Analyse transcriptionnelle des phases précoces de l'infection par le baculovirus AcMNPV et exploitation d'une banque d'ADN complémentaire issue de sa cellule-hôte, Sf9

Igor Landais

▶ **To cite this version:**

Igor Landais. Analyse transcriptionnelle des phases précoces de l'infection par le baculovirus AcMNPV et exploitation d'une banque d'ADN complémentaire issue de sa cellule-hôte, Sf9. Biochimie [q-bio.BM]. Universite Montpellier II; Université Montpellier II - Sciences et Techniques du Languedoc, 2002. Français. NNT : . tel-00472671

HAL Id: tel-00472671

https://theses.hal.science/tel-00472671

Submitted on 12 Apr 2010

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.

Université MONTPELLIER II
Sciences et Techniques du Languedoc

THESE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE MONTPELLIER II

Discipline: Biologie cellulaire
Formation Doctorale: Parasitologie
Ecole Doctorale: Sciences Chimiques et Biologiques pour la Santé

Présentée et soutenue publiquement

par

Igor LANDAIS

le 16 décembre 2002

Analyse transcriptionnelle des phases précoces de l'infection par le baculovirus AcMNPV et exploitation d'une banque d'ADN complémentaire issue de sa cellule-hôte, Sf9

JURY

M. Jacques MARTI (Professeur de l'Université Montpellier II)	Président du jury
Mme Martine CERUTTI (Directeur de Recherche, CNRS)	Directrice de Thèse
M. Pierre COUBLE (Directeur de Recherche, CNRS)	Rapporteur
M. Pierre JALINOT (Directeur de Recherche, CNRS)	Rapporteur
M. René FEYEREISEN (Directeur de Recherche, INRA)	Examinateur
Mme Mylène OGLIASTRO (Chargée de Recherche, INRA)	Examinatrice

RESUME en français

Le lépidoptère *Spodoptera frugiperda* est un ravageur important des cultures mais aussi l'organisme source de la lignée cellulaire *Sf*9, utilisée pour la production de protéines recombinantes dans le système d'expression baculovirus/cellule d'insecte.
Dans la première partie de ce travail, nous avons étudié la transcription chez le baculovirus *Ac*MNPV aux temps précoces de l'infection. Nous montrons tout d'abord que les régions homologues (hrs) portées par le génome de ce virus contiennent de nombreux motifs de reconnaissance pour des facteurs de transcription de type AP1/CREB, et que ces sites fixent spécifiquement des protéines de la cellule-hôte, *Sf*9. Par ailleurs, dans le contexte de l'infection, ces sites sont nécessaires à la transactivation médiée par le facteur viral très précoce IE1 qui se fixe lui aussi sur les hrs. Cette étude montre pour la première fois l'implication de facteurs cellulaires dans ce mécanisme de transactivation virale.
La progression de ces travaux a cependant été freinée par le peu de données disponibles sur les gènes exprimés par *Sf*9. Pour en faciliter l'accès, nous avons donc construit une banque d'ADNc, dont l'exploitation fait l'objet de la deuxième partie de ce mémoire. Son criblage a permis d'obtenir la séquence complète de l'ADN complémentaire du gène hsp90 chez *S. frugiperda*, et d'en déterminer les principales caractéristiques. Par ailleurs, le séquençage à grande échelle de la banque a conduit à la constitution d'une banque d'ESTs, permettant l'identification de la quasi-totalité des gènes de protéines ribosomales. L'analyse de leur séquence a révélé l'existence de particularités qui semblent restreintes aux insectes et aux lépidoptères, un résultat inattendu compte tenu de la grande conservation de ces gènes entre espèces.

TITRE en anglais

Transcriptional study of the early steps of the *Ac*MNPV baculovirus infection and analysis of a cDNA library belonging to its host-cell, *Sf*9

ABSTRACT

S. frugiperda (Lepidoptera) is both a major pest of cultivated plants and the source of the *Sf*9 cell line, widely used for the production of recombinant proteins in the baculovirus expression system. In the first part of this work, we studied the transcriptional events occuring during the early steps of the viral infection. We first show that homologous regions (hrs) of the *Ac*MNPV baculovirus contain a large number of binding sites for AP1/CREB transcription factors, that specifically bind proteins of the *Sf*9 host cell. Furthermore, in the infected context these sites are required for the transactivation mediated by the immediate-early viral factor IE1, that also binds hr sequences. This study shows for the first time the involvement of cellular factors in this mechanism of viral transactivation.
However, this work has been impaired by the fact that little is known about the genes expressed in the *Sf*9 cell line. We thus decided to construct a cDNA library, that was used for different purposes: screening of the library using a probe previously identified in the lab allowed us to obtain the complete hsp90 gene sequence and to determine its main features. Besides this work, large-scale sequencing of the library led to the construction of an EST library, from which we identified almost all the ribosomal protein genes. Sequence analysis revealed several features that seemed to be restricted to insect or lepidoptera species, an unexpected result given the large conservation of these genes througout the evolution.

MOTS-CLES

*Ac*MNPV, baculovirus, banque d'ADNc, CRE, CREB, EST, facteur de transcription, transactivation, génomique, hsp90, IE1, lépidoptère, protéine ribosomale, région homologue, *Sf*9, *Spodoptera frugiperda*.

