TATA et plus particulièrement avec TFIIB. TFIIF stabilisera alors ce facteur avec l'ARN pol II formant ainsi un pont. S'ajoute ensuite le facteur TFIIE qui interagit avec TFIIF et l'ARN pol II réalisant ainsi une structure particulière. TFIIE recrute alors TFIIH (constituée de onze sous-unités dont deux hélicases et une kinase) et stimule son activité ATPase. Ceci induit l'initiation de la transcription et ouvre l'ADN à l'aide de son activité hélicase. Cet enchaînement forme alors un méga complexe de plus de deux Méga Daltons nommé Complexe de Pré Initiation de la transcription (PIC) (Luse, 2014) (cf. figure n°2).

Une fois le complexe de transcription associé et le Domaine Carboxy-Terminal de l'ARN pol II phosphorylé (par les kinases de TFIIH), l'étape d'élongation débute. La majorité des composants du PIC sont alors dissociés pour que l'ARN polymérase II puisse interagir avec des facteurs d'élongation et ainsi continuer la transcription jusqu'à la séquence de terminaison.

C) Une diversification des facteurs et de complexes de transcription.

TBP a été initialement décrite comme le facteur général de la transcription car il était admis que cette protéine était nécessaire à l'expression de tous les gènes. Néanmoins, depuis quelques années, une nouvelle dimension émerge. Il a été démontré qu'il existait outre la TBP « classique », trois paralogues du gène codant pour la protéine TBP : TRF1 (TBP-Related

Factor 1), les TBP2 (ou TRF3 pour TBP-Related Factor 3) et les TBP-like factors (TLF ou TRF2) (Zehavi et al., 2015).

TRF1 est retrouvée majoritairement au niveau du système nerveux et des gonades, mais uniquement chez certains insectes tels que la drosophile et les moustiques (D'Alessio et al., 2009). TRF1 est plutôt retrouvée associée à l'ARN polymérase III afin de permettre la transcription des ARNt. Néanmoins, TRF1 peut remplacer TBP et interagir avec TFIIA, TFIIB et la TATA-box pour induire la transcription associée à l'ARN polymérase II.

Les TBP2 (ou TRF3) (Persengiev et al., 2003) sont très similaires aux TBP car elles possèdent des domaines de liaison à TFIIA, TFIIB et aux séquences TATA-box. L'ensemble de ces caractéristiques leur permet ainsi d'initier la transcription avec l'ARN pol II. Elles sont par ailleurs fréquemment associées aux protéines TAF3 en lieu et place du complexe TFIID sur les promoteurs TATA (Reina and Hernandez, 2007). TRF3 joue un rôle primordial dans l'embryogénèse et dans l'hématopoïèse (Hart et al., 2007; Jallow et al., 2004).

À l'inverse, les TBP-like factors (TLF ou TRF2) sont celles qui divergent le plus structurellement des facteurs TBP canoniques. Elles peuvent interagir avec TFIIA, TFIIB, TAF6 et TAF9 mais sont incapables de se lier aux TATA-box (Müller et al., 2010). Il a été démontré récemment que la protéine TRF2 de la drosophile est capable de se fixer au niveau des séquences DPE décrites précédemment, sans pour autant savoir si d'autres facteurs cellulaires peuvent y être associés (Kedmi et al., 2014). TRF2 fixe également de très nombreux promoteurs sans TATA-box, parmi lesquels celui du gène codant pour l'histone H1 chez la drosophile, ou encore des gènes codant pour les protéines ribosomales (Duttke, 2015; Isogai et al., 2007).

Parmi les protéines impliquées dans l'initiation de la transcription, outre TBP, nombreuses sont celles ayant des paralogues dans différents types cellulaires. C'est le cas par exemple de TAF4 qui a un paralogue spécifique dans les ovaires humains ou TAF7 dans les testicules de souris (D'Alessio et al., 2009). De plus, suivant le promoteur, le type cellulaire, l'espèce, il se peut que les complexes d'initiation de la transcription soient différents. Par exemple, TFIID peut être substitué par d'autres complexes tels que TAF3 associée à TRF3, qui vont recruter les facteurs généraux de la transcription sur le promoteur de la myogénine, impliquée dans la myogénèse chez la souris (Deato and Tjian, 2007).

Ces données démontrent que TBP de même que l'intégralité des TAF ne sont pas forcément toutes requises pour la transcription par l'ARN pol II (D'Alessio et al., 2009; Müller et al., 2007). Ainsi, la complexité des promoteurs est accompagnée par une multitude de combinaisons de facteurs initiateurs de la transcription, notamment au niveau de TFIID, permettant une régulation tissu-spécifique.

D) L'analyse tridimensionnelle du PIC, une nouvelle ère de compréhension.

Dans le courant des années 1990, des études biochimiques ont démontré la présence du domaine Histone Fold (HF) dans les dimères TAF6-TAF9 et TAF4-TAF12, ainsi qu'une interaction entre ces TAF et les histones H2A, H2B, H3 et H4. D'autres travaux ont également mis en évidence l'existence du domaine HF chez les TAF3, 8, 10, 11 et 13. Par ailleurs, des analyses de cristallographie chez la drosophile avaient également démontré que les homologues humains de TAF6 et TAF9 formaient un hétérodimère, capable de se dimériser pour générer un hétérotétramère, similaire à celui des histones H3-H4 (Xie et al., 1996). Néanmoins, depuis quelques années, de nouvelles études structurales viennent contredire ces résultats et ouvrent ainsi une nouvelle ère dans la compréhension de l'assemblage du complexe humain d'initiation de la transcription présent sur les promoteurs.

En effet, en 2013, l'équipe de Berger a caractérisé une partie du complexe TFIID humain par cryo-electromicroscopie. Son groupe a ainsi démontré que TFIID était constitué d'un premier core composé de deux copies du trio TAF5-TAF6-TAF9. Ils ont également constaté que celui-ci était symétrique, et que sa structure était comparable à celle du core-TFIID lorsque deux copies du duo TAF4-TAF12 sont ajoutées à l'hexamère. La structure de cet ensemble de TAF est différente de celle des histones-like qui fut décrite précédemment dans la littérature car la formation histone-like de l'octamère TAF6-9 et TAF4-12 n'est pas retrouvée dans le complexe TFIID (cf. figure 3). Cette donnée infirme donc les résultats précédents mais confirme l'interactome obtenu il y a une vingtaine d'années. Dans ce même papier, les auteurs démontrent que le recrutement de l'hétérodimère TAF8-TAF10 va entraîner un réarrangement de la forme symétrique du complexe core TFIID en structure asymétrique et générer ainsi un complexe nommé 7TAF. Cette modification permettra ensuite

le recrutement en un exemplaire des autres TAF et de TBP, formant ainsi le complexe holo-TFIID sur les promoteurs (Bieniossek et al., 2013).

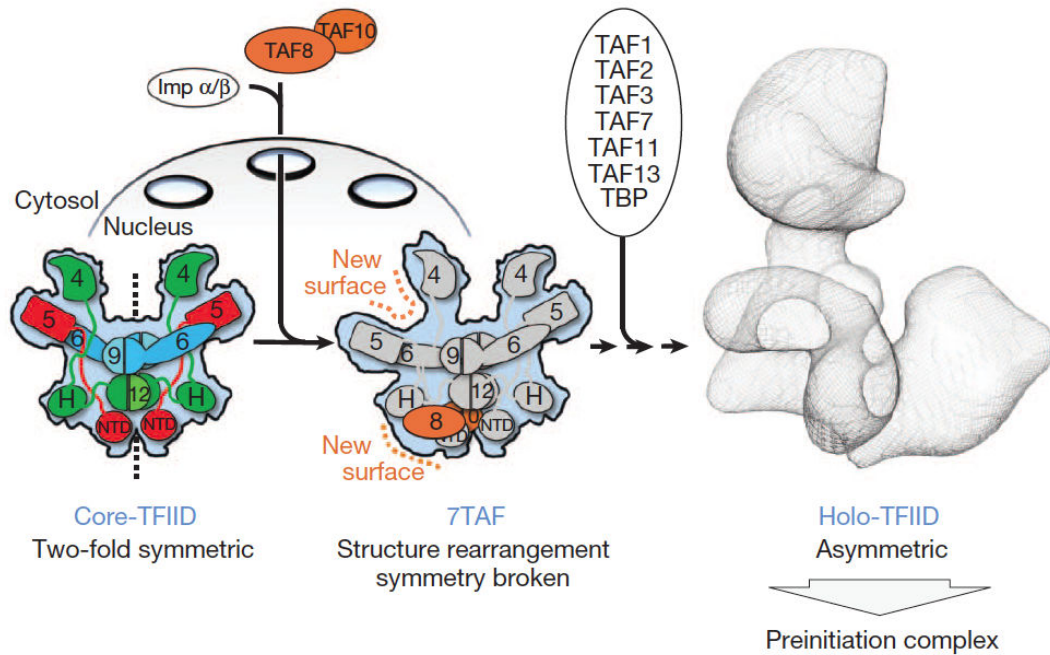


Figure 3 : Représentation schématique du complexe core TFIID, tiré de Bieniossek et al., 2013.

Le complexe core TFIID est composé des TAF4-TAF5-TAF6-TAF9-TAF12 en structure symétrique, différente de celle des histones décrites jusqu'à présent dans la littérature et la désorganisation asymétrique lors de l'ajout du dimère TAF8-10 au complexe core TFIID ; extrait de la publication de Bieniossek et al, 2013.

Plus récemment, la même équipe a constaté que seules les TAF2-TAF8-TAF10 pouvaient être retrouvées dans le cytoplasme pour s'assembler en un hétérotrimère stable contenant une copie unique de chaque TAF. Ces trois protéines possèdent également un domaine HF. Alors que TAF8 et TAF10 interagissent ensemble, seule TAF8 est capable de se lier à TAF2. Le recrutement dans le noyau de ces trois protéines se fait via le domaine NLS de TAF8 interagissant avec le complexe Importine α (Trowitzsch et al., 2015) (cf. figure 4).

En 2013, l'équipe de Nogales a tenté également par cryo-electromicroscopie d'observer la structure du complexe TFIID-TFIIA-ADN. Les auteurs de l'étude démontrent la présence en trois lobes de TFIID dont l'un est flexible en fonction de la présence de TFIIA et de l'ADN. Par ailleurs, la liaison de ce lobe avec TFIIA se fait à proximité de la boîte TATA. TFIIA aurait donc un rôle dans la stabilisation de TFIID au niveau de la TATA-box. Aussi, la structure des deux autres lobes est semblable à celle décrite par l'équipe de Berger, confirmant ainsi cette disposition. En revanche, il serait intéressant d'identifier les différentes

TAF présentes dans chacun de ces trois complexes, afin de mieux comprendre les rôles respectifs de chacune des ces protéines (Cianfrocco et al., 2013).

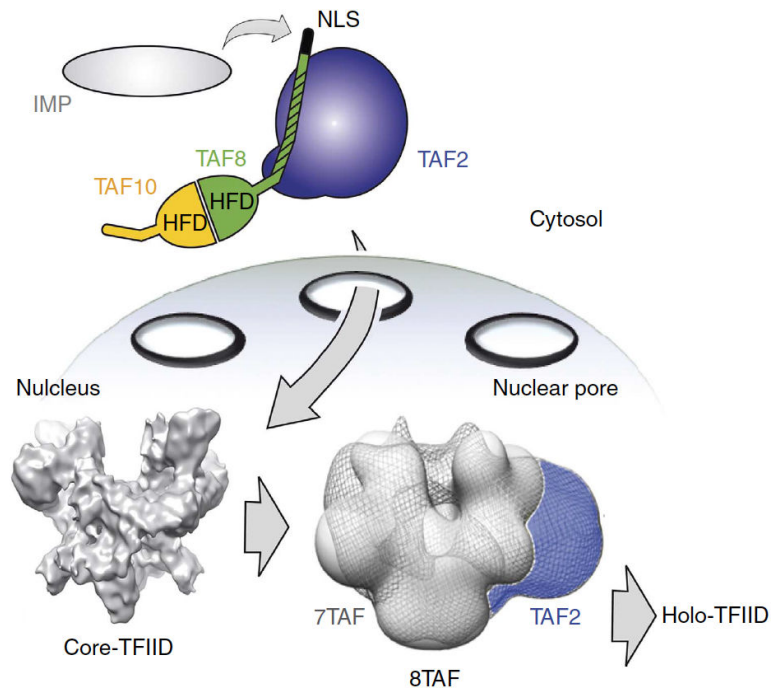


Figure 4 : Représentation schématique de l'assemblage du core TFIIID avec le trimère TAF2-TAF8-TAF10, tiré de Trowitzsch et al., 2015.

Le trimère TAF2-TAF8-TAF10 rentre dans le noyau avec le système l'Importine- α (IMP). Le complexe core TFIIID, composé de TAF4-TAF5-TAF6-TAF9-TAF12 va s'associer ensuite avec le trimère TAF2 - TAF8 - TAF10. Cette association va former une structure intermédiaire asymétrique avant de former le complexe TFIIID complet.

La même année, l'équipe de Nogales a, toujours par cryo-electromicroscopie, étudié la structure du complexe de pré-initiation de la transcription (PIC) au cours des différentes étapes d'assemblage (He et al., 2013). Les auteurs de cette étude l'ont réalisée à l'aide d'un promoteur contenant une boîte TATA entourée des éléments de réponse à TFIIB (BRE) ainsi qu'une séquence INR chevauchant le +1 de transcription. Leurs travaux ont permis de constater que TFIIF se fixait sur la séquence BRE en aval de la TATA-box avec TFIIB, entraînant une modification de la conformation du complexe, mais ont aussi confirmé que TFIIF stabilisait une partie de l'ADN dans le site catalytique de l'ARN pol II. Par ailleurs, un complexe formé de TBP-TFIIA-TFIIB-TFIIF-ARN pol II est à cheval entre la séquence BRE et INR afin d'ouvrir l'ADN au niveau du site catalytique de l'ARN pol II (à l'aide de TFIIF). Ils ont également constaté une stabilisation du PIC lors de l'ajout de TFIIE. Cette dernière interagit avec l'extrémité et le domaine de fixation à l'ADN (clamp) de l'ARN pol II ainsi qu'avec une sous-unité de TFIIF (cf. figure 5A). Par ailleurs, une partie de TFIIH interagit avec l'ARN pol II alors qu'une autre se situe sur l'ADN en aval de la TATA-box, qui en coopération avec TFIIB pourrait induire une ouverture de l'ADN et le séparer en deux brins distincts. Cependant, l'absence des TAF, notamment capables de reconnaître le site INR, pourrait infirmer cette donnée car l'encombrement stérique rendrait difficile ces liaisons. De

plus, le groupe a suivi l'ordre d'assemblage « connu » du complexe alors que les protéines pourraient intervenir précocement d'une autre manière.

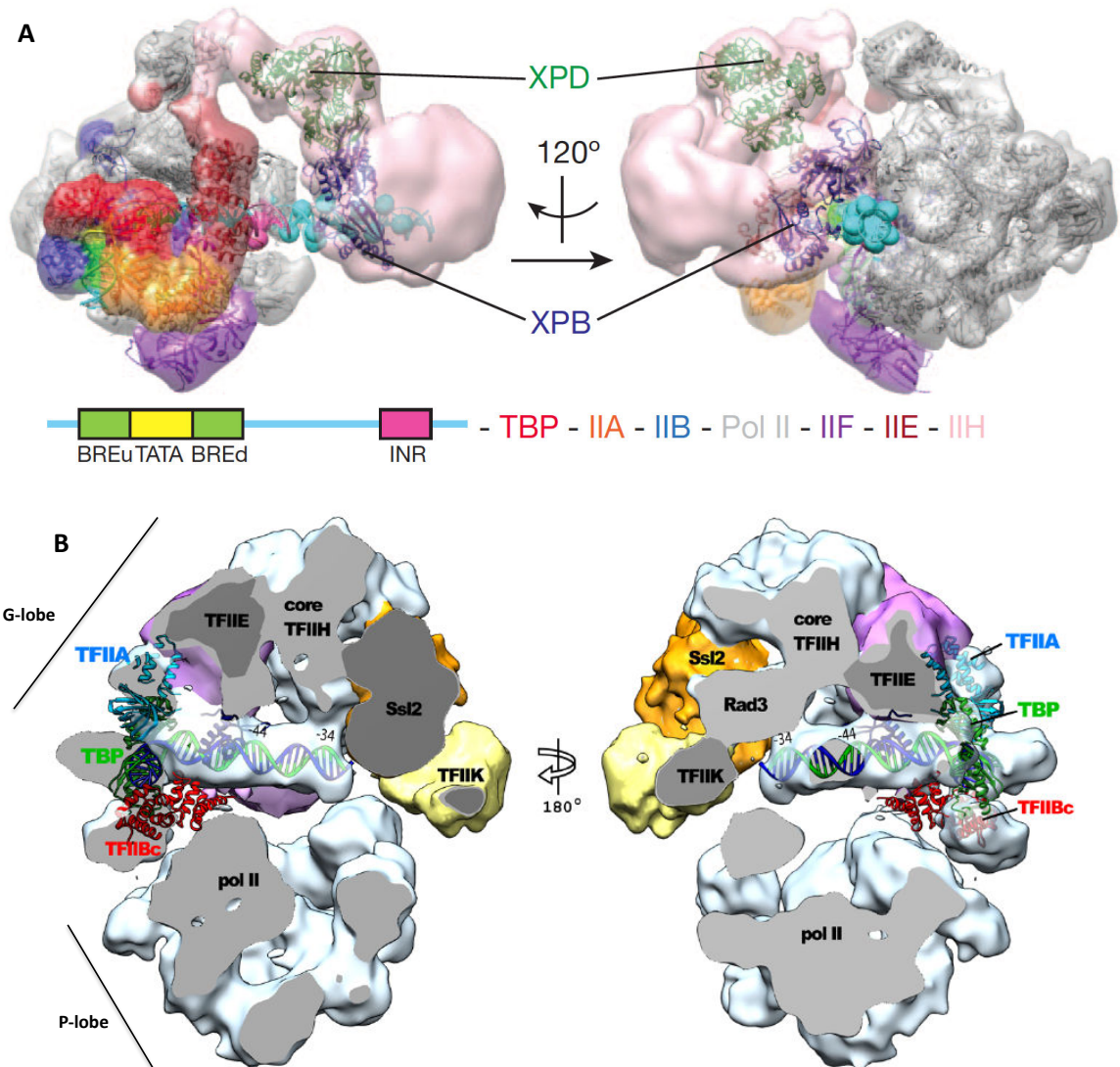


Figure 5 : Figures extraites des publications de He & al, 2013 (A) et Murakami & al (B) représentant les complexes de pré-initiation de la transcription par l'ARN polymérase II issus de cellules humaines (A) et de levures (B).

(A). Représentation tridimensionnelle du complexe cellulaire humain d'initiation de la transcription. Les couleurs correspondantes aux protéines impliquées sont référencées avec la sonde en dessous du modèle. Les sous-unités XPD et XPB sont les hélicases de la protéine TFIIH, situées à proximité de l'ARN polymérase II et de l'ADN. (B) Représentation tridimensionnelle du complexe de transcription par l'ARN polymérase II chez la levure en présence d'ADN. Sont présents les deux lobes mentionnés dans la publication : en haut le G-lobe contenant les facteurs de transcription et l'ARN polymérase constituant le P-lobe en bas (gris). A noter que Ssl2, Rad3 et TFIK sont trois des sous-unités de TFIIH.

Toujours en 2013, l'équipe de Kornberg a réussi à étudier, également par cryo-electromicroscopie, trente deux protéines constituant le complexe d'initiation de la

transcription de la levure (Murakami et al., 2013). Ils ont démontré que le complexe de pré-initiation de la transcription était dissocié en deux parties : le G-lobe (contenant les principaux facteurs de transcription) et le P-lobe (contenant l'ARN pol II). De plus, TFIIB, TFIID et une partie de TFIIIE sont présentes en amont de la boîte TATA alors qu'en aval, deux sous-unités de TFIIIE interagissant avec TFIIH y sont présentes ensemble. En outre, les auteurs confirment également la présence de la sous-unité kinase de TFIIH à proximité du CTD de l'ARN pol II (cf. figure 5B). Ces résultats valident une partie de ceux observés par l'équipe de Nogales. Néanmoins, des différences sont à souligner car l'une des structures est issue de cellules humaines et l'autre de levures. Par exemple, les deux lobes décrits chez la levure sont absents chez l'homme et les sites d'entrée de l'ADN dans le complexe de transcription sont différents (cf. figure 5).

Ces analyses structurales nous permettent ainsi de mieux comprendre les mécanismes mis en jeu lors de l'initiation de la transcription sur des promoteurs contenant des boîtes TATA. En effet, si la structure de ces complexes est différente, les modes de fonctionnement des PIC semblent globalement similaires d'une espèce à une autre.

E) La chromatine et son rôle dans la régulation du recrutement du PIC.

La chromatine est constituée d'octamères d'histones (nommés nucléosomes) autour desquels s'enroule l'ADN. Les nucléosomes sont constitués de deux copies de chaque histone : H2A, H2B, H3 et H4 (Thomas and Kornberg, 1975). Toutefois, il existe des histones variants dont H3.3/H2A.Z ou H2A.Z/H2A, possédant des propriétés physico-chimiques différentes des histones canoniques et peuvent être inclus dans les nucléosomes. Le nucléosome composé des histones variants est déplacé plus facilement par des facteurs trans-régulateurs, le rendant plus « mobile ». Ce type de nucléosomes est souvent retrouvé au niveau des régions promotrices riches en îlots CpG afin de faciliter la dynamique de recrutement de facteurs cellulaires impliqués dans la transcription. Par ailleurs, toutes les histones subissent des modifications post-traductionnelles sur leurs queues N-terminales telles que les acétylations, méthylation, ubiquitinylation. Elles ont pour rôle de réguler le recrutement du complexe de transcription (Müller and Tora, 2014).

Par exemple, il a été déjà démontré que la protéine TAF3 du complexe TFIID est capable de se lier sur le H3K4me3 (tri-méthylation de la lysine 4 de l'histone 3). Cette

interaction est même renforcée par l'acétylation des lysines 9 et 14 de l'histone H3 (H3K9Ac et H3K14Ac), alors qu'elle est inhibée par la diméthylation de l'arginine 2 de l'histone H3 (H3R2me2) (Vermeulen et al., 2007). Le rôle de cette interaction entre H3K4me3 et TAF3/TFIID consiste à faciliter la formation du PIC sur les promoteurs et induire la transcription (Lauberth et al., 2013). Malgré l'absence de quelque marque répressive de transcription, le site H3K4me3 n'est pas suffisant pour recruter le complexe ARN polymérase II.

Une étude plus ancienne a démontré que la protéine TAF1 du complexe TFIID est capable de lier l'histone H4 di-acétylé (marque active de la transcription) (Jacobson et al., 2000). Ces données suggèrent que le recrutement de TFIID et de l'ensemble du complexe PIC sur les promoteurs peut se faire par l'intermédiaire d'autres facteurs, notamment par les protéines de remodelage de la chromatine ou bien par les TAF capables de lier les nucléosomes qui pourraient stabiliser le complexe TFIID sur les promoteurs.

L'ensemble de ces éléments démontre ainsi que l'assemblage du complexe d'initiation de la transcription est un mécanisme très complexe. En effet, celui-ci est extrêmement dynamique au cours du temps et dépend de plusieurs éléments : des séquences promotrices et cis-régulatrices, des facteurs de transcription impliqués, de leurs modifications post-traductionnelles, du type cellulaire, mais aussi des marqueurs épigénétiques et de leur régulation.

II) Les virus à ADN.

Les virus, parasites intracellulaires obligatoires sont présents partout dans notre environnement. De tailles variables pouvant aller de quelques nanomètres à plusieurs micromètres, ils sont capables d'infecter de très nombreux organismes: eucaryotes, bactéries et archées. On estime que la présence de virus dans notre monde est quantitativement cent fois supérieure à celle des bactéries. Quant à savoir si ces organismes sont vivants ou non, il est possible de faire la différence entre les virions, libres dans la nature où ils sont incapables de se répliquer (donc potentiellement non vivants) et les virus à proprement parler, présents dans

leurs cellules hôtes où se déroule leur cycle viral (et donc potentiellement vivants). Cependant, la découverte en 2008 (La Scola et al., 2008) d'un virus nommé Sputnik capable d'infecter un autre virus, certes géant (Mimivirus découvert en 2003) pause quelques questions quant à la définition même de ce qu'est un virus.

De manière plus générale, le support de l'information génétique de ces micro-organismes peut être de l'ADN (simple ou double brin) ou de l'ARN (simple brin (positif ou négatif) ou double brin), ce qui détermine leurs classes (classification de Baltimore). Alors que de très nombreuses familles de virus à ARN sont répertoriées, nous nous focaliserons uniquement sur les virus à ADN double brin. Toutefois, nous n'évoquerons pas les poxvirus dont les étapes de réplication sont exclusivement cytoplasmiques, ni les virus géants à ADN tels que Mimivirus, Mamavirus, Pandoravirus, Pithovirus, et autres récemment découverts. Nous détaillerons plus précisément la régulation de l'expression des gènes viraux tardifs pour les comparer avec ceux du virus d'Epstein-Barr, sujet central de cette thèse.

A) Les Caudovirales (bactériophages).

Les bactériophages sont, comme leur nom l'indique, des virus infectant les bactéries. Ces virus sont classés en différentes familles compte tenu de leur très grande diversité. En effet, on estime qu'il existe plus de 5000 espèces de bactériophages. Ils sont de tailles variables (allant de quelques nanomètres à plus de 200 nanomètres) mais ont un support de l'information génétique basé soit sur l'ADN (pour la plupart) soit sur l'ARN. La longueur est également hétérogène (de quelques milliers de bases à plus de 700 000 paires de bases). Ces virus ont deux cycles :

- le cycle lysogénique, dans lequel le virus infecte la bactérie et s'intègre dans le génome bactérien sous forme de prophage.

- le cycle lytique, durant lequel le cycle viral productif est activé. La machinerie transcriptionnelle de la bactérie nécessaire à la synthèse de ses ARN et par extension pour se répliquer est détournée au profit du virus.

Afin de faciliter la compréhension des mécanismes régulant l'expression des gènes viraux des bactériophages, nous allons nous focaliser sur le phage T4, qui est probablement le mieux étudié parmi les 5000 espèces de bactériophages. Ce virus de bactérie est également un

des plus étudiés de part son intérêt en biologie moléculaire. Le bactériophage T4 est un virus à ADN double brin d'environ 169 000 paires de bases, linéaire dans le virion et circulaire une fois dans la bactérie. Son ADN a la particularité d'être composé à plus de 65% de bases A/T alors que son hôte (*Escherichia Coli* ou *E. Coli*) possède un génome composé à 50% de bases A/T. Cette différence lui permettrait de résister à la dégradation par les endonucléases bactériennes.

Le virus T4, comme la plupart des virus de sa famille, a une expression de ses gènes séquentielle qui est divisée en trois catégories : précoce, intermédiaire et tardive. Quelle que soit la classe de protéine étudiée, toutes ont un promoteur spécifique, dont la transcription est assurée par l'ARN polymérase bactérienne (RNAP), modifiée sur son extrémité Carboxy-Terminale (CTD) par certaines protéines virales.

1) Les promoteurs des gènes viraux précoces.

L'expression des gènes précoces est dépendante d'un type de promoteur bien défini : une première séquence conservée est retrouvée entre les positions -6 et -17 en amont du site d'initiation de la transcription (TSS) (TGTGGTA T/C T/A ATA) (Wilkens and Rüger, 1996) (cf. figure 6). Celle-ci conserve de nombreuses homologies avec certains promoteurs bactériens notamment au niveau de la région située entre -7 et -12 en amont du TSS. Une deuxième région est également hautement conservée entre les positions -29 et -38 en amont du site d'initiation de la transcription : GTTTAC. De plus, juste avant cette séquence, une région riche en A est présente entre les positions -40 et -52 avant le TSS. Cette dernière zone est aussi nommée élément UP et est retrouvée sur certains promoteurs bactériens (cf. figure 6). Cette séquence est connue pour augmenter la transcription de gènes hôtes et viraux précoces en recrutant une des sous-unités de la RNAP bactérienne.

Promoteur de gène précoce :

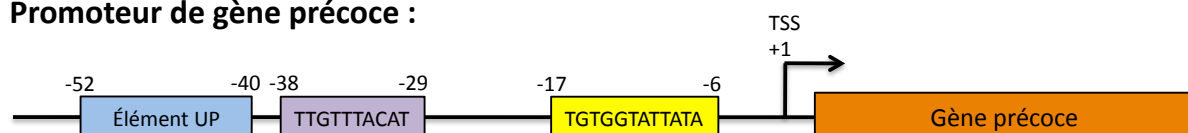


Figure 6 : Représentation schématique de la structure prototype des promoteurs précoces du bactériophage T4.

Les gènes précoces sont régulés par trois séquences majoritairement retrouvées au niveau de leurs promoteurs. L'élément UP, qui est une région riche en A, est présent sur de très nombreux promoteurs précoces viraux mais

aussi sur ceux des bactéries. Une autre région centrée vers -35 avant le TSS diffère de celle des promoteurs de l'hôte. Enfin, la troisième séquence, située entre -17 et -6 sur le promoteur viral, est également retrouvée sur ceux de la bactérie.

Cependant, quelques promoteurs précoces, différents de ceux évoqués précédemment, ressemblent largement à certains promoteurs d'*E. Coli* (constitués de deux séquences positionnées vers -35 et -10 en amont du TSS et reconnues par la protéine bactérienne sigma σ). Cependant, la structure de ce type de promoteur chez le bactériophage T4 n'est pas favorable à la transcription des gènes régulés par celui-ci. En effet, au début de l'infection, une protéine virale va entrer dans la bactérie en même temps que l'ADN viral et va modifier le fonctionnement de la RNAP de son hôte. Cette protéine, Alt, est une mono-ADP-ribosyltransférase. Elle va déposer efficacement un ADP-ribosyl (Adénosine Di-Phosphate) sur le CTD d'une des sous-unités de la RNAP bactérienne. Cette modification post-traductionnelle va engendrer une inhibition de la transcription des gènes de l'hôte et par conséquent celle des gènes viraux précoces sous contrôle du promoteur similaire à celui de la bactérie. À l'inverse, cet ajout de ribosyl va engendrer une augmentation de la transcription des autres gènes viraux précoces (Koch et al., 1995). De plus, la séquence nucléotidique située en position -33 et la région riche en nucléotides A contribuent à augmenter l'interaction avec la RNAP contenant l'ADP-ribosyl.

Peu de temps après l'infection, deux nouvelles protéines, ModA et ModB, vont être exprimées et vont modifier la polymérase bactérienne, entraînant la diminution voire l'inhibition de la transcription des promoteurs contenant les éléments UP : c'est le cas de la majorité des gènes précoces mais aussi bactériens. Ce changement va permettre le début de l'expression des gènes intermédiaires une minute après l'infection (Tiemann et al., 2004).

2) Les promoteurs des gènes intermédiaires.

Les gènes intermédiaires sont sous le contrôle d'un promoteur légèrement différent des gènes précoces avec deux sites identifiés. En effet, la séquence située entre -7 et -12 est globalement conservée (TATAAAT pour les intermédiaires et TATTATA pour les précoces). Cette similarité peut s'étendre jusqu'au -16 avant le TSS (cf. figure 7). En revanche, les régions situées entre -31 et -52 disparaissent au profit d'une séquence centrée entre -27 et -32

avant le TSS (TTGCTT), nommée Mot-box comme le nom de la protéine, MotA, reconnaissant cette boîte (cf. figure 7).

Promoteur de gène intermédiaire :

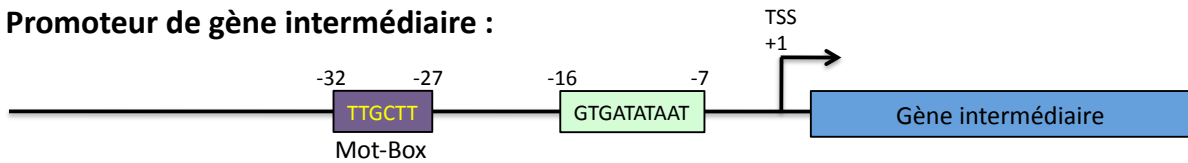


Figure 7 : Représentation schématique de la structure de promoteur de gènes intermédiaires du bactériophage T4.

Les promoteurs intermédiaires sont constitués de deux séquences bien distinctes. Une première retrouvée au niveau de la région -30 avant le TSS est un élément de réponse à la protéine précoce MotA qui recrute l'ARN polymérase bactérienne pour transcrire les gènes intermédiaires. La deuxième séquence est très proche de la boîte TATA retrouvée dans de nombreuses espèces dont les promoteurs bactériens.

En effet, MotA est une protéine capable de lier l'ADN au niveau de cette séquence pour faciliter la reconnaissance du site par l'ARN polymérase bactérienne et donc l'activation de la transcription (Schmidt and Kreuzer, 1992).

Cependant, une autre protéine est également impliquée dans l'expression des gènes intermédiaires : la protéine AsiA. Cette dernière va former un hétérodimère avec la RNAP de l'hôte, et ainsi modifier la conformation de cette dernière. Ce phénomène va bloquer l'interaction avec les promoteurs bactériens et inhiber la transcription des gènes précoces. L'interaction entre AsiA et la RNAP bactérienne va favoriser le lien entre MotA et la RNAP dans le but d'activer la transcription des gènes intermédiaires au détriment des précoces (Ouhammouch et al., 1995). Suite à l'expression des gènes intermédiaires, l'ADN viral sera répliqué par l'ADN polymérase virale gp43.

3) Les promoteurs des gènes viraux tardifs et leur régulation spécifique.

S'en suit alors l'expression des gènes viraux tardifs, qui codent pour des protéines structurales. À cette étape, l'expression des gènes viraux précoces va diminuer jusqu'à s'éteindre au profit des gènes tardifs. Les gènes tardifs ont également des promoteurs totalement différents de ceux des gènes viraux précoces et intermédiaires. En effet, seule une boîte TATAAATA est conservée entre les positions -13 et -6 en amont du TSS. En revanche, une petite région de cinq nucléotides A/T est retrouvée à cheval sur le TSS (-3 à +1) (cf. figure 8) (Miller et al., 2003).

Promoteur de gène tardif :

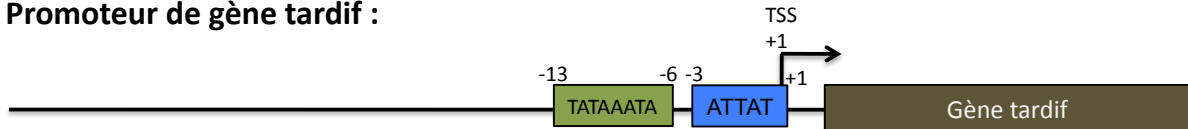


Figure 8 : Représentation schématique de la structure de promoteur tardif du bactériophage T4.

Les promoteurs des gènes tardifs sont constitués d'un seul et unique élément cis-régulateur identifié à ce jour. Il s'agit de la boîte TATA, proche de la séquence canonique sur laquelle sont recrutés des facteurs viraux.

Il est connu que la cassure de l'ADN matrice simple brin à proximité ou à distance du promoteur est essentielle à la transcription. Cette rupture joue le rôle d'« enhancer » dès l'instant où cela n'affecte pas le brin transcrit (Geiduschek and Kassavetis, 2010). De plus, trois protéines virales vont jouer un rôle majeur dans l'expression des gènes viraux tardifs : les protéines précoces gp45, gp55 et gp33. En effet, il est démontré que la protéine gp55 (homologue du facteur σ (sigma) bactérien qui recrute l'ARN polymérase bactérienne) initie la transcription des gènes tardifs en reconnaissant les séquences promotrices tardives (Kassavetis et al., 1983). Pourtant, cette protéine seule n'est pas suffisante à l'expression des gènes tardifs. Effectivement, la protéine gp33, essentielle dans cette étape, interagit avec l'ARN pol bactérienne et agit comme co-activateur de la transcription des gènes tardifs en faisant le pont entre la protéine gp55 et la protéine gp45 pour former un trimère capable de se déplacer sur l'ADN. L'absence d'une seule de ces protéines empêche la synthèse des ARNm viraux tardifs.

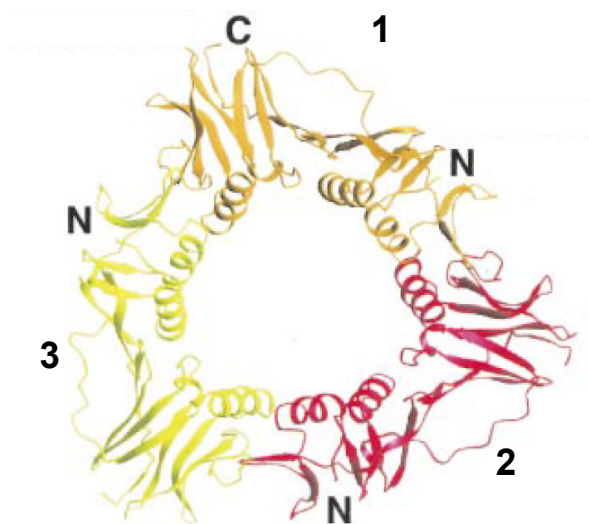


Figure 9 : Structure tridimensionnelle de la gp45, tiré de Moarefi et al., 2000.

La gp45 du bactériophage T4 s'assemble autour de l'ADN en trimère (numéroté de 1 à 3) afin de le faire passer au sein de son anneau. Le N et le C correspondent aux extrémités N-terminale et C-terminale des monomères.

La protéine gp45 est essentielle à l'expression des gènes tardifs car elle se lie à l'ADN et ouvre les promoteurs. Elle fait également le lien avec la réplication de l'ADN viral en tant que facteur de processivité du complexe de réplication viral. Elle est présente sous forme de

trimère afin de former un anneau autour de l'ADN (cf. figure 9) (Fu et al., 1996; Moarefi et al., 2000).

Cette protéine possède deux sites de liaison à l'ADN. Le premier est le site de cassure de l'ADN où le dimère protéique viral gp44 et gp62 va la recruter (qui sont des protéines de chargement de gp45 sur l'ADN). Elle sera alors stabilisée sur ce site par la présence de gp55 et gp33, en amont ou en aval du promoteur tardif (Kolesky et al., 2002). La protéine gp32 (protéine de liaison à l'ADN simple brin) n'est pas essentielle au chargement de gp45 sur l'ADN cassé. Toutefois, sa présence facilite le recrutement de gp45 sur les jonctions amorces-matrices (de la réplication) qui représentent le deuxième site de liaison à l'ADN de gp45. Ces sites particuliers favorisent l'interaction entre gp45 et l'ADN polymérase mais doivent être situés en aval du promoteur. Ces données suggèrent que gp45 pourrait faire le lien entre la T4 DNA polymérase et le complexe gp55-gp33-ARN pol bactérienne. Ces trois protéines virales, associées à l'ARN polymérase bactérienne forment alors un complexe d'initiation de la transcription des gènes viraux tardifs, dont la structure a été réalisée avec précision (cf. figure 10) (Twist et al., 2011). Cet ensemble, extrêmement dynamique, est toujours dépendant d'une ARN polymérase bactérienne au CTD subissant des modifications post-traductionnelles (ADP-ribosyl) induites par des protéines virales (Nechaev and Geiduschek, 2008).

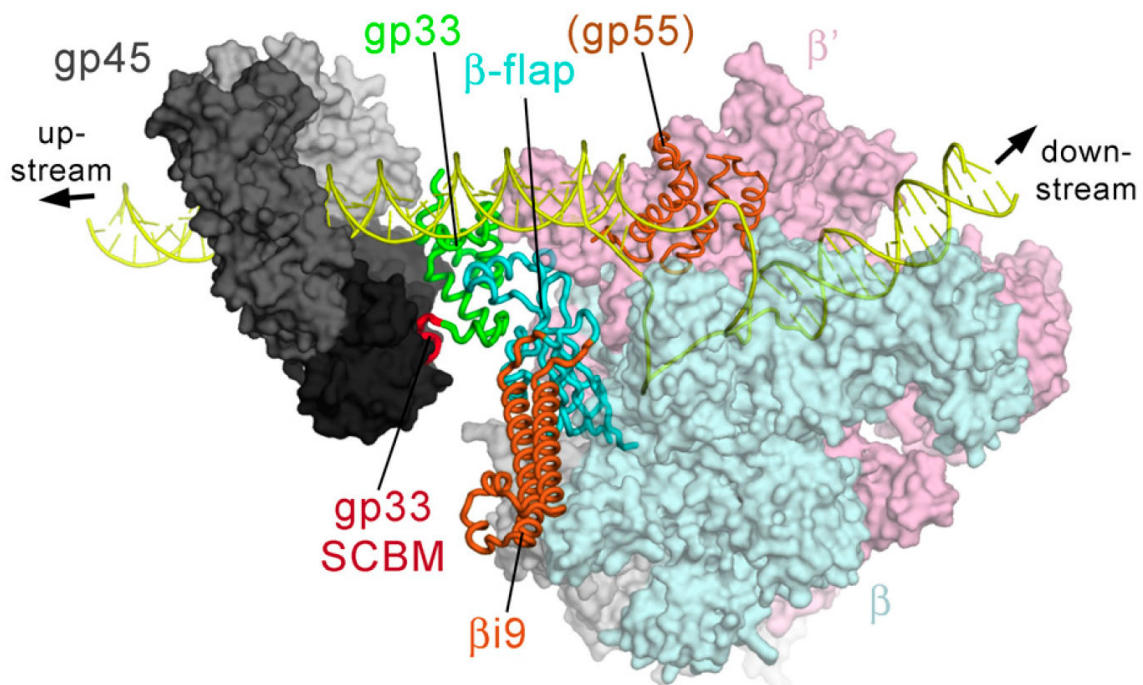


Figure 10 : Modèle structurel du complexe d'initiation de la transcription des gènes tardifs, tiré de Twist et al., 2011.

L'ARN polymérase bactérienne (les différents éléments β) est recrutée par la gp33 sur le promoteur des gènes

tardifs. Le domaine de fixation SCBM de gp33 interagit avec la gp45. Cette interaction permet à gp33 de faire le lien entre l'ARN pol bactérienne et la protéine gp45. La gp55 est retrouvée sur le promoteur tardif du bactériophage T4 est interagit avec la gp33. L'ADN en jaune est retrouvé dans l'anneau de gp45 et au niveau de la RNA pol bactérienne qui va transcrire de nouveaux ARN.

D'autres protéines virales, telles que RpbA ou Srh (qui se lient à l'ARN pol bactérienne) mais aussi DsbA ou Srd sont également importantes pour l'expression des gènes viraux tardifs sans pour autant être essentielles (Miller et al., 2003). Par exemple, il a été démontré que la protéine RpbA a un rôle important dans la transcription des gènes tardifs en se liant à l'ARN pol bactérienne en favorisant la transcription des promoteurs tardifs (Kolesky et al., 1999).

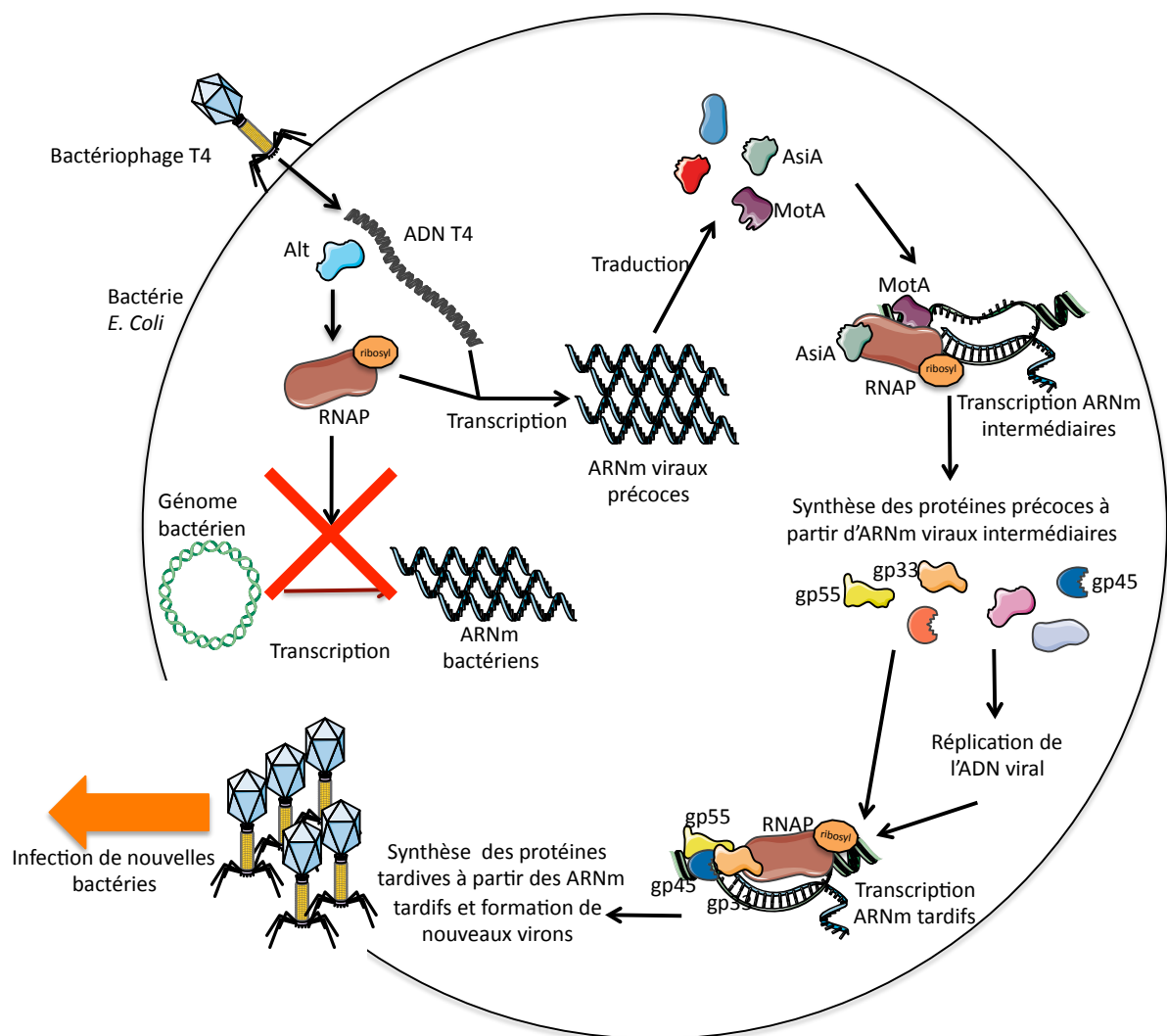


Figure 11: Représentation schématique du cycle viral du bactériophage T4 lors de l'infection de la bactérie *Escherichia Coli*.

Lorsque le virus s'attache à la bactérie, il va injecter son génome ainsi qu'une protéine virale Alt, encapsidée dans la particule virale. Ce facteur va alors modifier la structure tridimensionnelle de l'ARN polymérase

bactérienne (RNAP), entraînant l'extinction de la transcription de l'hôte tout en favorisant celle du phage. Ainsi, la RNAP va induire la synthèse des ARN viraux précoces qui codent pour des protéines dont certaines sont impliquées dans l'expression des gènes viraux intermédiaires. Parmi celles-ci, MotA et AsiA sont connues pour respectivement reconnaître le promoteur des gènes viraux précoces et recruter la RNAP sur les promoteurs où MotA est fixée. Ceci permettra la synthèse des ARNm viraux intermédiaires qui codent pour de nombreuses protéines dont certaines sont impliquées dans la réplication de l'ADN viral. Suite à cette étape, les gènes viraux tardifs du T4 seront transcrits par la RNAP bactérienne qui sera recrutée par gp33 sur les promoteurs tardifs où est positionnée gp55 et gp45. À partir des ARNm viraux tardifs traduits en protéines, il y aura formation de nouvelles particules virales qui iront infecter de nouvelles bactéries.

Ainsi, le bactériophage T4 a la particularité de coder pour des protéines qui vont recruter la machinerie transcriptionnelle bactérienne et inhiber la transcription de l'hôte afin de transcrire ses propres ARNm de manière séquentielle. Cette régulation s'effectue grâce à l'implication des protéines virales exprimées au cours du cycle productif. Ces dernières auront un rôle primordial dans le contrôle et la régulation des étapes qui suivront lors de l'expression des gènes viraux (cf. figure 11).

Les bactériophages étant très nombreux, les mécanismes de régulation de l'expression des gènes viraux sont par conséquent très diversifiés. En effet, certains ont des systèmes similaires dans lesquels le virus détourne la machinerie transcriptionnelle de la bactérie. Par exemple, le phage Lambda λ régule l'expression de ses gènes précoces à l'aide de la protéine virale N. Cette dernière va se lier à des séquences cis-régulatrices virales et interagir avec certaines protéines bactériennes, induisant un changement de conformation de la RNAP et son recrutement sur les promoteurs précoces. Ce complexe RNAP et protéine N est alors résistant aux séquences de terminaison du virus pour produire plus de transcrits viraux. Un mécanisme similaire est observé pour le contrôle de l'expression des gènes tardifs à l'aide de la protéine virale Q, conférant une résistance à la RNAP à ce même mécanisme de terminaison de transcription (Court et al., 2007).

D'autres phages ont opté pour un système totalement différent. En effet, certains codent pour leur propre ARN polymérase. C'est par exemple le cas pour les bactériophages T3 et T7. Le cycle lytique de ces virus débute par la transcription des gènes précoces à l'aide de la RNAP bactérienne. S'en suit alors la transcription des gènes intermédiaires puis des gènes tardifs à l'aide de la seule RNAP virale (constituée d'une seule sous-unité), qui

reconnaît les séquences spécifiques de ces deux classes de promoteurs viraux. De plus, une protéine intermédiaire, gp2 va s'associer à la RNAP bactérienne afin d'inhiber la machinerie transcriptionnelle de la bactérie (Yang et al., 2014).

Ainsi, la régulation de la transcription des gènes de bactériophages est réalisée à l'aide de différentes stratégies (variables d'un virus à l'autre) et est fréquemment couplée à une inhibition de la transcription des gènes de leurs hôtes.

B) Les *Baculoviridae* (baculovirus).

Les baculovirus font partie de la famille de *Baculoviridae*. Avec plus de six cent espèces recensées, ils sont divisés en quatre genres : *Alphabaculovirus*, *Betabaculovirus*, *Gammabaculovirus* et *Deltabaculovirus*, basés sur la phylogénie, et sont parfois même divisés en espèces. Ce sont des virus qui infectent exclusivement les insectes. Ce sont de très gros virus enveloppés avec un ADN double brin, circulaire, de taille variant entre 80000 et 200000 paires de bases, contenu dans une nucléocapside en forme de tige d'un diamètre d'environ 30 à 70 nm et d'une longueur d'environ 200 à 400 nm (cf. figure 12).

Ces virus codent en moyenne plus de 180 protéines (Jehle et al., 2006). Cette catégorie de virus à ADN est étudiée depuis plusieurs décennies de par leur utilisation comme pesticide non pathogène pour l'homme, mais aussi pour la production de protéines recombinantes en biologie moléculaire.

Comme pour les bactériophages, se sont des virus qui ont une expression séquentielle de leurs gènes, divisée en 4 catégories : immédiate précoce, précoce, tardive et très tardive. Parmi toutes les familles de virus à ADN infectant les eucaryotes, c'est la seule qui utilise deux ARN polymérase différentes: d'une part, ils détournent l'ARN polymérase II cellulaire pour transcrire les gènes viraux immédiats précoces et précoces et d'autre part ils codent pour une ARN polymérase virale (résistante à l' α -amanitine) qui transcrit les gènes viraux tardifs et très tardifs (Fuchs et al., 1983; Huh and Weaver, 1990). Comme pour les bactériophages, la quantité d'espèces de baculovirus est importante. Nous nous focaliserons majoritairement sur le plus étudié d'entre eux: le AcMNPV (*Autographa californica NucleoPolyhedrovirus*), membres des *Alphabaculovirus*.

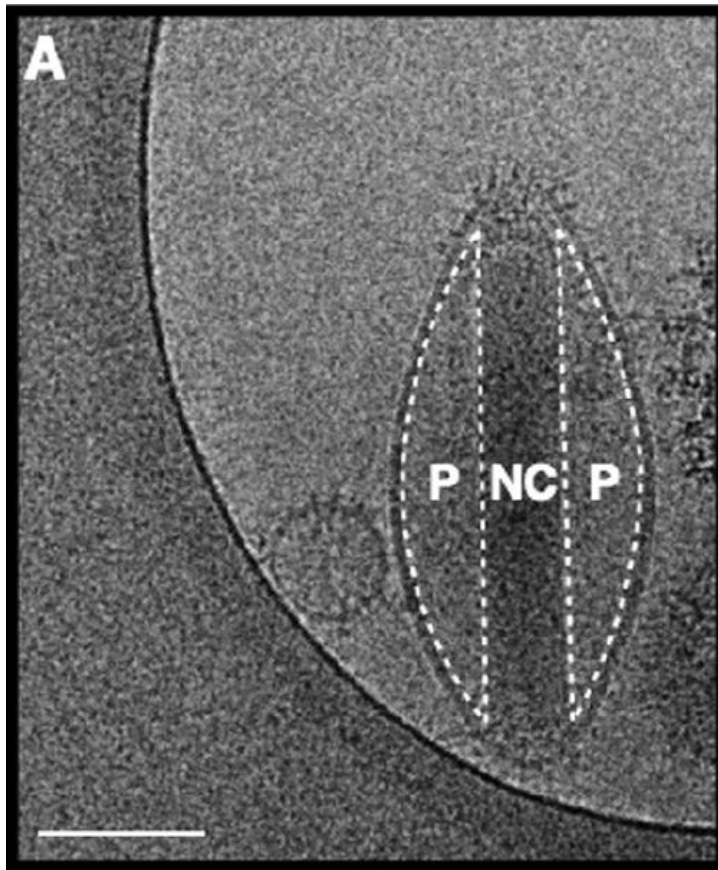


Figure 12 : Photographie d'une particule virale de baculovirus AcMNPV par cryo-electromicroscopie, extrait de Wang et al., 2016.

Le virus, de morphologie allongée contient deux poches latérales (P) et une longue nucléocapside en forme de tige (NC), occupant la majorité de la particule virale. Le virus est enveloppé et possède des petits filaments en apex du virion. Cette structure est assez conservée entre virus de cette sous-famille.

1) L'expression des gènes viraux précoces.

Lors de l'infection par baculovirus, il a été constaté que de nombreuses protéines cellulaires étaient beaucoup moins exprimées en raison d'une diminution de la quantité de leur ARN (Salem et al., 2011). Néanmoins, les protéines à activité enzymatique, notamment les kinases sont quant à elles, légèrement surexprimées. Le cycle viral productif débute par l'expression des gènes immédiats précoces dont l'un d'eux code pour la protéine virale IE1 (Immediate Early 1 ou Ac147). C'est une protéine impliquée dans l'expression des gènes viraux précoces et qui augmente sa propre expression. Par épissage alternatif, le transcrit IE1 peut donner naissance à une autre protéine IE-0, susceptible de former un hétérodimère avec IE1 (Sokal et al., 2014). Les deux autres protéines immédiates précoces impliquées dans l'expression des gènes viraux précoces sont les protéines IE-2 (Ac151) et PE-38 (Ac153). Contrairement à IE-1, ces deux protéines n'ont pas de sites de fixation à l'ADN, suggérant que des facteurs intermédiaires sont nécessaires à leur fonctionnement de trans-activation. Curieusement, il a été démontré que seuls ces gènes immédiats précoces pouvaient être transcrits par la machinerie cellulaire, alors que les autres gènes viraux semblent nécessiter la

présence de facteurs viraux pour être exprimés (Lin and Jarvis, 2013). Ceci est d'autant plus surprenant car les structures des promoteurs des gènes viraux immédiats précoces et précoces sont similaires.

Les promoteurs des gènes viraux précoces sont plutôt conservés entre eux et sont similaires à ceux retrouvés chez les eucaryotes. En effet, deux séquences distinctes ont été mise en lumière : une boîte canonique TATA située aux alentours de -30 avant le TSS et une seconde séquence consensus CAGT localisée au niveau du TSS. Elle est semblable aux séquences INR retrouvées chez les eucaryotes et est considérée comme telle dans le cas du baculovirus (Pullen and Friesen, 1995a) (cf. figure 13). Néanmoins, cette dernière séquence est variable en fonction des promoteurs du virus. Elle peut également avoir deux autres formes : CATT ou CAAT. Dans tous les cas, cet élément INR permet l'initiation de la transcription des gènes viraux précoces et ces deux éléments sont suffisants pour initier la transcription (Blissard et al., 1992; Chen et al., 2013; Xing et al., 2005). Par ailleurs, d'autres séquences cis-régulatrices sont impliquées dans la régulation de l'expression des gènes viraux précoces, comme par exemple, de la position +11 à +24 après le TSS du promoteur de IE-1. Cependant, les mécanismes et les régions exactes n'ont pas encore été caractérisés précisément (Pullen and Friesen, 1995b), démontrant ainsi que les phénomènes de régulation de l'expression des gènes de baculovirus sont encore mal définis.

Immédiat Précoce et Précoce :

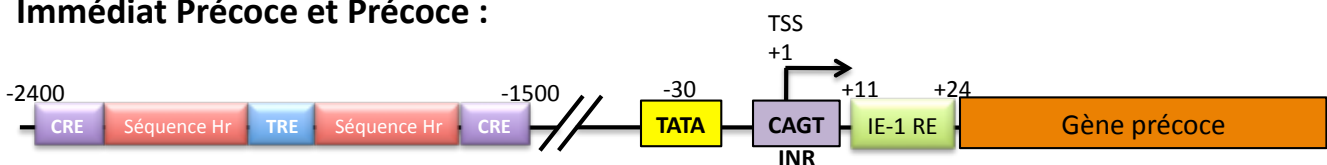


Figure 13 : Représentation schématique de la structure du promoteur des gènes viraux immédiats précoces et précoces.

Les promoteurs des gènes viraux précoces et immédiats précoces sont identiques. Ils sont constitués d'un promoteur contenant une TATA-box située environ trente nucléotides en amont du site d'initiation de la transcription, au niveau de laquelle se trouve une séquence conservée CAGT jouant le rôle d'initiateur. Bien en amont de ces promoteurs, des séquences « enhancers » sont composées d'éléments de réponse à des facteurs cellulaires et des séquences Hr sur lesquelles vient se fixer la protéine virale immédiate précoce IE-1. Cette dernière liaison va transactiver en coopération avec les facteurs cellulaires, la transcription virale.

Une autre séquence, répartie aléatoirement sur le génome viral pourrait jouer le rôle d' « enhancer » dans la régulation de l'expression des gènes viraux précoces : la séquence Hr (Homologous Region) (Guarino and Summers, 1986). Il s'agit d'une région palindromique

imparfaite, variable entre espèces de Baculovirus (Thumbi et al., 2011), d'une trentaine de paires de bases environ, riches en A/T. Dans le cas de AcNMPV, cette séquence est retrouvée à divers endroits, chacune contenant entre deux et huit répétitions. Ces séquences contiennent à leurs extrémités des sites de fixation pour le facteur de transcription cellulaire CREB (CRE) et AP-1 (TRE et répondant à l'activation par le TPA : tétradécanoylphorbol acétate). Il a été démontré que IE-1 est capable de lier sous forme de dimère ces séquences Hr. Cette interaction entraîne une augmentation importante de la transcription des gènes viraux précoces. Toutefois, IE-1 n'a pas besoin de ces éléments Hr pour activer la transcription. De façon similaire, les séquences TRE et CRE dans Hr peuvent activer la transcription en absence de IE-1 mais moins puissamment qu'en présence de IE-1. Ceci suggérerait que la protéine IE-1 pourrait recruter des facteurs cellulaires liant les séquences Hr, TRE et CRE à la machinerie transcriptionnelle de la cellule dans le but d'induire la transcription des gènes viraux précoces (cf. figure 14) (Landais et al., 2006).

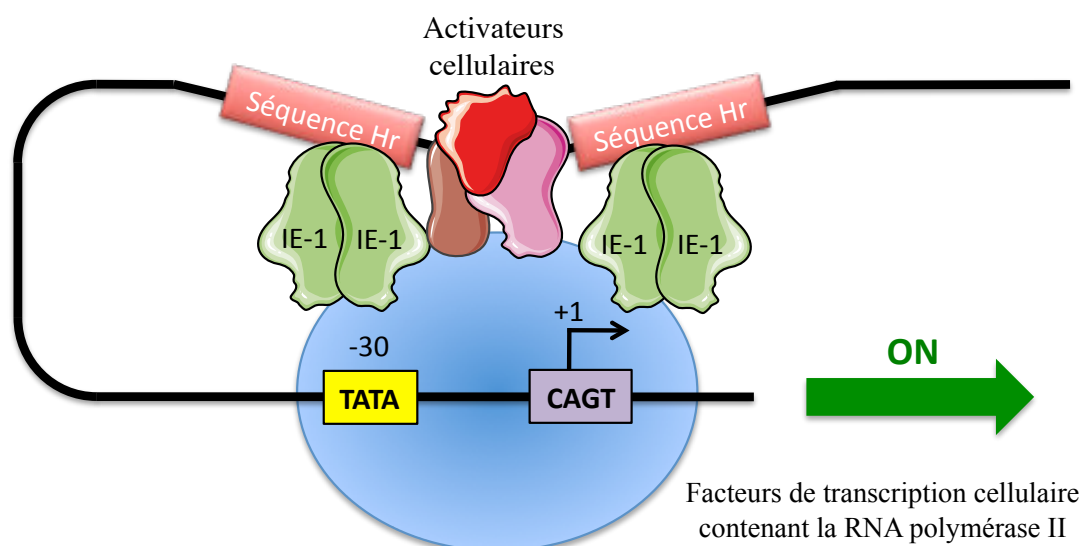


Figure 14 : Représentation schématique de l'activation des promoteurs de gènes viraux immédiats précoces et précoces.

Les séquences Hr situées à plusieurs milliers de paires de bases vont être liées par la protéine virale immédiate précoce IE-1. Probablement en coopération avec des facteurs cellulaires, la protéine IE-1 va aller interagir avec la machinerie de transcription cellulaire pour activer l'expression des gènes viraux sous influence de ce système « enhancer ».

Suite à l'expression des gènes viraux précoces, il y a réplication de l'ADN viral. Cette réplication est assurée grâce à une origine de réplication qui correspondrait aux séquences Hr (Pearson et al., 1992), bien que leur rôle soit encore débattu (Carstens and Wu, 2007). AcMNVP code pour ses propres protéines impliquées dans la réplication de l'ADN viral. À la

fin de la réplication, l'ADN synthétisé servira probablement de matrice à la synthèse des gènes viraux tardifs même si aucune publication ne le démontre réellement. En effet, seule une preuve indirecte tend à confirmer cette affirmation : la quantité d'ARNm viraux augmente significativement après le début de la réplication. Ceci pourrait refléter l'amplification de matrices virales ainsi que l'accessibilité des promoteurs tardifs nouvellement synthétisés, et libres de protéines associées (Rohrmann, 2013).

2) L'expression des gènes viraux tardifs.

L'expression des gènes viraux tardifs est divisée en deux catégories : les tardifs et les très tardifs. Bien que leurs expressions soient espacées temporellement, la structure de leur promoteur est assez similaire.

En effet, il a été démontré qu'une séquence conservée de quatre nucléotides, TAAG, était présente au niveau du site d'initiation de la transcription (cf. figure 15) et que celle-ci était suffisante pour initier la transcription des gènes viraux tardifs et très tardifs (Chen et al., 2013; Morris and Miller, 1994; Rankin et al., 1988, 1988; Xing et al., 2005). Aussi, afin d'activer la transcription de ces gènes viraux, le nucléotide en amont de la thymidine ne doit pas être une cytosine. Le positionnement de cette boîte est prédominant pour son activité transcriptionnelle.

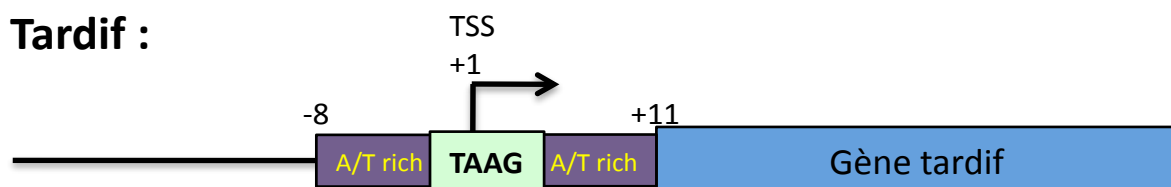


Figure 15 : Représentation schématique de la structure des promoteurs des gènes tardifs de baculovirus.

Les gènes viraux tardifs sont constitués eux d'un promoteur très simplifié dans lequel est conservée une région TAAG où est initiée la transcription. Les régions flanquantes à cette boîte sont riches en A/T. Ces éléments sont nécessaires au recrutement de l'ARN polymérase virale.

Par ailleurs, il semble que les séquences flanquantes (riches en A/T) de ce motif TAAG soient requises pour une transcription efficace des gènes viraux tardifs et très tardifs (Garrity et al., 1997; Morris and Miller, 1994; Xu et al., 1995), bien qu'aucune séquence consensus n'ait été caractérisée (cf. figure 15). Néanmoins, il a été mis en évidence que les nucléotides situés en amont de la boîte TAAG ainsi que cette dernière, sont absolument nécessaires pour la transcription car ils permettent la reconnaissance du promoteur par l'ARN

polymérase virale (Mans and Knebel-Mörsdorf, 1998). Enfin, l'initiation de la transcription se fait directement à partir de la première adénosine de cette boîte.

En plus des éléments cis-régulateurs, la transcription des gènes viraux tardifs de baculovirus est dépendante de facteurs viraux trans-régulateurs. En effet, une dizaine de protéines ont été démontrées comme impliquées dans l'expression des gènes viraux tardifs (Li et al., 1999; Rapp et al., 1998; Todd et al., 1995, 1996). Parmi celles-ci, quatre forment l'ARN polymérase virale, impliquée dans la transcription des gènes viraux tardifs et très tardifs. En effet, Weaver et son équipe ont tenté de purifier l'enzyme ARN polymérase virale en 1996, à partir de différentes techniques de chromatographies. Ils ont ainsi identifié de nombreux peptides, parmi lesquels sont présents les sous-unités formant l'ARN polymérase virale mais aussi des facteurs potentiellement associés à cette protéine (BENIYA et al., 1996). En effet, en 1998, quatre protéines ont été purifiées à partir d'un promoteur de gène viral tardif : p47, LEF4, LEF8 et LEF9. Ensemble, elles forment les différentes sous-unités de l'ARN polymérase virale, résistante à l'alpha-amanitine, et nécessaire à la synthèse des ARNm viraux tardifs. Ces quatre protéines sont présentes en quantité équimolaire et en deux exemplaires chacune (Guarino et al., 1998).

Parmi ces quatre facteurs, seuls LEF8 et LEF9 ont des domaines d'homologies avec les ARN polymérases, laissant supposer que les autres facteurs ont peut-être d'autres fonctions. De plus, les quatre sous-unités de l'ARN polymérase virale ont été identifiées et purifiées à partir d'un promoteur de gène viral tardif, ce qui prouve que cette protéine est capable de reconnaître de manière directe, uniquement les promoteurs des gènes viraux tardifs et très tardifs (Guarino et al., 1998). Elle reconnaît en particulier la boîte TAAG ainsi que les sept nucléotides situés en amont même s'ils ne sont pas conservés d'un gène à l'autre (Mans and Knebel-Mörsdorf, 1998) (cf. figures 15 et 16).

Les mécanismes impliqués dans cette régulation des gènes viraux tardifs et très tardifs sont encore méconnus car très peu d'études ont été réalisées pour identifier le rôle et la fonction d'autres protéines virales dans un contexte infectieux. De plus, le rôle potentiel des facteurs cellulaires dans l'expression des gènes viraux tardifs n'a été que peu observé. Toutefois, quelques équipes ont tenté de caractériser le rôle des différents facteurs impliqués dans l'expression des gènes tardifs.

Parmi ceux-ci, pp31 (protéine virale précoce), décrite dans l'étude de Todd et al., 1995 comme impliquée dans l'expression des gènes tardifs, augmenterait le niveau de la plupart des

transcrits viraux (Yamagishi et al., 2007) et se fixerait à l'ADN (Guarino et al., 2002). LEF-6 (Late Expression Factor-6 qui est une protéine précoce) accélérerait elle, l'expression des gènes viraux tardifs et très tardifs (Lin and Blissard, 2002). Une autre publication a démontré que p35 (protéine précoce) inhibe l'apoptose en bloquant la voie des caspases (Clem et al., 1991). Son rôle anti-apoptotique permettrait donc indirectement l'accumulation d'ARNm viraux. Néanmoins, une autre étude tend à démontrer qu'elle serait impliquée directement dans l'expression des gènes tardifs sans identifier le mécanisme (Hu and Li, 2013). LEF4, qui est une des sous-unités de l'ARN polymérase virale, possède deux activités enzymatiques impliquées dans la mise en place de la coiffe des ARNm (Gross and Shuman, 1998). L'absence de LEF4 conduit à la perte d'expression des ARNm viraux tardifs et très tardifs ce qui démontre l'importance de son activité de capping des ARNm viraux (Knebel-Mörsdorf et al., 2006).

Les gènes viraux très tardifs ont globalement le même promoteur que les gènes viraux tardifs. Cependant, une séquence, située après le TSS du gène très tardif, a été démontrée comme importante dans l'expression des gènes viraux très tardifs (Mans and Knebel-Mörsdorf, 1998; Ooi et al., 1989). Cette séquence répétée et inversée, très riche en A/T, joue un rôle d'« explosion » de la transcription des gènes viraux très tardifs, dans le sens où il y a une surexpression drastique de ces derniers en présence de cette séquence lors de l'infection (cf. figure 16). En effet, l'absence de cette région réduit la synthèse des gènes viraux très tardifs à un niveau comparable à ceux des gènes tardifs (Mans and Knebel-Mörsdorf, 1998; Yang and Miller, 1999). Ces données suggèrent que des facteurs viraux trans-régulateurs peuvent être impliqués dans ce mécanisme.

Promoteur très tardif :

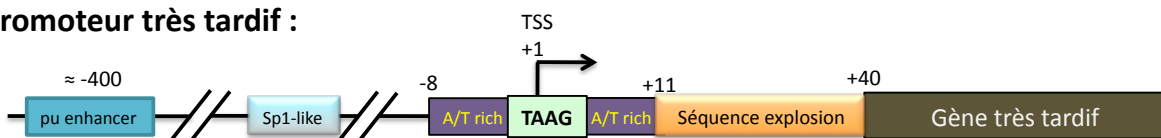


Figure 16 : Représentation schématique de la structure des promoteurs des gènes très tardifs de baculovirus.

Les gènes viraux très tardifs utilisent la même structure que les promoteurs tardifs avec des séquences supplémentaires : une séquence « enhancer » Pu (dont le rôle demeure incertain), des séquences Sp1-like où se fixent des facteurs cellulaires de la famille de Sp1. La région riche en A/T située juste avant la boîte TAAG, fixe la protéine cellulaire PPBP. Enfin, une dernière séquence dite séquence d'« explosion », qui est une séquence répétée et inversée, très riche en A/T, est située entre le +11 et +40 du TSS du gène viral très tardif. Le facteur VLF-1 pourrait interagir avec cet élément pour stimuler la transcription des gènes viraux tardifs.

Une autre séquence virale, nommée Pu (Polyhedrin Upstream sequence), située en amont du promoteur de la polyhedrine, aurait un effet « enhancer-like » sur l'expression de ce gène très tardif (Acharya and Gopinathan, 2001) à travers l'action de trois protéines virales précoces IE-1, IE-2 et pe38. Néanmoins, des expériences menées ultérieurement montrent que cet élément peut agir sur d'autres promoteurs et de manière synergique avec les séquences Hr pour augmenter l'expression de gènes viraux précoces en système rapporteur (Wu et al., 2008). Ces résultats surprenants sont donc à confirmer en contexte infectieux.

Il a été démontré en 1994 que le virus codait pour une protéine tardive nommée VLF-1 (Very Late Factor-1) qui est impliquée dans la surexpression des gènes viraux très tardifs p10 et polyhedrine (McLachlin and Miller, 1994). Cette protéine est étroitement liée à la séquence « d'explosion » car une mutation dans cette région entraîne une diminution drastique de l'expression des gènes viraux très tardifs, et ce malgré la présence de VLF-1. De plus, il semblerait que VLF-1 interagisse directement avec cette séquence cible, du moins, *in vitro* (Yang and Miller, 1999). Aussi, il semble que l'ARN polymérase virale purifiée et VLF-1 soient suffisantes à l'expression des gènes viraux très tardifs (Mistretta and Guarino, 2005) bien qu'aucune interaction entre ces facteurs n'ait été démontrée.

Il a été prouvé que d'autres facteurs viraux et cellulaires ont un rôle dans l'expression des gènes viraux très tardifs. Ainsi, LEF-2 (d'un autre baculovirus que celui massivement étudié), bien qu'impliqué dans la réplication de l'ADN viral, est primordiale à l'expression des gènes viraux très tardifs. Son extinction entraîne dans un système rapporteur une diminution de l'activité luciférase dont le gène est sous contrôle d'un promoteur de gène très tardif (Sriram and Gopinathan, 1998). Par ailleurs, on sait maintenant qu'une des sous-unités de l'ARN polymérase virale (LEF-8) est phosphorylée par la protéine virale pk-1 (Mishra et al., 2008a). Cette dernière est une sérine/thréonine kinase, déjà identifiée par Todd et al., en 1996 dans un système transitoire, comme nécessaire à l'expression des gènes tardifs. Des études plus récentes ont confirmé son rôle : cette protéine est retrouvée au niveau du promoteur d'un gène très tardif (Mishra et al., 2007) et son absence inhibe l'expression des gènes viraux très tardifs (Mishra et al., 2008b).

Hormis ces protéines virales, il a été montré que certains facteurs cellulaires pouvaient être impliqués dans l'expression des gènes viraux très tardifs. En effet, contrairement à ce que pouvait laisser suggérer la publication de Mainz en 2014 (Mainz et al., 2014), il a été prouvé

que TFIID ne se fixait pas sur les promoteurs des gènes viraux très tardifs, *in vitro*, bien qu'une TBP-like, retrouvée dans les cellules d'insectes puissent le faire (Ghosh et al., 1998). En revanche, il est probable que la protéine cellulaire PPBP (Polyhedrin Promoter Binding Protein) soit impliquée dans l'expression des gènes viraux très tardifs. En effet, celle-ci se lie à l'ADN sur la séquence AATAAA, juste avant le TSS des gènes (Burma et al., 1994) et ce, probablement avec d'autres facteurs cellulaires. Elle pourrait donc correspondre à cette TBP-like.

De plus, si TBP n'a pas de rôle dans l'expression des gènes tardifs, elle pourrait être impliquée dans l'expression des gènes viraux précoces. Il a été démontré que son expression protéique était augmentée par un mécanisme encore inconnu lors des étapes tardives du cycle viral productif. Ceci suggère qu'elle pourrait être impliquée dans l'expression des gènes tardifs et très tardifs. De plus, sa relocalisation au niveau des foyers de réplication de l'ADN viral peut être un argument allant dans cette direction, contredisant ainsi le papier de Ghosh en 1998. À l'inverse, il a été démontré par microscopie que l'ARN pol II de l'hôte disparaissait du noyau après la réplication de l'ADN viral, confirmant ainsi qu'elle n'est pas nécessaire à l'expression des gènes tardifs. En revanche, l'ARN pol II co-localise au niveau des spots de TBP au début de l'infection, tendant à confirmer son rôle dans l'expression des gènes viraux précoces (Mainz et al., 2014; Quadt et al., 2002).

Une autre classe de protéine cellulaire a été identifiée comme impliquée dans l'expression des gènes viraux très tardifs en jouant le rôle d'« enhancer » : les Sp-like. En effet, des motifs de liaison semblables à ceux de Sp1 ont été retrouvés bien en amont du promoteur de la polyhedrine. Ainsi, des facteurs cellulaires de la famille des Sp1 (qui sont conservés entre cellules humaines et cellules de drosophiles) sont impliqués dans la stimulation de l'expression du gène viral très tardif polyhedrine (Ramachandran et al., 2001).

Le fait que ces différentes protéines cellulaires soient recrutées sur les promoteurs des gènes viraux très tardifs et qu'elles puissent participer à la transcription de ces gènes apporte de nouveaux éclairages. En effet, ces derniers sont transcrits et surexprimés à l'aide de l'ARN polymérase virale et de VLF-1. Le recrutement de facteurs cellulaires et viraux sur les promoteurs viraux suggère une coopération entre ces différentes protéines pour activer l'expression des gènes viraux tardifs et très tardifs.

Ainsi, quelle que soit la catégorie de gène viral étudié chez le baculovirus, il est possible de constater que la régulation de leur expression est finement contrôlée. Ces mécanismes de transcription sont tous associés à un mélange de facteurs protéiques cellulaires et viraux. L'expression des gènes viraux tardifs est elle, dépendante de l'ARN polymérase virale et ses facteurs associés (cf. figure 17). Cependant, il est évident que tous les systèmes impliqués dans la régulation de l'expression génique des baculovirus ne soient pas encore identifiés. De nombreuses zones d'ombre demeurent dans la compréhension de ce modèle d'infection.

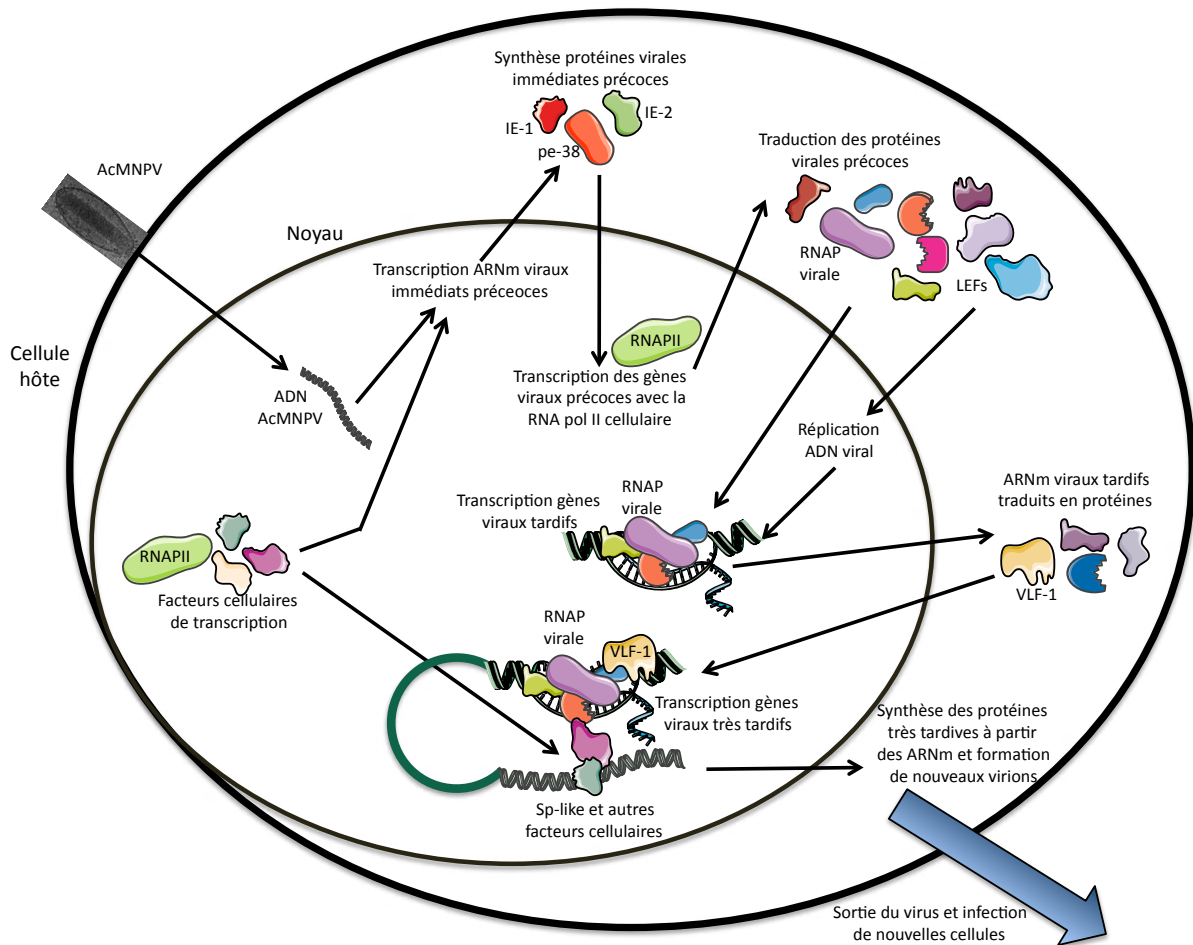


Figure 17 : Représentation schématique de l'expression séquentielle des gènes du baculovirus *Autographa californica Nucleopolyhedrovirus*.

Le cycle commence par l'expression des gènes viraux immédiats précoces qui sont transcrits par les facteurs de transcription cellulaire. Une fois les protéines synthétisées, IE-1 est capable d'augmenter sa propre expression. Suite à l'expression des gènes viraux immédiats précoces, il y a expression des gènes viraux précoces à l'aide de facteurs cellulaires, en coopération avec les facteurs viraux immédiats précoces. Les ARNm viraux précoces vont coder pour des protéines impliquées dans la réplication de l'ADN viral mais aussi dans la transcription des gènes viraux tardifs. Suite à la réplication de l'ADN viral par des facteurs du virus, il y aura transcription des gènes viraux tardifs. Parmi les protéines précoces exprimées, quatre vont former un multimère générant ainsi l'ARN polymérase virale qui va transcrire à l'aide d'autres facteurs les gènes viraux tardifs et très tardifs. Ces

gènes viraux tardifs vont alors coder pour différentes protéines dont l'une d'elles, VLF-1, va stimuler davantage la transcription des gènes viraux très tardifs via une séquence d'ADN, située en aval du TSS de ces derniers. Les ARNm ainsi produits seront traduits en protéines et permettront la synthèse de nouvelles particules virales qui sortiront de la cellule et iront infecter de nouvelles cibles.

C) Les *Adenoviridae* (Adénovirus).

Les adénovirus font partie de la famille des *Adenoviridae*, répartis en cinq genres. Ce sont des virus capables d'infecter un large spectre d'hôtes, allant des mammifères aux oiseaux mais aussi des reptiles aux amphibiens. De très nombreux types d'adénovirus ont été référencés et parmi ceux-ci, plus de soixante sont capables d'infecter l'homme. Ils font tous partie du genre *Mastadenovirus* et ont été divisés en sept espèces (A à G). Chez l'humain, les sérotypes 2 et 5 (Ad2 et Ad5), tous deux de classe C, sont les plus étudiés et seront utilisés comme base de description dans cette partie du manuscrit. Les adénovirus sont des virus non enveloppés et ont une capsidie icosaédrique (constituées de nombreuses protéines) d'un diamètre de 80 à 90 nm (cf. figure 18). Ces particules virales contiennent un ADN double brin, linéaire d'environ 30 à 40 kpb.

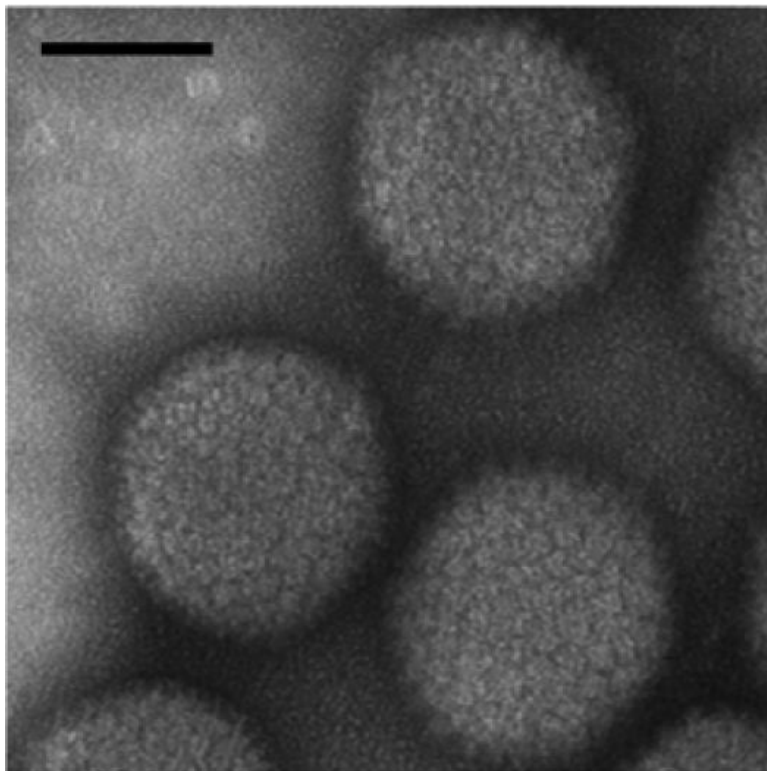


Figure 18 : Photographie d'une particule virale d'*Adenoviridae* humain, de type 5 par électromicroscopie, extrait de (Tang et al., 2009).

L'adénovirus est un petit virus non enveloppé, à capsidie icosaédrique, contenant un ADN double brin d'environ 40 kpb.

La majorité des adénovirus produit environ une quarantaine de protéines classées comme précoces ou tardives en fonction de leur expression avant ou après la réplication de l'ADN viral. Les gènes codants pour ces différentes protéines sont répartis en plusieurs régions sur le génome (en rose les précoces et en bleu les tardifs, cf. figure 19). Aux extrémités du génome, des répétitions terminales inversées (ITR) sont nécessaires à l'encapsidation de l'ADN (cf. figure 19).