Discipline: Biologie cellulaire

Adresse: INRA/CNRS UMR 5087, Laboratoire de Pathologie Comparée 30380 St Christol les Alès

REMERCIEMENTS

Je voudrais d'abord remercier les membres du jury pour avoir accepté d'évaluer ce travail:

Ma directrice de thèse, Martine Cerutti, qui m'a accueilli dans le laboratoire de Biologie Moléculaire et Cellulaire et qui a dirigé ces travaux, pour sa constante disponibilité et ses conseils avisés;

Pierre Couble et Pierre Jalinot, qui ont accepté sans la moindre hésitation d'être rapporteurs de ce mémoire;

René Feyereisen et Jacques Marti, pour avoir accepté d'être examinateurs de ce travail;

Et Mylène, pour m'avoir tout appris... au labo et à côté aussi! Ton éternel équilibre, ta bonne humeur communicative, ta générosité et ton talent de prof' de bio mol ont fait que ces quelques années sont passées bien trop vite.

Je suis très reconnaissant à Gérard Devauchelle pour m'avoir ouvert les portes du laboratoire de Pathologie Comparée, pour son soutien constant. Merci, Chef.

Merci à Rachel pour sa précieuse collaboration.

Merci à toutes les personnes du Labo de Pathologie Comparée, qui ont créé cette atmosphère particulière de "grande famille où l'on se sent chez soi". Merci en particulier à mes colocataires Sissi et Janick, à Nathalie, Philippe, Marie-Anne, Serafin, Miguel... et à tous ceux du volley et du tennis, ils se reconnaîtront...

... A StephApne, membre actif et militant de l'ANPMM (Association Nationale de Protection du Mérou en Méditerrannée) et à Cyraïl, spécialiste ès Balle et Volant, pour avoir partagé quelques tranches intenses de vie...

... A ma mère, pour tout ce que tu m'as donné et pour m'avoir construit.

A ma Titou

SOMMAIRE

REMERCIEMENTS 2

ABREVIATIONS 8

LISTE DES FIGURES ET TABLEAUX 9

AVANT-PROPOS 10

MATERIEL ET METHODES 12

A- MATERIEL BIOLOGIQUE 13

B- FACTEURS DE TRANSCRIPTION CELLULAIRES: MISE EN EVIDENCE DES INTERACTIONS ADN-PROTEINES 13

I- Préparation d'extraits nucléaires à partir de cellules Sf9 13

II- Marquage radioactif d'une sonde nucléique 14

III- Gel retard 14

IV- UV-crosslinking 15

C- FACTEURS DE TRANSCRIPTION CELLULAIRES ET VIRAUX: PURIFICATION ET RECHERCHE D'INTERACTIONS PROTEINE-PROTEINE 16

I- Purification 16

1. Protéines cellulaires ayant de l'affinité pour le motif CRE 16

1.1. Colonne d'affinité 16

1.2. Fractionnements successifs 16

2. Purification du facteur viral IE1 17

2.1 FLAG-IE1 produit en système hétérologue 17

2.1.1. Système bactérien 17

2.1.2. Système cellulaire Sf9 (transfection) 18

2.2. Purification partielle de IE1 produit en système baculovirus/cellule d'insecte 18

II- Caractérisation de l'interaction IE1/CBP par co-immunoprécipitation 18

D- BANQUE D'ADN COMPLEMENTAIRE DE CELLULES *Sf*9 EN PHAGE LAMBDA 19

I- Construction 19

II- Amplification 20

III- Conversion de la banque de phages en banque de plasmides 20

IV- Mise en collection et séquençage 21

V- Criblage de la banque de phages par sonde nucléique 21

DES FACTEURS CELLULAIRES SONT IMPLIQUES DANS LA TRANSACTIVATION PAR LE COUPLE IE1/HRS 22

A- INTRODUCTION ET RAPPELS 23

I- Introduction 23

II- Rappels 24

1. Les baculovirus 24

1.1. Histoire et importance 24

1.2. Caractéristiques et cycle viral 25

1.3. Contrôle général de l'expression des gènes 26

1.3.1. La phase très précoce 26

1.3.2. La phase précoce retardée 27

1.3.3. La phase tardive 27

1.3.4. La phase très tardive 288

2. Le couple IE1/hrs 28

2.1. Les protagonistes: présentation et structure 28

2.1.1. Les régions homologues (hrs) 28

2.1.2. La protéine Immediate Early 1 (IE1) 28

2.2. Les protagonistes: rôles connus 29

2.2.1. hrs 29

2.2.1.1. transactivation 29

2.2.1.2. réplication 30

2.2.2. IE1 30

2.2.2.1. transactivation 30

2.2.2.2. réplication 31

3. Implication de facteurs cellulaires dans la transactivation médiée par IE1/hrs 31

3.1. Arguments en faveur de l'implication de facteurs cellulaires 32

3.1.1. Les enhancers de nombreux virus portent des sites de fixation pour des facteurs de l'hôte 32

3.1.2. De nombreux transactivateurs viraux interagissent avec des facteurs de l'hôte 33

3.1.3. Les hrs activent la transcription même en l'absence de facteurs viraux 34

3.1.4. Les hrs semblent fixer des facteurs de l'hôte in vitro 34

3.1.5. La région hr5 complète transactive mieux que la somme de ses palindromes 34

3.1.6. Le domaine transactivateur de IE1 isolé interfère avec la protéine complète 34

3.1.7. Un facteur de l'hôte identifié se fixe sur les hrs 355

3.1.8. Les hrs contrôlent en partie la spécificité d'hôte de la transactivation par IE1 35

3.2. Les facteurs de transcription AP1 et CREB/CREM/ATF 35

3.2.1. Structure et caractéristiques 35

3.2.1.1. la superfamille bZip 35

3.2.1.2. les membres de la famille et leurs sites de fixation 36

3.2.1.3. combinatoire des facteurs AP1 et CREB/CREM/ATF 37

3.2.2. Activation de ces facteurs 37
3.2.3. Rôles 39
4. Structure et fonction d'un enhanceosome **40**
III- Objectifs de l'étude 41

B- RESULTATS **41**
I- Résumé de l'article 1 41
II- Article 1 43

C- RESULTATS COMPLEMENTAIRES ET PERSPECTIVES **63**
I- Affinité des facteurs de l'hôte pour le motif TRE 63
II- Purification des protéines fixant le motif CRE 64
1. Résultats **64**
1.1. Chromatographie d'affinité avec le motif CRE 64
1.2. Séparation sur colonnes successives 64
2. Discussion et perspectives **65**
III- Voie AMPc et activation transcriptionnelle médiée par les sites CRE 66
IV- Purification de IE1 exprimé en système homologue/hétérologue 67
V- Interaction entre IE1 et CBP 69

D- CONCLUSION **70**

MISE EN PLACE ET EXPLOITATION D'UNE BANQUE D'ADN COMPLEMENTAIRES DE CELLULES *Sf*9 ..... **71**
A- INTRODUCTION **72**
B- EXPLOITATION DE LA BANQUE D'ADNC PAR CRIBLAGE AVEC UNE SONDE NUCLEIQUE **72**
I- hsp90 72
II- CBP 74
C- EXPLOITATION DE LA BANQUE PAR SEQUENÇAGE A GRANDE ECHELLE **75**
I- Mise en place de la banque d'ESTs et choix de l'analyse des gènes de protéines ribosomales *75*
II- Rappels sur les protéines ribosomales 76
1. Caractéristiques, nombre et nomenclature **76**
2. Origine **77**
3. Expression **77**
4. Rôles **77**
5. Biais de codons **78**

III- Résultats *79*

1. Résumé en français de l'article 2 ..... **79**

2. Article 2 ..... **81**

IV- Perspectives 1099

1. Les caractéristiques particulières des protéines ribosomales chez les insectes **109**

1.1. Analyse des insertions dans la séquence des gènes L35A et L36 109

1.2. Biais de codons 109

1.3. Les gènes lépidoptères: des cibles pour de nouvelles molécules? 110

2. Des outils pour le programme génomique de S. frugiperda **111**

D- CONCLUSION **111**

CONCLUSIONS ET PERSPECTIVES GENERALES **113**

ANNEXE : CARACTERISATION DU GENE HSP90 CHEZ DEUX LEPIDOPTERES **116**

A- RESUME EN FRANÇAIS DE L'ARTICLE 3 ..... **117**

B- ARTICLE 3 ..... **118**

BIBLIOGRAPHIE **127**

ABREVIATIONS

Ac	*Autographa californica*
ADNc	ADN complémentaire
Ag	*Anticarsia gemmatalis*
AML	Acute myelogenous leukemia
AMPc	AMP cyclique
AP1	Activating protein 1
ARNi	ARN interférence
ATF	Activating transcription factor
BAC	Bacterial artificial chromosome
Bm	*Bombyx mori*
bZip	basic domain-leucin zipper
CaMK	Calmodulin kinase
C/EBP	CAAT box/Enhancer binding protein
CAT	Chloramphenicol acétyl transférase
CBP	CREB binding protein
Cf	*Choristoneura fumiferana*
CRE	cAMP response element
CREB	CRE binding protein
CREM	CRE modulator
DMSO	Diméthylsulfoxyde
DO	Densité optique
DTT	Dithiothreitol
EB	E-box
EDTA	Ethylène diamine tétraacétate
EST	Expressed sequence tag
HMG	High mobility group
Hr	homologous region
Hr1-BP	Hr1-binding protein
hsp	Heat shock protein
HTLV-1	Human T-cell leukemia virus
IE1	Immediate early 1
IPTG	Isopropyl β-D-thiogalactopyranoside
IRF	Interferon regulating factor
KID	Kinase inducible domain
KIX	KID interaction domain
LAT	Latency-associated transcript
Ld	*Lymantria dispar*
LEF	Lymphoid enhancer-binding factor
LTR	Long terminal repeat
MAPK	Mitogen-associated protein kinase
MOI	Multiciplicity of infection
NF-κB	Nuclear factor κB
NGF	Nerve growth factor
Op	*Orgya pseudotsugata*
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PKA	Protein kinase A
PM	Poids moléculaire
Px	*Plutella xylostella*
SCMV	Simian cytomegalovirus
SDS	Sodium dodécyl sulfate
Sf	*Spodoptera frugiperda*
TAF	TBP-associated factor
TBP	TATA-box binding protein
TRE	TPA response element
UV	Ultraviolet
Xc	*Xestia c-nigrum*

LISTE DES FIGURES ET TABLEAUX

Figure 1.....**18**
Figure 2.....**19**
Figure 3.....**21**
Figure 4.....**23**
Figure 5.....**24**
Figure 6.....**25**
Figure 7.....**26**
Figure 8.....**27**
Figure 9.....**28**
Figure 10.....**29**
Figure 11.....**30**
Figure 12.....**32**
Figure 13.....**35**
Figure 14.....**37**
Figure 15.....**38**
Figure 16.....**39**
Figure 17.....**40**
Figure 18.....**40**
Figure 19.....**63**
Figure 20.....**64**
Figure 21.....**64**
Figure 22.....**64**
Figure 23.....**65**
Figure 24.....**66**
Figure 25.....**67**
Figure 26.....**69**
Figure 27.....**72**
Figure 28.....**74**
Figure 29.....**78**

Tableau 1.....**36**
Tableau 2.....**36**

Avant-propos

L'une des thématiques développées par l'équipe de Martine Cerutti dans le Laboratoire de Biologie Cellulaire et Moléculaire de la station de Pathologie Comparée est l'étude des interactions entre le baculovirus *Ac*MNPV et sa cellule-hôte. Dans ce cadre, le projet de thèse qui m'a été proposé a pour objet l'étude de l'implication de facteurs cellulaires dans l'activation de la transcription virale aux temps précoces de l'infection. Les résultats de ce travail font l'objet de la première partie de ce mémoire et d'une publication en cours de soumission.