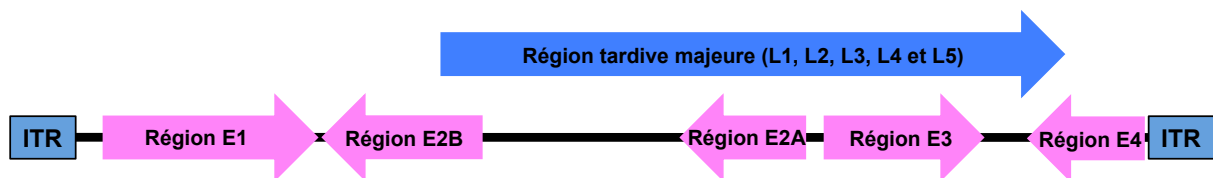


Figure 19 : Représentation schématique du génome d'adénovirus.

Le génome d'adénovirus contient à ses extrémités des séquences ITR (répétitions terminales inversées) et possède deux grandes régions : une région précoce et une région tardive. La région précoce (en rose) est elle-même répartie en quatre régions E1, E2, E3 et E4, toutes dépendantes d'un promoteur par bloc, hormis E2 qui est divisée en deux sous régions : E2A et E2B. La région tardive (en bleue) est dépendante d'un promoteur majeur tardif pour la plupart des gènes et est sous-divisée en cinq blocs.

L'infection par adénovirus commence par l'entrée du virus dans la cellule. L'ADN viral arrivé dans le noyau va alors se chromatiniser (Giberson et al., 2012) avec des histones H3.3 (marqueurs actifs de la transcription). Le cycle viral productif va débuter et durer entre 24 et 36h pour produire de nouvelles particules virales en commençant par l'expression des gènes viraux précoces.

1) L'expression des gènes viraux précoces.

Les gènes viraux précoces sont divisés en quatre régions (E1, E2, E3 et E4) dont chacune est sous la dépendance de son propre promoteur (hormis E2) (cf. figure 19). L'ARN synthétisé à partir de ces promoteurs va être épissé pour donner des profils d'ARNm viraux précoces complets qui seront ensuite traduits en protéines.

Les promoteurs des gènes viraux précoces, paraissent assez bien conservés entre adénovirus. Toutefois, en fonction des blocs de régions, la structure de ces promoteurs peut sensiblement varier. Globalement, ceux-ci sont régulés de la même manière que les promoteurs des gènes cellulaires en contenant de nombreuses séquences cis-régulatrices, fixés par différents facteurs cellulaires et viraux.

Par exemple, le promoteur de la région E3 est constitué d'une TATA-box localisée une trentaine de nucléotides en amont du TSS, d'une séquence liée par ATF entre -44 et -68 avant le TSS (cf. figure 20). Plus en amont, AP-1 se fixe sur des éléments situés entre -80 et -100 avant le TSS et NF-1 (Nuclear Factor 1) sur des séquences cibles localisées entre -154 et -183 en amont du TSS (Hurst and Jones, 1987; Leff et al., 1985). Ces deux derniers facteurs semblent coopérer avec les protéines virales E1A et E1B pour augmenter l'activité transcriptionnelle de ce promoteur (Kornuc et al., 1990; Lee et al., 1982).

Promoteur Précoce (E3) :

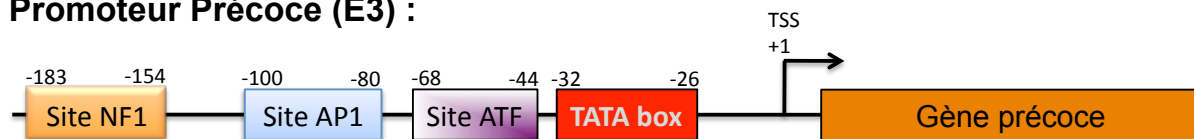


Figure 20 : Représentation schématique de la structure d'un promoteur précoce d'adénovirus.

Les promoteurs des gènes viraux précoces, notamment celui de E3, sont constitués de nombreux éléments cis-régulateurs. Les séquences les plus fréquemment retrouvées sont la TATA-box, les sites de fixation à AP-1 et ATF, sur lesquelles sont recrutées de nombreux facteurs cellulaires dont AP-1, ATF ainsi que les facteurs généraux de la transcription cellulaire, y compris TFIID qui se fixe sur la boîte TATA canonique. À travers les facteurs cellulaires recrutés, E1A est capable de stimuler la transcription.

Après l'expression des gènes viraux précoces, certaines des protéines issues de ces gènes vont permettre la réplication de l'ADN viral, notamment celles issues des ARNm provenant de la région E2. Ceux-ci codent pour : une protéine de liaison à l'ADN simple brin, une ADN polymérase virale et un précurseur protéique pTP (protéine pré-terminale, clivée après encapsidation) (Stillman et al., 1981). Cette dernière est associée aux deux extrémités de l'ADN viral : les séquences ITR servant d'origine de réplication de l'ADN viral. Cette protéine virale va servir d'amorce pour la réplication et sera associée à la protéine de liaison à l'ADN simple brin et à l'ADN polymérase ainsi qu'à trois facteurs cellulaires pour assurer la réplication (Hoeben and Uil, 2013). Par ailleurs, il semble que la protéine E4 orf3, probablement en association avec E4 orf6, facilite la réplication de l'ADN viral en réorganisant la structure nucléaire des domaines PML (qui sont des domaines associés à la matrice nucléaire) (Evans and Hearing, 2003). Cette étape va permettre le passage entre phase précoce et phase tardive et va favoriser l'expression des gènes viraux tardifs à partir d'un ADN probablement dépourvu de marquages épigénétiques (Paul Thomas and Mathews, 1980). L'ensemble sera accompagné d'une diminution de l'expression des gènes viraux précoces.

2) L'expression de gènes couplée à la réplication de l'ADN viral.

De manière concomitante à la réplication de l'ADN viral, certains gènes viraux vont être exprimés grâce au déplacement d'un facteur répresseur cellulaire. Par exemple le gène codant pour la protéine IVa2 est sous le contrôle d'un promoteur constitué d'une séquence située entre -60 et -38 en amont du TSS, riche en GC et d'une autre séquence comprise entre -20 et -35. De plus, un élément INR est retrouvé à cheval sur le site d'initiation de la transcription (-7/+13). Ce promoteur est également composé d'une région riche en TA : TATAGAA, qui ressemble à la TATA-box canonique, située entre +21 et +29 après le TSS (sur le brin non codant). Ce domaine est à priori fixé par le complexe général de transcription TFIID (Carcamo et al., 1990; Chen and Flint, 1992; Kasai et al., 1992). L'expression de l'ARNm de IVa2 est par ailleurs augmentée en présence de E1A (Natarajan and Salzman, 1985). En plus de posséder des régions activatrices, une séquence inhibitrice de la transcription a été identifiée ultérieurement entre les positions +14 et +19 du TSS sur le promoteur de IVa2 (Chen et al., 1994). En revanche, le facteur cellulaire répresseur n'a toujours pas été caractérisé. Il a été cependant démontré *in vitro* que la réplication jouait un rôle majeur dans la levée de l'inhibition de ce promoteur par un effet de titration, due à l'augmentation du nombre de matrice d'ADN (Huang et al., 2003). Ces résultats ont été confirmés *in vivo* par mutations de la région inhibitrice du promoteur IVa2, aboutissant à une augmentation de l'expression de l'ARNm correspondant. (Iftode and Flint, 2004).

En 2010, une découverte majeure a permis de mieux comprendre le passage entre phase précoce et tardive. Un nouveau promoteur, nommé L4P (promoteur L4) a été caractérisé (Morris et al., 2010). Ce promoteur est situé dans la région codante de L4-100K et contrôle l'expression d'un gène codant deux protéines essentielles au virus : L4-33K et L4-22K. Une séquence INR, située au niveau du site d'initiation de la transcription, tout comme pour le promoteur de IVa2 est présente (Morris et al., 2010). Une TATA-box non consensus TACGAAA, située en -30 avant le TSS a également été identifiée. Néanmoins, son rôle dans le promoteur L4P semble peu évident. En revanche, trois sites de fixation pour TFII-I ont été caractérisés : le premier chevauche la TATA-box, le second est situé sur la séquence INR et le dernier est situé trente nucléotides en aval du TSS (cf. figure 21). Cependant, seul les deux premiers sites semblent réprimer l'activité transcriptionnelle du L4P par la fixation de TFII-I

sur ces séquences cibles. La répression du promoteur L4P réalisée par TFII-I peut être levée par la protéine E4 orf3 (considérée comme activatrice du promoteur L4P), uniquement en présence de la séquence INR. Il a été également révélé que L4-33K inhibait l'activité transcriptionnelle de L4P dans un système transitoire, contrairement à L4-22K (Wright et al., 2015).

Outre L4-22K, d'autres protéines virales ont été démontrées comme activatrice du promoteur L4P : E1A, E4 orf3 et IVa2, sans en connaître le mécanisme (Morris et al., 2010). De plus, la protéine cellulaire p53 est retrouvée associée au niveau du promoteur L4P, sans pour autant connaître les séquences de fixation, et active ce dernier. La présence de E4 orf3 et IVa2 augmente d'autant plus le rôle activateur de p53 dans la transcription à partir du L4P. Cependant, l'association entre p53 et L4P est transitoire car elle est observée seulement entre 10 et 14h post-infection, en parfaite corrélation avec le début de l'expression des gènes viraux tardifs. Cette liaison peut être bloquée par les protéines L4-22K et L4-33K, toutes deux issues du gène sous contrôle de L4P. Cette observation suggère un rétro-contrôle négatif de la part de ces facteurs viraux sur leur promoteur (Wright and Leppard, 2013).

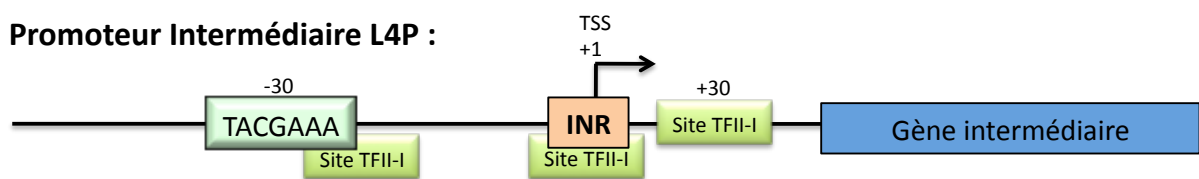


Figure 21 : Représentation schématique de la structure d'un promoteur intermédiaire d'adénovirus.

Les gènes viraux synthétisés en même temps que la réplication de l'ADN viral possèdent des promoteurs différents. Toutefois, le promoteur le plus important demeure probablement le L4P qui permet l'expression des facteurs viraux, essentiels au passage de la phase précoce à tardive. L4P est constitué d'une séquence INR ainsi que d'une boîte non canonique TATA, située une trentaine de nucléotides en amont du TSS. Le facteur cellulaire p53 stimule avec différents facteurs viraux (E1A, E4 orf3 et IVa2) ce promoteur. En revanche, trois séquences de fixation pour le facteur cellulaire TFII-I sont présentes et ont un rôle inhibiteur de la transcription. Cette inhibition est levée à l'aide de la protéine E4 orf3 qui semble contrecarrer cet effet.

Enfin, comme pour IVa2, il semble évident que la réplication de l'ADN viral active ce promoteur, probablement grâce aux modifications de la chromatine engendrées par la réplication de l'ADN viral mais aussi à l'augmentation du nombre de matrices promotrices.

Ainsi, la découverte du promoteur L4P régulant l'expression du gène codant pour les protéines virales L4-22K et L4-33K explique le changement de la phase précoce à la phase

tardive de l'infection. En effet, ces facteurs viraux, exprimés de manière concomitante avec la réplication de l'ADN viral ont un rôle dans l'expression des gènes viraux tardifs (Farley et al., 2004). D'ailleurs, de très nombreuses études ont démontré l'implication de ces deux facteurs dans l'expression des gènes viraux tardifs, ce que nous développerons ultérieurement.

3) La transcription des gènes tardifs.

Chez les adénovirus, les gènes viraux tardifs sont regroupés dans une unité de transcription majeure, produisant ainsi plus de vingt ARNm différents, divisés en cinq familles : L1 à L5. Chaque famille a son site de polyadénylation en 3'. L'expression des vrais ARNm viraux tardifs est dépendante d'un unique promoteur tardif majeur (nommé MLP). Au début de l'infection, tous ces ARNm sont transcrits à un niveau basal (mais non retrouvés dans le cytoplasme (Shaw and Ziff, 1980)), contrairement à la période tardive où ils sont efficacement exprimés et traduits.

Le promoteur tardif est constitué d'une boîte TATA canonique, d'une séquence riche en GC autour de la TATA-box (amont et aval) ainsi que d'un élément INR. Des sites de liaison à des facteurs de transcription sont également localisés en amont du TSS : CAAT box et UPE (Éléments en amont du promoteur), situés entre -80 et -52 avant le TSS (cf. figure 22) (Hen et al., 1982; Hu and Manley, 1981; Logan and Shenk, 1986; Lu et al., 1997; Sawadogo and Roeder, 1985).

Promoteur Tardif MLP :

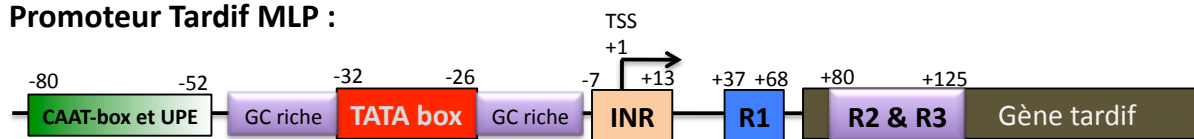


Figure 22 : Représentation schématique de la structure du promoteur tardif majeur d'adénovirus.

Le promoteur majeur des gènes viraux tardifs est le seul qui permette la synthèse d'une vingtaine d'ARNm. Celui-ci, massivement étudié pour comprendre les mécanismes de régulation de la transcription, est caractérisé par la présence d'une boîte canonique TATA, située trente nucléotides en amont du TS et cernée de séquences riches en GC. En amont de cette boîte sont retrouvés des éléments de réponse à des facteurs cellulaires, au rôle redondant. Une séquence INR, absolument nécessaire à la transcription est également présente au niveau du TSS. De plus, une région s'étendant du +37 au +125 après le TSS, divisée en trois régions (R1, R2 et R3) est fixée par différents facteurs : L4-22K et IVa2. L'ensemble de ces facteurs interagissent et conduisent à l'activation de la transcription des gènes viraux tardifs.

Un peu plus en aval, des séquences situées entre +37 / + 68 et +80 / +125 sont présentes. Elles sont nommées R1 (+37 à +68), R2 (+80 à +96) et R3 (+113 à +125), ou DE pour R2 et R3 (Downstream Elements, R2=DE1 et R3=D2). Ces régions semblent coopérer avec les éléments situés en amont du TSS (Leong et al., 1990; Mondésert and Kédinger, 1991) (cf. figure 22). La majorité de ces éléments sont conservés (hormis UPE, remplacé par un site de liaison à Sp1) entre les différents adénovirus, ce qui tend à confirmer leur implication dans l'activité transcriptionnelle du MLP (Song and Young, 1997; SONG et al., 1996).

Ce promoteur MLP est régulé par de nombreux facteurs cellulaires qui peuvent lier la TATA-box, en l'occurrence TFIID avec les facteurs associés et le reste de la machinerie transcriptionnelle, ARN pol II comprise (Burton et al., 1986). La région UPE est liée par le facteur cellulaire USF (Sawadogo and Roeder, 1985). Les séquences riches en GC, situées en amont du TSS, sont fixées par Maz et Sp1 (Parks and Shenk, 1997). La coopération entre facteurs cellulaires semble essentielle pour l'activité transcriptionnelle de ce promoteur car des mutations dans certaines séquences réduisent cette activité. Néanmoins, il semble que les rôles des sites CAAT et UPE soient redondants car ils peuvent se suppléer (Reach et al., 1991). Ainsi, le recrutement de ces facteurs de transcription cellulaire permet de comprendre comment les gènes viraux tardifs sont légèrement exprimés lors des étapes précoces du cycle productif d'adénovirus.

La forte activité transcriptionnelle observée lors des étapes tardives du cycle productif est due à l'activité trans-activatrice de E1A ainsi qu'à la présence des séquences DE qui correspondent aux sites de fixations de protéines virales trans-activatrices (Mondesert et al., 1992). La protéine IVa2 (DEF-B) fixe en dimère la région DE2 et la région DE1 (Tribouley et al., 1994) avec L4-22K (DEF-A) pour former un hétérodimère. L4-22K peut également s'attacher à la région DE2 mais avec une affinité moindre que pour DE1 (Lutz and Keding, 1996; Ostapchuk et al., 2006) (cf. figure 22). IVa2 n'est pas capable d'activer seule le promoteur tardif majeur, contrairement à la protéine L4-22K (Tribouley et al., 1994). Les résultats de ces expériences réalisées *in vitro*, ont été confirmés dans un contexte infectieux en mutant les régions DE (Pardo-Mateos and Young, 2004). Néanmoins, l'action de IVa2 et L4-22K dans un système rapporteur luciférase sous contrôle du MLP est modeste comparée à celle du virus complet, ce qui tend à prouver que d'autres facteurs viraux peuvent être impliqués dans cette régulation (Morris and Leppard, 2009). Par ailleurs, il semble que

d'autres séquences de fixation à L4-22K soient présentes dans ces régions DE (Backström et al., 2010).

En effet, il a été démontré très récemment que L4-22K était fixée à une séquence située juste après le TSS du promoteur tardif majeur, entre +37 et +68 (région R1) (cf. figure 23) (Lan et al., 2015). Cette région a déjà été montrée comme importante pour l'activité du MLP sans pour autant connaître le(s) facteur(s) associé(s) (Leong et al., 1990). Toutefois, L4-22K se fixe avec moins d'affinité à cet élément qu'aux séquences DE. La liaison de L4-22K à la région R1 inhibe légèrement l'activité transcriptionnelle du MLP, tout comme Sp1 et probablement avec d'autres facteurs cellulaires encore non identifiés (Lan et al., 2015). Cette étude ayant été réalisée avec un système rapporteur, il aurait été intéressant d'observer par Chromatine Immuno-Précipitation (ChIP) la fixation de Sp1 au cours du cycle viral sur cette séquence (comme avec p53 sur le L4P). Il est possible qu'en contexte d'infection, cette observation soit erronée puisqu'il a été démontré que Sp1 activait la transcription du promoteur tardif majeur en se fixant sur deux séquences riches en GC juste avant la TATA-box et -166 avant le TSS du MLP (Parks and Shenk, 1997).

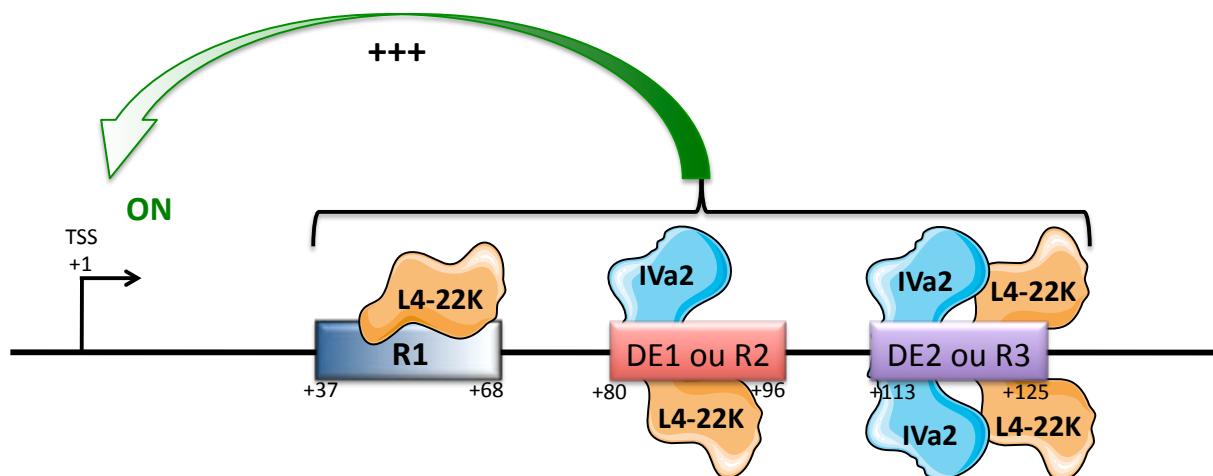


Figure 23 : Représentation schématique des régions présentes en aval du site d'initiation de la transcription du promoteur tardif majeur et des facteurs viraux associés.

Trois régions situées en aval du TSS sont nécessaires à l'expression des gènes viraux tardifs : R1, R2=DE1 et R3=DE2. Ces éléments sont fixés par des protéines virales intermédiaires. IVa2 (DEF-B) est capable de se fixer en dimère sur la partie 5' de DE2 et en hétérodimère avec L4-22K (DEF-A) sur la région DE1. Cette dernière protéine virale est elle aussi capable de fixer en dimère la partie 3' de la région DE2 et de lier la région R1. L'ensemble de ces éléments va stimuler la transcription des gènes viraux tardifs à un niveau très élevé.

Un autre facteur cellulaire, MAZ, lie ces deux séquences mais aussi deux autres sites plus en amont (deux vers la TATA-box et deux vers -160), et ce avec une meilleure affinité. La surexpression de ces deux facteurs cellulaires (Sp1 et MAZ) active le promoteur tardif

majeur dans les essais rapporteurs luciférase au même niveau qu'en contexte infectieux. L'activité du MLP est drastiquement augmentée lorsque ces deux facteurs sont en présence de E1A, d'autant plus que ces facteurs interagissent ensemble. (Parks and Shenk, 1997).

Outre son implication dans la transcription des gènes viraux tardifs, L4-22K induit une réduction légère de l'expression des gènes viraux intermédiaires IVa2 et IX alors qu'ils sont exprimés juste après le début de la réplication de l'ADN viral. Ces résultats laissent supposer un rôle de stabilisation de L4-22K à un niveau post-transcriptionnel sur ces protéines. En effet, la trans-complémentation d'un vecteur d'expression codant pour la protéine L4-22K restaure le profil d'expression des gènes viraux IVa2 et IX, contrairement à la protéine L4-33K. (Morris and Leppard, 2009). Hormis ce rôle sur les gènes intermédiaires, L4-22K stimule l'expression des gènes viraux tardifs (en se fixant aux séquences DE), engendrant par conséquent l'extinction des gènes viraux précoces (Wu et al., 2012).

D'autres études plus récentes ont confirmé que l'absence des deux protéines virales, issues du transcrit L4 (L4-22K et L4-33K), engendre une diminution drastique de l'expression des ARNm virus tardifs, corroborant des résultats plus anciens (Wu et al., 2012, 2013). Cette réduction peut être compensée pour certains ARNm par la surexpression de L4-100K qui stimule la traduction des ARNm tardifs.

Le rôle transcriptionnel de L4-22K a déjà été évoqué précédemment. Concernant L4-33K, il est connu pour être un facteur d'épissage des ARNm viraux tardifs et être impliqué dans leur expression. En effet, son absence entraîne une diminution drastique de l'expression des gènes viraux tardifs. Cette déficience mène également à l'accumulation des ARNm de L1-52/55K, L4-100K et L4-22K, exprimés juste avant les gènes tardifs. Ces résultats mettent en avant un problème de transition entre gènes viraux précoces et tardifs (Wu et al., 2013).

4) L'épissage des ARNm tardifs, une étape clé dans leur expression.

Une autre étape importante régule l'expression des ARNm viraux tardifs. En effet, il semble que les protéines E4 orf3 et E4 orf6 jouent un rôle prépondérant dans la stabilité des ARNm en agissant au niveau post-transcriptionnel. L'absence de ces deux protéines induit un défaut d'export des ARNm viraux tardifs dans le cytoplasme et les stabilisent dans le noyau (Bridge and Ketner, 1989; Sandler and Ketner, 1989) en agissant au niveau d'une région

située en 5' de l'ARNm (tripartite leader) (Nordqvist and Akusjärvi, 1990; Öhman et al., 1993). De plus, il est montré que E4 orf3 favorise l'épissage constitutif alors que E4 orf6 favorise l'épissage alternatif (Nordqvist et al., 1994).

L'ARNm de L1-52/55K est retrouvé très précocement au cours de l'infection dans le cytoplasme car il est traduit en protéine au contraire de IIIa, pourtant issu du même pré-ARNm. Cette différence s'explique par une régulation très fine de l'épissage. En effet, les ARNm de L1-52/55K et IIIa sont issus d'un pré-ARNm commun où le dernier intron est épissé en utilisant un site 5' donneur unique et deux sites accepteurs 3', générant ainsi deux ARNm distincts après épissage. Cependant, le pré-ARNm de IIIa est très faiblement voire pas du tout épissé au début du cycle viral productif. Ceci est dû à la présence d'un élément répresseur, nommé 3RE, situé au niveau du site de branchement de l'épissage, dans l'intron (cf. figure 24).

Les protéines SR (Serine/Arginin-rich protein), stimulent généralement l'épissage. Dans cette situation, SRSF2, 4 et 6 se fixent sur la région 3RE et bloquent le recrutement des facteurs du spliceosome au niveau de ce site 3' d'épissage. Cet enchaînement empêche l'épissage de IIIa et va favoriser la forme épissée de L1-52/55K (cf. figure 24). En effet, la présence de cette séquence 3RE dans l'exon de L1-52/55K stimule son épissage de la même manière que la surexpression des protéines SR (Kanopka et al., 1996). Néanmoins, ce blocage est réversible par l'intermédiaire de la protéine E4 orf4. Cette dernière interagit avec une des phosphatases cellulaires PP2A. Le complexe E4 orf4 — PP2A va conduire à la déphosphorylation des protéines SR, en interagissant notamment avec SRSF1 et d'autres protéines SR hyperphosphorylées. Cet événement va ainsi bloquer la liaison des SR sur le pré-ARNm et permettre l'accès au spliceosome. Ces données expliquent l'expression de IIIa uniquement dans les étapes tardives du cycle viral (Kanopka et al., 1998; Nilsson et al., 2001).

En plus de la région 3RE, une autre séquence est importante dans le contrôle de l'épissage de ce pré-ARNm : 3VDE (cf. figure 24). Il s'agit d'une séquence « enhancer » d'épissage fixée par la protéine L4-33K (Mühlemann et al., 2000). Cette dernière peut également lier la séquence 3RE. Elle est capable d'induire à elle seule la modification du profil d'épissage du pré-ARNm de L1 en faveur de IIIa, probablement en recrutant le spliceosome (Törmänen et al., 2006). En effet, il est possible d'imaginer le recrutement de facteurs cellulaires du spliceosome sur ces éléments cis-régulateurs (3RE et 3VDE) afin de

réguler l'épissage des gènes viraux tardifs car deux éléments vont dans ce sens. Premièrement, E4 orf4 relève l'inhibition de l'épissage induit par la protéine SRSF1 au niveau des séquences 3RE. Deuxièmement, le domaine C-Terminal de L4-33K est similaire aux protéines SR (Biasiotto and Akusjärvi, 2015). Néanmoins, aucune étude n'a pour l'instant identifié les facteurs du complexe cellulaire d'épissage recruté par L4-33K. Ainsi, l'ensemble de ces événements permettrait de contrôler et réguler l'expression des ARNm viraux tardifs lors des différentes phases du cycle viral.

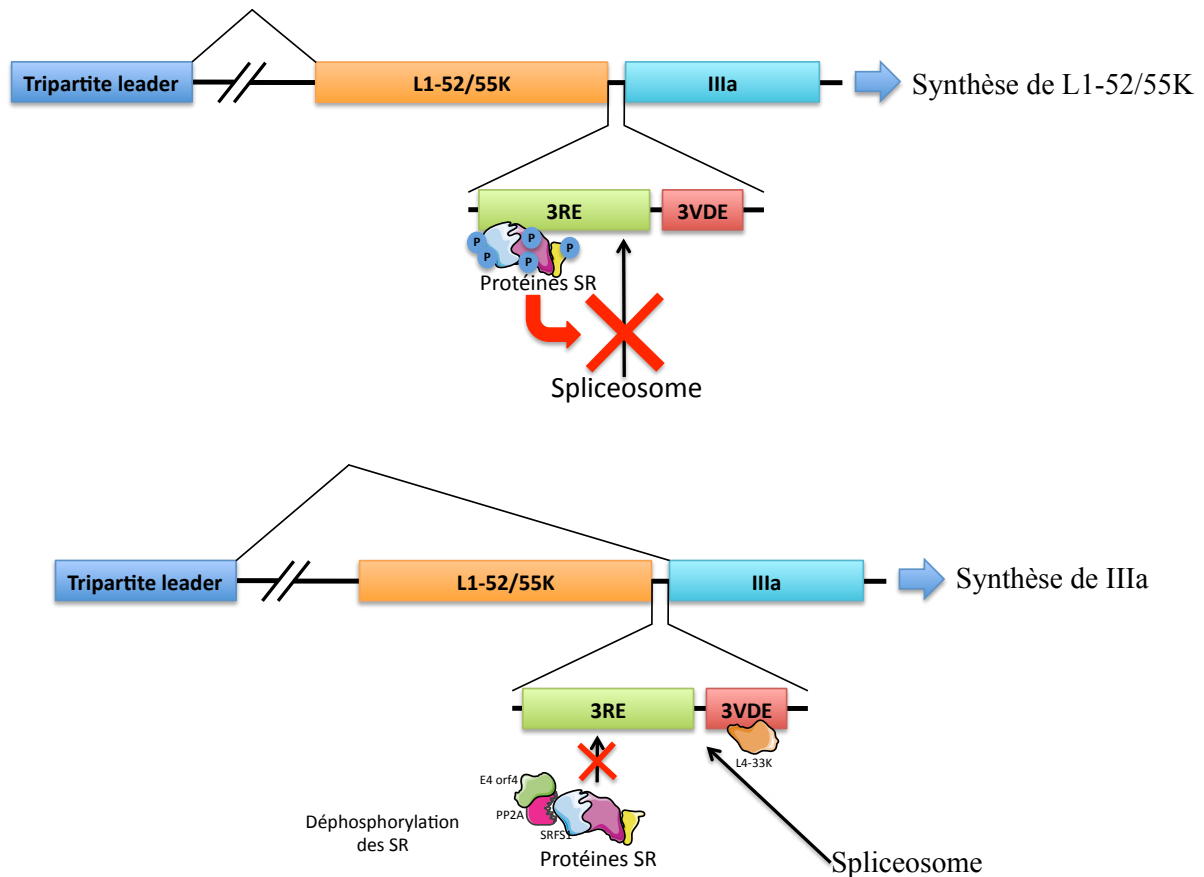


Figure 24 : Représentation schématique du contrôle de la régulation de l'épissage de la région tardive L1.

La région L1 code pour deux protéines virales (L1-52/55K et IIIa) issues d'un même pré-ARNm. Cet ARN va être contrôlé au cours du cycle viral afin d'obtenir un profil d'expression différentiel des facteurs viraux. Lors de l'étape précoce du cycle viral, seule L1-52/55K est exprimée. Ceci est dû à la présence de facteurs cellulaires nommés SR (protéines Sérine/Arginine riches) qui sont hyperphosphorylées (représentés par des P). Ces protéines vont se fixer au niveau de la région intronique de IIIa : 3RE (Elément Répresseur de IIIa). De ce fait, le recrutement de la machinerie d'épissage (spliceosome) est bloqué. Cette action va empêcher l'épissage de IIIa et va favoriser la forme épissée L1-52/55K. À l'inverse, lors de l'étape tardive du cycle viral, la protéine E4 orf4 va interagir avec la phosphatase cellulaire PP2A, entraînant une déphosphorylation des protéines SR et permettre le recrutement du spliceosome. De plus le facteur L4-33K, produit intermédiaire du virus, va aller se fixer sur le

site de branchement d'épissage en 3'. Il semble suffisant pour recruter à lui seul le spliceosome et induire l'épissage de IIIa.

5) La traduction des ARNm viraux est favorisée lors de l'étape tardive de l'infection.

En plus de subir un contrôle au niveau de l'épissage, les ARNm viraux tardifs sont efficacement traduits à l'aide d'une région nommée « tripartite leader ». Elle est composée de deux cent nucléotides et est située dans la région 5' non codante, juste derrière le TSS. Cette région « tripartite leader » est présente dans la majorité des transcrits viraux tardifs (Logan and Shenk, 1984). De plus, le facteur viral protéique L4-100K a été révélé comme étant un stimulateur de la traduction des ARNm viraux tardifs (Hayes et al., 1990) et possède un domaine de liaison à l'ARNm dans cette région tripartite en 5' non codante (Xi et al., 2004). Bien que coiffés comme les ARNm cellulaires, les ARNm viraux tardifs sont traduits de manière différente. En effet, la petite sous-unité ribosomique 40S est recrutée de manière coiffe-dépendante¹ mais ne scanne pas la partie 5' non codante (à contrario des ARNm cellulaires) et migre directement derrière le codon d'initiation de la traduction (Yueh and Schneider, 1996).

La fixation de L4-100K sur la séquence « tripartite-leader » est couplée avec l'interaction de L4-100K à eIF4G. Ceci permet le recrutement du ribosome 40S au site d'initiation de la traduction (cf. figure 25). De plus, L4-100K augmente le niveau des protéines eIF4E, eIF4G et PABP associées au polysomes, en présence des ARNm contenant la région tripartite leader, facilitant ainsi la traduction uniquement des ARNm viraux tardifs (Xi et al., 2004).

À cette stimulation de la traduction des ARNm viraux tardifs est associée une extinction de la traduction des ARNm cellulaires. Cet impact est dû à l'inhibition de l'interaction entre eIF4G et la kinase de eIF4E : Mnk1. En effet, cette dernière (Mnk-1) interagit avec L4-100K, qui elle-même est associée à eIF4G au niveau de son domaine de

¹ Le complexe d'initiation de la traduction cellulaire coiffe-dépendante est assuré par le complexe protéique eIF4F. Il est constitué de différentes protéines : eIF4E (protéine liant la coiffe de l'ARNm) et sa kinase Mnk1, eIF4A (l'ARN hélicase), eIF4G (protéine adaptatrice). eIF4G interagit avec la PABP (polyA-binding protein) et le facteur de traduction cellulaire eIF3 qui recrute la petite sous-unité ribosomique 40S. La phosphorylation de eIF4E par sa kinase Mnk1 est prépondérante pour l'initiation de la traduction et est assurée uniquement lorsque ces deux derniers facteurs sont associés à eIF4G (Pyronnet et al., 1999)

liaison habituel à Mnk1. Cette étape va empêcher la phosphorylation de eIF4E et semble réduire l'association entre eIF4A et eIF4G. Par conséquent, l'ensemble de ces événements sous-entend un défaut dans l'assemblage du complexe d'initiation de la traduction coiffe-dépendante des ARNm cellulaires au profit des ARNm viraux tardifs (Cuesta et al., 2000, 2004).

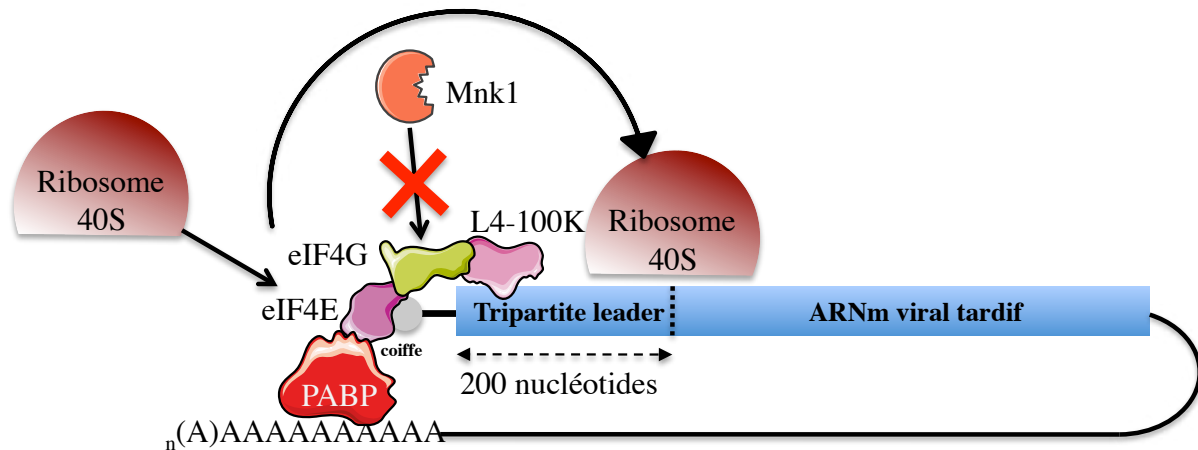


Figure 25 : Représentation schématique de la stimulation de la traduction des ARNm viraux tardifs par L4-100K.

La protéine virale L4-100K va se fixer sur la séquence tripartite leader et va interagir avec le facteur de traduction eIF4G. Cette interaction va bloquer le recrutement de Mnk1, empêchant la phosphorylation de eIF4E. Le ribosome 40S sera recruté de manière coiffe-dépendante et n'effectuera pas de scanning de la 5' non codante de l'ARNm viral. Il sera directement déposé au niveau du site d'initiation de la traduction pour plus d'efficacité.

Ainsi, les promoteurs d'adénovirus et les mécanismes les régulant sont bien plus complexes que ceux des bactériophages et des baculovirus car la simplification des promoteurs des gènes viraux au cours du cycle productif de ces derniers n'est pas retrouvée chez les adénovirus. En effet, les gènes viraux précoces sont sous le contrôle de promoteurs similaires à ceux des gènes cellulaires. Ils sont constitués de nombreuses séquences cis-régulatrices, fixées par différents facteurs de transcription. Les gènes viraux transcrits de manière concomitante à la réplication de l'ADN viral sont sous le contrôle de facteurs cellulaires répresseurs, titrés lors de la réplication du génome viral. Ces protéines intermédiaires vont stimuler la transcription d'autres facteurs intermédiaires, qui eux même vont jouer un rôle prépondérant dans l'augmentation de la transcription des gènes viraux tardifs (cf. figures 23, 24 et 26). Enfin, les gènes viraux tardifs sont très finement régulés au niveau transcriptionnel par des séquences cis-régulatrices sur lesquelles sont fixées des facteurs généraux de transcription cellulaires et viraux. Ils sont également sujets à des

mécanismes de régulation post-transcriptionnels tels que l'épissage, l'export et la traduction qui sont régulés par des facteurs viraux (cf. figure 26).

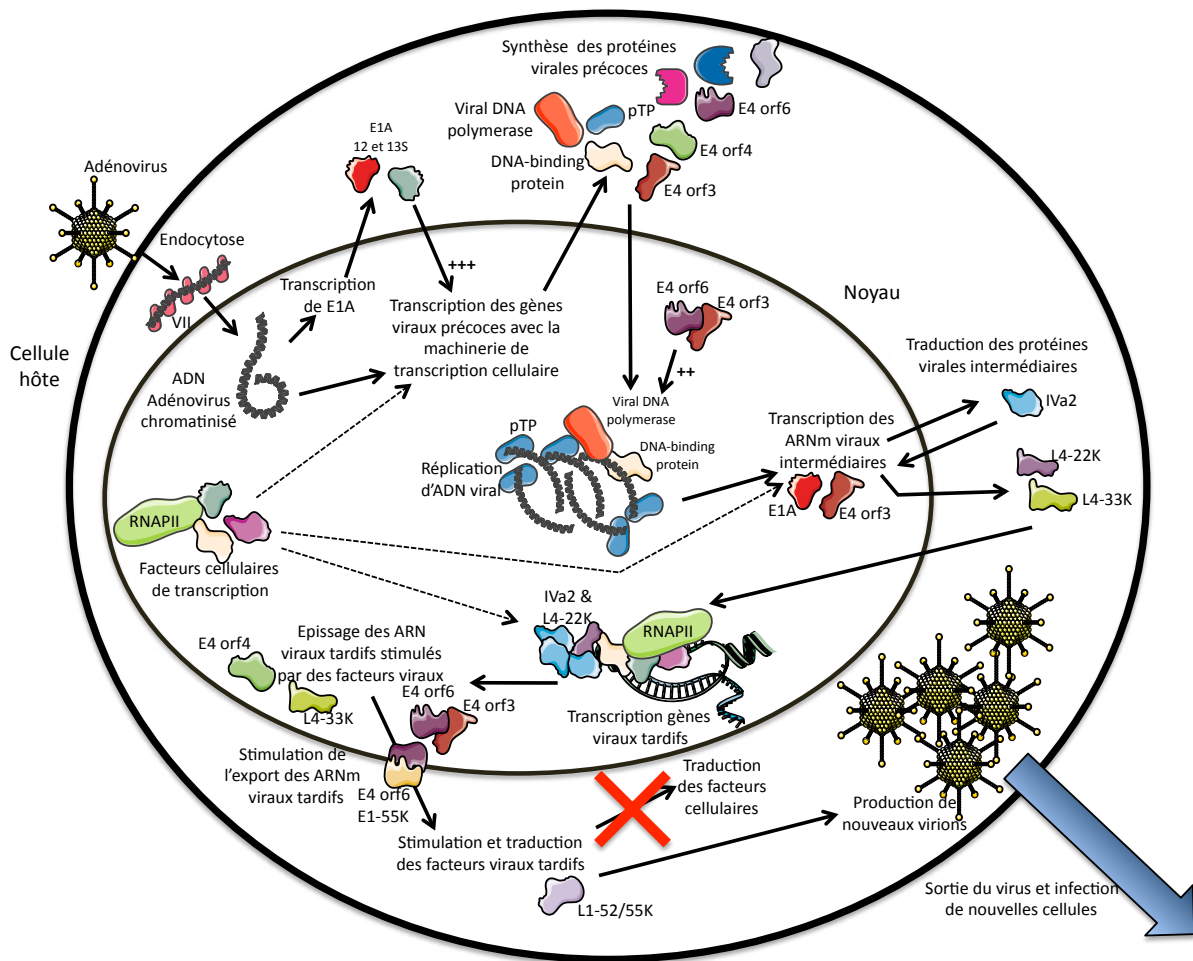


Figure 26 : Représentation schématique du cycle viral d'adénovirus lors de l'infection.

Le cycle commence par l'expression du gène viral précoce E1A qui va stimuler la transcription des autres facteurs viraux précoces. Il interagit avec de très nombreux facteurs cellulaires recrutés sur les promoteurs de ces gènes viraux. Les gènes viraux précoces codent pour des protéines impliquées dans la réplication de l'ADN viral, de l'épissage du contrôle cellulaire, etc.... Simultanément à la réplication de l'ADN, une phase de transition est observée avec la transcription puis l'expression de facteurs viraux intermédiaires IVa2, L4-22K et L4-33K dont les premiers vont stimuler de manière drastique la transcription à partir du promoteur majeur tardif. Au contraire, L4-33K va réguler l'épissage des gènes viraux tardifs. Ces derniers transcrits seront efficacement exportés dans cytoplasme par le complexe E4 orf6 – E1-55K. La protéine virale L1-52/55K va quant à elle, faciliter leur traduction en recrutant la machinerie cellulaire de traduction. Ce phénomène va induire la synthèse des protéines viraux tardives et l'extinction de l'expression des facteurs protéiques cellulaires. Les protéines viraux tardives codent majoritairement pour des protéines de structure qui vont former de nouvelles particules virales dans lesquelles sera encapsidé l'ADN viral à l'aide des facteurs IVa2, L4-22K, L1-52/55K, IIIa et TP. Ces virions seront alors capables d'infecter de nouvelles cellules hôtes.

D) Les *Polyomaviridae* (Polyomavirus).

Les polyomavirus font partis d'une seule et même famille mais sont divisés en de très nombreuses espèces. Ce sont de petits virus non enveloppés, à capsidie icosaédrique de 40 à 50 nm de diamètre, capables d'infecter un large spectre d'hôtes, dont l'homme. On estime que près de 90% de la population mondiale adulte est infectée de manière persistante par ce type de virus (Knowles et al., 2003). Douze polyomavirus humains ont été caractérisés à ce jour : dans les années 1970 le virus JC (JCV) et le virus BK (BKV), et au cours de années 2000 le KI polyomavirus, WU polyomavirus et polyomavirus des cellules de Merkel entre autres (Feng et al., 2008; van der Meijden et al., 2010; Schowalter et al., 2010; White et al., 2013). Ces virus ont un tropisme large mais infectent majoritairement de manière latente les cellules épithéliales rénales, urinaires (BKV), les lymphocytes du sang périphérique ou encore certaines cellules du système nerveux central (JCV). De manière générale, l'infection par ces virus est asymptomatique. Chez les personnes immunocompétentes, peu de pathologies leurs sont associées. À l'inverse, chez les personnes immunodéprimées (séropositives à HIV et les transplantées), les polyomavirus sont associés à des néphropathies, leuco-encéphalopathies (démyélinisation de neurones), ainsi que différents cancers. Des épisodes de réactivation virale sont observés dans la population, permettant aux virus de se propager un peu partout, y compris dans le système nerveux central (oligodendrocytes et astrocytes pour JCV) et d'être détecté dans les urines (JCV et BKV). Comme la diversité de polyomavirus est très large, nous nous focaliserons sur les trois virus les plus étudiés : chez l'homme JCV et BKV, chez le singe, le best-seller SV40 qui fut massivement étudié au cours des années 1980-1990. Ce dernier a grandement participé à la compréhension des mécanismes de régulation de la transcription (Eash et al., 2006; White et al., 2009).

1) Généralités.

Les polyomavirus ont un ADN double brin circulaire d'environ 5100 pb. Le génome de ces virus est sous forme de minichromosome, associé aux histones cellulaires H2A, H2B, H3 et H4, y compris dans la capsidie. Leurs génomes sont divisés en deux régions codantes : une précoce et une tardive, toute deux transcrites en direction opposée. Ces deux régions codantes sont très conservées et sont séparées l'une de l'autre par une région régulatrice non

codante (NCCR), hypervariable d'un type de virus à l'autre, permettant de distinguer les virus archétypes des réassortis (Frisque et al., 1984; Seif et al., 1979).

La région précoce de 2,4 kpb est la première à être transcrite en un unique pré-ARNm. Celui-ci sera épissé de telle manière qu'il donnera naissance à deux protéines virales majeures régulatrices retrouvées chez tous les polyomaovirus : l'antigène petit t (t-Ag) et l'antigène grand T (T-Ag). D'autres isoformes protéiques issus d'un même pré-ARNm différemment épissé ont été identifiés chez plusieurs polyomavirus mais le nombre de transcrits est variable d'un virus à l'autre (cf. figure 26) (Abend et al., 2009). Ces différentes isoformes protéiques ont pour rôle probable de contrôler l'expression de l'antigène T mais aussi d'avoir des fonctions redondantes avec ce dernier ou le suppléer ou encore de déréguler le cycle cellulaire. Un micro-ARN est également exprimé dans cette région et est parfaitement complémentaire aux ARNm viraux précoces, suggérant une régulation négative de leur expression, du moins chez SV40 (Sullivan et al., 2005). Une étude récente tend à démontrer que le micro-ARN de BKV serait exprimé à partir du promoteur tardif mais avant la réplication de l'ADN viral et régulerait négativement la réplication de l'ADN viral pour maintenir une infection persistante (Broekema and Imperiale, 2013). Ainsi, les micro-ARN de polyomavirus auraient pour rôle de moduler l'expression de l'antigène T et de contrôler la réplication de l'ADN viral (Lagatie et al., 2013).

Suite à l'expression des gènes viraux précoces et la modification de l'environnement cellulaire, la réplication de l'ADN viral débute au niveau du site d'initiation de la réplication (ori), présent dans la région non codante NCCR. La protéine virale T-Ag lie cette ori et initie la réplication. Certains facteurs cellulaires participent à ce processus : NF-1 chez BKV, ou NF-1 et Sp1 chez SV40 en modifiant la structure chromatinienne (Cheng and Kelly, 1989; Kelly, 1988; Liang et al., 2012). L'antigène T recrute la machinerie cellulaire (ADN polymérase et primase) pour répliquer l'ADN viral de manière bi-directionnelle. Une autre protéine virale (t-Ag) participe à ce mécanisme par un système encore mal connu. Suite à cette réplication de l'ADN viral, l'activation du promoteur viral tardif par T-Ag et des facteurs de transcription cellulaire associés conduit à l'expression des gènes viraux tardifs.

La région codante tardive s'étend sur 2,3 kpb et code pour les protéines de capsides VP1, VP2 et VP3 mais aussi pour une protéine régulatrice nommée agnoprotéine. VP1, VP2 et VP3 proviennent d'un même ARNm précurseur épissé (cf. figure 27).

La région régulatrice non codante (NCCR), d'environ trois cent à cinq cent nucléotides s'étend entre les régions précoces et tardives (cf. figure 27). Elle contient les promoteurs et « enhancers » nécessaires à l'expression des gènes viraux précoces et tardifs. Ces éléments sont composés de très nombreuses séquences cis-régulatrices sur lesquelles se fixe une variété de facteurs de transcription cellulaires, dont certains sont exprimés de manière ubiquitaire alors que d'autres sont exprimés de manière tissu spécifique (dans le cas de JCV (Beggs et al., 1988)). La région située du côté précoce est plutôt conservée entre virus et ne subit pas de réarrangements, contrairement à la région située du côté tardif. Les variants réarrangés sont caractérisés par des délétions et duplications de blocs de séquences chez les virus humains tels que JCV et BKV. Compte tenu du fait que certains facteurs sont utilisés pour la transcription précoce et tardive, nous allons décrypter le NCCR des polyomavirus, en particulier celui de SV40, en détaillant précisément les éléments présents dans chaque région afin de définir leur rôle.

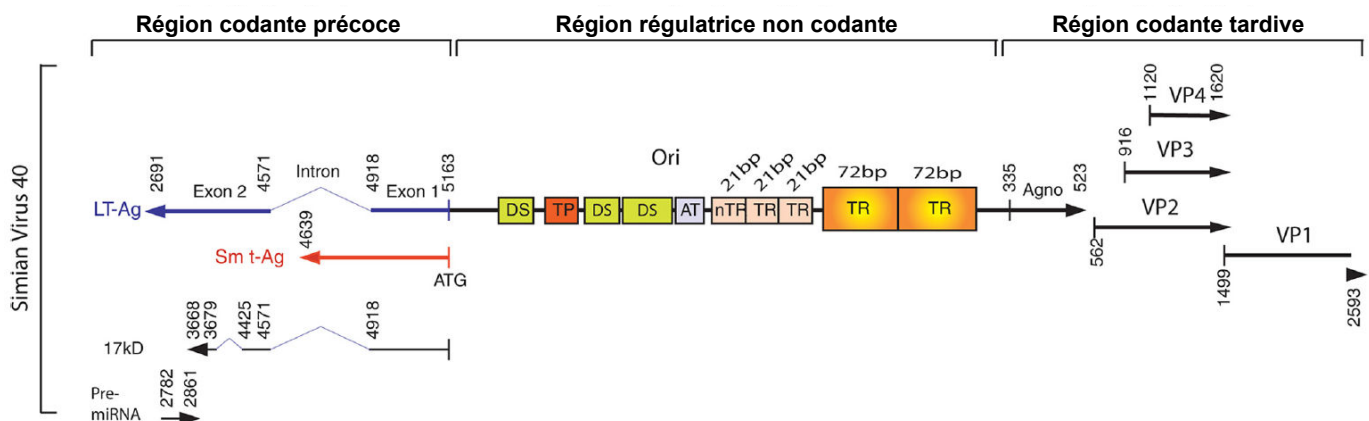


Figure 27 : Représentation schématique du génome du polyomavirus de primates SV40 (Simian Virus 40), tiré de White et al., 2009.

Les génomes de polyomavirus sont globalement constitués de la même manière : une région précoce représentée à gauche qui code pour différentes protéines issues d'un même pré-ARNm épissé. Ici, au milieu, est représentée la région régulatrice non codante (NCCR), différente entre espèces virales.. Il est constitué de séquences répétées, de palindromes (TP), d'une séquence riche en A/T correspond à la boîte TATA (AT), d'une origine de réplication (Ori), de sites de fixation pour des facteurs de transcription, et d'éléments « enhancers » de plusieurs dizaines de paires de bases nommés TR qui module la transcription virale de la région précoce et tardive. Chez les virus infectant les primates humains, cette région NCCR est hypervariable d'un même virus à l'autre, en fonction du tropisme et des pathologies associées Enfin, à droite figure la région codante tardive qui code pour trois protéines de structure et une protéine régulatrice Agnoprotéine.

2) La région précoce.

SV40

Dans le cas de SV40, la région NCCR côté précoce (située au centre, cf. figure 28), possède une TATA-box retrouvée juste avant le TSS (position -21 à -27 avant le TSS) nécessaire pour définir le site d'initiation de la transcription. Toutefois sa délétion n'entraîne pas de diminution de transcription. Un peu plus en amont, de la boîte TATA, trois blocs de répétitions de vingt-et-une paires de bases sont présents (cf. figure 28). Cette région très riche en G/C (Hertz and Mertz, 1988), est importante pour la fixation du facteur de transcription cellulaire Sp1 puisque chaque bloc possède deux sites de liaison à Sp1. Ces séquences sont impliquées dans la transcription des gènes viraux précoces et tardifs, mais de manière différente en fonction des sites, et est régulée par l'antigène T (Barrera-Saldana et al., 1985; Gidoni et al., 1985; May et al., 1992). De plus, le facteur Sp1 semble induire un remodelage de la chromatine virale permettant ainsi une transcription des gènes viraux précoces, tout comme le facteur AP-1, fixé sur une séquence cible située côté tardif (Milavetz, 2002).

De manière intéressante, cette région NCCR est également associée à une hyperacétylation des histones H3 et H4, en présence de l'ARN pol II, qui est réalisée par l'histone acétyltransférase p300 (Balakrishnan and Milavetz, 2007).

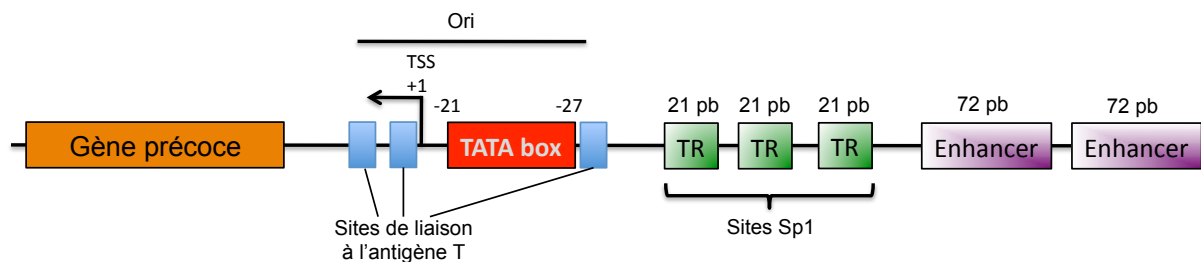


Figure 28 : Représentation schématique de la structure du promoteur précoce de SV40.

Le promoteur des gènes viraux précoces est sous la dépendance d'une boîte TATA qui recrute les facteurs généraux de la transcription dont TFIID. Plus loin, trois blocs de vingt-et-une paires de bases sont les cibles de fixation du facteur cellulaire Sp1. Toutefois, tous ne sont pas essentiels à la transcription du promoteur précoce. Plus en amont, deux séquences répétées de 72 pb jouent le rôle d'enhancer. De plus, trois sites de liaison pour l'antigène T sont présents autour de la TATA-box et vont être utilisés par l'antigène T pour bloquer la transcription des gènes viraux précoces et favoriser la réplication de l'ADN viral.

Du côté tardif (cf. figures 27 et 28), un « enhancer » constitué de deux séquences répétées de 72 pb chacune est présent et sont elles, absolument requises pour l'expression des gènes viraux précoces (Barrera-Saldana et al., 1985; Benoist and Chambon, 1981; Dynan and

Tjian, 1983a, 1983b). Une fois exprimé, l'antigène T se lie sur différents sites du NCCR, proches du TSS, et génère un rétro-contrôle négatif de son expression ainsi que des autres ARNm viraux précoces en créant probablement un encombrement stérique autour de la TATA-box (Myers et al., 1981; Rio and Tjian, 1983). Au contraire, T-Ag stimule la réplication de l'ADN viral et active l'expression des gènes tardifs de SV40 (indépendamment de la réplication) (Brady and Khoury, 1985; Brady et al., 1984, 1985; Keller and Alwine, 1984).

BKV & JCV

Dans le cas de BKV et JCV, la structure du promoteur est plus complexe du fait des réarrangements localisés dans l' « enhancer », situé du côté tardif. Cette structure a donc été divisée en cinq blocs distincts chez ces virus, nommés O-P-Q-R-S, où O est l'élément précoce conservé entre virus (Moens et al., 1995). Les éléments « enhancers » P, Q, R et S sont très variables entre souches (Moens and Van Ghelue, 2005; Moens et al., 1995). Dans le cas de JCV, la partie variable du côté tardif confère même une expression tissu-spécifique de ces gènes viraux précoces et tardifs (Lashgari et al., 1989).

Ces deux virus possèdent un promoteur précoce très complexe, composé d'une multitude de séquences de fixation pour de très nombreux facteurs cellulaires qui vont activer la transcription. Ces facteurs impliqués peuvent être présent de manière ubiquitaire alors que d'autres le sont uniquement dans certains types cellulaires. Une régulation au niveau de l'état de la chromatine permet en plus, de contrôler l'expression des gènes viraux précoces (Wollebo et al., 2013).

3) La région tardive.

Du côté tardif, c'est-à-dire au niveau de la région dégénérée, les polyomavirus possèdent des éléments répétés en tandem, variables, ayant un rôle d' « enhancer » de la transcription pour la région précoce et de promoteur/ « enhancer » des gènes tardifs.

SV40

Dans le cas de SV40, deux éléments répétés de 72 pb chacun servent de promoteur/ « enhancer » pour l'expression des gènes viraux précoces et tardifs. Ils ont été divisés en plusieurs domaines fonctionnels, parfois sous-divisés en différents motifs où coopèrent plusieurs facteurs spécifiques (Ernoul-Lange et al., 1987; Ondek et al., 1988; Zenke et al., 1986). Ces éléments répétés sont retrouvés juste après les répétitions de 21 pb (liées par Sp1 et nécessaires à l'expression des gènes viraux précoces et tardifs) et sont fixées par de nombreux facteurs de transcription cellulaire (cf. figure 29) : AP-1 (Lee et al., 1987), AP-2 (Mitchell et al., 1987), AP-4 (Mermod et al., 1988), NF- κ B (Kanno et al., 1989), TEF-1 ou encore NFAT3 (Nuclear Factor of activated T cells 3) NFAT4 (Manley et al., 2008). Ces deux derniers facteurs lient les deux sites de fixation de NF- κ B situés dans la région « enhancer ». Leur absence empêche l'expression des gènes viraux précoces et tardifs de manière différente en fonction du type cellulaire. L'activité transcriptionnelle induite par ces facteurs NFAT est augmentée par l'antigène T (Manley et al., 2008).

Dans le cas de AP-1, un site de fixation pour cette protéine est présent dans la région « enhancer » et stimule la transcription des gènes viraux précoces *in vitro* avec de la protéine purifiée (Lee et al., 1987).

AP-2 se fixe sur un site au niveau de l' « enhancer » et un des sites de Sp1 afin de coopérer ensemble. Même si son rôle semble limité, AP-2 est inhibé par T-Ag, bloquant de ce fait son rôle d'activateur de la transcription (Kanno et al., 1989; Mitchell et al., 1987) (cf. figure 29).

AP-4, elle se fixe à un site en amont de l'élément « enhancer » et coopère avec AP-1 pour activer l'expression des gènes viraux tardifs *in vitro* (Mermod et al., 1988). Bien que son rôle dans l'expression des gènes tardifs soit mal caractérisé, il semble que AP-4 soit impliquée dans le remodelage de la chromatine afin de permettre la transcription de ces gènes (Milavetz, 2002).

Plusieurs sous-unités du facteur protéique NF- κ B se fixent sur deux séquences similaires et adjacentes à celles de AP-2 et sont requises pour augmenter l'activité transcriptionnelle (Kanno et al., 1989; Macchi et al., 1989). Ainsi, l'interaction de tous ces facteurs de transcription à leurs motifs respectifs est complexe et implique une coopération et une organisation fonctionnelle pour réguler l'activité transcriptionnelle (Fromental et al., 1988).

Par ailleurs, le facteur « enhancer » de transcription cellulaire TEF-1 possède trois sites de liaison à l'ADN dans la région « enhancer » et promotrice de SV40. Ces trois sites

sont importants pour la transcription des gènes viraux tardifs assurée par t-Ag et T-Ag (Casaz et al., 1991). En effet, si TEF-1 réprime la transcription des gènes tardifs, l'antigène T en interagissant avec TEF-1, l'empêche de se fixer à l'ADN et lève ainsi son rôle inhibiteur (Berger et al., 1996).

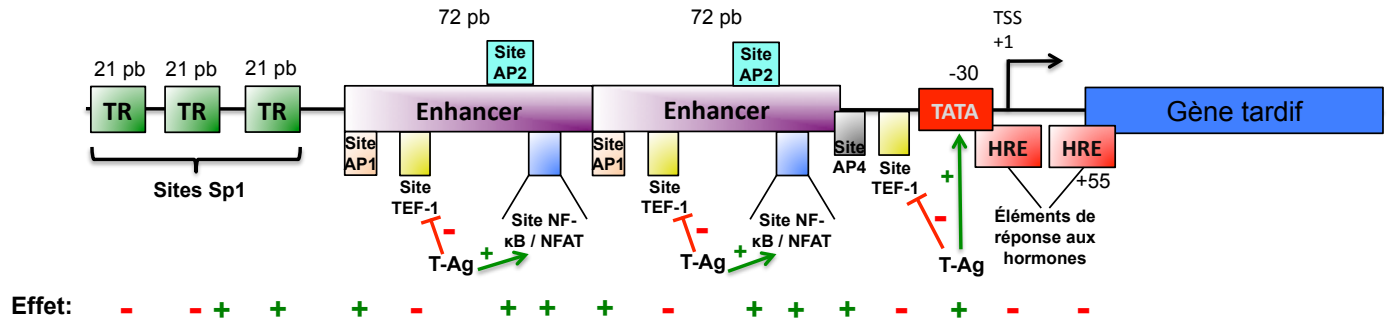


Figure 29 : Représentation schématique du promoteur/ « enhancer » tardif de SV40.

Le promoteur/ « enhancer » tardif de SV40 est très complexe. Il est composé de différents sites de fixation pour de nombreux facteurs cellulaires. Parmi ceux-ci, Sp1, AP-1, AP-2, AP-4, NF-κB et NFAT vont activer la transcription. La séquence TATA va permettre la fixation de TBP et l'assemblage du PIC, favorisé par le recrutement des facteurs de transcription. À l'inverse, TEF-1 et les éléments de réponse aux hormones HRE fixés par plusieurs facteurs (NR2F1, NR2F2, NR2C1, NR2C2, EE1 et TRα/RXRα) vont inhiber la transcription des gènes tardifs. Toutefois, l'antigène T va relever l'inhibition induite par la fixation en de TEF-1 en bloquant sa fixation alors que l'ajout d'hormone va relever la fixation des facteurs fixés sur les séquences HRE.

Toutefois, d'autres facteurs permettent à l'antigène T d'activer l'expression des gènes tardifs (Casaz et al., 1995). En effet, une boîte non canonique TATA est située une trentaine de nucléotides en amont du TSS (cf. figure 29) où sont recrutés les facteurs généraux de transcription, dont TFIID (Wiley et al., 1992). De manière intéressante, TBP et l'antigène T interagissent ensemble (Martin et al., 1993). Une étude plus récente a ainsi démontré que l'antigène T interagit efficacement avec TBP, TFIIB, l'ARN pol II et Sp1 afin de stabiliser le complexe. Ces interactions tendent à démontrer un mécanisme d'augmentation de la transcription des gènes viraux tardifs de SV40 menée par l'antigène T (Johnston et al., 1996). Une étude plus approfondie a démontré que l'antigène T était même capable de fixer plusieurs TAFs en plus de TBP, et semble pouvoir compléter fonctionnellement TAF1. À partir de ces données, les auteurs de cette étude en viennent à conclure que T-Ag est une TAF-like car incorporée dans TFIID (Damania and Alwine, 1996). De plus, il a été constaté que l'antigène T stabilisait le complexe TBP/TFII-A sur la TATA-box, suggérant ainsi une capacité à augmenter le taux de transcription et confirmant son rôle de TAF-like (Damania et al., 1998) en assemblant plus rapidement le PIC. Néanmoins, ce rôle dans la transcription se fait au

détriment de la réplication, puisque l'interaction entre TBP et l'antigène T inhibe la réplication de l'ADN viral (Herbig et al., 1999).

Deux autres régions, situées à cheval sur le TSS tardif et environ +55 nucléotides après ce même TSS sont fixées par des facteurs répresseurs (cf. figure 29). Cette région est en fait composée d'élément de réponse aux hormones (HRE) sur lesquelles peuvent se fixer les récepteurs nucléaires NR2F1 et NR2F2 (sur le TSS et le site + 55), NR2C1 et NR2C2 (sur le site +55), le récepteur d'hormone ERR1 (sur le +55), ou encore l'hétérodimère TR α /RXR α (sur le TSS) (respectivement Récepteur α d'hormone thyroïdienne et Récepteur α d'acide rétinoïque X) qui vont inhiber l'expression des gènes viraux tardifs en se fixant sur leurs séquences cibles (Lee and Chang, 1995; Lee et al., 1995; Zuo and Mertz, 1995; Zuo et al., 1997). L'inhibition de l'expression des gènes viraux tardifs induite par TR α /RXR α est dû à l'impossibilité de former le complexe de pré-initiation de la transcription fonctionnel sur les promoteurs des gènes tardifs (Zuo et al., 1997). Cette incapacité à transcrire peut être levée par l'ajout d'hormone thyroïdienne ou bien, par la réplication de l'ADN viral (Wiley et al., 1993). Néanmoins, une étude plus approfondie a démontré que l'absence de l'élément de réponse aux hormones au niveau du promoteur tardif conduit à un retard et une diminution de l'expression des gènes viraux précoces, de la réplication de l'ADN viral mais aussi à une accumulation des ARNm viraux tardifs au début de l'infection. Ces constatations peuvent s'expliquer par la levée de l'inhibition de la transcription des gènes tardifs, et par une compétition de recrutement des facteurs généraux de transcription entre les promoteurs précoce et tardif (Farrell and Mertz, 2002).

BKV & JCV

Contrairement à SV40, les éléments de réponses aux stéroïdes retrouvés chez BKV et JCV active la transcription des gènes tardifs (Moens et al., 1994). De plus, en prenant exemple sur JCV, les répétitions en tandem de 98 nucléotides servant de promoteur/ « enhancer » de la transcription (cf. figure 30) sont responsables, en partie, du tropisme viral dans les cellules gliales et sont nécessaires pour le virus (DANIEL et al., 1996). Des facteurs de transcription généraux (parmi lesquels YB-1 (Kerr et al., 1994), Pura α (Chen and Khalili, 1995)) et tissus spécifiques (tels que des membres de la famille de NF-1 (Amemiya et al., 1989, 1992) et Tst-1 (Renner et al., 1994; Wegner et al., 1993)) sont

retrouvés fixés sur les séquences « enhancers » de JCV (cf. figure 30) (Kenney et al., 1984; Khalili et al., 1988).

L'expression des gènes viraux tardifs est dépendante du tissu cellulaire et est stimulée par l'antigène T uniquement dans des cellules gliales à l'aide de facteurs cellulaires tissu-spécifique (Lashgari et al., 1989). En effet, l'antigène T coopère avec le facteur cellulaire Tst-1 (ou Oct-6) pour activer synergiquement l'expression des gènes tardifs dans les cellules gliales (cf. figure 30) (Renner et al., 1994). T-Ag est également capable d'interagir avec des facteurs cellulaires exprimés de manière ubiquitaire (YB-1 et Pur α (protéine de liaison à l'ADN simple brin, impliquée dans la réplication de l'ADN et l'expression des gènes précoces)) pour former un ménage à trois avec ces protéines. Si Pur α stimule l'expression des gènes précoces et réprime celle des gènes tardifs en recrutant T-Ag (Gallia et al., 1998), l'association de l'antigène T avec YB-1 augmente drastiquement l'expression des gènes tardifs en prenant la place de Pur α sur sa séquence cible (cf. figure 30) (Chen and Khalili, 1995; Safak et al., 1999a, 1999b).

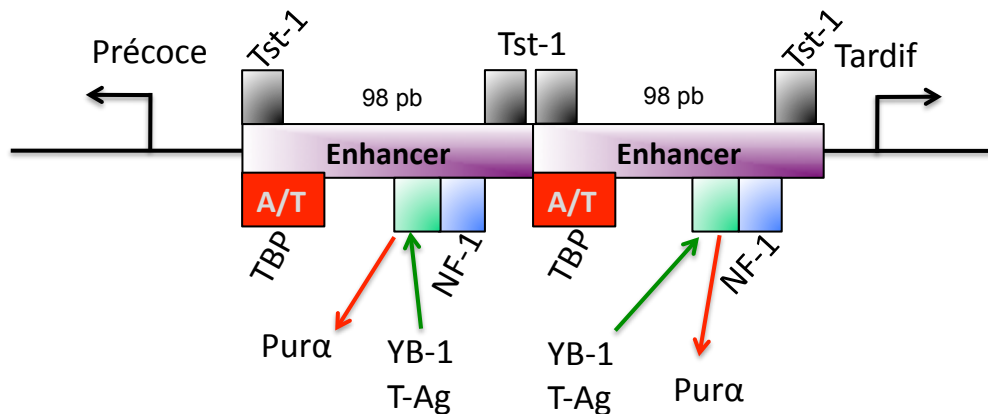


Figure 30 : Représentation schématique de la régulation du promoteur de JCV par différents facteurs cellulaires et viraux.

Plusieurs facteurs cellulaires vont se fixer dans la région « enhancer » de 98 pb pour activer l'expression des gènes tardifs (Tst-1, NF-1 et TBP sur la boîte TATA (A/T)). La fixation de Pur α sur sa séquence sera relevée grâce à l'arrivée de YB-1 associée à T-Ag afin d'activer l'expression des gènes viraux tardifs.

Outre des facteurs activateurs, certains inhibent l'expression des gènes tardifs comme par exemple des facteurs membres de la famille AP-1, se fixant dans les « enhancers » (Sadowska et al., 2003). Ils inhibent le rôle transcriptionnel de l'antigène T, bloquant ainsi l'expression des gènes tardifs (Kim et al., 2003). C'est aussi le cas d'un facteur membre de la famille de NF-1 en se fixant à des séquences adjacentes aux TATA-box (Tada et al., 1991),

tout comme SRSF1 qui se fixe dans l' « enhancer » (Sariyer and Khalili, 2011). Toutefois, l'effet de SRSF1 peut être levé par la présence de T-Ag qui bloque son expression (Craigie et al., 2015).

En plus d'être régulé par des facteurs cellulaires, il semble que la protéine virale non structurale Agnoprotéine réprime l'expression des gènes viraux tardifs de BKV et JCV. En effet l'agnoprotéine de JCV est capable d'interagir avec les protéines YB-1 et l'antigène T afin de bloquer la transcription des gènes viraux tardifs ainsi que la réplication de l'ADN viral (Safak et al., 2001, 2002). De plus, les agnoprotéines de BKV et JCV sont phosphorylées par la serines thréonines kinase cellulaire PKC (Johannessen et al., 2008; Sariyer et al., 2006). Pour BKV, la phosphorylation d'un des sites de l'agnoprotéine entraîne une inhibition de la transcription des gènes viraux tardifs (Johannessen et al., 2008). En revanche, pour celle de JCV, c'est l'absence de phosphorylation qui entraîne un défaut de réplication d'ADN viral et de production de particules virales infectieuses (Sariyer et al., 2006). Dans la continuité de cette régulation, la phosphatase cellulaire PP2A interagit avec l'agnoprotéine pour la déphosphoryler. Ensemble, elles interagissent avec t-Ag au niveau du même domaine, suggérant ainsi une compétition entre ces deux premiers facteurs. De manière intéressante, l'antigène t est capable d'inhiber directement la déphosphorylation de l'agnoprotéine réalisée par PP2A. Ainsi, l'antigène t bloque le rôle de PP2A et permet la phosphorylation de l'agnoprotéine. Cet enchainement permet de restaurer la réplication de l'ADN viral ainsi que l'expression des gènes tardifs (Bollag et al., 2010; Sariyer et al., 2008).

Ainsi, à travers l'ensemble de ces données, il apparaît que les mécanismes de régulation de l'expression des gènes de polyomavirus sont complexes et s'effectuent au niveau transcriptionnel et épigénétique. Les promoteurs des gènes viraux sont situés dans une région non codante, dans laquelle se fixe de très nombreux facteurs de transcription cellulaires activateurs ou répresseurs. L'expression des facteurs viraux précoces va être induite par différentes protéines cellulaires recrutées sur le promoteur précoce, qui vont également jouer sur l'état de la chromatine pour mener à un environnement transcriptionnellement actif. Les produits précoces, dont l'antigène T, vont réaliser un rétro-contrôle négatif de leur expression. L'antigène T va encombrer stériquement le promoteur précoce en se liant à ses sites de fixation au niveau de l'Ori qui chevauche le promoteur. Cette fixation va induire la réplication de l'ADN viral qui est un tremplin pour l'expression des gènes viraux tardifs. L'augmentation du nombre de copies de génomes viraux va titrer les facteurs cellulaires inhibiteurs de la

transcription présents sur l'ADN. Cet évènement va ensuite permettre le recrutement de protéines activatrices, dont les facteurs généraux de la transcription cellulaire, et avec en maître d'œuvre l'antigène T, qui va amplifier la transcription des gènes viraux tardifs. Ces derniers codent pour des protéines de structure et vont permettre la production de particules virales infectieuses.

E) Les Papillomaviridae (Papillomavirus).

Classés à l'origine avec les polyomavirus dans la famille des *Papovaviridae*, les Papillomavirus font partie d'une branche distincte depuis l'an 2000. Les papillomavirus infectent la plupart des mammifères et des insectes. Actuellement, plus de deux cent types de papillomavirus sont référencés dont un peu plus d'une centaine est susceptible d'infecter l'homme.

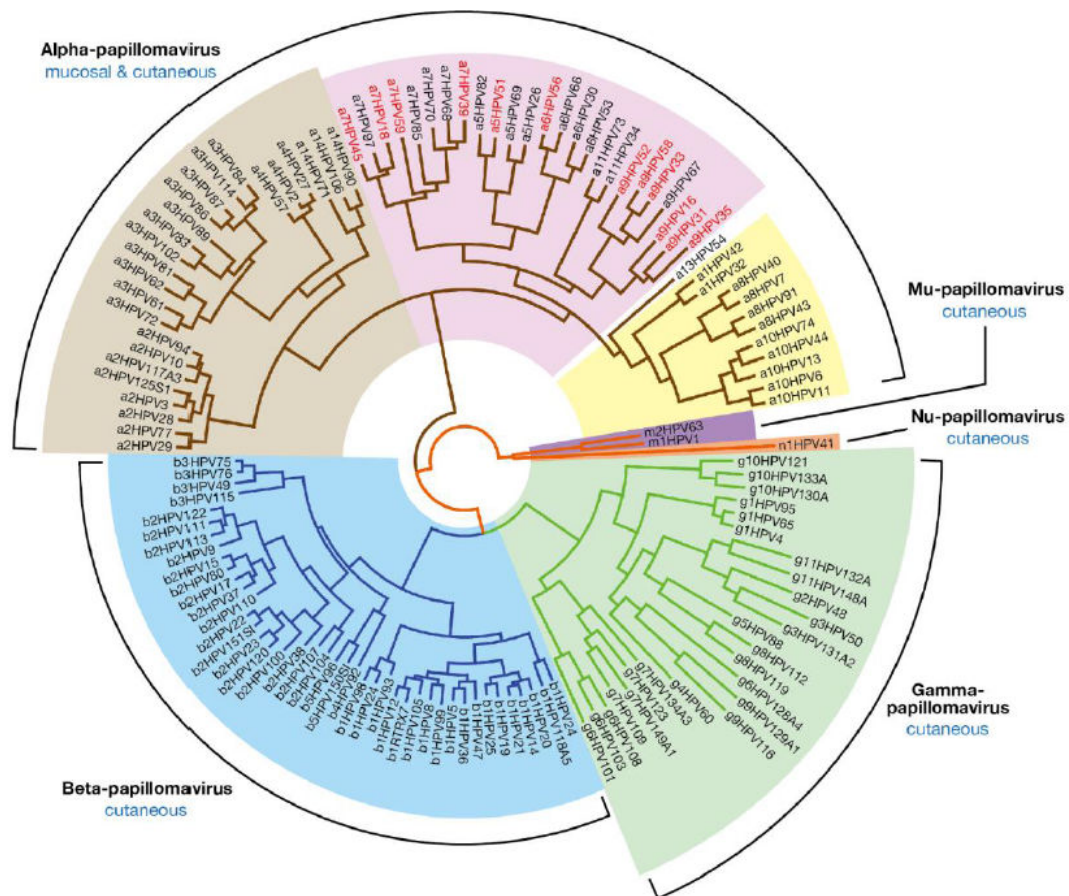


Figure 31 : Arbre phylogénétique des papillomavirus humains, basé sur l'alignement des gènes E1, E2, L1 et L2, tiré de (Egawa et al., 2015).

Ils sont classés en cinq genres avec les Alpha-, Beta- (en bleu) et Gammapapillomavirus (en vert) représentant les plus grands groupes. Les Alpha sont plutôt muqueux (en jaune pour les bas-risques et en rose pour les haut-risques), même si des virus cutanés existent (en marron).

Ces virus ont été divisés en plusieurs genres (cf. figure 31). Parmi ceux-ci, trois contiennent la majorité des papillomavirus infectant l'homme : les Alpha-, les Beta- et les Gamma-papillomavirus, avec un tropisme principalement muqueux (pour les alpha), transmis par voie sexuelle ; et cutané (pour les beta et gamma) qui se transmettent par simple contact cutané (de Villiers et al., 2004).

Les papillomavirus sont les agents causals d'un large panel de pathologies cutanées et épithéliales. Celles-ci s'étendent de la simple verrue jusqu'aux cancers invasifs : cancer du col de l'utérus (dont 55% associés uniquement à HPV-16 ; HPV-18, -31 et -45 étant également des hauts-risques) (cf. figure 31), cancer anal, vulvaire, au niveau de la tête ou encore de la peau avec une association plus ou moins importante à différents papillomavirus (zur Hausen, 2009).

1) Généralités.

Les papillomavirus sont des petits virus non enveloppés, à capsidie icosaédrique. Leur génome est un ADN double brin d'environ 8 kbp, circulaire, organisé de manière assez similaire à celui des polyomavirus. L'expression de leurs gènes est séquentielle et semble être régulée par la différenciation des cellules épithéliales. La plupart des papillomavirus codent pour huit protéines majeures : six protéines précoces et deux protéines tardives. Il existe néanmoins d'autres isoformes protéiques issus de l'épissage d'ARNm polycistroniques. Les protéines précoces majeures E1, E2, E4, E5, E6 et E7 sont des protéines non structurales, impliquées dans diverses fonctions telles que la réplication du génome viral, la transcription, ou encore le contrôle du cycle cellulaire, ou de l'apoptose. Les protéines L1 et L2 sont les deux protéines de structure du virus (cf. figure 32). Les papillomavirus possèdent également une région non codante d'environ 800 nucléotides, nommée LCR (Long Control Region) contenant l'origine de réplication de l'ADN viral ainsi que des sites de liaisons à des facteurs de transcription ayant un rôle d'« enhancer » de la transcription pour le promoteur précoce (Bernard, 2013). Cette région est localisée entre les gènes viraux tardifs et précoces (cf. figure 32). Afin de simplifier la compréhension des HPV et compte tenu de la multitude de ces

virus, nous nous focaliserons sur le plus étudié en raison de son importance médicale : HPV-16.

L'infection par HPV débute par l'entrée de la particule virale dans les cellules épithéliales basales (Raff et al., 2013). Une fois arrivé dans le noyau de la cellule, le virus se maintient sous forme latente avec un faible nombre d'épisomes viraux. Les protéines virales E1 et E2 permettent une ségrégation correcte du génome viral lors de la division cellulaire. La protéine E2, par son interaction avec certains facteurs cellulaires permet le maintien du génome de HPV sur les chromosomes lors de la division cellulaire (You et al., 2004). Dans certains cas (notamment associé aux pathologies), le génome HPV peut s'intégrer dans les chromosomes de l'hôte au niveau de la région de E2, bloquant de ce fait son expression. La perte du régulateur E2 conduit à une surexpression des protéines E6 et E7 qui vont induire la prolifération cellulaire, inhiber l'apoptose, etc... et ainsi mener au développement de cancer (Doorbar et al., 2012).

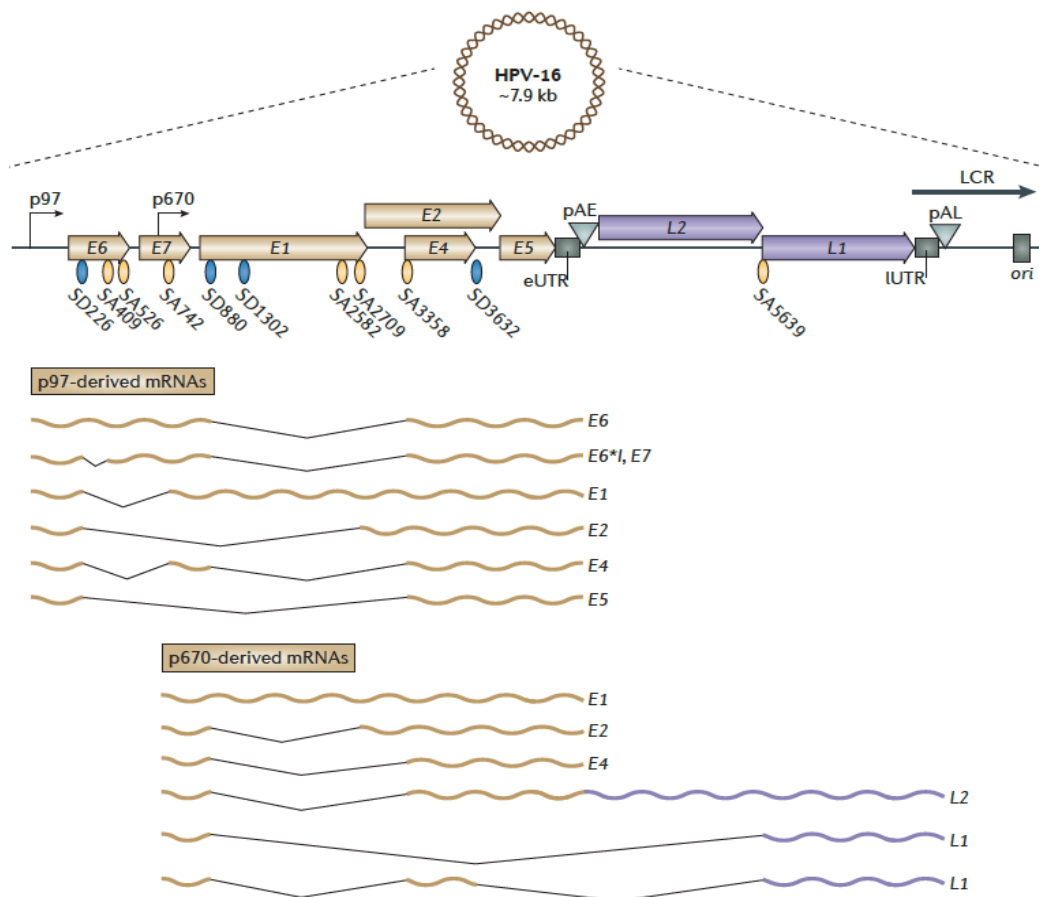


Figure 32 : Représentation schématique linéaire de l'organisation du génome du papillomavirus humain 16 et les ARNm majoritairement synthétisés, tiré de Johansson and Schwartz, 2013.

Le génome de HPV-16 d'environ 8 kpb est circulaire dans les cellules infectées. Le virus code pour six protéines précoces (E1, E2, E4, E5, E6 et E7) et deux protéines tardives (L1 et L2). La région non codante (LCR) sépare le codon stop de L1 de l'ATG de E6 et comprend l'origine de réplication ainsi que de nombreux sites de fixation de facteurs cellulaires de transcription pour initier la transcription à partir du promoteur précoce p97. Le promoteur tardif de HPV-16 est lui localisé en position 670 (p670), dans la séquence codante de E7. Deux sites de polyadénylation, un précoce (pAE) et un tardif (pAL) sont localisés respectivement juste après l'ORF de E5 et la région 3' non codante précoce (eUTR) et juste après l'ORF de L1 et la région 3' non codante tardive (lUTR). Les cercles en jaunes correspondent aux différents sites d'épissages accepteurs (SA) les cercles bleus aux sites donneurs d'épissage (SD). Les ARNm viraux majoritairement transcrits sont représentés en dessous du génome de HPV16. Les régions enlevées par épissage sont représentées avec des traits fins noirs tandis que les séquences codantes sont sous forme de « vagues ».

Dans le cas d'une infection bénigne, le cycle viral est mieux régulé. En effet, l'infection débute toujours au niveau de couche basale de l'épithélium. Chez HPV16, la transcription à partir du promoteur précoce (p97) génère des ARNm codant pour tous les gènes viraux précoces qui sont polyadénylés au niveau du signal précoce de polyadénylation (nommé pAE) (cf. figure 32). Ces ARN polycistroniques, donnent naissance par épissage alternatif aux différents ARNm codant pour les protéines virales précoces. Certains sites d'épissage sont utilisés uniquement pour les gènes viraux précoces alors que d'autres exclusivement pour les tardifs (cf. figure 32).