Les projets de l'équipe ont tous un point commun : ils ont révélé l'implication de protéines cellulaires dans les processus transcriptionnels viraux étudiés. Cependant, très peu de gènes ont été identifiés et caractérisés chez *Spodoptera frugiperda*, l'hôte du baculovirus *Ac*MNPV, ce qui constitue un frein considérable à l'analyse des interactions moléculaires entre la cellule et le baculovirus. Pour accéder plus facilement à ces gènes, nous avons donc construit une banque d'ADNc à partir de cellules *Sf*9, utilisées en routine au laboratoire. Cette stratégie a rapidement porté ses fruits, puisque le criblage de la banque avec un court fragment de gène identifié lors d'une précédente étude a permis d'obtenir la séquence complète de l'ADNc du gène hsp90, sa caractérisation donnant lieu à une publication.

L'exploitation de la banque s'est poursuivie en collaboration avec un laboratoire japonais par une campagne de séquençage à grande échelle, et a débouché sur la constitution d'une banque d'ESTs. Son annotation a permis d'identifier une grande quantité de nouveaux gènes de *S. frugiperda*, dont la quasi-totalité des gènes de protéines ribosomales. Leur analyse est présentée dans une troisième publication.

Les travaux réalisés durant cette thèse sont donc issus de deux approches différentes, l'une transcriptionnelle et l'autre génomique, qui sont complémentaires par leur contribution à la connaissance de la cellule d'un lépidoptère et des gènes qu'elle exprime.

Le plan du mémoire suit cette dichotomie, avec une première partie consacrée à l'étude de l'implication de protéines cellulaires dans l'activation de la transcription par un facteur viral très précoce, IE1. La deuxième partie du mémoire et l'annexe présentent les résultats issus des travaux effectués à partir de la banque d'ADNc *Sf*9, en particulier l'analyse de la famille des gènes de protéines ribosomales chez *S. frugiperda*.

Matériel et méthodes

Nous exposerons dans ce chapitre essentiellement les principes des méthodes employées et non les protocoles dans leur intégralité (quand ceux-ci sont disponibles soit dans les articles présentés, soit dans les manuels indiqués dans chaque section). Les stratégies et protocoles particuliers de préparation d'extraits nucléaires, d'expression et de purification de protéines, de gel retard et de co-immunoprécipitation seront également détaillés car ayant été optimisés pour cette étude.

Enfin, les protocoles "classiques" en biologie moléculaire (Southern, northern, western blot, gel polyacrylamide), décrits dans les current protocols (Ausubel *et al*., 1996), ne seront abordés que dans les articles.

A- MATERIEL BIOLOGIQUE

Virus : baculovirus *Ac*MNPV, souche C6 (Ayres *et al*., 1994)

Cellules : *Sf*9 ATCC, CRL1711 (Vaughn *et al*., 1977)

Souches bactériennes : E.coli JM25.8, Xl1-Blue, BM25-8

B- FACTEURS DE TRANSCRIPTION CELLULAIRES: MISE EN EVIDENCE DES INTERACTIONS ADN-PROTEINES

I- PRÉPARATION D'EXTRAITS NUCLÉAIRES À PARTIR DE CELLULES *Sf*9

La technique utilisée suit le protocole de Dignam avec quelques modifications (Dignam, 1990). Les cellules en phase exponentielle de croissance (cultivées en milieu liquide ou sur boîte de culture) sont rincées trois fois dans du PBS et une fois dans du tampon A hypotonique (Hepes 10 mM pH 7,9 ; $MgCl_2$ 1,5 mM ; KCl 10 mM ; DTT 0,5 mM ; Complete 1X, Roche). Le culot cellulaire est ensuite resuspendu dans 10 fois son volume de tampon A et incubé 10 minutes sur glace, pour faire gonfler les cellules. Celles-ci sont ensuite lysées mécaniquement au Dounce piston B (15 allers-retours lents), qui laisse intacts les noyaux. Le lysat est centrifugé 5 minutes à 500 g et le culot rincé dans 10 volumes de tampon A. Une deuxième centrifugation (10 min à 10000 g) permet de récupérer le culot de noyaux, qui sont resuspendus dans une fois leur volume de tampon C hypertonique (Hepes 20 mM pH 7,9 ; glycérol 25% ; NaCl 0,8 à 1,2 M ; MgCl2 1,5 mM ; EDTA 0,2 mM ; DTT 0,5 mM ; Complete 1X, Roche). Du fait de la dilution au 1/2, la concentration finale en NaCl du tampon d'extraction s'établit à 0,4 - 0,6 M NaCl. Après 30 minutes à 4°C en agitation lente, les noyaux sont centrifugés 10 minutes à 10000 g. Le surnageant contient les extraits nucléaires, composés des protéines nucléaires ayant

diffusé du noyau vers le tampon hypertonique par le flux osmotique établi entre les deux compartiments.

II- MARQUAGE RADIOACTIF D'UNE SONDE NUCLÉIQUE

Nous avons utilisé deux méthodes de marquage en fonction des expérimentations: pour les expériences de gel retard, les sondes étant de courts fragments d'ADN (moins de 80 pb), la méthode employée a été le marquage terminal de l'extrémité 5' au γ^{32}P ATP grâce à la polynucléotide kinase du phage T4 (T4 PNK, Promega). L'utilisation d'oligonucléotides de synthèse, non phosphorylés à l'extrémité 5', permet d'éviter l'étape de déphosphorylation (nécessaire dans le cas de fragments issus de digestions enzymatiques).

Pour les sondes ADN de taille supérieure à 200 pb, nous avons employé la méthode de l'amorçage aléatoire (random priming). Nous avons utilisé le kit "Megaprime DNA labelling system" (Amersham Biosciences).