Lors de la différenciation des kératinocytes, les cellules migrent dans les couches supérieures de l'épithélium. Cette différenciation cellulaire est couplée à une diminution de l'expression de facteurs cellulaires requis pour assurer la réplication de l'ADN viral ou encore de ceux impliqués dans les mécanismes d'épissage et de polyadénylation (cf. figure 33).

Cette diminution est contrecarrée par la protéine virale E7 qui lie la protéine suppresseur de tumeur associée du rétinoblastome (pRB), libérant de ce fait le facteur de transcription E2F pour restaurer l'expression des facteurs de réplication de l'ADN (tels que MCM ou PCNA) (Doorbar et al., 2012). Des mécanismes de rétro-contrôle négatif générés par E2 induisent une diminution de l'expression des protéines précoces E6 et E7, permettant ainsi aux cellules de se différencier et continuer leur cycle. L'induction des gènes viraux tardifs est observée uniquement dans les cellules différenciées de la couche supérieure de l'épithélium. À ce moment, le promoteur tardif p670 de HPV-16 (situé dans la région de E7) est activé. L'activation de ce promoteur bloque l'expression des gènes viraux E6 et E7 et

induit fortement celle des ARNm de E1, E2 et E4 en utilisant les sites précoces d'épissage. Ces ARNm sont également polyadénylés en utilisant le signal du site pAE. La différenciation terminale des cellules infectées diminue l'activité de pAE, conduisant à la transcription de la région tardive du virus, en l'occurrence des ARNm de L1 et L2, suivit par la polyadénylation au niveau du signal tardif de polyadénylation (pAL) pour générer ces ARNm tardifs (cf. figure 32). Chaque signal de polyadénylation est précédé d'une région non codante contenant des éléments qui détermineront l'efficacité de polyadénylation, de traduction et de demi-vie des ARNm transcrits. L'ensemble de ces événements aboutira à la formation de nouvelles particules virales qui seront libérées par lyse mécanique des cellules épithéliales superficielles (cf. figures 32 et 33).

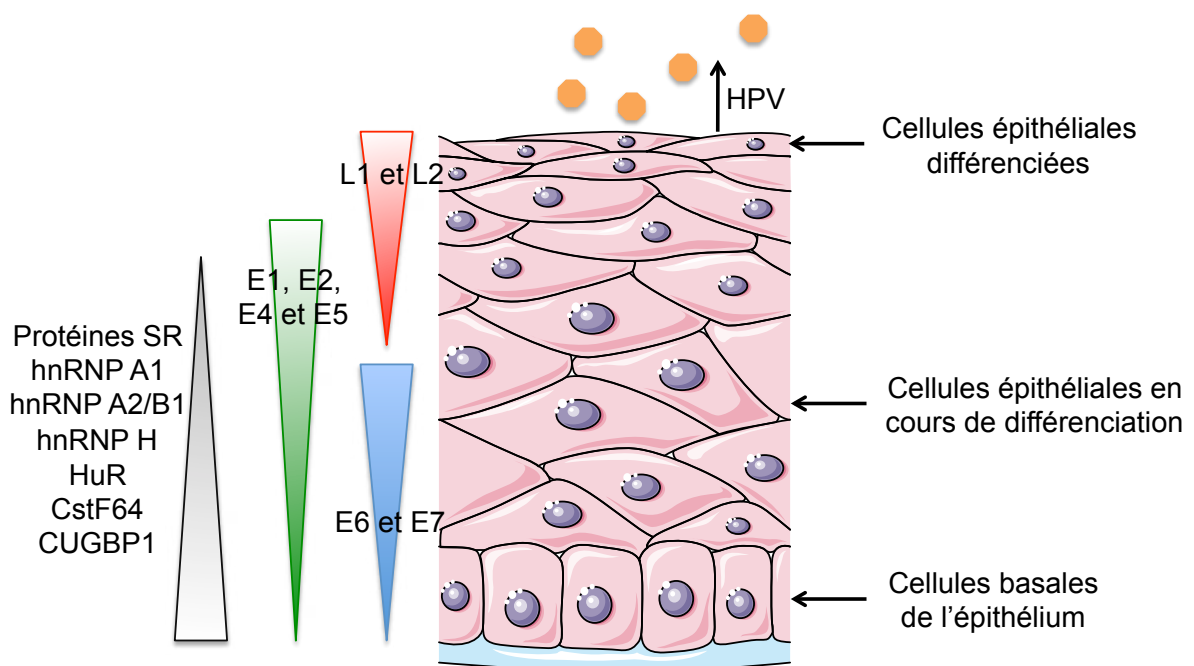


Figure 33 : Représentation schématique du cycle productif des papillomavirus au cours de la différenciation cellulaire des kératinocytes.

Le virus débute son infection au niveau des cellules basales en entrant dans le noyau des cellules. Les protéines précoces E1 et E2, exprimées dans les couches basales de l'épithélium, maintiennent le génome viral et permettent sa réplication. Les protéines précoces oncogènes E6 et E7, également retrouvées dans les couches basales, perturbent le cycle cellulaire et la différenciation. La protéine E4 est exprimée un peu plus tardivement dans les cellules en cours de différenciation pendant que la protéine E2 va réprimer l'expression du promoteur précoce et favoriser l'expression des protéines tardives. Dans les cellules différenciées, certaines protéines précoces sont encore exprimées en compagnie des protéines tardives L1 et L2. Ce cycle viral est régulé par différents facteurs cellulaires d'épissage et de polyadénylation, qui ont tendance à favoriser l'expression des ARNm viraux précoces. Leur extinction au cours de la différenciation cellulaire et donc du cycle viral va permettre l'expression des gènes viraux tardifs.

À travers ces généralités sur le cycle cellulaire viral, on comprend que les mécanismes de régulation de l'expression des gènes viraux précoces et tardifs est très complexe et s'effectuent à plusieurs niveaux : transcriptionnel, de l'épissage et de la polyadénylation.

2) La transcription virale.

a) Structure du promoteur précoce.

Le cycle viral productif des papillomavirus débute par l'expression des gènes viraux précoces, sous contrôle de leurs promoteurs. Le promoteur précoce a été cartographié pour différents papillomavirus, toujours en amont de E6. Pour HPV-16 (cf. figure 34), un promoteur majeur (p97) a été identifié avec une TATA-box localisée une trentaine de nucléotides en amont de ce dernier (Smotkin and Wettstein, 1986), mais aussi pour HPV-31 avec p99 (Hummel et al., 1992) et HPV-18 avec le p105 (Thierry et al., 1987; Wang et al., 2011b) bien que d'autres TSS aient été observés.

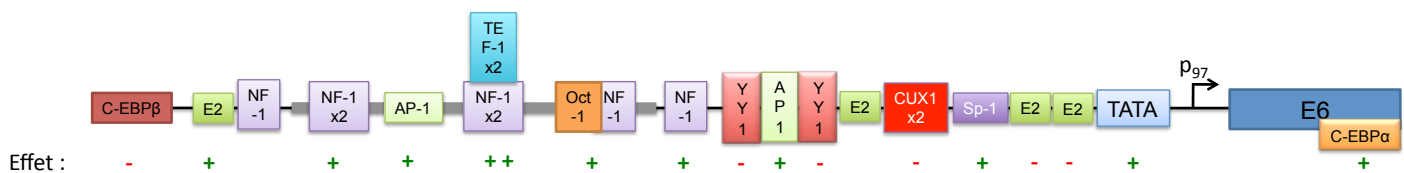


Figure 34 : Représentation schématique du promoteur précoce p97 de HPV-16.

Le promoteur de HPV-16 possède de très nombreux sites de fixation pour des facteurs de transcription cellulaires, activateurs ou répresseurs. Chaque boîte représente la fixation du facteur correspondant sauf pour la TATA-box recrutant la TBP et le reste des facteurs généraux de transcription. Les éléments de E2 situés proches de la TATA-box inhibent la transcription en encombrant le promoteur au contraire des deux autres. La région épaisse en gris correspond à la séquence « enhancer » de la transcription. En dessous, les «+» signifient activateur alors que les «-» correspondent à un effet inhibiteur de la transcription.

Ces promoteurs sont, quels que soient le type de HPV, composés de très nombreux sites de fixation, pour des facteurs de transcription cellulaire activateurs, situés en amont du TSS, tels que AP-1 (Chan et al., 1990; Chong et al., 1991), dont le site est superposé à ceux des éléments de réponse aux glucocorticoïdes. Sp-1 qui permet un meilleur recrutement des facteurs de transcription généraux ainsi qu'une modification de la chromatine (Wang et al., 2011a). Des sites pour NF-1 (Apt et al., 1993), NFAT, Oct-1 (O'Connor and Bernard, 1995), TEF-1 (Ishiji et al., 1992), Sp-1 ou encore YY1 (Dong et al., 1994b) ont été caractérisés (cf.

figure 34). En aval, d'autres facteurs activateurs se fixent sur des séquences précises : C/EBP α (Krüppel et al., 2008). Le promoteur précoce de HPV-16 contient également dans le LCR, des séquences de fixation pour le facteur C/EBP β qui inhibe la transcription dans cette situation (Kyo et al., 1993) tout comme pour CUX-1 (cf. figure 34) (O'Connor et al., 2000). De plus, TBP travaille en coopération avec E2 et Sp-1 pour activer la transcription efficacement (Ham et al., 1994). Par ailleurs, une faible expression de E2 peut stimuler la transcription du promoteur précoce en interagissant avec p300, qui est une histone acétyltransférase (Krüppel et al., 2008), d'autant plus que la protéine p300 est fortement exprimée dans les cellules basales de l'épithélium infectées. Aussi, au cours de la différenciation cellulaire, le niveau d'acétylation des histones H3 et H4 présents sur le génome de HPV est augmenté et est corrélié à l'augmentation de la fixation des facteurs C/EBP α (autour du promoteur tardif) et C/EBP β plus YY1 autour du promoteur précoce (Wooldridge and Laimins, 2008). Cette publication suggère donc une modification du recrutement des facteurs cellulaires au cours de la différenciation cellulaire et probablement une modification de la chromatine nécessaire à l'activation du promoteur tardif.

Toutefois, la protéine virale E2 n'est pas qu'activatrice. En effet, il a été démontré que E2 est capable de se fixer sur des séquences d'ADN en dimère via son domaine de liaison à l'ADN, et possède quatre sites de fixation chez HPV-16. Deux sont localisés entre le site Sp-1 et la TATA-box chez HPV-16 et 18, un est situé environ 100 pb en amont du TSS et un dernier environ 500 nucléotides en amont (Romanczuk et al., 1990). De manière intéressante, la fixation de E2 sur les séquences adjacentes à la TATA-box réprime drastiquement le promoteur précoce p97 de HPV-16 en déplaçant la protéine Sp-1 (Tan et al., 1992, 1994). Ce mécanisme est retrouvé et conservé entre papillomavirus (Dong et al., 1994a; Dostatni et al., 1991). Toutefois, une faible expression de E2 peut activer le promoteur précoce p107 de HPV-18 (Steger and Corbach, 1997), suggérant que E2 encombre stériquement la TATA-box pour empêcher le recrutement de TBP sur son site de liaison.

Il a été observé récemment que les facteurs de transcription TBX2 et TBX3 (qui lie des motifs d'ADN palindromiques) interagissent avec la protéine tardive L2 de HPV-16 et régulent négativement l'expression des gènes viraux précoces. Cette répression est accentuée lorsque la protéine tardive L2 est exprimée, démontrant ainsi leur rôle dans la régulation de l'expression des gènes viraux précoces (Schneider et al., 2013).

b) Structure du promoteur tardif.

De très nombreux sites d'initiation de la transcription tardifs ont été caractérisés sur une centaine de nucléotides chez les HPV. Toutefois, un promoteur majoritaire a toujours été mis en évidence dans l'ORF de E7 : p670 pour HPV-16 (Grassmann et al., 1996); p742 pour HPV-31 (Hummel et al., 1992) et p765 pour HPV-18 (Wang et al., 2011b). Quoiqu'il en soit, le promoteur tardif ne possède pas de TATA-box canonique et est activé de manière dépendante de la différenciation cellulaire car il est régulé par différents facteurs cellulaires exprimés différemment au cours de la différenciation.

Par exemple, deux sites de fixations pour la protéine Oct-11 ont été identifiés 190 et 170 nucléotides en amont du TSS et vont permettre d'activer le promoteur tardif p670 (Kukimoto and Kanda, 2001) (cf. figure 35). Cette protéine partage les mêmes sites de fixation que la protéine YY-1 qui a été démontrée comme inhibitrice de l'expression des gènes tardifs de HPV-6 en se fixant sur ces mêmes séquences (Ai et al., 2000). Néanmoins, aucune étude n'a confirmé son rôle inhibiteur chez HPV-16.

De plus, YY-1 est exprimée dans les cellules épithéliales non différenciées, contrairement à Oct-11 qui est exprimée dans les kératinocytes différenciés. Au cours de la différenciation, le facteur Oct-11 remplacerait YY-1 de ses séquences cibles (Kukimoto and Kanda, 2001) et activerait ainsi l'expression des gènes viraux tardifs.

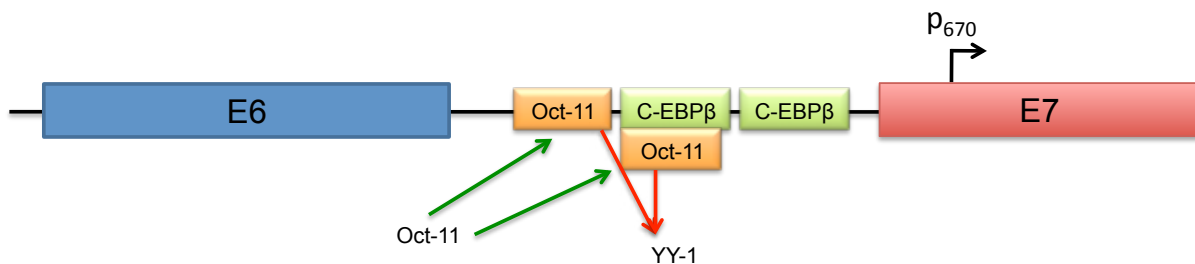


Figure 35 : Représentation schématique du promoteur tardif p670 de HPV-16.

Le promoteur tardif de HPV-16 qui permet l'expression de tous les gènes viraux hormis E6 et E7 est composé d'éléments de fixation pour le facteur C/EBP β dont l'isoforme activateur LAP est capable d'activer la transcription dans les cellules différenciées de l'épithélium. Il est également constitué de séquences de fixation pour Oct-11, partagées avec YY-1. Dans les cellules basales YY-1 est présente sur ces séquences. Au cours de la différenciation cellulaire Oct-11 va être exprimé et va chasser YY-1 pour prendre sa place et activer l'expression des gènes tardifs.

Par ailleurs, deux sites de fixation pour C/EBP β ont été caractérisés en amont du p670 de HPV-16 et permettent à cette protéine d'activer le promoteur tardif (cf. figure 35). De manière intéressante, C/EBP β est impliquée dans la différenciation terminale de kératinocytes ce qui expliquerait son rôle activateur (Kukimoto et al., 2006). Il existe plusieurs isoformes de C/EBP β provenant de son pré-ARNm : LIP (facteur de transcription inhibiteur) et LAP (facteur de transcription activateur). Au cours de la différenciation de cellules infectées, le niveau de LIP diminue au contraire de LAP. De manière intéressante, la surexpression de LIP dans des cellules différenciées diminue significativement l'expression des gènes viraux tardifs. Si l'extinction de l'ensemble des formes de C/EBP β empêche l'expression des ARNm tardifs, elle augmente légèrement la quantité d'ARNm précoces. Ces résultats suggèrent un rôle inhibiteur de C/EBP β dans la transcription des gènes précoces et tardifs dans les cellules basales, et activateur des gènes tardifs dans les cellules différenciées (Gunasekharan et al., 2012).

Compte tenu des difficultés de culture cellulaire et du chevauchement des gènes, très peu d'études ont pu identifier les facteurs cellulaires ou viraux impliqués dans l'expression des ARNm tardifs. Néanmoins, il semble qu'un élément « enhancer » indépendant de la différenciation cellulaire stimule la transcription des gènes viraux tardifs sans connaître les facteurs cellulaires associés (Bodily and Meyers, 2005). De plus, l'activité du promoteur tardif n'est pas dépendante de la réplication du génome viral puisque lorsque la réplication virale est bloquée, la quantité d'ARNm produits par le promoteur tardif n'est pas impactée (Bodily and Meyers, 2005).

Cependant, malgré cette régulation transcriptionnelle due à la différenciation cellulaire, les ARNm tardifs sont toutefois exprimés dans les couches basales de l'épithélium sans pour autant être traduit en protéines. Ceci peut suggérer une transcription basale induit par des facteurs de transcription exprimés de manière ubiquitaire ou bien une régulation par un mécanisme post-transcriptionnel. Parmi ces modes, l'épissage et la polyadénylation des ARNm viraux sont deux des étapes primordiales à l'expression (temporelle) correcte des ARNm tardifs des papillomavirus.

3) La régulation de l'expression des gènes viraux tardifs par l'épissage et la polyadénylation.

Chez les HPV, tous les ARNm subissent un épissage alternatif (cf. figure 32). Ces sites d'épissages sont contrôlés par des éléments cis- et trans-régulateurs présents dans l'ARN. Outre l'épissage, l'expression des gènes viraux tardifs est également dépendante de la polyadénylation des ARNm viraux. Ces deux mécanismes (épissage et polyadénylation) sont étroitement liés chez ces virus.

Lorsque les pré-ARNm précoces de HPV sont transcrits de manière tissu-spécifique par la machinerie de transcription cellulaire, l'épissage alternatif va générer différents ARNm (Milligan et al., 2007), tous polyadénylés au site pAE. L'épissage alternatif génère de ce fait des ARNm polycistroniques dans lesquels la robustesse du codon d'initiation AUG de chaque gène détermine l'efficacité de traduction de cet ARNm. Cependant, un codon d'initiation en contexte favorable dans le premier cadre de lecture peut efficacement défavoriser la traduction de l'ORF suivant car la majorité des ribosomes initie la traduction au premier ORF avant de se dissocier au niveau du codon stop, localisé en amont des signaux de polyadénylation. Dans la situation de HPV-16, l'épissage alternatif des ARN est par conséquent nécessaire car il permet à chaque gène d'être transcrit en un ARNm dans lequel le gène en question est en première position pour être traduit (cf. figure 32). Dans le cas des ARNm précoces synthétisés à partir du p97, le premier AUG est celui de E6. Cependant, la très courte séquence séparant l'AUG de E6 et l'extrémité 5' de l'ARNm rend la reconnaissance de ce site inefficace. Ainsi, les ribosomes continuent le balayage de l'ARNm jusqu'à rencontrer l'AUG suivant : celui de E7 (cf. figure 32).

a) Les mécanismes d'inhibition de l'expression des gènes viraux tardifs dans la région précoce.

a.1) La régulation de l'expression des gènes viraux tardifs par épissage dans la région précoce.

L'épissage chez HPV-16 est régulé par différents facteurs cellulaires. Par exemple, un site d'épissage accepteur majeur (SA3358 chez HPV-16) est situé juste en amont de l'ORF de E4. Ce site est régulé de manière à bloquer l'expression des gènes viraux tardifs et n'exprimer que les ARNm précoces en particulier E4, E5, E6 et E7 (cf. figures 32 et 36). Ainsi, dans les

cellules basales, un ensemble de mécanisme va favoriser l'utilisation de ce site 3' accepteur pour inhiber l'expression des gènes tardifs. Ce site est dépendant d'une séquence « enhancer » d'épissage, riche en A/C, et localisée une centaine de nucléotides en aval de ce dernier (Rush et al., 2005). Elle est fixée par le facteur SRSF1 qui va stimuler l'épissage (Li et al., 2013a; Somberg and Schwartz, 2010). Une autre séquence (de quarante nucléotides) retrouvée plus en aval du site de fixation de SRSF1 est fixée par un autre facteur, SRSF3, qui bloque l'épissage de ce site (SA3358) (cf. figure 36) pour favoriser l'expression des ARNm de E6 et E7 (Jia et al., 2009). Effectivement, la surexpression de SRSF3 augmente l'expression de E6 et E7 alors qu'un siRNA contre cette protéine cellulaire augmente l'expression de l'ARNm tardif L1. Une dernière protéine SR, SRSF9, est impliquée dans la régulation de ce site d'épissage (Somberg et al., 2011). SRSF9 semble inhiber l'épissage sur le site accepteur pour favoriser celui des gènes tardifs, d'autant plus que cette protéine est toujours exprimée dans les cellules différenciées (Somberg et al., 2011), contrairement à SRSF3.

De plus, la protéine virale précoce E2 augmente l'expression de SRSF1 au niveau transcriptionnel en se fixant sur son promoteur, favorisant de ce fait l'épissage des ARNm tardifs précocement dans l'infection (Mole et al., 2009). Ainsi, SRSF1 semble être une protéine importante dans le cycle de HPV-16 car elle stimule l'expression des gènes précoces dans les cellules basales et des gènes tardifs dans les cellules différenciées.

Dans les couches épithéliales supérieures, les cellules sont exposées aux signaux de différenciation cellulaire qui activent le promoteur viral tardif p670. Ce promoteur ne permet plus l'expression de E6 et E7 au contraire de E1, E2 et E4, qui sont toujours polyadénylés au site pAE. Le changement de promoteur ouvre la voie à l'expression de L1 et L2 et cette étape transitoire est accompagnée par une activité importante de réplication de l'ADN viral dans les cellules infectées.

a.2) La régulation de l'expression des gènes viraux tardifs par le signal de polyadénylation précoce.

Le signal de polyadénylation précoce de HPV-16 régule l'expression des gènes viraux tardifs en bloquant efficacement l'expression de L1 et L2 durant les étapes précoces du cycle viral. En effet, pAE divise le génome viral entre région précoce et tardive (cf. figure 36). Ce site de polyadénylation est donc très important et est finement régulé. Une mutation du pAE

induit quand même une polyadénylation efficace des ARNm précoces à partir d'un site cryptique en amont du pAE, suggérant que des éléments régulateurs forçent la polyadénylation des ARNm viraux précoces avant les ORF de L1 et L2 (Zhao et al., 2005). Effectivement, il semble que deux séquences distinctes, situées en amont et en aval du pAE soient impliquées dans ce mécanisme.

L'épissage du site accepteur 3358, situé avant le pAE, régule le site de polyadénylation précoce car une mutation de ce site inhibe la polyadénylation à partir du pAE. De manière intéressante, des éléments « enhanceurs » de polyadénylation en amont du pAE sont situés dans la partie 3' non codante des gènes précoces (cf. figure 36). Ils sont constitués de régions riches en A/U et U et favorise la polyadénylation. Ceci est probablement dû à un facteur cellulaire interagissant avec la poly(A) polymérase 1 nommé FIP1, qui se fixe dans la région riche en U pour stimuler la polyadénylation. Les éléments riches en AU et U sont également liés par des protéines associées à la machinerie de polyadénylation telles que : CstF64 (facteur stimulateur de clivage), hnRNP C1/C2, hnRNP I (protéine de liaison aux pyrimidines, PTB), et CPEB1 (protéine impliquée dans la polyadénylation) qui semblent tous travailler de concert pour la polyadénylation du pAE (Zhao et al., 2005).

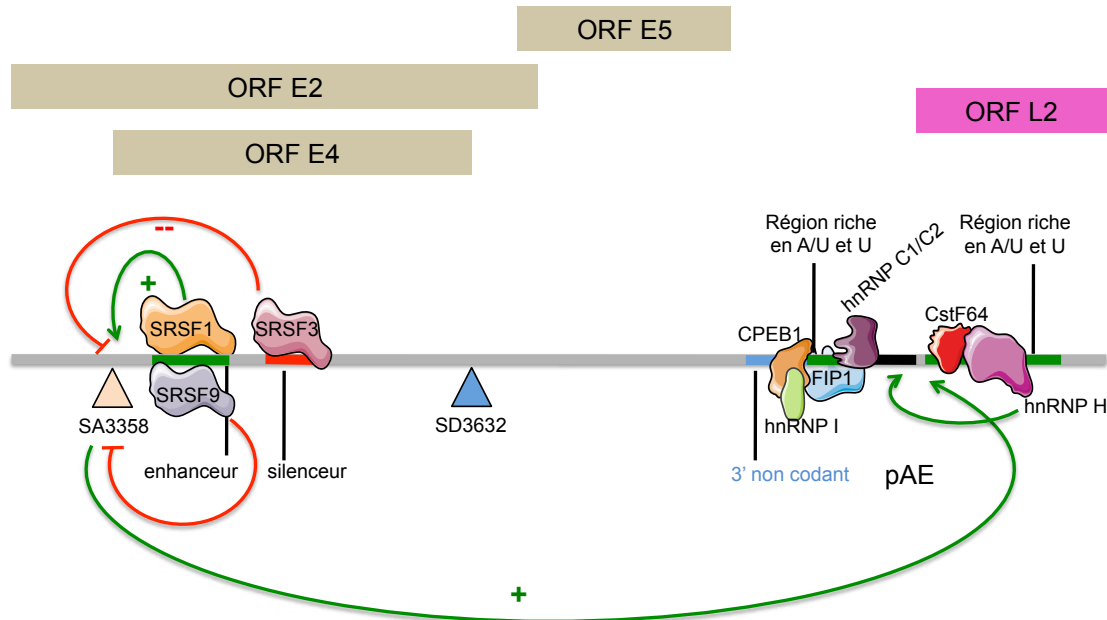


Figure 36 : Représentation schématique de la régulation de l'épissage du site accepteur 3358 et du signal de polyadénylation précoce de HPV-16, menant à l'inhibition de l'expression des gènes tardifs dans les cellules épithéliales basales, adapté de Johansson and Schwartz, 2013.

En haut sont représentés les ORF des différentes protéines virales. Les triangles représentent les sites d'épissage accepteur (en ocre) et donneur (bleu). La protéine cellulaire SRSF1, en se fixant sur sa séquence cible va favoriser l'épissage du site SA3358 afin de former des ARNm précoces. La liaison de SRSF3 sur sa séquence

« silencer » va inhiber l'épissage de ce site, donnant ainsi priorité à l'expression de E6 et E7. Comme SRSF3, SRSF9 bloque le site SA3358 afin de favoriser le recrutement du spliceosome sur d'autres sites accepteurs en faveur de l'expression des gènes viraux tardifs. Du côté de l'ORF de E2 se situe la région 3' non codante des gènes précoces suivit du signal de polyadénylation précoce pAE. De nombreux facteurs cellulaires jouxtent le pAE afin de favoriser la polyadénylation précoce et ainsi l'expression des gènes de même catégorie. La différenciation cellulaire va permettre de limiter l'expression de ces facteurs cellulaires. Cette étape est primordiale puisqu'elle permettra de passer outre le signal de polyadénylation et donc de former les ARNm tardifs.

L'activité de pAE est également régulée par des éléments situés dans la région codante de L2, localisés 174 nucléotides après le pAE et riche en motifs GGG. Ces éléments sont liés par CstF64 qui favorise la polyadénylation, mais aussi par hnRNP H afin de stimuler la polyadénylation à partir du site pAE (cf. figure 36). Une mutation de ces séquences GGG abolit la polyadénylation précoce et permet la lecture des ORF tardifs L1 et L2 et donc l'expression de leurs ARNm. La diminution de l'expression de hnRNP H et CstF64 est corrélée à la différenciation cellulaire terminale, ce qui tend à prouver que ces facteurs régulent le site pAE et empêchent la production des ARNm tardifs L1 et L2 dans les cellules basales de l'épithélium. En effet, dans les cellules différenciées, ces facteurs ne sont plus exprimés efficacement, ce qui permet de lire à travers le pAE et donc d'exprimer les gènes tardifs L1 et L2 (Öberg et al., 2005).

b) La régulation de l'expression des gènes tardifs dans la région tardive.

La diminution de la polyadénylation au site pAE mène à l'étape tardive du cycle viral productif, caractérisé par l'expression des protéines tardives L1 et L2. Les ARNm codant pour ces protéines sont épissés et polyadénylés au niveau du signal de polyadénylation tardifs pAL (cf. figure 37). Un des systèmes favorisant ce passage serait assuré par la protéine virale précoce E2. En effet, l'expression très élevée de cette protéine induit l'expression des gènes viraux tardifs. De plus, E2 semble capable de se fixer sur une région de 300 nucléotides chevauchant le pAE et d'inhiber la polyadénylation à partir de ce site. E2 interfère probablement avec les membres du complexe de polyadénylation² (CPSF) formé sur ce site. Effectivement, elle est capable d'interagir avec certains facteurs de ce complexe. Toutefois, il

² Cet ensemble est constitué de nombreuses protéines cellulaires formant un complexe sur le signal de polyadénylation AAUAAA de l'ARNm. Il est impliqué dans la réaction de polyadénylation de ARNm.

est possible que le rôle de E2 soit d'inhiber de manière globale la polyadénylation afin de diminuer le recrutement du CPSF sur le pAE à la faveur du pAL (Johansson et al., 2012).

Ainsi, E2 participe au changement d'expression entre précoces et tardifs en inhibant le promoteur précoce p97 mais aussi en réprimant la polyadénylation sur le pAE afin de permettre la lecture des ORF de L1 et L2 et donc l'expression de ces gènes viraux tardifs (cf. figure 37).

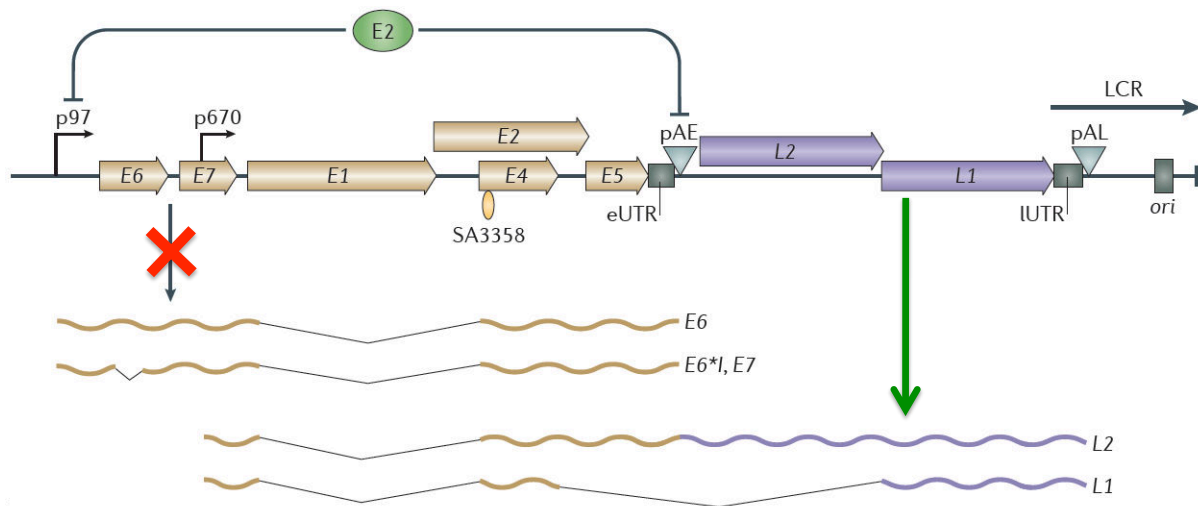


Figure 37 : Représentation schématique du rôle de E2 dans le changement d'expression entre gènes précoces à tardifs, tiré de Johansson and Schwartz, 2013.

La protéine précoce E2, exprimée très tôt au cours du cycle viral va permettre le maintien du génome viral ainsi que sa réplication. Elle va également réprimer le promoteur précoce p97, empêchant de ce fait l'expression des gènes viraux précoces E6 et E7 et favoriser indirectement l'utilisation du promoteur tardif p670. Celui-ci va contrôler l'expression de certains gènes viraux précoces tels que E1, E2, E4 et E5 mais aussi les gènes tardifs L1 et L2. De plus, E2 va bloquer le signal de polyadénylation précoce en interférant avec les membres du complexe de polyadénylation. Ainsi, les ARNm tardifs pourront être exprimés.

b.1) La régulation de l'expression des gènes viraux tardifs par épissage dans la région tardive.

Bien que la régulation de l'expression de l'ARNm L2 demeure floue, il a déjà été démontré que certaines protéines cellulaires telles que hnRNP K et PCBP (protéine de liaison au poly(C)) bloquaient la traduction de cet ARNm (Collier et al., 1998). De plus, différentes séquences cis-régulatrices ont été identifiées avec pour rôle de déstabiliser l'ARNm de L2 (Sokolowski et al., 1998). À l'inverse, la régulation de la production des ARNm de L1 est mieux comprise. Elle est réalisée à l'aide de deux sites d'épissage (un 5' et un 3'), inhibés par des séquences « silenciers » d'épissage où se fixent des facteurs cellulaires.

Comme évoqué précédemment, l'expression de l'ARNm de L1 est régulée de manière dépendante de la différenciation cellulaire. Il a déjà été démontré que SRSF1 et SRSF9 stimulent l'épissage du site 3' utilisé pour l'ARNm tardif de L1, facilitant de ce fait son expression (Somberg et al., 2011). Ceci est observé malgré le fait que le niveau d'expression de SRSF1 diminue fortement lorsque les cellules sont en phase terminale de différenciation, au contraire de SRSF9. De plus, alors que hnRNP A2/B1 induit l'expression des gènes viraux tardifs, hnRNP A1 la bloque (Zhao et al., 2004). Ces données suggèrent donc que l'épissage des gènes tardifs chez HPV-16 est régulé par des facteurs cellulaires à la fois répresseurs et activateurs.

L'utilisation du site 5' donneur tardif (SD3632) est inhibée par des éléments situés juste en amont de ce site donneur (Rush et al., 2005). Ils sont fixés par les protéines cellulaires hnRNP D et hnRNP A2/B1 sur deux motifs AUAGUA chez HPV-16 (Li et al., 2013b) et vont jouer le rôle d'inhibiteur d'épissage (cf. figure 38). À l'inverse, la protéine hnRNP I (ou PTB) bloque l'action de hnRNP D et hnRNP A2/B1, favorisant ainsi l'épissage au niveau de ce site SD3632. Cette protéine induit donc l'expression des ARNm tardifs (cf. figure 38) (Somberg et al., 2008).

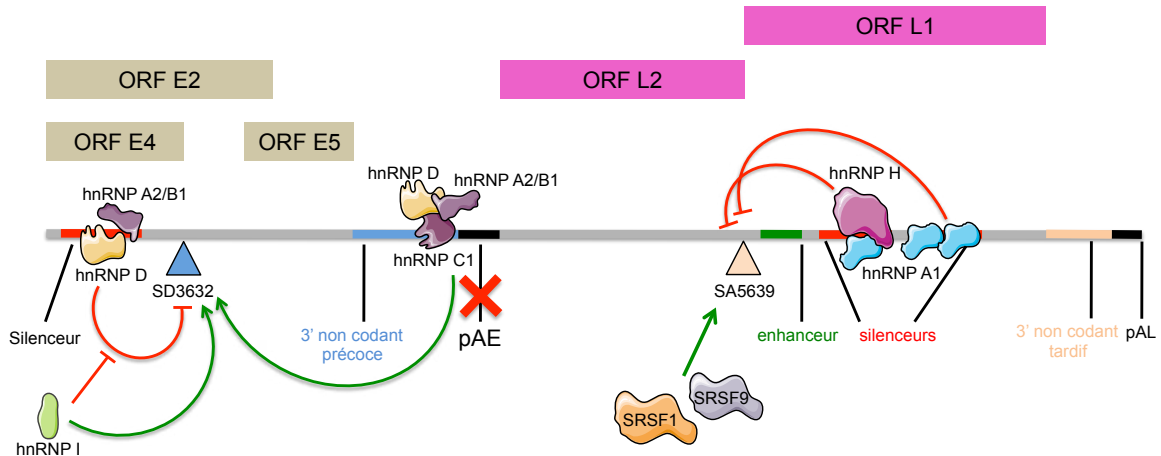


Figure 38 : Représentation schématique de la régulation de l'expression des gènes viraux tardifs par l'épissage, adapté de Johansson and Schwartz, 2013.

L'ARNm tardif de L2 est épissé à partir d'un site donneur situé dans l'ORF de E1 et le site accepteur 3358 (évoqué précédemment). L'ARNm de L1 est lui épissé à partir du site donneur 3632 et du site accepteur 5639. Le site donneur est négativement régulé dans les cellules basales de l'épithélium par les facteurs cellulaires hnRNP A2/B1 et hnRNP D qui agissent en se fixant à leur séquence cible. La différenciation cellulaire va permettre de lever l'inhibition. De plus, le facteur cellulaire hnRNP I va contrecarrer l'effet des deux facteurs cités précédemment dans le but d'activer l'épissage à partir de ce site. Le site accepteur est contrôlé par différents mécanismes. Il est réprimé par la fixation sur les « silencers » des facteurs cellulaires hnRNP H et A1. Inversement, les protéines SRSF1 et SRSF9 vont elles, stimuler l'épissage de site accepteur tout comme la

séquence « enhancer », localisée juste après le site 5639 (sans en connaître pour l'instant la protéine cellulaire impliquée).

Le site 3' SA5639 est lui sous le contrôle d'une séquence nucléotidique de 17 pb, située juste après ce site d'épissage dans la région codante de L1, augmentant son utilisation. À l'inverse, une séquence riche en purine (A et G) dans le région codante de L1 est retrouvée derrière la séquence « enhancer » et joue un rôle de « silencer » d'épissage (cf. figure 38). Cette dernière est fixée par hnRNP A1 et hnRNP H qui bloquent l'épissage et contrecarrent l'effet de l'élément « enhancer » (Zhao et al., 2004, 2007a). D'autres séquences « silencers » d'épissage, situées dans la région 5' de L1, inhibent l'expression des ARNm tardifs (Collier et al., 2002) et sont fixées par la protéine hnRNP A (cf. figure 38) (Zhao and Schwartz, 2008). Toutefois, hnRNP A1 est absent dans les cellules différenciées, permettant ainsi l'épissage au niveau de ce site accepteur tardif et donc l'expression de l'ARNm de L1 (Zhao et al., 2007a).

L'épissage de HPV-16 est également régulé par une séquence située dans la partie 3' non codante précoce où se fixe la protéine hnRNP C1. Celle-ci se fixe sur sa séquence cible et semble favoriser l'épissage au niveau du site 5' donneur (SD3632) en interagissant avec d'autres facteurs cellulaires tels que hnRNP D et hnRNP A2/B1 (Dhanjal et al., 2015).

b.2) La régulation de l'expression des gènes viraux tardifs par le signal de polyadénylation tardif.

Comme pour les gènes précoces, les gènes tardifs de papillomavirus sont polyadénylés mais cette fois à partir du signal de polyadénylation tardif pAL. La région 3' non codante des ARNm tardifs contient des séquences inhibant l'expression des gènes tardifs, soit en réduisant la stabilité des ARNm via des séquences riches en nucléotides A/U (Kennedy et al., 1991) soit en inhibant la polyadénylation ou encore la traduction. Par exemple les 3' non codantes des gènes tardifs de HPV16 contiennent plusieurs séquences UGUUUGU, conservées entre papillomavirus (Zhao et al., 2007b) et sont similaires aux motifs AUUUA, retrouvés dans les éléments instables des 3' non codants des ARNm cellulaires à durée de vie courte. Dans le cas de HPV-16, ils sont probablement fixés par différents facteurs cellulaires tels que CUGBP1 (protéine de liaison à l'ARN des motifs riches en CUG) (Goracznik and Gunderson, 2008) ou HuR (protéine de navette cytoplasme noyau) (Cumming et al., 2009) qui vont les stabiliser.

Ainsi, les mécanismes de régulation de la transcription impliqués dans l'expression des gènes de papillomavirus précoces sont assez bien compris, contrairement à ceux des gènes viraux tardifs. Toutefois, il est évident que le passage entre expression des gènes précoces et tardifs est étroitement couplé à la différenciation cellulaire des kératinocytes et ce, via les facteurs cellulaires impliqués dans les mécanismes d'épissage et de polyadénylation. En effet, la différenciation des cellules épithéliales est corrélée avec la diminution d'expression des facteurs cellulaires impliqués dans ces deux mécanismes (cf. figure 33). Ce changement de profil va alors inhiber le site de polyadénylation précoce ainsi que l'utilisation des sites d'épissage favorable à l'expression des gènes viraux précoces. À l'inverse, cette situation sera propice à l'expression des ARNm tardifs de L1 et L2 en avantageant la polyadénylation tardive ainsi que l'utilisation des sites donneurs d'épissage exclusivement utilisés pour les ARNm tardifs.

F) Synthèse sur les virus à ADN.

À travers cette introduction sur ces virus à ADN double brin transitant par le noyau, il est possible de constater que ces grandes familles de virus ont toutes des mécanismes de régulation de l'expression de leurs gènes variés. En effet, certains virus comme ceux ayant un petit génome tels que les papillomavirus ou les polyomavirus utilisent massivement la machinerie transcriptionnelle et d'épissage cellulaire pour réguler l'expression de tous leurs gènes. En effet la structure de leurs promoteurs est semblable à celle des gènes cellulaires. De plus, l'antigène T (polyo) et E2 (HPV), qui sont toutes deux des protéines précoces, vont moduler ces mécanismes pour favoriser l'avancée du cycle viral productif et la synthèse des ARNm tardifs. Par ailleurs, la différenciation des cellules épithéliales est couplée à l'expression des gènes uniquement chez les papillomavirus.

Les virus ayant un génome de taille intermédiaire tels que les adénovirus possèdent également des promoteurs complexes, très similaires à ceux des gènes cellulaires. Ils utilisent d'ailleurs toutes les machineries cellulaires à leur avantage. À la différence des virus à petit génome, les adénovirus codent pour différentes protéines ayant une fonction bien définie dans le détournement de ces machineries. Par exemple, certaines vont faciliter la transcription des

gènes précoces alors que d'autres vont stimuler celle des tardifs. Certains facteurs viraux vont spécifiquement moduler l'épissage des ARNm tardifs, d'autres stimuler l'export de ces ARNm et d'autres, la traduction de ces derniers.

Les virus ayant un génome de grande taille (baculovirus et bactériophages) possèdent des promoteurs pour leurs gènes précoces similaires à ceux de leurs hôtes respectifs. Toutefois, une protéine virale va favoriser la transcription des gènes viraux précoces en facilitant le recrutement de l'ARN polymérase. L'expression des gènes tardifs est beaucoup plus complexe et variée. Certains virus comme les baculovirus ou le bactériophage T7 codent pour leur propre ARN polymérase qui transcrita spécifiquement leurs ARNm tardifs. D'autres comme le bactériophage T4, codent pour plusieurs protéines qui vont mimer un complexe d'initiation de la transcription en recrutant l'ARN polymérase de leur hôte qui assurera la transcription uniquement des ARNm viraux tardifs.

Les Herpesvirus sont des virus ayant un génome du même ordre de grandeur que les bactériophages T4 ou les baculovirus. Quelle stratégie permet donc aux Herpesvirus de réguler l'expression de leurs gènes et notamment la synthèse de leurs ARNm tardifs ?

G) Les *Herpesviridae* (Herpesvirus).

Les Herpesvirus sont des virus enveloppés, à ADN double brin de plus de 150000 paires de bases. Ils font partie d'une grande famille de virus, dont certains signes pathologiques ont été décrits dès l'antiquité grecque. Plus d'une centaine d'Herpesvirus ont été recensés parmi lesquels neuf infectent l'homme (cf. tableau 1). Ils sont divisés en trois grandes sous-familles : Alpha-Herpesvirus (α -Herpesvirus), Beta-Herpesvirus (β -Herpesvirus) et Gamma-Herpesvirus (γ -Herpesvirus). Ils peuvent être encore divisés par genre et sont classés en fonction de leur tropisme mais aussi en fonction de la vitesse de répllication et de leur séquence nucléotidique. Les Alpha-Herpesvirus humains comprennent l'Herpès simplex virus 1 (HSV-1), l'Herpes simplex virus 2 (HSV-2) et Varicella Zona Virus (VZV) qui infectent les cellules épithéliales, mais se maintiennent en latence dans les cellules neuronales (notamment les ganglions) et se répliquent rapidement. Les Beta-Herpesvirus humains sont composés du Cytomégalovirus humain (HCMV), des Herpesvirus humain 6A et 6B (HHV-6A, HHV-6B) et l'Herpesvirus Humain 7 (HHV-7) qui infectent les

cellules épithéliales, les monocytes et se répliquent très lentement. Le virus d'Epstein-Barr et le virus du sarcome de Kaposi (KSHV) font partie des Gamma-Herpesvirus humains et infectent principalement les lymphocytes B en se répliquant assez lentement.

Alpha-Herpesvirus	Beta-Herpesvirus	Gamma-Herpesvirus
Human Herpes Simplex virus 1 (HSV-1 ou HHV-1)	Human Cytomegalovirus (HCMV ou HHV-5)	Epstein-Barr virus (EBV ou HHV-4)
Human Herpes simplex virus 2 (HSV-2 ou HHV-2)	Human Herpesvirus 6A et 6B (HHV-6A et HHV-6B)	Kaposi sarcoma-associated Herpesvirus (KSHV ou HHV-8)
Varicella Zoster Virus (VZV ou HHV-3)	Human Herpesvirus 7 (HHV-7)	

Tableau 1 : Classification des Herpesvirus humains.

1) Le virus d'Epstein-Barr.

Le virus d'Epstein-Barr (EBV) est, comme évoqué précédemment, un γ -herpesvirinae. Il possède une capsidie icosaédrique et une enveloppe d'environ 150 à 200 nm de diamètre. C'est un virus à ADN double brin, d'environ 172 kpb. Il porte à chacune de ses extrémités des séquences répétées en tandem de plus de 500 pb (TR) (cf. figure 39). Ces éléments lui permettent de se circulariser une fois dans le noyau de la cellule hôte. Il forme ainsi un épisode viral, recouvert d'histones cellulaires. Entre ces deux régions, des séquences uniques (U) sont séparées par des séquences internes répétées (IR : Internal Repeat). Ces éléments IR séparent les fragments d'EBV en régions U1 à U5. Les séquences répétées IR 2 et 4 contiennent deux origines de réplication de l'ADN viral utilisées lors du cycle productif (OriLyt) alors que la région U1 contient l'origine de réplication virale utilisée lors de la latence (OriP) (cf. figure 39).



Figure 39 : Représentation schématique du génome du virus d'Epstein-Barr.

Le virus d'Epstein-Barr est composé de séquences répétées (TR) aux deux extrémités de son génome. À l'intérieur, 4 zones de répétitions (IR) séparent le génome en cinq régions nommées U1 à U5. La région U1 comprend l'origine de réplication virale utilisée lors de la latence (OriP) alors que l'IR 2 et 4 contiennent les origines de réplication virale utilisées lors du cycle productif (OriLyt).

2) Nomenclature.

Le virus d'Epstein-Barr fut le premier Herpesvirus dont le génome a été complètement séquencé en 1984, à partir de la souche de référence B95.8 issue de lymphocytes B de marmousets infectés par l'EBV provenant de patients atteints de la mononucléose infectieuse (Baer et al., 1984). Ce séquençage a permis d'identifier une centaine de cadres de lecture ouverts (Open Reading Frame : ORF), répertoriés d'après les fragments de restriction obtenus avec l'enzyme BamH1. Les gènes d'EBV sont nommés selon le fragment de restriction BamH1 dans lequel ils se trouvent, l'orientation de leur transcription (vers la gauche : Leftward (L), ou vers la droite : Right (R)) et selon le cadre de lecture (Frame : F). Par exemple, BamH1 Z Leftward Frame 1 s'écrit BZLF1.

3) Infection et pathologies associées.

Le virus d'Epstein-Barr se transmet principalement par la salive. Toutefois des infections par transfusions sanguines ou contacts sexuels ont été décrits. In fine, on estime qu'environ 95% de la population mondiale adulte est infectée de manière persistante par EBV. La primo-infection survient fréquemment lors de la petite enfance et ce, de manière asymptomatique. Toutefois, elle peut engendrer une mononucléose infectieuse dans un tiers des cas si elle a lieu chez les jeunes adultes. EBV est aussi associé à de nombreux cancers tels que le lymphome de Burkitt, la maladie de Hodgkin, des cancers du rhinopharynx ou encore des carcinomes gastriques chez les individus immunocompétents. Chez les personnes immunodéprimées (HIV positives ou transplantées), il est associé à des nombreuses lymphoproliférations et des lymphomes.

Le virus d'Epstein-Barr infecte majoritairement les lymphocytes B mais peut également être retrouvé dans les cellules épithéliales oropharyngées. L'infection par EBV peut arriver au niveau des amygdales lors de l'échange de salive. Pour ce faire, le virus passe

dans les cryptes d'amygdales à travers la mince couche d'épithélium recouvrant les lymphocytes B pour y pénétrer (Babcock et al., 2000), et ce, par des mécanismes encore mal définis.

4) Le cycle viral.

Le génome d'EBV présent dans la nucléocapside est libre de tout marquage épigénétique. Lors de la libération de l'ADN viral dans le noyau, cet ADN va être circularisé, recouvert des nucléosomes et méthylé. Il formera ainsi un épisode viral, semblable à un chromosome cellulaire. EBV rentre donc dans les lymphocytes B naïfs quiescents. Il va alors activer les lymphoblastes qui vont migrer jusqu'au centre germinatif. À l'aide de l'expression d'un nombre restreint de protéines virales, nommées protéines de latence (cf. tableau 2), il va induire la prolifération des blastes puis leur différenciation en cellules B mémoires. Ces dernières vont alors voyager dans le sang périphérique et porter un virus en latence 0 où aucune protéine virale n'est s'exprimée (cf. tableau 2).

Latence	Gènes exprimés	Fonction
Latence de type III	EBNA-1, EBNA-2, EBNA-3A, EBNA-3B, EBNA-3C, EBNA-LP, LMP-1, LMP-2A, LMP-2B, EBER et BART	Active la prolifération des lymphocytes B
Latence de type II	EBNA-1, LMP-1, LMP2A, EBER et BART	Induction de la différenciation des lymphocytes B
Latence de type I	EBNA-I, EBER et BART	Maintient du génome viral au cours des divisions cellulaires
Latence de type 0	EBER et BART	Persistance du virus dans les cellules infectées.

Tableau 2 : Les différents types de latence avec les gènes exprimés durant ces étapes ainsi que leur fonction dans le cycle.

Cette extinction des gènes viraux va permettre ainsi au virus d'échapper au système immunitaire. Toutefois, lors des divisions cellulaires, le virus entre en latence de type I pour exprimer uniquement EBNA-1 qui va maintenir et permettre la ségrégation du génome viral au cours de la mitose à l'aide de facteurs cellulaires (Deschamps et al, unpublished datas).

C'est ainsi que le virus, sous forme épisomique, persiste à vie dans les cellules B mémoires du sang périphérique. Suite à un stimuli, les cellules B mémoires vont se différencier en plasmocytes. Ces derniers vont migrer jusqu'aux cellules épithéliales de l'oropharynx pour sécréter des anticorps. Cette stimulation de la différenciation en plasmocyte va induire la réactivation du virus et donc la production de nouvelles particules virales (cf. figure 40). Celles-ci seront alors libérées au niveau des muqueuses, et donc au niveau de la salive lorsqu'il s'agit des amygdales (Thorley-Lawson, 2005; Thorley-Lawson et al., 2013). Néanmoins, les mécanismes de réactivation du virus sont pour l'heure mal connus *in vivo*.

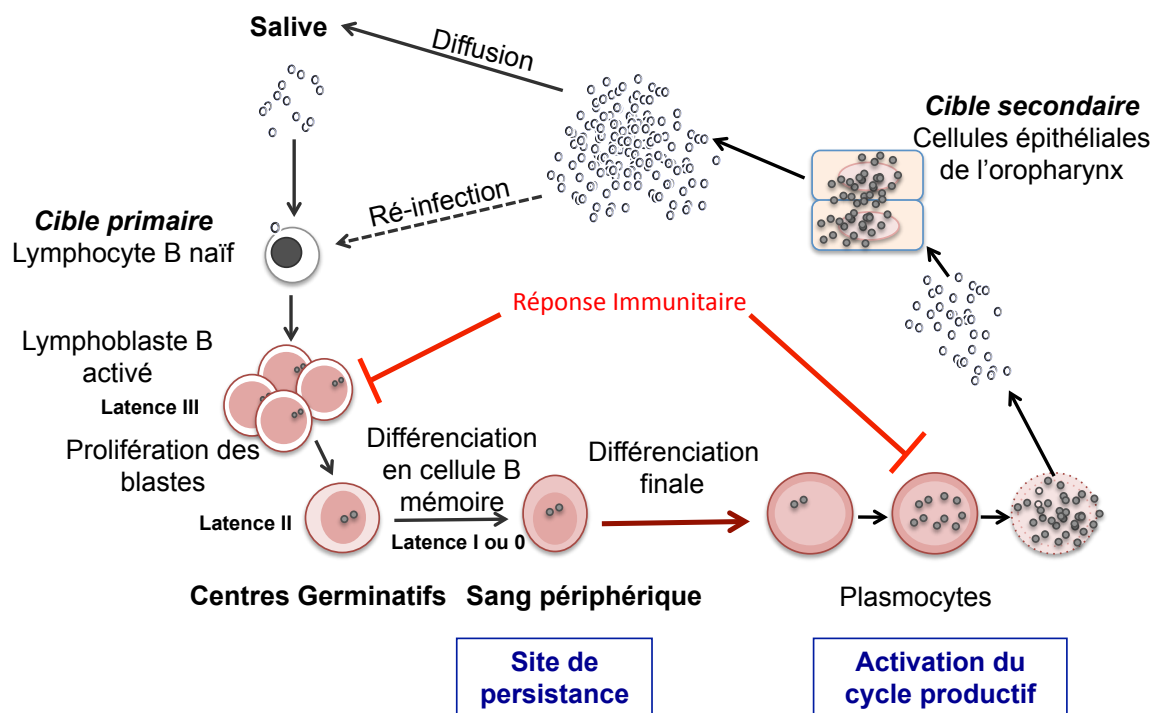


Figure 40 : Représentation schématique du cycle viral *in vivo*, adapté de (Thorley-Lawson, 2005).

Le virus présent dans la salive infecte un nouvel individu. Ces particules virales passeront à travers la barrière de l'épithélium pour infecter des lymphocytes B naïfs. Ces cellules vont alors être activées par le virus en latence III et vont proliférer avant de migrer jusqu'au centre germinatif. De là, le programme de latence virale va passer au niveau II. Les cellules B vont alors se différencier en B mémoire pour voyager dans le sang périphérique. À ce moment, aucune protéine virale ne sera exprimée hormis lors de la division cellulaire. Le virus est alors en latence 0 ou I. Le virus va alors persister dans cet état jusqu'à ce qu'un signal vienne induire la différenciation des B mémoires en plasmocytes. La différenciation va alors induire le cycle viral productif qui engendra la formation de nouvelles particules virales, libérées au niveau des muqueuses. Ces virions iront alors infecter de nouvelles cellules épithéliales où seront produits de nouveaux virions qui eux même iront infecter de nouvelles cellules B ou un autre individu à travers la salive.

5) Le cycle viral productif.

Les signaux et mécanismes moléculaires impliqués dans cette réactivation sont, comme évoqués précédemment, encore mal connus *in vivo*. En revanche, *in vitro*, il est possible d'induire le cycle viral productif grâce à l'existence de différentes molécules chimiques et biologiques susceptible d'induire cette réactivation. Parmi ces éléments, le 12-O-tétradécanoyl-phorbol-13-acétate (TPA) associé au butyrate de sodium (TPA-BA), les inhibiteurs de HDAC (histones déacétylase), les anti-immunoglobulines ou encore le facteur de croissance de tumeur TGF- β sont couramment utilisés pour y parvenir. Dans toutes les situations, ces évènements auront pour conséquence l'activation de l'expression des gènes immédiats précoces qui initieront le cycle viral productif.

Comme pour tous les Herpesvirus, le cycle viral productif est séquentiel et l'expression des gènes se fait de manière coordonnée, en cascade, et est toujours assurée par l'ARN polymérase II cellulaire. Celui-ci est divisé en quatre phases : une phase immédiate précoce, une phase précoce, une phase de réplication de l'ADN viral et enfin une phase tardive (cf. figure 41).

Ce cycle productif débute par l'expression de deux gènes immédiats précoces BZLF1 et BRLF1 qui codent pour deux facteurs de transcription, respectivement EB1 et R. Ces derniers vont coopérer, se fixer indépendamment l'un de l'autre en homodimère sur leurs séquences promotrices respectives, présentent dans les promoteurs des gènes viraux précoces (BZLF1 responsive element (ZRE) et BRLF1 responsive element (RRE)) (Farrell et al., 1989; Flemington and Speck, 1990a; Gruffat et al., 1990a; Manet et al., 1991; Rooney et al., 1989; Urier et al., 1989), permettant ainsi l'expression de ces gènes. Toutefois, EB1 est capable d'activer seule le cycle productif (Chevallier-Greco et al., 1986). EB1 a la particularité de posséder une structure similaire à la famille des facteurs de transcription « b-ZIP » à laquelle appartient notamment AP-1 (Chang et al., 1990). De plus, EB1 stabilise l'association des facteurs de transcription cellulaires TFIID-TFIIA avec la TATA-box présente dans les promoteurs des gènes viraux précoces dans le but d'activer l'expression de ces derniers (Lieberman, 1994; Lieberman and Berk, 1991, 1994). Aussi, EB1 semble travailler de concert avec la protéine CREB pour augmenter son activité transcriptionnelle (Zerby et al., 1999) mais également avec le facteur cellulaire Ikaros (Iempridee et al., 2014). L'autre singularité de EB1 est que cette protéine se fixe sur des séquences cibles (ZRE) variables. En effet,

celles-ci peuvent être divisées en deux catégories : ZRE classe 1, ZRE classe 2. La première (classe 1) comporte ou non des îlots CpG et est similaire aux sites de liaison de AP-1. La seconde comporte des îlots CpG méthylés (ZREme) (Bergbauer et al., 2010; Bhende et al., 2004). Étrangement, si les sites ZRE sont non méthylés lors de l'induction du cycle productif, celui-ci n'aboutit pas : seuls quelques gènes précoces sont exprimés (Kalla et al., 2012). Or, la méthylation des CpG demeure maintenue lors de la phase précoce du cycle productif sans entraver la transcription des gènes précoces (Woellmer et al., 2012). Cette particularité permet ainsi à EB1 de stimuler l'expression de R et d'activer rapidement l'expression des gènes viraux précoces (Bhende et al., 2004). De manière intéressante, le promoteur de BZFL1 est lui, pas ou peu méthylé, au contraire des autres promoteurs viraux (Woellmer et al., 2012).

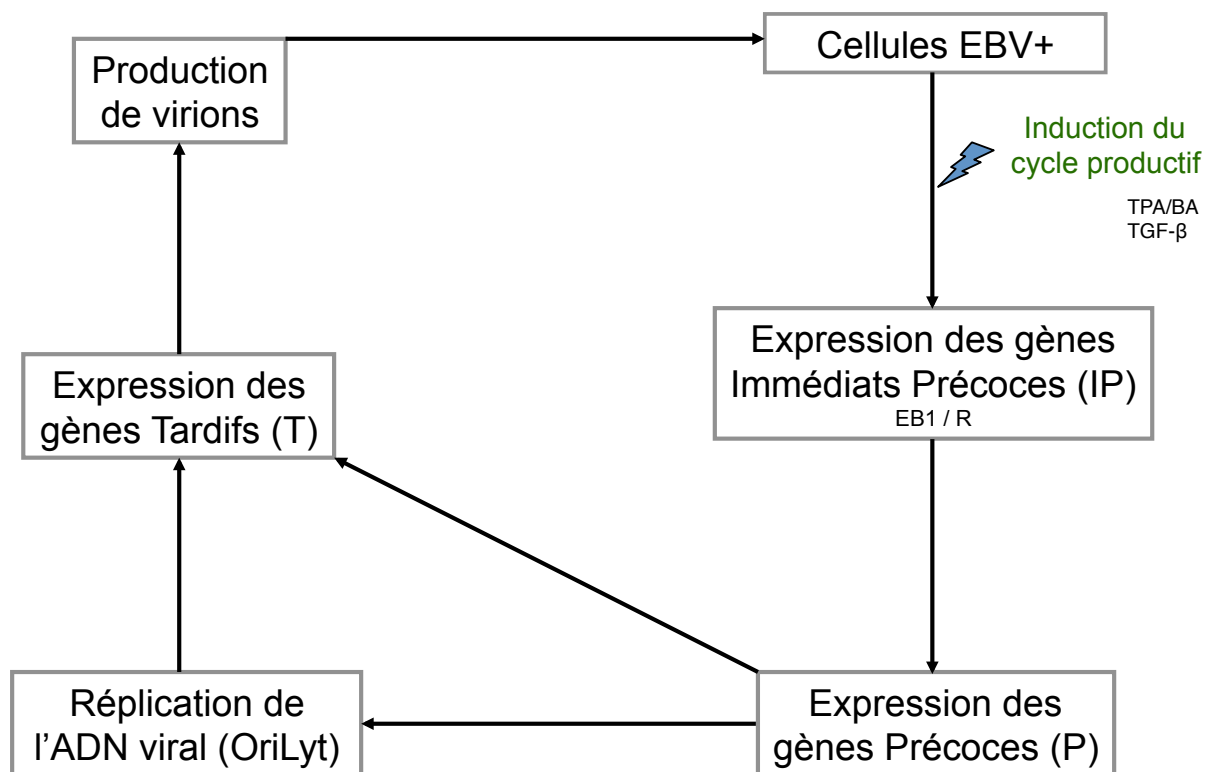


Figure 41 : Représentation schématique du cycle productif du virus d'Epstein-Barr.

Le cycle viral productif est décomposé en quatre phases : une immédiate précoce, une précoce, une de réplication d'ADN viral et une phase tardive. La phase immédiate précoce implique l'expression de deux facteurs viraux de transcription EB1 et R, impliqués dans l'expression des gènes viraux précoces. Ces derniers vont coder pour de nombreux facteurs impliqués dans la réplication de l'ADN viral mais aussi dans l'expression des gènes viraux tardifs. Suite à la réplication de l'ADN viral, les gènes viraux tardifs vont être exprimés et vont coder pour des protéines de structure du virus qui seront assemblées pour former de nouvelles particules virales infectieuses.

Parmi les gènes viraux précoces exprimés, certains codent pour des protéines impliquées dans la réplication de l'ADN viral. Ces protéines vont former un complexe de

réplication de l'ADN viral qui va assurer la copie du génome viral en multiples exemplaires. Les protéines incriminées sont : l'ADN polymérase virale BALF5 associée à son facteur de processivité BMRF1, la protéine de liaison à l'ADN simple brin BALF2, l'hélicase BBLF4, la primase BSLF1 et le facteur de processivité associée au complexe hélicase/primase BBLF2/3. Une dernière protéine, EB1, participe à la réplication de l'ADN viral. Cette réplication sera assurée en utilisant les deux origines de réplication d'ADN viral nommé Ori-Lyt (Hammerschmidt and Sugden, 2013). La réplication se déroule dans des compartiments bien différenciés nommés compartiments de réplication qui incluent BMRF1. Ce sont de larges structures globuleuses où l'ADN viral en cours de réplication est stocké (Daikoku et al., 2005).

Outre ces facteurs, le virus code pour une protéine exprimée précocement lors du cycle : le facteur EB2 issu du gène BSLF2/BMLF1. Tout d'abord, EB2 est essentielle à la formation de nouvelles particules virales (Gruffat et al., 2002a). Cette protéine a la particularité de pouvoir lier les ARNm viraux (Hiriart et al., 2003) et surtout de réguler leur expression au niveau post-transcriptionnel. En effet, EB2 favorise l'accumulation cytoplasmique d'ARNm viraux précoces et tardifs sans intron (Juillard et al., 2009, 2012), parmi lesquels deux ARNm précoces codant pour des protéines impliquées dans la réplication de l'ADN viral (BALF2 et BALF5) (Gruffat et al., 2002a) en recrutant des facteurs d'exports ou en bloquant l'action de certaines protéines d'épissage. Elle stabilise également de nombreux autres ARNm viraux, notamment tardifs (Thompson et al., 2016). En plus de ces fonctions, EB2 stimule la traduction des ARNm viraux en recrutant probablement la poly(A) binding protein (PABP) et eIF4G (Mure et al., unpublished data ; Ricci et al., 2009). Enfin, EB2 semble empêcher la dégradation des ARNm viraux sans intron (Mure et al, unpublished datas).

Hormis ces facteurs viraux, d'autres sont exprimés lors de la phase précoce et ont un rôle prépondérant dans l'expression des gènes viraux tardifs. Ce dernier point est l'élément central du travail de recherche effectué durant cette thèse.

Suite à la réplication de l'ADN viral (assurée par le complexe de réplication viral), les gènes viraux tardifs qui codent majoritairement pour des protéines de structure vont être exprimés. L'ensemble de ces événements va permettre la formation de nouvelles particules virales qui iront infecter de nouvelles cellules et de nouveaux hôtes (cf. figure 41).

6) Structure des promoteurs viraux immédiats précoces et précoces.

Au cours des années 1990-2000, de nombreuses études ont révélés les mécanismes de régulation de l'expression des gènes viraux immédiats précoces et précoces. En effet, la structure de leur promoteur a été caractérisée dans de nombreuses publications.

Par exemple, le promoteur du gène immédiat précoce BZLF1 est constitué de très nombreux éléments cis-régulateurs où se fixent les facteurs trans-régulateurs associés. Il a été divisé en trois régions cis-régulatrices activatrices situées entre -200 et +7 par rapport au TSS de BZLF1 (cf. figure 42). La première (ZI) est fixée par des facteurs de transcription de la famille de Sp-1 (Sp-1 ; Sp-3 et KLF4) (LIU et al., 1997; Nawandar et al., 2015), de MEF2 (MEF2D) (Gruffat et al., 2002b; Liu et al., 1997; McDonald et al., 2010) et de Smad (sensible à l'action du TGF- β) (Liang et al., 2002) et est riche en séquences GC, GACC et A/T. Cette région est fixée par des facteurs cellulaires qui activent la transcription du promoteur immédiat précoce. Le deuxième motif (ZII) contient différents sites de fixation aux facteurs de la famille « b-ZIP » (Ruf and Rawlins, 1995) tels que CREB, ATF (Wang et al., 1997), AP-1 (c-Jun et c-Fos), XBP1 (Bhende et al., 2007; Sun and Thorley-Lawson, 2007) et C/EBP α (Wu et al., 2004), afin d'augmenter la transcription. Le dernier site de fixation des protéines appartenant à la famille des « b-ZIP » est proche de la boîte TATA (cf. figure 42). Cette TATA-box est elle, fixée par les facteurs généraux de la transcription cellulaire dont TFIID. Enfin, la troisième région (ZIII) est fixée par EB1 elle même et effectue un rétrocontrôle positif (Flemington and Speck, 1990a, 1990b; Lieberman and Berk, 1990) (cf. figure 42).

Hormis ces séquences cis-activatrices fixées par les facteurs de transcription correspondants, le promoteur de BZLF1 est également régulé négativement par différentes protéines cellulaires. Parmi celles-ci, la famille des protéines ZEB (protéines à doigt de zinc fixant les E-box), ZEB1 et ZEB2 qui se fixe entre la TATA-box et le TSS (cf. figure 42) (Ellis et al., 2010; Yu et al., 2007). MEF2 réprime également la transcription lors de la latence en étant associé à ses séquences cibles et en recrutant des histones déacétylases sur les nucléosomes. Lors de la réactivation, des modifications post-traductionnelles vont induire un changement d'activité de MEF2 qui va alors recruter les histones acétylase, qui vont alors acétyler les histones H3 et H4 pour activer la transcription (Gruffat et al., 2002b; Jenkins et

al., 2000). Une autre séquence cis-inhibitrice se trouve entre le site de fixation à MEF2D et à la famille des b-ZIP, sans en connaître le facteur associé (Yu et al., 2011). Aussi, le facteur YY1 (Yin Yang 1 protein) fixant ces séquences cibles tout comme E2F2 ou encore les sites Sp1/NF1 situés en amont du nucléotide -360 par rapport au TSS de BZLF1 sont impliqués dans l'inhibition de l'expression de l'ARNm de BZLF1 (cf. figure 42) (Montalvo et al., 1995; Thomas et al., 2003).

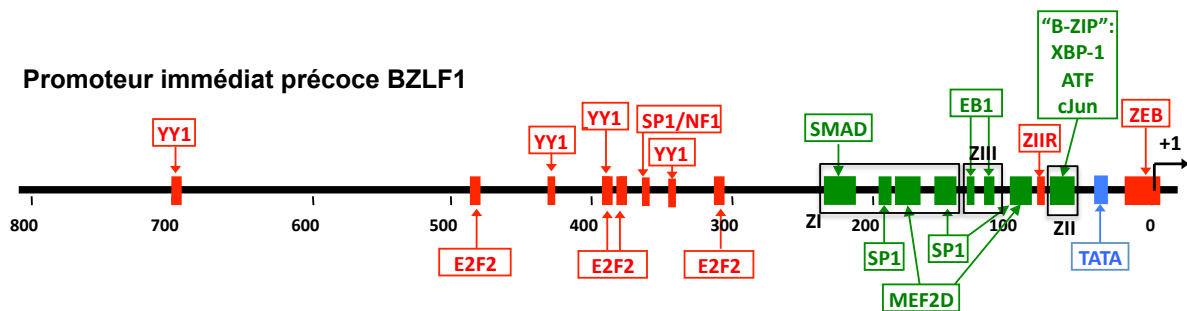


Figure 42 : Représentation schématique d'un promoteur de gène immédiat précoce d'EBV.

Le promoteur du gène viral immédiat précoce BZLF1 est constitué de très nombreux éléments cis-régulateurs où se fixent des facteurs cellulaires qui vont activer (en vert) ou réprimer (en rouge) la transcription mais aussi une TATA-box qui recrute la machinerie transcriptionnelle cellulaire. Aussi, le gène codant pour EB1 effectue un rétrocontrôle positif sur son promoteur en se fixant sur ces éléments de réponses.

En plus d'être régulé par des facteurs de transcription se fixant sur leurs séquences cis-régulatrices, le promoteur immédiat précoce est contrôlé par des marques épigénétiques suppressives de la transcription (H3K27me3, H3K9me2/3 et H4K20me3). Effectivement, ceux-ci sont retrouvées au niveau de la région régulatrice du gène immédiat précoce BZLF1 en latence (Murata et al., 2012). La réactivation du virus passe donc par une modification des marques épigénétiques présentes sur le promoteur de BZLF1 mais qui sont toutefois insuffisantes pour activer ce dernier (Countryman et al., 2008; Ramasubramanian et al., 2012).

Ainsi, il est clair que le promoteur de BZLF1 est constitué de très nombreuses séquences cis-régulatrices où se fixent des facteurs cellulaires trans-activateurs et répresseurs. Des marques épigénétiques répressives de la transcription sont également observées sur le promoteur de BZLF1. De plus, EB1 effectue un rétrocontrôle positif sur son promoteur pour augmenter son expression en se liant à ses propres éléments de réponse (ZRE). L'ensemble de ces éléments permet ainsi de contrôler la réactivation du virus.

Les promoteurs des gènes viraux précoces sont eux, plus simples. Par exemple, le

promoteur du gène précoce BMLF1/BSLF2 codant pour la protéine virale EB2 est dépendant d'une TATA-box. De plus, un site de fixation pour EB1 est situé environ une centaine de nucléotide en amont du TSS. Plus loin, environ 150 nucléotides avant le TSS, un élément de réponse pour R est présent (Buisson et al., 1989; Urier et al., 1989). Dans le cas de BMRF1 (facteur de processivité de l'ADN polymérase virale BALF5), la structure de ce promoteur est la même avec une TATA-box localisée une trentaine de nucléotides en amont du TSS, précédée de quatre séquences de fixation pour EB1 vers -42, -59, -106 et -170 par rapport au TSS. Également des sites de fixation pour R sont retrouvés entre les positions -158 et -170 (cf. figure 43). Le dernier suggère donc une coopération entre EB1 et R pour la stimulation de la transcription des gènes viraux précoces (Holley-Guthrie et al., 1990; Quinlivan et al., 1993).

Ainsi, la structure des promoteurs précoces est extrêmement simplifiée puisque des éléments de réponse aux facteurs de transcription immédiats précoces EB1 et R sont retrouvés en amont de ces gènes ainsi qu'une TATA-box recrutant les facteurs généraux de transcription permettant d'initier la synthèse d'ARNm.

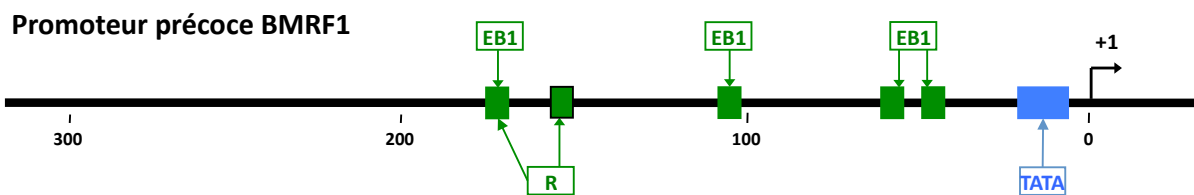


Figure 43 : Représentation schématique d'un promoteur de gène précoce d'EBV.

Les promoteurs des gènes précoces sont globalement structurés de la même manière. Ils sont constitués d'éléments de réponse aux protéines immédiates précoces activatrices ainsi que d'une TATA-box, permettant ainsi la transcription de ces gènes.

7) Les gènes viraux tardifs.

À l'inverse, les mécanismes de régulation de l'expression des gènes viraux tardifs sont beaucoup moins bien connus. Chez EBV, deux promoteurs de gènes tardifs ont été étudiés : BcLF1 (gène codant pour la protéine majeure de capsid MCP) et BFRF3 (gène codant pour la petite protéine de capsid VP26). Ces analyses ont démontré que les promoteurs de ces deux gènes sont radicalement différents de ceux des gènes viraux précoces et immédiats précoces. En effet, les promoteurs des gènes tardifs ne possèdent pas d'éléments cis-régulateurs en amont ou en aval du site d'initiation de la transcription. En revanche, seule une séquence TATT, située environ à 30 nucléotides en amont du TSS est présente (cf. figure 44)

(Chang et al., 1998; Serio et al., 1998). Cette séquence diffère de la boîte TATA canonique au niveau du dernier nucléotide ou une thymidine remplace une adénosine. De manière intéressante, la séquence TATT est localisée au même endroit que la TATA canonique et semble être le seul élément cis-régulateur capable de réguler ces gènes tardifs.

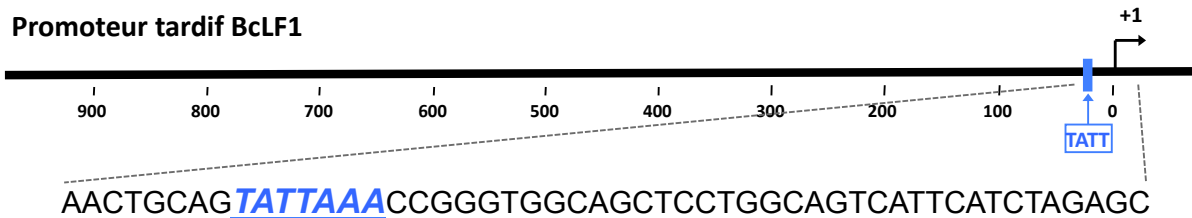


Figure 44 : Représentation schématique d'un promoteur de gène tardif d'EBV.

Les promoteurs des gènes viraux tardifs sont constitués d'une seule séquence cis-régulatrice : une séquence TATT où le dernier nucléotide diffère de la boîte TATA canonique. Ce seul élément est capable de réguler l'expression des gènes viraux tardifs.

Outre la structure du promoteur de ces gènes, il a été prouvé par plusieurs expériences que la réplication de l'ADN viral était nécessaire pour activer l'expression des gènes viraux tardifs. En effet, l'utilisation d'acide phosphono-acétique (PAA) qui cible l'ADN polymérase virale BALF5 bloque la réplication d'ADN viral. Cette inhibition empêche l'expression des protéines tardives d'EBV (Summers and Klein, 1976). De la même manière, des cellules dérivées de Lymphome de Burkitt (P3HR-1) capables de produire des particules virales ont été traitées à la cytosine arabinoside (qui est un analogue nucléotidique). L'utilisation de ce composé a bloqué la réplication de l'ADN viral mais aussi l'expression des protéines tardives (Gergely et al., 1971a). L'utilisation de cette même drogue dans des cellules Raji (dérivées de Lymphome de Burkitt), entraîne le même profil d'inhibition (Gergely et al., 1971b). Toutefois, ces cellules ne sont pas capables de produire de particules virales lorsque le cycle viral productif est induit. Effectivement, le génome viral présent dans ces cellules est amputé dans deux grandes régions dont l'une inclut le gène codant pour la protéine de liaison à l'ADN simple brin BALF2. De manière intéressante, l'induction du cycle viral productif conjugué à la transfection des cellules Raji avec un vecteur d'expression codant pour la protéine BALF2 restaure la réplication de l'ADN viral mais également l'expression des gènes viraux tardifs (Decaussin et al., 1995).

Enfin, l'utilisation d'une lignée cellulaire transfectée stablement avec un plasmide rapporteur luciférase dont le gène est sous contrôle du promoteur de gène tardif TATT a confirmé le rôle de la réplication dans l'expression des gènes viraux tardifs (Amon et al., 2004). Ce plasmide contient une origine de réplication virale OriP (latence) et une OriLyt

(cycle productif) qui permettent au vecteur de mimer le génome viral. Les auteurs de cette étude ont constaté que l'expression des gènes tardifs nécessitait la réplication de l'ADN viral, suggérant donc un rôle important des modifications chromatinienne (Amon et al., 2004).

Cependant, un vecteur rapporteur (sous contrôle d'un promoteur de gène tardif) transfecté transitoirement dans des cellules B infectées par EBV, dans lesquelles le cycle viral productif a été induit, suggère que la réplication n'est pas forcément le seul élément nécessaire à l'expression des gènes tardifs. En effet, en absence de réplication du plasmide, l'expression du gène rapporteur est quand même détectée lorsque le cycle viral productif est induit. Ces résultats suggèrent fortement l'implication de facteurs viraux trans-régulateurs exprimés après EB1 (Serio et al., 1997).

Ainsi, plus le cycle productif avance, plus la structure des promoteurs des gènes immédiats précoces à tardifs d'EBV est simplifiée. Effectivement, une régulation très maîtrisée des gènes immédiats précoces est nécessaire pour bien contrôler le maintien en latence ou la réactivation. Ensuite, les promoteurs de gènes viraux précoces sont constitués d'éléments de réponse à ces facteurs immédiats précoces activateurs qui faciliteront l'assemblage du complexe de transcription cellulaire. Enfin les promoteurs des gènes viraux tardifs sont constitués uniquement d'une boîte TATT. De plus, l'expression des ces derniers gènes viraux est dépendante de la réplication de l'ADN viral.

8) Caractérisation de BcRF1, une avancée majeure dans la compréhension de la régulation de l'expression des gènes tardifs.

De manière intéressante, une étude publiée en 2007 par l'équipe de Wyrwicz a suggéré par une analyse bio-informatique que la structure tridimensionnelle prédictive de la protéine BcRF1 était superposable à celle de la TBP d'archae *Pyrococcus Woesei* (cf. figure 45), suggérant un rôle de TBP-like pour BcRF1 (Wyrwicz and Rychlewski, 2007).

Des analyses réalisées au laboratoire ont confirmé le rôle de TBP-like de BcRF1 (Gruffat et al., 2012). En effet, la protéine se fixe *in vitro* sur la séquence TATT mais aussi TATA sans aucune distinction. En revanche, dans un contexte viral *in vivo*, la protéine virale BcRF1 se fixe uniquement sur la séquence TATT des gènes tardifs. De plus, des cellules

HEK293 contenant sous forme de bacmide le virus EBV amputé du gène codant pour la protéine BcRF1 (HEK293_{EBVΔBcRF1}), et dans lesquelles le cycle viral productif est induit en transfectant un plasmide d'expression codant pour la protéine virale EB1 ne sont pas capables de produire de nouvelles particules virales infectieuses ni d'exprimer les gènes viraux tardifs. En revanche, lorsque ce même type de cellules est transfecté avec un plasmide d'expression codant pour la protéine EB1 et transcomplémenté avec un vecteur d'expression codant pour la protéine BcRF1 (afin de restaurer son expression), des particules virales sont de nouveau produites. Par ailleurs, l'expression des ARNm tardifs est restaurée dans cette condition. Ces données suggèrent que BcRF1 est nécessaire à l'expression des gènes tardifs.

Néanmoins, BcRF1 n'est pas capable d'activer seule l'expression des gènes viraux tardifs. En effet, lorsque des cellules HEK293_{EBVΔBcRF1} sont transfectées avec un système rapporteur luciférase dont le gène est sous contrôle d'un promoteur tardif TATT, en présence du plasmide d'expression codant uniquement pour BcRF1, alors la luciférase n'est pas activée. En revanche, lorsque BcRF1 est co-exprimée avec l'ensemble complet de protéines viraux induites par EB1 la luciférase est fortement exprimée (Gruffat et al., 2012). L'ensemble de ces données suggère que d'autres protéines, notamment viraux sont impliquées dans l'expression des gènes viraux tardifs.

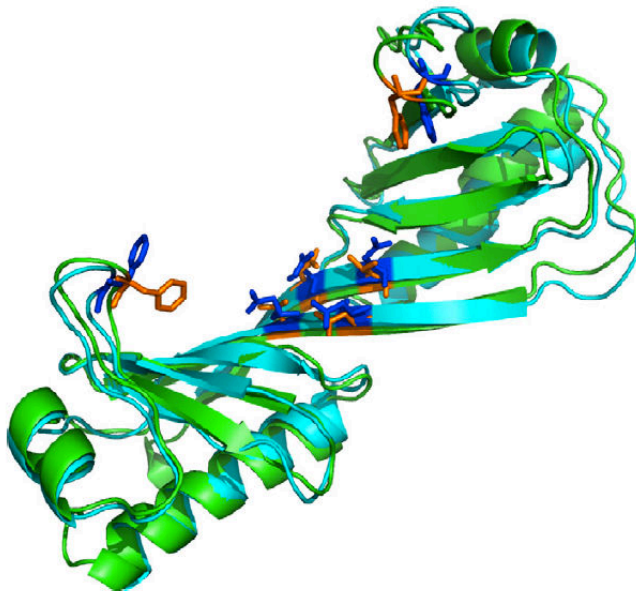


Figure 45 : Représentation tridimensionnelle de la structure prédictive de la TBP d'archaea et de BcRF1, tirée de (Wyrwicz and Rychlewski, 2007).

La superposition des structures de la TBP d'archaea (en vert) et prédictive de BcRF1 (en cyan) révèle une homologie importante de ces domaines.

9) Homologies entre facteurs protéiques de Beta et Gamma-Herpesvirus.

De manière intéressante, la protéine BcRF1 possède des homologues chez d'autres Herpesvirus. En effet, il a été prouvé que les homologues de BcRF1 chez l'Herpesvirus murin

68 (MHV68, γ -Herpesvirus) et le Cytomégalo­virus Human (HCMV, β -Herpesvirus), respectivement ORF24 et UL87 étaient impliquées dans l'expression des gènes tardifs de ces virus sans pour autant connaître leur fonction (Isomura et al., 2011; Wong et al., 2007). Curieusement, il a été démontré chez ces Herpesvirus (MHV68 et CMV), que d'autres protéines virales étaient impliquées dans l'expression des gènes tardifs de ces virus sans avoir caractérisé leur fonction. Ainsi, chez MHV68, outre ORF24 qui est l'homologue de BcRF1, quatre autres protéines ont été démontrées comme impliquées dans l'expression des gènes viraux tardifs : ORF18 (Arumugaswami et al., 2006), ORF31 (Jia et al., 2004), ORF30 et ORF34 (Wu et al., 2009) (cf. tableau 3). Chez CMV, trois protéines ont été caractérisées comme impliquées dans l'expression des gènes tardifs : UL79 (Perng et al., 2011), UL87 et UL95 (Isomura et al., 2011). Ces protéines sont les homologues respectives de ORF18, ORF24 et ORF34. Étrangement, des homologues de ces protéines sont également retrouvées chez EBV : BVL­F1 (pour ORF18 et UL79), BGLF3 (pour ORF34 et UL95), BDLF4 (pour ORF31) et BDLF3.5 (pour ORF30) (cf. tableau 3).

<u>γ-Herpesvirus</u>		<u>β-Herpesvirus</u>
EBV	MHV68	HCMV
BVL­F1 ?	ORF18	UL79
BcRF1 (TBP-like)	ORF24	UL87
BDLF3.5 ?	ORF30	
BDLF4 ?	ORF31	
BGLF3 ?	ORF34	UL95

Tableau 3 : Listes des protéines virales de MHV68 et HCMV démontrées comme impliquées dans l'expression des gènes tardifs de ces virus et leurs homologues retrouvées chez EBV.

Nous avons constaté que ces protéines avaient toutes des homologues chez d'autres Herpesvirus appartenant aux sous-familles des β - γ -Herpesvirus. Toutefois, ces protéines n'ont pas d'homologues avec les α -Herpesvirus. Cette constatation suggère fortement que la régulation de l'expression des gènes viraux tardifs des β - et γ -Herpesvirus est distincte de celle des α -Herpesvirus. En effet, les promoteurs et les facteurs protéiques impliqués dans

l'expression des gènes tardifs d' α -Herpesvirus sont différents des deux autres sous-familles d'Herpesvirus.

10) Mécanismes de régulation de l'expression des gènes tardifs d'Alpha-Herpesvirus.

La structure des promoteurs immédiats précoces (IP) des α -Herpesvirus est similaire à celle des β - γ -Herpesvirus. Ils sont constitués de très nombreux éléments cis-régulateurs (liés par des facteurs cellulaires et viraux), situés en amont d'une TATA-box. Les promoteurs précoces sont similaires aux IP à la différence que le nombre de sites de fixation pour des protéines cellulaires est moins élevé (les différenciant ainsi de ceux des β - γ -Herpesvirus).

À contrario, la sous-famille des α -Herpesvirus possède un système de régulation pour ses gènes tardifs bien distinct des β - γ -Herpesvirus. Si la réplication est aussi nécessaire pour permettre l'expression des gènes tardifs (Mavromara-Nazos and Roizman, 1987), la structure des promoteurs est en revanche légèrement différente. Effectivement, les promoteurs tardifs de cette catégorie de virus possèdent trois éléments. Effectivement, ils disposent tous d'une TATA-box canonique (située une trentaine de nucléotide en amont du TSS). Ils sont également constitué d'une séquence INR chevauchant le site d'initiation de la transcription (Homa et al., 1986, 1988; Johnson and Everett, 1986) et des séquences situées jusqu'à trente nucléotides en aval du TSS (DAS, similaires aux éléments DPE cellulaires) (Guzowski and Wagner, 1993; Huang et al., 1993; Weir and Narayanan, 1990). Ces dernières confèrent la régulation temporelle de la transcription des gènes tardifs uniquement après la réplication (Kibler et al., 1991). Cette région (DAS) serait fixée par le facteur cellulaire TAF6 bien que cela ne soit pas confirmé (Guzowski et al., 1994; Petroski et al., 2001)

De manière intéressante, ICP4 qui est une protéine virale immédiate précoce, semble être le seul facteur viral réellement nécessaire à l'expression des gènes viraux tardifs (Cook et al., 1995). En effet, ICP4 stimule l'expression des gènes tardifs *in vitro* en augmentant l'association du PIC sur le promoteur (Gu and DeLuca, 1994). Elle est également associée stablement *in vitro* avec TFIID (TBP et TAFs) et TFIIB (Smith et al., 1993).

In vivo, les mécanismes sont plus détaillés et complexes. En effet, l'absence de ICP4 sur les promoteurs des gènes viraux tardifs empêche la fixation de TBP et de l'ARN pol II sur

le promoteur (Sampath and DeLuca, 2008). C'est en se fixant à l'ADN que ICP4 va stabiliser la liaison de TFIID à la TATA-box des gènes viraux tardifs, permettant le recrutement efficace du PIC, notamment TFIIB et l'ARN pol II (Grondin and DeLuca, 2000). Plus précisément, ICP4 va stabiliser la TBP sur sa TATA-box uniquement en présence de l'élément INR (Kim et al., 2002), suggérant ainsi un rôle prépondérant de ce dernier avec ICP4. De plus, ICP4 interagit avec TBP et TAF1 (Carrozza and DeLuca, 1996). Or TAF1 est impliquée dans la reconnaissance des éléments INR (Verrijzer et al., 1995). Ces résultats suggèrent que la stabilisation de TBP sur la TATA-box passe par l'interaction de ICP4 et TAF1, fixées sur l'élément INR (Kim et al., 2002). Ainsi, l'élément INR et la TATA-box sont indispensables pour l'activation efficace des promoteurs tardifs par ICP4.

De plus, la présence de ICP4 sur les promoteurs des gènes viraux tardifs ne requière pas la présence de TFIIA dans le PIC (Zabierowski and DeLuca, 2004). De manière intéressante, la quantité de TFIIA au cours de l'infection diminue drastiquement. In fine, ICP4 est capable de substituer TFIIA dans la stabilisation de la liaison de TFIID à sa séquence TATA en présence de l'élément INR. En revanche, la mutation de ce dernier (INR) nécessite le recrutement de TFIIA pour stabiliser TFIID sur la TATA-box des gènes tardifs (Zabierowski and DeLuca, 2008).

11) Contexte et hypothèse.

Ainsi, à partir de l'ensemble des données de la littérature, nous savons que l'expression des gènes tardifs du virus d'Epstein-Barr est dépendante de la réplication de l'ADN viral. La structure de leur promoteur est simplifiée à son maximum puisqu'une seule séquence TATT régule leur expression. De plus, notre laboratoire a caractérisé la fonction de BcRF1 comme étant une TBP-like, se fixant uniquement sur les séquences des gènes tardifs en contexte infectieux. Nous avons également constaté que cinq protéines conservées exclusivement entre β - γ -Herpesvirus et non retrouvées chez les α -Herpesvirus étaient impliquées dans l'expression des gènes tardifs. Ces protéines chez HCMV sont même retrouvées dans les compartiments de réplifications (Isomura et al., 2011)

L'objectif général de la thèse fut d'étudier la régulation de l'expression des gènes tardifs du virus d'Epstein Barr.

Nous avons commencé par émettre l'hypothèse que toutes les protéines virales conservées entre β et γ -Herpesvirus et absentes chez les α -Herpesvirus pourraient avoir un rôle dans l'expression des gènes viraux tardifs.

À partir de cette hypothèse, nous avons effectué une analyse *in silico* au laboratoire pour découvrir tous ces facteurs. Ainsi, sept protéines ont été identifiées comme conservées entre β et γ -Herpesvirus et sont listées dans le tableau ci-dessous (cf. tableau 4).

γ-Herpesvirus			β-Herpesvirus	
EBV	KSHV	MHV68	HCMV	HHV-6
BVLF1	ORF18	ORF18	UL79	U52
BcRF1 (TBP-like)	ORF24	ORF24	UL87	U58
BDLF3.5	ORF30	ORF30	UL91	U62
BDLF4	ORF31	ORF31	UL92	U63
BGLF3	ORF34	ORF34	UL95	U67
BFRF2	ORF66	ORF66	UL49	U33
BTRF1	ORF23	ORF23	UL88	U59

Tableau 4 : Conservation des protéines virales entre β - et γ -Herpesvirinae.

Parmi ces protéines, outre les cinq déjà démontrées chez MHV68 (ORF18, ORF24, ORF30, ORF31 et ORF34) comme impliquées dans l'expression des gènes tardifs de ce virus, deux nouvelles correspondent à ces critères. Chez EBV ces protéines ont pour nom : BFRF2 et BTRF1. Toutefois, une publication venait de traiter le rôle de BTRF1 en démontrant que l'homologue de cette protéine chez MHV68 (ORF23), n'était pas impliquée dans l'expression des gènes tardifs (Ohno et al., 2012). En revanche, aucune publication ne s'était intéressée à la protéine BFRF2 ni à ses homologues chez les autres β et γ -Herpesvirus. C'est pourquoi nous avons décidé d'étudier et de caractériser cette protéine qui s'avérait être la dernière pièce nécessaire à l'expression des gènes viraux tardifs (Aubry et al., 2014).

PARTIE II :

RESULTATS

I) Matériel et Méthode :

Lignées cellulaires et transfections :

La lignée cellulaire HEK293 dérive de cellules de rein embryonnaire humain. Les cellules HEK293_{EBV} dérivent toutes de cellules HEK293 infectées avec un virus recombinant EBV présent dans un bacmide (donné par le Pr Hammerschmidt). Les lignées cellulaires HEK293, HEK293_{EBV} et les HEK293_{EBVΔX} sont cultivées dans du milieu RPMI 1640, supplémenté avec 10% de sérum de veau fœtal (SVF) et 10 µg/µl de Penicilline/Streptomycine. Les transfections transitoires de ces lignées cellulaires sont réalisées avec du polyéthylèneimine (PEI/NaCl).

Plasmides recombinants :

Un plasmide d'expression codant pour la protéine EB1 est utilisé pour l'initiation du cycle viral productif. Le cadre de lecture entier des gènes BFRF2, BGLF3, BVLF1, BDLF4, BDLF3.5, BALF2, BALF5, BBLF2/3, BBLF4, BMRF1, BSLF1, BTRF1 et BGLF4 a été amplifié par PCR avec les primers listés dans le tableau 5, purifié sur gel d'agarose par freeze-squeeze et cloné dans un vecteur donneur pDONR207 par la technique du Gateway® BP clonase® II enzyme mix (Invitrogen™). Ils ont ensuite été recombinés dans des plasmides pCi-3xFlag et pDEST27-GST par Gateway® LR clonase® II enzyme mix (Invitrogen™), sous contrôle d'un promoteur CMV et possèdent soit trois épitopes Flag soit un épitope GST à l'extrémité 5' du gène.

Le cadre entier de lecture de la GST a été amplifié par PCR avec des oligonucléotides spécifiques contenant des sites de restriction unique aux extrémités (listés dans le tableau 5). Ces produits sont ensuite digérés avec les enzymes de restriction appropriées, puis clonés par le système In-Fusion (Clontech) dans les plasmides pCI-Flag (Promega) portant les ORF de BFRF2, BGLF3, BVLF1, BDLF4 et BDLF3.5, déjà décrit précédemment (Aubry et al., 2014). La GST est insérée entre l'épitope Flag et l'ORF du gène après digestion enzymatique donnant ainsi un vecteur pCi-Flag-GST-X.

Le clonage dans le vecteur d'expression pCI-3xFlag, codant pour BcRF1, et du plasmide rapporteur pTATT_{BcLF}-Luc est déjà décrit (Gruffat et al., 2012). Ce dernier contient les séquences entre les positions -38 à -4 du gène viral tardif BcLF1, incluant la TATT-box. Le

clonage dans le vecteur d'expression bactérien pGEX-5X-2 des gènes codant pour les six protéines du vPIC est déjà décrit (Aubry et al., 2014)

Le cadre de lecture entier du gène codant pour BcRF1 a été amplifié par PCR à l'aide des primers listés dans le tableau 5. Le plasmide commercial pNTAP (Interplay Mammalian TAP System, Agilent) est digéré par les mêmes enzymes puis purifié. Une réaction de ligation entre le plasmide pNTAP et l'insert BcRF1 est réalisée sur la nuit à 4°C. Cette réaction permet de générer un plasmide codant pour la protéine BcRF1 contenant des épitopes Streptavidine et Calmoduline dans sa partie N-Terminale.

Génération de virus EBV recombinant KO pour BVLf1, BDLf4 ou BDLf3.5 :

Pour générer un mutant n'exprimant pas BVLf1 ou BDLf4 ou BDLf3.5 dans le contexte du génome viral, nous avons utilisé la technologie décrite par le Pr. Delecluse (Delecluse et al., 1998). Les virus mutants EBV Δ BVLf1, EBV Δ BDLf4 et EBV Δ BDLf3.5 sont générés pour remplacer une partie du cadre de lecture de BVLf1, BDLf4 et BDLf3.5 par un gène de résistance à la kanamycine par recombinaison homologue en bactérie. Afin d'enlever la cassette kanamycine, les bactéries *Escherichia coli* portant les bacmides EBV Δ BVLf1, EBV Δ BDLf4 et EBV Δ BDLf3.5 ont été transformées avec un plasmide pCP20, codant pour la red recombinase, qui va enlever la cassette. Le bacmide recombinant EBV (10 μ g) est ensuite transfecté dans des cellules HEK293 en utilisant le polyéthylèneimine (PEI/NaCl). 24h post-transfection, 100 μ g/ml d'hygromycine sont ajoutées dans le milieu de culture pour sélectionner les lignées HEK293 portant le virus recombinant stablement. Les colonies cellulaires GFP positives se développant sont amplifiées pour les analyser. Les clones sélectionnés dans cette thèse seront nommés HEK293_{EBV Δ BVLf1}, HEK293_{EBV Δ BDLf4} et HEK293_{EBV Δ BDLf3.5}. La lignée cellulaire HEK293_{EBV Δ BcRF1}, déjà décrite (Gruffat et al., 2012) est également utilisée.

Induction du cycle viral productif :

Les cellules HEK293_{EBV}, HEK293_{EBV Δ BVLf1}, HEK293_{EBV Δ BDLf4} et HEK293_{EBV Δ BDLf3.5} sont transfectées avec un plasmide d'expression codant pour la protéine EB1 (1 μ g par point) afin d'induire le cycle viral productif (Chevallier-Greco et al., 1986). Lors de la trans-complémentation, les cellules HEK293_{EBV Δ BVLf1}, HEK293_{EBV Δ BDLf4} et HEK293_{EBV Δ BDLf3.5} sont transfectées avec 1 μ g de plasmide pCI-Flag-BVLf1, pCI-Flag-BDLf4 et pCI-Flag

BDLF3.5 respectivement ou un plasmide vide. 72h post-transfection, les cellules sont récoltées pour analyser l'expression des gènes viraux.

Types Plasmides	Nom Primers	Séquences
pDONR207	BFRF2s	GGGGACAACCTTTGTACAAAAAAGTTGGCATGGCGTTATTCTTGGCGCGCC
	BFRF2as	GGGGACAACCTTTGTACAAGAAAAGTTGGTTAGGAAGCAGGGGACTGTCTG
	BGLF3s	GGGGACAACCTTTGTACAAAAAAGTTGGCATGTTCAACCGGTCAAGGCCG
	BGLF3as	GGGGACAACCTTTGTACAAGAAAAGTTGGCTACTCATCTTCATAAGTCACCA
	BVLF1s	GGGGACAACCTTTGTACAAAAAAGTTGGCATGTTAATGGGACTGGGGGTGC
	BVLF1as	GGGGACAACCTTTGTACAAGAAAAGTTGGCTACCTGGCCTCCCCGGTGG
	BDLF4s	GGGGACAACCTTTGTACAAAAAAGTTGGCATGTCAGACCAAGGCCGATTGAG
	BDLF4as	GGGGACAACCTTTGTACAAGAAAAGTTGGTCAACACTTGGTTGTCAATGTGGA
	BDLF3.5s	GGGGACAACCTTTGTACAAAAAAGTTGGCATGTCTGCCCCCGGATGCTCT
	BDLF3.5as	GGGGACAACCTTTGTACAAGAAAAGTTGGTCAATCGGCCTTGGTCTGACATT
	BTRF1s	GGGGACAACCTTTGTACAAAAAAGTTGGCATGCTCAAGTGAAGCAGCCC
	BTRF1as	GGGGACAACCTTTGTACAAGAAAAGTTGGTTAGATGCGCAGCAGGTAATAAA
	BGLF4s	GGGGACAACCTTTGTACAAAAAAGTTGGCATGGATGTGAATATGGCTGCG
	BGLF4as	GGGGACAACCTTTGTACAAGAAAAGTTGGTCATCCACGTCGGCCATCTG
	BALF2s	GGGGACAACCTTTGTACAAAAAAGTTGGCATGCAGGGTGCACAGACTAG
	BALF2as	GGGGACAACCTTTGTACAAGAAAAGTTGGCTAGACCTCGAGTCCGGGG
	BALF5s	GGGGACAACCTTTGTACAAAAAAGTTGGCATGTCTGGGGGACTCTTCTATA
	BALF5as	GGGGACAACCTTTGTACAAGAAAAGTTGGTTAGAATGGTGGCCGGGCTG
	BMRF1s	GGGGACAACCTTTGTACAAAAAAGTTGGCATGGAAACCACTCAGACTCTC
	BMRF1as	GGGGACAACCTTTGTACAAGAAAAGTTGGTTAAATGAGGGGGTTAAAGGC
	BBLF2/3s	GGGGACAACCTTTGTACAAAAAAGTTGGCATGTTACCCAGACCGCAGGG
	BBLF2/3as	GGGGACAACCTTTGTACAAGAAAAGTTGGCTAGAATAAACTGAGAACAGTCAT
	BBLF4s	GGGGACAACCTTTGTACAAAAAAGTTGGCATGGCCGAGGAGCCGAGG
	BBLF4as	GGGGACAACCTTTGTACAAGAAAAGTTGGTCAGTAAACCAGTAGTGCGCG
BSLF1s	GGGGACAACCTTTGTACAAAAAAGTTGGCATGTCCGCCCCCGTCGTC	
BSLF1as	GGGGACAACCTTTGTACAAGAAAAGTTGGCTAGTTCGGGAGAGTCTCTGA	
pNTAP	NTAP-BcRF1s	GCGCCCGGGCGGATCCATGGACTACAAAGACCATG
	NTAP-BcRF1as	CGGGCCCCCTCGAGTTACTTGGAGCATCAC
In fusion pCi-Flag-GST	GSTin-fusion s	AAGGGATCCCCTCGAGAAATGTCCCCTATACTAGG
	GSTin-BFRF2as	CCGGGTCGACTCTAGATTAGGAAGCAGGGGACTGTC
	GSTin-BGLF3as	CCGGGTCGACTCTAGACTACTCATCTTCATAAGTC
	GSTin-BVLF1as	CCGGGTCGACTCTAGACTACCTGGCCTCCCCGGTG
	GSTin-BDLF4as	CCGGGTCGACTCTAGATCAACACTTGGTTGTCAATGTGG
	GSTin-BDLF3.5as	CCGGGTCGACTCTAGATCAATCGGCCTTGGTCTGAC
Bacmide EBV	BVLF1KanaS	CACCCACCACCAGGCGGTTAGGGAGGACACTGCCAAGGTTGAAGAGCAGACAAAGTGCC ACCTGCAGAT
	BVLF1KanaAS	CCCGGTAACCGCTCACGGTCTCTTTGTCTGTTTTTGCATGACCGGCCTGCAGGAACACTTA ACGGCTGA
	BDLF4KanaS	GCCGATTGAGCCTCCGCGGGGGAGGGGGCACGGATGAGCCCAATCCTGAAAAAGTGCC ACCTGCAGAT
	BDLF4KanaAS	ATCCATCACATACGTGGTACTCGCCACAGCCCCAGCAGGCAAACACGGACAGGAACACTT AACGGCTGA
	BDLF3.5KanaS	CACTATTGGGGAGAGAGAATTTGAGAATTGCTGAGCTGGGACCCACTGGAAAAAGTGCC ACCTGCAGAT

BDLF3.5KanaAS	GCTCAAAGAGACCTCCCACCGCCACGTAGACGCGGGCCACCGTCCGGGGACAGGAACACT TAACGGCTGA
---------------	--

Tableau 5 : Liste des différents primers utilisés pour cette étude

Western-Blot :

Les anticorps primaires utilisés pour les Western-Blot sont : un anticorps monoclonal anti-BZLF1 (MAb), Z125, un anticorps polyclonal lapin dirigé contre la protéine EB2, un anticorps monoclonal anti-Flag M2 (Sigma, F3165), un anticorps anti-ARN pol II polyclonal lapin (Santa Cruz, N-20 sc-899), un anticorps monoclonal (Covalab) dirigé contre la GST.

Les anticorps secondaires utilisés sont : un anti-souris et un anti-lapin couplés à la peroxydase (GE Healthcare Life Sciences). Les complexes anticorps-protéines formés sont révélés par chimioluminescence (ECL, Pierce) et visualisé sur un film autoradiographique.

Essais Luciférase :

L'activité firefly luciférase (Luc) des cellules transfectées a été mesurée avec le luminomètre GloMax (Promega) en utilisant le système Firefly Luciferase Assay (Promega). L'activité luciférase a été mesurée pour une quantité identique de protéines préalablement déterminée par Bradford.

Extraction d'ADN viral et analyse par qPCR:

Les cellules infectées par EBV sont récoltées 72h post-transfection, resuspendues et lysées dans du TNE selon la méthode Hirt, à 4°C, sur la nuit. L'ADN est ensuite extrait par phénol/chloroforme, traité à la RNase A (100ug/ml), précipité à l'éthanol pur et resuspendu dans de l'eau versol. Les niveaux d'ADN viraux sont déterminés par PCR quantitative (qPCR). L'ADN complémentaire (cDNA) de la luciférase firefly a été quantifié comme contrôle de purification.

Extraction d'ARN et analyse par RT-qPCR :

Les ARNm totaux ont été extraits et purifiés au Trizol, phénol/chloroforme, traités 1h à 37°C avec la DNase RQ1 puis à nouveau purifiés par phénol/chloroforme et précipités à l'éthanol pur. Ils ont ensuite subi une étape de transcription reverse en utilisant la qScript cDNA SuperMix (Quanta Biosciences). La RT-qPCR a été réalisée en utilisant le Fast-Start SYBR green master mix (Roche Molecular Biochemicals). Les primers utilisés sont listés dans le tableau 5 bis. L'ARNm du gène de ménage GAPDH (Glycéraldéhyde-3-phosphate

déshydrogénase) a été quantifié comme un contrôle interne des quantités d'ARN. La quantification des RT-qPCR a été effectuée comme déjà décrit (Juillard et al., 2012).

Nom Primers	Séquences	Nom Primers	Séquences
BMRF1s qPCR	AGGAGTGCTGCAGGTAAACC	BcLF1s qPCR	GGCTCAGTCTAAGG
BMRF1as qPCR	GCTCTGGTGATTCTGCCACT	BcLF1as qPCR	AGGTGGGCTGACACAGACTT
BDLF1s qPCR	CGCAGACATGCTCGATGTA	GAPDHs qPCR	AGCCACATCGTCTCAGACAC
BDLF1as qPCR	GTAGTGGTGCCCCAGGTATG	GAPDHas qPCR	GCCCAATACGACCAAATCC

Tableau 5 bis : liste des primers de qPCR.

GST pull-down *in vivo* :

Les plasmides d'expression pCi-Flag-GST ou pCI-Flag-GST-X (X = protéine virale) ont été transfectés dans des cellules HEK293 avec les plasmides d'expression codant pour les différentes protéines du vPIC et/ou du complexe de réplication contenant trois épitopes Flag chacune. 48h post-transfection, les cellules sont récoltées et lysées dans du tampon HEGN (20mM Hepses pH7.9, 300 mM KCl, 10% Glycérol, 0.1 mM EDTA, 0.1% NP40, 1mM DTT et 1mM PMSF). Les billes Glutathione-Sépharose (GE Healthcare) sont préalablement lavées cinq fois dans du HEGN avant d'être ajoutées aux lysats cellulaires afin de purifier les protéines chimériques virales couplée à la glutathione-S-transférase ainsi que les facteurs associés. L'incubation se fait à 4°C durant la nuit et les billes sont ensuite lavées 5 fois avec le tampon HEGN. Les protéines accrochées sont éluées et séparées par SDS-PAGE avant d'être analysées par Western-Blot. Les protéines du vPIC et du complexe de réplication sont révélées avec un anticorps anti-Flag, les GST-X avec un anticorps anti-GST et l'ARN pol II est visualisée avec un anticorps polyclonal anti-ARN pol II.

GST pull-down *in vitro* :

Les plasmides d'expression codant pour les différentes protéines du complexe de réplication (BALF2, BALF5, BMRF1, BBLF2/3, BBLF4, BSLF1 et EB1) ont été transcrits et traduits *in vitro* en lysat de réticulocytes (TNT coupled réticulocyte lysate System, Promega), en présence de méthionine marquée au soufre 35 (³⁵S). Ces lysats ont été incubés avec les différentes protéines chimériques GST-X (X = protéines du vPIC), préalablement produites en bactéries BL21 (Agilent) et purifiées sur billes Glutathione Sépharose. Les incubations ont été réalisées sur la nuit à 4°C. Les billes et complexes protéiques associés ont été lavés 5 fois dans du tampon MTPBS (150 mM NaCl, 100 mM EDTA, 16 mM Na₂HPO₄, 4mM NaH₂PO₄,

1% Triton, 1mM PMSF). Les protéines accrochées sont éluées et séparées par SDS-PAGE avant d'être révélée à l'aide d'un film autoradiographique.

Tandem Affinity Purification (TAP-tag) :

Les plasmides d'expression pNTAP-BcRF1 ou pCi-3xFlag-BcRF1 ont été transfectés dans des cellules HEK293 EBV BcRF1 avec le plasmide d'expression codant pour la protéine EB1 dans le but d'induire le cycle viral productif. Soixante heures post-transfection, les cellules ont été récoltées et lysées dans du tampon lyse (Agilent) en présence de 0.5X d'un cocktail d'inhibiteurs de protéase (Roche) et 1mM PMSF et ce, en les congelant à l'azote liquide et les décongelant trois fois successives. Les billes de Streptavidine du kit InterPlay Mammalian TAP System (Agilent) saturées à la BSA sont ajoutées aux lysats afin de purifier la protéine BcRF1 fusionnée à la Streptavidine. Les billes sont lavées avec le tampon de lavage (Agilent) après une incubation d'une nuit à 4°C. Les complexes associés sont élués de la colonne avec le tampon d'éluion Streptavidine (Agilent) avant d'être de nouveau incubé sur des billes Calmoduline du kit InterPlay Mammalian TAP System (Agilent) saturées à la BSA, à 4°C sur la nuit. Les billes calmodulines sont lavées dans le tampon de lavage calmoduline (Agilent) avant d'être éluées avec le tampon d'éluion Calmoduline (Agilent). Les protéines sont séparées par SDS-PAGE afin d'être coloré au nitrate d'argent (Promega). Les bandes sont découpées, décolorées avant d'être analysées par spectrométrie de masse.

II) Résultats :

A) L'expression des gènes viraux tardifs est dépendante d'un complexe d'initiation viral de la transcription : résumé de l'article scientifique publié en 2014 (Aubry et al., 2014).

Au cours de cette étude (cf. article), nous avons démontré que BFRF2 était impliquée *in vitro* et *in vivo* dans l'expression des gènes tardifs. Nous avons également caractérisé un complexe viral nécessaire et suffisant à l'expression des gènes viraux tardifs. Ce complexe est composé de six protéines différentes : BcRF1, BFRF2, BGLF3, BVLF1, BDLF4 et BDLF3.5. De plus, ce complexe est capable d'interagir avec l'ARN polymérase II cellulaire.

Nous avons ainsi décidé de nommer ce complexe vPIC pour « viral Pre-Initiation Complex » par homologie avec le PIC cellulaire. En effet, le vPIC contient une TBP-like virale (BcRF1), associée à cinq autres facteurs viraux capables de recruter l'ARN pol II. Nous démontrons également que l'activité fonctionnelle du vPIC est dépendante de la réplication de l'ADN viral.

De plus, les composants du vPIC d'EBV ont des homologues chez les β et γ -Herpesvirus mais pas chez les α -Herpesvirus. Nos résultats révèlent non seulement que ces virus codent pour leur propre complexe d'initiation de la transcription mais suggèrent aussi une évolution fonctionnelle de complexe entre ces sous-familles virales. L'ensemble de ces données est présenté dans l'article qui suit.

De part ces résultats, le mécanisme de régulation de l'expression des gènes tardifs d'EBV (que nous avons mis en évidence) se rapproche étrangement de celui du bactériophage T4 qui code pour des facteurs viraux mimant un complexe de transcription avec une « TBP-like ».

Epstein-Barr Virus Late Gene Transcription Depends on the Assembly of a Virus-Specific Preinitiation Complex

Valentin Aubry, Fabrice Mure, Bernard Mariamé, Thibaut Deschamps, Lucjan S. Wyrwicz, Evelyne Manet and Henri Gruffat

J. Virol. 2014, 88(21):12825. DOI: 10.1128/JVI.02139-14.
Published Ahead of Print 27 August 2014.

Updated information and services can be found at:
<http://jvi.asm.org/content/88/21/12825>

These include:

REFERENCES

This article cites 42 articles, 32 of which can be accessed free at: <http://jvi.asm.org/content/88/21/12825#ref-list-1>

CONTENT ALERTS

Receive: RSS Feeds, eTOCs, free email alerts (when new articles cite this article), [more»](#)

Information about commercial reprint orders: <http://journals.asm.org/site/misc/reprints.xhtml>
To subscribe to to another ASM Journal go to: <http://journals.asm.org/site/subscriptions/>

Epstein-Barr Virus Late Gene Transcription Depends on the Assembly of a Virus-Specific Preinitiation Complex

Valentin Aubry,^{a,b,c,d,e} Fabrice Mure,^{a,b,c,d,e} Bernard Mariamé,^f Thibaut Deschamps,^{a,b,c,d,e} Lucjan S. Wyrwicz,^g Evelyne Manet,^{a,b,c,d,e} Henri Gruffat^{a,b,c,d,e}

CIRI, Centre International de Recherche en Infectiologie, Oncogenic Herpesviruses Team, Université de Lyon, Lyon, France^a; INSERM, U1111, Lyon, France^b; Ecole Normale Supérieure de Lyon, Lyon, France^c; Université Lyon 1, Centre International de Recherche en Infectiologie, Lyon, France^d; CNRS, UMR5308, Lyon, France^e; CNRS, UMR 5099-LBME (Laboratoire de Biologie Moléculaire des Eucaryotes), USR 3505-ITAV (Institut des Technologies Avancées des Sciences du Vivant), Toulouse, France^f; Laboratory of Bioinformatics and Biostatistics and Department of Gastrointestinal Oncology, M. Skłodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland^g

ABSTRACT

During their productive cycle, herpesviruses exhibit a strictly regulated temporal cascade of gene expression that has three general stages: immediate early (IE), early (E), and late (L). Promoter complexity differs strikingly between IE/E genes and L genes. IE and E promoters contain *cis*-regulating sequences upstream of a TATA box, whereas L promoters comprise a unique *cis* element. In the case of the gammaherpesviruses, this element is usually a TATT motif found in the position where the consensus TATA box of eukaryotic promoters is typically found. Epstein-Barr virus (EBV) encodes a protein, called BcRF1, which has structural homology with the TATA-binding protein and interacts specifically with the TATT box. However, although necessary for the expression of the L genes, BcRF1 is not sufficient, suggesting that other viral proteins are also required. Here, we present the identification and characterization of a viral protein complex necessary and sufficient for the expression of the late viral genes. This viral complex is composed of five different proteins in addition to BcRF1 and interacts with cellular RNA polymerase II. During the viral productive cycle, this complex, which we call the vPIC (for viral preinitiation complex), works in concert with the viral DNA replication machinery to activate expression of the late viral genes. The EBV vPIC components have homologs in beta- and gammaherpesviruses but not in alphaherpesviruses. Our results not only reveal that beta- and gammaherpesviruses encode their own transcription preinitiation complex responsible for the expression of the late viral genes but also indicate the close evolutionary history of these viruses.

IMPORTANCE

Control of late gene transcription in DNA viruses is a major unsolved question in virology. In eukaryotes, the first step in transcriptional activation is the formation of a permissive chromatin, which allows assembly of the preinitiation complex (PIC) at the core promoter. Fixation of the TATA box-binding protein (TBP) is a key rate-limiting step in this process. This study provides evidence that EBV encodes a complex composed of six proteins necessary for the expression of the late viral genes. This complex is formed around a viral TBP-like protein and interacts with cellular RNA polymerase II, suggesting that it is directly involved in the assembly of a virus-specific PIC (vPIC).

Herpesviruses are enveloped viruses containing relatively large, double-stranded DNA genomes. They are divided into three subfamilies (alpha-, beta-, and gammaherpesviruses) according to sequence homology, cellular tropism, and productive cycle behavior under laboratory culture conditions. Nine herpesviruses have been identified in humans. Herpes simplex virus 1 (HSV-1) and 2 (HSV-2) and varicella-zoster virus (VZV) are alphaherpesviruses that due to neurotropism cause recurrent skin lesions, meningitis, and rare but very serious encephalitis in the case of HSV-1; human cytomegalovirus (HCMV), human herpesviruses 6A and 6B (HHV-6A and HHV-6B), and human herpesvirus 7 (HHV-7) are betaherpesviruses that cause severe diseases in patients with compromised immune function; and Epstein-Barr virus (EBV) and Kaposi's sarcoma-associated herpesvirus (KSHV) are gammaherpesviruses that are associated with a variety of cancers.

EBV is an orally transmitted herpesvirus that infects over 90% of the world's adult population. It is the causative agent of infectious mononucleosis and oral hairy leukoplakia, which results from EBV infection of epithelial cells along the side of the tongue (1). Moreover, EBV is also associated with several types of cancer, including Burkitt's lymphoma (BL), nasopharyngeal carcinoma

(NPC), Hodgkin's disease (HD), T/NK cell lymphoma, and B cell lymphoproliferations and lymphoma in immunocompromised patients. Following infection, EBV persists for life in the host by establishing latency in memory B cells (2). In EBV-associated cancers, the virus is present in a latent form of viral infection; the viral genome is replicated once per cell cycle by the host cell machinery, and only a subset of the viral genes is expressed. In contrast, infection of normal oropharyngeal epithelial cells by EBV results in a productive or lytic infection that is similar to that observed in oral hairy leukoplakia lesions. EBV also periodically reactivates in B cells, particularly when they are stimulated by antigen, and B cells

Received 21 July 2014 Accepted 19 August 2014

Published ahead of print 27 August 2014

Editor: R. M. Longnecker

Address correspondence to Henri Gruffat, henri.gruffat@ens-lyon.fr, or Evelyne Manet, evelyne.manet@ens-lyon.fr.

Copyright © 2014, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JVI.02139-14

consequently differentiate into plasma cells (3). The switch from latent to productive infection is mediated by the EBV immediate early (IE) proteins EB1 (also called Zta or ZEBRA) and Rta (4–6). EB1 and Rta are transcription factors that cooperatively activate expression of the EBV early (E) genes, many products of which are involved in viral DNA replication (7). The virally encoded replication complex amplifies the viral genome via the lytic origin of replication (OriLyt) (8). Following viral DNA replication, the late (L) viral genes, encoding mainly proteins required for assembly, maturation, and release of infectious particles, are expressed (9). This coordinated cascade of viral gene expression during the viral productive cycle relies on the host RNA polymerase II (RNAP II) and is observed in all herpesviruses.

Whereas the regulation of immediate early and early gene expression has been extensively studied in herpesviruses, very little is known about the mechanisms regulating late gene expression. Immediate early and early gene promoters have the same structure as cellular gene promoters transcribed by RNAP II; they typically consist of distal regulatory sequences upstream of the TATA box which recruit TBP (TATA box-binding protein) at the origin of the formation of the preinitiation complex (PIC) on the promoters. The structure of the late gene promoters appears to be less complex, since it has been shown (in all cases studied so far) that a small proximal region is sufficient to control late gene expression (10–15). Interestingly, for the gammaherpesviruses, an unusual TATT sequence is often present where early genes and many cellular genes display a TATA box (11, 12, 15–17). In addition, while early genes are expressed independently of viral DNA synthesis, late genes are not expressed when viral DNA synthesis is blocked by specific inhibitors (10).

Until now, the mechanisms by which viral proteins are implicated in the regulation of the late gene promoters have not been precisely defined. In the case of HSV-1, several viral proteins (ICP4, ICP8, and ICP27) have been shown to be necessary for the efficient expression of late genes. These factors interact with the general transcription machinery and may thus function to facilitate the assembly of the transcription preinitiation complex on the late promoters. However, since these viral proteins also play an important role in the regulation of the early promoters, the specific molecular mechanisms by which they control the delayed expression of the late genes is unknown. For betaherpesviruses like CMV and gammaherpesviruses like mouse herpesvirus 68 (MHV-68), several viral proteins have been found to be involved in regulating late gene expression. These factors are expressed by beta- and gammaherpesviruses but not by the alphaherpesviruses. We have recently identified and characterized the function of one of these proteins: BcRF1 of EBV. BcRF1 was found, by *in silico* studies, to have structural homology with TBP from the thermophilic archaeon *Pyrococcus woesei* (17), and we recently showed that BcRF1 interacts specifically with the TATT sequence present on the promoters of late EBV genes, demonstrating for the first time that EBV encodes a TBP-like protein that is absolutely required for late viral gene expression. Moreover, we showed that BcRF1 alone had no effect on the expression of the late viral genes, suggesting that, to be functional, its coexpression of other viral proteins expressed during the viral productive cycle is required (18). These results, together with the finding that several HCMV or MHV-68 factors of unknown function are essential for late viral gene expression, suggest that control of late transcription via

RNAP II in beta- and gammaherpesviruses relies on a specific late transcription preinitiation complex (19).

Here, we present the identification of the complete set of EBV genes required for late gene expression. We show that six viral proteins are necessary and sufficient for activation of the late viral genes. These six proteins, which include BcRF1, the viral TBP-like protein, form a complex that is required for EBV late gene transcription. This complex also recruits the cellular RNAP II enzyme. We have called this complex the viral preinitiation complex (vPIC), which is analogous to the cellular PIC. Moreover, we show that the HCMV proteins analogous to EBV vPIC factors are also sufficient to activate transcription from a model late gene promoter, although they are not able to complement an EBV vPIC from which individual proteins are missing. These data indicate that conserved beta- and gammaherpesvirus gene products control the transcription of late viral gene expression and suggest that the viral TBP-like proteins (BcRF1 in the case of EBV) recruit a virus-specific preinitiation complex on the TATT sequence in a manner similar to the recruitment of the cellular preinitiation complex by cellular TBP.

MATERIALS AND METHODS

Cell lines and transfections. HEK293-EBV cells are derived from HEK293 cells infected with a bacterial artificial chromosome in EBV (BAC-EBV; kindly provided by W. Hammerschmidt). All cell lines were routinely grown in RPMI 1640 medium supplemented with 10% fetal calf serum. Transient transfections of the cells and derivatives were performed by electroporation (950 μ F, 220 mV) using a Bio-Rad electroporator.

Recombinant plasmids. An EB1 expression plasmid was used for initiation of the EBV productive cycle (20). The BcRF1 expression vector has previously been described (18). The entire BFRF2, BGLF3, BVL1, BDLF4, and BDLF3.5 open reading frames (ORFs) were PCR amplified, cloned using the in-fusion system (Clontech) into the pCI expression plasmid (Promega) under the control of a CMV promoter, and fused at their 5' ends to the Flag epitope. The entire HCMV UL49, UL79, UL87, UL91, UL92, and UL95 genomic sequences were PCR amplified with specific primers containing unique restriction sites, and the FLAG epitope was fused to their 5' ends. PCR products were digested with the appropriate restriction enzymes and cloned into the corresponding sites of the pCIneo vector (Promega) under the control of a CMV promoter. The primer sequences are given in Table 1. The pTATT_{BcLF1}-Luc reporter plasmid has previously been described (18). It contains sequences from position –38 to –4 from the viral late gene BcLF1, including the TATT element.

Generation of a BFRF2-KO recombinant EBV. To generate a BFRF2 knockout mutant in the context of the EBV genome, we used the maxi-EBV technology described in the work of Delecluse et al. (21). The EBV Δ BFRF2 mutant was constructed by replacing part of the BFRF2 ORF (B95.8 coordinates 47668 to 48700; GenBank accession number AIE88895) with the kanamycin (*kan*) resistance gene using homologous recombination (22). To remove the kanamycin cassette, the *Escherichia coli* bacteria harboring BAC-EBV Δ BFRF2 were transformed with the plasmid pCP20, encoding *red* recombinase, which allowed the cassette's elimination. Recombinant EBV plasmid DNA (10 μ g) was then transfected into HEK293 cells using polyethylenimine (PEI) reagent. One day post-transfection, hygromycin (100 μ g/ml) was added to the culture medium for selection of stable HEK293 clones carrying the EBV recombinant plasmid. Outgrowing green fluorescent protein (GFP)-positive colonies were amplified for further investigation. The cell clone used in this study is referred to as HEK293_{EBV Δ BFRF2}.

Induction of the virus productive cycle. Producer cell clones HEK293_{EBV} (kindly provided by W. Hammerschmidt) and HEK293_{EBV Δ BFRF2} were transfected with an EB1 expression plasmid (1 μ g/plate) to induce the viral productive cycle. In *trans*-complementation assays, HEK293_{EBV Δ BFRF2}

TABLE 1 List of the different primers used in this study

Primer name	Sequence
BFRF2 ORFs	AAGGGATCCCCTCGAGAAATGGCGTTATTCTTGCGCGC
BFRF2 ORFas	CCGGGTCGACTCTAGATTAGGAAGCAGGGGACTGTC
BGLF3 ORFs	AAGGGATCCCCTCGAGAAATGTTCAACGCGGTG
BGLF3 ORFas	CCGGGTCGACTCTAGACTACTCATCTTCATAAGTC
BVLF1 ORFs	AAGGGATCCCCTCGAGAAATGTTAATGGGACTGG
BVLF1 ORFas	CCGGGTCGACTCTAGACTACCTGGCCTCCCCGGTG
BDLF4 ORFs	AAGGGATCCCCTCGAGAAATGTCAGACCAAGGCC
BDLF4 ORFas	CCGGGTCGACTCTAGATCAACACTTGGTTGTCAATGTGG
BDLF3.5 ORFs	AAGGGATCCCCTCGAGAAATGTCTGCCCCGGATGCTC
BDLF3.5 ORFas	CCGGGTCGACTCTAGATCAATCGGCCTTGGTCTGAC
Virus Δ BFRF2s ^d	CAGGCCCTTGGGGAGCGGGGGTCTCCAGGCTCTGGATCTGGGGCTGGCGAAAAGTGCCACCTGCAGA
Virus Δ BFRF2as ^d	GGAATTGAGCCTGCCTCTACACCTGTCTGCTTCCAAAGTTCAAGCAGTCAGGAACACTTAACGGCTG
BFRF2 a	ATGGCGTTATTCTTGCGCGC
BFRF2 b	CTCACCGGATTACAGTCG
BFRF2 c	TAGATTGTCGCACCTG
BFRF2 d	ATTAGGAAGCAGGGGACTGTC
BFRF2 e	GCCTCGTGAAGAGGTTCTTG
BMRF1s qPCR	AGGAGTGCTGCAGGTAAACC
BMRF1as qPCR	GCTCTGGTGATTCTGCCACT
BDLF1s qPCR	CGCAGACATGCTCGATGTA
BDLF1as qPCR	GTAGTGGTGCCCCAGGTATG
BcLF1s qPCR	GGCTCAGTCTAAGG
BcLF1as qPCR	AGGTGGGCTGACACAGACTT
GAPDHs qPCR	AGCCACATCGCTCAGACAC
GAPDHas qPCR	GCCCAATACGACCAAATCC
BFRF3s qPCR	AGAGGCAGAGAGCCAGTGTG
BFRF3as qPCR	CCGGAGGCTGCTAATAGATG
FIUL49-S	CGCCGCCACCATTGGACTACAAAGACGATGACGATAAGGCCAGTCGTCGTCTCCGAC
UL49-AS	AAAAGCGGCCGCGGTGTTAGACATGGGG
FIUL79S	AAAAACGCGTGCCGCCACCATTGGACTACAAAGACGATGACGATAAGATGATGGCCCGCGACG
UL79AS	ATGAGCGGCCGCGCTGTAC
FIUL87S	AAAAGAATTGCGCGCCACCATTGGACTACAAAGACGATGACGATAAGGCCGCGCTGCGC
UL87AS	AAAAGTCGACGCTCCTCCGGACGAAAC
FIUL91S	TGCCGCCACCATTGGACTACAAAGACGATGACGATAAAGAACTCGTTGCTGGCGGAAAC
UL91AS	AAAAGCGGCCGCGTCTGGACGTGCC
FIUL92-S	CGCCGCCACCATTGGACTACAAAGACGATGACGATAAGTGCGACGCTCCGGG
UL92-AS	AAAAGCGGCCGCGCTTCAAACGCC
FIUL95S	AAAAGCTAGCCGCCACCATTGGACTACAAAGACGATGACGATAAGATGATGGCGCGGGC
UL95AS	AAAACGCGTATCACGCTCTTAAAGAGCTGTTGTGG
IGUSB/F29	GTGCTGGGGAATAAAAAGGGG
IGUSB/F29	ACTCGGGGAGGAAGGGACAC

^d For the Δ BFRF2 virus primers, the underlined sequences allowed PCR-mediated amplification of the kanamycin resistance gene. The other primer sequences are homologous to BFRF2 sequences and are used for recombination of the amplified PCR product within the BFRF2 coding sequence.

cells were cotransfected with 1 μ g of plasmid pCi-Flag-BFRF2 or the empty plasmid. Two days posttransfection, cells were harvested for analysis of viral gene expression.

Viral DNA replication analysis. EBV-infected cells were collected 72 h posttransfection, resuspended in lysis buffer (23), and incubated at 4°C overnight. DNA was extracted with phenol-chloroform, treated with RNase A (100 μ g/ml), precipitated with ethanol, and then resuspended in nuclease-free water. Viral DNA levels were determined by quantitative PCR (qPCR).

Western blot analysis. Antibodies used for Western blotting were an anti-BZLF1 monoclonal antibody (MAb), Z125 (20), a rabbit polyclonal serum directed against the EB2 protein (24), an anti-Flag M2 MAb (Sigma; F3165), an anti-Flag polyclonal antibody (Sigma; F7425), and an anti-RNAP II polyclonal antibody (Santa Cruz; N-20 sc-899), as well as anti-mouse or anti-rabbit horseradish peroxidase-conjugated secondary antibodies (GE Healthcare Life Sciences). Immune complexes were visualized using enhanced-chemiluminescence (ECL) reagent (Pierce).

RNA extraction and RT-qPCR analysis. Total RNA was extracted and purified using the NucleoSpin RNA kit (Macherey-Nagel) and reverse transcribed using qScript cDNA SuperMix (Quanta Biosciences). Real-time reverse transcription-PCR (RT-qPCR) was performed using Fast-Start universal SYBR green master mix (Rox; Roche Molecular Biochemicals) and primers listed in Table 1. Cellular glyceraldehyde-3-phosphate dehydrogenase (GAPDH) RNA was quantified as an internal control for the amounts of RNA. Quantification of the RT-qPCR results was done as previously described (25).

Luc assays. Firefly luciferase (Luc) activity from transfected cells was measured in a Veritas luminometer (Turner Biosystems) using the Firefly luciferase assay system (Promega). Luciferase activity was measured for identical amounts of total protein as evaluated by the Bradford assay.

In vivo GST pulldown assays. The pCi-GST.Flag or pCi-GST.Flag-BcRF1 expression plasmid was transfected into HEK293 cells together, or not, with expression vectors for the vPIC factors. Forty-eight hours after transfection, cells were lysed in HNTG buffer (20 mM HEPES, pH 7.9, 180

TABLE 2 List of analogous genes between beta- and gammaherpesviruses which are not present in alphaherpesviruses^a

EBV strain	Analogous gene in:				Reference(s)
	KSHV	MHV-68	CMV	HHV-6	
BVLF1	ORF18	ORF18	UL79	U52	33, 26
BcRF1	ORF24	ORF24	UL87	U58	18, 28
BDLF3.5	ORF30	ORF30	UL91	U62	30, 27
BDLF4	ORF31	ORF31	UL92	U63	31, 29
BGLF3	ORF34	ORF34	UL95	U67	27
BFRF2	ORF66	ORF66	UL49	U33	This publication
BTRF1	ORF23	ORF23	UL88	U59	34

^a For more information, see reference 45. References for the genes' implications in late viral gene expression are indicated in the extreme-right column.

mM NaCl, 10% glycerol, 0.1 mM EDTA, 0.1% NP-40, 1 mM dithiothreitol [DTT], 1 mM phenylmethylsulfonyl fluoride [PMSF]). Glutathione-Sepharose beads saturated with bovine serum albumin (BSA) were added to the lysate in order to purify the glutathione S-transferase (GST) fusion proteins and their associated proteins. After 4 h of incubation, the beads were washed 5-fold with HNTG buffer and bound proteins were fractionated by SDS-PAGE and analyzed by Western blotting. The vPIC proteins were visualized with anti-Flag antibodies, and the RNAP II protein was visualized with rabbit polyclonal antibody sc-899 (Santa Cruz).

RESULTS

Identification of viral proteins required for late gene expression.

We have previously shown that BcRF1 is necessary, but not sufficient alone, to activate late gene transcription (18), suggesting that other viral proteins are required. In the case of MHV-68, five proteins (ORF18, ORF24, ORF30, ORF31, and ORF34) were found to be necessary for the expression of the MHV-68 late genes. Analogs for these proteins in CMV (UL79, UL87, UL91, UL92, and UL95, respectively) were also found to be necessary for late gene expression (26–33). Interestingly, EBV also encodes five putative analogs (BVLF1.5, BcRF1, BDLF3.5, BDLF4, and BGLF3, respectively). We have already shown by the generation of a recombinant virus (EBV_{ΔBcRF1}) that BcRF1 is required for late gene expression (18). Together, these data suggest that a conserved set of proteins implicated in the control of late gene expression is conserved between beta- and gammaherpesviruses. Interestingly, these genes are not conserved in the alphaherpesviruses. Based on this observation, we established a list of the early viral proteins that are conserved in the betaherpesvirus (CMV and HHV6) and gammaherpesvirus (EBV, KSHV, and MHV-68) subfamilies but absent in HSV-1 (Table 2). From this comparison, we identified two proteins in addition to those cited above, called BTRF1 and BFRF2 in the case of EBV. The analog of BTRF1 in MHV-68 (ORF23) has previously been shown to be the product of a late gene and dispensable for virus production both in cell culture and *in vivo* (34). Interestingly, however, BFRF2 is an uncharacterized protein of unknown function. We thus decided to further explore the role of this protein in the expression of the late viral genes of EBV.

Construction of a BAC-EBV recombinant clone carrying a deletion in the BFRF2 gene. In order to study the function of BFRF2 in the EBV productive cycle, a BAC-EBV recombinant named BAC-EBV_{ΔBFRF2} was constructed in a manner that did not interfere with the expression or functions of adjacent genes. For this, we introduced a large deletion in the BFRF2 ORF (Fig. 1A). The BFRF2 early gene is located between the early BFRF1 gene and the late BFRF3 gene. The BFRF2 gene is transcribed from the same promoter as the BFRF1 gene, and the ORF of the BFRF2 gene has

been sized to 1,775 bp, coding for a 64-kDa protein. To generate the BAC-EBV_{ΔBFRF2} mutant, a 1,032-bp fragment from the BFRF2 gene (B95.8 coordinates 47668 to 48700) was exchanged for the kanamycin resistance gene sequence by homologous recombination in *E. coli* while leaving the first 148 bp from the BFRF2 ORF intact in order to avoid disruption of the transcriptional regulation of its neighboring genes. In a second step, the kanamycin resistance gene was deleted to generate the BAC-EBV_{ΔBFRF2ΔKana} recombinant (Fig. 1A). The genome integrity of all mutants and control wild-type BAC-EBV was verified by BamHI restriction enzyme digestion. As expected from the cloning strategy, the restriction profile confirmed that the 7,396-bp BamHI fragment that contains the BFRF2 gene had been replaced by a 7,442-bp fragment in the ΔBFRF2 mutant and reduced to a 6,365-bp fragment after deletion of the kanamycin gene (Fig. 1B). Moreover, PCR amplification of the BFRF2 region with primers located on each side of the BFRF2 ORF gave a fragment of 1,821 bp for BAC-EBV_{ΔBFRF2} and 744 bp for BAC-EBV_{ΔBFRF2ΔKana}, compared with the 1,780-bp fragment coming from wild-type BAC-EBV (Fig. 1C, primers a and d). As expected, PCR amplification with primers inside the kanamycin gene gave a fragment only with the BAC-EBV_{ΔBFRF2} recombinant (1,288 bp with primers c and d or 591 bp with primers a and b). BFRF2 ORF internal primers amplified only the wild-type BAC-EBV genome, giving a fragment of 230 bp, as expected (Fig. 1C, primers e and f). BAC-EBV_{ΔBFRF2ΔKana} DNA was prepared and transfected into HEK293 cells in order to establish an HEK293 cell line harboring this mutated virus, hereinafter referred to as HEK293_{EBVΔBFRF2}. We screened several clones for their permissivity for the EBV productive cycle. The cell clones were transfected with the EB1 expression plasmids in order to induce the viral productive cycle, and after 48-h cell extracts were prepared, we verified by Western blotting expression of the viral early protein EB2 (the product of the BMLF1 gene). EB2 was used as a marker for the induction of the viral productive cycle. We then selected the cell clone with the highest expression level of EB2 to perform further experiments.

BFRF2 is required during the late phase of the viral productive cycle. In order to test the production of progeny virus by HEK293_{EBVΔBFRF2} cells, they were transfected with an expression plasmid encoding the EBV transcription factor EB1 to induce the viral productive cycle, either alone or together with an expression plasmid for the BFRF2 protein. Two days after transfection, cell supernatants were used to infect Raji cells. The presence of infectious EBV was visualized indirectly by GFP expression in infected Raji cells as previously described (21, 35). As shown in Fig. 2A, 2 days postinfection, no green Raji cells were detected with super-

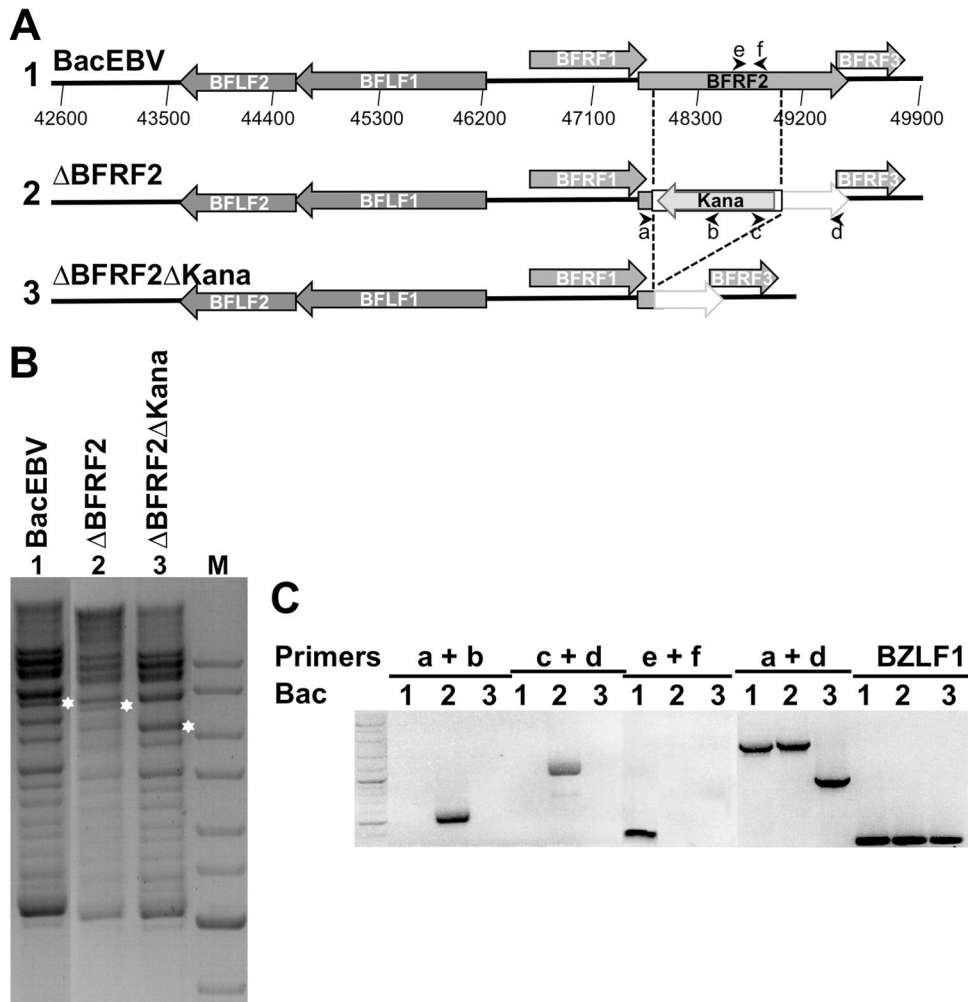


FIG 1 Construction of the BacEBV Δ BFRF2 recombinant. (A) Schematic representation of the EBV region surrounding the BFRF2 gene (1). Homologous recombination between the wild-type EBV genome (p2089) and a PCR fragment carrying the kanamycin gene flanked by 50 bp of BCRF1 sequences was performed in the DH10B *E. coli* strain containing the pKD46 plasmid. Clones carrying the recombinant EBV were selected for chloramphenicol and kanamycin resistance. The resulting BAC-EBV Δ BFRF2 recombinant had the kanamycin gene inserted between positions 47668 and 48700 in the BFRF2 gene (2). The kanamycin gene was then deleted by homologous recombination using the pCP20 vector in order to generate the BAC-EBV Δ BFRF2 Δ Kana recombinant used in this study (3). (B) Restriction enzyme analysis of the BAC-EBV Δ BFRF2 plasmid DNA. The BamHI restriction pattern of BAC-EBV Δ BFRF2 (lane 2) and BAC-EBV Δ BFRF2 Δ Kana (lane 3) mutant DNA was compared to that of wild-type BAC-EBV DNA (lane 1). An asterisk indicates the modified DNA pattern. Lane M, molecular markers. (C) PCR amplification analysis of the BAC-EBV Δ BFRF2 (lanes 2) and BAC-EBV Δ BFRF2 Δ Kana (lanes 3) (Δ F2) clones compared to the BAC-EBV clone (lane 1) carrying the wild-type EBV genome. PCRs were performed using primers indicated by small arrowheads in panel A (a, b, c, d, e, and f) and listed in Table 1.

natants from cells transfected with the EB1 expression vector alone. However, with the supernatants from cells transfected with expression plasmids for both EB1 and BFRF2, a large number of green Raji cells was observed (Fig. 2A), indicating that ectopic expression of BFRF2 can rescue virus production from the HEK293 Δ BFRF2 cells. Quantification by fluorescence-activated cell sorting (FACS) analysis showed that an average of 8% of the Raji cells were GFP positive after infection with the supernatant of HEK293 Δ BFRF2 cells transfected with expression plasmids encoding EB1 and BFRF2, whereas no GFP-positive Raji cells were detected after infection with the supernatant of HEK293 Δ BFRF2 cells transfected with expression plasmids encoding EB1 alone. These results clearly demonstrate that the EBV BFRF2 gene is essential for the production of infectious EBV particles.

We hypothesized that, as a putative late gene expression regu-

lator, BFRF2 should affect neither the expression of the viral early genes nor the OriLyt-dependent replication of the viral genome. In order to test this, HEK293 Δ BFRF2 cells or HEK293 Δ EBV cells as a control were either mock transfected or transfected with an expression vector for EB1 so as to activate EBV's productive cycle, together, or not, with a BFRF2 expression plasmid. Forty-eight hours posttransfection, expression of the viral early protein EB2 (BMLF1 or Mta) was monitored by Western blotting (Fig. 2A) and viral DNA was purified and quantified by qPCR (Fig. 2B). As expected, the viral early protein EB2 was expressed in cells induced in the productive cycle by ectopic expression of the viral immediate early protein EB1, and its expression was not significantly modified when BFRF2 was coexpressed in HEK293 Δ BFRF2 cells (Fig. 2A). Quantification of the viral DNA showed that induction of the productive cycle by EB1 led to a strong increase in the

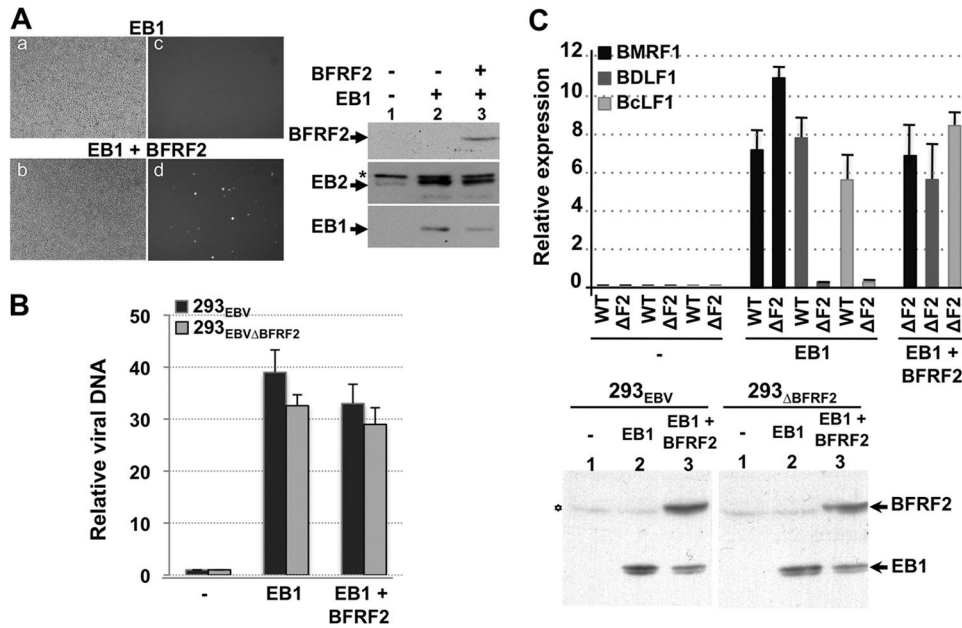


FIG 2 BFRF2 is necessary for virion production. (A) Raji cells were incubated with supernatants from HEK293_{EBVΔBFRF2} cells transfected with an expression plasmid encoding the EB1 protein in order to induce the productive cycle together (or not) with an expression plasmid encoding the Flag.BFRF2 protein to complement BFRF2 deficiency. (Left side) The GFP fluorescence of Raji cells was analyzed 48 h after infection under UV light. (A, sub panels a and b) Phase contrast; (A, sub panels c and d) UV light. (Right side) Immunoblot analysis of the EB2, Flag.BFRF2, and EB1 proteins expressed in HEK293_{EBVΔBFRF2} cells that either had not been transfected (lane 1), had been transfected with an EB1 expression plasmid (lane 2), or had been cotransfected with expression plasmids for both EB1 and BFRF2 (lane 3). Western blots were revealed using a polyclonal anti-Flag antibody, an anti-EB2 polyclonal rabbit antibody, or an anti-EB1 MAb. * indicates an unspecified protein recognized by the anti-EB2 serum. (B) BFRF2 is not required for viral DNA replication. Viral DNA in HEK293_{EBV} and HEK293_{EBVΔBFRF2} cells transfected with EB1 and Flag.BFRF2 expression plasmids, as indicated in the figure, was quantitated by qPCR. The amount of viral DNA in the induced cells (transfected with the EB1 expression plasmid) is expressed relative to the amount of viral DNA present in the uninduced cells. (C, top) BFRF2 enhances expression of the late viral genes. Total RNA from HEK293_{EBV} (wild type [WT]) or HEK293_{EBVΔBFRF2} (Δ F2) cells that either had not been transfected (-), had been transfected with an EB1 expression plasmid (EB1), or had been cotransfected with expression plasmids for EB1 and Flag.BFRF2 (EB1 + BFRF2) was purified, reverse transcribed, and analyzed by qPCR using specific primers for either the BMRF1 early gene or the BDLF1 and BcLF1 late genes. (Bottom) Western blot of HEK293_{EBV} or HEK293_{EBVΔBFRF2} cells either not transfected (-) or transfected with an EB1 expression plasmid or with expression plasmids for both EB1 and Flag.BFRF2. Protein extracts were analyzed by Western blotting using a polyclonal anti-Flag antibody and an anti-EB1 MAb. The asterisk indicates an unspecified protein detected by the anti-Flag antibody.

amount of viral DNA present in HEK293_{EBVΔBFRF2} cells, comparable to that observed in HEK293_{EBV} cells. However, coexpression of BFRF2 in HEK293_{EBV} or HEK293_{EBVΔBFRF2} cells did not change the amount of newly replicated viral DNA, demonstrating that the BFRF2 protein is not required for OriLyt-dependent viral DNA replication (Fig. 2B). Taken together, these results demonstrate that despite being an early protein, BFRF2 neither affects expression of the early viral proteins nor participates in viral DNA replication.

Next, in order to define BFRF2's impact on viral late gene transcript levels, we analyzed the accumulation of viral RNA for some early and late genes. Total RNA was extracted from HEK293_{EBV} or HEK293_{EBVΔBFRF2} cells that were mock transfected, transfected with an EB1 expression vector, or transfected with both an EB1 and a BFRF2 expression vector. RNAs were then quantified by RT-qPCR using primers specific for the early BMRF1 gene or the late BDLF1 and BcLF1 genes. In mock-transfected HEK293_{EBV} or HEK293_{EBVΔBFRF2} cells, the early or late mRNAs were, as expected, not detectable, whereas in cells transfected with an EB1 expression plasmid, the early BMRF1 transcript accumulated at comparable levels for the mutant and control viruses (Fig. 2C). In contrast, the late BDLF1 and BcLF1 mRNA levels were remarkably low in cells harboring the BFRF2-mutated virus compared to those in cells infected with the wild-type BAC-EBV. However, after coexpres-

sion of BFRF2 in HEK293_{EBVΔBFRF2} cells, expression of the late BDLF1 and BcLF1 mRNAs was clearly enhanced (Fig. 2C).

The implication of BFRF2 in late viral gene expression was then confirmed in a transient-transfection assay by using the reporter vector pTATT_{BcLF1}-Luc, which carries typical late promoter sequences (Fig. 3A). The results (Fig. 3B) showed that in HEK293_{EBVΔBFRF2} cells, induction of the pTATT_{BcLF1}-Luc reporter required both the activation of the viral productive cycle by EB1 and the expression of the BFRF2 protein. These results are very similar to those obtained using HEK293_{EBVΔBcRF1} cells, where induction of the pTATT_{BcLF1}-Luc reporter required both activation of the viral productive cycle by EB1 and expression of the BcRF1 protein (Fig. 3B) (18). Taken together, the above-described results demonstrate that BFRF2, like BcRF1, is necessary but not sufficient for efficient transcription of the late viral genes.

Six viral proteins are necessary and sufficient for the activation of the late viral genes of EBV. The data previously obtained with different beta- and gammaherpesviruses and the above-described results suggest that six viral proteins are required for the activation of the late viral genes of EBV. In order to show that the products of these six viral genes are not only necessary but also sufficient for the transcriptional activation of the late viral genes, we transfected their corresponding expression vectors into EBV-negative HEK293 cells, together with the pTATT_{BcLF1}-Luc

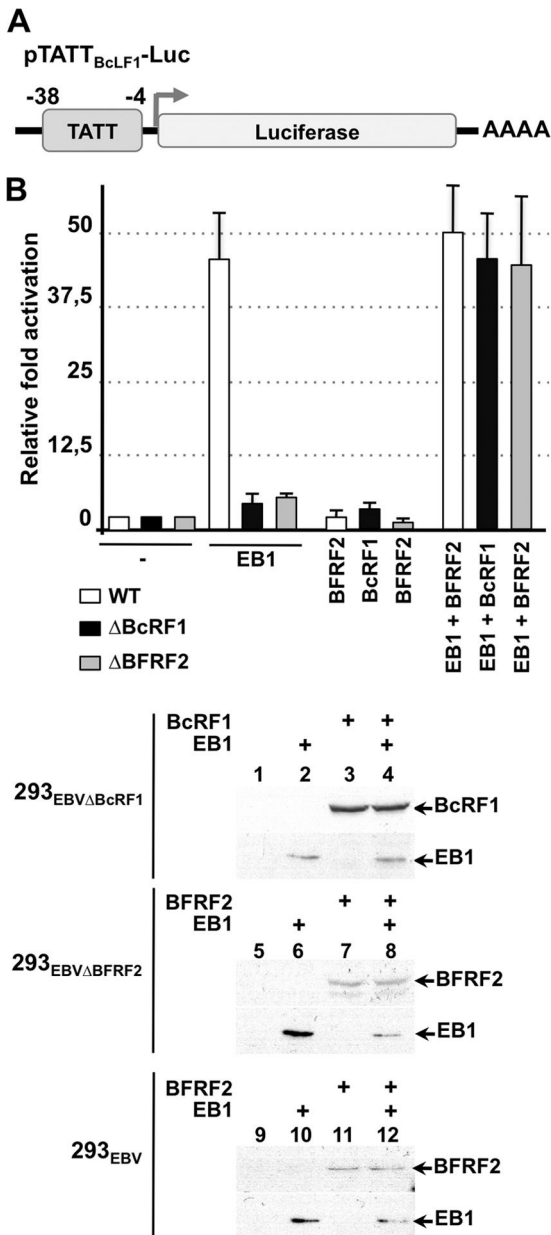


FIG 3 BFRF2 is required for late gene expression. (A) Schematic representation of the luciferase reporter vector used (pTATT_{BcLF1}-Luc). (B) HEK293_{EBV}, HEK293_{EBVΔBcRF1}, or HEK293_{EBVΔBFRF2} cells were either transfected with the reporter vector alone (-) or cotransfected with an expression plasmid for EB1, Flag.BcRF1, or Flag.BFRF2, as indicated. The amount of luciferase expressed was measured 24 h after transfection. The relative luciferase fold activation was calculated relative to the value obtained with the pTATT_{BcLF1}-Luc reporter transfected alone. The experiment was performed 3 times, and the error bars represent standard deviations. Protein extracts from the experiment whose results are shown in panel B were tested for EB1, BcRF1, and BFRF2 expression by immunoblotting using an anti-EB1 MAb, an anti-Flag MAb (BcRF1), and a polyclonal anti-Flag antibody (BFRF2).

porter plasmid. The results of the luciferase assays (Fig. 4A) show that none of these viral proteins alone are able to activate the expression of the luciferase reporter. However, when the cells were cotransfected with the six different expression vectors, an average 10-fold induction of luciferase activity was observed, suggesting

that, together, the six proteins can activate late viral gene expression. When the cells were cotransfected with various combinations of only five of the expression vectors, the luciferase activity was not induced (Fig. 4C). All six viral proteins are thus required for activation of viral late gene expression. In contrast, coexpression of the six viral proteins had no effect on the level of luciferase expressed under the control of the viral early pBMRF1 promoter, whereas EB1 strongly induced this expression (Fig. 4E). However, EB1 was unable to induce the expression of the luciferase gene controlled by the BcLF1 minimal promoter (Fig. 4A). Taken together, these results suggest that the six viral proteins are required for the activation of the late viral genes of EBV.

Further analysis of the proteins expressed from cells transfected with the different expression vectors separately showed that most of the proteins expressed in the assay are individually unstable. Except for BcRF1 and BDLF4, all the other proteins were barely detectable when expressed alone. Furthermore, although detectable, BcRF1 was very sensitive to degradation (Fig. 4B). However, when coexpressed, all the proteins, except BDLF3.5, likely due to its small size, were easily detectable (Fig. 4B, lane 8), suggesting that these proteins form a complex which allows their stabilization. Interestingly, when HEK293 cells were transfected with various combinations of five of the viral proteins, we found that the expression of BFRF2 was required to observe detectable levels of BGLF3 (but not the reverse) and that BVLf1 was stabilized by the coexpression of BDLF3.5 (Fig. 4D and F). It is important to note that BFRF2 and BDLF3.5 do not act at the transcriptional level on the accumulation of, respectively, BGLF3 and BVLf1, as shown by RT-PCR analysis (Fig. 4G). Taken together, these results suggest that interactions between the different proteins are necessary for their stabilization and also that these six proteins are required for the formation of a basic transcription initiation complex able to activate the expression of the late viral genes. Because of the presence in the complex of BcRF1, the viral TBP-like protein, and by analogy to the cellular preinitiation complex formed around the cellular TBP, we decided to call this complex vPIC, for viral preinitiation complex.

The six viral proteins of the vPIC form a complex associated with RNAP II. In order to further analyze the interactions between the vPIC components, we transfected HEK293 cells with a GST-Flag or a GST-Flag.BcRF1 expression plasmid together with expression vectors for the vPIC viral components (each tagged with the Flag epitope). Forty-eight hours after transfection, the cell lysates were used in a GST pull-down assay. Five of the six vPIC components were found associated with GST-Flag.BcRF1 but not with GST (Fig. 5A). However, it was not possible to conclude this for the sixth protein, BDLF3.5, as this protein was already barely detectable in the input. It is interesting to note that BcRF1 itself was pulled down with GST-Flag.BcRF1, suggesting that BcRF1 either forms homodimers or is engaged in interactions with several components of the complex. This result confirms that the six viral proteins are involved in the same complex. This was also seen in *in vitro* GST pull-down assays in which *in vitro*-translated proteins labeled with [³⁵S]methionine were incubated with the different vPIC proteins fused to GST and produced in bacteria (Fig. 5B). The results of these *in vitro* GST pull-down assays show that each protein appears to interact both with itself and with all the others, although interactions detected with BVLf1 were much weaker. The fact that BVLf1 was very efficiently pulled down with GST-Flag-BcRF1, together with the rest of the vPIC components from

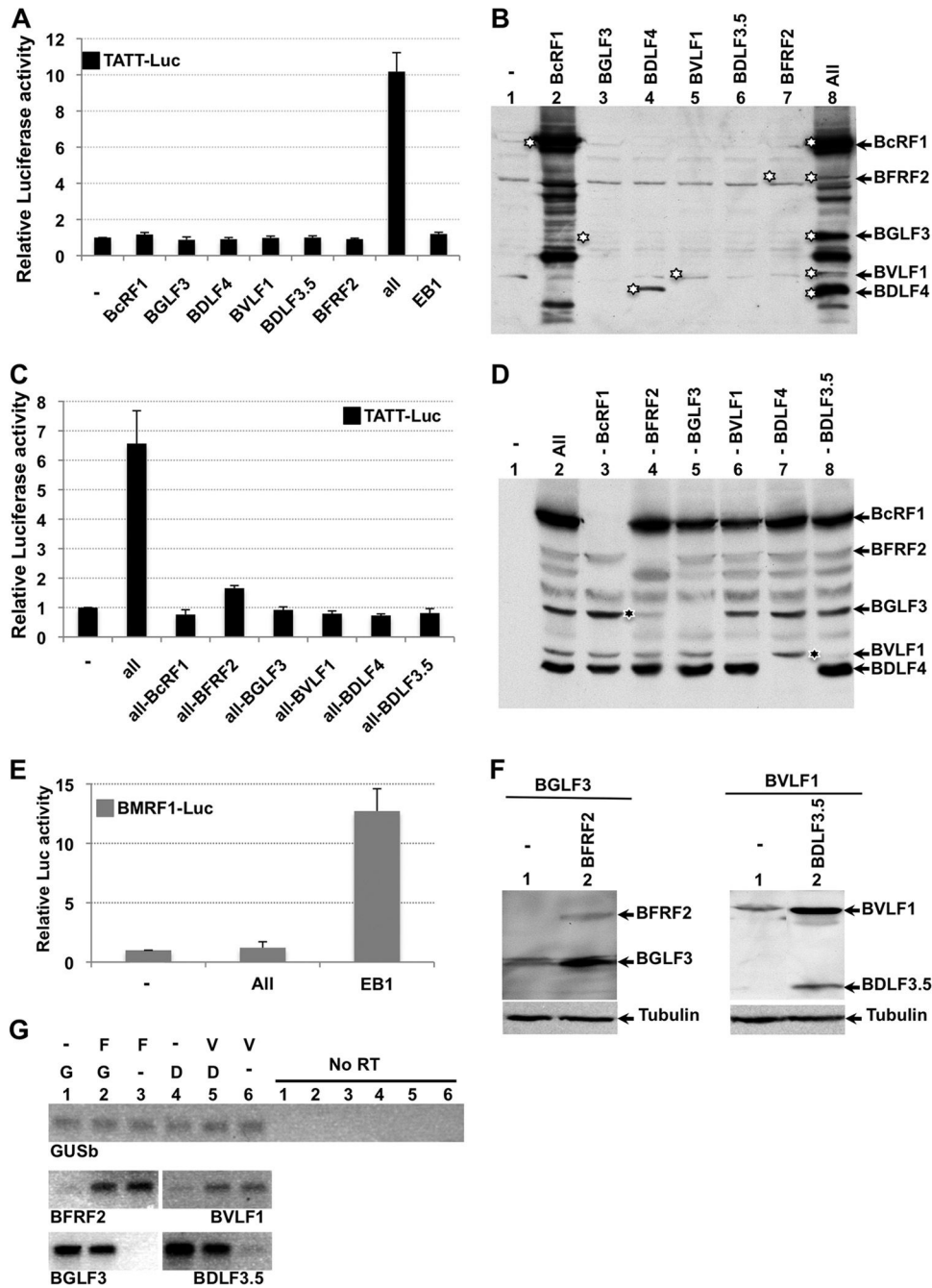


FIG 4 Six viral proteins are required for activation of the late viral genes. (A) The luciferase (Luc) reporter vector pTATT_{B_{CLF1}}-Luc was transfected into the EBV-negative cell line HEK293, together with expression vectors for the six different proteins, separately or combined. An EB1 expression vector was also used in order to show that the response of the TATT_{B_{CLF1}} promoter was specific for the other six proteins. (B) Proteins expressed in HEK293 cells transfected with the different expression plasmids encoding the Flag-tagged proteins used in panel A were analyzed by Western blotting using a monoclonal anti-Flag antibody. White asterisks label the expected positions of the different proteins on the gel. — indicates the missing protein. (C) HEK293 cells were transfected, together with the pTATT_{B_{CLF1}}-Luc reporter, with expression vectors for the six different viral proteins combined (all) or with only five expression vectors as indicated (“all-” indicates the missing expression vector). (D) Western blot corresponding to HEK293 cells transfected with expression vectors for all 6 proteins or with different combinations of five of the proteins. All the proteins were tagged with a Flag epitope, and the blot was incubated with an anti-Flag MAb. Black asterisks label the positions of proteins destabilized in the absence of one component (the monoclonal anti-Flag antibody does not react with Flag-BDLF3.5). (E) The pBMRF1-Luc reporter was transfected into the EBV-negative cell line HEK293, together with combined expression vectors for the six different proteins (all) or an EB1 expression vector. (A, C, and E) The amount of luciferase expressed was measured 24 h after transfection. The relative luciferase fold activation was calculated relative to the value obtained with the pTATT_{B_{CLF1}}-Luc or the pBMRF1-Luc reporter transfected alone. The experiment was performed 3 times, and the error bars represent standard deviations. (F) BGLF3 and BVLF1 are stabilized by coexpression of BFRF2 and BDLF3.5, respectively. HEK293 cells were transfected with a Flag.BGLF3 expression vector either alone or in combination with a Flag.BFRF2 expression vector or with a Flag-BVLF1 expression vector either alone or in combination with a Flag-BDLF3.5 expression vector. The immunoblot was revealed with a polyclonal anti-Flag antibody. (G) Total RNA from the transfected cells was extracted, and RT-PCR was performed in order to show that expression of BFRF2 and BDLF3.5 did not increase the amount of BGLF3 and BVLF1 mRNA, respectively. Primers specific for the cellular GusB mRNA were used to control for the absence of contaminating DNA in the assay and to ensure that the same amount of mRNA was used for the RT.

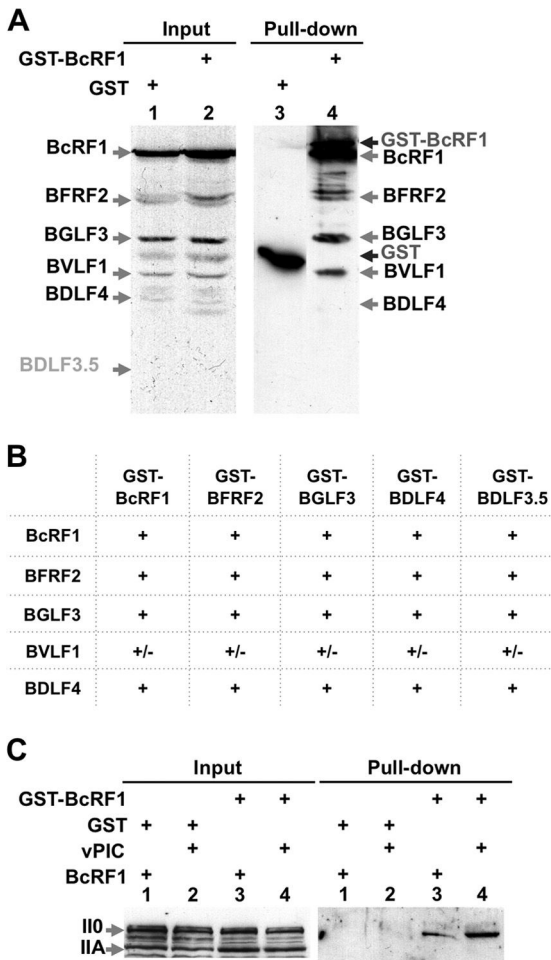


FIG 5 GST pull-down of the vPIC. Extracts from HEK293 cells transfected with a GST-Flag (GST) or a GST-Flag.BcRF1 (GST-BcRF1) expression plasmid together with expression vectors for components of the vPIC (all tagged with the Flag epitope) were incubated with glutathione-Sepharose beads. The purified proteins were analyzed by immunoblotting with a mix of monoclonal and polyclonal anti-Flag antibodies in order to identify the components of the vPIC associated with GST-Flag.BcRF1 (A) or with an anti-RNAP II antibody in order to analyze the interaction between BcRF1 and the RNAP II complex (C). GST-Flag and GST-Flag.BcRF1, which are very weakly expressed in the cells, are seen only after GST pull-down. Dark-gray arrows indicate their positions. Again, Flag.BDLF3.5 was not efficiently detected. II0 indicates the position of the hyperphosphorylated form of the RNAP II enzyme and IIA its unphosphorylated form. (B) Summary of the interactions found between the different components of the vPIC as determined by *in vitro* GST pull-down assay. *In vitro*-translated ³⁵S-labeled BcRF1, BGLF3, BVLF1, BDLF4, and BFRF2 proteins were incubated with purified GST, GST-BcRF1, GST-BGLF3, GST-BDLF4, GST-BFRF2, or GST-BDLF3.5 bound to glutathione-Sepharose beads. The bound proteins were analyzed by SDS-PAGE and visualized by autoradiography. None of the *in vitro*-translated ³⁵S-labeled proteins interacted with GST.

HEK293 cells (Fig. 5A), suggests that BVLF1 has concomitant interactions with several vPIC components.

In order to analyze the interaction between the vPIC and the RNA polymerase II machinery, proteins eluted from a GST-Flag-BcRF1 pull-down assay, performed with transfected HEK293 cells lysates, were analyzed by Western blotting using a specific anti-RNAP II antibody. In these experiments (Fig. 5C), we found that the cellular RNAP II large subunit was specifically pulled down

with the GST-Flag.BcRF1 protein (lane 3) but not with GST (lane 1). Interestingly, the interaction appeared to be more important when the full vPIC was expressed in the cells (lane 4). Moreover, the hyperphosphorylated II0 form of RNAP II (the active form) was the only RNAP II form found associated with GST-Flag.BcRF1.

Taken together, these results demonstrate that the vPIC components are involved in a complex network of mutual interactions and, most importantly, that this viral complex associates with the active form of RNAP II.

Viral DNA replication is required for expression of the endogenous late viral genes. As reported above, in transient-transfection assays, the vPIC can induce the expression of a reporter gene under the control of a late viral promoter. However, in the context of the whole EBV genome, it is well known that viral DNA replication is a prerequisite for the expression of the late viral genes. Accordingly, when HEK293_{EBV} or Raji cells were transfected with the vPIC alone, the late BDLF1 and BFRF3 viral mRNAs were not expressed (Fig. 6A) despite the vPIC being functional, as could be demonstrated by cotransfection with the pTATT_{BcRF1}-Luc reporter (Fig. 6B). As expected, expression of the early BMRF1 gene was also not induced by the vPIC. In contrast, when HEK293_{EBV} cells were transfected with an EB1 expression vector in order to induce the viral productive cycle, both the early BMRF1 gene and the late BDLF1 and BFRF3 genes were expressed. However, in the case of Raji cells, which are unable to replicate the viral genome, only the early BMRF1 gene was efficiently expressed. The late BDLF1 and BFRF3 genes were not expressed, although components of the vPIC are produced as part of the early genes. From these results, we conclude that vPIC expression is not sufficient for the induction of the late viral genes from the EBV genome. Another event, probably replication of the viral DNA, is required for the vPIC to be fully functional on the endogenous late genes.

This was confirmed by the finding that treatment of HEK293_{EBV} cells in which the viral productive cycle had been induced with phosphonoacetic acid (PAA), an inhibitor of EBV DNA replication, completely inhibited the expression of the late viral genes even when the vPIC was ectopically expressed (Fig. 7B and C). However, it had only a limited effect (50% decrease) on the level of luciferase expressed from the pTATT_{BcRF1}-Luc reporter (Fig. 7A). This effect of PAA on the amount of luciferase expressed from the pTATT_{BcRF1}-Luc reporter could be the result of either a toxic effect of the drug on the cells or a decrease in the expression of the endogenous vPIC components due to inhibition of viral DNA replication. In effect, it has been shown, in the CMV model, that expression of some of the early genes whose products are involved in late gene expression are enhanced following viral DNA replication (31). Taken together, these data show that although the vPIC is sufficient for late gene expression in transient assays, viral DNA replication is necessary for late gene expression from the endogenous genome, probably because a change in the structure of the viral chromatin is required before late gene expression is activated.