Après marquage, toutes les sondes ont été purifiées sur colonne de gel filtration (Chroma Spin columns, Clontech) pour éliminer les nucléotides non incorporés.

III- GEL RETARD

Le principe consiste à incuber un fragment d'ADN marqué avec des extraits protéiques. Si des facteurs ayant de l'affinité pour cet ADN sont présents dans les extraits, et si les conditions d'accrochage sont favorables (pH, salinité...), un complexe nucléoprotéique va se former, et migrer dans un gel plus lentement que le fragment d'ADN seul. L'ajout de fragments d'ADN compétiteurs non marqués en fort excès (typiquement 50-100 fois plus) dans la réaction permet de mieux caractériser la spécificité de l'interaction entre les facteurs protéiques et la sonde marquée. En effet, si ces facteurs ont de l'affinité pour le compétiteur, leur probabilité d'être recrutés sur le fragment non marqué en excès est plus grande que sur la sonde marquée. Cette compétition diminue, voire abolit, l'accrochage des facteurs sur la sonde marquée, diminuant d'autant l'intensité du complexe retardé.

Toutes les expérience de gel retard sont réalisées dans les conditions suivantes : 1 à 5 µl d'extraits protéiques (2 à 10 µg), 2 µl de tampon de " binding " 10X (Tris-HCl 50 mM pH 8,0 ; NaCl 250 mM ; Ficoll 10% ; EDTA 5 mM ; DTT 5mM), 2 µl poly [dI-dC] (1 µg/µl, Sigma), 1 à 2 µl de sonde (soit 10-20 ng), 1 à 2 µg de compétiteur éventuel, qsp 20 µl d'eau distillée. Après 20 minutes d'incubation à température ambiante et ajout de 2 µl de bleu de dépôt (Bleu de bromophénol, xylène cyalol, Ficoll 20%), les échantillons sont déposés sur un gel non

dénaturant (polyacrylamide 19/1, 4%). La durée de la migration (effectuée à 170-200V) dépend de la taille de la sonde : jusqu'au premier bleu (bleu de bromophénol) pour les sondes courtes (20 pb), presque jusqu'au second (xylène cyanol) pour les sondes longues (80 pb). Après la migration, le gel est transféré sur papier Whatman 3MM, enveloppé dans du Saran et directement autoradiographié (sans séchage) une nuit à -80°C sur film Kodak Biomax MS.
Dans certains cas (sonde issue d'oligonucléotides de synthèse de plus de 60 pb), une étape de purification supplémentaire de la sonde est nécessaire pour éliminer le voile de migration (smear): pour cela, la sonde fraîchement marquée est déposée sur un gel polyacrylamide 4% non dénaturant, qui est autoradiographié après électrophorèse. Le signal correspondant à la sonde sur le film permet de repérer et d'exciser celle-ci du gel. Le fragment de gel est écrasé dans 300 µl de TE 0,1X (Tris HCl 1mM pH 7,5; EDTA 0,1 mM) et incubé une nuit à 37°C sous forte agitation, avant d'être passé sur fibre de verre siliconée. La sonde purifiée est alors précipitée à l'alcool absolu puis resuspendue dans du TE 0,1X.

IV- UV-CROSSLINKING

Si le gel retard est une technique intéressante pour analyser la fixation de facteurs protéiques sur un fragment d'ADN, il ne permet pas en revanche de déterminer le nombre et la taille de ces facteurs car la migration du complexe se fait sur un gel non dénaturant (les protéines migrant essentiellement en fonction de leur pH isoélectrique). La technique d'UV-crosslinking pallie cet inconvénient en formant des liaisons covalentes entre la molécule d'ADN et les protéines qui interagissent avec elle. De plus, cet évènement survenant très rarement, une seule protéine établit statistiquement des liaisons covalentes avec le fragment d'ADN. Ainsi, même dans le cas où plusieurs molécules fixent la même sonde, les complexes covalents produits par irradiation UV contiennent un seul facteur protéique. Ces complexes sont ensuite séparés sur un gel dénaturant (SDS-PAGE) en présence d' un marqueur de taille, ce qui permet la détermination du nombre et de la taille des protéines fixant la sonde (voir aussi les "Current Protocols in Molecular Biology", section 12.5; Chodosh, 1996).
Pratiquement, les conditions d'interaction entre ADN et protéines sont identiques à celles du gel retard, mais avant leur migration, les échantillons sont irradiés aux UV (longueur d'onde 254 nm, lampe UV 4W à 5 cm des échantillons) de 30 minutes à 2 heures, puis bouillis dans du tampon de Laemmli 1X (Laemmli, 1970) et déposés sur gel SDS-PAGE 10%. Après migration, le gel est transféré sur papier Whatman 3MM, enveloppé dans du Saran et autoradiographié (alternativement, il peut être séché avant exposition).

C- FACTEURS DE TRANSCRIPTION CELLULAIRES ET VIRAUX: PURIFICATION ET RECHERCHE D'INTERACTIONS PROTEINE-PROTEINE

I- PURIFICATION

1. Protéines cellulaires ayant de l'affinité pour le motif CRE

1.1. Colonne d'affinité

Pour identifier les protéines présentes dans les extraits nucléaires *Sf*9 qui fixent le motif CRE, nous avons au départ adopté une stratégie de purification par affinité avec le motif CRE. Pour cela, nous avons utilisé des oligonucléotides CRE/EB3 et CREmut1 biotinylés en 5' sur un seul brin, dont la séquence est identique à celle présentée dans la figure 2 de l'article 1. Des tests préliminaires en gel retard ont montré que ces oligonucléotides sont capables de fixer les facteurs cellulaires de manière analogue aux oligonucléotides CRE/EB3 non biotinylés. Nous avons donc effectué plusieurs essais de purification par billes magnétiques couplées à des molécules de streptavidine (Magnesphere, Promega). Brièvement, les oligos biotinylés ont été incubés avec des extraits nucléaires de cellules *Sf*9 pendant 1h à 4°C, puis capturés par les billes magnétiques. Après 4 rinçages, les billes ont été reprises dans du tampon de Laemmli 1X, bouillies et le surnageant déposé sur gel SDS-PAGE, celui-ci étant révélé au bleu de Coomassie et au nitrate d'argent. Nous avons également testé une variante de cette méthode, en couplant préalablement les oligos aux billes, puis en incubant l'ensemble avec les extraits nucléaires.