Analogous proteins from HCMV behave like those from EBV. Since an analogy exists between the EBV components of the vPIC and other beta- and gammaherpesviruses proteins putatively involved in a vPIC, especially HCMV, we tested the effect of a reconstituted HCMV vPIC on the pTATT_{BcRF1}-Luc reporter. The result of the luciferase assay obtained from HEK293 cells

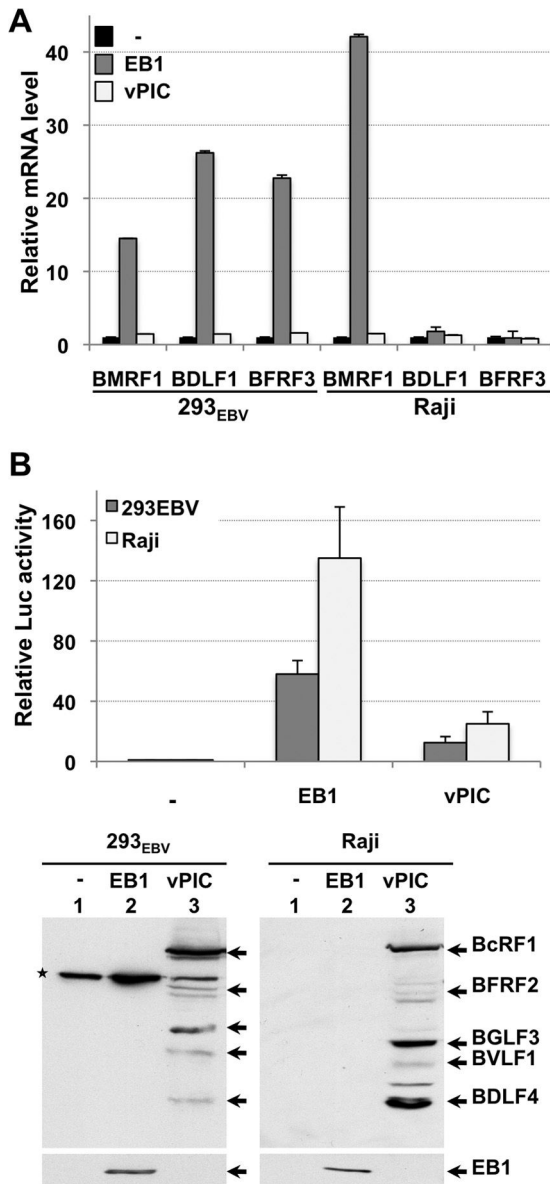


FIG 6 The vPIC was unable to induce late gene expression from the endogenous viral genome. (A) Total RNA from HEK293_{EBV} or Raji cells either mock transfected (-) or transfected with the EB1 expression vector or the six expression plasmids encoding the vPIC factors was quantified by RT-qPCR using primers specific for either the BMRF1 early gene or the BDLF1 and BFRF3 late genes. The experiment was performed 3 times, and the error bars represent standard deviations. (B) EB1 and vPIC induce late gene expression in a transient-transfection assay of EBV-positive cells. HEK293_{EBV} or Raji cells were transfected with the pTATT_{BcRF1}-Luc reporter plasmid either alone (-) or together with the EB1 expression vector or the six expression plasmids encoding the vPIC components. The experiment was performed 3 times, and the error bars represent standard deviations. Protein extracts from the experiment whose results are shown in panel B were tested for expression of EB1, and the vPIC components were tested by immunoblotting using an anti-EB1 MAb or a mix of monoclonal and polyclonal anti-Flag antibodies (for the vPIC proteins). A star indicates an unspecified band recognized by the anti-Flag antibody.

cotransfected with expression vectors for the six HCMV proteins, UL79, UL87, UL91, UL92, UL95, and UL49, was very similar to that of HEK293 cells transfected with the six EBV vPIC protein expression vectors (Fig. 8A). However, none of the HCMV pro-

teins involved in the vPIC could be efficiently replaced with their EBV equivalent (Fig. 8A). The reverse was also true (data not shown).

The functional complementation between the EBV and the HCMV vPICs was also demonstrated by transfecting HEK293_{EBVΔBFRF2} cells with the complete set of EBV or HCMV vPIC expression vectors and analyzing the expression of the late EBV genes following virus reactivation. As already shown, after virus was reactivated in these cells, the late viral genes BDLF1 and BFRF3 were not expressed, but the early gene BMRF1 was (Fig. 8B, lane 4). Moreover, without reactivation of the viral productive cycle, neither the EBV nor the HCMV vPIC was able to activate expression of the early or late viral genes (Fig. 8B, lanes 2 and 3). As expected, the EBV vPIC efficiently complemented the cells for the expression of the late genes when EBV was reactivated (Fig. 8B, lane 5), and interestingly, the HCMV vPIC gave the same result (Fig. 8B, lane 6).

Taken together, these observations demonstrate that, although each individual EBV vPIC protein cannot be functionally replaced by its HCMV counterpart, the complete HCMV vPIC is able to *trans*-complement the defect in EBV late gene expression induced by the absence of the EBV vPIC component BFRF2. These results strongly suggest that, as in EBV, the analogous HCMV vPIC proteins form a functional complex that plays a crucial role in the activation of late viral gene expression.

DISCUSSION

Expression of the late genes is an essential step for herpesviruses to complete their productive cycle. Although the initiation of viral late gene expression by viral and/or host factors has remained largely unexplored, previous works have demonstrated that several factors from beta- and gammaherpesviruses, in addition to viral DNA replication, are necessary for the expression of late viral genes. In this study, we have determined the role of the EBV protein BFRF2 during virus infection. We found that when expression of BFRF2 was abolished in cells infected by EBV, the accumulation of early gene products was not affected during the viral productive cycle and viral DNA synthesis proceeded normally. However, viral production was blocked by failure to accumulate late viral transcripts. BFRF2, by supporting the efficient transcription of late viral genes, is absolutely crucial for this accumulation. It appears that the phenotype of the EBV_{ΔBFRF2} mutant is reminiscent of that of the EBV_{ΔBcRF1} mutant (18), and like BcRF1, BFRF2 has analogs in beta- and gammaherpesviruses. These observations place the BFRF2 protein together with other EBV, CMV, and MHV-68 genes that have been characterized as being essential for late gene expression (18, 26–33).

The BFRF2 protein product consists of two domains with a variable N-terminal part and a C terminus conserved between beta- and gammaherpesviruses. The latter domain contains several conserved cysteine residues, which potentially form a zinc finger protein. Although the detailed activity of BFRF2 needs to be further explored, there are several indications that this protein may be directly involved in binding to DNA.

Studies on beta- and gammaherpesviruses, such as CMV, MHV-68 and EBV, have provided direct insights into the function of a distinct set of conserved genes that are not present in alpha-herpesviruses. These genes are called the beta-gamma genes; ORFs 18, 23, 24, 30, 31, 34, and 66 from MHV-68 are analogs of HCMV UL79, UL88, UL87, UL91, UL92, UL95, and UL49 and EBV BVLf1, BcRF1, BTRF1, BDLF3.5, BDLF4, BGLF3, and BFRF2,

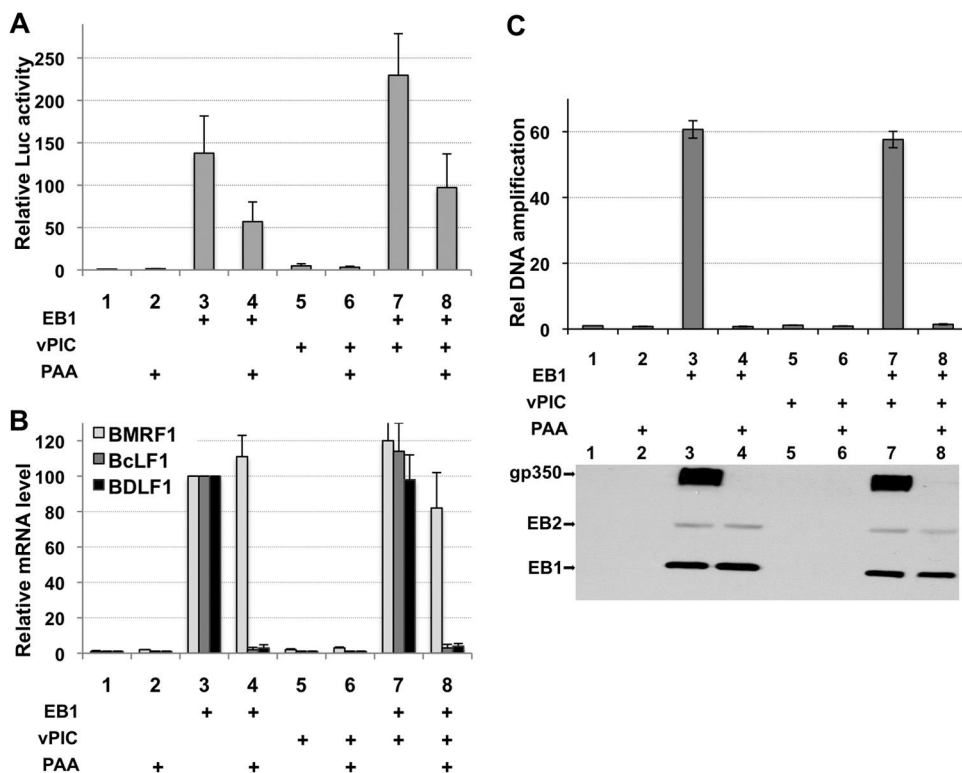


FIG 7 Activation of endogenous late gene expression is dependent on viral DNA replication. HEK293_{EBV} cells were transfected with the pTATT_{BcLF1}-Luc reporter plasmid together with an EB1 expression vector or the six expression plasmids encoding the vPIC components in the presence or absence of PAA, a specific inhibitor of viral DNA replication. (A) Both EB1 and the vPIC activate luciferase expression from the pTATT_{BcLF1}-Luc reporter, and this expression is not highly sensitive to PAA treatment. (B) EB1 is unable to activate expression of the endogenous viral late genes in the presence of PAA (50 ng/ml), even when vPIC components are expressed. Following transfection of expression plasmids encoding EB1 or the vPIC components, as indicated in the figure, RNA expression of the endogenous early BMRF1 gene and the late BcLF1 or BDLF1 gene was quantified by RT-qPCR. (C) Control showing complete inhibition of viral DNA replication by PAA (upper panel) and expression of the early BMLF1 and late gp350 proteins analyzed by immunoblotting (bottom panel) in cells used in the experiment whose results are shown in panels A and B of the figure. The experiment was performed 2 times, and the error bars represent standard deviations.

respectively. All these genes, except for MHV-68 ORF23 (34) and, by analogy, EBV BTRF1 and CMV UL88, appear to be involved in the regulation of late gene expression. Furthermore, we have now demonstrated that the six viral factors interact with one another and that a functional complex comprising these six proteins can be ectopically formed in uninfected cells. This suggests that during EBV infection, a complex containing these viral components forms to promote late gene expression. Moreover, our study provides additional evidence that the regulatory mechanism of viral late gene expression is conserved between EBV and HCMV. The analogous HCMV proteins have the same role in the activation of a late promoter reporter construct. However, it was not possible to substitute any HCMV protein for its EBV analog and vice versa, suggesting that the homology between each protein is not sufficient to allow *trans*-complementation in the complex. Importantly, however, the complete HCMV complex is able to *trans*-complement the defect in late gene expression of EBV_{ΔBFRF2}. Taken together, our results suggest that beta- and gammaherpesviruses encode their own transcription complex responsible for the expression of late viral genes.

The existence of a conserved vPIC between beta- and gammaherpesviruses raises the question of how the alphaherpesviruses regulate their late gene expression without using analogs of the vPIC despite their late promoters being organized in a manner very similar to those of the beta- and gammaherpesviruses. It is

interesting to note that for alphaherpesviruses, the viral immediate early proteins ICP4, ICP0, and ICP27 are involved in the expression of late viral genes. In particular, ICP4, which has been shown to be important for the recruitment of the cellular TBP to late gene promoters (36), contains a unique uracil-DNA glycosylase-like domain potentially involved in this process (37). Until now, a direct implication of the CMV or EBV immediate early factors in the regulation of late viral genes has not been described, except for particular genes like that encoding the EBV BLLF1 protein (gp350), whose expression is induced by Rta (6, 38). It is thus tempting to speculate that the EBV BcRF1 protein and its analogs in beta- and gammaherpesviruses play the role of cellular TBP (18). The exact recognition mechanism of minimal promoters in CMV is not known. As previously suggested (17), the CMV complex should not be able to perform a proper molecular recognition of altered TATA box motifs, in coherence with the evolutionary history of CMV promoters, in which no altered consensus TATA box is observed. We are tempted to speculate that the vPIC in CMV is in fact essential for late transcription but needs to be complemented with either cellular or viral factors in order to recognize proper promoters. Additionally, since a differential domain composition of the vPIC elements is observed in beta- and gammaherpesviruses, there is the possibility that these variants contain additional DNA-binding factors responsible for the process of recognition. Based on the conservation of the vPICs be-

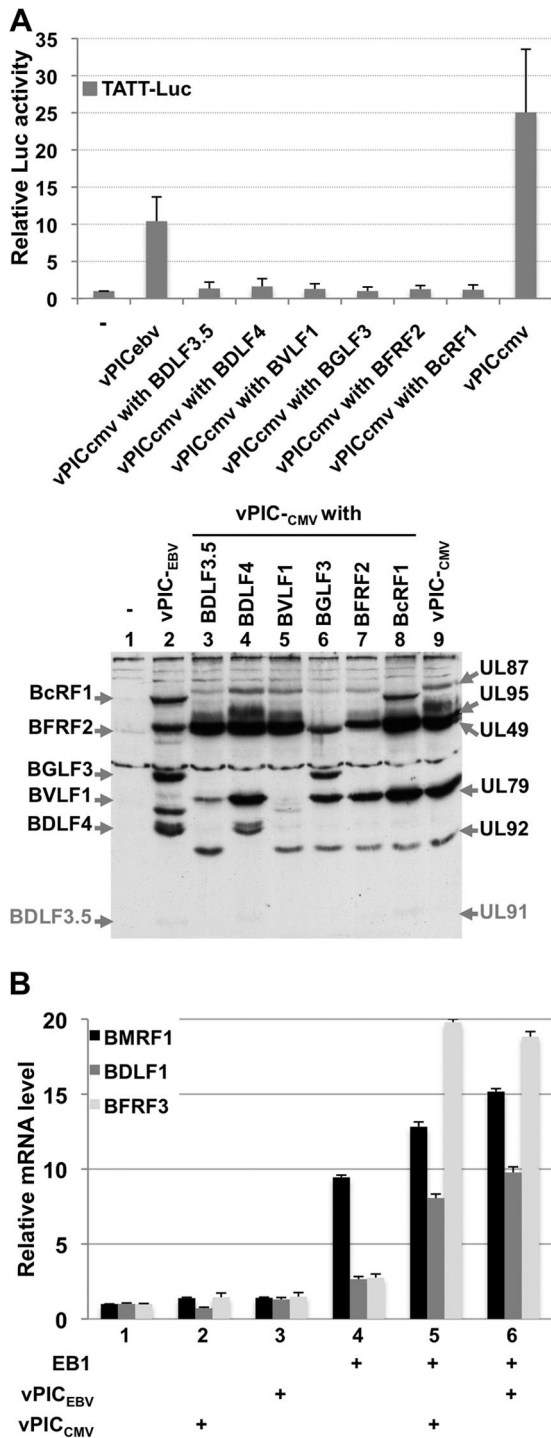


FIG 8 Betaherpesviruses also encode a vPIC. (A) HEK293 cells were transfected with the pTATT_{BcLF1}-Luc reporter plasmid together with the expression vectors encoding either the vPIC EBV components (BcRF1, BFRF2, BGLF3, BVLF1, BDLF4, and BDLF3.5), the vPIC CMV components (UL87, UL49, UL95, UL79, UL92, and UL91), or the vPIC CMV components with one component replaced by the EBV analog, as indicated. The relative luciferase fold activation was calculated relative to the value obtained with the pTATT_{BcLF1}-Luc reporter transfected alone (-). The experiment was performed 3 times, and the error bars represent standard deviations. Protein expression from transfected cells in panel A was controlled by immunoblotting using a mix of monoclonal and polyclonal anti-Flag antibodies. Neither BDLF3.5 nor UL91 was detected. (B) The CMV vPIC can substitute for the EBV vPIC.

tween beta- and gammaherpesviruses and the differences with alphaherpesviruses, we can speculate that beta- and gammaherpesviruses have diverged more recently.

A reporter luciferase plasmid driven by a minimal late promoter containing a TATT box was efficiently activated by transient expression of six EBV or HCMV proteins involved in the vPIC in both uninfected and EBV-infected cells not induced into the productive cycle. However, when the productive cycle was induced by transient expression of the viral immediate early transcription factor EB1, activation of the pTATT_{BcLF1}-Luc reporter was much more important. This difference in the activation level may be the result of the technical difficulty of simultaneously introducing seven expression plasmids in cells by transfection. Thus, the participation of the endogenous vPIC components expressed following induction of the productive cycle by EB1 might explain this difference in activation of the pTATT_{BcLF1}-Luc reporter. Alternatively, additional accessory viral proteins might be required for the complete activation of the late viral genes. This possibility is reminiscent of what is known for the viral DNA replication complex. Herpesviruses encode a minimum of seven proteins necessary for DNA replication, BALF2, BALF5, BMRF1, BBLF2/BBLF3, BSLF1, BZLF1, and BBLF4 in the case of EBV (7, 39), but additional accessory proteins are also necessary to obtain a more efficient replication complex.

Activation of the BcLF1 and BFRF3 late gene promoters was observed in transiently transfected EBV-infected cells even when the reporter plasmids were devoid of the OriLyt sequence. However, this reconstitution was dependent on EBV DNA replication occurring in the same cells (10), suggesting either that *trans*-acting factors produced after OriLyt-dependent DNA replication are required for late gene expression or that DNA replication increases the copy number of the viral episome, allowing titration of a putative repressor. In contrast, in a reconstituted stable reporter system, the same late viral promoters showed a specific dependence on OriLyt DNA replication in *cis* (11), suggesting that during or after DNA replication, the structure of the viral chromatin changes, allowing an increased accessibility to the late promoters for the transcription machinery. If late viral gene transcription is clearly tightly coupled to viral DNA replication, the underlying mechanism is still poorly understood. In accordance with these previous data, we show here that although the vPIC can induce transcriptional activation of the viral late genes in a transient assay, it is completely inefficient at inducing expression of the endogenous late genes from the viral genome in the absence of viral DNA replication. These results may suggest that the vPIC controls late viral gene transcription in association with the viral DNA replication complex, although there is currently no direct evidence that any of the beta-gamma gene products interact with proteins from the viral DNA replication complex. However, Sugimoto et al. (40) have reported that the EBV BcRF1 protein localizes inside the replication foci (BMRF1 cores) in late stages of lytic replication, suggesting that late gene expression is initiated from the newly replicated DNA in the replication foci.

HEK293_{EBVΔBFRF2} cells were transfected with the different expression plasmids indicated at the bottom of the graph. Total RNA extracted from the transfected cells was quantified by RT-qPCR using primers specific for either the BMRF1 early gene or the BDLF1 and BFRF3 late genes. The experiment was performed 3 times, and the error bars represent standard deviations.

How the vPIC functions and which additional protein components are present in this complex remain important questions. We can speculate, as previously suggested (32), that the vPIC may activate late gene transcription by remodeling the chromatin structure of the viral genome. Genomes of *Herpesviridae* associate with histones during infection and require epigenetic regulation/modification for gene expression (41–43). However, in contrast to the original episomic viral genome, the newly synthesized DNA is probably not covered by nucleosomes (44). This point may explain why, in transient transfection, it is possible to activate late viral gene transcription, whereas this is not possible with the endogenous viral genome. Alternatively, and more likely, the vPIC may play a more direct role in linking viral DNA replication and transcription. We have previously shown that BcRF1, the homolog of CMV UL87 and MHV-68 ORF24, acts as a TATA-binding protein, preferentially interacting with TATT sequences, to regulate late gene expression during infection (17, 18). Furthermore, it has been shown that MHV-68 ORF30 and ORF34, homologs of CMV UL91 and UL95 and EBV BDLF3.5 and BGLF3, respectively, regulate the binding of RNAP II to the late gene promoters (27). All this evidence points toward a direct contribution by the beta-gamma late gene regulators in the assembly of a late gene RNAP II transcription complex, a multicomponent enzyme that requires a host accessory scaffold and regulatory proteins for its activity. The vPIC may play an essential role in the recruitment of these components to or their assembly on viral late gene promoters as cellular TBP does for the recruitment of TFIID and the cellular preinitiation complex. In accordance with this, we have shown that BcRF1 is associated with the hyperphosphorylated form of RNAP II. Further studies on the formation of this viral complex as well as its regulation and recruitment to the TATT element will be an important basis for dissection of the mechanism of beta- and gammaherpesvirus late gene regulation. How the vPIC interacts with viral and/or host factors and its relationship to core components of the RNAP II transcriptional machinery as well as viral DNA synthesis will be the subject of future investigations.

This study advances our understanding of the mechanisms of EBV late gene expression, an important question that has long remained unresolved. Furthermore, this very specific vPIC, present in beta- and gammaherpesviruses and required for productive infection, is potentially a very attractive target for antiviral strategies to combat HCMV, KSHV, and EBV infection as well as their reactivation and associated diseases.

ACKNOWLEDGMENTS

We thank W. Hammerschmidt for providing the HEK293_{EBV} cells. We thank R. Buckland for reading the manuscript.

This work was supported by the Institut National de la Santé et de la Recherche Médicale (INSERM) and the Association pour la Recherche sur le Cancer (grant 3420). We acknowledge the contribution of AniRA Genetic Analysis, the AniRA flow cytometry platforms, and the PLATIM microscopy facilities of SFR Biosciences Gerland-Lyon Sud, France (US8/UMS3444). V.A. and T.D. are recipients of a fellowship from the Ministère de la Recherche et de la Technologie (MRT).

REFERENCES

- Rickinson AB, Kieff E. 2001. Epstein-Barr virus, p 2575–2627. *In* Knipe DM, Howley PM, Griffin DE, Martin MA, Lamb RA, Roizman B, Straus SE (ed), *Fields virology*, 4th ed. Lippincott Williams & Wilkins, Philadelphia, PA.
- Babcock GJ, Decker LL, Volk M, Thorley-Lawson DA. 1998. EBV persistence in memory B cells in vivo. *Immunity* 9:395–404. [http://dx.doi.org/10.1016/S1074-7613\(00\)80622-6](http://dx.doi.org/10.1016/S1074-7613(00)80622-6).
- Laichalk LL, Thorley-Lawson DA. 2005. Terminal differentiation into plasma cells initiates the replicative cycle of Epstein-Barr virus in vivo. *J. Virol.* 79:1296–1307. <http://dx.doi.org/10.1128/JVI.79.2.1296-1307.2005>.
- Chevallier-Greco A, Manet E, Chavrier P, Mosnier C, Daillie J, Sergeant A. 1986. Both Epstein-Barr virus (EBV) encoded trans-acting factors, EB1 and EB2, are required to activate transcription from an EBV early promoter. *EMBO J.* 5:3243–3249.
- Hardwick JM, Lieberman PM, Hayward SD. 1988. A new Epstein-Barr virus transactivator, R, induces expression of a cytoplasmic early antigen. *J. Virol.* 62:2274–2284.
- Feederle R, Kost M, Baumann M, Janz A, Drouet E, Hammerschmidt W, Delecluse HJ. 2000. The Epstein-Barr virus lytic program is controlled by the co-operative functions of two transactivators. *EMBO J.* 19:3080–3089. <http://dx.doi.org/10.1093/emboj/19.12.3080>.
- Fixman ED, Hayward GS, Hayward SD. 1992. Trans-acting requirements for replication of Epstein-Barr virus ori-Lyt. *J. Virol.* 66:5030–5039.
- Hammerschmidt W, Sugden B. 1988. Identification and characterization of ORLyt, a lytic origin of DNA replication of Epstein-Barr virus. *Cell* 55:427–433. [http://dx.doi.org/10.1016/0092-8674\(88\)90028-1](http://dx.doi.org/10.1016/0092-8674(88)90028-1).
- Israel BF, Kenney SC. 2005. EBV lytic infection, p 571–611. *In* Robertson ES (ed), *Epstein-Barr virus*. Caister Academic Press, Norfolk, United Kingdom.
- Serio TR, Kolman JL, Miller G. 1997. Late gene expression from the Epstein-Barr virus BcLF1 and BFRF3 promoters does not require DNA replication in cis. *J. Virol.* 71:8726–8734.
- Amon W, Binne UK, Bryant H, Jenkins PJ, Karstegl CE, Farrell PJ. 2004. Lytic cycle gene regulation of Epstein-Barr virus. *J. Virol.* 78:13460–13469. <http://dx.doi.org/10.1128/JVI.78.24.13460-13469.2004>.
- Chang PJ, Chang YS, Liu ST. 1998. Characterization of the BcLF1 promoter in Epstein-Barr virus. *J. Gen. Virol.* 79:2003–2006.
- Homa FL, Otal TM, Glorioso JC, Levine M. 1986. Transcriptional control signals of a herpes simplex virus type 1 late (gamma 2) gene lie within bases –34 to +124 relative to the 5' terminus of the mRNA. *Mol. Cell. Biol.* 6:3652–3666.
- Shapira M, Homa FL, Glorioso JC, Levine M. 1987. Regulation of the herpes simplex virus type 1 late (gamma 2) glycoprotein C gene: sequences between base pairs –34 to +29 control transient expression and responsiveness to transactivation by the products of the immediate early (alpha) 4 and 0 genes. *Nucleic Acids Res.* 15:3097–3111. <http://dx.doi.org/10.1093/nar/15.7.3097>.
- Wong-Ho E, Wu T-T, Davis ZH, Zhang B, Huang J, Gong H, Deng H, Liu F, Glaunsinger B, Sun R. 2014. Unconventional sequence requirement for viral late gene core promoters of murine gammaherpesvirus 68. *J. Virol.* 88:3411–3422. <http://dx.doi.org/10.1128/JVI.01374-13>.
- Serio TR, Cahill N, Prout ME, Miller G. 1998. A functionally distinct TATA box required for late progression through the Epstein-Barr virus life cycle. *J. Virol.* 72:8338–8343.
- Wyrwicz LS, Rychlewski L. 2007. Identification of herpes TATT-binding protein. *Antiviral Res.* 75:167–172. <http://dx.doi.org/10.1016/j.antiviral.2007.03.002>.
- Gruffat H, Kadjouf F, Mariamé B, Manet E. 2012. The Epstein-Barr virus BcRF1 gene product is a TBP-like protein with an essential role in late gene expression. *J. Virol.* 86:6023–6032. <http://dx.doi.org/10.1128/JVI.00159-12>.
- Isomura H, Stinski MF, Murata T, Yamashita Y, Kanda T, Toyokuni S, Tsurumi T. 2011. The human cytomegalovirus gene products essential for late viral gene expression assemble into pre-replication complexes before viral DNA replication. *J. Virol.* 85:6629–6644. <http://dx.doi.org/10.1128/JVI.00384-11>.
- 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.* 3:141–146. <http://dx.doi.org/10.1093/embo-reports/kvf031>.
- Delecluse HJ, Hilsendegen T, Pich D, Zeidler R, Hammerschmidt W. 1998. Propagation and recovery of intact, infectious Epstein-Barr virus from prokaryotic to human cells. *Proc. Natl. Acad. Sci. U. S. A.* 95:8245–8250. <http://dx.doi.org/10.1073/pnas.95.14.8245>.
- Feederle R, Bartlett EJ, Delecluse HJ. 2010. Epstein-Barr virus genetics: talking about the BAC generation. *Herpesviridae* 1:6. <http://dx.doi.org/10.1186/2042-4280-1-6>.

23. Hirt B. 1969. Replicating molecules of polyoma virus DNA. *J. Mol. Biol.* 40:141–144. [http://dx.doi.org/10.1016/0022-2836\(69\)90302-7](http://dx.doi.org/10.1016/0022-2836(69)90302-7).
24. Batisse J, Manet E, Middeldorp J, Sergeant A, Gruffat H. 2005. Epstein-Barr virus mRNA export factor EB2 is essential for intranuclear capsid assembly and production of gp350. *J. Virol.* 79:14102–14111. <http://dx.doi.org/10.1128/JVI.79.22.14102-14111.2005>.
25. Juillard F, Bazot Q, Mure F, Tafforeau L, Macri C, Rabourdin-Combe C, Lotteau V, Manet E, Gruffat H. 2012. Epstein-Barr virus protein EB2 stimulates cytoplasmic mRNA accumulation by counteracting the deleterious effects of SRp20 on viral mRNAs. *Nucleic Acids Res.* 40:6834–6849. <http://dx.doi.org/10.1093/nar/gks319>.
26. Arumugaswami V, Wu TT, Martinez-Guzman D, Jia Q, Deng H, Reyes N, Sun R. 2006. ORF18 is a transfactor that is essential for late gene transcription of a gammaherpesvirus. *J. Virol.* 80:9730–9740. <http://dx.doi.org/10.1128/JVI.00246-06>.
27. Wu T-T, Park T, Kim H, Tran T, Tong L, Martinez-Guzman D, Reyes N, Deng H, Sun R. 2009. ORF30 and ORF34 are essential for expression of late genes in murine gammaherpesvirus 68. *J. Virol.* 83:2265–2273. <http://dx.doi.org/10.1128/JVI.01785-08>.
28. Wong E, Wu TT, Reyes N, Deng H, Sun R. 2007. Murine gammaherpesvirus 68 open reading frame 24 is required for late gene expression after DNA replication. *J. Virol.* 81:6761–6764. <http://dx.doi.org/10.1128/JVI.02726-06>.
29. Jia Q, Wu TT, Liao HI, Chernishof V, Sun R. 2004. Murine gammaherpesvirus 68 open reading frame 31 is required for viral replication. *J. Virol.* 78:6610–6620. <http://dx.doi.org/10.1128/JVI.78.12.6610-6620.2004>.
30. Omoto S, Mocarski ES. 2013. Cytomegalovirus UL91 is essential for transcription of viral true late (γ_2) genes. *J. Virol.* 87:8651–8664. <http://dx.doi.org/10.1128/JVI.01052-13>.
31. Omoto S, Mocarski ES. 2014. Transcription of true late (γ_2) cytomegalovirus genes requires UL92 function that is conserved among beta- and gammaherpesviruses. *J. Virol.* 88:120–130. <http://dx.doi.org/10.1128/JVI.02983-13>.
32. Chapa TJ, Perng Y-C, French AR, Yu D. 2014. Murine cytomegalovirus protein pM92 is a conserved regulator of viral late gene expression. *J. Virol.* 88:131–142. <http://dx.doi.org/10.1128/JVI.02684-13>.
33. Perng YC, Qian Z, Fehr AR, Xuan B, Yu D. 2011. The human cytomegalovirus gene UL79 is required for the accumulation of late viral transcripts. *J. Virol.* 85:4841–4852. <http://dx.doi.org/10.1128/JVI.02344-10>.
34. Ohno S, Steer B, Sattler C, Adler H. 2012. ORF23 of murine gammaherpesvirus 68 is non-essential for in vitro and in vivo infection. *J. Gen. Virol.* 93:1076–1080. <http://dx.doi.org/10.1099/vir.0.041129-0>.
35. Gruffat H, Batisse J, Pich D, Neuhierl B, Manet E, Hammerschmidt W, Sergeant A. 2002. Epstein-Barr virus mRNA export factor EB2 is essential for production of infectious virus. *J. Virol.* 76:9635–9644. <http://dx.doi.org/10.1128/JVI.76.19.9635-9644.2002>.
36. Grondin B, DeLuca N. 2000. Herpes simplex virus type 1 ICP4 promotes transcription preinitiation complex formation by enhancing the binding of TFIID to DNA. *J. Virol.* 74:11504–11510. <http://dx.doi.org/10.1128/JVI.74.24.11504-11510.2000>.
37. Wyrwicz LS, Rychlewski L. 2007. Fold recognition insights into function of herpes ICP4 protein. *Acta Biochim. Pol.* 54:551–559.
38. Ramasubramanian S, Kanhere A, Osborn K, Flower K, Jenner RG, Sinclair AJ. 2012. Genome-wide analyses of Zta binding to the Epstein-Barr virus genome reveals interactions in both early and late lytic cycle and an epigenetic switch leading to an altered binding profile. *J. Virol.* 86:12494–12502. <http://dx.doi.org/10.1128/JVI.01705-12>.
39. Fixman ED, Hayward GS, Hayward SD. 1995. Replication of Epstein-Barr virus oriLyt: lack of a dedicated virally encoded origin-binding protein and dependence on Zta in cotransfection assays. *J. Virol.* 69:2998–3006.
40. Sugimoto A, Sato Y, Kanda T, Murata T, Narita Y, Kawashima D, Kimura H, Tsurumi T. 2013. Different distributions of Epstein-Barr virus early and late gene transcripts within viral replication compartments. *J. Virol.* 87:6693–6699. <http://dx.doi.org/10.1128/JVI.00219-13>.
41. Cuevas-Bennett C, Shenk T. 2008. Dynamic histone H3 acetylation and methylation at human cytomegalovirus promoters during replication in fibroblasts. *J. Virol.* 82:9525–9536. <http://dx.doi.org/10.1128/JVI.00946-08>.
42. Chen H-S, Lu F, Lieberman PM. 2013. Epigenetic regulation of EBV and KSHV latency. *Curr. Opin. Virol.* 3:251–259. <http://dx.doi.org/10.1016/j.coviro.2013.03.004>.
43. Lieberman PM. 2013. Keeping it quiet: chromatin control of gammaherpesvirus latency. *Nat. Rev. Microbiol.* 11:863–875. <http://dx.doi.org/10.1038/nrmicro3135>.
44. Chiu Y-F, Sugden AU, Sugden B. 2013. Epstein-Barr viral productive amplification reprograms nuclear architecture, DNA replication, and histone deposition. *Cell Host Microbe* 14:607–618. <http://dx.doi.org/10.1016/j.chom.2013.11.009>.
45. Mocarski ES, Jr. 2007. Betaherpes viral genes and their functions, p 204–230. *In* Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, Yamanishi K (ed), *Human herpesviruses: biology, therapy, and immunoprophylaxis*. Cambridge University Press, Cambridge, United Kingdom.

À la suite de cette publication, nous avons décidé de continuer à caractériser le rôle des autres protéines virales potentiellement impliquées dans l'expression des gènes viraux tardifs par analyse génétique. Effectivement, aucune étude à ce moment là, n'avait confirmé le rôle de BVLF1, BDLF4 et BDLF3.5 dans l'expression des gènes viraux tardifs.

Nous avons montré dans l'article que le vPIC est capable d'interagir avec l'ARN polymérase II cellulaire, *in vivo*. Nous avons voulu déterminer quel(s) membre(s) de ce complexe est susceptible de recruter l'ARN pol II.

Par ailleurs, nous avons réalisé les interactions des membres du vPIC uniquement *in vitro*. Nous avons décidé de réaliser les interactions par GST pull-down *in vivo*, afin de dévoiler l'interactome du vPIC.

De plus, seules six protéines virales et une protéine cellulaire ont été identifiées comme nécessaires et suffisantes à l'expression des gènes viraux tardifs. Nous avons dès lors émis l'hypothèse que d'autres facteurs pourraient être impliqués dans l'expression des gènes viraux tardifs. Afin de caractériser l'intégralité du complexe, nous avons, par double purification sur chromatographie d'affinité puis analyse par spectrométrie de masse, tenté de découvrir les potentiels facteurs cellulaires et viraux impliqués dans la synthèse des ARNm tardifs. Les résultats obtenus lors de cette expérience ont mis en lumière un éventuel lien entre le vPIC d'EBV et le complexe de réplication de l'ADN viral de ce même virus.

B) BVLF1, BDLF4 et BDLF3.5 sont nécessaires à l'expression des gènes viraux tardifs.

L'article publié en 2014 par notre équipe a démontré que la protéine BFRF2 était impliquée dans l'expression des gènes viraux tardifs sans connaître sa fonction. Par ailleurs, les orthologues des cinq autres protéines du vPIC ont tous été démontrés comme impliqués dans l'expression des gènes viraux tardifs de ces virus (HCMV et MHV68). N'ayant pas prouvé leur implication chez EBV, il était nécessaire de le confirmer. Pour ce faire, trois bacmides recombinants portant le génome entier d'EBV ont été mutés afin d'enlever une partie du gène codant pour les protéines d'intérêts, générant ainsi trois virus mutants. Ils ont ensuite été transfectés dans des cellules HEK293 puis mis sous sélection afin d'engendrer trois nouvelles lignées cellulaires nommées HEK293_{EBVΔBVLF1}, HEK293_{EBVΔBDLF4} et HEK293_{EBVΔBDLF3.5}. Dans le but de démontrer que ces trois protéines virales sont impliquées dans l'expression des gènes viraux tardifs, le cycle productif du virus a été induit par

transfection du plasmide d'expression codant pour la protéine EB1 et transcomplémenté ou non avec un vecteur d'expression codant pour la protéine enlevée dans le génome viral afin de restaurer son expression dans les lignées cellulaires correspondantes. La réplication de l'ADN viral, l'expression des ARNm viraux précoces et tardifs ont été analysées au cours du cycle dans ces différentes conditions pour les trois virus recombinants.

1) BVLf1, BDLf4 et BDLf3.5 sont nécessaires à l'expression des gènes viraux tardifs *in vitro*.

Afin de démontrer l'implication de BVLf1, BDLf4 et BDLf3.5 dans l'expression des gènes viraux tardifs, une première approche basée sur un système transitoire a été utilisée. Pour ce faire, un vecteur rapporteur pTATT_{BcLF1}-Luc dont le gène luciférase est sous contrôle du promoteur tardif du gène BcLF1 (cf. figure 46) a été transfecté dans les lignées cellulaires HEK293 portant un virus recombinant amputé d'un des gènes d'intérêt.

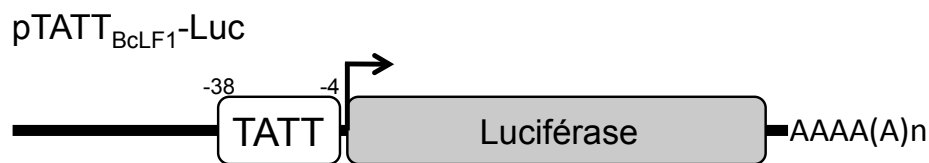


Figure 46 : Représentation schématique du rapporteur luciférase utilisé dans l'étude (pTATT_{BcLF1}-Luc).

Ce vecteur rapporteur est composé d'un gène codant pour la luciférase. L'expression de ce gène est sous contrôle du promoteur tardif de BcLF1. L'activité luciférase sera mesurée pour déterminer rapidement si les protéines d'intérêts sont impliquées dans l'expression des gènes viraux tardifs.

Ces cellules ont également été transfectées avec, soit un plasmide vide, soit un vecteur d'expression codant pour la protéine virale EB1 (capable d'induire le cycle viral productif), et transcomplémentées ou non avec un vecteur d'expression codant pour une des protéines virales manquante. Soixante heures post-transfection, l'activité luciférase a été mesurée à partir des cellules transfectées qui ont été récoltées.

Comme attendu, lorsque les cellules HEK293_{EBVΔBVLf1}, HEK293_{EBVΔBDLf4} et HEK293_{EBVΔBDLf3.5} sont en latence, l'activité luciférase n'est pas induite dans les trois lignées cellulaires (cf. figure 47). Ces différentes lignées, transfectées avec un vecteur d'expression codant pour la protéine virale EB1 (qui induit le cycle viral productif), ne sont également pas capable d'induire l'activité luciférase (cf. figure 47). En revanche, lorsque les cellules HEK293_{EBVΔBVLf1} (figure 47A), sont transfectées avec un vecteur d'expression codant pour la

protéine EB1 et transcomplémentées avec un vecteur d'expression codant pour la protéine BVLF1 alors, l'activité luciférase est induite de manière significative (cf. figure 47A). Des résultats similaires sont observés lorsque les cellules HEK293_{EBV Δ BDLF4} (figure 47B) et HEK293_{EBV Δ BDLF3.5} (figure 47C) sont transfectées dans les mêmes conditions mais transcomplémentées avec des vecteurs d'expression codants respectivement pour BDLF4 (cf. figure 47B) et BDLF3.5 (cf. figure 47C). Ainsi, ces résultats démontrent que les protéines BVLF1, BDLF4 et BDLF3.5 sont nécessaires à l'expression des gènes tardifs.

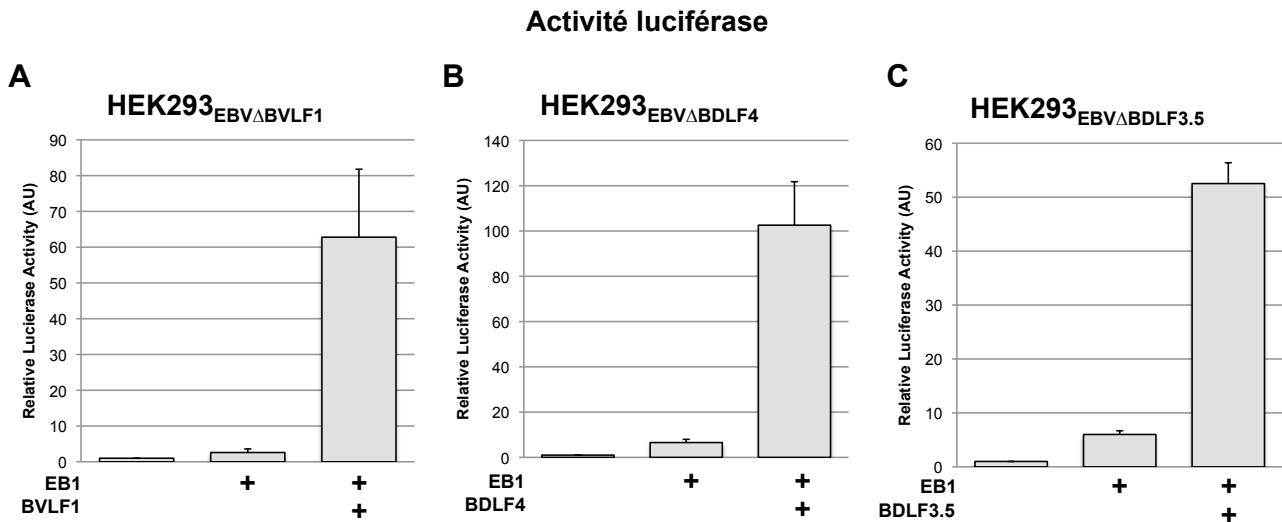


Figure 47 : BVLF1, BDLF4 et BDLF3.5 sont nécessaires *in vitro*, à l'expression des gènes viraux tardifs.

Les cellules HEK293_{EBV Δ BVLF1} (A), HEK293_{EBV Δ BDLF4} (B) et HEK293_{EBV Δ BDLF3.5} (C) ont été transfectées avec : soit un vecteur vide soit un vecteur d'expression codant pour la protéine EB1 avec ou sans les plasmides d'expression codant pour respectivement la protéine BVLF1 ou BDLF4 ou BDLF3.5. L'activité luciférase a été mesurée soixante heures post-transfection. L'activation relative de l'activité luciférase a été normalisée avec l'activité luciférase obtenue lorsque le plasmide vide est transfecté seul.

2) BVLF1, BDLF4 et BDLF3.5 ne sont pas impliquées dans l'expression des protéines virales précoces, *in vivo*.

Afin de confirmer cette première approche, des analyses *in vivo* ont été réalisées à l'aide des différentes lignées cellulaires établies : HEK293_{EBV Δ BVLF1}, HEK293_{EBV Δ BDLF4} et HEK293_{EBV Δ BDLF3.5}. Tout d'abord, ces différentes lignées ont été transfectées avec soit un plasmide vide, soit un vecteur d'expression codant pour la protéine EB1 (qui induit le cycle productif) avec ou sans un plasmide d'expression codant pour la protéine viral dont le gène a été enlevé dans le génome viral présent dans les cellules. L'expression de différentes protéines virales a ensuite été analysée soixante heures post-transfection par Western-Blot

(WB) (cf. figure 48). L'expression de la protéine immédiate précoce (IP) EB1 permettra de contrôler la transfection dans les différentes conditions.

Comme attendu, l'expression de la protéine immédiate précoce EB1 est bien détectée par Western-Blot quand elle est transfectée seule. Son expression n'est pas modifiée lorsque BVLV1 (cf. figure 48A) et BDLF4 (cf. figure 48B) sont transcomplémentées dans les cellules. En revanche, la transcomplémentation des cellules HEK293_{EBV Δ BDLF3.5} avec le vecteur d'expression codant pour la protéine BDLF3.5 est couplée à une légère augmentation de la détection de EB1 uniquement dans cette expérience (cf. figure 48C).

L'expression de la protéine virale précoce (P) EB2 a également été observée. Elle est efficacement exprimée lorsque le cycle viral productif est induit. L'expression de cette dernière n'est pas modifiée lorsque BVLV1 et BDLF4 sont transcomplémentées dans les lignées cellulaires correspondantes (cf. figures 48 A et B). Toutefois, comme avec EB1, la protéine EB2 est légèrement mieux détectée probablement car EB1 est un peu plus exprimée dans les cellules HEK293_{EBV Δ BDLF3.5}.

La protéine virale tardive (T) gp350 n'est pas exprimée lorsque le cycle viral productif est induit dans les différentes lignées cellulaires portants les virus EBV mutants. En revanche, lorsque les cellules HEK293_{EBV Δ BVLV1}, HEK293_{EBV Δ BDLF4} et HEK293_{EBV Δ BDLF3.5} sont transfectées avec un plasmide codant pour la protéine EB1 et transcomplémentées avec des vecteurs d'expression codant pour la protéine dont le gène a été enlevé du génome viral, alors la protéine tardive gp350 est exprimée dans les trois cas (cf. figure 48 A, B, C).

Expression des protéines virales

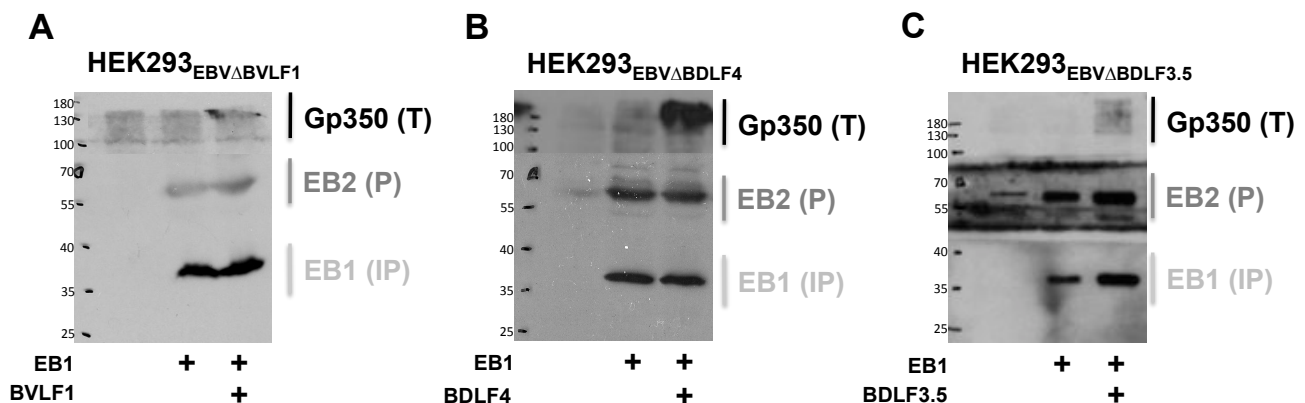


Figure 48 : Les protéines virales BVLV1, BDLF4 et BDLF3.5 sont nécessaires à l'expression de protéines virales tardives mais pas précoces.

L'expression des protéines virales immédiate précoce (IP) EB1, précoce (P) EB2 et tardive (T) gp350 a été analysée par Western-Blot, soixante heures après transfection des cellules HEK293_{EBV Δ BVLV1} (A),

HEK293_{EBV Δ BDLF4} (B) et HEK293_{EBV Δ BDLF3.5} (C).

L'expression de la protéine gp350 varie entre lignées cellulaires. En effet, elle est faiblement détectée dans les lignées HEK293_{EBV Δ BVLF1} et HEK293_{EBV Δ BDLF3.5} contrairement aux cellules HEK293_{EBV Δ BDLF4}, peut être à cause d'un problème de production ou de détection de cette protéine.

Ces résultats démontrent donc que les protéines BVLF1, BDLF4 et BDLF3.5 ne sont pas impliquées dans l'expression des protéines virales précoces mais sont nécessaires à celle des protéines tardives du virus d'Epstein-Barr.

3) BVLF1, BDLF4 et BDLF3.5 n'ont pas de rôle *in vivo*, dans la réplication de l'ADN viral.

L'expression des protéines virales tardives est dépendante de la réplication de l'ADN viral. Afin de savoir si BVLF1, BDLF4 et BDLF3.5 sont impliquées dans cette étape primordiale du cycle viral productif, l'ADN viral présent dans les différentes lignées HEK293_{EBV Δ BVLF1}, HEK293_{EBV Δ BDLF4} et HEK293_{EBV Δ BDLF3.5} a été quantifié par qPCR (cf. figure 49).

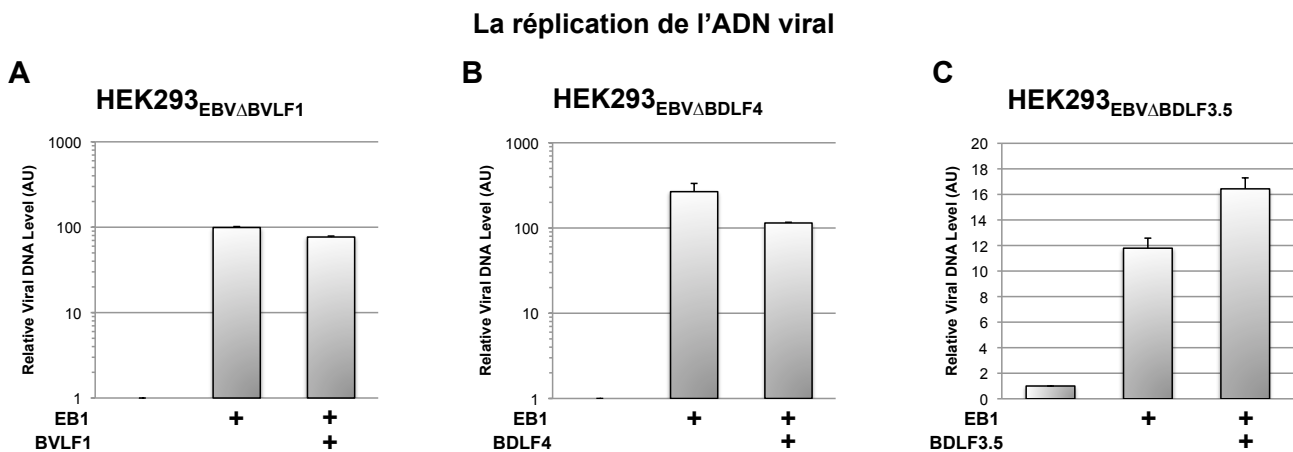


Figure 49 : Les protéines virales BVLF1, BDLF4 et BDLF3.5 ne sont pas impliquées dans la réplication de l'ADN viral.

L'ADN viral présent dans les cellules HEK293_{EBV Δ BVLF1} (A), HEK293_{EBV Δ BDLF4} (B) et HEK293_{EBV Δ BDLF3.5} (C), transfectées avec soit un plasmide vide, soit un vecteur d'expression codant pour la protéine EB1 avec ou sans un vecteur d'expression codant pour la protéine manquante, respectivement BVLF1, BDLF4 et BDLF3.5 a été quantifié par qPCR. La quantité d'ADN viral présent dans les cellules (où le cycle viral est induit) a été normalisée par rapport à l'ADN viral présent dans les cellules non induite en cycle viral productif.

Comme attendu, l'induction du cycle viral productif par EB1 mène à une forte augmentation d'ADN viral présent dans les cellules HEK293_{EBV Δ BVLF1} et HEK293_{EBV Δ BDLF4} (cf. figures 49 A et B). Toutefois, cette augmentation est beaucoup faible dans les cellules HEK293_{EBV Δ BDLF3.5} (cf. figure 49C). La co-expression de BVLF1, BDLF4 et BDLF3.5 dans les lignées cellulaires correspondantes, ne modifie pas significativement la réplication de l'ADN viral. Ces résultats sont valables pour toutes les conditions (cf. figures 49 A, B et C). Ainsi, les protéines virales BVLF1, BDLF4 et BDLF3.5 ne sont pas impliquées dans la réplication de l'ADN viral (cf. figure 49).

4) BVLF1, BDLF4 et BDLF3.5 sont nécessaires à l'expression des ARNm tardifs du virus d'Epstein-Barr.

Afin de confirmer le potentiel impact de BVLF1, BDLF4 et BDLF3.5 sur l'expression des ARNm viraux tardifs, certains ARNm viraux des différentes lignées cellulaires utilisées (EK293_{EBV Δ BVLF1}, HEK293_{EBV Δ BDLF4} et HEK293_{EBV Δ BDLF3.5}) ont été quantifiés par RT-qPCR. Les cellules HEK293_{EBV Δ BVLF1}, HEK293_{EBV Δ BDLF4} et HEK293_{EBV Δ BDLF3.5} ont été transfectées avec soit un plasmide vide soit un vecteur d'expression codant pour la protéine EB1 avec ou sans un vecteur d'expression codant pour la protéine dont le gène fut enlevé du génome viral : respectivement BVLF1, BDLF4 et BDLF3.5. Les ARNm ont ensuite été extraits puis quantifiés par RT-qPCR en utilisant des primers spécifiques pour l'ARNm précoce BMRF1 et les ARNm tardifs de BcLF1 et BDLF1 (cf. figure 50). Des oligonucléotides spécifiques pour la GAPDH ont été utilisés pour normaliser la quantité d'ARNm viraux par rapport au gène de ménage.

Dans les trois lignées cellulaires utilisées, transfectées avec un plasmide vide, les ARNm viraux précoces et tardifs ne sont pas détectés comme attendu (cf. figure 50). Ces même cellules transfectées avec un vecteur d'expression codant pour la protéine EB1, expriment efficacement l'ARNm précoce BMRF1 et très faiblement voire pas du tout les ARNm tardifs de BcLF1 et BDLF1. À contrario, l'induction du cycle productif avec EB1 et la transcomplémentation avec les vecteurs d'expression codant pour l'une des protéines virales dont le gène a été enlevé du génome viral, respectivement BVLF1 (A), BDLF4 (B) et BDLF3.5 (C) induit une augmentation drastique de l'expression des ARNm tardifs de BcLF1 et BDLF1. De plus, ces conditions ne perturbent pas l'expression de l'ARNm précoce BMRF1.

Lorsque BDLF3.5 est transcomplémentée dans les cellules HEK293_{EBV Δ BDLF3.5}, la quantité d'ARNm viraux tardifs synthétisés est bien plus faible qu'avec BVLFL1 et BDLF4 (cf. figure 50). Ces données peuvent être corroborées avec le fait que l'ADN viral est mal répliqué dans cette lignée cellulaire. Il faut savoir que la lignée utilisée dans cette expérience correspond à un clone cellulaire préalablement sélectionné pour la manipulation car elle répondait parfaitement aux caractéristiques attendues lors des essais luciférase. Peut être qu'un autre clone cellulaire aurait mieux répondu à la réplication de l'ADN viral. Néanmoins, il est possible aussi que la région codante de BDLF3.5 que nous avons enlevé du génome viral contienne des séquences d'ADN cis-régulatrices impliquées d'une manière ou d'une autre dans la réplication de l'ADN viral. Ainsi, la transcomplémentation du vecteur d'expression codant pour la protéine BDLF3.5 dans les cellules HEK293_{EBV Δ BDLF3.5} ne permettrait pas une réplication optimale de l'ADN viral. Une autre explication est que la cellule considère le virus comme non essentiel pour son cycle et entraîne sa chromatinisation de manière drastique, menant à un blocage de l'expression des gènes viraux.

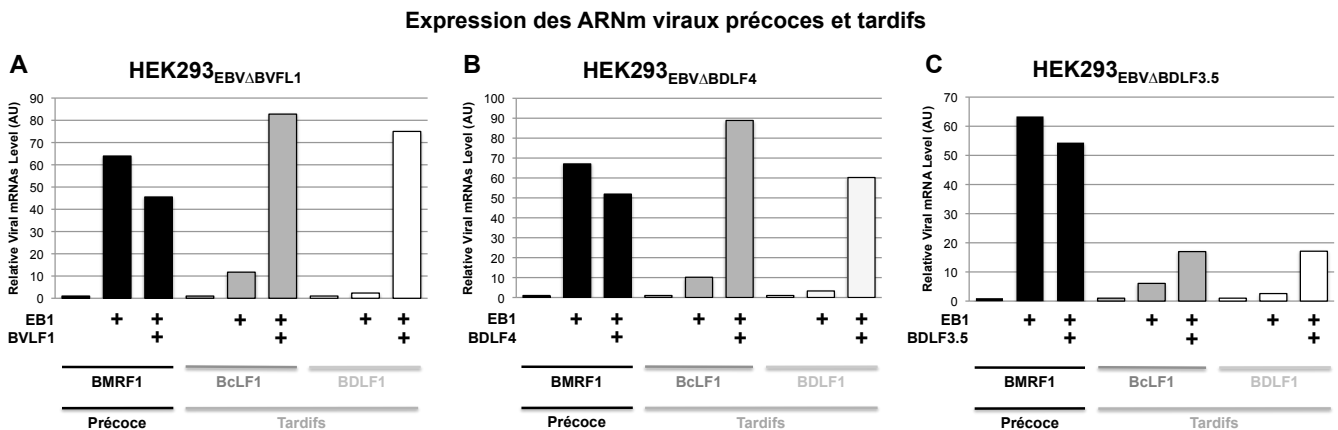


Figure 50 : BVLFL1, BDLF4 et BDLF3.5 augmentent l'expression des ARNm viraux tardifs.

Les ARNm totaux des cellules HEK293_{EBV Δ BVLFL1} (A), HEK293_{EBV Δ BDLF4} (B) et HEK293_{EBV Δ BDLF3.5} (C), transfectées avec soit un vecteur vide soit un vecteur d'expression codant pour la protéine virale EB1 avec ou sans un plasmide d'expression codant pour la protéine virale dont le gène a été enlevé du génome viral ; respectivement BVLFL1 (A), BDLF4 (B) et BDLF3.5 (C), ont été purifiés. Ils ont ensuite subi une étape de transcription inverse et ont été analysés par qPCR avec des primers spécifiques pour le gène précoce BMRF1 et pour les gènes tardifs BcLF1 et BDLF1.

Ainsi, l'ensemble de ces expériences (essais luciférase, expression protéique, réplication de l'ADN viral et expression des ARNm) permet de conclure que BVLFL1, BDLF4 et BDLF3.5 sont absolument nécessaires à l'expression des gènes tardifs du virus d'Epstein-Barr. Ces protéines n'ont pas de rôle dans l'expression des gènes viraux précoces testés ni

dans la réplication de l'ADN.

5) Conclusion sur le rôle de BVLF1, BDLF4 et BDLF3.5.

Ainsi, en utilisant une approche de système rapporteur (TATT-Luc) et une analyse génétique avec un virus recombinant amputé des gènes d'intérêts, nous avons démontré que les protéines virales BVLF1, BDLF4 et BDLF3.5 sont bien impliquées dans l'expression des gènes viraux de ces virus. De plus, nous avons déjà démontré qu'elles étaient insuffisantes seules pour ce mécanisme (Aubry et al., 2014).

Si les résultats avec BVLF1 et BDLF4 sont très encourageants et concluants, ceux avec BDLF3.5 sont à prendre avec recul. En effet, même si les orthologues de ce gène ont été démontrés comme impliqués dans l'expression des gènes tardifs de leurs virus, nos analyses génétiques avec BDLF3.5 sont plus nuancées, probablement à cause d'un problème viral.

En effet, la détection de la glycoprotéine tardive gp350 est très faible. Ceci peut s'expliquer par le fait que l'ADN viral est très faiblement répliqué, engendrant de ce fait, moins d'ARNm tardifs synthétisés puisque l'expression de ces derniers est dépendante de la réplication du génome viral. Effectivement, alors qu'avec BVLF1 et BDLF4 la quantité d'ARNm viraux est multipliée par des facteurs allant de x60 à x90, ceux du clone HEK293_{EBVΔBDLF3.5} ne sont que de x18.

Une question demeure toutefois, pourquoi le virus n'arrive-t-il pas à répliquer alors que EB1 est bien exprimé et que le cycle viral est productif est bien induit (en témoigne l'expression efficace de EB2, figure 48) ? La première idée serait de tester une autre lignée cellulaire HEK293_{EBVΔBDLF3.5}. Ceci permettrait de vérifier si ces données sont reproductibles ou non entre clones. Si des résultats similaires sont observés dans ces différents clones, alors il pourrait être envisagé que l'amputation de la région codante de BDLF3.5 enlève une séquence d'ADN impliquée dans la réplication du génome viral lors du cycle productif.

Concernant la dernière protéine d'EBV potentiellement impliquée dans l'expression des gènes viraux tardifs, BGLF3, aucune lignée cellulaire stable n'a pour l'instant pu être établie pour étudier la fonctionnalité de cette protéine.

C) L'interactome du vPIC.

Ainsi, nous avons démontré le rôle de BcRF1 (Gruffat et al., 2012), BFRF2 (Aubry et

al., 2014), BVL1, BDLF4 et BDLF3.5 (cette thèse) dans l'expression des gènes viraux tardifs, et imaginer que BGLF3 est elle aussi impliquée dans ce mécanisme. Nous avons également publié que le vPIC, en particulier BcRF1, est capable d'interagir avec une des sous-unités de l'ARN polymérase II cellulaire. De plus, lors de cette même étude, nous avons réalisé les interactions *in vitro* des différentes protéines du vPIC. Afin de vérifier si ces interactions étaient également valables ou différentes *in vivo*, plusieurs GST pull-down dans des cellules HEK293 ont été réalisés. Ces données permettront peut-être d'établir un modèle d'assemblage du vPIC.

1) Le vPIC et l'ARN polymérase II cellulaire.

Comme évoqué précédemment, nous avons démontré que l'ARN polymérase II cellulaire interagissait *in vivo* avec BcRF1. Nous avons émis l'hypothèse que d'autres membres du vPIC pouvaient être liées à cette protéine cellulaire. Pour y répondre, un vecteur d'expression codant pour une protéine chimérique GST-X (X = protéine du vPIC) a été transfecté dans des cellules HEK293. Un GST pull-down a ensuite été réalisé comme décrit dans la partie matériel et méthode. La partie révélation a été effectuée avec un anticorps dirigé contre la grande sous-unité de l'ARN pol II cellulaire : Rpb1.

Le panel en haut à gauche de la figure 51 révèle l'expression de la protéine cellulaire et permet de contrôler indirectement la quantité de protéines déposées. Rpb1 est efficacement détectée dans tous les points (cf. figure 51, panel en haut à gauche).

Après purification, et de manière inattendue, seule une protéine semble capable dans un contexte non infectieux de recruter l'ARN polymérase II cellulaire : la TBP-like virale BcRF1 (cf. figure 51, panel en haut à droite). Bien que les autres protéines du vPIC couplées à la GST soient correctement exprimées, aucune d'elles n'est capable de retenir Rpb1 lors de cette chromatographie d'affinité (cf. figure 51, panels du bas). Lorsque toutes les protéines du vPIC sont présentes, l'interaction avec l'ARN pol II n'est pas modifiée.

Ce résultat démontre clairement que seule la TBP-like virale BcRF1 permet le recrutement de l'ARN polymérase II cellulaire pour transcrire les gènes viraux tardifs.

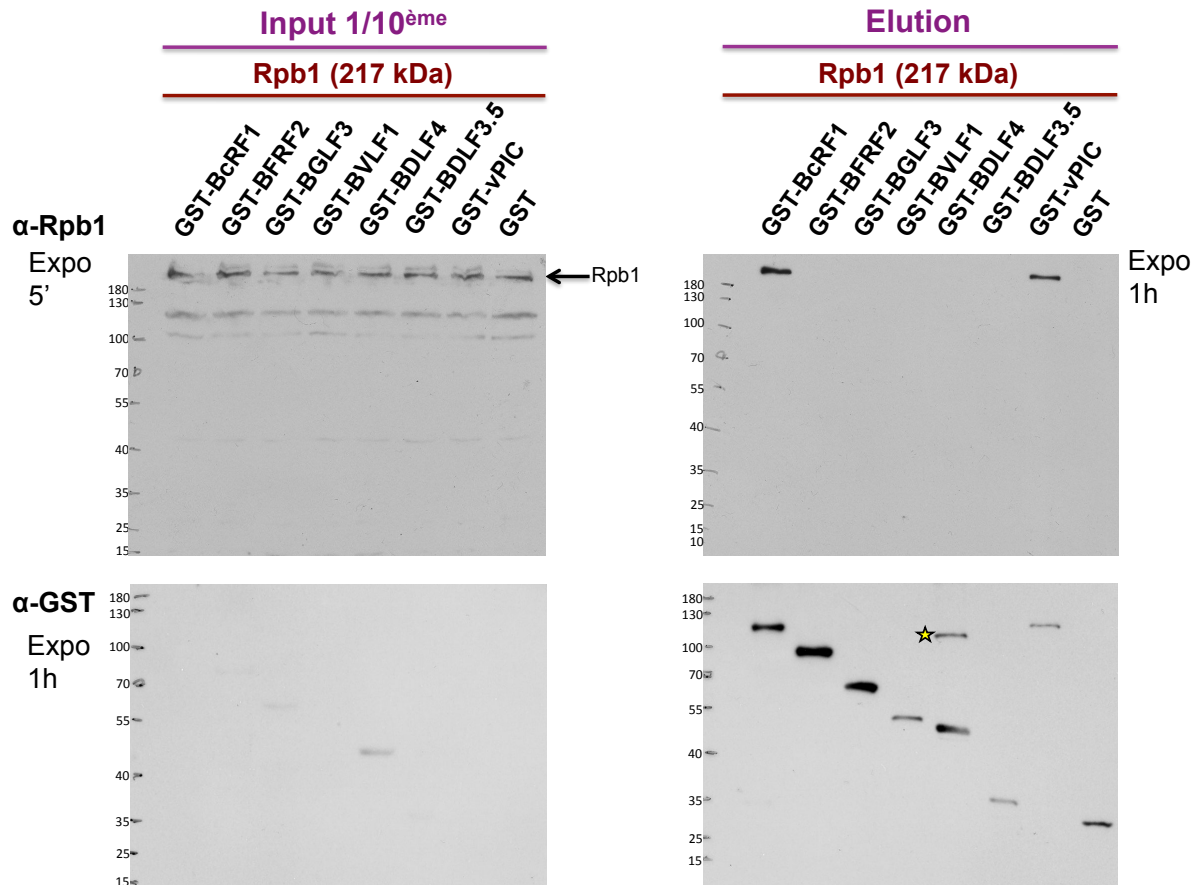


Figure 51 : La TBP-like virale BcRF1 est la seule protéine du vPIC interagissant avec l'ARN pol II.

Les plasmides d'expression codant pour les protéines chimériques GST-X (X = protéine du vPIC) ou Flag-GST ont été transfectés dans des cellules HEK293 dans le but de vérifier l'interaction de ces protéines avec l'ARN polymérase II cellulaire. Les cellules sont ensuite récoltées et lysées. L'input, correspondant à 10% du lysat cellulaire est représenté à gauche. Les 90 % restant du lysat ont été incubés sur billes Glutathione-Sépharose. Les protéines retenues sur la colonne ont été lavées, éluées et séparées sur SDS-PAGE avant d'être transférées sur membrane de nitrocellulose. Ces protéines ont été détectées après Western Blot à l'aide d'un film autoradiographique (panel de droite). Le panel du haut correspond à la détection de l'ARN pol II cellulaire (Rpb1) et le panel du bas fait suite à une révélation avec un anticorps anti-GST. L'astérisque correspond à la dimérisation de la protéine GST-BDLF4. La bande observée dans la colonne GST-vPIC correspond à la protéine GST-BcRF1 qui a été utilisée comme appât en présence des cinq autres protéines du vPIC.

2) Interactions individuelles entre protéines du vPIC.

Après avoir publié les interactions *in vitro* entre les membres du vPIC, des GST pull-down *in vivo* dans des cellules HEK293 ont été effectués afin de confirmer ou infirmer les données obtenues en vue d'établir un ordre d'assemblage du vPIC.

Pour ce faire, un vecteur d'expression codant pour une protéine chimérique GST couplée à l'une des protéines du vPIC a été transfecté avec un vecteur d'expression codant pour une autre protéine du vPIC couplée à un épitope 3xFlag en N-ter de la protéine virale ou toutes ensemble. Un GST pull-down a ensuite été réalisé comme indiqué dans la partie matériel et méthodes avant de révéler les interactions par Western-Blot et un film autoradiographique. Pour une meilleure visibilité et compréhension compte tenu de la multitude de GST pull-down effectués, les résultats des différentes expériences *in vivo* ont été résumés sous forme d'un tableau. À titre d'exemple, une seule expérience sera exposée. Les autres figures seront présentées en annexe 1. De plus, afin d'identifier rapidement les protéines sur le WB, le poids moléculaire théorique arrondi de chacune des protéines utilisées est résumé dans le tableau 6, ci-dessous.

Protéines	BcRF1	BFRF2	BGLF3	BVLF1	BDLF4	BDLF3.5
3xFlag	88 KDa	68 KDa	41 KDa	35 KDa	30 KDa	16 KDa
Flag-GST	110 KDa					

Tableau 6 : Poids moléculaire théorique en kilo Dalton, des protéines du vPIC possédant un épitope 3xFlag ou un épitope Flag-GST dans leur région N-terminale.

a) GST-BcRF1

Afin de contrôler l'expression des protéines après la transfection, 10% du lysat cellulaire total est déposé sur gel SDS-PAGE 10% (Input), correspondant au panel gauche de la figure 52. Sur le panel de droite sont représentées les révélations avec un anticorps anti-Flag, des interactions avec Flag-GST-BcRF1.

Aucune protéine composant le vPIC n'est retenue par la protéine Flag-GST. Flag-GST-BcRF1, d'une masse moléculaire approximative de 110 KDa est bien détectée et est capable d'interagir avec tous les membres du vPIC mais de manière différente. En effet, BcRF1 interagit plus fortement avec BGLF3 qu'avec les autres protéines du vPIC. Flag-GST-BcRF1 interagit également avec elle-même et BFRF2 de manière efficace. L'interaction de Flag-GST-BcRF1 avec BVLF1 et BDLF4 est plus faible mais clairement identifiable. En revanche, compte tenu de la faible expression de la protéine BDLF3.5, il est difficile de conclure sur la qualité de l'interaction même s'il semble, sur un temps long d'exposition, que BDLF3.5 et GST-BcRF1 interagissent.

Lorsque les six protéines du vPIC sont exprimées ensemble, BFRF2, BGLF3, BVLF1 et BDLF4 restent accrochées sur la colonne de la même manière que lorsqu'elles sont exprimées individuellement. À contrario, BDLF3.5 semble mieux retenue avec Flag-GST-BcRF1 quand le vPIC est reconstitué que lorsqu'elle est présente individuellement (cf. figure 52).

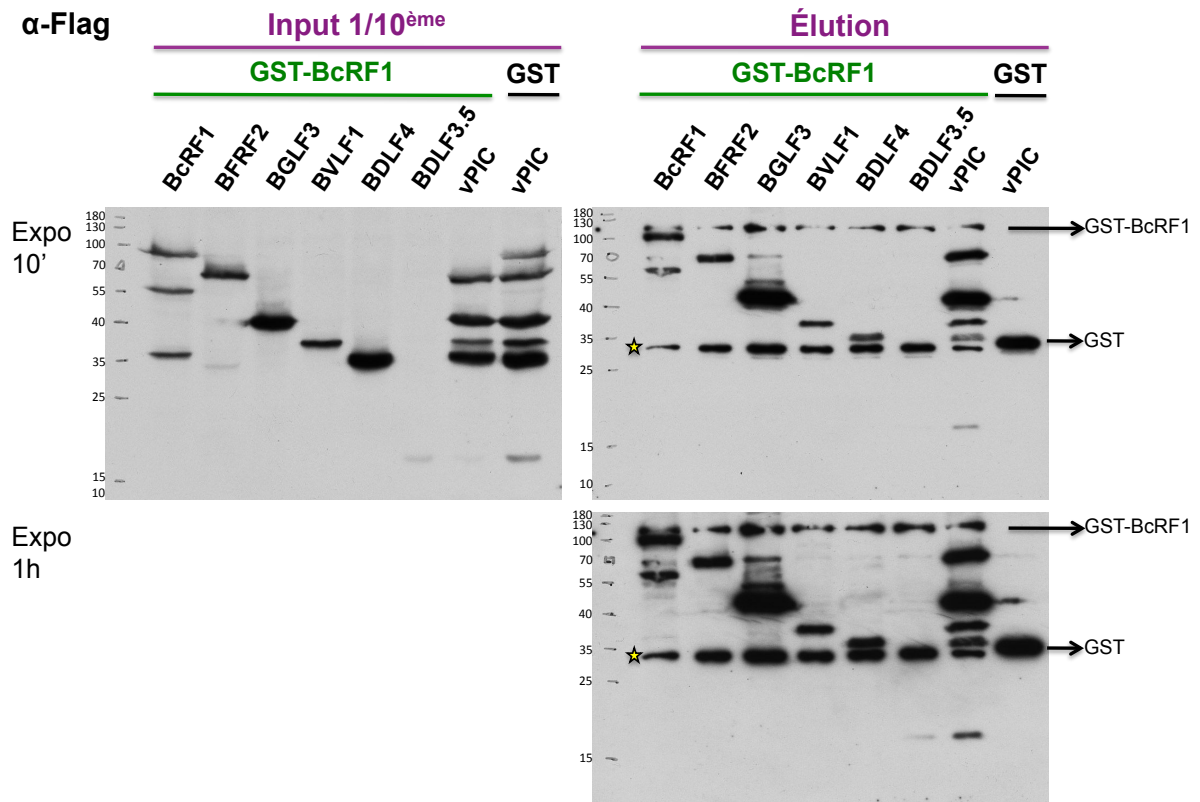


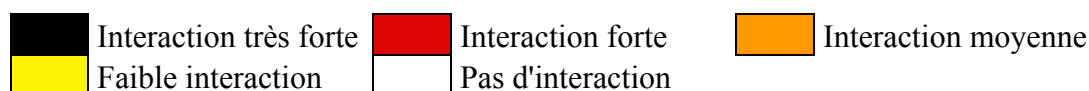
Figure 52 : La protéine Flag-GST-BcRF1 interagit différemment avec tous les composants du vPIC.

Les vecteurs d'expression codant pour les protéines chimériques Flag-GST-BcRF1 ou Flag-GST ont été transfectés dans des cellules HEK293 avec soit un plasmide d'expression codant pour une des protéines du vPIC soit avec tous les plasmides codant pour les cinq protéines du vPIC restantes. Les cellules ont été récoltées et lysées. L'input, correspondant à 10% du lysat cellulaire est représenté à gauche. Les 90 % restant du lysat ont été incubés sur billes Glutathione-Sépharose. Les protéines retenues sur la colonne ont été lavées, éluées et séparées sur SDS-PAGE avant d'être transférées sur membrane de nitrocellulose. Ces protéines ont été détectées après Western Blot à l'aide d'un film autoradiographique (panel de droite). Un temps d'exposition court (en haut, 10 minutes) et un temps long (en bas, 1h), permettent d'observer les différentes interactions. L'astérisque jaune marque une bande aspécifique pouvant être la protéine Flag-GST qui serait clivée à partir de la protéine Flag-GST-BcRF1.

b) Synthèse des interactions individuelles des protéines du vPIC.

Afin de synthétiser toutes les interactions observées par GST pull-down, celles-ci ont été regroupées dans un tableau (cf. tableau 7). La colonne la plus à gauche correspond aux protéines d'appâts GST-X utilisées pour précipiter les partenaires viraux. La ligne du haut correspond aux différentes protéines détectées à l'aide de l'épitope 3xFlag.

Les couleurs utilisées pour les interactions sont associées à l'intensité de l'interaction entre la protéine d'appât (GST-X) et la protéine proie. Elle est à relativiser avec la quantité de protéines détectées dans l'input. Plus la couleur est foncée (noire, rouge), plus l'interaction est forte. À l'inverse, plus la couleur est claire (jaune, blanche), plus l'interaction est faible. Le code couleur est donc :



Ainsi, les résultats présentés précédemment avec GST-BcRF1 sont résumés dans la première ligne colorée du tableau 7. Par exemple, la protéine GST-BcRF1 interagit correctement avec elle-même et BFRF2. L'interaction entre GST-BcRF1 et BGLF3 est très forte. BVLF1 et BDLF4 interagissent modérément avec GST-BcRF1. Enfin, BDLF3.5 interagit très faiblement avec GST-BcRF1.

Protéines	BcRF1	BFRF2	BGLF3	BVLF1	BDLF4	BDLF3.5
GST-BcRF1	Interaction forte	Interaction forte	Interaction très forte	Interaction moyenne	Interaction moyenne	Faible interaction
GST-BFRF2	Faible interaction	Interaction moyenne	Interaction très forte	Faible interaction	Faible interaction	Pas d'interaction
GST-BGLF3	Interaction très forte	Interaction très forte	Faible interaction	Faible interaction	Faible interaction	Pas d'interaction
GST-BVLF1	Interaction moyenne	Interaction forte	Faible interaction	Interaction moyenne	Interaction moyenne	Pas d'interaction
GST-BDLF4	Interaction moyenne	Interaction moyenne	Interaction moyenne	Faible interaction	Interaction forte	Pas d'interaction
GST-BDLF3.5	Interaction forte	Interaction forte	Faible interaction	Interaction moyenne	Interaction moyenne	Pas d'interaction

Tableau 7 : Résumé des interactions des différents membres du vPIC entre eux.

Parmi les interactions très fortes, celle entre GST-BcRF1 et BGLF3 est réciproque. De la même manière, GST-BFRF2 interagit fortement avec BGLF3 et inversement. En revanche, GST-BFRF2 interagit assez faiblement avec BcRF1, BVLF1 et BDLF4 et pas du tout avec BDLF3.5. GST-BGLF3 interagit assez faiblement avec elle-même, BVLF1 et BDLF4 ; très fortement avec BcRF1 et BFRF2 et pas du tout avec BDLF3.5. Concernant GST-BVLF1, elle semble avoir une meilleure affinité pour BFRF2 que pour les autres protéines virales. De plus, la liaison entre GST-BVLF1 et BDLF3.5 est inexistante dans ces conditions alors qu'elles semblent se stabiliser l'une l'autre au niveau protéique. Pour GST-BDLF4, seule une interaction avec elle-même semble forte. Les autres protéines virales interagissent modérément avec elle. En revanche, il semble que BVLF1 soit très faiblement retenue avec

GST-BDLF4 et que BDLF3.5 n'interagissent pas avec. Si la protéine BDLF3.5 est très difficilement détectée par WB, la protéine GST-BDLF3.5 permet d'apporter quelques éléments de réponse. Effectivement, cette dernière interagit fortement avec BcRF1 et BFRF2, modérément avec BVLF1 et BDLF4 et très faiblement avec BGLF3. Cette expérience avec GST-BDLF3.5 est plus pertinente qu'avec 3xFlag-BDLF3.5 car cette dernière est à chaque fois très faiblement détectée dans les inputs au contraire de la protéine GST-BDLF3.5.

Ainsi, l'intégralité de ces GST pull-down réalisés *in vivo* démontre la complexité des interactions entre les différents membres du vPIC. En effet, la quasi-totalité des protéines interagissent les unes avec les autres, rendant impossible toute interprétation dans l'assemblage du vPIC. Nous avons simplifié sous la forme d'un schéma les différentes interactions du vPIC (cf. figure 53) avec toutes celles qui ont été observées au cours des multiples expériences.

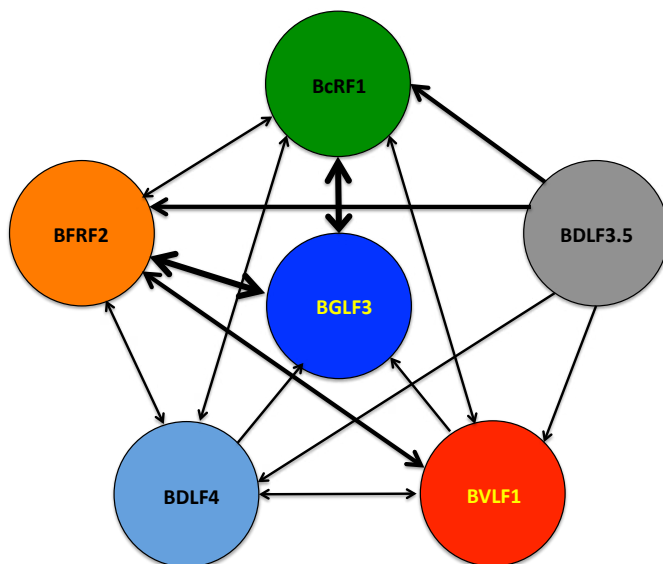


Figure 53 : Représentation schématique des différentes interactions constatées dans le vPIC.

Toutes les interactions observées lors des GST pull-down sont représentées avec des flèches de tailles variables. Les flèches fines correspondent aux interactions faibles et les épaisses aux interactions fortes.

Par ailleurs, afin d'avoir un ordre d'idée des différentes interactions, les figures des GST pull-down de cette partie sont présentées en annexe 1.

3) L'interactome du vPIC : toutes les protéines du vPIC moins une.

a) GST-BcRF1 + vPIC-1.

Le vPIC étant un complexe, il est intéressant d'observer l'interaction des protéines en association. Pour ce faire, un plasmide d'expression codant pour une protéine GST ou un

plasmide d'expression codant pour une protéine chimérique GST-X (X=protéine virale) a été transfecté dans des cellules HEK293 avec les vecteurs d'expression codant pour seulement quatre des cinq protéines virales du vPIC. Ainsi, les différences de recrutement des protéines du vPIC pourront être observées.

Comme précédemment, pour une meilleure visibilité et compréhension, tous les résultats d'interactions ont été regroupés sous forme de tableau synthétique. Toutefois, nous exposerons uniquement les résultats avec GST-BcRF1 pour illustrer nos données. Ceux obtenus pour les autres protéines GST-X sont regroupés en annexe 2.

Afin de contrôler l'expression des protéines après transfection, 10% du lysat cellulaire total est déposé sur gel SDS-PAGE 10% (Input), correspondant au panel gauche de la figure 54. Sur le panel de droite sont représentées les révélations des interactions avec la protéine chimérique Flag-GST-BcRF1. Ceci a été réalisé après incubation des deux membranes avec un anticorps anti-Flag.

Aucune protéine composant le vPIC n'est retenue par la protéine Flag-GST. Flag-GST-BcRF1, d'une masse moléculaire approximative de 110 KDa est bien détectée une fois purifiée sur billes Glutathione Sépharose. Quand le vPIC est reconstitué au complet, elle est capable de retenir tous les membres du complexe mais de manière différente (cf. figure 54).

En effet, BGLF3 est très bien retenue sur la colonne, alors que BFRF2 l'est de manière moins efficace. BVLf1 est également retenue correctement sur la colonne GST-BcRF1 comparé à l'input au contraire de BDLF4 qui est très faiblement purifiée sur la colonne (cf. figure 54). Pour BDLF3.5, il est compliqué de conclure car la protéine est faiblement exprimée. Toutefois, sur une exposition longue, BDLF3.5 est retenue de la même manière que BVLf1 sur la colonne, suggérant la présence de BDLF3.5 dans le complexe dans cette situation.

Dès qu'une protéine est enlevée volontairement du complexe, les résultats de purification varient d'une protéine à l'autre. Par exemple, lorsque BFRF2 est « oubliée », BGLF3 est retenue aussi efficacement que lorsque le vPIC est complet. En revanche, BDLF4 est mieux retenue dans ce complexe, alors que BVLf1 et BDLF3.5 ne sont quasiment plus présentes après purification.

Quand BGLF3 est enlevée, seule BFRF2 est maintenue de la même manière que lorsque le vPIC est complet. De plus, BDLF4 est très faiblement retenue sur la colonne alors que BVLf1 et BDLF3.5 ne sont plus accrochées.

L'absence de BDLF4 dans le complexe n'engendre aucune modification dans le complexe hormis un meilleur recrutement de BDLF3.5.

L'« oubli » de BVLV1 n'entraîne que la non rétention de BDLF3.5. Les autres protéines sont retenues aussi efficacement que lorsque le vPIC est complet. Ces résultats sont également observables lorsque BDLF3.5 est enlevée ; à savoir que seule BVLV1 n'est plus retenue.

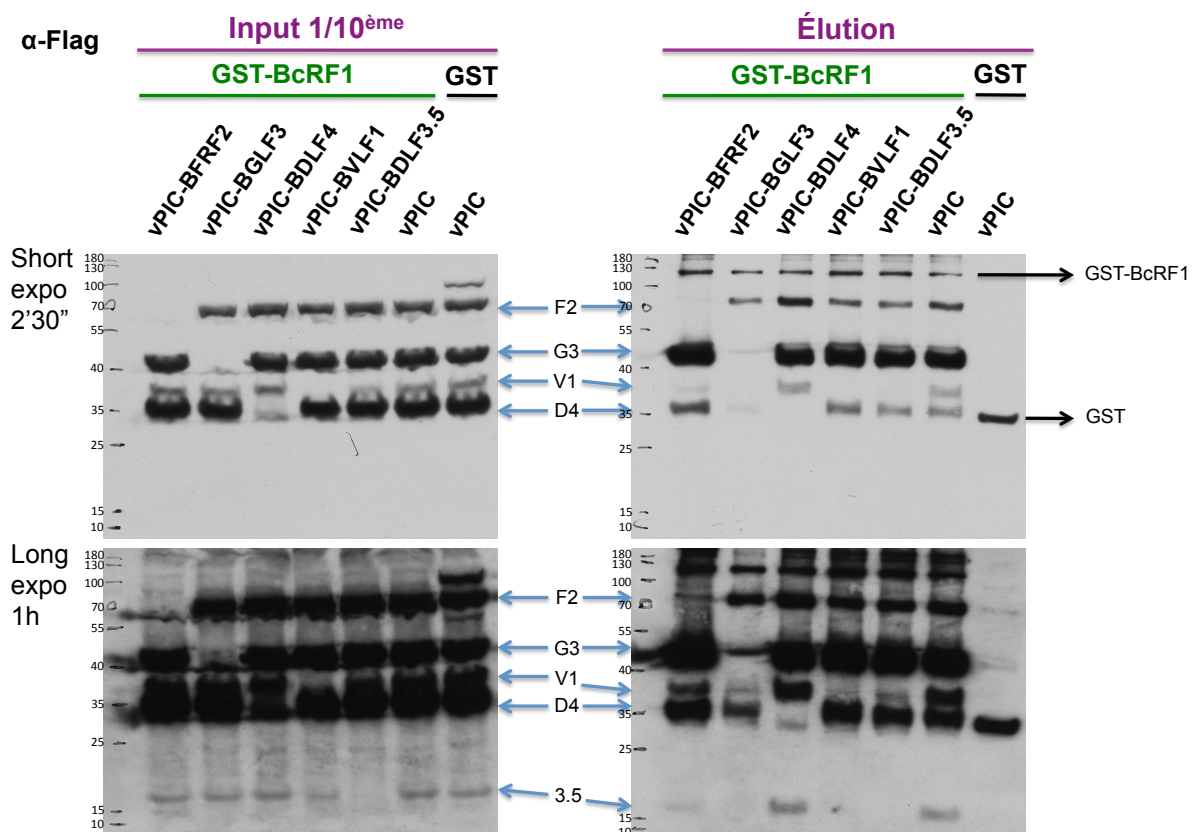








Figure 54 : La protéine Flag-GST-BcRF1 maintient différemment plusieurs composants du vPIC.