1.2. Fractionnements successifs

Les purifications suivantes ont été effectuées avec un appareil de chromatographie basse pression (Äkta Prime, Amersham) qui comprend un module de mesure de la densité optique en sortie de colonne et un module mélangeur permettant l'obtention d'un gradient continu entre deux solutions A et B.

Les extraits nucléaires *Sf*9 sont passés sur une colonne de gel filtration permettant un fractionnement des protéines selon la taille, de 20 à 350 kDa (résine AcA 34, Bioserpa). La colonne (Ø 16 mm, 100 ml de résine) a été préalablement équilibrée par passage de 2 volumes de tampon C, avec un débit de 0,5 ml/min. Le débit est établi à 0,2 ml/min pendant toute la durée du fractionnement, le passage des extraits en sortie de colonne étant contrôlé par les variations de DO à 280 nm. Les extraits nucléaires fractionnés (2 ml par 2 ml, soit environ 50 tubes) sont alors testés en gel retard avec la sonde CRE. Les fractions contenant des protéines ayant de l'affinité pour le motif CRE sont dialysées dans 100 volumes de tampon D (Hepes 20

mM pH 7,9 ; glycérol 20% ; KCl 0,1 M; EDTA 0,2 mM; DTT 0,5 mM; Complete 1X, Roche), en boudin de dialyse (limite d'exclusion : 8 kDa; Spectra/Por, Spectrum Laboratories). Après centrifugation, le précipité insoluble est éliminé et le surnageant passé sur une colonne d'affinité (colonne d'héparine 1 ml, HiTrap Heparin HP column, Amersham Bioscience) selon les recommandations du fournisseur et le protocole décrit dans l'article 1. Les fractions issues de cette colonne sont testées en gel retard, et celles qui contiennent des protéines affines pour CRE sont dialysées et passées sur une colonne échangeuse de cations (HiTrap SP HP column, Amersham Bioscience) également selon les recommandations du fournisseur et le protocole décrit dans l'article 1. Ces fractions sont alors soumises à un dernier gel retard pour sélectionner celles qui gardent l'activité de fixation sur CRE. Parallèlement, à chaque étape de purification, les fractions contenant des protéines affines pour CRE sont testées en gel SDS-PAGE 10% avec une révélation au nitrate d'argent.

2. Purification du facteur viral IE1

2.1 FLAG-IE1 produit en système hétérologue

2.1.1. Système bactérien

La séquence codante de IE1 a été amplifiée par PCR à partir du plasmide pIE1 (Ohresser *et al*., 1994) en utilisant des amorces 5' (CGTGAACAACC**AAGCTT**ATGACGCA) et 3' (TGTA**AGATCT**TTAAATTCAAAATAT) portant les sites *Hind*III et *Bgl*II, respectivement (marqués en gras). Le fragment PCR obtenu, préalablement digéré par ces deux enzymes, a ensuite été cloné dans le vecteur d'expression bactérien pFLAG-MAC (Sigma) aux mêmes sites. La souche bactérienne XlI-Blue a été transformés avec ce plasmide recombinant (pFLAG-MACIE1) et l'expression de FLAG-IE1 induite 2h à température ambiante par ajout d'IPTG (2 mM final) dans le milieu de culture (3ml, DO_{600}: 1). Le culot bactérien a été soniqué sur glace (deux pulses de 20 secondes, puissance 3) dans 400 µl de tampon TEDG (Tris-HCl pH 7,5 50 mM; EDTA 1,5 mM; glycérol 10%; NaCl 0,4 M; DTT 1 mM; Complete 1X) puis centrifugé 10 min. à 10000 g. IE1 a ensuite été purifié à partir du surnageant par passage sur colonne d'affinité anti-FLAG (80 µl d'ANTI-FLAG M2 Affinity Gel, Sigma) préalablement lavée/rincée selon les recommandations du fournisseur: 300 µl de surnageant bactérien ont été ajoutés sur la colonne puis, après 3 lavages avec 12 volumes de tampon TBS, IE1 a été élué soit (i) par compétition avec le peptide FLAG (100 µg/ml; FLAG peptide, Sigma), soit (ii) par clivage du FLAG avec une entérokinase (Sigma) selon les recommandations du fournisseur.

2.1.2. Système cellulaire Sf9 (transfection)

Nous avons utilisé le plasmide pIE1, qui a déjà fait ses preuves pour l'expression transitoire de IE1 (Ohresser *et al*., 1994), dans lequel nous avons inséré au niveau de l'ATG du gène la séquence codante du peptide FLAG. Pour cela, deux sites de restriction (*Spe*I et *Sal*I) ont été créés par mutagenèse dirigée au niveau de la boîte de Kozak et de l'ATG du gène IE1 (figure 1A et B), générant le plasmide pIE1-SpeSal. La séquence correspondant au peptide FLAG a ensuite été reconstituée par hybridation d'oligonucléotides complémentaires puis clonée au niveau de ces sites (avec restauration à l'identique de la séquence de Kozak et de l'ATG) en fusion avec la séquence codante du gène IE1 (plasmide pFLAG-IE1, figure 1C). Le protocole de transfection de ce plasmide dans la cellule *Sf*9 est identique à celui décrit dans l'article 1, et la purification de IE1 analogue à celle décrite dans la section précédente.

Notons que le plasmide pIE1-SpeSal peut être utilisé pour l'expression transitoire d'autres gènes que IE1, avec trois précaution à prendre lors du clonage: (i) un deuxième site *Spe*I a été généré entre les sites *Spe*I et *Sal*I initiaux (figure 1B), ce qui nécessite une digestion en deux temps du plasmide, d'abord par *Sal*I puis par *Spe*I, (ii) un codon sérine est ajouté juste en aval de l'ATG (figure 1C), et (iii) le fragment à cloner doit comprendre la séquence codante du gène à exprimer + la boîte de Kosak (GTGACT), l'ATG et le codon stop.

2.2. Purification partielle de IE1 produit en système baculovirus/cellule d'insecte

Les extraits nucléaires issus de cellules *Sf*9 infectées par le virus *Ac*MNPV (récupérées 4 h post-infection) sont préparés selon le protocole décrit dans la section B.1. Après dialyse dans le tampon D, ils sont passés directement sur colonne échangeuse de cations (HiTrap SP HP column, Amersham Bioscience). Les fractions obtenues, testées par western blot en utilisant un anticorps polyclonal anti IE1 (Ohresser *et al*., 1994), sont passées après dialyse sur colonne d'héparine. La fraction contenant la plus grande quantité de IE1 (estimée par western blot) est alors conservée.

II- CARACTÉRISATION DE L'INTERACTION IE1/CBP PAR CO-IMMUNOPRÉCIPITATION

Des cellules *Sf*9 à 70% de confluence (1 boîte Falcon F75, soit environ 12 millions de cellules) sont infectées à une MOI de 10 soit par le virus *Ac*MNPV sauvage, soit par un baculovirus recombinant exprimant sous le contrôle du promoteur de la polyédrine la CBP

humaine portant un peptide FLAG à son extrémité C-terminale (Kurokawa *et al*., 1998). Après 48 heures d'infection, les cellules sont collectées, rincées dans du tampon PBS et lysées par trois cycles de congélation dans l'azote liquide /décongélation à 37°C dans 100 µl de PBS + Complete 1X (cocktail d'antiprotéases, Roche). Après élimination des débris cellulaires par centrifugation, 90 µl de chacun des lysats de cellules infectées par le virus sauvage ou recombinant sont incubés avec 20 µl d'anticorps monoclonal anti-FLAG couplé à des billes d'agarose (Anti-FLAG M2 affinity gel, Sigma) 2 heures à 4°C. Après 5 lavages dans 10 volumes de tampon TBS froid (Tris HCl 50 mM pH 7,4 ; NaCl 150 mM), les billes sont bouillies dans 20 µl de tampon de tampon de Laemmli 1X (Laemmli, 1970), et les surnageants analysés en western blot avec un anticorps polyclonal anti IE1 (Ohresser *et al*., 1994).

D- BANQUE D'ADN COMPLEMENTAIRE DE CELLULES *Sf*9 EN PHAGE LAMBDA

I- CONSTRUCTION

Nous avons utilisé le kit "SMART cDNA library construction kit" de Clontech pour réaliser la banque d'ADN complémentaires *Sf*9. Pour les protocoles précis et la composition des milieux, on peut se référer au manuel de l'utilisateur (# K1051-1 du catalogue Clontech). Le principe de la construction de la banque est schématisé sur la figure 2 : 1 µg d'ARN messagers extraits de la lignée cellulaire *Sf*9 sont rétrotranscrits par une transcriptase inverse (Superscript II, Gibco BRL) à partir d'une amorce oligo dT hybridée sur la queue polyadénylée des ARN messagers. Cette transcriptase inverse ajoute à l'extrémité 5' du brin d'ADN néosynthétisé une série de 3 C, qui permet l'hybridation d'une deuxième amorce oligo dG et la synthèse du brin complémentaire d'ADN. Les deux amorces contenant un site de restriction *Sfi*I, la digestion des ADNc par cette enzyme permet leur clonage dans le phage λTriplEx2 au même site. Le choix de cette enzyme est astucieux car son site de restriction particulier (GGCCNNNNGGCC, où les 5 nucléotides centraux sont indifférents) permet, en n'utilisant qu'une seule enzyme de restriction, un clonage directionnel des ADNc dans le phageλ: les sites *Sfi*IA (en 5') et *Sfi*IB (en 3') ont en effet une séquence différente (GGCC**ATTAT**GGCC et GGCC**GCCTC**GGCC, respectivement). De plus, ce site de restriction est rare, ce qui limite les risques de coupures au milieu de la séquence de l'ADNc. Les ADNc ainsi digérés sont ligués à l'ADN phagique (prédigéré par *Sfi*I) par une ADN ligase. Cette étape de ligation aboutit à la formation d'une banque d'ADN de phages recombinants, chacun portant un ADNc différent. Leur encapsidation (kit Gigapack Gold III, Amersham) aboutit à la constitution d'une banque de phage non amplifiée.

Cette banque non amplifiée est titrée par la méthode des plages de lyse, ce qui permet de vérifier sa qualité: une banque représentative de la population d'ARN messagers de départ doit avoir un titre supérieur à 10^6 plages de lyse par ml. Dans notre cas, le titre était très satisfaisant, avec environ $4x10^6$ plages de lyse/ml et 98% de recombinants (pourcentage estimé par criblage blanc/bleu).