Les vecteurs d'expression codant pour les protéines chimériques Flag-GST-BcRF1 ou Flag-GST ont été transfectés dans des cellules HEK293 avec soit les différents plasmides d'expression codant pour les cinq autres protéines du vPIC restantes soit seulement quatre d'entre elles. Les cellules ont été récoltées et lysées. L'input, correspondant à 10% du lysat cellulaire est représenté à gauche. Les 90 % restant du lysat ont été incubés sur billes Glutathione-Sépharose. Les protéines retenues sur la colonne ont été lavées, éluées et séparées sur SDS-PAGE avant d'être transférées sur membrane de nitrocellulose. Ces protéines ont été détectées après Western Blot à l'aide d'un film autoradiographique (panel de droite). Un temps d'exposition court (en haut, 2 minutes et 30 secondes) et un temps long (en bas, 1h), permettent d'observer les différentes interactions.

L'ensemble des résultats observés a été résumé sous forme de tableau (cf. tableau 8) avec, comme précédemment, un code couleur pour représenter l'intensité des interactions (par

rapport aux inputs). Les cellules barrées correspondent aux conditions non testées. La couleur bleu ciel est elle associée à l'absence volontaire d'une des protéines virales.

Code couleur :

	Interaction très forte		Interaction forte		Interaction moyenne
	Faible interaction		Pas d'interaction		Absente

Ainsi, les résultats obtenus avec GST-BcRF1 sont représentés comme ci-dessous :

GST-BcRF1








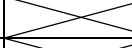
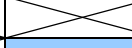





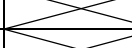













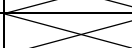






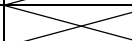






	vPIC-c1	vPIC-F2	vPIC-G3	vPIC-V1	vPIC-D4	vPIC-3.5	vPIC
BcRF1							
BFRF2							
BGLF3							
BVLF1							
BDLF4							
BDLF3.5							

Tableau 8 : Résultats des interactions par GST pull-down avec GST-BcRF1+vPIC-1.

b) Synthèse des interactions des protéines du vPIC.

Ces expériences ont également été réalisées avec GST-BGLF3, GST-BDLF4 et GST-BDLF3.5, ont été répertoriées sous forme de tableaux pour chaque protéine GST-X (cf. tableaux 9, 10, 11). Les figures correspondantes sont présentées en annexe 2.

GST-BGLF3










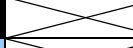
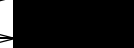
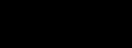
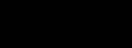
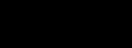


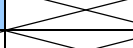





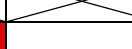
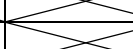


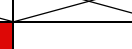



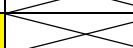






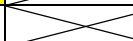




	vPIC-c1	vPIC-F2	vPIC-G3	vPIC-V1	vPIC-D4	vPIC-3.5	vPIC
BcRF1							
BFRF2							
BGLF3							
BVLF1							
BDLF4							
BDLF3.5							

Tableau 9 : Résultats des interactions par GST pull-down avec GST-BGLF3+vPIC-1.

GST-BDLF4












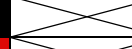






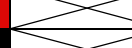






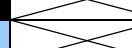






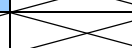
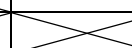





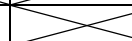


	vPIC-c1	vPIC-F2	vPIC-G3	vPIC-V1	vPIC-D4	vPIC-3.5	vPIC
BcRF1							
BFRF2							
BGLF3							
BVLF1							
BDLF4							
BDLF3.5							

Tableau 10 : Résultats des interactions par GST pull-down avec GST-BDLF4+vPIC-1.

GST-BDLF3.5

	vPIC-c1	vPIC-F2	vPIC-G3	vPIC-V1	vPIC-D4	vPIC-3.5	vPIC
BcRF1						X	
BFRF2						X	
BGLF3						X	
BVLF1						X	
BDLF4						X	
BDLF3.5	X	X	X	X	X	X	X

Tableau 11 : Résultats des interactions par GST pull-down avec GST-BDLF3.5+vPIC-1.

L'intégralité de ces données confirme plusieurs éléments. Tout d'abord, il semble qu'un trimère constitué de BcRF1, BFRF2 et BGLF3 s'associe très fortement avec BGLF3. Celle-ci semble former un pont entre les deux autres protéines virales. De plus, elle semble être au cœur du complexe car elle interagit avec tous les composants du vPIC (hormis BDLF3.5). De plus, son absence dans toutes les conditions testées mène à la perte de la majorité des autres facteurs, et ce, quelque soit la protéine d'appât (cf. tableaux 8, 9, 10 et 11).

Ensuite, la protéine BDLF4, qui interagit modestement avec BcRF1 et BGLF3 exprimée individuellement, nécessite la présence de ces deux protéines pour être présente dans le vPIC (cf. tableau 8, colonne vPIC-G3 et tableau 9, colonne vPIC-c1).

Bien qu'aucune interaction évidente entre BVLF1 et BDLF3.5 (cf. tableau 7) n'ai pu être détectée, il semble que ces deux protéines soient associées d'une manière différente. Effectivement, l'absence de l'une d'elles conduit à la perte de l'autre et ce, dans toutes les conditions testées (cf. tableaux 8 colonnes vPIC-V1 et vPIC-3.5 ; 9, colonnes vPIC-V1 et vPIC-3.5 ; et 10, colonnes vPIC-V1 et vPIC-3.5). De plus, BVLF1 et BDLF3.5 ont besoin que BFRF2 et surtout BGLF3 soient présents pour se maintenir dans le vPIC (cf. tableaux 8, colonnes vPIC-F2 et vPIC-G3 ; 9, colonne vPIC-F2 ; 10, colonne vPIC-G3).

Enfin, il semble exister une compétition entre BDLF4 et le complexe BVLF1-BDLF3.5. En effet, lorsque BDLF4 est présente, BVLF1 et BDLF3.5 sont difficilement retenues dans le vPIC. À l'inverse, l'absence de BDLF4 conduit à un meilleur maintien de BVLF1 et BDLF3.5 dans le vPIC (cf. tableaux 8, colonne vPIC-D4 ; 9, colonne vPIC-D4 ; et 11, colonne vPIC-4).

Ainsi, l'intégralité de nos données laissent entendre qu'il n'y aurait pas qu'un seul complexe vPIC mais plusieurs sous complexes avec d'un côté BDLF4 associée à BcRF1-BGLF3 et de l'autre BVLF1-BDLF3.5 associées à BFRF2 et BGLF3. Ce raisonnement laisse

sous-entendre que BGLF3 serait une pierre angulaire du vPIC. En effet, il semble que ce soit la seule protéine capable de purifier l'intégralité des membres du vPIC.

Néanmoins, l'absence de résultats avec GST-BFRF2 et GST-BVLF1 est toutefois préjudiciable car ils auraient permis de mieux comprendre comment BDLF4 est retenue dans ce complexe.

4) L'association avec BDLF4 dans le complexe viral.

Afin d'obtenir plus d'informations sur le recrutement de BDLF4 dans le vPIC, des vecteurs d'expression codant pour les protéines BcRF1, BFRF2, et BDLF4 en présence du vecteur d'expression codant pour la protéine GST-BGLF3 ont été transfectés dans des cellules HEK293 sous différentes conditions. Ceci permettra d'observer son incorporation dans le complexe. Les résultats obtenus après purification sont présentés dans la figure 55.

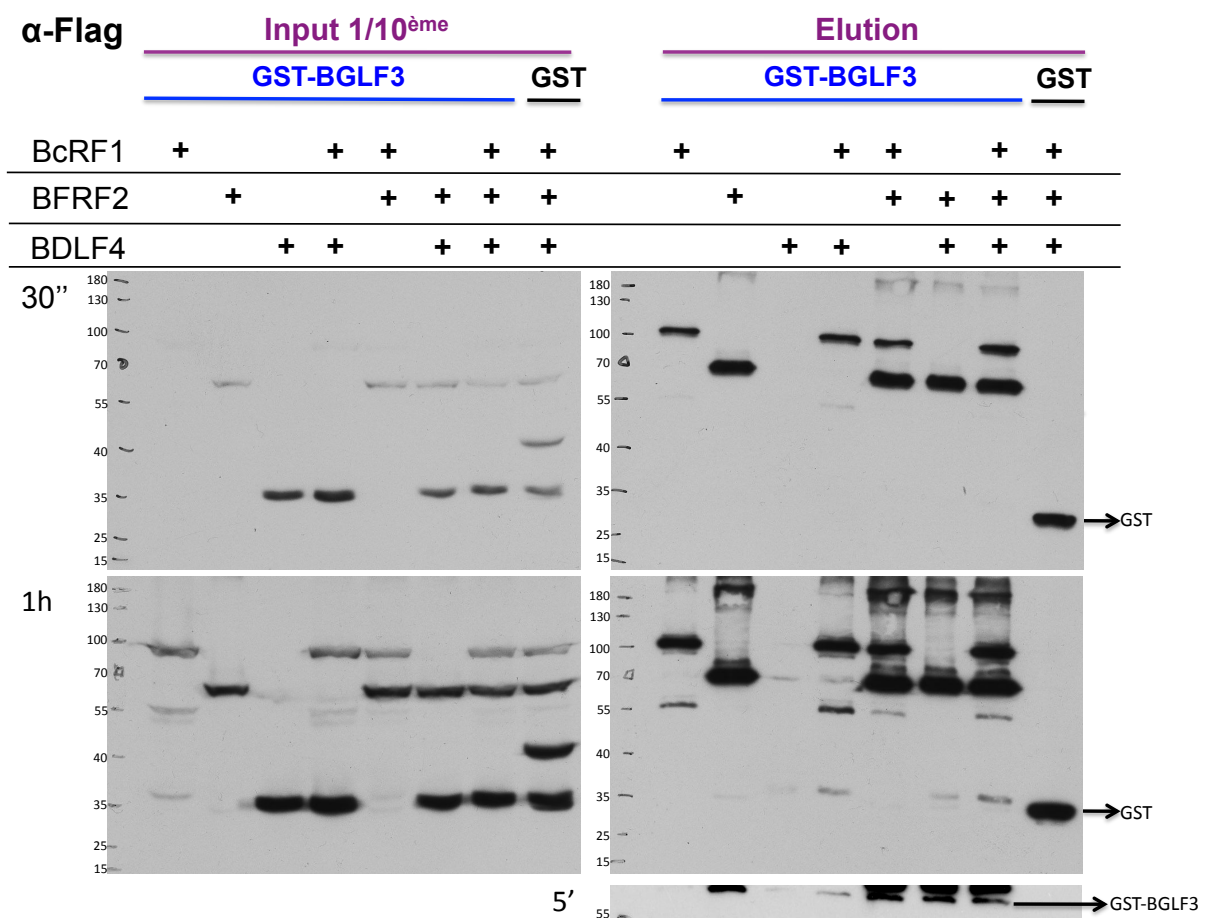


Figure 55 : Le recrutement de BDLF4 dans le vPIC est plus efficace en présence de trois autres facteurs viraux : BcRF1, BFRF2 et BGLF3.

Les plasmides codant pour les protéines chimériques GST-BGLF3 ou Flag-GST ont été transfectés dans des cellules HEK293 avec différents plasmides d'expression codant pour les différentes protéines du vPIC susceptibles de faciliter le recrutement de BDLF4 dans le complexe : BcRF1 et BFRF2. Les cellules sont ensuite récoltées et lysées. L'input, correspondant à 10% du lysat cellulaire est représenté à gauche. Les 90 % restant du lysat ont été incubés sur billes Glutathione-Sépharose. Les protéines retenues sur la colonne ont été lavées, éluées et séparées sur SDS-PAGE avant d'être transférées sur membrane de nitrocellulose. Ces protéines ont été détectées après Western Blot à l'aide d'un film autoradiographique (panel de droite). Un temps d'exposition court (en haut, 30 secondes) et un temps long (au milieu, 1h) permettent d'observer les différentes interactions. Le panel du bas montre l'expression de la protéine GST-BGLF3 dans les différents points après purification.

Comme attendu, la purification de BcRF1 ou BFRF2 seule avec GST-BGLF3 est très efficace alors que celle avec BDLF4 seule est faible (cf. figure 55). La co-expression de BcRF1 et BDLF4 permet de mieux retenir BDLF4 sur la colonne. À l'inverse, la co-expression de BFRF2 avec BDLF4 n'engendre qu'un faible enrichissement de BDLF4 sur la colonne GST-BGLF3. En revanche, lorsque BcRF1, BFRF2 et BDLF4 sont exprimées ensemble, il semble que cette dernière protéine (BDLF4) soit mieux retenue sur la colonne. Ceci suggère donc que BDLF4 nécessite la présence de BcRF1 et BGLF3 ainsi que la coopération de BFRF2 pour être présente dans le vPIC.

Toutefois, ces résultats sont à relativiser. En effet, GST-BGLF3 n'est pas exprimée de manière homogène dans toutes les conditions (cf. figure 55) : BGLF3 est mieux exprimée en présence de BFRF2. Or BDLF4 est mieux retenue lorsque BFRF2 est présente en compagnie de BcRF1 et GST-BGLF3. Il sera donc important de réaliser un autre GST pull-down avec la protéine GST-BcRF1 pour observer d'éventuelle différence de recrutement de BDLF4 dans le vPIC et de confirmer le rôle de BFRF2 dans ce mécanisme.

5) Conclusions sur l'interactome vPIC.

À travers l'intégralité de ces GST pull-down, il est compliqué d'avoir une vision globale de ce système. Toutefois, l'ensemble des résultats semble indiquer que la protéine BDLF4 est maintenue dans le vPIC grâce à la protéine BcRF1 en coopération avec BGLF3. De la même manière, il semble que BDLF3.5 et BVL1 soient recrutées dans le vPIC grâce aux protéines BFRF2 et BGLF3. De plus, malgré l'absence évidente d'interaction entre BVL1 et BDLF3.5, ces deux protéines sont étroitement liées car l'absence d'une de ces deux protéines est préjudiciable pour son partenaire. Ces constatations sont résumées dans la figure 56 et permettent d'émettre un modèle d'assemblage du vPIC (cf. figure 56B). Par ailleurs, il

se pourrait que le vPIC ne soit pas formé que d'un seul complexe mais de plusieurs sous structure. Effectivement, BDLF4 semble être en compétition avec BVLV1 et BDLF3.5 et la présence de la première induirait l'absence des deux autres facteurs, et inversement (cf. tableaux 8, 9, 10 et 11).

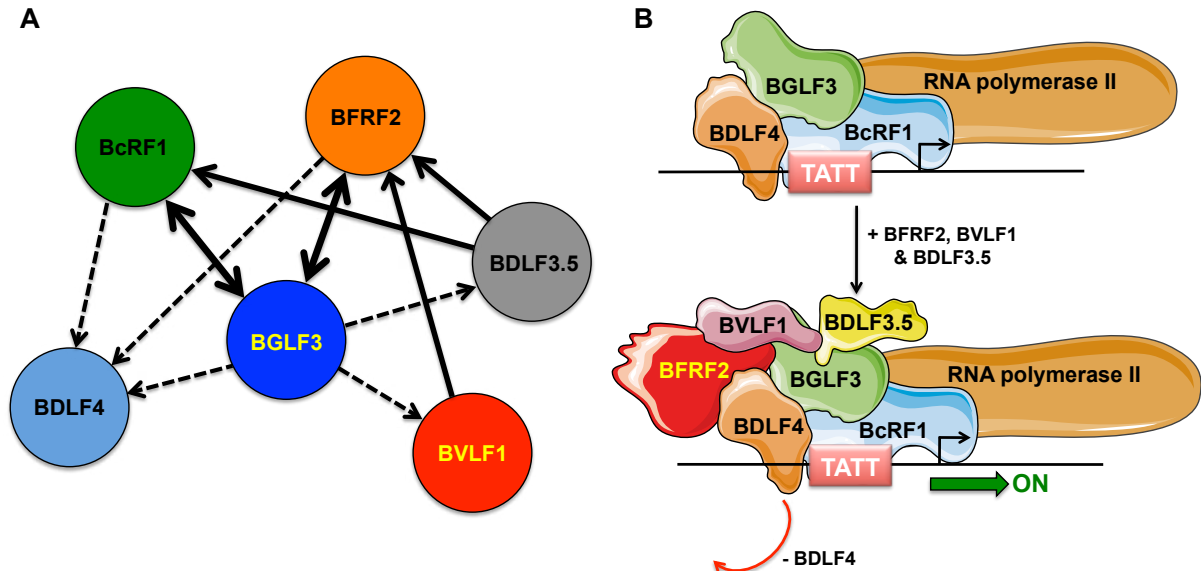


Figure 56 : Représentation schématique du potentiel modèle d'assemblage des membres du vPIC.

(A) Les différentes protéines du vPIC interagissent globalement toutes entre elles mais de manière différentes. Les interactions les plus fortes sont représentées sur ce schéma à l'aide des flèches pleines et épaisses. Les flèches en pointillés montrent des interactions faibles voire inexistantes mais qui semblent nécessaires pour que les protéines telles que BDLF4 puissent être présentes dans le vPIC ou BVLV1 et BDLF3.5. (B) Modèle d'assemblage du vPIC. BcRF1 se fixe sur la TATT-box avec BDLF4 qui lui donne la spécificité de liaison. BGLF3 fait le pont entre les deux protéines pour les maintenir assemblées et servirait de cargo pour l'arrivée de BFRF2, BVLV1 et BDLF3.5. L'assemblage de ces promoteurs permettrait simultanément d'initier la transcription et d'éjecter BDLF4 du complexe pour induire l'élongation de l'ARNm.

D) Recherche des partenaires éventuels du vPIC.

Nous avons émis l'hypothèse que d'autres protéines virales et cellulaires pourraient être impliquées dans l'expression des gènes viraux tardifs. Nous avons eu pour cela plusieurs approches. Tout d'abord, nous avons supposé que la protéine BTRF1, conservée uniquement chez les β - et γ -Herpesvirus puisse être impliquée dans l'expression des gènes viraux tardifs. Bien que son orthologue chez MHV68 (ORF23) ait été démontré comme non nécessaire à l'expression des gènes tardifs de ce virus (Ohno et al., 2012) nous avons supputé qu'elle puisse avoir un rôle auxiliaire dans le vPIC.

Deuxièmement, plusieurs études semblent mettre en évidence un rôle de la sérine/thréonine kinase virale BGLF4 dans l'expression des gènes viraux tardifs (El-Guindy et al., 2014; Whitehurst et al., 2013).

Enfin, dans le but d'avoir un panel plus large des cibles potentielles du vPIC, nous avons décidé d'utiliser une technique de double purification par chromatographie d'affinité (Tandem Affinity Purification : TAP), décrite par l'équipe du Dr. Séraphin (Puig et al., 2001). Les protéines purifiées par cette technique seront analysées par spectrométrie de masse ce qui nous permettra d'identifier les protéines virales et cellulaires associées au vPIC.

1) BTRF1 et BGLF4 ont-elles un rôle dans l'expression des gènes tardifs d'EBV ?

Pour répondre à cette question, nous avons cloné les séquences codantes pour les protéines BTRF1 et BGLF4, par Gateway, dans le plasmide pCI-3xFlag et testé en GST pull-down une éventuelle interaction de ces protagonistes avec le vPIC. D'autre part, le système rapporteur luciférase TATT-Luc, décrit précédemment (cf. figure 46), a été utilisé en lignée cellulaire HEK293. Les plasmides d'expression codant pour les différents composants du vPIC ont été transfectés dans ces cellules avec ou sans BTRF1 ou BGLF4, le tout en présence du vecteur rapporteur luciférase.

Dans le cas de BTRF1, le ratio de l'activité luciférase obtenu lors des expériences n'a pas été suffisant pour conclure un éventuel rôle de cette protéine dans l'expression des gènes viraux tardifs. Ces expériences devront donc être répétées afin d'obtenir une réponse plus claire. Pour compléter ce petit test fonctionnel, un GST pull-down *in vivo* a été réalisé. Un vecteur d'expression codant pour la protéine 3xFlag-BTRF1 a été transfecté dans des cellules HEK293 avec un vecteur d'expression codant pour une protéine chimérique Flag-GST ou Flag-GST-X, X étant une protéine du vPIC. Quarante-huit post-transfection, les cellules ont été récoltées, lysées et les extraits ont été incubés sur billes Glutathione-Sépharose. Un GST pull-down a été ainsi réalisé avant de révéler les interactions par Western-Blot et un film autoradiographique (cf. figure 57).

La protéine BTRF1, d'une taille théorique approximative de 44 KDa, est efficacement détectée et exprimée dans cette lignée cellulaire (cf. figure 57, panel de gauche). Malgré l'expression efficace de la majorité des composants du vPIC, la protéine BTRF1 n'est pas retenue sur les différentes colonnes (cf. figure 57, panel de droite). Après purification, la

majorité des protéines chimériques Flag-GST-X sont détectées correctement hormis GST-BVLF1 et GST-BDLF4. En effet, ces deux dernières protéines n'expriment pas le tag Flag d'où l'absence de détection après une incubation avec un anticorps anti-Flag (cf. figure 57, panel de droite).

L'absence d'interaction avec les membres du vPIC *in vivo*, et les résultats publiés en 2012 selon lesquels l'orthologue de BTRF1 chez MHV68 n'aurait pas de rôle dans l'expression des gènes viraux tardifs suggèrent fortement que BTRF1 n'est pas impliqué dans ce mécanisme.

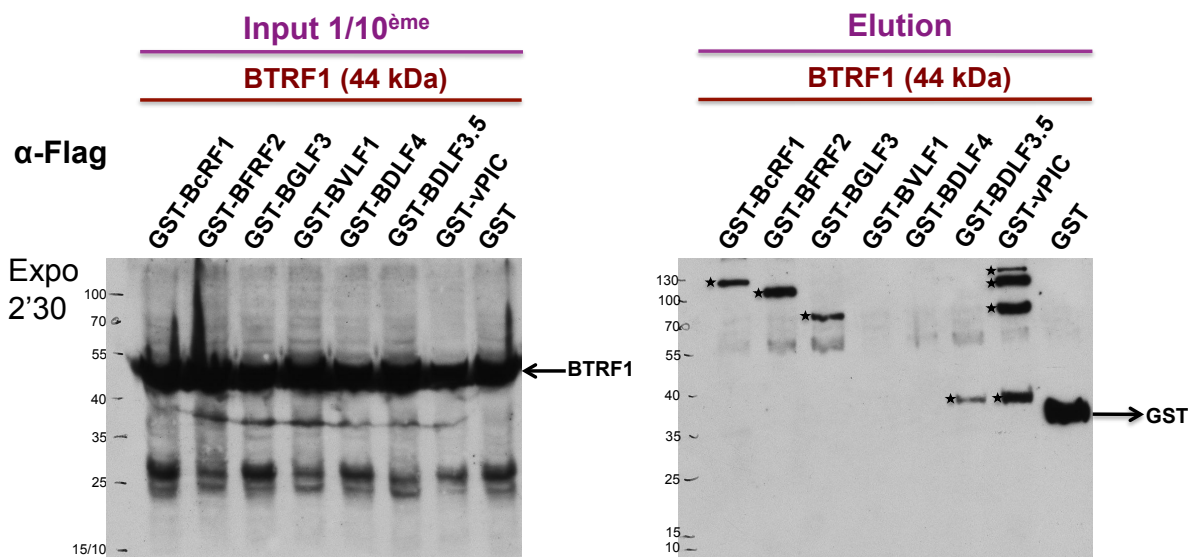


Figure 57: BTRF1 n'interagit pas avec les membres du vPIC *in vivo*.

Les cellules HEK293 ont été transfectées avec un vecteur d'expression codant pour la protéine BTRF1 avec les différents plasmides d'expression codants pour les différentes protéines du vPIC couplées à la GST. Les cellules sont ensuite récoltées et lysées. L'input, correspondant à 10% du lysat cellulaire est présenté sur le panel de gauche. Les 90% du lysat restant ont été incubés sur des billes Glutathione Sépharose. Les protéines retenues sur la colonne ont été lavées, éluées et séparées par SDS-PAGE avant d'être transférées sur membrane de nitrocellulose (panel de droite). Les différentes protéines sont détectées par Western-Blot à l'aide d'un anticorps anti-Flag avant d'être révélées sur film autoradiographique. Les astérisques mettent en évidence les protéines du vPIC couplées à la GST.

Concernant la sérine/thréonine kinase virale BGLF4, les résultats obtenus sont plus complexes. En effet, pour démontrer le rôle de BGLF4 dans l'expression des gènes viraux tardifs, un système rapporteur luciférase dont le gène est sous contrôle du promoteur tardif de BcLF1 a été utilisé.

Le plasmide pTATT_{BcLF1}-Luc a été transfecté dans des cellules HEK293 en présence du vPIC avec ou sans le vecteur d'expression codant pour la protéine BGLF4 en quantité

décroissante (de 1 µg à 10 ng). L'activité luciférase a été analysée 48h post-transfection et normalisée par rapport à l'activité luciférase obtenue lorsque le pTATT_{BGLF1}-Luc est transfecté seul.

Comme attendu, le vPIC est capable d'activer seul, l'expression de la luciférase comme déjà démontré (cf. figure 58). Cette activité luciférase demeure basale lorsque BGLF4 est exprimée uniquement avec le vecteur rapporteur (cf. figure 58). En revanche, lorsqu'une quantité croissante de BGLF4 est exprimée en présence du vPIC, il semble que l'activité luciférase diminue à mesure que cette quantité de plasmide codant pour la protéine BGLF4 croît. Cette observation n'est pas liée à une différence d'expression des protéines du vPIC dans les différentes conditions (datas non montrées). Ces résultats *in vitro* suggèrent que la protéine BGLF4 n'a pas de rôle dans l'expression des gènes viraux tardifs.

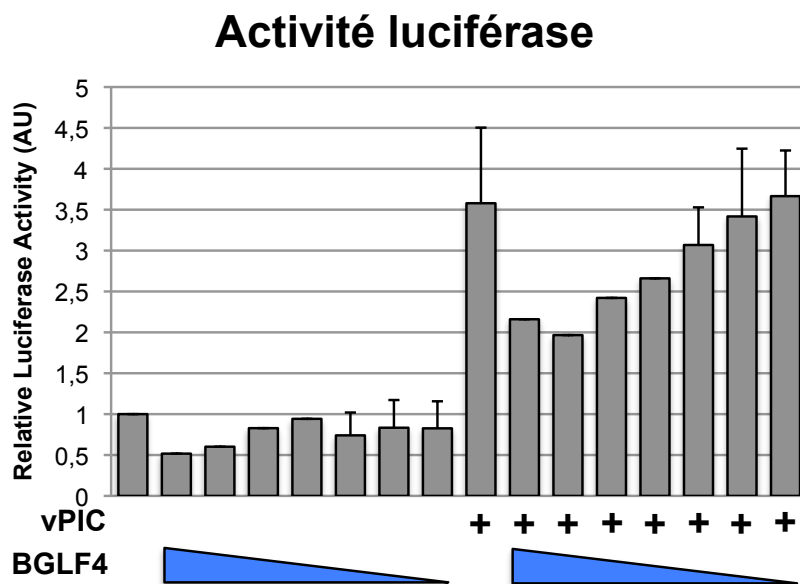


Figure 58 : BGLF4 n'est pas nécessaire à l'expression des gènes viraux tardifs *in vitro*.

Les cellules HEK293 ont été transfectées en quantité décroissante avec un vecteur d'expression codant pour la protéine BGLF4 en absence ou en présence des membres du vPIC. L'activité luciférase a été mesurée 48h post-transfection. L'activation relative de l'activité

luciférase a été normalisée avec l'activité luciférase obtenue lorsque le plasmide pTATT-Luc est transfecté seul.

Nous avons ensuite voulu savoir si la protéine BGLF4 était capable d'interagir avec les membres du vPIC. Pour réaliser ceci, un vecteur d'expression codant pour la protéine 3xFlag-BGLF4 a été transfecté dans des cellules HEK293 avec un vecteur d'expression codant pour une protéine chimérique Flag-GST ou Flag-GST-X, X étant une protéine du vPIC. Quarante-huit post-transfection, les cellules ont été récoltées, lysées et les extraits ont été incubés sur billes Glutathione-Sépharose. Un GST pull-down a été ensuite réalisé avant de détecter les interactions par Western-Blot à l'aide d'un anticorps-anti-Flag et révélé avec un film autoradiographique (cf. figure 59).

La protéine BGLF4, d'une taille théorique approximative de 48 KDa, est efficacement détectée dans cette lignée cellulaire (cf. figure 59, panel de gauche). La protéine BGLF4 ne semble pas interagir avec les différents composants du vPIC (cf. figure 59, panel de droite). Après purification, les différentes protéines chimériques Flag-GST-X sont détectées correctement dans toutes les conditions hormis pour GST-BVLF1 et GST-BDLF4. En effet, ces deux dernières protéines ne sont pas « flaguées » au contraire des quatre autres protéines du vPIC. Ainsi, l'absence d'interaction entre BGLF4 et le vPIC n'est pas liée à un problème d'expression des protéines virales couplées à la GST.

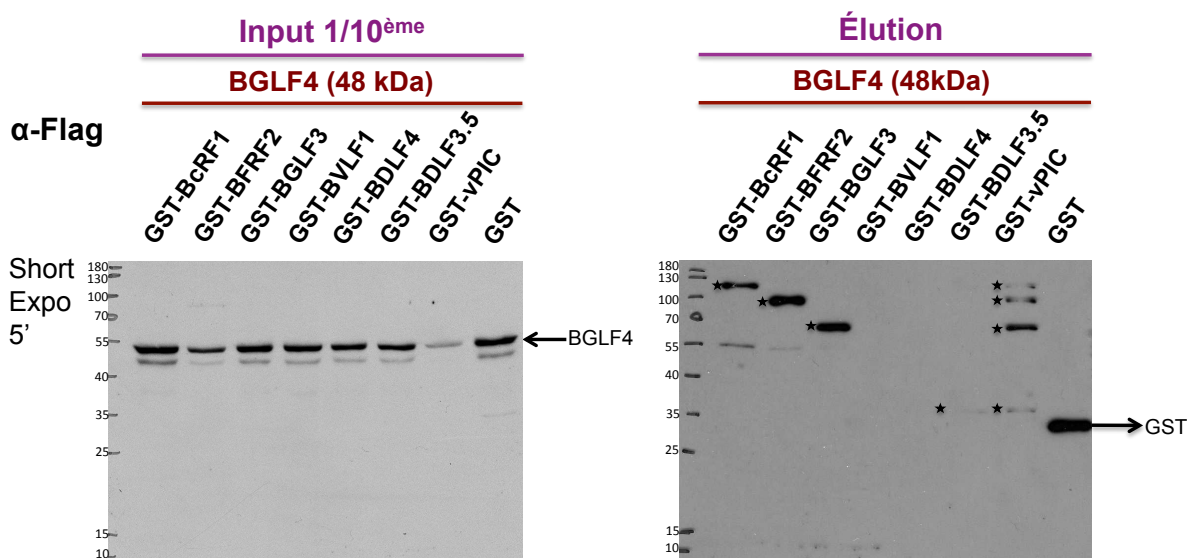


Figure 59: BGLF4 interagit faiblement avec les membres du vPIC *in vivo*.

Les cellules HEK293 ont été transfectées avec un vecteur d'expression codant pour la protéine BGLF4 avec les différents vecteurs d'expression codants pour les protéines du vPIC couplées à la GST. Les cellules sont ensuite récoltées et lysées. L'input, correspondant à 10% du lysat cellulaire est présenté sur le panel de gauche. Les 90% du lysat restant ont été incubés sur des billes Glutathione Sépharose. Les protéines retenues sur la colonne ont été lavées, éluées et séparées par SDS-PAGE avant d'être transférées sur membrane de nitrocellulose (panel de droite). Les différentes protéines sont détectées par Western-Blot à l'aide d'un anticorps anti-Flag avant d'être révélées sur film autoradiographique. Les astérisques mettent en évidence les protéines du vPIC couplées à la GST (appâts).

Ainsi, ces expériences suggèrent que BGLF4 ne semble pas avoir de rôle majeur dans l'expression des gènes viraux tardifs avec notre système, contrairement aux travaux publiés par l'équipe du Dr Miller, (El-Guindy et al., 2014). Toutefois rien n'empêche l'idée que BGLF4 puisse avoir un rôle au niveau post-transcriptionnel sur les protéines du vPIC.

2) Identification des partenaires potentiels du vPIC en condition virale.

Afin d'identifier de réels partenaires du vPIC, une technique de double purification par chromatographie d'affinité (TAP-Tag) fut envisagée. Pour ce faire, le cadre entier de lecture du gène codant pour la protéine BcRF1 a été inséré par ligation dans le vecteur commercial pNATP (Agilent) afin de générer une protéine recombinante possédant à son extrémité N-Terminale deux étiquettes : Calmoduline (Calm) et Streptavidine (Strep). Ainsi, une protéine chimérique Calm-Strep-3xFlag-BcRF1 peut s'exprimer dans les cellules (cf. figure 60). Le plasmide codant pour cette protéine a été transfecté dans la lignée cellulaire HEK293_{EBVΔBcRF1} déjà établie (Gruffat et al., 2012) avec le vecteur d'expression codant pour la protéine EB1 afin d'induire le cycle viral productif. En guise de contrôle négatif, un vecteur d'expression codant pour la protéine EB1 pour induire le cycle viral productif et un plasmide codant pour la protéine 3xFlag-BcRF1 ont été transfectés dans les cellules HEK293_{EBVΔBcRF1}.

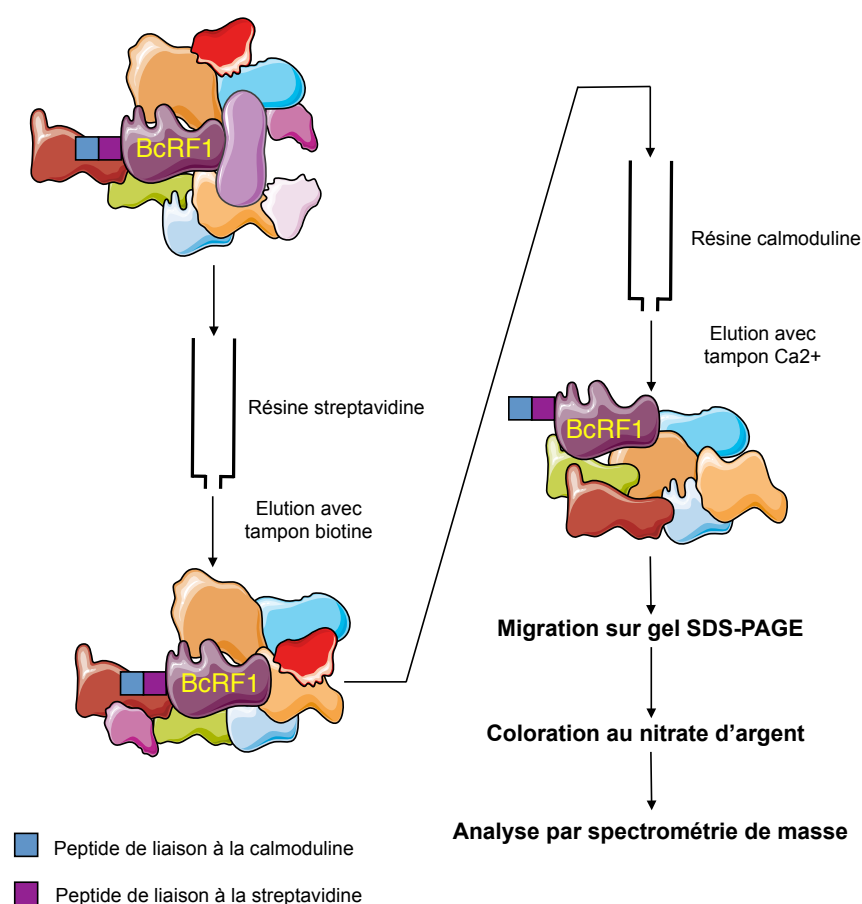


Figure 60 : Schéma représentant la double purification par chromatographie d'affinité

Le lysat cellulaire extrait des cellules HEK293_{EBVΔBcRF1} est incubé sur une résine Streptavidine. Après plusieurs lavages, les protéines sont éluées à l'aide d'un tampon à base de biotine. L'éluat est à nouveau incubé mais sur une colonne Calmoduline. Après différents lavages, les protéines encore associées à la protéine Calm-Strep-3xFlag-BcRF1 sont éluées à l'aide d'un tampon à base de

calcium. Les protéines sont alors séparées sur gel SDS-PAGE. Ce dernier sera ensuite coloré au nitrate d'argent, les bandes sélectionnées découpées, les protéines seront trypsinées et enfin analysées par spectrométrie de masse.

Des expériences préliminaires ont révélé par qPCR que la quantité d'ADN viral répliqué atteignait un plateau soixante heures post-transfection. Nous avons donc estimé que ce temps était le point où un maximum de transcrits viraux tardifs pouvaient être synthétiser. Par conséquent, il est possible que les protéines du vPIC soient exprimées au mieux à ce moment précis. Les cellules ont donc été récoltées soixante heures post-transfection, lysées et incubées sur billes Streptavidine dont le protocole est résumé schématiquement sur la figure 60.

Après une incubation à 4°C pendant la nuit, les protéines accrochées à l'appât retenu sur la colonne Strep sont lavées, purifiées puis éluées à l'aide d'un tampon à base de biotine (cf. figure 60). Les protéines éluées sont alors ré-incubées sur billes Calmoduline. Le peptide de liaison à la Calmoduline que possède BcRF1 va lui permettre de s'accrocher sur les billes et maintenir les protéines qui lui sont associées sur la colonne. Après plusieurs lavages, les différentes protéines sont éluées à partir d'un tampon riche en calcium. En théorie, cette double purification permet d'obtenir un appât enrichi en protéines capables d'interagir avec celui-ci. L'ensemble de l'éluât est déposé sur gel SDS-PAGE, et est concentré dans le stacking. Le gel est ensuite coloré au nitrate d'argent afin de détecter les différentes bandes protéiques. Ces dernières sont découpées, décolorées et enfin les protéines sont digérées à la trypsine et analysées par spectrométrie de masse.

Ainsi, l'analyse de spectrométrie de masse a révélé quelques éléments intéressants mais somme toute insuffisants (cf. tableau 12).

Accession	Coverage	# Peptides	MW [kDa]	Description
K9URF1	62,40	283	83,8	Uncharacterized protein OS=Epstein-Barr virus (strain GD1) GN= BcRF1 PE=4 SV=1 - [K9URF1_EBVG]
P24928	9,80	21	217,0	DNA-directed RNA polymerase II subunit RPB1 OS=Homo sapiens GN=POLR2A PE=1 SV=2 - [RPB1_HUMAN]
P03227	7,45	8	123,0	Major DNA-binding protein OS=Epstein-Barr virus (strain B95-8) GN= BALF2 PE=1 SV=1 - [DNBI_EBVB9]
P30876	6,64	11	133,8	DNA-directed RNA polymerase II subunit RPB2 OS=Homo sapiens GN=POLR2B PE=1 SV=1 - [RPB2_HUMAN]
P19387	10,18	3	31,4	DNA-directed RNA polymerase II subunit RPB3 OS=Homo sapiens GN=POLR2C PE=1 SV=2 - [RPB3_HUMAN]

Tableau 12 : Résultats des protéines identifiées par spectrométrie de masse.

Tout d'abord, la protéine d'appât BcRF1 sort en majorité, avec 283 peptides recouvrant plus de 62% de l'ORF (cf. tableau 12). Ceci semble logique puisqu'il s'agit de la protéine d'appât utilisée pour cette étude. Ensuite, avec 21 peptides, la plus grosse sous-unité

de l'ARN pol II (Rpb1) est détectée. Ces derniers recouvrent globalement 9,8% de la séquence codante de la protéine. Ce résultat est cohérent et corroborent nos données publiées en 2014 ainsi que les résultats de la figure 51. Deux autres sous-unités de l'ARN pol II : Rpb2 et Rpb3 sont détectées par spectrométrie de masse avec respectivement 11 et 3 peptides qui permettent de recouvrir plus de 6 et 10% de la séquence codante de ces protéines. Ainsi, ces données démontrent que la protéine BcRF1 est capable de recruter différentes sous-unités de l'ARN pol II cellulaire. Enfin, et ce de manière inattendue, huit peptides de la protéine BALF2 recouvrent environ 7,5% de sa séquence. Cette dernière est la protéine de liaison à l'ADN simple brin, impliquée dans la réplication de l'ADN viral. En revanche, aucune protéine du vPIC autre que BcRF1 n'est détectée. Cette double purification ne fut donc pas une franche réussite, c'est pourquoi elle devra être reproduite dans d'autres conditions.

L'interaction entre BcRF1 et BALF2 semble intéressante à creuser. Nous avons donc tenté une première approche, *in vitro*, afin de voir si les sept protéines du complexe de réplication étaient capables d'interagir avec le vPIC. Pour ce faire, les différents plasmides codant pour les protéines du complexe de réplication ont été transcrits puis traduits en lysat de réticulocytes en protéines marquées radioactivement au soufre 35. Parallèlement, des protéines chimériques du vPIC, couplées à la GST, ont été produites en bactéries et purifiées sur billes Glutathione Sépharose afin de servir d'appât.

Ces protéines ont été incubées sur des billes Glutathione Sépharose purifiées avec les différentes protéines chimériques du vPIC couplées à la GST, produites préalablement en bactéries. Après plusieurs lavages, élution et migration sur gel SDS-PAGE, les protéines sont transférées sur membranes de nitrocellulose et révélées à l'aide d'un film autoradiographique. Les résultats obtenus sont représentés sous forme de tableau (cf. tableau 13) pour simplifier la compréhension. Les couleurs utilisées sont les mêmes que celles pour les interactions *in vivo* réalisées précédemment. Toutefois, les figures correspondantes sont présentées en Annexe 3.

	BALF2	BALF5	BMRF1	BBLF2/3	BBLF4	BSLF1	EB1
GST							
GST-BcRF1							
GST-BFRF2							
GST-BGLF3							
GST-BDLF4							
GST-BDLF3.5							

Tableau 13 : Résultats des interactions par GST pull-down *in vitro* entre les protéines du vPIC et les membres du complexe de réplication.

Les résultats de GST pull-down démontrent que différentes protéines semblent interagir directement avec les membres du vPIC. Parmi celles-ci, BALF5, BMRF1, BBLF4 et BSLF1 sont retenues sur toutes les colonnes contenant les protéines chimériques du vPIC alors que le contrôle négatif GST demeure vierge (cf. tableau 13). À contrario, il semble que les protéines BALF2, BBLF2/3 et EB1 n'interagissent pas directement avec les différents membres du vPIC ou du moins que très faiblement. (cf. tableau 13).

Ces données sont surprenantes car la protéine BALF2 qui a été identifiée en spectrométrie de masse après double purification par TAP-tag (cf. tableau 12), n'interagit pas *in vitro* avec les différentes protéines du vPIC. Cela reviendrait à dire que soit, les résultats de spectrométrie de masse sont biaisés, soit que d'autre(s) protéine(s) virale(s) ou cellulaire(s), non détectée(s) en spectrométrie de masse, permettrai(en)t de recruter BALF2 sur BcRF1. En revanche, l'interaction probablement directe observée par GST pull-down *in vitro* entre les quatre autres protéines du complexe de réplication et les membres du vPIC permettrait d'expliquer la présence de BALF2 en spectrométrie de masse.

Pour vérifier cette hypothèse et afin de confirmer la reproductibilité des interactions obtenues *in vitro*, des GST pull-down *in vivo* ont été réalisés. Pour cela, l'intégralité du cadre de lecture des six différents gènes codant pour les protéines du complexe de réplication a été cloné dans des vecteurs pCi-3xFlag et pDEST27-GST. Pour commencer, l'ensemble des vecteurs d'expression codant pour les protéines du complexe de réplication de l'ADN viral a été transfectée dans des cellules HEK293 avec soit un vecteur d'expression codant pour la protéine GST soit un vecteur d'expression codant pour une protéine GST-X, X étant une des protéines du vPIC. Quarante huit heures post-transfection, les cellules ont été récoltées, lysées, et incubées sur billes Glutathione Sépharose. Après incubation, les différentes protéines retenues sur la colonne ont été lavées, éluées et séparées sur gel SDS-PAGE avant d'être transférées sur membrane de nitrocellulose pour être détectées à l'aide d'un anticorps anti-Flag par Western-Blot et révélées à l'aide d'un film autoradiographique.

Compte tenu de la faible différence de masse moléculaire des protéines, il est compliqué de bien différencier les composants du complexe de réplication. Toutefois, il est possible de distinguer dans les inputs une bande aux alentours de 130 KDa qui pourrait correspondre à BALF2 (cf. figure 61, panels de gauche). Juste en dessous de 100 KDa, une bande fine est visible et pourrait correspondre à celle de BSLF1 alors que la bande majoritaire

située juste en dessous pourrait être BBLF4 ou BBLF2/3 (de masse moléculaire approximatives respectives 90 KDa et 80 KDa). Enfin, la protéine BMRF1 est efficacement exprimée et détectée à 50 KDa (cf. figure 61, panels de gauche). Pour plus de clarté, le poids moléculaire théorique de chacune des protéines du complexe de réplication est répertorié dans le tableau ci-dessous :

	BALF2	BALF5	BMRF1	BBLF2/3	BBLF4	BSLF1
Masse Moléculaire Théorique	125 KDa	114 KDa	44 KDa	79 KDa	90 KDa	99 KDa

Tableau 14 : Masse moléculaire théorique attendue pour les protéines du complexe de réplication

Après trente secondes d'exposition, seule une bande, aux alentours de 95 KDa, apparaît assez nettement sur le Western dans toutes les conditions hormis avec GST-BVLF1 (cf. figure 61, panel en haut à droite). Cette bande pourrait correspondre à la protéine BBLF4. En revanche, sur un temps de détection long d'une heure (cf. figure 61, panel en bas à droite), plusieurs bandes sont mises en évidence. Parmi celles-ci, une bande aux alentours de 130 KDa, pourrait correspondre à la protéine BALF2. Une autre, juste en dessous, de 130 KDa est également détectable et pourrait être celle de BALF5. Plus intéressante, une bande migrant à peine au dessus de 100 KDa est détectée. Celle-ci pourrait correspondre à la protéine BSLF1. La bande apparente à 95 KDa serait donc bien BBLF4. À l'inverse, aucune bande spécifique à BBLF2/3 et BMRF1 ne semble être mise en lumière lors de cette expérience.

De plus, les protéines chimériques couplées à la GST du vPIC sont détectées uniquement après purification. Elles sont annotées sur les panels de droite avec des astérisques. Toutefois, aucune bande correspondante à GST-BFRF2 et GST-BVLF1 n'est détectée. Dans le premier cas, la protéine est bien exprimée car des protéines du complexe sont retenues mais ne peut être détectée car elle ne présente pas de tag-Flag en son extrémité N-Terminale. En revanche, il est impossible de conclure pour GST-BVLF1. Effectivement, soit la protéine n'est pas exprimée, soit les protéines du complexe de réplication n'interagissent pas avec.

Ces résultats mettent ainsi en évidence une interaction forte avec probablement les protéines BSLF1 et BBLF4 qui sont respectivement la primase et l'hélicase virale (confirmant ainsi les résultats obtenus *in vitro*) et une interaction faible, modérée avec BALF2 et BALF5 qui sont respectivement la protéine de liaison à l'ADN simple brin et l'ADN polymérase virale. Ces deux derniers facteurs sont peut-être recrutés par BSLF1 et BBLF4 pour être

détectés dans cette expérience. Cette manipulation permet aussi de démontrer que les résultats de spectrométrie de masse ne sont pas biaisés mais simplement que BALF2 est peut être associée au vPIC à l'aide des protéines virales du complexe de réplication.

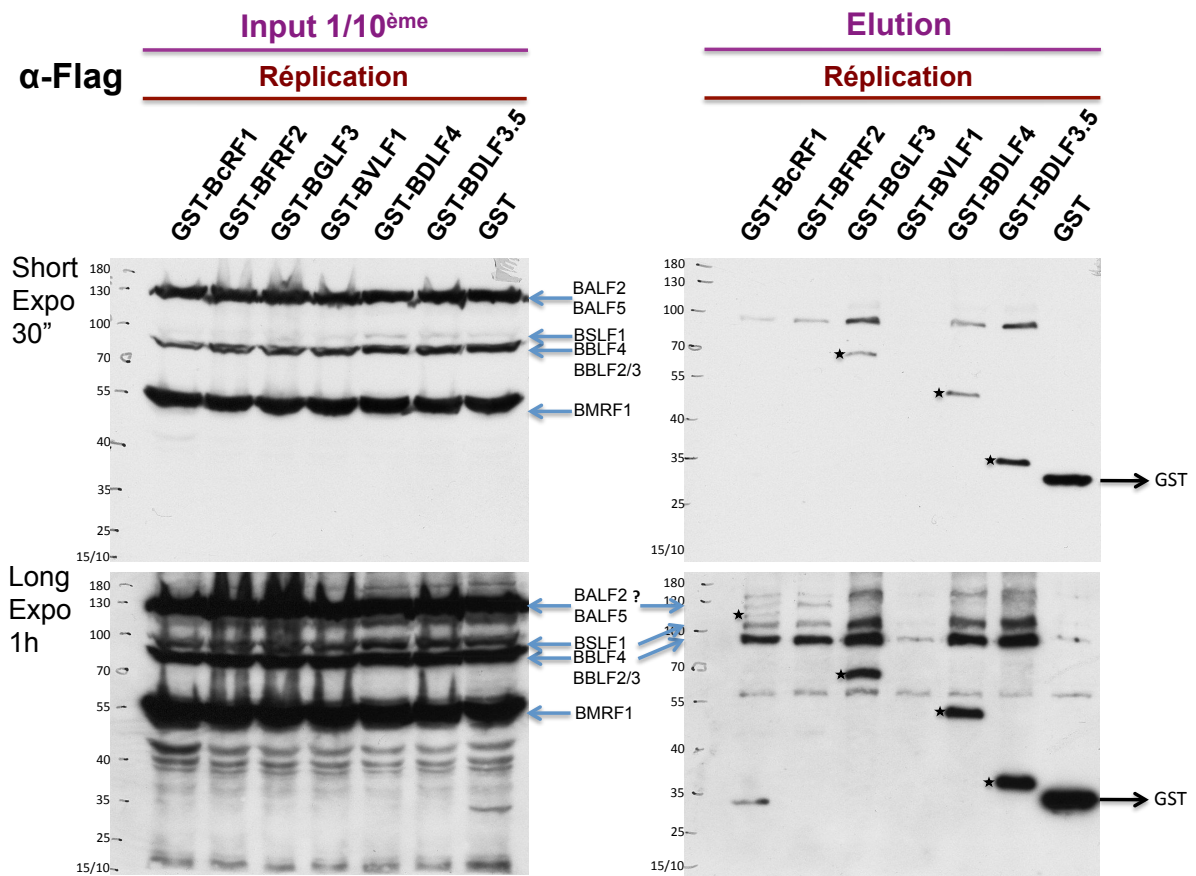


Figure 61 : BSLF1 et BBLF4 semblent interagir avec différents composants du vPIC.

Les cellules HEK293 ont été transfectées avec les vecteurs d'expression codant pour les différentes protéines du complexe de réplication de l'ADN viral avec un vecteur d'expression codant soit pour la protéine GST soit pour une protéine chimérique du vPIC couplée à la GST. Les cellules sont ensuite récoltées et lysées. L'input, correspondant à 10% du lysat cellulaire est présenté sur le panel de gauche. Les 90% du lysat restant ont été incubés sur des billes Glutathione Sépharose. Les protéines retenues sur la colonne ont été lavées, éluées et séparées par SDS-PAGE avant d'être transférées sur membrane de nitrocellulose (panel de droite). Les différentes protéines sont détectées par Western-Blot à l'aide d'un anticorps anti-Flag avant d'être révélées sur film autoradiographique. Les astérisques mettent en évidence les protéines du vPIC couplées à la GST.

Pour confirmer l'identité des protéines interagissant efficacement avec les différents membres du vPIC, des GST pull-down ont été réalisés dans un premier temps avec la protéine GST-BBLF4 en présence des différentes protéines du vPIC (cf. figure 62) et avec les différentes protéines du vPIC couplées à la GST avec la protéine 3xFlag-BSLF1 (cf. figure 63). Ces GST pull-down ont été réalisés dans les mêmes conditions que les autres présentés précédemment.

Pour commencer avec le GST pull-down de GST-BBLF4, toutes les protéines du vPIC sont efficacement détectées dans les inputs à l'exception de BDLF3.5 seule (cf. figure 62, panel de gauche). Après purification, différentes protéines du vPIC sont retenues sur la colonne GST-BBLF4 (cf. figure 62, panel de droite). Parmi ces différentes protéines, BcrRF1, BFRF2, BGLF3 et enfin BDLF4 semblent interagir avec la protéine GST-BBLF4. De manière intéressante, quand le vPIC est reconstitué, seule BGLF3 et dans une moindre mesure BFRF2 semblent rester associées à GST-BBLF4 (cf. figure 62, panel de droite). Malgré l'absence de détection de la protéine GST-BBLF4, celle-ci doit être exprimée car certaines protéines du vPIC interagissent avec (cf. figure 62, panel de droite).

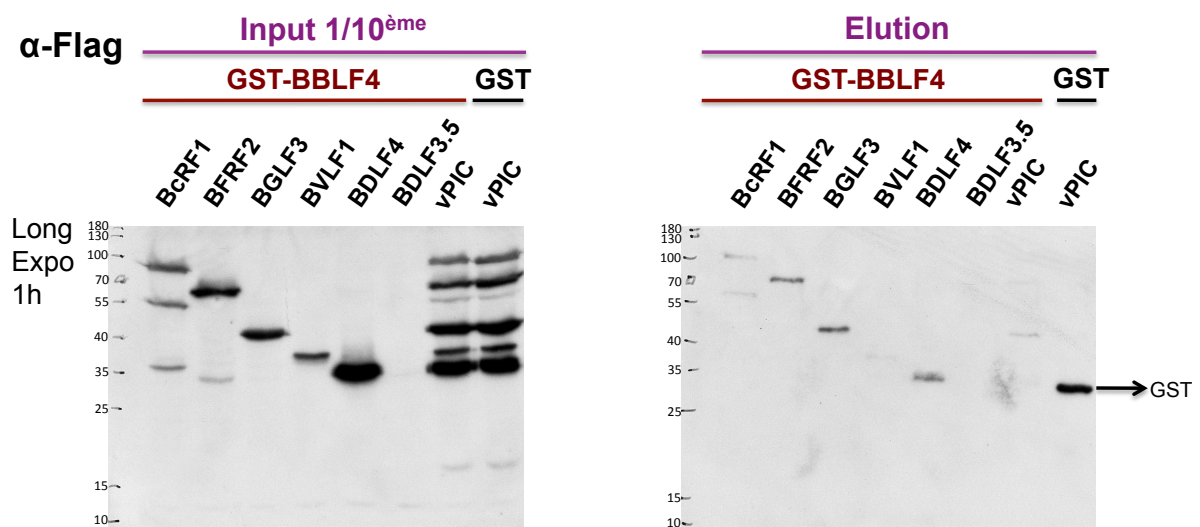


Figure 62 : GST-BBLF4 interagit avec plusieurs composants du vPIC.

Les cellules HEK293 ont été transfectées avec les vecteurs d'expression codant pour les différentes protéines du vPIC avec un vecteur d'expression codant soit pour la protéine GST soit pour la protéine chimérique GST-BBLF4. Les cellules sont ensuite récoltées et lysées. L'input, correspondant à 10% du lysat cellulaire est présenté sur le panel de gauche. Les 90% du lysat restant ont été incubés sur des billes Glutathione Sépharose. Les protéines retenues sur la colonne ont été lavées, éluées et séparées par SDS-PAGE avant d'être transférées sur membrane de nitrocellulose (panel de droite). Les différentes protéines sont détectées par Western-Blot à l'aide d'un anticorps anti-Flag avant d'être révélées sur film autoradiographique.

Enfin, pour observer l'interaction entre la protéine BSLF1 et les différentes protéines du vPIC, les cellules HEK293 ont été transfectées avec un vecteur d'expression codant pour la protéine BSLF1 avec un vecteur d'expression codant pour une protéine chimérique GST-X, X étant une protéine du vPIC. Un GST pull-down comme décrit dans la partie Matériel et méthodes a ensuite été réalisé. La protéine BSLF1 est bien détectée et exprimée dans les cellules HEK293 (cf. figure 63, panel de gauche). Comme attendue, la protéine BSLF1

n'interagit pas avec la protéine GST (cf. figure 63, panel de droite). En revanche, elle est retenue sans ubiquité après purification sur les colonnes GST-BFRF2, GST-BGLF3 et GST-BDLF3.5 (cf. figure 63, panel de droite). Malheureusement, il est difficile d'analyser le résultat avec GST-BcRF1 et GST-vPIC. En effet, la protéine GST-BcRF1 a une masse moléculaire similaire à celle de BSLF1. De plus, elles sont toutes deux révélées à l'aide d'un anticorps dirigé contre le peptide Flag, présent en leur extrémité N-Ter.

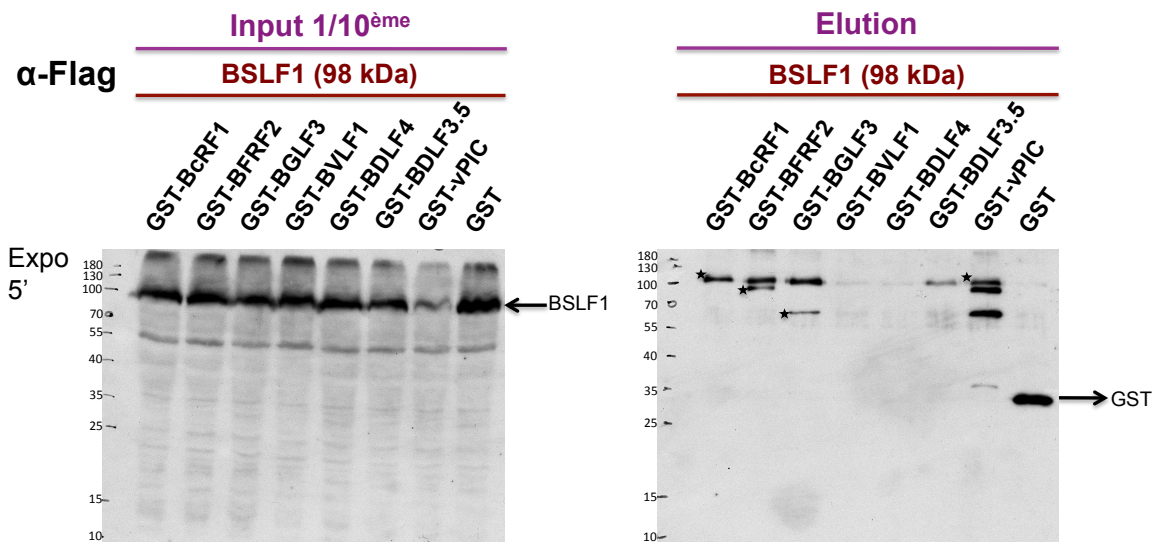


Figure 63 : BSLF1 interagit avec plusieurs composants du vPIC.

Les cellules HEK293 ont été transfectées avec le vecteur d'expression codant pour la protéine BSLF1 avec le vecteur d'expression codant soit pour la protéine GST soit pour une des protéines chimériques du vPIC couplée à la GST. Les cellules sont ensuite récoltées et lysées. L'input, correspondant à 10% du lysat cellulaire est présenté sur le panel de gauche. Les 90% du lysat restant ont été incubés sur des billes Glutathione Sépharose. Les protéines retenues sur la colonne ont été lavées, éluées et séparées par SDS-PAGE avant d'être transférées sur membrane de nitrocellulose (panel de droite). Les différentes protéines sont détectées par Western-Blot à l'aide d'un anticorps anti-Flag avant d'être révélées sur film autoradiographique. Les astérisques mettent en évidence les protéines du vPIC couplées à la GST.

Aucune interaction entre BSLF1 et GST-BVLF1 ou GST-BDLF4 n'est visible. Ceci est probablement dû à un défaut d'expression des protéines du vPIC couplées à la GST. Une expérience inverse avec GST-BSLF1 devra être réalisée afin de compléter et confirmer ces interactions *in vivo*.

Ainsi ces deux dernières expériences démontrent que certaines protéines du vPIC, notamment BcRF1, BFRF2, BGLF3, BDLF4 et BDLF3.5 pourraient interagir avec l'hélicase (BBLF4) et la primase virale (BSLF1) (cf. figures 62 et 63). De plus, d'autres GST pull-down

in vivo, réalisés avec les protéines du vPIC couplées à la GST en présence de BZLF1 ou de BMRF1 ne révèlent aucune interaction (données non montrées).

La découverte de l'interaction entre les protéines BBLF4-BSLF1 (hélicase-primase) et celles du vPIC, met en lumière un lien concret entre la réplication de l'ADN viral et l'expression des gènes tardifs du virus d'Epstein-Barr. Toutefois, des expériences complémentaires devront être réalisées pour confirmer ces interactions mais aussi pour comprendre la liaison entre les deux complexes viraux, d'autant plus qu'il est connu que l'expression des gènes tardifs d'EBV est dépendante de la réplication de l'ADN viral...

PARTIE III :

Conclusions & Discussions

I) La découverte d'un complexe de pré-initiation de la transcription des gènes viraux tardifs, conservé entre β - et γ -Herpesvirus.

Ainsi, l'objectif central de cette thèse était d'étudier et mieux comprendre la régulation de l'expression des gènes tardifs du virus d'Epstein-Barr. Au début de la thèse, les mécanismes impliqués dans la régulation de l'expression des gènes tardifs étaient peu connus. Tout d'abord, l'expression de ces gènes est dépendante de la réplication car l'absence de réplication empêche l'expression des gènes tardifs. Ensuite, la structure des promoteurs des gènes tardifs est différente des promoteurs des gènes viraux immédiats précoces ou cellulaires. En effet, seule une boîte TATT localisée environ trente nucléotides en amont du TSS est nécessaire à l'expression des gènes viraux tardifs. Enfin, il a été démontré au laboratoire que la protéine virale BcRF1 est une TBP-like se fixant exclusivement sur la séquence TATT en contexte infectieux. Toutefois, si cette protéine est nécessaire à l'expression des gènes viraux tardifs, elle n'est pas suffisante. Ceci suggère que d'autres facteurs viraux sont nécessaires à ce mécanisme. De manière intéressante, cinq protéines virales ont été démontrées chez MHV68 et trois chez HCMV comme essentielles à l'expression des gènes tardifs de ces virus sans toutefois identifier leurs fonctions. Parmi celles-ci, des homologues de BcRF1 sont retrouvés.

Nous avons constaté que ces cinq protéines avaient des homologues uniquement chez les β - et γ -Herpesvirus mais pas chez les α -Herpesvirus. Nous avons alors émis l'hypothèse que seules les protéines conservées entre β - et γ -Herpesvirus et absentes chez les Alpha, pouvaient être impliquées dans l'expression des gènes viraux tardifs. Après une analyse *in silico*, seules sept protéines correspondaient à ces critères. Outre les cinq protéines déjà caractérisées chez MHV68 comme impliquées dans l'expression des gènes viraux tardifs, deux furent identifiées : BTRF1 et BFRF2 chez EBV. L'homologue de la première a déjà été démontré comme non essentiel à l'expression des gènes tardifs chez MHV68. En revanche, aucune étude ne s'était focalisée sur BFRF2.

Nous avons donc tout d'abord démontré comme espéré que cette protéine (BFRF2) est bien impliquée dans l'expression des gènes viraux tardifs, sans toutefois connaître sa fonction (Aubry et al., 2014). Nous avons également démontré qu'un ensemble de six protéines virales était nécessaire et suffisant à l'expression des gènes viraux tardifs. Ce complexe, composé des

protéines virales BcRF1 (TBP-like), BFRF2, BGLF3, BVLF1, BDLF4 et BDLF3.5 forme une structure qui est associée à l'ARN polymérase II cellulaire. Cette interaction est probablement réalisée grâce à la TBP-like virale BcRF1. Effectivement, cette enzyme cellulaire n'est retenue (par GST pull-down *in vivo*) que sur une colonne GST-BcRF1.

Par homologie avec le complexe de pré-initiation de la transcription cellulaire (PIC), nous avons décidé de nommer le complexe d'EBV vPIC (viral pre-initiation complex of transcription). En effet, une TBP-like, un ensemble de cinq autres protéines virales et l'ARN polymérase II cellulaire forment ce complexe (cf. figure 64)

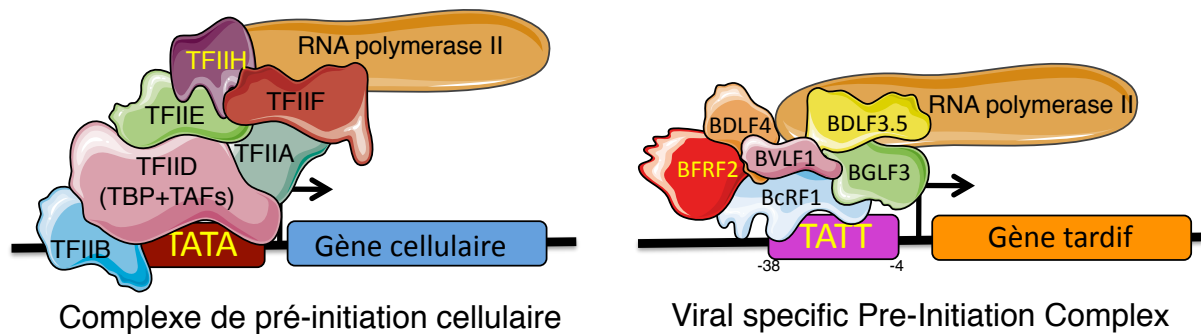


Figure 64 : Représentation schématique de l'homologie entre les deux complexes de pré-initiation de la transcription, l'un viral et l'autre cellulaire.

Le complexe de pré-initiation de la transcription cellulaire représenté à gauche est constitué de plusieurs facteurs cellulaires appelés facteurs généraux de la transcription dont la protéine TBP qui se fixe sur la boîte TATA et l'ARN polymérase II qui réalise la synthèse des ARNm. À droite est représenté le complexe de pré-initiation de la transcription des gènes tardifs du virus d'Epstein-Barr. Celui-ci est composé d'une TBP-like se fixant sur une boîte TATT qui sert de promoteurs aux gènes tardifs, de cinq autres protéines virales (BFRF2, BGLF3, BVLF1, BDLF4 et BDLF3.5), et de l'ARN polymérase II.

Nous avons également démontré que l'activité fonctionnelle du vPIC est dépendante de la réplication de l'ADN viral (Aubry et al., 2014). En effet, la transfection de cellules Raji (défectueuse dans la réplication de l'ADN) ou HEK293_{EBV} avec le vecteur rapporteur TATT-Luc et les plasmides d'expression codants pour les protéines du vPIC (sans induction du cycle productif) ne permet pas d'induire l'expression endogène des gènes viraux tardifs. À contrario, l'activité luciférase sous contrôle du promoteur de gène tardif BcLF1 est elle détectée. De plus des cellules HEK293_{EBV} dont le cycle viral productif a été induit, transfectées ou non avec les vecteurs d'expression codant pour les protéines du vPIC et traitées au PAA ne sont pas capables d'exprimer les gènes viraux tardifs. Toutes ces expériences démontrent ainsi que le vPIC nécessite l'étape de réplication de l'ADN viral pour transcrire les ARNm tardifs.

De plus, nous avons prouvé que les six protéines homologues de HCMV sont capables d'induire l'activité luciférase à partir du vecteur rapporteur TATT-Luc, de la même intensité que le vPIC d'EBV. Ce résultat tend à démontrer une conservation fonctionnelle de ce complexe entre EBV et HCMV. Néanmoins, une protéine du vPIC de HCMV, par exemple UL87, ne peut pas être substituée par son homologue d'EBV (BcRF1), pour induire l'activité luciférase sous contrôle d'un promoteur de gène tardif. De manière intéressante, des cellules HEK293_{EBVΔBcRF1} transfectées avec un vecteur d'expression codant pour la protéine EB1 (qui induit le cycle viral productif) avec les plasmides d'expression codant pour les différentes protéines du vPIC du Cytomégalo virus Humain sont capables d'exprimer les ARNm tardifs d'EBV de manière aussi efficace que le vPIC d'EBV. Ainsi, nous démontrons pour la première fois que six protéines virales, conservées entre EBV et HCMV et probablement entre tous les β - et γ -Herpesvirus sont nécessaires à l'expression des gènes tardifs. De plus, le fait que le vPIC de HCMV soit capable d'activer l'expression des gènes tardifs endogènes du virus d'Epstein-Barr de la même manière que le vPIC d'EBV signifie que les six protéines virales sont suffisantes à l'expression des gènes viraux tardifs. Ces données suggèrent fortement une co-évolution fonctionnelle du vPIC entre les β - et γ -Herpesvirus.

Ainsi, l'ensemble de ces résultats a permis d'avancer dans la compréhension des mécanismes régulant l'expression des gènes tardifs du virus d'Epstein-Barr et plus globalement des β - et γ -Herpesvirus. En effet, l'expression des gènes tardifs d'EBV est dépendante de la réplication de l'ADN viral, d'une séquence cis-régulatrice TATT sur laquelle se fixe différents facteurs trans-régulateurs dont BcRF1, en association avec un ensemble de cinq autres protéines virales, ainsi qu'avec l'ARN pol II cellulaire formant ainsi le vPIC.

Ce complexe est probablement conservé chez tous les β - et γ -Herpesvirus mais n'est pas retrouvé chez les α -Herpesvirus. Il est donc possible d'imaginer une divergence assez récente entre les deux premières sous-familles virales. En effet, α -, β - et γ -Herpesvirus auraient pu avoir un ancêtre commun possédant un mécanisme de régulation pour les gènes tardifs basé sur celui des Beta- et Gammaherpesvirus. Les Alpha-Herpesvirus auraient ensuite divergés et adaptés un système proche de celui des gènes cellulaires alors que les Beta et Gamma l'auraient conservés.

À l'inverse, α -, β - et γ -Herpesvirus auraient pu avoir un ancêtre commun possédant des mécanismes de régulation similaires à ceux des TATA-box avant que les Beta- et

Gammaherpesvirus ne co-divergent pour acquérir un système indépendant basé sur une boîte TATT et des facteurs viraux trans-régulateurs. Néanmoins, à partir de quand cette divergence aurait pu intervenir et surtout de quel organisme, système, proviennent ces facteurs protéiques sachant que ceux-ci n'ont aucune homologie structurelle connue mise à part BcRF1, structurellement proche d'une TBP-d'archae... ? Ces questions d'évolution démontrent la complexité de l'origine de ces virus.

II) Caractérisation du rôle de BVLf1, BDLf4 et BDLf3.5 dans l'expression des gènes viraux tardifs.

Suite la publication sur BFRF2 et le rôle des six protéines virales nécessaires à l'expression des gènes tardifs, nous avons continué nos investigations par rapport à ce complexe. Tout d'abord, nous avons confirmé, à l'instar des homologues de HCMV et MHV-68 que BVLf1, BDLf4 et BDLf3.5 sont absolument nécessaires à l'expression des gènes viraux tardifs. Néanmoins, il reste encore à caractériser la protéine BGLf3.

Depuis la publication et même le début de cette thèse, de nombreux groupes ont publiés des données similaires. Par exemple, les homologues de BDLf4 et BDLf3.5 chez HCMV (respectivement UL92 et UL91) et de BVLf1 et BDLf4 chez le CMV murin (respectivement pM79 et pM92), ont été démontrés comme impliqués dans l'expression des gènes tardifs (Chapa et al., 2013, 2014; Omoto and Mocarski, 2013, 2014). Ceci fut également démontré pour les protéines ORF18, 24, 30, 31 et 34 de KSHV (Brulois et al., 2015; Davis et al., 2015a; Gong et al., 2014). Par ailleurs, l'équipe de Murata a réussi à publier le rôle de BDLf4 chez EBV sans identifier sa fonction (Watanabe et al., 2015). De plus, l'équipe de Glaunsinger a démontré par spectrométrie de masse que la protéine ORF24 de KSHV (qui est l'homologue de BcRF1) interagissait spécifiquement avec dix des treize sous-unités protéiques qui constituent l'ARN polymérase II cellulaire (Davis et al., 2015b), confirmant ainsi nos résultats publiés.

Ainsi, l'ensemble de ces données et publications mettent en évidence un rôle conservé de chacune des protéines formant le vPIC et impliquées dans l'expression des gènes tardifs de probablement tous les β et γ -Herpesvirus. En plus de la conservation des facteurs trans-

régulateurs, différentes études ont démontré que le promoteur TATT des gènes tardifs est retrouvé chez plusieurs β - et γ -Herpesvirus et est nécessaire à l'expression de cette catégorie de gènes (Isomura et al., 2008; McWatters et al., 2002; Tang et al., 2004; Wong-Ho et al., 2014), donnant encore plus de crédit à cette co-évolution entre β - et γ -Herpesvirus.

Cependant, aucune fonction n'a, pour l'instant été associée aux différentes protéines du vPIC, hormis pour BcRF1 (Gruffat et al., 2012) et peut être l'homologue de BVLF1 chez HCMV : UL79 (Perng et al., 2014). En effet, concernant la protéine UL79, il a été démontré que cette protéine interagissait avec plusieurs sous-unités de l'ARN pol II (dont Rpb1 et Rpb2). Ces résultats divergent de ceux que nous avons obtenus au laboratoire, élément sur lequel nous reviendrons ultérieurement.

In fine, UL79 est associée à la sous-unité Rpb1 (du CTD de la RNAP II), phosphorylée sur la Sérine 2, suggérant un rôle dans l'élongation. Toutefois, les auteurs de cette étude ont démontré que UL79 n'était pas capable de recruter l'ARN pol II sur les promoteurs tardifs (Perng et al., 2014). Si UL79 n'est pas nécessaire au recrutement de l'ARN pol II, l'article prouve avec une expérience de « nuclear run-on » que l'absence de UL79 mène à une diminution sévère dans l'étape d'élongation de la transcription des gènes viraux tardifs (Perng et al., 2014). En revanche, l'élongation des transcrits viraux précoces et cellulaires n'est que très faiblement impactée.

Les résultats concernant UL79 sont intrigants car ils suggèrent un rôle de cette protéine dans l'élongation et non l'initiation. La liaison de UL79 avec la sous-unité Rpb1, phosphorylée sur la sérine 2 qui est une marque souvent associée à l'élongation (Bataille et al., 2012), ainsi que le défaut d'élongation des transcrits tardifs sont des arguments allant dans ce sens. Il serait intéressant à l'avenir de tester par CHIP la présence de BVLF1 ou de ses homologues sur l'ADN au niveau des gènes tardifs pour vérifier la répartition de la protéine au cours de la transcription.

D'autres travaux seront donc nécessaires pour identifier le rôle des quatre autres protéines du vPIC. Par exemple, il est fort probable que l'une d'elles confère la spécificité de liaison de BcRF1 à la TATT-box. En effet, BcRF1 est capable de se fixer sur les séquences TATA canoniques et TATT sans distinction. En revanche, en contexte infectieux, BcRF1 est retrouvée uniquement sur les boîtes TATT, suggérant qu'une protéine virale est nécessaire à cette sélection (Gruffat et al., 2012). Pour répondre à cette énigme, des expériences d'EMSA (retard sur gel) seraient à réaliser en utilisant une sonde TATA ou TATT en présence de

BcRF1 plus l'une des protéines du vPIC. Si l'une d'elle est impliquée dans la spécificité de liaison de BcRF1 à la boîte TATT, alors aucun retard de migration sur gel ne sera observable avec la sonde comportant la TATA-box.

III) Interactome du vPIC.

Après avoir caractérisé un complexe viral impliqué dans l'initiation de la transcription des gènes viraux tardifs, nous avons voulu identifier les interactions au sein de ce complexe. Pour ce faire, deux approches à partir de GST pull-down *in vivo* ont été réalisées. La première consiste à observer les interactions protéine par protéine. La seconde est de reconstituer tout le vPIC et à chaque fois « en oublier une » pour examiner d'éventuelles différences d'interaction lorsque le vPIC est présent en entier. Nous avons ainsi pu constater que presque toutes les protéines du vPIC interagissaient entre elles mais de manières différentes. De plus, la reconstitution du vPIC « en oubliant » une des protéines a permis de mettre en évidence que le vPIC ne serait pas qu'un complexe mais, a priori, divisé au minimum en deux parties. En effet, nos résultats semblent suggérer que la présence de BDLF4 dans le complexe engendre un défaut de recrutement de BVLF1 et BDLF3.5 et ce, réciproquement. Ces résultats permettent ainsi d'établir un modèle d'assemblage théorique du vPIC, évoqué plus tôt (cf. figure 56). Cependant, l'absence de résultats avec GST-BFRF2 et GST-BVLF1 en présence de seulement quatre des cinq autres protéines du vPIC est préjudiciable pour valider ce modèle.

En tout cas, ce résultat pourrait aller dans le sens de la publication de 2014 (Perng et al., 2014) qui propose que l'homologue de BVLF1 chez HMCV (UL79) est un facteur d'élongation, comme évoqué précédemment. Nous aurions alors dans notre complexe, un ensemble de protéines impliqué dans l'initiation de la transcription et peut être un autre groupe de protéines plutôt engagé dans l'élongation. À l'interface de ces deux complexes, il est possible que BGLF3 et BFRF2 participent au changement de conformation et de fonction du vPIC. En effet, ces deux protéines retiennent en coopération avec BcRF1 la protéine BDLF4. De plus, l'absence de BGLF3 ou BFRF2 semble préjudiciable au maintien de BVLF1 et BDLF3.5 dans le complexe. Ainsi BcRF1 et BDLF4 participeraient à l'initiation de la transcription alors que BVLF1 et BDLF3.5 pourraient être associées l'élongation. Toutefois

ce modèle nécessite de nombreuses manipulations supplémentaires pour être validé, a commencé par faire le GST pull-down avec GST-BcRF1 et GST-BFRF2 pour vérifier comment BDLF4 est bien maintenue dans le complexe par ces protéines. Des expériences de purification sur gradient de sucrose, pourront peut-être séparer les différentes associations présentes dans le vPIC. Des essais de CHIP et de CHIP-re-CHIP sur les promoteurs et sur différentes parties des gènes viraux tardifs permettraient d'observer un éventuellement enrichissement des protéines du vPIC sur les sites testés. Enfin, des manipulations de FRET en contexte infectieux amèneraient plus de réponses quant à l'association des membres du vPIC.

Nous démontrons aussi dans cette étude que seule la protéine BcRF1 (TBP-like) semble capable d'interagir avec l'ARN polymérase II cellulaire. Cependant, nos résultats divergent avec la publication démontrant le rôle de l'homologue de BVL1 (UL79) dans l'élongation. En effet, les auteurs de cette étude prouvent que la protéine virale est capable d'interagir avec différentes sous-unités de l'ARN pol II (Perng et al., 2014). Cette peut probablement s'expliquer par le fait que nos expériences de GST pull-down n'ont pas été réalisés en contexte infectieux au contraire de l'article de Perng. Toutefois, les auteurs de cette étude ont démontré que UL79 n'était pas capable de recruter l'ARN pol II sur les promoteurs tardifs. Ceci pourrait suggérer que ce rôle reviendrait à BcRF1 puisqu'il s'agit de la seule protéine virale capable d'interagir avec cette enzyme cellulaire en contexte non infectieux (cf. figure 51). Effectivement, l'équipe de Glaunsinger a confirmé l'interaction entre ORF24 (homologue de BcRF1) et plusieurs sous-unités de l'ARN pol II au niveau des promoteurs des gènes viraux tardifs (Davis et al., 2015b). In fine, UL79 serait donc liée à l'ARN pol II uniquement après avoir été intégrée dans le vPIC, probablement au niveau des promoteurs de gènes viraux tardifs.

De plus, TFIIB et TFIIH sont présentes au niveau des promoteurs des gènes tardifs mais n'interagissent pas avec ORF24 qui est l'homologue de BcRF1 (Davis et al., 2015b). Ceci pourrait suggérer que d'autres protéines virales recrutent ces facteurs de transcription sur les promoteurs. Il conviendrait d'analyser les interactions des autres membres du vPIC avec les facteurs cellulaires (facteurs de transcription et TAF). Aussi, l'hypothèse que différents facteurs généraux de la transcription cellulaire soient présents sur les promoteurs des gènes tardifs laisse imaginer que les six protéines virales du vPIC seraient plutôt des TATA-Associated Factors-like (TAFs-like) plutôt que des facteurs de transcription à proprement

parler. En effet, BcRF1 est une TBP-like autour de laquelle s'assemble un minimum de cinq autres protéines virales.

IV) Identification des partenaires potentiels du vPIC.

1) BTRF1 et BGLF4.

Si BTRF1 ne semble pas être impliquée dans l'expression des gènes viraux tardifs, plusieurs articles mettent en avant un rôle de BGLF4 (qui est la sérine/thréonine kinase) dans l'expression de ces gènes.

Tout d'abord, BGLF4 a été démontrée comme impliquée dans la réplication de l'ADN viral en phosphorylant le facteur de processivité de l'ADN polymérase virale : BMRF1 (Chen et al., 2000; Gershburg and Pagano, 2002; Yang et al., 2008). De plus, BGLF4 a un rôle dans l'import nucléaire des protéines virales, notamment des protéines du complexe hélicase/primase (Chang et al., 2015) en modulant la structure des pores nucléaires. Aussi, BGLF4 est retrouvée dans les compartiments de réplication de l'ADN viral (Wang et al., 2005) et semble être impliquée dans la synthèse des ARNm viraux tardifs codant pour les protéines de structure (El-Guindy et al., 2014). L'ensemble de ces éléments nous laissait penser que BGLF4 pourrait être un candidat de choix.

Le système rapporteur luciférase (décrit figure 44) a été utilisé avec les plasmides codants pour les protéines du vPIC, plus une quantité décroissante de plasmide codant pour la kinase BGLF4. Les résultats obtenus démontrent qu'une quantité croissante de BGLF4 inhibe légèrement l'activité transcriptionnelle du vPIC alors que les composants de ce dernier sont efficacement exprimés dans toutes les conditions et ne semblent pas interagir avec la kinase virale.

Si nos données suggèrent un rôle négligeable de BGLF4 dans l'expression des gènes viraux tardifs, ceux de l'équipe de Miller prouvent le contraire (El-Guindy et al., 2014). Ces différences de résultats sont très probablement liées aux systèmes utilisés. En effet, nos expériences (essais luciférase) ont été réalisées *in vitro* alors que celles d'El-Guindy (quantification de génomes viraux répliqués et ARNm viraux endogènes exprimés) ont été produites en contexte infectieux.

Néanmoins, il a été démontré que BGLF4 phosphoryle BMRF1, entraînant une légère augmentation de la réplication de l'ADN viral via l' OriLyt (Yang et al., 2008), ce qui est également observé dans l'article de l'équipe de Miller (El-Guindy et al., 2014). Ce résultat est conjugué à une augmentation de l'expression des transcrits viraux tardifs. De manière intéressante, l'extinction de BGLF4 abolit l'expression des gènes viraux tardifs sans modifier drastiquement la quantité de génome viraux répliqués (El-Guindy et al., 2014). Or, l'extinction de BGLF4 mène aussi à une inhibition drastique de l'expression de BGLF3 (El-Guindy et al., 2014) ainsi que de BFRF2 (Whitehurst et al., 2013), qui sont toutes deux impliquées dans l'expression des gènes viraux tardifs. Il est donc possible que l'effet de BGLF4 observé par le groupe de Miller soit lié à ce phénomène remarqué avec BFRF2 et/ou BGLF3 ; ou bien à son rôle dans la modulation de l'entrée des protéines virales dans le noyau. Les résultats obtenus par l'équipe de Miller ne seraient donc pas dus à un rôle direct de la kinase virale dans l'expression des gènes viraux tardifs.

2) Analyse par spectrométrie de masse.

Dans le but d'identifier d'autres facteurs protéiques potentiellement associés au vPIC, nous avons décidé d'utiliser une technique de double purification par chromatographie d'affinité (TAp-Tag) suivi d'une analyse en spectrométrie de masse des échantillons purifiés. Après la double purification et élution sur colonnes Streptavidine puis Calmoduline, les éluas ont été déposés sur gel SDS-PAGE. Les protéines uniquement retrouvées en présence de la protéine d'appât Streptavidine-Calmoduline-BcRF1 ont été prélevées avant d'être analysées par spectrométrie de masse (MS).

Malheureusement, les résultats escomptés furent très minces. Plusieurs raisons peuvent être invoquées. Tout d'abord, la protéine BcRF1 a été utilisée en système transitoire et est donc surreprésentée dans notre analyse. D'autre part nos résultats de GST pull-down *in vivo* démontrent que BGLF3 serait une meilleure protéine d'appât car elle semble être au cœur du vPIC. Par ailleurs l'absence de détection de protéines du vPIC (hormis BcRF1) alors qu'elles sont efficacement précipitées par GST pull-down *in vivo* s'expliquerait par une très faible expression de ces protéines en conditions d'infection.

En effet, il a été publié en 2012 que les protéines du vPIC de HCMV n'étaient que très peu exprimées par le virus (Stern-Ginossar et al., 2012). De manière similaire, l'induction du cycle productif dans des cellules de lymphome de Burkitt ne permet pas de détecter les

différentes protéines du vPIC (Traylen et al., 2015). Nos résultats, combinés à ceux de la littérature laissent sous entendre que les protéines du vPIC ne sont pas ou que très peu exprimées, expliquant pourquoi il est presque impossible de les observer en conditions physiologiques.

Ainsi, afin d'optimiser ces résultats, il faudrait à l'avenir utiliser une protéine étiquetée directement dans le virus afin d'obtenir un échantillon en condition physiologique et prendre en protéine d'appât une protéine centrale du vPIC telle que BGLF3.

Enfin, malgré le peu de protéines purifiées et l'absence de détection des membres du vPIC après analyse par MS, les résultats obtenus ont permis de confirmer l'interaction de BcRF1 avec l'ARN pol II mais ont également démontré que la protéine de liaison à l'ADN simple brin, BALF2 était associée à la TBP-like virale ! Cette interaction suggère pour la première fois un lien entre le complexe de transcription et le complexe de réplication.

V) Le complexe de réplication de l'ADN viral, partenaire du vPIC ?

Afin d'approfondir les résultats obtenus par spectrométrie de masse, la première approche a été de réaliser des GST pull-down *in vitro*. Ceux-ci n'ont pas confirmé d'interaction directe entre BcRF1 et BALF2. En revanche, différents membres du vPIC interagissent efficacement avec BMRF1 (facteur de processivité de l'ADN polymérase virale), BBLF4 (hélicase) et BSLF1 (primase). Ces résultats laissent ainsi supposer que BALF2 interagirait indirectement avec BcRF1. En revanche, il semble que deux des trois protéines du complexe hélicase/primase soient associées au vPIC.

Dans le but de valider cette hypothèse, nous avons réalisé des GST pull-down *in vivo* entre le complexe de réplication de l'ADN viral et le complexe d'initiation de la transcription des gènes viraux tardifs. Les résultats obtenus démontrent que deux protéines du complexe de réplication : BBLF4 et BSLF1, interagissent avec plusieurs facteurs du vPIC. Cette interaction est assez faible BBLF4 et BSLF1 sont exprimées individuellement alors qu'elle est renforcée lorsque le complexe de réplication est reconstitué. Ceci suggérerait que le complexe hélicase/primase doit être fonctionnel, et donc être composé de ces trois facteurs pour interagir avec le vPIC. En effet, il a été démontré que le complexe hélicase/primase est

constitué d'une copie de chacune des trois protéines pour être à priori fonctionnel (Thierry et al., 2015; Yokoyama et al., 1999).

Afin de confirmer ces résultats, des expériences des GST pull-down inverses devront être réalisés. De plus, ce même type d'expérience devra être testé en reconstituant le complexe hélicase/primase pour observer d'éventuelles différences d'interactions avec le vPIC.

Cette découverte est une étape clef dans la compréhension des mécanismes régulant l'expression des gènes viraux tardifs. En effet l'expression de ces gènes est dépendante de la réplication de l'ADN viral. Elle est assurée par des protéines virales parmi lesquelles BBLF4 et BSLF1. L'interaction de ces deux protéines avec plusieurs membres du vPIC pourrait ainsi faire le lien entre réplication et transcription des gènes tardifs. En quoi l'interaction entre BBLF4-BSLF1 et le vPIC est-elle importante pour l'expression des gènes tardifs ? Plusieurs hypothèses peuvent être avancées.

1) L'hélicase/primase servirait-elle de cargo pour le transport du vPIC dans le noyau ?

Tout d'abord, des données préliminaires au laboratoire mettent en évidence par immunofluorescence que les protéines du vPIC sont réparties aléatoirement (noyau et cytoplasme) dans des cellules HEK293 ou HEK293_{EBV}. De manière intéressante, les protéines du vPIC sont retrouvées majoritairement dans le noyau uniquement dans les cellules HEK293_{EBV} dans lesquelles le cycle viral productif a été induit. Ceci suggère donc que des protéines virales permettraient le recrutement du vPIC dans le noyau des cellules infectées.

Le complexe hélicase/primase, constitué des protéines BBLF4-BSLF1-BBLF2/3, pourrait être l'élément recrutant le vPIC dans le noyau. Effectivement, si l'ADN polymérase rentre dans le noyau à l'aide de BMRF1 qui possède un signal de localisation nucléaire (NLS) (Kawashima et al., 2013), l'entrée du complexe hélicase/primase dans ce compartiment est mal connu. Il semble que la protéine EB1 permette la relocalisation de chacune des protéines du complexe hélicase/primase dans le noyau (Gao et al., 1998). Néanmoins, les auteurs de cette étude démontrent que la reconstitution du complexe hélicase/primase permet aux trois protéines d'être relocalisées dans le noyau alors qu'elles ne possèdent pas, à priori, de NLS.

Aussi, ces trois facteurs protéiques (BBLF4-BSLF1-BBLF2/3) interagissent avec l'ADN polymérase virale BALF5 (Fujii et al., 2000) et pourraient donc rentrer dans le noyau à l'aide du cargo BALF5-BMRF1. Dans tous les cas, cette donnée permettrait d'expliquer pourquoi BALF5 semble être purifiée sur certaines colonnes GST-vPIC.

Afin de tester cette hypothèse, il serait intéressant de regarder par immunofluorescence la localisation du vPIC en présence ou en absence du complexe hélicase/primase et avec ou sans EB1. Des extinctions de gènes par siRNA, dirigés contre les gènes codant pour les protéines cargos telles que les Importines, permettraient de caractériser la voie empruntée par les différents facteurs des complexes hélicase/primase et vPIC pour rentrer dans le noyau.

2) L'hélicase/primase recruterait-elles le vPIC sur l'ADN ?

Une autre hypothèse serait une relocalisation du vPIC dans le compartiment de réplication de l'ADN viral mais aussi un positionnement sur l'ADN en cours de synthèse, assuré par le complexe hélicase/primase. De manière intéressante, il a été démontré que la protéine virale BcRF1 (TBP-like) est retrouvée dans l'usine virale dans les étapes tardives de l'infection, en compagnie des ARNm tardifs (Sugimoto et al., 2013). Ceci signifie donc que la synthèse des ARNm viraux tardifs se fait dans le compartiment de réplication de l'ADN viral lorsque l'ADN viral est répliqué en quantité. Il serait alors intéressant de regarder par fluorescence la reconstitution de l'usine virale afin de visualiser la localisation du vPIC en présence ou en absence du complexe hélicase/primase dans des cellules EBV négatives. Pour le recrutement sur l'ADN, des expériences de chromatine immunoprécipitation (ChIP-re-ChIP) avec une des sous-unités de l'hélicase/primase et une du vPIC seraient à réaliser. Pour ce faire, il est possible d'utiliser une lignée cellulaire stable contenant un vecteur possédant l'OriP (pour se maintenir sous forme de minichromosome dans les cellules), l'OriLyt (pour la réplication de l'ADN viral lors du cycle productif) et un promoteur de gène tardif TATT, comme décrit dans l'article de l'équipe de Farrell (Amon et al., 2004). Il serait alors possible de reconstituer le complexe de réplication de l'ADN ainsi que le vPIC en transfectant dans cette lignée cellulaire les plasmides correspondants. Ainsi, un potentiel enrichissement du vPIC sur l'ADN pourrait être observé au niveau de sites préalablement sélectionnés. Une

autre approche serait par microscopie de localiser les deux complexes dans des cellules HEK293_{EBV} et visualiser leur positionnement par rapport à l'ADN.

3) L'expression des gènes viraux tardifs est-elle dépendante du complexe de réplication de l'ADN viral ou bien nécessite-t-elle uniquement un changement de conformation de la chromatine ?

Il est connu depuis plusieurs décennies que l'expression des gènes tardifs du virus d'Epstein-Barr est dépendante de la réplication. L'ADN nouvellement synthétisé est encapsidé dans les nouvelles particules virales, libre de toute marque de chromatine au contraire de l'ADN parental présent sous forme d'épisome dans les cellules infectées (Hammerschmidt and Sugden, 2013; Kalla et al., 2010; Shaw et al., 1979). Par ailleurs, bien que le vPIC soit capable d'activer l'expression des gènes viraux tardifs dans un système rapporteur transitoire, il est incapable de le faire dans un contexte infectieux si le cycle viral productif n'est pas induit. Ceci signifie donc qu'il est dépendant de l'étape de réplication de l'ADN viral. Ces données suggèrent soit un rôle du complexe de réplication (via l'hélicase/primase), soit un changement d'état de la chromatine, soit des deux. Le changement d'état de la chromatine peut être induit de plusieurs manières. Il est possible que les facteurs de remodelage de la chromatine par exemple engendrent des modifications post-traductionnelles sur les nucléosomes, favorables à la transcription. L'autre option est l'absence de chromatine sur l'ADN viral nouvellement synthétisé (tel qu'il est retrouvé dans les particules virales) qui servirait de matrice au vPIC pour la synthèse des ARNm viraux tardifs.

Pour répondre à ces différentes possibilités, plusieurs expériences pourront être réalisées. Tout d'abord, en utilisant la lignée cellulaire évoquée précédemment (Amon et al., 2004). Celle-ci contient un vecteur rapporteur luciférase dont le gène est sous contrôle d'un promoteur de gène tardif. Ce plasmide possède également les origines de réplication de latence (OriP) et du cycle productif (OriLyt). Ainsi, le vecteur est maintenu dans la cellule sous forme de mini chromosome, comme EBV.

Par exemple, la transfection de cette lignée cellulaire avec les différents plasmides d'expression codant pour les protéines du vPIC n'induit pas l'activation de la luciférase (Gruffat H, datas non montrées). Ceci suggère donc un rôle prépondérant de la chromatine. Il

serait intéressant, au vu des interactions observées précédemment, de transfecter soit les vecteurs d'expression codants pour le complexe hélicase/primase soit ceux codant pour tous les membres du complexe de réplication de l'ADN viral, en présence ou en absence du vPIC. Ainsi, si l'activité luciférase est induite, ces résultats démontreraient le couplage entre la réplication et la transcription des gènes tardifs.

L'autre approche serait d'utiliser une lignée cellulaire HEK293_{EBV} dans laquelle le cycle viral productif serait induit par transfection avec le vecteur d'expression codant pour la protéine EB1. Les cellules seraient traitées quelques heures post-transfection avec un analogue nucléotidique tel que la cytosine arabinoside. Cette substance bloquerait alors l'incorporation de nucléotides dans l'ADN en cours de synthèse et permettrait aussi de conserver un complexe de réplication fonctionnel, au contraire de l'acyclovir ou du PAA (acide phosphonoacétique). La quantité d'ADN viral serait alors quantifiée par qPCR et les ARNm tardifs par RT-qPCR. L'utilisation de cette drogue a déjà démontrée que la réplication de l'ADN viral et l'expression de protéines virales tardives étaient inhibées dans des cellules B (Gergely et al., 1971a, 1971b). L'utilisation de cette drogue donnerait probablement des résultats identiques dans des cellules épithéliales EBV positive, en l'occurrence une absence d'expression des ARNm viraux tardifs, mais ceci reste à confirmer. L'autre approche serait d'utiliser la lignée cellulaire décrite précédemment (Amon et al., 2004) portant un vecteur rapporteur luciférase mais avec une OriLyt mutée. Ces dernières cellules seront ensuite transfectées avec les plasmides d'expression codant pour les protéines du complexe de réplication et du vPIC.

Si les ARNm viraux tardifs ne sont pas détectés après traitement ou transfection (dans la lignée cellulaire ayant l'OriLyt mutée), cette manipulation sera un argument allant en faveur de la nécessité d'une modification de l'état de la chromatine pour la transcription des gènes viraux tardifs. À l'inverse, si les ARNm viraux tardifs sont exprimés, alors ceci signifierait que le complexe de réplication de l'ADN viral est nécessaire pour ouvrir l'ADN viral et recruter le vPIC (peut être via l'hélicase/primase) sur l'ADN parental qui servirait de matrice à la synthèse de ARNm tardifs.

De plus, des expériences de ChIP-re-ChIP entre des histones et le vPIC permettraient de voir si le complexe viral est préférentiellement présent sur de l'ADN parental ou bien de l'ADN nouvellement synthétisé et nu de tous marquage chromatinien.

Ainsi, l'ensemble de ces expériences permettrait de comprendre le lien impliquant le complexe hélicase/primase et le vPIC et surtout le couplage entre la réplication de l'ADN viral et l'expression des gènes viraux tardifs.

VI) Modèle.

Ainsi, au cours de cette thèse, nous avons démontré qu'un ensemble de six protéines virales, associées à l'ARN polymérase II cellulaire formaient un complexe que nous avons nommé vPIC. Ce complexe est dépendant de la réplication de l'ADN viral et, de manière intéressante interagit avec le complexe hélicase/primase impliqué dans la synthèse de l'ADN viral. Nous émettons l'hypothèse que le complexe hélicase/primase permette de faire le cargo pour amener le complexe de transcription des gènes viraux tardifs dans le noyau puis de le conduire dans le compartiment de réplication. Le début de la réplication de l'ADN viral va permettre de synthétiser de nombreuses matrices d'ADN, qui seront libres de marque chromatinienne, sur lesquelles iront se fixer les protéines du vPIC pour assurer la transcription des gènes viraux tardifs (cf. figure 65).

Ce modèle est très semblable à celui du bactériophage T4 évoqué en introduction. En effet, la transcription des gènes tardifs de ce virus se fait à partir d'un promoteur simplifié sur lequel est présente une protéine virale capable de reconnaître la séquence spécifique à ce promoteur (à la manière de BcRF1 chez EBV). Une autre protéine du bactériophage recrute l'ARN pol bactérienne (comme BcRF1) alors que le facteur de processivité de l'ADN polymérase virale (homologue à BMRF1) permet de faire le lien entre la « TBP-like » du T4 et la protéine virale qui détourne l'enzyme bactérienne. Le virus d'Epstein-Barr utilise lui, une TBP-like autour de laquelle sont présentes cinq autres protéines virales ainsi que l'ARN pol II. Le complexe hélicase/primase permettrait de faire le lien entre la réplication et la transcription des gènes tardifs.

Cette similarité de régulation de gènes tardifs entre bactériophages T4 et Herpesvirus pourrait insinuer que ces deux familles virales auraient eu un ancêtre commun avant de diverger. En effet, en plus de cette similitude dans la régulation de l'expression des gènes tardifs, des homologies entre certaines protéines de capsid (Dai et al., 2015) ; entre l'hélicase du T4 et celle d'EBV (BSLF1) (Thierry et al., 2015) ainsi que dans le modèle d'assemblage

de la capside virale (Brown and Newcomb, 2011) sont autant d'arguments suggérant une probable origine commune de ces deux familles.

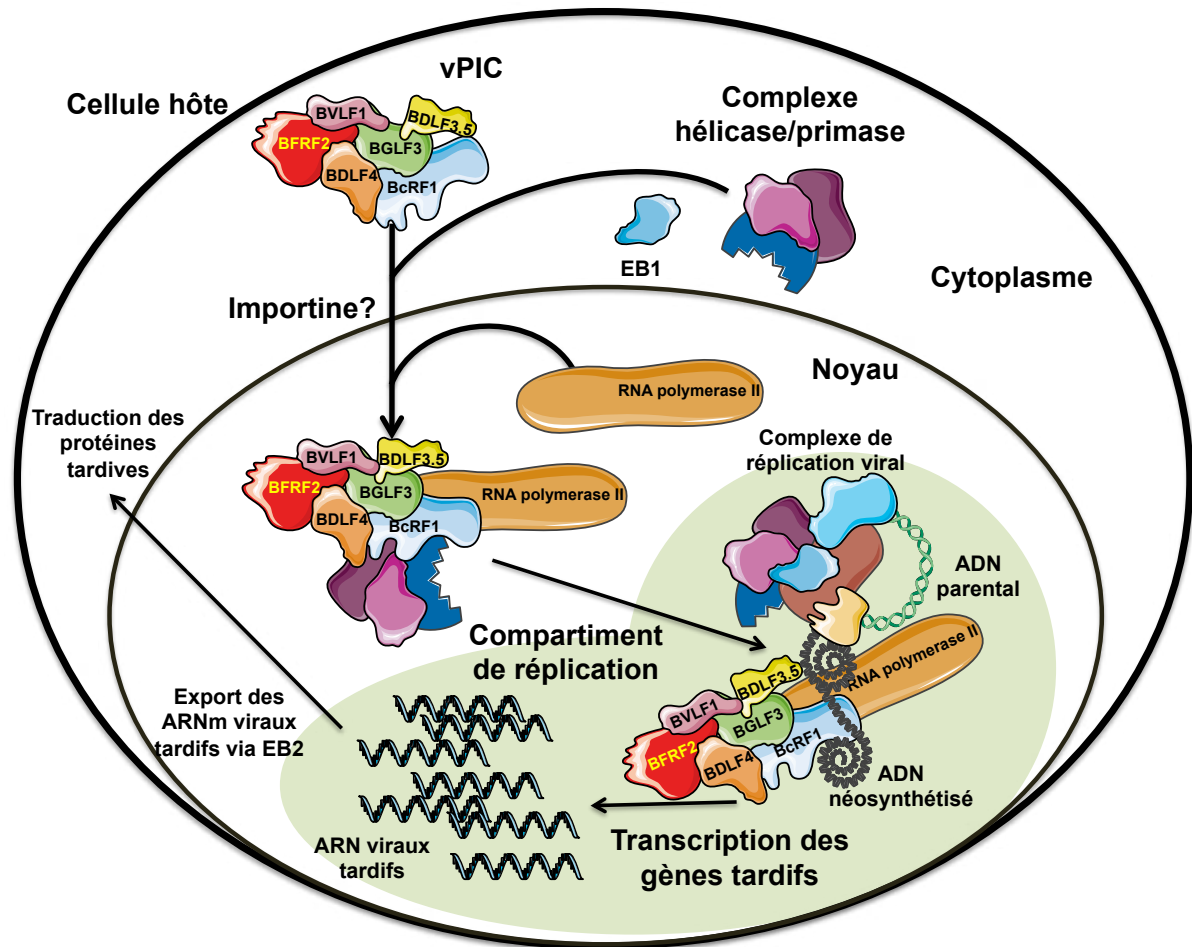


Figure 65: Représentation schématique du recrutement du vPIC dans l'usine virale et son action sur la transcription des gènes viraux tardifs.

Les protéines du vPIC sont traduites dans le cytoplasme et sont peut être transportées dans le noyau en complexe avec l'hélicase/primase. Le vPIC est ensuite recruté dans l'usine virale (nommée compartiment de réplication) et va assurer la transcription des gènes viraux tardifs. Cette synthèse est probablement réalisée sur l'ADN nouvellement synthétisé et est peut être couplée voire associée à certains composants du complexe de réplication de l'ADN viral (notamment le complexe hélicase/primase). Les ARNm viraux tardifs transcrits vont alors être exportés du noyau dans le cytoplasme à l'aide de EB2 pour être traduits en protéines de structure qui permettront la formation de nouvelles particules virales infectieuses.

Ainsi, l'idée que les α -, β - et γ -Herpesvirus proviendraient d'un même ancêtre commun, proche du bactériophage T4, ne semble pas extravagant. En effet, la présence d'un promoteur de gène tardif simplifié (juste une boîte TATA) chez le bactériophage T4, fixé par une TBP-like virale qui est elle-même associée à une protéine de la réplication, est un modèle proche du mécanisme de régulation des β - et γ -Herpesvirus. Il est alors possible de supputer

que les α -Herpesvirus possédaient ce système de régulation pour leurs gènes tardifs avant de diverger des β - et γ -Herpesvirus précocement au cours de l'Histoire pour acquérir le mécanisme décrits aujourd'hui dans la littérature.

VII) Perspectives du projet.

Comme évoqué précédemment, le virus est très répandu dans la population (>95% de séropositifs), et persiste sous forme latente dans les cellules B mémoires des individus infectés. Par intermittence, des épisodes de production de particules virales au niveau de l'oropharynx sont détectées, ce qui suggère qu'un cycle viral productif a lieu *in vivo* dans des cellules de l'épithélium oropharyngé. Outre le fait qu'elles soient responsables d'une propagation très efficace du virus entre individus, ces phases de réactivation virale semblent être aussi un facteur de risques pour l'émergence de certains cancers (Rickinson et al., 1974; Timms et al., 2003). Par ailleurs, certains travaux suggèrent que l'émergence de pathologies associées à EBV pourrait être générée à la suite d'épisodes de productions virales importantes accompagnées d'infection de nouvelles cellules B (Rooney, 2003). De plus, des LCL (Lymphoblastoid Cell Line : cellules B immortalisées par EBV *in vitro*) infectées avec un virus défectif pour le déroulement du cycle productif développent moins de tumeurs chez des souris SCID comparée à des LCL infectées avec un virus non muté (Hong et al., 2005). La présence de cellules productrices au sein des tumeurs pourrait donc être un facteur important de la cancérogénèse.

C'est pourquoi, comprendre les mécanismes régulant l'expression des gènes du cycle productif d'EBV est essentiel afin de sélectionner des cibles potentielles pour un traitement. En effet, en l'absence de vaccin et d'antidrogue spécifiquement dirigés contre EBV, ainsi que la multiplication des cas de résistances aux traitements antiviraux tels que l'acyclovir et ses analogues sont un problème pour prévenir et soigner les personnes atteintes de pathologies associées à EBV (notamment chez les personnes immunodéprimées).

Dans ce contexte, le vPIC est une cible de choix. Effectivement, ce complexe viral est conservé entre tous les β - et γ -Herpesvirus et est propre à la synthèse des gènes viraux tardifs. En bloquant spécifiquement ce complexe viral à l'aide d'une molécule inhibitrice, la production de particules virales serait annihilée. Afin, de procéder à un criblage de molécule,

il est important de comprendre le fonctionnement du vPIC d'une part, mais aussi de connaître la structure de ce complexe dans le but de mieux caractériser l'action d'un éventuel inhibiteur. Cette dernière partie est actuellement en cours de réalisation par le Dr Marchione Roberta à Grenoble, en collaboration avec les équipes des Drs Carlo Petosa et Imre Berger. La production des différentes protéines du vPIC se fait à partir du système multi-Bac (dérivé de baculovirus) en cellules d'insectes (Berger et al., 2004; Bieniossek et al., 2012). Après l'optimisation des procédures de purification, la structure du vPIC pourra être analysée par cryo-électromicroscopie. Les domaines d'interactions entre les différents membres du vPIC devront également être étudiés afin de déterminer des sites potentiellement accessibles à des molécules inhibitrices, susceptibles d'interférer spécifiquement dans les associations du vPIC. Ces manipulations devront également être réalisées sur d'autres β - et γ -Herpesvirus pour confirmer la conservation de ce complexe viral. Une fois le crible effectué, des tests de toxicité sur les cellules et de doses devront être réalisés afin de vérifier l'action de cette drogue ou ce peptide inhibiteur spécifiquement contre le vPIC. Ces éléments serviront de base pour l'élaboration d'un médicament efficace contre la production de nouvelles particules virales de Beta- et Gamma-Herpesvirus.

PARTIE IV :

Références Bibliographiques

- Abend, J.R., Joseph, A.E., Das, D., Campbell-Cecen, D.B., and Imperiale, M.J. (2009). A truncated T antigen expressed from an alternatively spliced BK virus early mRNA. *J. Gen. Virol.* *90*, 1238–1245.
- Acharya, A., and Gopinathan, K.P. (2001). Identification of an enhancer-like element in the polyhedrin gene upstream region of *Bombyx mori* nucleopolyhedrovirus. *J. Gen. Virol.* *82*, 2811–2819.
- Ai, W., Narahari, J., and Roman, A. (2000). Yin Yang 1 Negatively Regulates the Differentiation-Specific E1 Promoter of Human Papillomavirus Type 6. *J. Virol.* *74*, 5198–5205.
- Amemiya, K., Traub, R., Durham, L., and Major, E.O. (1989). Interaction of a nuclear factor-1-like protein with the regulatory region of the human polyomavirus JC virus. *J. Biol. Chem.* *264*, 7025–7032.
- Amemiya, K., Traub, R., Durham, L., and Major, E.O. (1992). Adjacent nuclear factor-1 and activator protein binding sites in the enhancer of the neurotropic JC virus. A common characteristic of many brain-specific genes. *J. Biol. Chem.* *267*, 14204–14211.
- Amon, W., Binné, U.K., Bryant, H., Jenkins, P.J., Elgueta Karstegl, C., and Farrell, P.J. (2004). Lytic Cycle Gene Regulation of Epstein-Barr Virus. *J. Virol.* *78*, 13460–13469.
- Apt, D., Chong, T., Liu, Y., and Bernard, H.U. (1993). Nuclear factor I and epithelial cell-specific transcription of human papillomavirus type 16. *J. Virol.* *67*, 4455–4463.
- Arumugaswami, V., Wu, T.-T., Martinez-Guzman, D., Jia, Q., Deng, H., Reyes, N., and Sun, R. (2006). ORF18 Is a Transfactor That Is Essential for Late Gene Transcription of a Gammaherpesvirus. *J. Virol.* *80*, 9730–9740.
- Aubry, V., Mure, F., Mariamé, B., Deschamps, T., Wyrwicz, L.S., Manet, E., and Gruffat, H. (2014). Epstein-Barr Virus Late Gene Transcription Depends on the Assembly of a Virus-Specific Preinitiation Complex. *J. Virol.* *88*, 12825–12838.
- Babcock, G.J., Hochberg, D., and Thorley-Lawson, D.A. (2000). The Expression Pattern of Epstein-Barr Virus Latent Genes In Vivo Is Dependent upon the Differentiation Stage of the Infected B Cell. *Immunity* *13*, 497–506.
- Backström, E., Kaufmann, K.B., Lan, X., and Akusjärvi, G. (2010). Adenovirus L4-22K stimulates major late transcription by a mechanism requiring the intragenic late-specific transcription factor-binding site. *Virus Res.* *151*, 220–228.
- 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., Séguin, C., et al. (1984). DNA sequence and expression of the B95-8 Epstein—Barr virus genome. *Nature* *310*, 207–211.
- Balakrishnan, L., and Milavetz, B. (2007). Histone Hyperacetylation in the Coding Region of Chromatin Undergoing Transcription in SV40 Minichromosomes Is a Dynamic Process Regulated Directly by the Presence of RNA Polymerase II. *J. Mol. Biol.* *365*,

- Barrera-Saldana, H., Takahashi, K., Vigneron, M., Wildeman, A., Davidson, I., and Chambon, P. (1985). All six GC-motifs of the SV40 early upstream element contribute to promoter activity in vivo and in vitro. *EMBO J.* *4*, 3839–3849.
- Bataille, A.R., Jeronimo, C., Jacques, P.-É., Laramée, L., Fortin, M.-È., Forest, A., Bergeron, M., Hanes, S.D., and Robert, F. (2012). A Universal RNA Polymerase II CTD Cycle Is Orchestrated by Complex Interplays between Kinase, Phosphatase, and Isomerase Enzymes along Genes. *Mol. Cell* *45*, 158–170.
- Beggs, A.H., Frisque, R.J., and Scangos, G.A. (1988). Extinction of JC virus tumor-antigen expression in glial cell–fibroblast hybrids. *Proc. Natl. Acad. Sci. U. S. A.* *85*, 7632–7636.
- BENIYA, H., FUNK, C.J., ROHRMANN, G.F., and WEAVER, R.F. (1996). Purification of a Virus-Induced RNA Polymerase from *Autographa californica* Nuclear Polyhedrosis Virus-Infected *Spodoptera frugiperda* Cells That Accurately Initiates Late and Very Late Transcription in Vitro. *Virology* *216*, 12–19.
- Benoist, C., and Chambon, P. (1981). In vivo sequence requirements of the SV40 early promoter region. *Nature* *290*, 304–310.
- Bergbauer, M., Kalla, M., Schmeinck, A., Göbel, C., Rothbauer, U., Eck, S., Benet-Pagès, A., Strom, T.M., and Hammerschmidt, W. (2010). CpG-Methylation Regulates a Class of Epstein-Barr Virus Promoters. *PLoS Pathog.* *6*.
- Berger, I., Fitzgerald, D.J., and Richmond, T.J. (2004). Baculovirus expression system for heterologous multiprotein complexes. *Nat. Biotechnol.* *22*, 1583–1587.
- Berger, L.C., Smith, D.B., Davidson, I., Hwang, J.J., Fanning, E., and Wildeman, A.G. (1996). Interaction between T antigen and TEA domain of the factor TEF-1 derepresses simian virus 40 late promoter in vitro: identification of T-antigen domains important for transcription control. *J. Virol.* *70*, 1203–1212.
- Bernard, H.-U. (2013). Regulatory elements in the viral genome. *Virology* *445*, 197–204.
- Bhende, P.M., Seaman, W.T., Delecluse, H.-J., and Kenney, S.C. (2004). The EBV lytic switch protein, Z, preferentially binds to and activates the methylated viral genome. *Nat. Genet.* *36*, 1099–1104.
- Bhende, P.M., Dickerson, S.J., Sun, X., Feng, W.-H., and Kenney, S.C. (2007). X-Box-Binding Protein 1 Activates Lytic Epstein-Barr Virus Gene Expression in Combination with Protein Kinase D. *J. Virol.* *81*, 7363–7370.
- Biasiotto, R., and Akusjärvi, G. (2015). Regulation of Human Adenovirus Alternative RNA Splicing by the Adenoviral L4-33K and L4-22K Proteins. *Int. J. Mol. Sci.* *16*, 2893–2912.
- Bieniossek, C., Imasaki, T., Takagi, Y., and Berger, I. (2012). MultiBac: expanding the

- research toolbox for multiprotein complexes. *Trends Biochem. Sci.* 37, 49–57.
- Bieniossek, C., Papai, G., Schaffitzel, C., Garzoni, F., Chaillet, M., Scheer, E., Papadopoulos, P., Tora, L., Schultz, P., and Berger, I. (2013). The architecture of human general transcription factor TFIID core complex. *Nature* 493, 699–702.
- Blissard, G.W., Kogan, P.H., Wei, R., and Rohrmann, G.F. (1992). A synthetic early promoter from a baculovirus: Roles of the TATA box and conserved start site CAGT sequence in basal levels of transcription. *Virology* 190, 783–793.
- Bodily, J.M., and Meyers, C. (2005). Genetic Analysis of the Human Papillomavirus Type 31 Differentiation-Dependent Late Promoter. *J. Virol.* 79, 3309–3321.
- Bollag, B., Hofstetter, C.A., Reviriego-Mendoza, M.M., and Frisque, R.J. (2010). JC Virus Small t Antigen Binds Phosphatase PP2A and Rb Family Proteins and Is Required for Efficient Viral DNA Replication Activity. *PLoS ONE* 5, e10606.
- Brady, J., and Khoury, G. (1985). trans Activation of the simian virus 40 late transcription unit by T-antigen. *Mol. Cell. Biol.* 5, 1391–1399.
- Brady, J., Bolen, J.B., Radonovich, M., Salzman, N., and Khoury, G. (1984). Stimulation of simian virus 40 late gene expression by simian virus 40 tumor antigen. *Proc. Natl. Acad. Sci. U. S. A.* 81, 2040–2044.
- Brady, J., Loeken, M.R., and Khoury, G. (1985). Interaction between two transcriptional control sequences required for tumor-antigen-mediated simian virus 40 late gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 82, 7299–7303.
- Bridge, E., and Ketner, G. (1989). Redundant control of adenovirus late gene expression by early region 4. *J. Virol.* 63, 631–638.
- Broekema, N.M., and Imperiale, M.J. (2013). miRNA regulation of BK polyomavirus replication during early infection. *Proc. Natl. Acad. Sci. U. S. A.* 110, 8200–8205.
- Brown, J.C., and Newcomb, W.W. (2011). Herpesvirus Capsid Assembly: Insights from Structural Analysis. *Curr. Opin. Virol.* 1, 142–149.
- Brulois, K., Wong, L.-Y., Lee, H.-R., Sivadas, P., Ensser, A., Feng, P., Gao, S.-J., Toth, Z., and Jung, J.U. (2015). Association of Kaposi's Sarcoma-Associated Herpesvirus ORF31 with ORF34 and ORF24 Is Critical for Late Gene Expression. *J. Virol.* 89, 6148–6154.
- Buisson, M., Manet, E., Trescol-Biemont, M.C., Gruffat, H., Durand, B., and Sergeant, A. (1989). The Epstein-Barr virus (EBV) early protein EB2 is a posttranscriptional activator expressed under the control of EBV transcription factors EB1 and R. *J. Virol.* 63, 5276–5284.
- Burma, S., Mukherjee, B., Jain, A., Habib, S., and Hasnain, S.E. (1994). An unusual 30-kDa protein binding to the polyhedrin gene promoter of *Autographa californica* nuclear polyhedrosis virus. *J. Biol. Chem.* 269, 2750–2757.

- Burton, Z.F., Ortolan, L.G., and Greenblatt, J. (1986). Proteins that bind to RNA polymerase II are required for accurate initiation of transcription at the adenovirus 2 major late promoter. *EMBO J.* *5*, 2923–2930.
- Carcamo, J., Maldonado, E., Cortes, P., Ahn, M.H., Ha, I., Kasai, Y., Flint, J., and Reinberg, D. (1990). A TATA-like sequence located downstream of the transcription initiation site is required for expression of an RNA polymerase II transcribed gene. *Genes Dev.* *4*, 1611–1622.
- Carrozza, M.J., and DeLuca, N.A. (1996). Interaction of the viral activator protein ICP4 with TFIID through TAF250. *Mol. Cell. Biol.* *16*, 3085–3093.
- Carstens, E.B., and Wu, Y. (2007). No single homologous repeat region is essential for DNA replication of the baculovirus *Autographa californica* multiple nucleopolyhedrovirus. *J. Gen. Virol.* *88*, 114–122.
- Casaz, P., Sundseth, R., and Hansen, U. (1991). trans activation of the simian virus 40 late promoter by large T antigen requires binding sites for the cellular transcription factor TEF-1. *J. Virol.* *65*, 6535–6543.
- Casaz, P., Rice, P.W., Cole, C.N., and Hansen, U. (1995). A TEF-1-independent mechanism for activation of the simian virus 40 (SV40) late promoter by mutant SV40 large T antigens. *J. Virol.* *69*, 3501–3509.
- Chan, W.K., Chong, T., Bernard, H.U., and Klock, G. (1990). Transcription of the transforming genes of the oncogenic human papillomavirus-16 is stimulated by tumor promoters through AP1 binding sites. *Nucleic Acids Res.* *18*, 763–769.
- Chang, C.-W., Lee, C.-P., Su, M.-T., Tsai, C.-H., and Chen, M.-R. (2015). BGLF4 Kinase Modulates the Structure and Transport Preference of the Nuclear Pore Complex To Facilitate Nuclear Import of Epstein-Barr Virus Lytic Proteins. *J. Virol.* *89*, 1703–1718.
- Chang, P.J., Chang, Y.S., and Liu, S.T. (1998). Characterization of the BcLF1 promoter in Epstein-Barr virus. *J. Gen. Virol.* *79*, 2003–2006.
- Chang, Y.N., Dong, D.L., Hayward, G.S., and 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*, 3358–3369.
- Chapa, T.J., Johnson, L.S., Affolter, C., Valentine, M.C., Fehr, A.R., Yokoyama, W.M., and Yu, D. (2013). Murine Cytomegalovirus Protein pM79 Is a Key Regulator for Viral Late Transcription. *J. Virol.* *87*, 9135–9147.
- Chapa, T.J., Perng, Y.-C., French, A.R., and Yu, D. (2014). Murine Cytomegalovirus Protein pM92 Is a Conserved Regulator of Viral Late Gene Expression. *J. Virol.* *88*, 131–142.
- Chen, H., and Flint, S.J. (1992). Mutational analysis of the adenovirus 2 IVa2 initiator and downstream elements. *J. Biol. Chem.* *267*, 25457–25465.

- Chen, N.N., and Khalili, K. (1995). Transcriptional regulation of human JC polyomavirus promoters by cellular proteins YB-1 and Pur alpha in glial cells. *J. Virol.* *69*, 5843–5848.
- Chen, H., Vinnakota, R., and Flint, S.J. (1994). Intragenic activating and repressing elements control transcription from the adenovirus IVa2 initiator. *Mol. Cell. Biol.* *14*, 676–685.
- Chen, M.-R., Chang, S.-J., Huang, H., and Chen, J.-Y. (2000). A Protein Kinase Activity Associated with Epstein-Barr Virus BGLF4 Phosphorylates the Viral Early Antigen EA-D In Vitro. *J. Virol.* *74*, 3093–3104.
- Chen, Y.-R., Zhong, S., Fei, Z., Hashimoto, Y., Xiang, J.Z., Zhang, S., and Blissard, G.W. (2013). The Transcriptome of the Baculovirus *Autographa californica* Multiple Nucleopolyhedrovirus in *Trichoplusia ni* Cells. *J. Virol.* *87*, 6391–6405.
- Cheng, L., and Kelly, T.J. (1989). Transcriptional activator nuclear factor I stimulates the replication of SV40 minichromosomes in vivo and in vitro. *Cell* *59*, 541–551.
- Chevallier-Greco, A., Manet, E., Chavrier, P., Mosnier, C., Daillie, J., and Sergeant, A. (1986). Both Epstein-Barr virus (EBV)-encoded trans-acting factors, EB1 and EB2, are required to activate transcription from an EBV early promoter. *EMBO J.* *5*, 3243–3249.
- Chong, T., Apt, D., Gloss, B., Isa, M., and Bernard, H.U. (1991). The enhancer of human papillomavirus type 16: binding sites for the ubiquitous transcription factors oct-1, NFA, TEF-2, NF1, and AP-1 participate in epithelial cell-specific transcription. *J. Virol.* *65*, 5933–5943.
- Choy, B., and Green, M.R. (1993). Eukaryotic activators function during multiple steps of preinitiation complex assembly. *Nature* *366*, 531–536.
- Cianfrocco, M.A., Kassavetis, G.A., Grob, P., Fang, J., Juven-Gershon, T., Kadonaga, J.T., and Nogales, E. (2013). Human TFIID Binds to Core Promoter DNA in a Reorganized Structural State. *Cell* *152*, 120–131.
- Clem, R.J., Fechheimer, M., and Miller, L.K. (1991). Prevention of apoptosis by a baculovirus gene during infection of insect cells. *Science* *254*, 1388–1390.
- Collier, B., Goobar-Larsson, L., Sokolowski, M., and Schwartz, S. (1998). Translational Inhibition in Vitro of Human Papillomavirus Type 16 L2 mRNA Mediated through Interaction with Heterogenous Ribonucleoprotein K and Poly(rC)-binding Proteins 1 and 2. *J. Biol. Chem.* *273*, 22648–22656.
- Collier, B., Öberg, D., Zhao, X., and Schwartz, S. (2002). Specific Inactivation of Inhibitory Sequences in the 5' End of the Human Papillomavirus Type 16 L1 Open Reading Frame Results in Production of High Levels of L1 Protein in Human Epithelial Cells. *J. Virol.* *76*, 2739–2752.
- Comai, L., Zomerdijk, J.C., Beckmann, H., Zhou, S., Admon, A., and Tjian, R. (1994). Reconstitution of transcription factor SL1: exclusive binding of TBP by SL1 or TFIID

- subunits. *Science* 266, 1966–1972.
- Cook, W.J., Gu, B., DeLuca, N.A., Moynihan, E.B., and Coen, D.M. (1995). Induction of transcription by a viral regulatory protein depends on the relative strengths of functional TATA boxes. *Mol. Cell. Biol.* 15, 4998–5006.
- Countryman, J.K., Gradoville, L., and 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, 4706–4719.
- Court, D.L., Oppenheim, A.B., and Adhya, S.L. (2007). A New Look at Bacteriophage λ Genetic Networks. *J. Bacteriol.* 189, 298–304.
- Craigie, M., Regan, P., Otolara, Y.-L., and Sariyer, I.K. (2015). Molecular interplay between T-Antigen and splicing factor, arginine/serine-rich 1 (SRSF1) controls JC virus gene expression in glial cells. *Virol. J.* 12.
- Cuesta, R., Xi, Q., and Schneider, R.J. (2000). Adenovirus-specific translation by displacement of kinase Mnk1 from cap-initiation complex eIF4F. *EMBO J.* 19, 3465–3474.
- Cuesta, R., Xi, Q., and Schneider, R.J. (2004). Structural Basis for Competitive Inhibition of eIF4G-Mnk1 Interaction by the Adenovirus 100-Kilodalton Protein. *J. Virol.* 78, 7707–7716.
- Cumming, S.A., Chuen-Im, T., Zhang, J., and Graham, S.V. (2009). The RNA stability regulator HuR regulates L1 protein expression in vivo in differentiating cervical epithelial cells. *Virology* 383, 142–149.
- Dai, X., Gong, D., Xiao, Y., Wu, T.-T., Sun, R., and Zhou, Z.H. (2015). CryoEM and mutagenesis reveal that the smallest capsid protein cements and stabilizes Kaposi's sarcoma-associated herpesvirus capsid. *Proc. Natl. Acad. Sci. U. S. A.* 112, E649–E656.
- Daikoku, T., Kudoh, A., Fujita, M., Sugaya, Y., Isomura, H., Shirata, N., and Tsurumi, T. (2005). Architecture of Replication Compartments Formed during Epstein-Barr Virus Lytic Replication. *J. Virol.* 79, 3409–3418.
- D'Alessio, J.A., Wright, K.J., and Tjian, R. (2009). Shifting Players and Paradigms in Cell-Specific Transcription. *Mol. Cell* 36, 924–931.
- Damania, B., and Alwine, J.C. (1996). TAF-like function of SV40 large T antigen. *Genes Dev.* 10, 1369–1381.
- Damania, B., Lieberman, P., and Alwine, J.C. (1998). Simian Virus 40 Large T Antigen Stabilizes the TATA-Binding Protein-TFIIA Complex on the TATA Element. *Mol. Cell. Biol.* 18, 3926–3935.
- DANIEL, A.M., SWENSON, J.J., REDDY MAYREDDY, R.P., KHALILI, K., and

- FRISQUE, R.J. (1996). Sequences within the Early and Late Promoters of Archetype JC Virus Restrict Viral DNA Replication and Infectivity. *Virology* 216, 90–101.
- Davis, Z.H., Hesser, C., Park, J., and Glaunsinger, B.A. (2015a). Interaction Between ORF24 and ORF34 in the Kaposi's Sarcoma-Associated Herpesvirus Late Gene Transcription Factor Complex is Essential For Viral Late Gene Expression. *J. Virol.* JVI.02157–15.
- Davis, Z.H., Verschuere, E., Jang, G.M., Kleffman, K., Johnson, J.R., Park, J., Von Dollen, J., Maher, M.C., Johnson, T., Newton, W., et al. (2015b). Global mapping of herpesvirus-host protein complexes reveals a transcription strategy for late genes. *Mol. Cell* 57, 349–360.
- Deato, M.D.E., and Tjian, R. (2007). Switching of the core transcription machinery during myogenesis. *Genes Dev.* 21, 2137–2149.
- Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. *Genes Dev.* 25, 1010–1022.
- Decaussin, G., Leclerc, V., and Ooka, T. (1995). The lytic cycle of Epstein-Barr virus in the nonproducer Raji line can be rescued by the expression of a 135-kilodalton protein encoded by the BALF2 open reading frame. *J. Virol.* 69, 7309–7314.
- Delecluse, H.-J., Hilsendegen, T., Pich, D., Zeidler, R., and Hammerschmidt, W. (1998). Propagation and recovery of intact, infectious Epstein-Barr virus from prokaryotic to human cells. *Proc. Natl. Acad. Sci. U. S. A.* 95, 8245–8250.
- Dhanjal, S., Kajitani, N., Glahder, J., Mossberg, A.-K., Johansson, C., and Schwartz, S. (2015). Heterogeneous Nuclear Ribonucleoprotein C Proteins Interact with the Human Papillomavirus Type 16 (HPV16) Early 3'-Untranslated Region and Alleviate Suppression of HPV16 Late L1 mRNA Splicing. *J. Biol. Chem.* 290, 13354–13371.
- Dieci, G., Fiorino, G., Castelnuovo, M., Teichmann, M., and Pagano, A. (2007). The expanding RNA polymerase III transcriptome. *Trends Genet.* 23, 614–622.
- Dong, G., Broker, T.R., and Chow, L.T. (1994a). Human papillomavirus type 11 E2 proteins repress the homologous E6 promoter by interfering with the binding of host transcription factors to adjacent elements. *J. Virol.* 68, 1115–1127.
- Dong, X.P., Stubenrauch, F., Beyer-Finkler, E., and Pfister, H. (1994b). Prevalence of deletions of YY1-binding sites in episomal HPV 16 DNA from cervical cancers. *Int. J. Cancer* 58, 803–808.
- Doorbar, J., Quint, W., Banks, L., Bravo, I.G., Stoler, M., Broker, T.R., and Stanley, M.A. (2012). The Biology and Life-Cycle of Human Papillomaviruses. *Vaccine* 30, Supplement 5, F55–F70.
- Dostatni, N., Lambert, P.F., Sousa, R., Ham, J., Howley, P.M., and Yaniv, M. (1991). The functional BPV-1 E2 trans-activating protein can act as a repressor by preventing formation of the initiation complex. *Genes Dev.* 5, 1657–1671.

- Dubrovskaya, V., Lavigne, A.C., Davidson, I., Acker, J., Staub, A., and Tora, L. (1996). Distinct domains of hTAFII100 are required for functional interaction with transcription factor TFIIF beta (RAP30) and incorporation into the TFIID complex. *EMBO J.* *15*, 3702–3712.
- Duttke, S.H.C. (2015). Evolution and Diversification of the Basal Transcription Machinery. *Trends Biochem. Sci.* *40*, 127–129.
- Dynan, W.S., and Tjian, R. (1983a). Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. *Cell* *32*, 669–680.
- Dynan, W.S., and Tjian, R. (1983b). The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell* *35*, 79–87.
- Eash, S., Manley, K., Gasparovic, M., Querbes, W., and Atwood, W.J. (2006). The human polyomaviruses. *Cell. Mol. Life Sci. CMLS* *63*, 865–876.
- Egawa, N., Egawa, K., Griffin, H., and Doorbar, J. (2015). Human Papillomaviruses; Epithelial Tropisms, and the Development of Neoplasia. *Viruses* *7*, 3863–3890.
- El-Guindy, A., Lopez-Giraldez, F., Delecluse, H.-J., McKenzie, J., and Miller, G. (2014). A Locus Encompassing the Epstein-Barr Virus bglf4 Kinase Regulates Expression of Genes Encoding Viral Structural Proteins. *PLoS Pathog.* *10*.
- Ellis, A.L., Wang, Z., Yu, X., and Mertz, J.E. (2010). Either ZEB1 or ZEB2/SIP1 Can Play a Central Role in Regulating the Epstein-Barr Virus Latent-Lytic Switch in a Cell-Type-Specific Manner. *J. Virol.* *84*, 6139–6152.
- Ernault-Lange, M., Omilli, F., O'Reilly, D.R., and May, E. (1987). Characterization of the simian virus 40 late promoter: relative importance of sequences within the 72-base-pair repeats differs before and after viral DNA replication. *J. Virol.* *61*, 167–176.
- Evans, J.D., and Hearing, P. (2003). Distinct Roles of the Adenovirus E4 ORF3 Protein in Viral DNA Replication and Inhibition of Genome Concatenation. *J. Virol.* *77*, 5295–5304.
- Farley, D.C., Brown, J.L., and Leppard, K.N. (2004). Activation of the Early-Late Switch in Adenovirus Type 5 Major Late Transcription Unit Expression by L4 Gene Products. *J. Virol.* *78*, 1782–1791.
- Farrell, M.L., and Mertz, J.E. (2002). Hormone Response Element in SV40 Late Promoter Directly Affects Synthesis of Early as Well as Late Viral RNAs. *Virology* *297*, 307–318.
- Farrell, P.J., Rowe, D.T., Rooney, C.M., and 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.
- Feng, H., Shuda, M., Chang, Y., and Moore, P.S. (2008). Clonal Integration of a Polyomavirus in Human Merkel Cell Carcinoma. *Science* *319*, 1096–1100.

- Flemington, E., and Speck, S.H. (1990a). Autoregulation of Epstein-Barr virus putative lytic switch gene BZLF1. *J. Virol.* *64*, 1227–1232.
- Flemington, E., and Speck, S.H. (1990b). Evidence for coiled-coil dimer formation by an Epstein-Barr virus transactivator that lacks a heptad repeat of leucine residues. *Proc. Natl. Acad. Sci. U. S. A.* *87*, 9459–9463.
- Frisque, R.J., Bream, G.L., and Cannella, M.T. (1984). Human polyomavirus JC virus genome. *J. Virol.* *51*, 458–469.
- Fromental, C., Kanno, M., Nomiya, H., and Chambon, P. (1988). Cooperativity and hierarchical levels of functional organization in the SV40 enhancer. *Cell* *54*, 943–953.
- Fu, T.J., Sanders, G.M., O'Donnell, M., and Geiduschek, E.P. (1996). Dynamics of DNA-tracking by two sliding-clamp proteins. *EMBO J.* *15*, 4414–4422.
- Fuchs, L.Y., Woods, M.S., and Weaver, R.F. (1983). Viral Transcription During *Autographa californica* Nuclear Polyhedrosis Virus Infection: a Novel RNA Polymerase Induced in Infected *Spodoptera frugiperda* Cells. *J. Virol.* *48*, 641–646.
- Fujii, K., Yokoyama, N., Kiyono, T., Kuzushima, K., Homma, M., Nishiyama, Y., Fujita, M., and Tsurumi, T. (2000). The Epstein-Barr Virus Pol Catalytic Subunit Physically Interacts with the BBLF4-BSLF1-BBLF2/3 Complex. *J. Virol.* *74*, 2550–2557.
- Gallia, G.L., Safak, M., and Khalili, K. (1998). Interaction of the Single-stranded DNA-binding Protein Pura with the Human Polyomavirus JC Virus Early Protein T-antigen. *J. Biol. Chem.* *273*, 32662–32669.
- Gao, Z., Krithivas, A., Finan, J.E., Semmes, O.J., Zhou, S., Wang, Y., and Hayward, S.D. (1998). The Epstein-Barr Virus Lytic Transactivator Zta Interacts with the Helicase-Primase Replication Proteins. *J. Virol.* *72*, 8559–8567.
- Garrity, D.B., Chang, M.-J., and Blissard, G.W. (1997). Late Promoter Selection in the Baculovirusgp64 Envelope Fusion ProteinGene. *Virology* *231*, 167–181.
- Geiduschek, E.P., and Kassavetis, G.A. (2010). Transcription of the T4 late genes. *Virol. J.* *7*, 288.
- Gergely, L., Klein, G., and Ernberg, I. (1971a). The action of dna antagonists on epstein-barr virus (EBV)-associated early antigen (EA) in burkitt lymphoma lines. *Int. J. Cancer* *7*, 293–302.
- Gergely, L., Klein, G., and Ernberg, I. (1971b). Appearance of Epstein-Barr virus-associated antigens in infected Raji cells. *Virology* *45*, 10–21.
- Gershburg, E., and Pagano, J.S. (2002). Phosphorylation of the Epstein-Barr Virus (EBV) DNA Polymerase Processivity Factor EA-D by the EBV-Encoded Protein Kinase and Effects of the l-Riboside Benzimidazole 1263W94. *J. Virol.* *76*, 998–1003.

- Ghosh, S., Jain, A., Mukherjee, B., Habib, S., and Hasnain, S.E. (1998). The Host Factor Polyhedrin Promoter Binding Protein (PPBP) Is Involved in Transcription from the Baculovirus Polyhedrin Gene Promoter. *J. Virol.* *72*, 7484–7493.
- Giberson, A.N., Davidson, A.R., and Parks, R.J. (2012). Chromatin structure of adenovirus DNA throughout infection. *Nucleic Acids Res.* *40*, 2369–2376.
- Gidoni, D., Kadonaga, J.T., Barrera-Saldana, H., Takahashi, K., Chambon, P., and Tjian, R. (1985). Bidirectional SV40 transcription mediated by tandem Sp1 binding interactions. *Science* *230*, 511–517.
- Gong, D., Wu, N.C., Xie, Y., Feng, J., Tong, L., Brulois, K.F., Luan, H., Du, Y., Jung, J.U., Wang, C., et al. (2014). Kaposi's sarcoma-associated herpesvirus ORF18 and ORF30 are essential for late gene expression during lytic replication. *J. Virol.* JVI.00793–14.
- Goodrich, J.A., Hoey, T., Thut, C.J., Admon, A., and Tjian, R. (1993). *Drosophila* TAFII40 interacts with both a VP16 activation domain and the basal transcription factor TFIIB. *Cell* *75*, 519–530.
- Goracznik, R., and Gunderson, S.I. (2008). The Regulatory Element in the 3'-Untranslated Region of Human Papillomavirus 16 Inhibits Expression by Binding CUG-binding Protein 1. *J. Biol. Chem.* *283*, 2286–2296.
- Grassmann, K., Rapp, B., Maschek, H., Petry, K.U., and Iftner, T. (1996). Identification of a differentiation-inducible promoter in the E7 open reading frame of human papillomavirus type 16 (HPV-16) in raft cultures of a new cell line containing high copy numbers of episomal HPV-16 DNA. *J. Virol.* *70*, 2339–2349.
- Grondin, B., and DeLuca, N. (2000). Herpes Simplex Virus Type 1 ICP4 Promotes Transcription Preinitiation Complex Formation by Enhancing the Binding of TFIID to DNA. *J. Virol.* *74*, 11504–11510.
- Gross, C.H., and Shuman, S. (1998). RNA 5'-Triphosphatase, Nucleoside Triphosphatase, and Guanylyltransferase Activities of Baculovirus LEF-4 Protein. *J. Virol.* *72*, 10020–10028.
- Gruffat, H., Manet, E., Rigolet, A., and Sergeant, A. (1990a). The enhancer factor R of Epstein-Barr virus (EBV) is a sequence-specific DNA binding protein. *Nucleic Acids Res.* *18*, 6835–6843.
- Gruffat, H., Moreno, N., and Sergeant, A. (1990b). The Epstein-Barr virus (EBV) ORI_{lyt} enhancer is not B-cell specific and does not respond synergistically to the EBV transcription factors R and Z. *J. Virol.* *64*, 2810–2818.
- Gruffat, H., Batisse, J., Pich, D., Neuhierl, B., Manet, E., Hammerschmidt, W., and Sergeant, A. (2002a). Epstein-Barr Virus mRNA Export Factor EB2 Is Essential for Production of Infectious Virus. *J. Virol.* *76*, 9635–9644.
- Gruffat, H., Manet, E., and Sergeant, A. (2002b). MEF2-mediated recruitment of class II HDAC at the EBV immediate early gene BZLF1 links latency and chromatin

- remodeling. *EMBO Rep.* **3**, 141–146.
- Gruffat, H., Kadjouf, F., Mariamé, B., and Manet, E. (2012). The Epstein-Barr Virus BcRF1 Gene Product Is a TBP-Like Protein with an Essential Role in Late Gene Expression. *J. Virol.* **86**, 6023–6032.
- Gu, B., and DeLuca, N. (1994). Requirements for activation of the herpes simplex virus glycoprotein C promoter in vitro by the viral regulatory protein ICP4. *J. Virol.* **68**, 7953–7965.
- Guarino, L.A., and Summers, M.D. (1986). Interspersed Homologous DNA of *Autographa californica* Nuclear Polyhedrosis Virus Enhances Delayed-Early Gene Expression. *J. Virol.* **60**, 215–223.
- Guarino, L.A., Xu, B., Jin, J., and Dong, W. (1998). A Virus-Encoded RNA Polymerase Purified from Baculovirus-Infected Cells. *J. Virol.* **72**, 7985–7991.
- Guarino, L.A., Mistretta, T.-A., and Dong, W. (2002). DNA binding activity of the baculovirus late expression factor PP31. *Virus Res.* **90**, 187–195.
- Gunasekharan, V., Haché, G., and Laimins, L. (2012). Differentiation-Dependent Changes in Levels of C/EBP β Repressors and Activators Regulate Human Papillomavirus Type 31 Late Gene Expression. *J. Virol.* **86**, 5393–5398.
- Guzowski, J.F., and Wagner, E.K. (1993). Mutational analysis of the herpes simplex virus type 1 strict late UL38 promoter/leader reveals two regions critical in transcriptional regulation. *J. Virol.* **67**, 5098–5108.
- Guzowski, J.F., Singh, J., and Wagner, E.K. (1994). Transcriptional activation of the herpes simplex virus type 1 UL38 promoter conferred by the cis-acting downstream activation sequence is mediated by a cellular transcription factor. *J. Virol.* **68**, 7774–7789.
- Ham, J., Steger, G., and Yaniv, M. (1994). Cooperativity in vivo between the E2 transactivator and the TATA box binding protein depends on core promoter structure. *EMBO J.* **13**, 147–157.
- Hammerschmidt, W., and Sugden, B. (2013). Replication of Epstein–Barr Viral DNA. *Cold Spring Harb. Perspect. Biol.* **5**.
- Hart, D.O., Raha, T., Lawson, N.D., and Green, M.R. (2007). Initiation of zebrafish hematopoiesis by the TATA-box-binding protein-related factor, Trf3. *Nature* **450**, 1082–1085.
- Zur Hausen, H. (2009). Papillomaviruses in the causation of human cancers - a brief historical account. *Virology* **384**, 260–265.
- Hayes, B.W., Telling, G.C., Myat, M.M., Williams, J.F., and Flint, S.J. (1990). The adenovirus L4 100-kilodalton protein is necessary for efficient translation of viral late mRNA species. *J. Virol.* **64**, 2732–2742.

- He, Y., Fang, J., Taatjes, D.J., and Nogales, E. (2013). Structural visualization of key steps in human transcription initiation. *Nature* 495, 481–486.
- Hen, R., Sassone-Corsi, P., Corden, J., Gaub, M.P., and Chambon, P. (1982). Sequences upstream from the T-A-T-A box are required in vivo and in vitro for efficient transcription from the adenovirus serotype 2 major late promoter. *Proc. Natl. Acad. Sci. U. S. A.* 79, 7132–7136.
- Herbig, U., Weisshart, K., Taneja, P., and Fanning, E. (1999). Interaction of the Transcription Factor TFIID with Simian Virus 40 (SV40) Large T Antigen Interferes with Replication of SV40 DNA In Vitro. *J. Virol.* 73, 1099–1107.
- Hernandez, N. (1993). TBP, a universal eukaryotic transcription factor? *Genes Dev.* 7, 1291–1308.
- Hertz, G.Z., and Mertz, J.E. (1988). The enhancer elements and GGGCGG boxes of SV40 provide similar functions in bidirectionally promoting transcription. *Virology* 163, 579–590.
- Hiriart, E., Bardouillet, L., Manet, E., Gruffat, H., Penin, F., Montserret, R., Farjot, G., and Sergeant, A. (2003). A region of the Epstein-Barr virus (EBV) mRNA export factor EB2 containing an arginine-rich motif mediates direct binding to RNA. *J. Biol. Chem.* 278, 37790–37798.
- Hisatake, K., Ohta, T., Takada, R., Guermah, M., Horikoshi, M., Nakatani, Y., and Roeder, R.G. (1995). Evolutionary conservation of human TATA-binding-polypeptide-associated factors TAFII31 and TAFII80 and interactions of TAFII80 with other TAFs and with general transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* 92, 8195–8199.
- Hoeben, R.C., and Uil, T.G. (2013). Adenovirus DNA Replication. *Cold Spring Harb. Perspect. Biol.* 5.
- Hoffmann, A., and Roeder, R.G. (1996). Cloning and characterization of human TAF20/15. Multiple interactions suggest a central role in TFIID complex formation. *J. Biol. Chem.* 271, 18194–18202.
- Holley-Guthrie, E.A., Quinlivan, E.B., Mar, E.C., and Kenney, S. (1990). The Epstein-Barr virus (EBV) BMRF1 promoter for early antigen (EA-D) is regulated by the EBV transactivators, BRLF1 and BZLF1, in a cell-specific manner. *J. Virol.* 64, 3753–3759.
- Homa, F.L., Otal, T.M., Glorioso, J.C., and Levine, M. (1986). Transcriptional control signals of a herpes simplex virus type 1 late (gamma 2) gene lie within bases -34 to +124 relative to the 5' terminus of the mRNA. *Mol. Cell. Biol.* 6, 3652–3666.
- Homa, F.L., Glorioso, J.C., and Levine, M. (1988). A specific 15-bp TATA box promoter element is required for expression of a herpes simplex virus type 1 late gene. *Genes Dev.* 2, 40–53.

- Hong, G.K., Gulley, M.L., Feng, W.-H., Delecluse, H.-J., Holley-Guthrie, E., and Kenney, S.C. (2005). Epstein-Barr Virus Lytic Infection Contributes to Lymphoproliferative Disease in a SCID Mouse Model. *J. Virol.* *79*, 13993–14003.
- Hu, S.L., and Manley, J.L. (1981). DNA sequence required for initiation of transcription in vitro from the major late promoter of adenovirus 2. *Proc. Natl. Acad. Sci. U. S. A.* *78*, 820–824.
- Hu, Y., and Li, L.-L. (2013). The p35 and iel1 of *Autographa californica* multiple nucleopolyhedrovirus could rescue late gene expression of *Plutella xylostella* granulovirus in nonpermissive cell lines. *Virus Genes* *48*, 343–355.
- Huang, C.J., Goodart, S.A., Rice, M.K., Guzowski, J.F., and Wagner, E.K. (1993). Mutational analysis of sequences downstream of the TATA box of the herpes simplex virus type 1 major capsid protein (VP5/UL19) promoter. *J. Virol.* *67*, 5109–5116.
- Huang, W., Kiefer, J., Whalen, D., and Flint, S.J. (2003). DNA synthesis-dependent relief of repression of transcription from the adenovirus type 2 IVa2 promoter by a cellular protein. *Virology* *314*, 394–402.
- Huh, N.E., and Weaver, R.F. (1990). Identifying the RNA Polymerases that Synthesize Specific Transcripts of the *Autographa Californica* Nuclear Polyhedrosis Virus. *J. Gen. Virol.* *71*, 195–201.
- Hummel, M., Hudson, J.B., and Laimins, L.A. (1992). Differentiation-induced and constitutive transcription of human papillomavirus type 31b in cell lines containing viral episomes. *J. Virol.* *66*, 6070–6080.
- Hurst, H.C., and Jones, N.C. (1987). Identification of factors that interact with the E1A-inducible adenovirus E3 promoter. *Genes Dev.* *1*, 1132–1146.
- Iempridee, T., Reusch, J.A., Riching, A., Johannsen, E.C., Dovat, S., Kenney, S.C., and Mertz, J.E. (2014). Epstein-Barr Virus Utilizes Ikaros in Regulating Its Latent-Lytic Switch in B Cells. *J. Virol.* *88*, 4811–4827.
- Iftode, C., and Flint, S.J. (2004). Viral DNA synthesis-dependent titration of a cellular repressor activates transcription of the human adenovirus type 2 IVa2 gene. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 17831–17836.
- Ishiji, T., Lace, M.J., Parkkinen, S., Anderson, R.D., Haugen, T.H., Cripe, T.P., Xiao, J.H., Davidson, I., Chambon, P., and Turek, L.P. (1992). Transcriptional enhancer factor (TEF)-1 and its cell-specific co-activator activate human papillomavirus-16 E6 and E7 oncogene transcription in keratinocytes and cervical carcinoma cells. *EMBO J.* *11*, 2271–2281.
- Isogai, Y., Keles, S., Prestel, M., Hochheimer, A., and Tjian, R. (2007). Transcription of histone gene cluster by differential core-promoter factors. *Genes Dev.* *21*, 2936–2949.
- Isomura, H., Stinski, M.F., Kudoh, A., Murata, T., Nakayama, S., Sato, Y., Iwahori, S., and Tsurumi, T. (2008). Noncanonical TATA Sequence in the UL44 Late Promoter of

- Human Cytomegalovirus Is Required for the Accumulation of Late Viral Transcripts. *J. Virol.* *82*, 1638–1646.
- Isomura, H., Stinski, M.F., Murata, T., Yamashita, Y., Kanda, T., Toyokuni, S., and Tsurumi, T. (2011). The Human Cytomegalovirus Gene Products Essential for Late Viral Gene Expression Assemble into Prereplication Complexes before Viral DNA Replication. *J. Virol.* *85*, 6629–6644.
- Jacobson, R.H., Ladurner, A.G., King, D.S., and Tjian, R. (2000). Structure and Function of a Human TAFII250 Double Bromodomain Module. *Science* *288*, 1422–1425.
- Jacq, X., Brou, C., Lutz, Y., Davidson, I., Chambon, P., and Tora, L. (1994). Human TAFII30 is present in a distinct TFIID complex and is required for transcriptional activation by the estrogen receptor. *Cell* *79*, 107–117.
- Jallow, Z., Jacobi, U.G., Weeks, D.L., Dawid, I.B., and Veenstra, G.J.C. (2004). Specialized and redundant roles of TBP and a vertebrate-specific TBP paralog in embryonic gene regulation in *Xenopus*. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 13525–13530.
- Jehle, J.A., Blissard, G.W., Bonning, B.C., Cory, J.S., Herniou, E.A., Rohrmann, G.F., Theilmann, D.A., Thiem, S.M., and Vlcek, J.M. (2006). On the classification and nomenclature of baculoviruses: a proposal for revision. *Arch. Virol.* *151*, 1257–1266.
- Jenkins, P.J., Binné, U.K., and Farrell, P.J. (2000). Histone Acetylation and Reactivation of Epstein-Barr Virus from Latency. *J. Virol.* *74*, 710–720.
- Jia, Q., Wu, T.-T., Liao, H.-I., Chernishof, V., and Sun, R. (2004). Murine Gammaherpesvirus 68 Open Reading Frame 31 Is Required for Viral Replication. *J. Virol.* *78*, 6610–6620.
- Jia, R., Liu, X., Tao, M., Kruhlak, M., Guo, M., Meyers, C., Baker, C.C., and Zheng, Z.-M. (2009). Control of the Papillomavirus Early-to-Late Switch by Differentially Expressed SRp20. *J. Virol.* *83*, 167–180.
- Johannessen, M., Myhre, M.R., Dragset, M., Tümmler, C., and Moens, U. (2008). Phosphorylation of human polyomavirus BK agnoprotein at Ser-11 is mediated by PKC and has an important regulative function. *Virology* *379*, 97–109.
- Johansson, C., and Schwartz, S. (2013). Regulation of human papillomavirus gene expression by splicing and polyadenylation. *Nat. Rev. Microbiol.* *11*, 239–251.
- Johansson, C., Somberg, M., Li, X., Backström Winquist, E., Fay, J., Ryan, F., Pim, D., Banks, L., and Schwartz, S. (2012). HPV-16 E2 contributes to induction of HPV-16 late gene expression by inhibiting early polyadenylation. *EMBO J.* *31*, 3212–3227.
- Johnson, P.A., and Everett, R.D. (1986). The control of herpes simplex virus type-1 late gene transcription: a “TATA-box”/cap site region is sufficient for fully efficient regulated activity. *Nucleic Acids Res.* *14*, 8247–8264.
- Johnston, S.D., Yu, X.M., and Mertz, J.E. (1996). The major transcriptional transactivation

- domain of simian virus 40 large T antigen associates nonconcurrently with multiple components of the transcriptional preinitiation complex. *J. Virol.* *70*, 1191–1202.
- Juillard, F., Hiriart, E., Sergeant, N., Vingtdoux-Didier, V., Drobecq, H., Sergeant, A., Manet, E., and Gruffat, H. (2009). Epstein-Barr virus protein EB2 contains an N-terminal transferable nuclear export signal that promotes nucleocytoplasmic export by directly binding TAP/NXF1. *J. Virol.* *83*, 12759–12768.
- Juillard, F., Bazot, Q., Mure, F., Tafforeau, L., Macri, C., Rabourdin-Combe, C., Lotteau, V., Manet, E., and Gruffat, H. (2012). Epstein-Barr virus protein EB2 stimulates cytoplasmic mRNA accumulation by counteracting the deleterious effects of SRp20 on viral mRNAs. *Nucleic Acids Res.* *40*, 6834–6849.
- Kadonaga, J.T. (2002). The DPE, a core promoter element for transcription by RNA polymerase II. *Exp. Mol. Med.* *34*, 259–264.
- Kadonaga, J.T. (2012). Perspectives on the RNA Polymerase II Core Promoter. *Wiley Interdiscip. Rev. Dev. Biol.* *1*, 40–51.
- Kalla, M., Schmeinck, A., Bergbauer, M., Pich, D., and 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.
- Kalla, M., Göbel, C., and Hammerschmidt, W. (2012). The Lytic Phase of Epstein-Barr Virus Requires a Viral Genome with 5-Methylcytosine Residues in CpG Sites. *J. Virol.* *86*, 447–458.
- Kanno, M., Fromental, C., Staub, A., Ruffenach, F., Davidson, I., and Chambon, P. (1989). The SV40 TC-II(kappa B) and the related H-2Kb enhansons exhibit different cell type specific and inducible proto-enhancer activities, but the SV40 core sequence and the AP-2 binding site have no enhanson properties. *EMBO J.* *8*, 4205–4214.
- Kanopka, A., Mühlemann, O., and Akusjärvi, G. (1996). Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature* *381*, 535–538.
- Kanopka, A., Mühlemann, O., Petersen-Mahrt, S., Estmer, C., Öhrmalm, C., and Akusjärvi, G. (1998). Regulation of adenovirus alternative RNA splicing by dephosphorylation of SR proteins. *Nature* *393*, 185–187.
- Kasai, Y., Chen, H., and Flint, S.J. (1992). Anatomy of an unusual RNA polymerase II promoter containing a downstream TATA element. *Mol. Cell. Biol.* *12*, 2884–2897.
- Kassavetis, G.A., Elliott, T., Rabussay, D.P., and Geiduschek, E.P. (1983). Initiation of transcription at phage T4 late promoters with purified RNA polymerase. *Cell* *33*, 887–897.
- Kawashima, D., Kanda, T., Murata, T., Saito, S., Sugimoto, A., Narita, Y., and Tsurumi, T. (2013). Nuclear Transport of Epstein-Barr Virus DNA Polymerase Is Dependent on the BMRF1 Polymerase Processivity Factor and Molecular Chaperone Hsp90. *J. Virol.* *87*, 6482–6491.

- Kedmi, A., Zehavi, Y., Glick, Y., Orenstein, Y., Ideses, D., Wachtel, C., Doniger, T., Waldman Ben-Asher, H., Muster, N., Thompson, J., et al. (2014). *Drosophila* TRF2 is a preferential core promoter regulator. *Genes Dev.* *28*, 2163–2174.
- Keller, J.M., and Alwine, J.C. (1984). Activation of the SV40 late promoter: Direct effects of T antigen in the absence of viral DNA replication. *Cell* *36*, 381–389.
- Kelly, T.J. (1988). SV40 DNA replication. *J. Biol. Chem.* *263*, 17889–17892.
- Kennedy, I.M., Haddow, J.K., and Clements, J.B. (1991). A negative regulatory element in the human papillomavirus type 16 genome acts at the level of late mRNA stability. *J. Virol.* *65*, 2093–2097.
- Kenney, S., Natarajan, V., Strike, D., Khoury, G., and Salzman, N.P. (1984). JC virus enhancer-promoter active in human brain cells. *Science* *226*, 1337–1339.
- Kerr, D., Chang, C.F., Chen, N., Gallia, G., Raj, G., Schwartz, B., and Khalili, K. (1994). Transcription of a human neurotropic virus promoter in glial cells: effect of YB-1 on expression of the JC virus late gene. *J. Virol.* *68*, 7637–7643.
- Khalili, K., Rappaport, J., and Khoury, G. (1988). Nuclear factors in human brain cells bind specifically to the JCV regulatory region. *EMBO J.* *7*, 1205–1210.
- Kibler, P.K., Duncan, J., Keith, B.D., Hupel, T., and Smiley, J.R. (1991). Regulation of herpes simplex virus true late gene expression: sequences downstream from the US11 TATA box inhibit expression from an unreplicated template. *J. Virol.* *65*, 6749–6760.
- Kim, D.-B., Zabierowski, S., and DeLuca, N.A. (2002). The Initiator Element in a Herpes Simplex Virus Type 1 Late-Gene Promoter Enhances Activation by ICP4, Resulting in Abundant Late-Gene Expression. *J. Virol.* *76*, 1548–1558.
- Kim, J.L., Nikolov, D.B., and Burley, S.K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* *365*, 520–527.
- Knebel-Mörsdorf, D., Quadt, I., Li, Y., Montier, L., and Guarino, L.A. (2006). Expression of Baculovirus Late and Very Late Genes Depends on LEF-4, a Component of the Viral RNA Polymerase Whose Guanyltransferase Function Is Essential. *J. Virol.* *80*, 4168–4173.
- Knowles, W.A., Pipkin, P., Andrews, N., Vyse, A., Minor, P., Brown, D.W.G., and Miller, E. (2003). Population-based study of antibody to the human polyomaviruses BKV and JCV and the simian polyomavirus SV40. *J. Med. Virol.* *71*, 115–123.
- Koch, T., Raudonikiene, A., Wilkens, K., and Rüger, W. (1995). Overexpression, purification, and characterization of the ADP-ribosyltransferase (gpAlt) of bacteriophage T4: ADP-ribosylation of *E. coli* RNA polymerase modulates T4 “early” transcription. *Gene Expr.* *4*, 253–264.
- Kolesky, S., Ouhammouch, M., Brody, E.N., and Geiduschek, E.P. (1999). Sigma Competition: The Contest between Bacteriophage T4 Middle and Late Transcription.

- J. Mol. Biol. 291, 267–281.
- Kolesky, S.E., Ouhammouch, M., and Peter Geiduschek, E. (2002). The Mechanism of Transcriptional Activation by the Topologically DNA-linked Sliding Clamp of Bacteriophage T4. *J. Mol. Biol.* 321, 767–784.
- Kornuc, M., Kliewer, S., Garcia, J., Harrich, D., Li, C., and Gaynor, R. (1990). Adenovirus early region 3 promoter regulation by E1A/E1B is independent of alterations in DNA binding and gene activation of CREB/ATF and AP1. *J. Virol.* 64, 2004–2013.
- Krüppel, U., Müller-Schiffmann, A., Baldus, S.E., Smola-Hess, S., and Steger, G. (2008). E2 and the co-activator p300 can cooperate in activation of the human papillomavirus type 16 early promoter. *Virology* 377, 151–159.
- Kukimoto, I., and Kanda, T. (2001). Displacement of YY1 by differentiation-specific transcription factor hSkn-1a activates the P(670) promoter of human papillomavirus type 16. *J. Virol.* 75, 9302–9311.
- Kukimoto, I., Takeuchi, T., and Kanda, T. (2006). CCAAT/enhancer binding protein β binds to and activates the P670 promoter of human papillomavirus type 16. *Virology* 346, 98–107.
- Kyo, S., Inoue, M., Nishio, Y., Nakanishi, K., Akira, S., Inoue, H., Yutsudo, M., Tanizawa, O., and Hakura, A. (1993). NF-IL6 represses early gene expression of human papillomavirus type 16 through binding to the noncoding region. *J. Virol.* 67, 1058–1066.
- Lagatie, O., Tritsmans, L., and Stuyver, L.J. (2013). The miRNA world of polyomaviruses. *Virol. J.* 10, 268.
- Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D., and Ebright, R.H. (1998). New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev.* 12, 34–44.
- Lan, S., Östberg, S., Punga, T., and Akusjärvi, G. (2015). A suppressive effect of Sp1 recruitment to the first leader 5' splice site region on L4-22K-mediated activation of the adenovirus major late promoter. *Virus Res.* 210, 133–140.
- Landais, I., Vincent, R., Bouton, M., Devauchelle, G., Duonor-Cerutti, M., and Ogliastro, M. (2006). Functional analysis of evolutionary conserved clustering of bZIP binding sites in the baculovirus homologous regions (hrs) suggests a cooperativity between host and viral transcription factors. *Virology* 344, 421–431.
- Lashgari, M.S., Tada, H., Amini, S., and Khalili, K. (1989). Regulation of JCVL promoter function: Transactivation of JCVL promoter by JCV and SV40 early proteins. *Virology* 170, 292–295.
- Lauberth, S.M., Nakayama, T., Wu, X., Ferris, A., Tang, Z., Hughes, S.H., and Roeder, R.G. (2013). H3K4me3 Interactions with TAF3 Regulate Preinitiation Complex Assembly and Selective Gene Activation. *Cell* 152, 1021–1036.

- Lee, H.-J., and Chang, C. (1995). Identification of Human TR2 Orphan Receptor Response Element in the Transcriptional Initiation Site of the Simian Virus 40 Major Late Promoter. *J. Biol. Chem.* *270*, 5434–5440.
- Lee, D.C., Roeder, R.G., and Wold, W.S. (1982). DNA sequences affecting specific initiation of transcription in vitro from the EIII promoter of adenovirus 2. *Proc. Natl. Acad. Sci. U. S. A.* *79*, 41–45.
- Lee, H.-J., Lee, Y., Burbach, J.P.H., and Chang, C. (1995). Suppression of Gene Expression on the Simian Virus 40 Major Late Promoter by Human TR4 Orphan Receptor A MEMBER OF THE STEROID RECEPTOR SUPERFAMILY. *J. Biol. Chem.* *270*, 30129–30133.
- Lee, W., Mitchell, P., and Tjian, R. (1987). Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. *Cell* *49*, 741–752.
- Leff, T., Corden, J., Elkaim, R., and Sassone-Corsi, P. (1985). Transcriptional analysis of the adenovirus-5 EIII promoter: absence of sequence specificity for stimulation by E1a gene products. *Nucleic Acids Res.* *13*, 1209–1221.
- Leong, K., Lee, W., and Berk, A.J. (1990). High-level transcription from the adenovirus major late promoter requires downstream binding sites for late-phase-specific factors. *J. Virol.* *64*, 51–60.
- Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. *Nature* *424*, 147–151.
- Levine, M., Cattoglio, C., and Tjian, R. (2014). Looping Back to Leap Forward: Transcription Enters a New Era. *Cell* *157*, 13–25.
- Li, L., Harwood, S.H., and Rohrmann, G.F. (1999). Identification of Additional Genes That Influence Baculovirus Late Gene Expression. *Virology* *255*, 9–19.
- Li, X., Johansson, C., Cardoso Palacios, C., Mossberg, A., Dhanjal, S., Bergvall, M., and Schwartz, S. (2013a). Eight nucleotide substitutions inhibit splicing to HPV-16 3'-splice site SA3358 and reduce the efficiency by which HPV-16 increases the life span of primary human keratinocytes. *PloS One* *8*, e72776.
- Li, X., Johansson, C., Glahder, J., Mossberg, A.-K., and Schwartz, S. (2013b). Suppression of HPV-16 late L1 5'-splice site SD3632 by binding of hnRNP D proteins and hnRNP A2/B1 to upstream AUAGUA RNA motifs. *Nucleic Acids Res.* *41*, 10488–10508.
- Liang, B., Tikhonovich, I., Nasheuer, H.P., and Folk, W.R. (2012). Stimulation of BK Virus DNA Replication by NFI Family Transcription Factors. *J. Virol.* *86*, 3264–3275.
- Liang, C.-L., Chen, J.-L., Hsu, Y.-P.P., Ou, J.T., and Chang, Y.-S. (2002). Epstein-Barr virus BZLF1 gene is activated by transforming growth factor- β through cooperativity of Smads and c-Jun/c-Fos proteins. *J. Biol. Chem.*
- Lieberman, P. (1994). Identification of functional targets of the Zta transcriptional activator

- by formation of stable preinitiation complex intermediates. *Mol. Cell. Biol.* *14*, 8365–8375.
- Lieberman, P.M., and Berk, A.J. (1990). In vitro transcriptional activation, dimerization, and DNA-binding specificity of the Epstein-Barr virus Zta protein. *J. Virol.* *64*, 2560–2568.
- Lieberman, P.M., and Berk, A.J. (1991). The Zta trans-activator protein stabilizes TFIID association with promoter DNA by direct protein-protein interaction. *Genes Dev.* *5*, 2441–2454.
- Lieberman, P.M., and Berk, A.J. (1994). A mechanism for TAFs in transcriptional activation: activation domain enhancement of TFIID-TFIIA--promoter DNA complex formation. *Genes Dev.* *8*, 995–1006.
- Lin, C.-H., and Jarvis, D.L. (2013). Utility of temporally distinct baculovirus promoters for constitutive and baculovirus-inducible transgene expression in transformed insect cells. *J. Biotechnol.* *165*, 11–17.
- Lin, G., and Blissard, G.W. (2002). Analysis of an *Autographa californica* multicapsid nucleopolyhedrovirus *lef-6*-null virus: LEF-6 is not essential for viral replication but appears to accelerate late gene transcription. *J. Virol.* *76*, 5503–5514.
- LIU, S., BORRAS, A.M., LIU, P., SUSKE, G., and SPECK, S.H. (1997). Binding of the Ubiquitous Cellular Transcription Factors Sp1 and Sp3 to the ZI Domains in the Epstein-Barr Virus Lytic Switch BZLF1 Gene Promoter. *Virology* *228*, 11–18.
- Liu, S., Liu, P., Borrás, A., Chatila, T., and Speck, S.H. (1997). Cyclosporin A-sensitive induction of the Epstein-Barr virus lytic switch is mediated via a novel pathway involving a MEF2 family member. *EMBO J.* *16*, 143–153.
- Logan, J., and Shenk, T. (1984). Adenovirus tripartite leader sequence enhances translation of mRNAs late after infection. *Proc. Natl. Acad. Sci. U. S. A.* *81*, 3655–3659.
- Logan, J., and Shenk, T. (1986). In vivo identification of sequence elements required for normal function of the adenovirus major late transcriptional control region. *Nucleic Acids Res.* *14*, 6327–6335.
- Lu, H., Reach, M.D., Minaya, E., and Young, C.S. (1997). The initiator element of the adenovirus major late promoter has an important role in transcription initiation in vivo. *J. Virol.* *71*, 102–109.
- Luse, D.S. (2014). The RNA polymerase II preinitiation complex. *Transcription* *5*, e27050.
- Lutz, P., and Kedinger, C. (1996). Properties of the adenovirus IVa2 gene product, an effector of late-phase-dependent activation of the major late promoter. *J. Virol.* *70*, 1396–1405.
- Macchi, M., Bornert, J.M., Davidson, I., Kanno, M., Rosales, R., Vigneron, M., Xiao, J.H., Fromental, C., and Chambon, P. (1989). The SV40 TC-II(kappa B) enhancer binds ubiquitous and cell type specifically inducible nuclear proteins from lymphoid and

- non-lymphoid cell lines. *EMBO J.* 8, 4215–4227.
- Mainz, D., Quadt, I., Stranzenbach, A.K., Voss, D., Guarino, L.A., and Knebel-Mörsdorf, D. (2014). Expression and nuclear localization of the TATA-box-binding protein during baculovirus infection. *J. Gen. Virol.* 95, 1396–1407.
- Manet, E., Rigolet, A., Gruffat, H., Giot, J.F., and Sergeant, A. (1991). Domains of the Epstein-Barr virus (EBV) transcription factor R required for dimerization, DNA binding and activation. *Nucleic Acids Res.* 19, 2661–2667.
- Manley, K., O’Hara, B.A., and Atwood, W.J. (2008). Nuclear Factor of Activated T-cells (NFAT) plays a role in SV40 infection. *Virology* 372, 48–55.
- Mans, R.M.W., and Knebel-Mörsdorf, D. (1998). In Vitro Transcription of pe38/Polyhedrin Hybrid Promoters Reveals Sequences Essential for Recognition by the Baculovirus-Induced RNA Polymerase and for the Strength of Very Late Viral Promoters. *J. Virol.* 72, 2991–2998.
- Marbach-Bar, N., Bahat, A., Ashkenazi, S., Golan-Mashiach, M., Haimov, O., Wu, S.-Y., Chiang, C.-M., Puzio-Kuter, A., Hirshfield, K.M., Levine, A.J., et al. (2015). DTIE, a novel core promoter element that directs start site selection in TATA-less genes. *Nucleic Acids Res.* gkv1032.
- Martin, D.W., Subler, M.A., Muñoz, R.M., Brown, D.R., Deb, S.P., and Deb, S. (1993). p53 and SV40 T antigen bind to the same region overlapping the conserved domain of the TATA-binding protein. *Biochem. Biophys. Res. Commun.* 195, 428–434.
- Mavromara-Nazos, P., and Roizman, B. (1987). Activation of herpes simplex virus 1 γ 2 genes by viral DNA replication. *Virology* 161, 593–598.
- May, E., Omilli, F., Borde, J., and Scieller, P. (1992). Simian virus 40 T antigen activates the late promoter by modulating the activity of negative regulatory elements. *J. Virol.* 66, 3347–3354.
- McDonald, C., Karstegl, C.E., Kellam, P., and Farrell, P.J. (2010). Regulation of the Epstein-Barr virus Zp promoter in B lymphocytes during reactivation from latency. *J. Gen. Virol.* 91, 622–629.
- McLachlin, J.R., and Miller, L.K. (1994). Identification and characterization of vlf-1, a baculovirus gene involved in very late gene expression. *J. Virol.* 68, 7746–7756.
- McWatters, B.J.P., Stenberg, R.M., and Kerry, J.A. (2002). Characterization of the Human Cytomegalovirus UL75 (Glycoprotein H) Late Gene Promoter. *Virology* 303, 309–316.
- Van der Meijden, E., Janssens, R.W.A., Lauber, C., Bouwes Bavinck, J.N., Gorbalenya, A.E., and Feltkamp, M.C.W. (2010). Discovery of a New Human Polyomavirus Associated with Trichodysplasia Spinulosa in an Immunocompromized Patient. *PLoS Pathog.* 6.
- Mengus, G., May, M., Jacq, X., Staub, A., Tora, L., Chambon, P., and Davidson, I. (1995).

- Cloning and characterization of hTAFII18, hTAFII20 and hTAFII28: three subunits of the human transcription factor TFIID. *EMBO J.* *14*, 1520–1531.
- Mermod, N., Williams, T.J., and Tjian, R. (1988). Enhancer binding factors AP-4 and AP-1 act in concert to activate SV40 late transcription in vitro. *Nature* *332*, 557–561.
- Milavetz, B.I. (2002). SP1 and AP-1 Elements Direct Chromatin Remodeling in SV40 Chromosomes during the First 6 Hours of Infection. *Virology* *294*, 170–179.
- Miller, E.S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., and Rüger, W. (2003). Bacteriophage T4 Genome. *Microbiol. Mol. Biol. Rev.* *67*, 86–156.
- Milligan, S.G., Veerapraditsin, T., Ahamet, B., Mole, S., and Graham, S.V. (2007). Analysis of novel human papillomavirus type 16 late mRNAs in differentiated W12 cervical epithelial cells. *Virology* *360*, 172–181.
- Mishra, G., Gautam, H.K., and Das, R.H. (2007). Serine/Threonine kinase dependent transcription from the polyhedrin promoter of SpltNPV-I. *Biochem. Biophys. Res. Commun.* *358*, 942–947.
- Mishra, G., Chadha, P., and Das, R.H. (2008a). Serine/threonine kinase (pk-1) is a component of *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV) very late gene transcription complex and it phosphorylates a 102 kDa polypeptide of the complex. *Virus Res.* *137*, 147–149.
- Mishra, G., Chadha, P., Chaudhury, I., and Das, R.H. (2008b). Inhibition of *Autographa californica* nucleopolyhedrovirus (AcNPV) polyhedrin gene expression by DNAzyme knockout of its serine/threonine kinase (pk1) gene. *Virus Res.* *135*, 197–201.
- Mistretta, T.-A., and Guarino, L.A. (2005). Transcriptional Activity of Baculovirus Very Late Factor 1. *J. Virol.* *79*, 1958–1960.
- Mitchell, P.J., Wang, C., and Tjian, R. (1987). Positive and negative regulation of transcription in vitro: Enhancer-binding protein AP-2 is inhibited by SV40 T antigen. *Cell* *50*, 847–861.
- Moarefi, I., Jeruzalmi, D., Turner, J., O'Donnell, M., and Kuriyan, J. (2000). Crystal structure of the DNA polymerase processivity factor of T4 bacteriophage. *J. Mol. Biol.* *296*, 1215–1223.
- Moens, U., and Van Ghelue, M. (2005). Polymorphism in the genome of non-passaged human polyomavirus BK: implications for cell tropism and the pathological role of the virus. *Virology* *331*, 209–231.
- Moens, U., Subramaniam, N., Johansen, B., Johansen, T., and Traavik, T. (1994). A steroid hormone response unit in the late leader of the noncoding control region of the human polyomavirus BK confers enhanced host cell permissivity. *J. Virol.* *68*, 2398–2408.
- Moens, U., Johansen, T., Johnsen, J.I., Seternes, O.M., and Traavik, T. (1995). Noncoding control region of naturally occurring BK virus variants: Sequence comparison and

- functional analysis. *Virus Genes* *10*, 261–275.
- Mole, S., Milligan, S.G., and Graham, S.V. (2009). Human Papillomavirus Type 16 E2 Protein Transcriptionally Activates the Promoter of a Key Cellular Splicing Factor, SF2/ASF. *J. Virol.* *83*, 357–367.
- Mondésert, G., and Kédinger, C. (1991). Cooperation between upstream and downstream elements of the adenovirus major late promoter for maximal late phase-specific transcription. *Nucleic Acids Res.* *19*, 3221–3228.
- Mondesert, G., Tribouley, C., and Keding, C. (1992). Identification of a novel downstream binding protein implicated in late-phase-specific activation of the adenovirus major late promoter. *Nucleic Acids Res.* *20*, 3881–3889.
- Montalvo, E.A., Cottam, M., Hill, S., and Wang, Y.J. (1995). YY1 binds to and regulates cis-acting negative elements in the Epstein-Barr virus BZLF1 promoter. *J. Virol.* *69*, 4158–4165.
- Morris, S.J., and Leppard, K.N. (2009). Adenovirus Serotype 5 L4-22K and L4-33K Proteins Have Distinct Functions in Regulating Late Gene Expression. *J. Virol.* *83*, 3049–3058.
- Morris, T.D., and Miller, L.K. (1994). Mutational analysis of a baculovirus major late promoter. *Gene* *140*, 147–153.
- Morris, S.J., Scott, G.E., and Leppard, K.N. (2010). Adenovirus Late-Phase Infection Is Controlled by a Novel L4 Promoter. *J. Virol.* *84*, 7096–7104.
- Mühlemann, O., Yue, B.-G., Petersen-Mahrt, S., and Akusjärvi, G. (2000). A Novel Type of Splicing Enhancer Regulating Adenovirus Pre-mRNA Splicing. *Mol. Cell. Biol.* *20*, 2317–2325.
- Müller, F., and Tora, L. (2014). Chromatin and DNA sequences in defining promoters for transcription initiation. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* *1839*, 118–128.
- Müller, F., Demény, M.A., and Tora, L. (2007). New Problems in RNA Polymerase II Transcription Initiation: Matching the Diversity of Core Promoters with a Variety of Promoter Recognition Factors. *J. Biol. Chem.* *282*, 14685–14689.
- Müller, F., Zaucker, A., and Tora, L. (2010). Developmental regulation of transcription initiation: more than just changing the actors. *Curr. Opin. Genet. Dev.* *20*, 533–540.
- Murakami, K., Elmlund, H., Kalisman, N., Bushnell, D.A., Adams, C.M., Azubel, M., Elmlund, D., Levi-Kalisman, Y., Liu, X., Gibbons, B.J., et al. (2013). Architecture of an RNA Polymerase II Transcription Pre-Initiation Complex. *Science* *342*, 1238724.
- Murata, T., Kondo, Y., Sugimoto, A., Kawashima, D., Saito, S., Isomura, H., Kanda, T., and Tsurumi, T. (2012). Epigenetic Histone Modification of Epstein-Barr Virus BZLF1 Promoter during Latency and Reactivation in Raji Cells. *J. Virol.* *86*, 4752–4761.

- Myers, R.M., Rio, D.C., Robbins, A.K., and Tjian, R. (1981). SV40 gene expression is modulated by the cooperative binding of T antigen to DNA. *Cell* 25, 373–384.
- Natarajan, V., and Salzman, N.P. (1985). Cis and trans activation of adenovirus IVa2 gene transcription. *Nucleic Acids Res.* 13, 4067–4083.
- Nawandar, D.M., Wang, A., Makielski, K., Lee, D., Ma, S., Barlow, E., Reusch, J., Jiang, R., Wille, C.K., Greenspan, D., et al. (2015). Differentiation-Dependent KLF4 Expression Promotes Lytic Epstein-Barr Virus Infection in Epithelial Cells. *PLoS Pathog.* 11.
- Nechaev, S., and Geiduschek, E.P. (2008). Dissection of the Bacteriophage T4 Late Promoter Complex. *J. Mol. Biol.* 379, 402–413.
- Nilsson, C.E., Petersen-Mahrt, S., Durot, C., Shtrichman, R., Krainer, A.R., Kleinberger, T., and Akusjärvi, G. (2001). The adenovirus E4-ORF4 splicing enhancer protein interacts with a subset of phosphorylated SR proteins. *EMBO J.* 20, 864–871.
- Nordqvist, K., and Akusjärvi, G. (1990). Adenovirus early region 4 stimulates mRNA accumulation via 5' introns. *Proc. Natl. Acad. Sci. U. S. A.* 87, 9543–9547.
- Nordqvist, K., Ohman, K., and Akusjärvi, G. (1994). Human adenovirus encodes two proteins which have opposite effects on accumulation of alternatively spliced mRNAs. *Mol. Cell. Biol.* 14, 437–445.
- Öberg, D., Fay, J., Lambkin, H., and Schwartz, S. (2005). A Downstream Polyadenylation Element in Human Papillomavirus Type 16 L2 Encodes Multiple GGG Motifs and Interacts with hnRNP H. *J. Virol.* 79, 9254–9269.
- O'Connor, M., and Bernard, H.-U. (1995). Oct-1 Activates the Epithelial-Specific Enhancer of Human Papillomavirus Type 16 via a Synergistic Interaction with NFI at a Conserved Composite Regulatory Element. *Virology* 207, 77–88.
- O'Connor, M.J., Stünkel, W., Koh, C.-H., Zimmermann, H., and Bernard, H.-U. (2000). The Differentiation-Specific Factor CDP/Cut Represses Transcription and Replication of Human Papillomaviruses through a Conserved Silencing Element. *J. Virol.* 74, 401–410.
- Oelgeschläger, T., Chiang, C.-M., and Roeder, R.G. (1996). Topology and reorganization of a human TFIID-promoter complex. *Nature* 382, 735–738.
- Öhman, K., Nordqvist, K., and Akusjärvi, G. (1993). Two Adenovirus Proteins with Redundant Activities in Virus Growth Facilitates Tripartite Leader mRNA Accumulation. *Virology* 194, 50–58.
- Ohno, S., Steer, B., Sattler, C., and Adler, H. (2012). ORF23 of murine gammaherpesvirus 68 is non-essential for in vitro and in vivo infection. *J. Gen. Virol.* 93, 1076–1080.
- Omoto, S., and Mocarski, E.S. (2013). Cytomegalovirus UL91 Is Essential for Transcription of Viral True Late (γ 2) Genes. *J. Virol.* 87, 8651–8664.

- Omoto, S., and Mocarski, E.S. (2014). Transcription of True Late (γ 2) Cytomegalovirus Genes Requires UL92 Function That Is Conserved among Beta- and Gammaherpesviruses. *J. Virol.* *88*, 120–130.
- Ondek, B., Gloss, L., and Herr, W. (1988). The SV40 enhancer contains two distinct levels of organization. *Nature* *333*, 40–45.
- Ooi, B.G., Rankin, C., and Miller, L.K. (1989). Downstream sequences augment transcription from the essential initiation site of a baculovirus polyhedrin gene. *J. Mol. Biol.* *210*, 721–736.
- Ostapchuk, P., Anderson, M.E., Chandrasekhar, S., and Hearing, P. (2006). The L4 22-Kilodalton Protein Plays a Role in Packaging of the Adenovirus Genome. *J. Virol.* *80*, 6973–6981.
- Ouhammouch, M., Adelman, K., Harvey, S.R., Orsini, G., and Brody, E.N. (1995). Bacteriophage T4 MotA and AsiA proteins suffice to direct Escherichia coli RNA polymerase to initiate transcription at T4 middle promoters. *Proc. Natl. Acad. Sci. U. S. A.* *92*, 1451–1455.
- Pardo-Mateos, A., and Young, C.S.H. (2004). Adenovirus IVa2 protein plays an important role in transcription from the major late promoter in vivo. *Virology* *327*, 50–59.
- Parks, C.L., and Shenk, T. (1997). Activation of the adenovirus major late promoter by transcription factors MAZ and Sp1. *J. Virol.* *71*, 9600–9607.
- Paul Thomas, G., and Mathews, M.B. (1980). DNA replication and the early to late transition in adenovirus infection. *Cell* *22*, 523–533.
- Pearson, M., Bjornson, R., Pearson, G., and Rohrmann, G. (1992). The *Autographa californica* baculovirus genome: evidence for multiple replication origins. *Science* *257*, 1382–1384.
- Perng, Y.-C., Qian, Z., Fehr, A.R., Xuan, B., and Yu, D. (2011). The Human Cytomegalovirus Gene UL79 Is Required for the Accumulation of Late Viral Transcripts ∇ . *J. Virol.* *85*, 4841–4852.
- Perng, Y.-C., Campbell, J.A., Lenschow, D.J., and Yu, D. (2014). Human cytomegalovirus pUL79 is an elongation factor of RNA polymerase II for viral gene transcription. *PLoS Pathog.* *10*, e1004350.
- Persengiev, S.P., Zhu, X., Dixit, B.L., Maston, G.A., Kittler, E.L.W., and Green, M.R. (2003). TRF3, a TATA-box-binding protein-related factor, is vertebrate-specific and widely expressed. *Proc. Natl. Acad. Sci. U. S. A.* *100*, 14887–14891.
- Peter Verrijzer, C., and Tjian, R. (1996). TAFs mediate transcriptional activation and promoter selectivity. *Trends Biochem. Sci.* *21*, 338–342.
- Petroski, M.D., Devi-Rao, G.B., Rice, M.K., and Wagner, E.K. (2001). The Downstream Activation Sequence of the Strict Late Herpes Simplex Virus Type 1 UL38 Promoter

- Interacts With hTAFII70, a Component of TFIID. *Virus Genes* 22, 299–310.
- Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Séraphin, B. (2001). The Tandem Affinity Purification (TAP) Method: A General Procedure of Protein Complex Purification. *Methods* 24, 218–229.
- Pullen, S.S., and Friesen, P.D. (1995a). The CAGT motif functions as an initiator element during early transcription of the baculovirus transregulator ie-1. *J. Virol.* 69, 3575–3583.
- Pullen, S.S., and Friesen, P.D. (1995b). Early transcription of the ie-1 transregulator gene of *Autographa californica* nuclear polyhedrosis virus is regulated by DNA sequences within its 5' noncoding leader region. *J. Virol.* 69, 156–165.
- Pyronnet, S., Imataka, H., Gingras, A.C., Fukunaga, R., Hunter, T., and Sonenberg, N. (1999). Human eukaryotic translation initiation factor 4G (eIF4G) recruits mnk1 to phosphorylate eIF4E. *EMBO J.* 18, 270–279.
- Quadt, I., Mainz, D., Mans, R., Kremer, A., and Knebel-Mörsdorf, D. (2002). Baculovirus Infection Raises the Level of TATA-Binding Protein That Colocalizes with Viral DNA Replication Sites. *J. Virol.* 76, 11123–11127.
- Quinlivan, E.B., Holley-Guthrie, E.A., Norris, M., Gutsch, D., Bachenheimer, S.L., and 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.
- Raff, A.B., Woodham, A.W., Raff, L.M., Skeate, J.G., Yan, L., Da Silva, D.M., Schelhaas, M., and Kast, W.M. (2013). The Evolving Field of Human Papillomavirus Receptor Research: a Review of Binding and Entry. *J. Virol.* 87, 6062–6072.
- Ramachandran, A., Jain, A., Arora, P., Bashyam, M.D., Chatterjee, U., Ghosh, S., Parnaik, V.K., and Hasnain, S.E. (2001). Novel Sp Family-like Transcription Factors Are Present in Adult Insect Cells and Are Involved in Transcription from the Polyhedrin Gene Initiator Promoter. *J. Biol. Chem.* 276, 23440–23449.
- Ramasubramanian, S., Osborn, K., Flower, K., and Sinclair, A.J. (2012). Dynamic Chromatin Environment of Key Lytic Cycle Regulatory Regions of the Epstein-Barr Virus Genome. *J. Virol.* 86, 1809–1819.
- Rankin, C., Ooi, B.G., and Miller, L.K. (1988). Eight base pairs encompassing the transcriptional start point are the major determinant for baculovirus polyhedrin gene expression. *Gene* 70, 39–49.
- Rapp, J.C., Wilson, J.A., and Miller, L.K. (1998). Nineteen Baculovirus Open Reading Frames, Including LEF-12, Support Late Gene Expression. *J. Virol.* 72, 10197–10206.
- Reach, M., Xu, L.X., and Young, C.S. (1991). Transcription from the adenovirus major late promoter uses redundant activating elements. *EMBO J.* 10, 3439–3446.

- Reina, J.H., and Hernandez, N. (2007). On a roll for new TRF targets. *Genes Dev.* *21*, 2855–2860.
- Renner, K., Leger, H., and Wegner, M. (1994). The POU domain protein Tst-1 and papovaviral large tumor antigen function synergistically to stimulate glia-specific gene expression of JC virus. *Proc. Natl. Acad. Sci.* *91*, 6433–6437.
- Ricci, E.P., Mure, F., Gruffat, H., Decimo, D., Medina-Palazon, C., Ohlmann, T., and Manet, E. (2009). Translation of intronless RNAs is strongly stimulated by the Epstein-Barr virus mRNA export factor EB2. *Nucleic Acids Res.* *37*, 4932–4943.
- Rickinson, A.B., Jarvis, J.E., Crawford, D.H., and Epstein, M.A. (1974). Observations on the type of infection by Epstein-Barr virus in peripheral lymphoid cells of patients with infectious mononucleosis. *Int. J. Cancer* *14*, 704–715.
- Rio, D.C., and Tjian, R. (1983). SV40 T antigen binding site mutations that affect autoregulation. *Cell* *32*, 1227–1240.
- Rohrmann, G.F. (2013). DNA replication and genome processing.
- Romanczuk, H., Thierry, F., and Howley, P.M. (1990). Mutational analysis of cis elements involved in E2 modulation of human papillomavirus type 16 P97 and type 18 P105 promoters. *J. Virol.* *64*, 2849–2859.
- Rooney, C.M. (2003). New insights into EBV-associated post-transplant lymphoproliferative disease. *The Lancet* *361*, 192–193.
- Rooney, C.M., Rowe, D.T., Ragot, T., and Farrell, P.J. (1989). The spliced BZLF1 gene of Epstein-Barr virus (EBV) transactivates an early EBV promoter and induces the virus productive cycle. *J. Virol.* *63*, 3109–3116.
- Ruf, I.K., and Rawlins, D.R. (1995). Identification and characterization of ZIIBC, a complex formed by cellular factors and the ZII site of the Epstein-Barr virus BZLF1 promoter. *J. Virol.* *69*, 7648–7657.
- Rush, M., Zhao, X., and Schwartz, S. (2005). A Splicing Enhancer in the E4 Coding Region of Human Papillomavirus Type 16 Is Required for Early mRNA Splicing and Polyadenylation as Well as Inhibition of Premature Late Gene Expression. *J. Virol.* *79*, 12002–12015.
- Russell, J., and Zomerdijk, J.C.B.M. (2006). The RNA polymerase I transcription machinery. *Biochem. Soc. Symp.* 203–216.
- Sadowska, B., Barrucco, R., Khalili, K., and Safak, M. (2003). Regulation of Human Polyomavirus JC Virus Gene Transcription by AP-1 in Glial Cells. *J. Virol.* *77*, 665–672.
- Safak, M., Gallia, G.L., Ansari, S.A., and Khalili, K. (1999a). Physical and Functional Interaction between the Y-Box Binding Protein YB-1 and Human Polyomavirus JC Virus Large T Antigen. *J. Virol.* *73*, 10146–10157.

- Safak, M., Gallia, G.L., and Khalili, K. (1999b). Reciprocal Interaction between Two Cellular Proteins, Pur α and YB-1, Modulates Transcriptional Activity of JCVCY in Glial Cells. *Mol. Cell. Biol.* *19*, 2712–2723.
- Safak, M., Barrucco, R., Darbinyan, A., Okada, Y., Nagashima, K., and Khalili, K. (2001). Interaction of JC Virus Agno Protein with T Antigen Modulates Transcription and Replication of the Viral Genome in Glial Cells. *J. Virol.* *75*, 1476–1486.
- Safak, M., Sadowska, B., Barrucco, R., and Khalili, K. (2002). Functional Interaction between JC Virus Late Regulatory Agnoprotein and Cellular Y-Box Binding Transcription Factor, YB-1. *J. Virol.* *76*, 3828–3838.
- Salem, T.Z., Zhang, F., Xie, Y., and Thiem, S.M. (2011). Comprehensive analysis of host gene expression in *Autographa californica* nucleopolyhedrovirus-infected *Spodoptera frugiperda* cells. *Virology* *412*, 167–178.
- Sampath, P., and DeLuca, N.A. (2008). Binding of ICP4, TATA-Binding Protein, and RNA Polymerase II to Herpes Simplex Virus Type 1 Immediate-Early, Early, and Late Promoters in Virus-Infected Cells. *J. Virol.* *82*, 2339–2349.
- Sandler, A.B., and Ketner, G. (1989). Adenovirus early region 4 is essential for normal stability of late nuclear RNAs. *J. Virol.* *63*, 624–630.
- Sariyer, I.K., and Khalili, K. (2011). Regulation of Human Neurotropic JC Virus Replication by Alternative Splicing Factor SF2/ASF in Glial Cells. *PLoS ONE* *6*.
- Sariyer, I.K., Akan, I., Palermo, V., Gordon, J., Khalili, K., and Safak, M. (2006). Phosphorylation Mutants of JC Virus Agnoprotein Are Unable To Sustain the Viral Infection Cycle. *J. Virol.* *80*, 3893–3903.
- Sariyer, I.K., Khalili, K., and Safak, M. (2008). Dephosphorylation of JC virus agnoprotein by protein phosphatase 2A: Inhibition by small t antigen. *Virology* *375*, 464–479.
- Sawadogo, M., and Roeder, R.G. (1985). Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. *Cell* *43*, 165–175.
- Schmidt, R.P., and Kreuzer, K.N. (1992). Purified MotA protein binds the -30 region of a bacteriophage T4 middle-mode promoter and activates transcription in vitro. *J. Biol. Chem.* *267*, 11399–11407.
- Schneider, M.A., Scheffer, K.D., Bund, T., Boukhallouk, F., Lambert, C., Cotarelo, C., Pflugfelder, G.O., Florin, L., and Spoden, G.A. (2013). The Transcription Factors TBX2 and TBX3 Interact with Human Papillomavirus 16 (HPV16) L2 and Repress the Long Control Region of HPVs. *J. Virol.* *87*, 4461–4474.
- Schowalter, R.M., Pastrana, D.V., Pumphrey, K.A., Moyer, A.L., and Buck, C.B. (2010). Merkel Cell Polyomavirus and Two Novel Polyomaviruses Are Chronically Shed from Human Skin. *Cell Host Microbe* *7*, 509–515.

- La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G., Merchat, M., Suzan-Monti, M., Forterre, P., Koonin, E., et al. (2008). The virophage as a unique parasite of the giant mimivirus. *Nature* 455, 100–104.
- Seif, I., Khoury, G., and Dhar, R. (1979). The genome of human papovavirus BKV. *Cell* 18, 963–977.
- Serio, T.R., Kolman, J.L., and Miller, G. (1997). Late gene expression from the Epstein-Barr virus BcLF1 and BFRF3 promoters does not require DNA replication in cis. *J. Virol.* 71, 8726–8734.
- Serio, T.R., Cahill, N., Prout, M.E., and Miller, G. (1998). A functionally distinct TATA box required for late progression through the Epstein-Barr virus life cycle. *J. Virol.* 72, 8338–8343.
- Shaw, A.R., and Ziff, E.B. (1980). Transcripts from the adenovirus-2 major late promoter yield a single early family of 3' coterminal mRNAs and five late families. *Cell* 22, 905–916.
- Shaw, J.E., Levinger, L.F., and Carter, C.W. (1979). Nucleosomal structure of Epstein-Barr virus DNA in transformed cell lines. *J. Virol.* 29, 657–665.
- Smale, S.T., and Kadonaga, J.T. (2003). The RNA Polymerase II Core Promoter. *Annu. Rev. Biochem.* 72, 449–479.
- Smith, C.A., Bates, P., Rivera-Gonzalez, R., Gu, B., and DeLuca, N.A. (1993). ICP4, the major transcriptional regulatory protein of herpes simplex virus type 1, forms a tripartite complex with TATA-binding protein and TFIIB. *J. Virol.* 67, 4676–4687.
- Smotkin, D., and Wettstein, F.O. (1986). Transcription of human papillomavirus type 16 early genes in a cervical cancer and a cancer-derived cell line and identification of the E7 protein. *Proc. Natl. Acad. Sci. U. S. A.* 83, 4680–4684.
- Sokal, N., Nie, Y., Willis, L.G., Yamagishi, J., Blissard, G.W., Rheault, M.R., and Theilmann, D.A. (2014). Defining the roles of the baculovirus regulatory proteins IE0 and IE1 in genome replication and early gene transactivation. *Virology* 468-470, 160–171.
- Sokolowski, M., Tan, W., Jellne, M., and Schwartz, S. (1998). mRNA Instability Elements in the Human Papillomavirus Type 16 L2 Coding Region. *J. Virol.* 72, 1504–1515.
- Somberg, M., and Schwartz, S. (2010). Multiple ASF/SF2 Sites in the Human Papillomavirus Type 16 (HPV-16) E4-Coding Region Promote Splicing to the Most Commonly Used 3'-Splice Site on the HPV-16 Genome. *J. Virol.* 84, 8219–8230.
- Somberg, M., Zhao, X., Fröhlich, M., Evander, M., and Schwartz, S. (2008). Polypyrimidine Tract Binding Protein Induces Human Papillomavirus Type 16 Late Gene Expression by Interfering with Splicing Inhibitory Elements at the Major Late 5' Splice Site, SD3632. *J. Virol.* 82, 3665–3678.

- Somberg, M., Li, X., Johansson, C., Orru, B., Chang, R., Rush, M., Fay, J., Ryan, F., and Schwartz, S. (2011). Serine/arginine-rich protein 30c activates human papillomavirus type 16 L1 mRNA expression via a bimodal mechanism. *J. Gen. Virol.* *92*, 2411–2421.
- Song, B., and Young, C.S.H. (1997). Functional Characterization of the Major Late Promoter of Mouse Adenovirus Type 1. *Virology* *235*, 109–117.
- SONG, B., HU, S.-L., DARAI, G., SPINDLER, K.R., and YOUNG, C.S.H. (1996). Conservation of DNA Sequence in the Predicted Major Late Promoter Regions of Selected Mastadenoviruses. *Virology* *220*, 390–401.
- Sriram, S., and Gopinathan, K.P. (1998). The potential role of a late gene expression factor, *lef2*, from *Bombyx mori* nuclear polyhedrosis virus in very late gene transcription and DNA replication. *Virology* *251*, 108–122.
- Steger, G., and Corbach, S. (1997). Dose-dependent regulation of the early promoter of human papillomavirus type 18 by the viral E2 protein. *J. Virol.* *71*, 50–58.
- Stern-Ginossar, N., Weisburd, B., Michalski, A., Le, V.T.K., Hein, M.Y., Huang, S.-X., Ma, M., Shen, B., Qian, S.-B., Hengel, H., et al. (2012). Decoding Human Cytomegalovirus. *Science* *338*, 1088–1093.
- Stillman, B.W., Lewis, J.B., Chow, L.T., Mathews, M.B., and Smart, J.E. (1981). Identification of the gene and mRNA for the adenovirus terminal protein precursor. *Cell* *23*, 497–508.
- Sugimoto, A., Sato, Y., Kanda, T., Murata, T., Narita, Y., Kawashima, D., Kimura, H., and Tsurumi, T. (2013). Different distributions of Epstein-Barr virus early and late gene transcripts within viral replication compartments. *J. Virol.* *87*, 6693–6699.
- Sullivan, C.S., Grundhoff, A.T., Tevethia, S., Pipas, J.M., and Ganem, D. (2005). SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells. *Nature* *435*, 682–686.
- Summers, W.C., and Klein, G. (1976). Inhibition of Epstein-Barr virus DNA synthesis and late gene expression by phosphonoacetic acid. *J. Virol.* *18*, 151–155.
- Sun, C.C., and Thorley-Lawson, D.A. (2007). Plasma Cell-Specific Transcription Factor XBP-1s Binds to and Transactivates the Epstein-Barr Virus BZLF1 Promoter. *J. Virol.* *81*, 13566–13577.
- Suzuki, Y., Tsunoda, T., Sese, J., Taira, H., Mizushima-Sugano, J., Hata, H., Ota, T., Isogai, T., Tanaka, T., Nakamura, Y., et al. (2001). Identification and Characterization of the Potential Promoter Regions of 1031 Kinds of Human Genes. *Genome Res.* *11*, 677–684.
- Tada, H., Lashgari, M.S., and Khalili, K. (1991). Regulation of JCV promoter function: Evidence that a pentanucleotide “silencer” repeat sequence AGGGAAGGGA down-regulates transcription of the JC virus late promoter. *Virology* *180*, 327–338.

- Takada, R., Nakatani, Y., Hoffmann, A., Kokubo, T., Hasegawa, S., Roeder, R.G., and Horikoshi, M. (1992). Identification of human TFIID components and direct interaction between a 250-kDa polypeptide and the TATA box-binding protein (TFIID tau). *Proc. Natl. Acad. Sci. U. S. A.* *89*, 11809–11813.
- Tan, S.H., Gloss, B., and Bernard, H.U. (1992). During negative regulation of the human papillomavirus-16 E6 promoter, the viral E2 protein can displace Sp1 from a proximal promoter element. *Nucleic Acids Res.* *20*, 251–256.
- Tan, S.H., Leong, L.E., Walker, P.A., and Bernard, H.U. (1994). The human papillomavirus type 16 E2 transcription factor binds with low cooperativity to two flanking sites and represses the E6 promoter through displacement of Sp1 and TFIID. *J. Virol.* *68*, 6411–6420.
- Tang, S., Yamanegi, K., and Zheng, Z.-M. (2004). Requirement of a 12-base-pair TATT-containing sequence and viral lytic DNA replication in activation of the Kaposi's sarcoma-associated herpesvirus K8.1 late promoter. *J. Virol.* *78*, 2609–2614.
- Tang, Y., Wu, H., Ugai, H., Matthews, Q.L., and Curiel, D.T. (2009). Derivation of a Triple Mosaic Adenovirus for Cancer Gene Therapy. *PLoS ONE* *4*.
- Tao, Y., Guermah, M., Martinez, E., Oelgeschläger, T., Hasegawa, S., Takada, R., Yamamoto, T., Horikoshi, M., and Roeder, R.G. (1997). Specific Interactions and Potential Functions of Human TAFII100. *J. Biol. Chem.* *272*, 6714–6721.
- Thierry, E., Brennich, M., Round, A., Buisson, M., Burmeister, W.P., and Hutin, S. (2015). Production and characterisation of Epstein–Barr virus helicase–primase complex and its accessory protein BBLF2/3. *Virus Genes* *51*, 171–181.
- Thierry, F., Heard, J.M., Dartmann, K., and Yaniv, M. (1987). Characterization of a transcriptional promoter of human papillomavirus 18 and modulation of its expression by simian virus 40 and adenovirus early antigens. *J. Virol.* *61*, 134–142.
- Thomas, J.O., and Kornberg, R.D. (1975). An octamer of histones in chromatin and free in solution. *Proc. Natl. Acad. Sci. U. S. A.* *72*, 2626–2630.
- Thomas, C., Dankesreiter, A., Wolf, H., and 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*, 959–964.
- Thompson, J., Verma, D., Li, D., Mosbrugger, T., and Swaminathan, S. (2016). Identification and Characterization of the Physiological Gene Targets of the Essential Lytic Replicative Epstein-Barr Virus SM Protein. *J. Virol.* *90*, 1206–1221.
- Thorley-Lawson, D.A. (2005). EBV the prototypical human tumor virus—just how bad is it? *J. Allergy Clin. Immunol.* *116*, 251–261.
- Thorley-Lawson, D.A., Hawkins, J.B., Tracy, S.I., and Shapiro, M. (2013). The Pathogenesis of Epstein-Barr Virus Persistent Infection. *Curr. Opin. Virol.* *3*, 227–232.

- Thumbi, D.K., Eveleigh, R.J.M., Lucarotti, C.J., Lapointe, R., Graham, R.I., Pavlik, L., Lauzon, H.A.M., and Arif, B.M. (2011). Complete Sequence, Analysis and Organization of the *Orgyia leucostigma* Nucleopolyhedrovirus Genome. *Viruses* 3, 2301–2327.
- Tiemann, B., Depping, R., Gineikiene, E., Kaliniene, L., Nivinskas, R., and Ruger, W. (2004). ModA and ModB, two ADP-ribosyltransferases encoded by bacteriophage T4: catalytic properties and mutation analysis. *J. Bacteriol.* 186, 7262–7272.
- Timms, J.M., Bell, A., Flavell, J.R., Murray, P.G., Rickinson, A.B., Traverse-Glehen, A., Berger, F., and Delecluse, H.-J. (2003). Target cells of Epstein-Barr-virus (EBV)-positive post-transplant lymphoproliferative disease: similarities to EBV-positive Hodgkin’s lymphoma. *The Lancet* 361, 217–223.
- Todd, J.W., Passarelli, A.L., and Miller, L.K. (1995). Eighteen baculovirus genes, including lef-11, p35, 39K, and p47, support late gene expression. *J. Virol.* 69, 968–974.
- Todd, J.W., Passarelli, A.L., Lu, A., and Miller, L.K. (1996). Factors regulating baculovirus late and very late gene expression in transient-expression assays. *J. Virol.* 70, 2307–2317.
- Tora, L. (2002). A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. *Genes Dev.* 16, 673–675.
- Tormanen, H., Backstrom, E., Carlsson, A., and Akusjarvi, G. (2006). L4-33K, an Adenovirus-encoded Alternative RNA Splicing Factor. *J. Biol. Chem.* 281, 36510–36517.
- Traylen, C., Ramasubramanian, S., Zuo, J., Rowe, M., Almohammad, R., Heesom, K., Sweet, S.M.M., Matthews, D.A., and Sinclair, A.J. (2015). Identification of Epstein-Barr Virus Replication Proteins in Burkitt’s Lymphoma Cells. *Pathogens* 4, 739–751.
- Tribouley, C., Lutz, P., Staub, A., and Kedinger, C. (1994). The product of the adenovirus intermediate gene IVa2 is a transcriptional activator of the major late promoter. *J. Virol.* 68, 4450–4457.
- Trowitzsch, S., Viola, C., Scheer, E., Conic, S., Chavant, V., Fournier, M., Papai, G., Ebong, I.-O., Schaffitzel, C., Zou, J., et al. (2015). Cytoplasmic TAF2-TAF8-TAF10 complex provides evidence for nuclear holo-TFIID assembly from preformed submodules. *Nat. Commun.* 6, 6011.
- Twist, K.-A.F., Campbell, E.A., Deighan, P., Nechaev, S., Jain, V., Geiduschek, E.P., Hochschild, A., and Darst, S.A. (2011). Crystal structure of the bacteriophage T4 late-transcription coactivator gp33 with the β -subunit flap domain of *Escherichia coli* RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* 108, 19961–19966.
- Urier, G., Buisson, M., Chambard, P., and Sergeant, A. (1989). The Epstein-Barr virus early protein EB1 activates transcription from different responsive elements including AP-1 binding sites. *EMBO J.* 8, 1447–1453.

- Vannini, A., and Cramer, P. (2012). Conservation between the RNA Polymerase I, II, and III Transcription Initiation Machineries. *Mol. Cell* *45*, 439–446.
- Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W.M.P., van Schaik, F.M.A., Varier, R.A., Baltissen, M.P.A., Stunnenberg, H.G., Mann, M., and Timmers, H.T.M. (2007). Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4. *Cell* *131*, 58–69.
- Verrijzer, C.P., Chen, J.-L., Yokomori, K., and Tjian, R. (1995). Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell* *81*, 1115–1125.
- De Villiers, E.-M., Fauquet, C., Broker, T.R., Bernard, H.-U., and zur Hausen, H. (2004). Classification of papillomaviruses. *Virology* *324*, 17–27.
- Wang, J.-T., Yang, P.-W., Lee, C.-P., Han, C.-H., Tsai, C.-H., and Chen, M.-R. (2005). Detection of Epstein–Barr virus BGLF4 protein kinase in virus replication compartments and virus particles. *J. Gen. Virol.* *86*, 3215–3225.
- Wang, Q., Bosch, B.-J., Vlak, J.M., van Oers, M.M., Rottier, P.J., and van Lent, J.W.M. (2016). Budded baculovirus particle structure revisited. *J. Invertebr. Pathol.* *134*, 15–22.
- Wang, W.-M., Wu, S.-Y., Lee, A.-Y., and Chiang, C.-M. (2011a). Binding Site Specificity and Factor Redundancy in Activator Protein-1-driven Human Papillomavirus Chromatin-dependent Transcription. *J. Biol. Chem.* *286*, 40974–40986.
- Wang, X., Meyers, C., Wang, H.-K., Chow, L.T., and Zheng, Z.-M. (2011b). Construction of a Full Transcription Map of Human Papillomavirus Type 18 during Productive Viral Infection. *J. Virol.* *85*, 8080–8092.
- Wang, Y.-C.J., Huang, J.-M., and Montalvo, E.A. (1997). Characterization of Proteins Binding to the ZII Element in the Epstein–Barr Virus BZLF1 Promoter: Transactivation by ATF1. *Virology* *227*, 323–330.
- Watanabe, T., Narita, Y., Yoshida, M., Sato, Y., Goshima, F., Kimura, H., and Murata, T. (2015). The Epstein-Barr Virus BDLF4 Gene Is Required for Efficient Expression of Viral Late Lytic Genes. *J. Virol.* *89*, 10120–10124.
- Wegner, M., Drolet, D.W., and Rosenfeld, M.G. (1993). Regulation of JC virus by the POU-domain transcription factor Tst-1: implications for progressive multifocal leukoencephalopathy. *Proc. Natl. Acad. Sci. U. S. A.* *90*, 4743–4747.
- Weinzierl, R.O., Ruppert, S., Dynlacht, B.D., Tanese, N., and Tjian, R. (1993). Cloning and expression of Drosophila TAFII60 and human TAFII70 reveal conserved interactions with other subunits of TFIID. *EMBO J.* *12*, 5303–5309.
- Weir, J.P., and Narayanan, P.R. (1990). Expression of the herpes simplex virus type 1 glycoprotein C gene requires sequences in the 5' noncoding region of the gene. *J. Virol.* *64*, 445–449.

- White, M.K., Safak, M., and Khalili, K. (2009). Regulation of Gene Expression in Primate Polyomaviruses. *J. Virol.* *83*, 10846–10856.
- White, M.K., Gordon, J., and Khalili, K. (2013). The Rapidly Expanding Family of Human Polyomaviruses: Recent Developments in Understanding Their Life Cycle and Role in Human Pathology. *PLoS Pathog.* *9*.
- Whitehurst, C.B., Sanders, M.K., Law, M., Wang, F.-Z., Xiong, J., Dittmer, D.P., and Pagano, J.S. (2013). Maribavir inhibits Epstein-Barr virus transcription through the EBV protein kinase. *J. Virol.* *87*, 5311–5315.
- Wiley, S.R., Kraus, R.J., and Mertz, J.E. (1992). Functional binding of the “TATA” box binding component of transcription factor TFIID to the -30 region of TATA-less promoters. *Proc. Natl. Acad. Sci. U. S. A.* *89*, 5814–5818.
- Wiley, S.R., Kraus, R.J., Zuo, F., Murray, E.E., Loritz, K., and Mertz, J.E. (1993). SV40 early-to-late switch involves titration of cellular transcriptional repressors. *Genes Dev.* *7*, 2206–2219.
- Wilkins, K., and Rüger, W. (1996). Characterization of Bacteriophage T4 Early Promoters in Vivo with a New Promoter Probe Vector. *Plasmid* *35*, 108–120.
- Woellmer, A., Arteaga-Salas, J.M., and Hammerschmidt, W. (2012). BZLF1 Governs CpG-Methylated Chromatin of Epstein-Barr Virus Reversing Epigenetic Repression. *PLoS Pathog.* *8*.
- Wollebo, H.S., Woldemichael, B., Khalili, K., Safak, M., and White, M.K. (2013). Epigenetic regulation of polyomavirus JC. *Virol. J.* *10*, 264.
- Wong, E., Wu, T.-T., Reyes, N., Deng, H., and Sun, R. (2007). Murine Gammaherpesvirus 68 Open Reading Frame 24 Is Required for Late Gene Expression after DNA Replication. *J. Virol.* *81*, 6761–6764.
- Wong-Ho, E., Wu, T.-T., Davis, Z.H., Zhang, B., Huang, J., Gong, H., Deng, H., Liu, F., Glaunsinger, B., and Sun, R. (2014). Unconventional Sequence Requirement for Viral Late Gene Core Promoters of Murine Gammaherpesvirus 68. *J. Virol.* *88*, 3411–3422.
- Wooldridge, T.R., and Laimins, L.A. (2008). Regulation of human papillomavirus type 31 gene expression during the differentiation-dependent life cycle through histone modifications and transcription factor binding. *Virology* *374*, 371–380.
- Wright, J., and Leppard, K.N. (2013). The Human Adenovirus 5 L4 Promoter Is Activated by Cellular Stress Response Protein p53. *J. Virol.* *87*, 11617–11625.
- Wright, J., Atwan, Z., Morris, S.J., and Leppard, K.N. (2015). The Human Adenovirus Type 5 L4 Promoter Is Negatively Regulated by TFII-I and L4-33K. *J. Virol.* *89*, 7053–7063.
- Wu, C.P., Wang, J.-Y., Huang, T.-Y., Lo, H.-R., and Chao, Y.-C. (2008). Identification of baculoviral factors required for the activation of enhancer-like polyhedrin upstream

- (pu) sequence. *Virus Res.* 138, 7–16.
- Wu, F.Y., Wang, S.E., Chen, H., Wang, L., Hayward, S.D., and Hayward, G.S. (2004). CCAAT/Enhancer Binding Protein α Binds to the Epstein-Barr Virus (EBV) ZTA Protein through Oligomeric Interactions and Contributes to Cooperative Transcriptional Activation of the ZTA Promoter through Direct Binding to the ZII and ZIIIB Motifs during Induction of the EBV Lytic Cycle. *J. Virol.* 78, 4847–4865.
- Wu, K., Orozco, D., and Hearing, P. (2012). The Adenovirus L4-22K Protein Is Multifunctional and Is an Integral Component of Crucial Aspects of Infection. *J. Virol.* 86, 10474–10483.
- Wu, K., Guimet, D., and Hearing, P. (2013). The Adenovirus L4-33K Protein Regulates both Late Gene Expression Patterns and Viral DNA Packaging. *J. Virol.* 87, 6739–6747.
- Wu, T.-T., Park, T., Kim, H., Tran, T., Tong, L., Martinez-Guzman, D., Reyes, N., Deng, H., and Sun, R. (2009). ORF30 and ORF34 Are Essential for Expression of Late Genes in Murine Gammaherpesvirus 68. *J. Virol.* 83, 2265–2273.
- Wyrwicz, L.S., and Rychlewski, L. (2007). Identification of Herpes TATT-binding protein. *Antiviral Res.* 75, 167–172.
- Xi, Q., Cuesta, R., and Schneider, R.J. (2004). Tethering of eIF4G to adenoviral mRNAs by viral 100k protein drives ribosome shunting. *Genes Dev.* 18, 1997–2009.
- Xie, X., Kokubo, T., Cohen, S.L., Mirza, U.A., Hoffmann, A., Chait, B.T., Roeder, R.G., Nakatani, Y., and Burley, S.K. (1996). Structural similarity between TAFs and the heterotetrameric core of the histone octamer. *Nature* 380, 316–322.
- Xing, K., Deng, R., Wang, J., Feng, J., Huang, M., and Wang, X. (2005). Analysis and prediction of baculovirus promoter sequences. *Virus Res.* 113, 64–71.
- Xu, B., Yoo, S., and Guarino, L.A. (1995). Differential transcription of baculovirus late and very late promoters: fractionation of nuclear extracts by phosphocellulose chromatography. *J. Virol.* 69, 2912–2917.
- Yamagishi, J., Burnett, E.D., Harwood, S.H., and Blissard, G.W. (2007). The AcMNPV pp31 gene is not essential for productive AcMNPV replication or late gene transcription but appears to increase levels of most viral transcripts. *Virology* 365, 34–47.
- Yang, S., and Miller, L.K. (1999). Activation of Baculovirus Very Late Promoters by Interaction with Very Late Factor 1. *J. Virol.* 73, 3404–3409.
- Yang, H., Ma, Y., Wang, Y., Yang, H., Shen, W., and Chen, X. (2014). Transcription regulation mechanisms of bacteriophages. *Bioengineered* 5, 300–304.
- Yang, P.-W., Chang, S.-S., Tsai, C.-H., Chao, Y.-H., and Chen, M.-R. (2008). Effect of phosphorylation on the transactivation activity of Epstein–Barr virus BMRF1, a major target of the viral BGLF4 kinase. *J. Gen. Virol.* 89, 884–895.

- Yokoyama, N., Fujii, K., Hirata, M., Tamai, K., Kiyono, T., Kuzushima, K., Nishiyama, Y., Fujita, M., and Tsurumi, T. (1999). Assembly of the Epstein–Barr virus BBLF4, BSLF1 and BBLF2/3 proteins and their interactive properties. *J. Gen. Virol.* *80*, 2879–2887.
- You, J., Croyle, J.L., Nishimura, A., Ozato, K., and Howley, P.M. (2004). Interaction of the Bovine Papillomavirus E2 Protein with Brd4 Tethers the Viral DNA to Host Mitotic Chromosomes. *Cell* *117*, 349–360.
- Yu, X., Wang, Z., and Mertz, J.E. (2007). ZEB1 Regulates the Latent-Lytic Switch in Infection by Epstein-Barr Virus. *PLoS Pathog.* *3*.
- Yu, X., McCarthy, P.J., Lim, H.-J., Iempridee, T., Kraus, R.J., Gorlen, D.A., and 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*, 5081–5090.
- Yueh, A., and Schneider, R.J. (1996). Selective translation initiation by ribosome jumping in adenovirus-infected and heat-shocked cells. *Genes Dev.* *10*, 1557–1567.
- Zabierowski, S., and DeLuca, N.A. (2004). Differential Cellular Requirements for Activation of Herpes Simplex Virus Type 1 Early (tk) and Late (gC) Promoters by ICP4. *J. Virol.* *78*, 6162–6170.
- Zabierowski, S.E., and DeLuca, N.A. (2008). Stabilized Binding of TBP to the TATA Box of Herpes Simplex Virus Type 1 Early (tk) and Late (gC) Promoters by TFIIA and ICP4. *J. Virol.* *82*, 3546–3554.
- Zehavi, Y., Kedmi, A., Ideses, D., and Juven-Gershon, T. (2015). TRF2: TRansForming the view of general transcription factors. *Transcription* *6*, 1–6.
- Zenke, M., Grundström, T., Matthes, H., Wintzerith, M., Schatz, C., Wildeman, A., and Chambon, P. (1986). Multiple sequence motifs are involved in SV40 enhancer function. *EMBO J.* *5*, 387–397.
- Zerby, D., Chen, C.-J., Poon, E., Lee, D., Shiekhatar, R., and Lieberman, P.M. (1999). The Amino-Terminal C/H1 Domain of CREB Binding Protein Mediates Zta Transcriptional Activation of Latent Epstein-Barr Virus. *Mol. Cell. Biol.* *19*, 1617–1626.
- Zhao, X., and Schwartz, S. (2008). Inhibition of HPV-16 L1 expression from L1 cDNAs correlates with the presence of hnRNP A1 binding sites in the L1 coding region. *Virus Genes* *36*, 45–53.
- Zhao, X., Rush, M., and Schwartz, S. (2004). Identification of an hnRNP A1-Dependent Splicing Silencer in the Human Papillomavirus Type 16 L1 Coding Region That Prevents Premature Expression of the Late L1 Gene. *J. Virol.* *78*, 10888–10905.
- Zhao, X., Öberg, D., Rush, M., Fay, J., Lambkin, H., and Schwartz, S. (2005). A 57-Nucleotide Upstream Early Polyadenylation Element in Human Papillomavirus Type 16 Interacts with hFip1, CstF-64, hnRNP C1/C2, and Polypyrimidine Tract Binding

Protein. *J. Virol.* 79, 4270–4288.

Zhao, X., Fay, J., Lambkin, H., and Schwartz, S. (2007a). Identification of a 17-nucleotide splicing enhancer in HPV-16 L1 that counteracts the effect of multiple hnRNP A1-binding splicing silencers. *Virology* 369, 351–363.

Zhao, X., Rush, M., Carlsson, A., and Schwartz, S. (2007b). The presence of inhibitory RNA elements in the late 3'-untranslated region is a conserved property of human papillomaviruses. *Virus Res.* 125, 135–144.

Zuo, F., and Mertz, J.E. (1995). Simian virus 40 late gene expression is regulated by members of the steroid/thyroid hormone receptor superfamily. *Proc. Natl. Acad. Sci.* 92, 8586–8590.

Zuo, F., Kraus, R.J., Gulick, T., Moore, D.D., and Mertz, J.E. (1997). Direct modulation of simian virus 40 late gene expression by thyroid hormone and its receptor. *J. Virol.* 71, 427–436.

PARTIE V :

Annexes Résultats

Annexe 1 : Interactions individuelles entre protéines du vPIC.

GST pull-down GST-BFRF2.

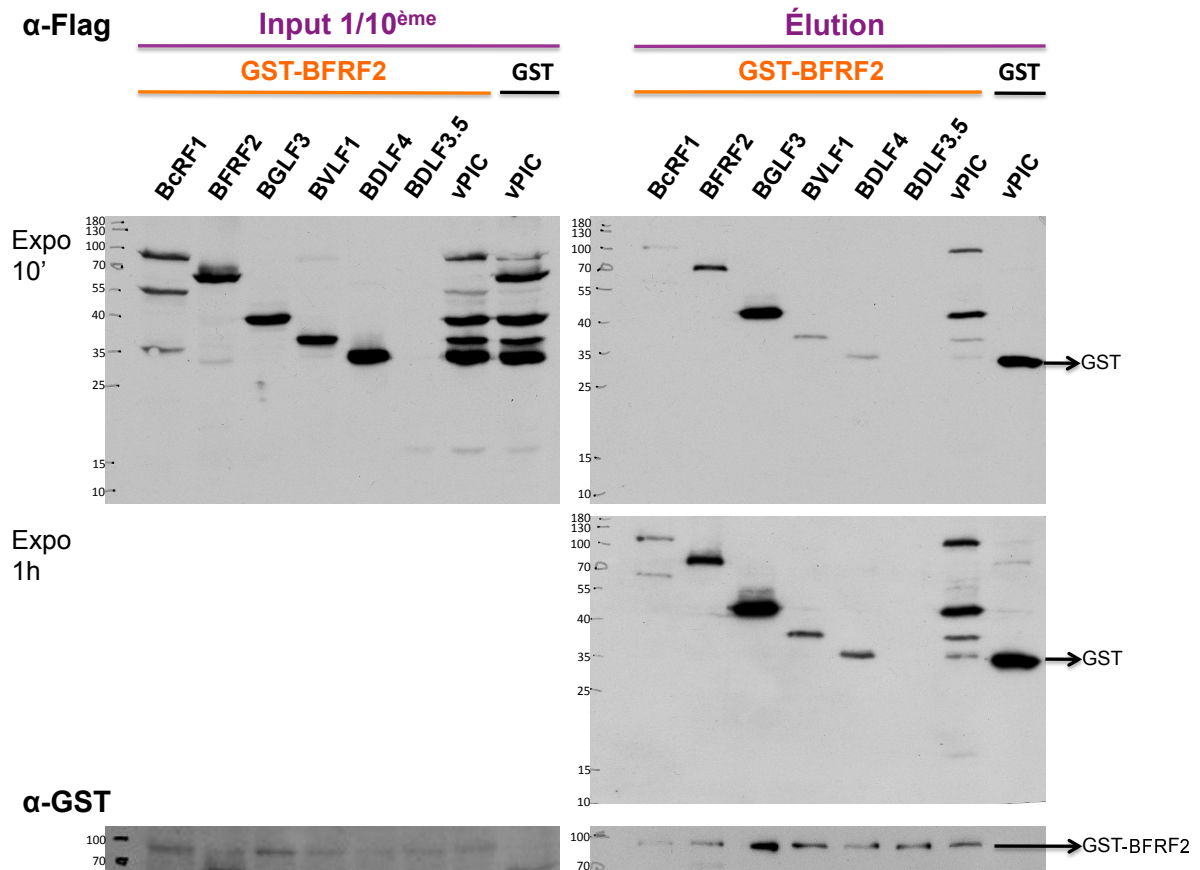


Figure S1 : La protéine GST-BFRF2 interagit différemment avec la plupart des composants du vPIC.

Les vecteurs d'expression codant pour les protéines chimériques GST-BFRF2 ou Flag-GST ont été transfectés dans des cellules HEK293 avec soit un plasmide d'expression codant pour une des protéines du vPIC soit avec tous les plasmides codant pour les cinq protéines du vPIC restantes. Les cellules ont été récoltées et lysées quarante huit heures post-transfection. L'input, correspondant à 10% du lysat cellulaire est représenté à gauche. Les 90 % restant du lysat ont été incubés sur billes Glutathione-Sépharose. Les protéines retenues sur la colonne ont été lavées, éluées et séparées sur SDS-PAGE avant d'être transférées sur membrane de nitrocellulose. Les membranes ont ensuite été incubées avec un anticorps anti-Flag M2 de Sigma (souris) puis avec un anti-GST Covalab (souris). Les différentes protéines ont été détectées après le Western Blot à l'aide d'un film autoradiographique (panel de droite). Un temps d'exposition court (en haut, 10 minutes) et un temps long (au milieu, 1h), permettent d'observer les différentes interactions. Le panel du bas démontre le niveau d'expression de la protéine GST-BFRF2.

GST pull-down GST-BGLF3.

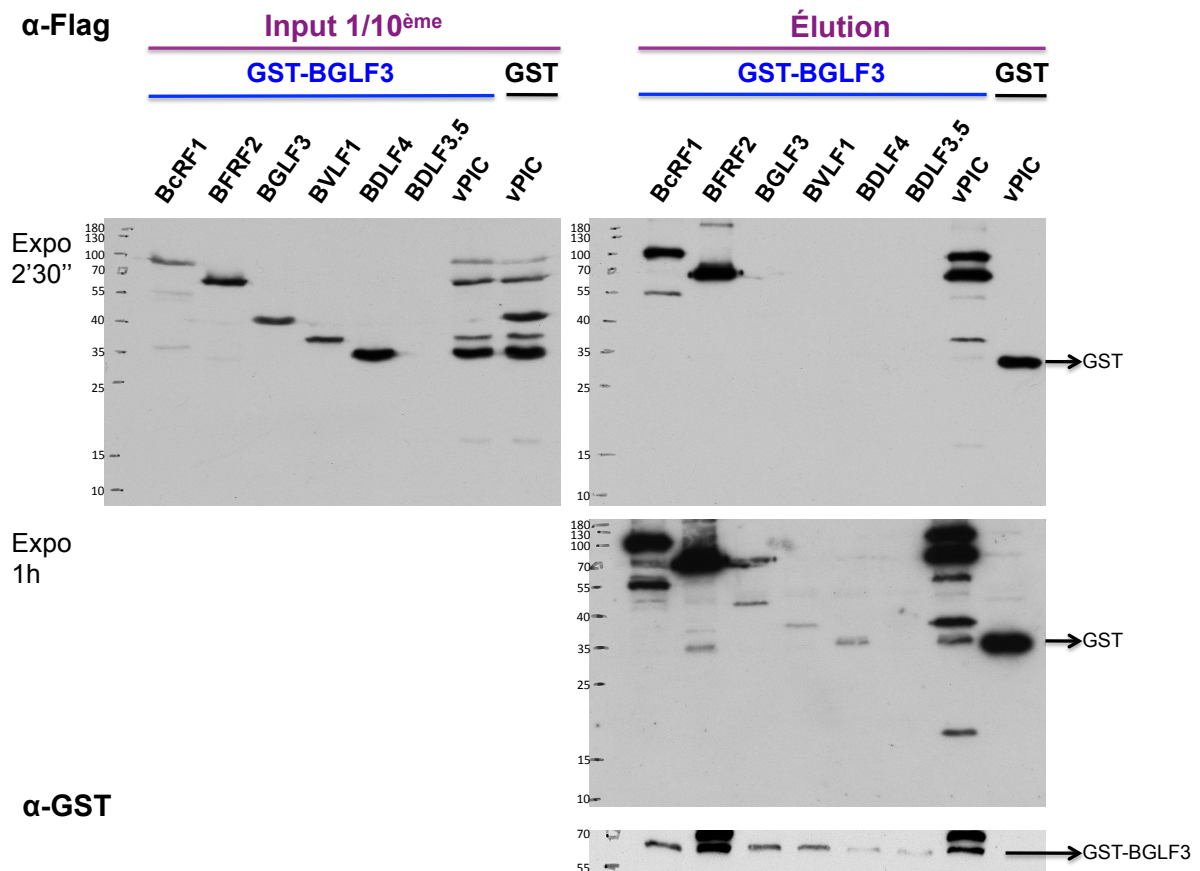


Figure S2 : La protéine Flag-GST-BGLF3 interagit avec la plupart des composants du vPIC.

Les vecteurs d'expression codant pour les protéines chimériques GST-BGLF3 ou Flag-GST ont été transfectés dans des cellules HEK293 avec soit un plasmide d'expression codant pour une des protéines du vPIC soit avec tous les plasmides codant pour les cinq protéines du vPIC restantes. Les cellules ont été récoltées et lysées quarante huit heures post-transfection. L'input, correspondant à 10% du lysat cellulaire est représenté à gauche. Les 90 % restant du lysat ont été incubés sur billes Glutathione-Sépharose. Les protéines retenues sur la colonne ont été lavées, éluées et séparées sur SDS-PAGE avant d'être transférées sur membrane de nitrocellulose. Ces protéines ont été détectées après Western Blot à l'aide d'un film autoradiographique (panel de droite). Un temps d'exposition court (en haut, 10 minutes) et un temps long (au milieu, 1h), permettent d'observer les différentes interactions. Le panel du bas permet de détecter le niveau d'expression de la protéine GST-BGLF3, indiquée à l'aide d'une flèche. La bande située à 70 KDa dans ce panel correspond à la protéine BFRF2.

GST pull-down GST-BVLF1.

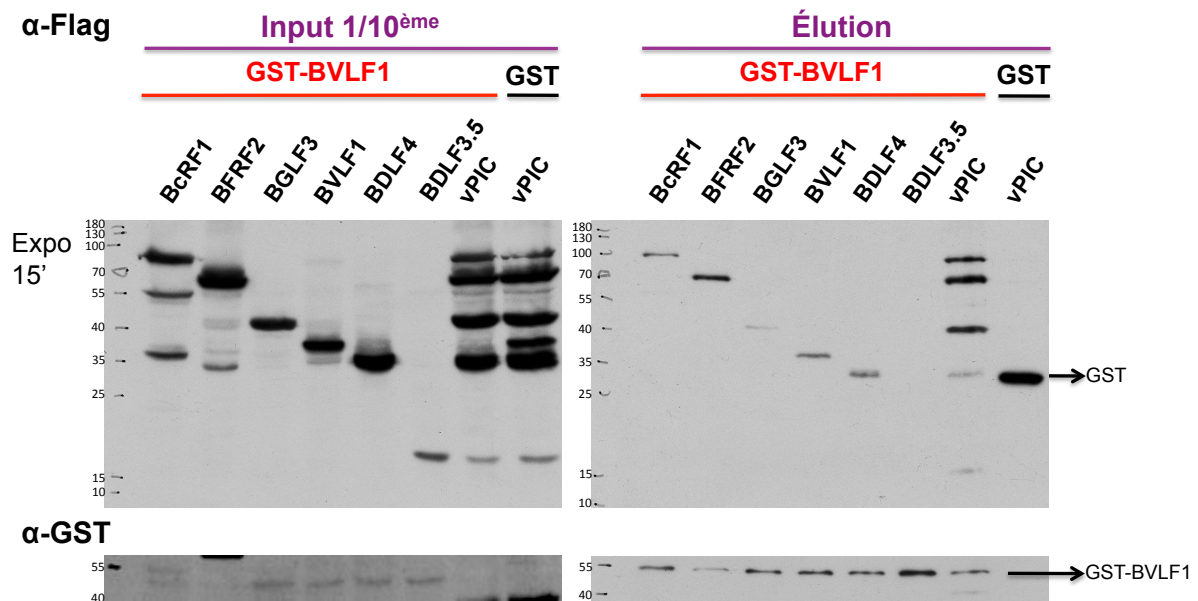


Figure S3 : La protéine GST-BVLF1 interagit avec plusieurs composants du vPIC.

Les vecteurs d'expression codant pour les protéines chimériques GST-BVLF1 ou Flag-GST ont été transfectés dans des cellules HEK293 avec soit un plasmide d'expression codant pour une des protéines du vPIC soit avec tous les plasmides codant pour les cinq protéines du vPIC restantes. Les cellules ont été récoltées et lysées quarante huit heures post-transfection. L'input, correspondant à 10% du lysat cellulaire est représenté à gauche. Les 90 % restant du lysat ont été incubés sur billes Glutathione-Sépharose. Les protéines retenues sur la colonne ont été lavées, éluées et séparées sur SDS-PAGE avant d'être transférées sur membrane de nitrocellulose. Ces protéines ont été détectées après Western Blot à l'aide d'un film autoradiographique (panel de droite). Un temps d'exposition médian (en haut, 15 minutes) permet d'observer les différentes interactions. Le panel du bas révèle le niveau d'expression de la protéine GST-BVLF1 aussi bien dans l'input qu'après purification sur billes Glutathione Sépharose.

GST pull-down GST-BDLF4.

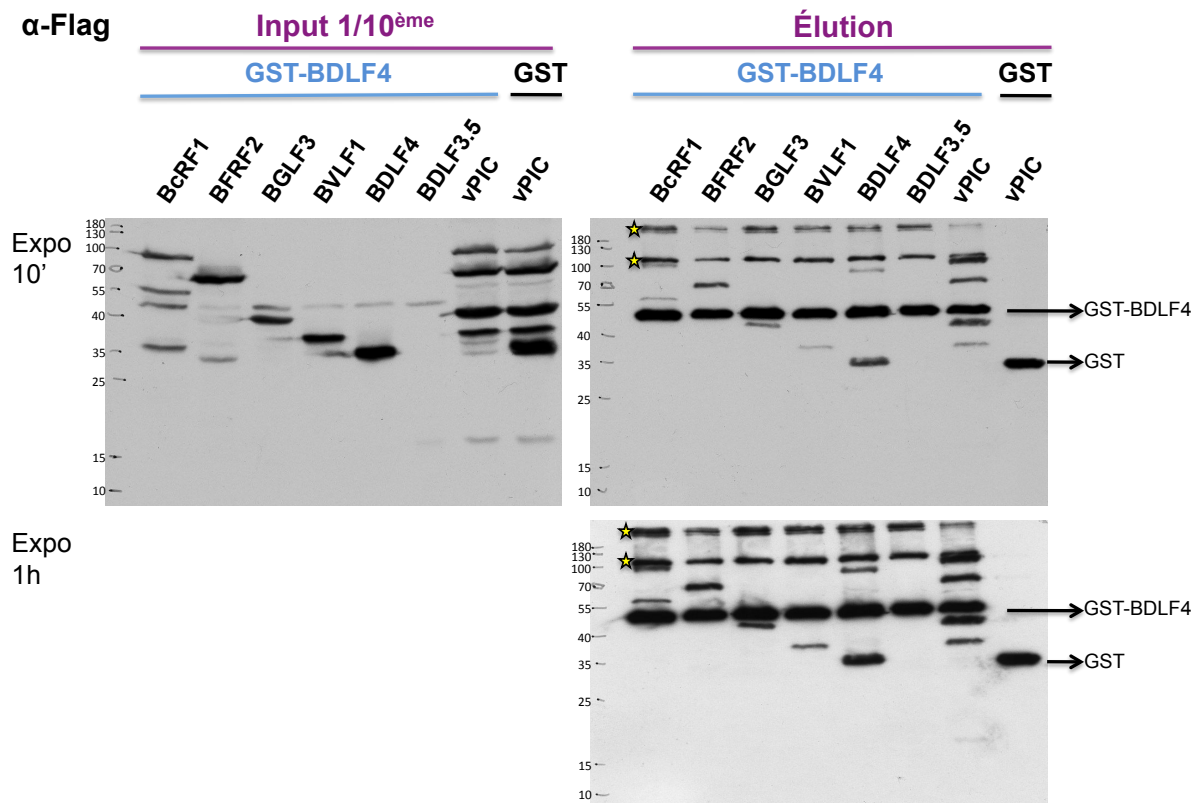


Figure S4 : La protéine Flag-GST-BDLF4 interagit différemment avec plusieurs composants du vPIC.

Les vecteurs d'expression codant pour les protéines chimériques Flag-GST-BDLF4 ou Flag-GST ont été transfectés dans des cellules HEK293 avec soit un plasmide d'expression codant pour une des protéines du vPIC soit avec tous les plasmides codant pour les cinq protéines du vPIC restantes. Les cellules ont été récoltées et lysées quarante huit heures post-transfection. L'input, correspondant à 10% du lysat cellulaire est représenté à gauche. Les 90 % restant du lysat ont été incubés sur billes Glutathione-Sépharose. Les protéines retenues sur la colonne ont été lavées, éluées et séparées sur SDS-PAGE avant d'être transférées sur membrane de nitrocellulose. Les membranes ont ensuite été incubées avec un anticorps anti-Flag M2 Sigma (souris). Les protéines éluées ont été détectées après ce Western Blot à l'aide d'un film autoradiographique (panel de droite). Un temps d'exposition court (en haut, 10 minutes) et un temps long (en bas, 1h), permettent d'observer les différences d'interactions. Les astérisques correspondent à la formation d'oligomères de la protéine Flag-GST-BDLF4.

GST pull-down GST-BDLF3.5.

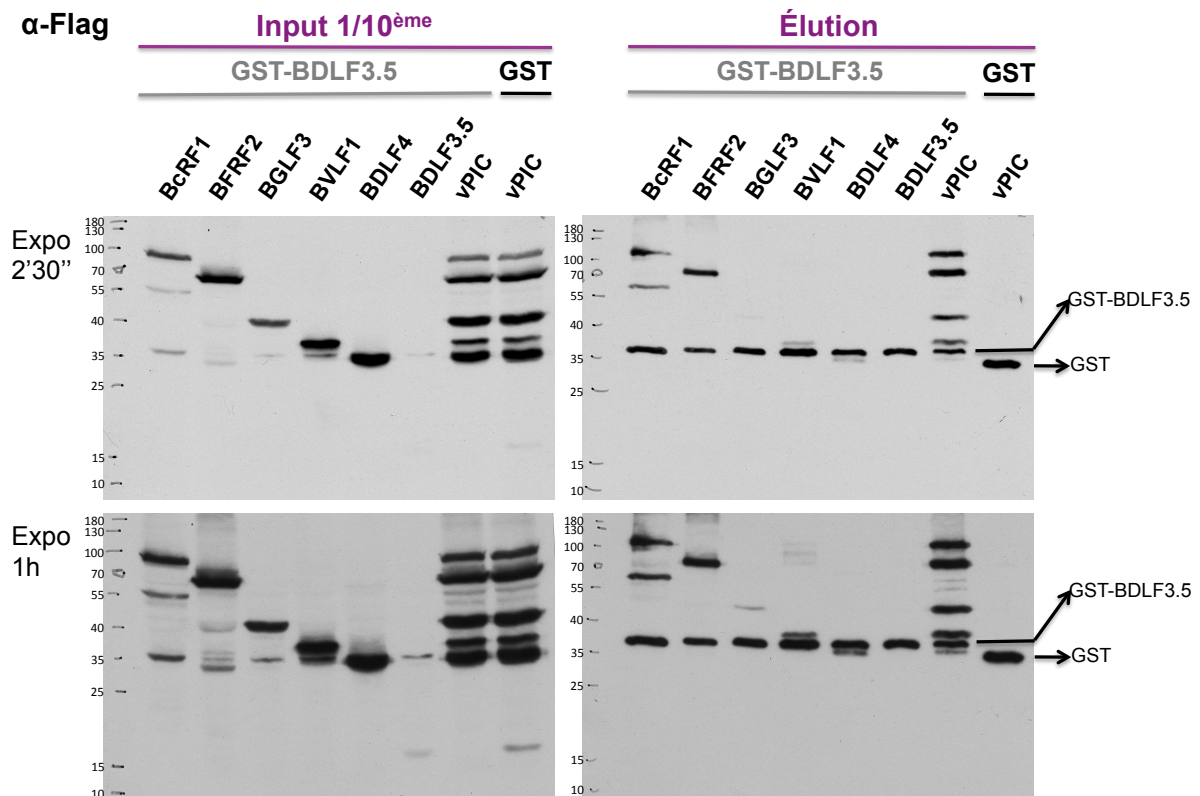


Figure S5 : La protéine Flag-GST-BDLF3.5 interagit différemment avec les composants du vPIC.

Les vecteurs d'expression codant pour les protéines chimériques Flag-GST-BDLF3.5 ou Flag-GST ont été transfectés dans des cellules HEK293 avec soit un plasmide d'expression codant pour une des protéines du vPIC soit avec tous les plasmides codant pour les cinq protéines du vPIC restantes. Les cellules ont été récoltées et lysées quarante huit heures post-transfection. L'input, correspondant à 10% du lysat cellulaire est représenté à gauche. Les 90 % restant du lysat ont été incubés sur billes Glutathione-Sépharose. Les protéines retenues sur la colonne ont été lavées, éluées et séparées sur SDS-PAGE avant d'être transférées sur membrane de nitrocellulose. Ces protéines ont été détectées après Western Blot à l'aide d'un film autoradiographique (panel de droite). Un temps d'exposition court (en haut, 2 minutes et 30 secondes) et un temps long (en bas, 1h), permettent de visualiser les différentes protéines interagissant entre elles.

Annexe 2 : Interactions des membres du vPIC en l'absence de l'un d'entre eux.

GST pull-down GST-BGLF3.

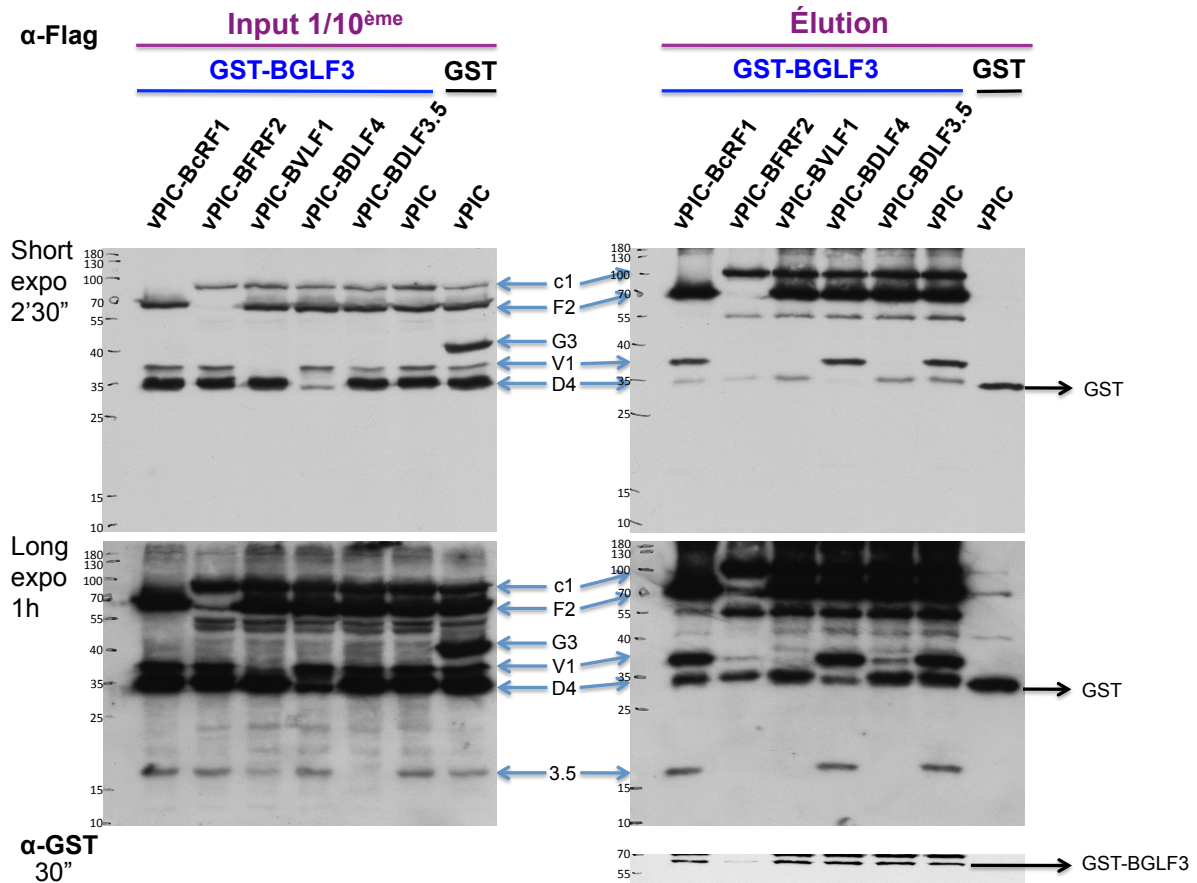


Figure S6 : La protéine GST-BGLF3 recrute différemment plusieurs composants du vPIC.

Les vecteurs d'expression codant pour les protéines chimériques GST-BGLF3 ou Flag-GST ont été transfectés dans des cellules HEK293 avec soit les différents plasmides d'expression codant pour les cinq autres protéines du vPIC restantes soit seulement quatre d'entre elles. Les cellules ont été récoltées et lysées quarante huit heures post-transfection. L'input, correspondant à 10% du lysat cellulaire est représenté à gauche. Les 90 % restant du lysat ont été incubés sur billes Glutathione-Sépharose. Les protéines retenues sur la colonne ont été lavées, éluées et séparées sur SDS-PAGE avant d'être transférées sur membrane de nitrocellulose. Ces protéines ont été détectées après Western Blot à l'aide d'un film autoradiographique (panel de droite). Un temps d'exposition court (en haut, 2 minutes et 30 secondes) et un temps long (au milieu, 1h), permettent d'observer les différentes interactions. Le panel du bas dévoile le niveau d'expression de la protéine GST-BGLF3 uniquement après purification sur billes Glutathione Sépharose.

GST pull-down GST-BDLF4.

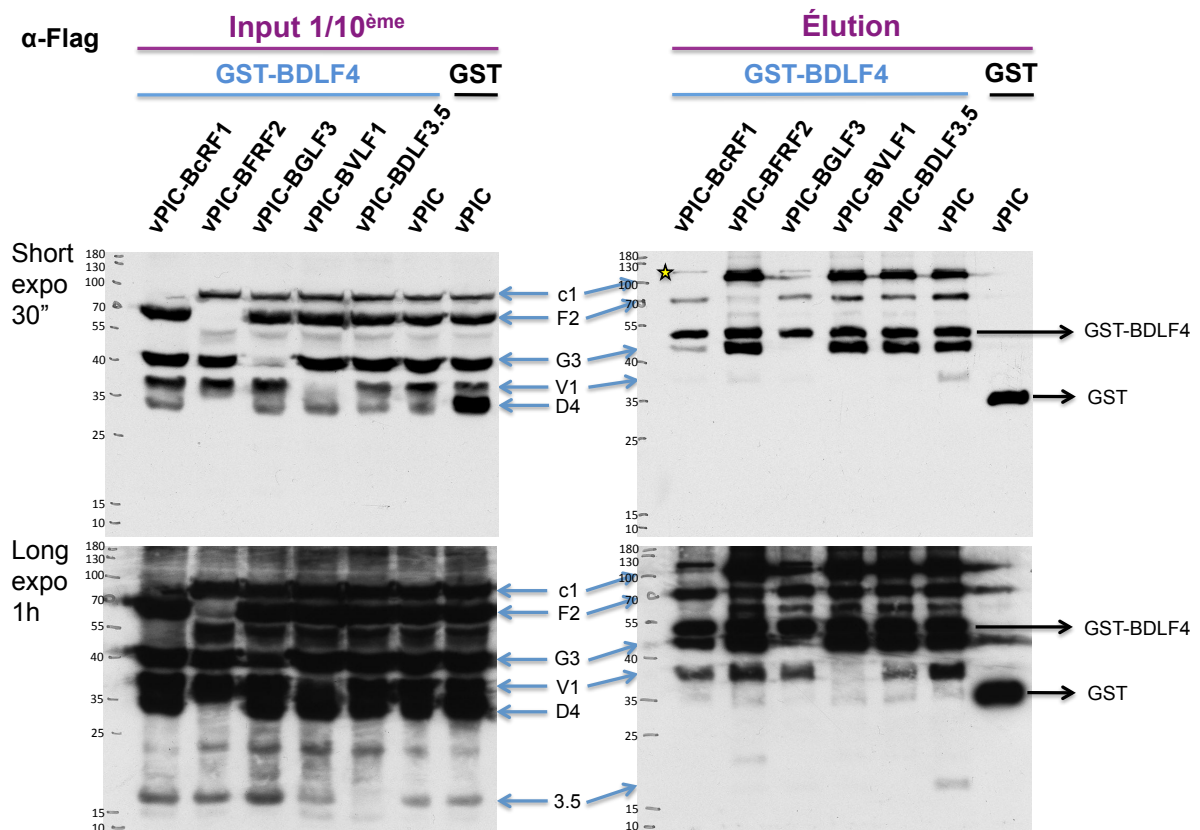


Figure S7 : La protéine Flag-GST-BDLF4 retient différemment plusieurs composants du vPIC.

Les vecteurs d'expression codant pour les protéines chimériques Flag-GST-BDLF4 ou Flag-GST ont été transfectés dans des cellules HEK293 avec soit les différents plasmides d'expression codant pour les cinq autres protéines du vPIC restantes soit seulement quatre d'entre elles. Les cellules ont été récoltées et lysées quarante huit heures post-transfection. L'input, correspondant à 10% du lysat cellulaire est représenté à gauche. Les 90 % restant du lysat ont été incubés sur billes Glutathione-Sépharose. Les protéines retenues sur la colonne ont été lavées, éluées et séparées sur SDS-PAGE avant d'être transférées sur membrane de nitrocellulose. Ces protéines ont été détectées après Western Blot à l'aide d'un film autoradiographique (panel de droite). Un temps d'exposition court (en haut, 30 secondes) et un temps long (en bas, 1h), permettent d'observer les différences d'interactions entre les partenaires. L'astérisque jaune correspond au dimère de la protéine Flag-GST-BDLF4.

GST pull-down GST-BDLF3.5.

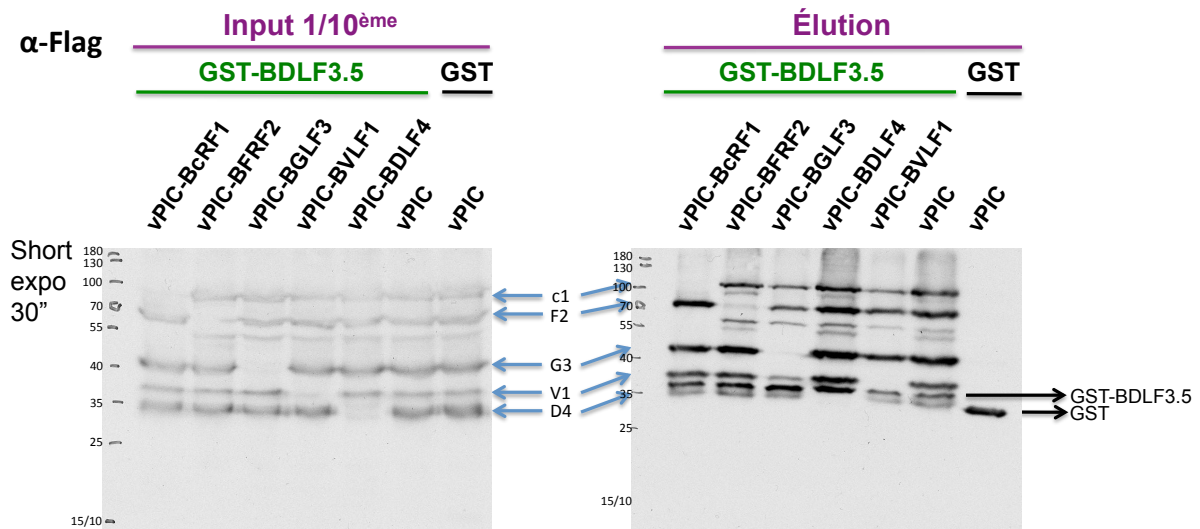


Figure S8 : La protéine Flag-GST-BDLF3.5 retient différemment plusieurs composants du vPIC.

Les vecteurs d'expression codant pour les protéines chimériques Flag-GST-BDLF3.5 ou Flag-GST ont été transfectés dans des cellules HEK293 avec soit les différents plasmides d'expression codant pour les cinq autres protéines du vPIC restantes soit seulement quatre d'entre elles. Les cellules ont été récoltées et lysées. L'input, correspondant à 10% du lysat cellulaire est représenté à gauche. Les 90 % restant du lysat ont été incubés sur billes Glutathione-Sépharose. Les protéines retenues sur la colonne ont été lavées, éluées et séparées sur SDS-PAGE avant d'être transférées sur membrane de nitrocellulose. Ces protéines ont été détectées après Western Blot à l'aide d'un film autoradiographique (panel de droite). Un temps d'exposition court (30 secondes) est suffisant pour observer les inputs et les différentes interactions.

Annexe 3 : GST pull-down *in vitro* des composants du complexe de réplication de l'ADN viral avec les membres du vPIC.

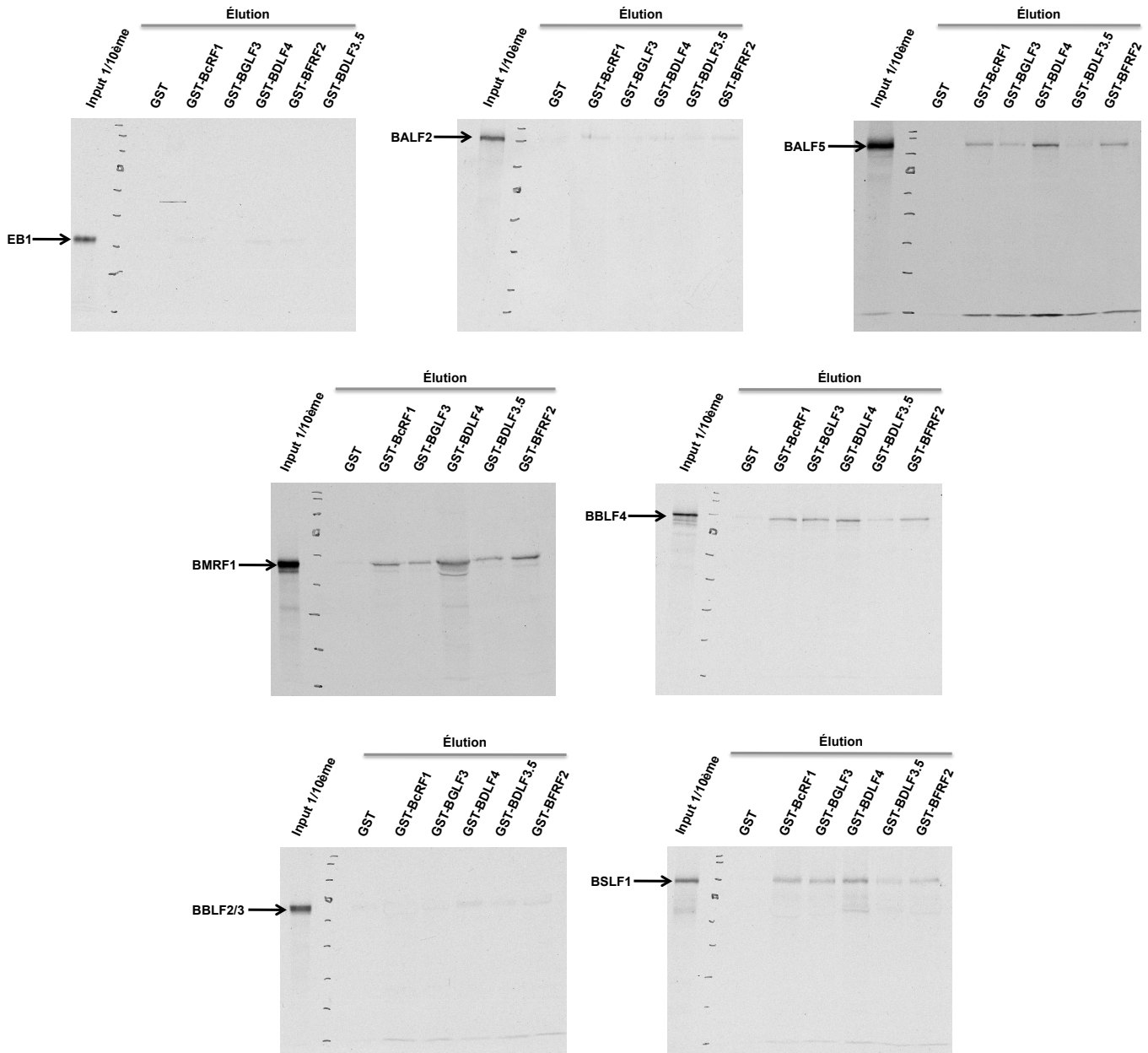


Figure S9 : Interactions, *in vitro*, entre protéines du complexe de réplication de l'ADN viral et du vPIC.

Les plasmides d'expression codant pour les différentes protéines du complexe de réplication ont été incubés dans du lysat de réticulocytes pour être transcrits et traduits (en présence de méthionine marquée au soufre 35) en protéines. Parallèlement, les protéines GST et GST-X (X étant une protéine du vPIC), préalablement produites en bactéries ont été purifiées sur billes Glutathione Sépharose. Ces dernières sont alors incubées avec les protéines du complexe de réplication traduites *in vitro* en lysat de réticulocytes. Après plusieurs lavages, les protéines ont été éluées, séparées sur gel SDS-PAGE et transférées sur membranes de nitrocellulose. Ces dernières ont été colorées au rougeponceau pour contrôler qualitativement la quantité de protéines chimériques GST utilisées pour la purification. Les protéines marquées radioactivement ont été révélées à l'aide d'un film autoradiographique.