II- AMPLIFICATION

L'intégralité de cette banque est utilisée pour une amplification en masse qui se réalise également par la méthode des plages de lyse, les phages présents dans les plage étant ensuite élués à 4°C pendant toute la nuit. Après récupération et clarification, le surnageant est titré, une amplification correcte donnant un titre au moins égal à 10^{10} plages de lyse/ml. Dans notre cas, la banque *Sf*9 amplifiée présentait un bon titre, de l'ordre de $2x10^{10}$ plages/ml. La banque est stable six mois à 4°C et un an à -80°C lorsqu'elle est conservée dans du DMSO 7%. Après un an, le titre décroît, ce qui nécessite une réamplification périodique.

III- CONVERSION DE LA BANQUE DE PHAGES EN BANQUE DE PLASMIDES

Les phages sont des vecteurs performants pour réaliser une banque d'ADNc (taille d'inserts très large, encapsidation de l'ADN indépendante d'une étape de transformation souvent peu efficace, titres élevés, recombinaison limitée, stabilité génétique, facilité de clonage et de manipulation). Par contre, l'extraction et la manipulation de l'ADN phagique, long et linéaire, est beaucoup plus délicate que l'ADN plasmidique. Par conséquent, il n'est pas envisageable d'utiliser directement la banque de phages pour toutes les opérations impliquant une purification à grande échelle de l'ADN, comme par exemple pour le séquençage d'un nombre élevé de clones. Dans cette optique, λTriplEx2 possède un système performant permettant une conversion aisée des phages en plasmides. La conversion du λTriplEx2 en pTriplEx2 se fait grâce au système enzymatique cre recombinase/loxP, par lequel le phage recombinant est excisé au niveau de sites loxP placés de part et d'autre de l'insert (figure 1), le fragment résultant étant circularisé pour former un plasmide. Ce mécanisme bactérien naturel de défense contre les phages est présent dans la souche bactérienne BM25-8, qui exprime la CRE recombinase à 31°C. Le protocole de conversion est donc extrêmement simple (détaillé dans le manuel du kit de construction de la banque), puisqu'il suffit d'infecter cette souche bactérienne avec les phages à convertir et de l'incuber ensuite à 31°C. On obtient ainsi des clones bactériens contenant le plasmide recombinant issu de l'excision du phage, qui sont facilement criblables par antibiotique

puisque le plasmide contient un gène de résistance à la carbenicilline/ampicilline. L'extraction et le traitement des plasmides ainsi obtenus est tout à fait classique, et donc automatisable.

IV- MISE EN COLLECTION ET SÉQUENÇAGE

Après conversion de la banque *Sf*9 amplifiée, 10000 clones bactériens isolés ont été repiqués manuellement en milieu LB, amplifiés et mis en collection après ajout de glycérol (20% final) à -80°C. Dans le même temps, chacun d'entre eux a été repiqué sur boîte de Pétri LB/ampicilline (100 clones par boîte) et envoyé au Dr Mita (Institut national de Sciences agrobiologiques, Japon) pour séquençage (Okano *et al*., 2001).

V- CRIBLAGE DE LA BANQUE DE PHAGES PAR SONDE NUCLÉIQUE

Cette technique permet de rechercher rapidement dans la banque d'ADNc les clones correspondant à un gène dont une partie au moins de la séquence est connue. Le protocole est détaillée dans le manuel "Lambda Library Protocol Handbook" (# PT1010-1 de Clontech). 10^6 plages de lyse (20 boîtes de Petri de 13 cm de diamètre, soit 50000 plages par boîte) sont produites à partir de la banque amplifiée, et les phages qu'elles contiennent sont transférés sur membrane de nitrocellulose (Schleicher and Schuell), qui sont ensuite traitées à la soude (pour dénaturer l'ADN) et fixées 3 minutes aux UV. Ces 20 membranes sont ensuite hybridées avec la sonde idoine et autoradiographiées (figure 3A), ce qui permet de repérer sur les boîtes de Petri les plages de lyse hybridant avec la sonde. Le milieu gélosé qui les contient est carotté, et les phages, élués dans un tampon d' élution, sont soumis à un deuxième tour de purification par plage de lyse. Une nouvelle hybridation permet de contrôler la présence et surtout l'enrichissement du clone d'intérêt. Cette procédure est répétée jusqu' à ce que la population soit clonale (figure 3B), c'est-à-dire que toutes les plages de lyse issues de l'élution précédente présentent un signal positif. Une plage de lyse positive est alors carottée, les phages qu'elle contient sont amplifiés et convertis en plasmides bactériens par le système CRE recombinase/loxP précédemment décrit. Le séquençage permet finalement de sélectionner un ou plusieurs plasmides portant l'ADNc recherché.

Des facteurs cellulaires sont impliqués dans la transactivation par le couple IE1/hrs

A- INTRODUCTION ET RAPPELS

I- INTRODUCTION

Le développement d'un eucaryote complexe nécessite la transcription coordonnée dans le temps et dans l'espace d'au moins 20000 gènes. Dans ce cadre, comment un organisme peut-il gérer une telle diversité d'évènements tout en maintenant la spécificité cellulaire et en répondant dynamiquement à l'environnement? L'une des solutions est d'utiliser un répertoire restreint d'activateurs pour réduire la complexité des voies de signalisation cellulaire et de l'intégration des signaux. Dans ce cadre, la cellule utilise des mécanismes impliquant la coopérativité et la synergie transcriptionnelle, par lesquels les différentes combinaisons possibles à partir d'un nombre limité d'activateurs (aussi bien ubiquitaires que tissu- ou signal- spécifiques) augmente les possibilités de régulation. Ainsi, les régions activatrices en *cis* (enhancers), qui possèdent des sites de fixation pour un petit nombre de facteurs de transcription, vont répondre à un signal donné par leur recrutement, la combinaison particulière de ceux-ci favorisant leurs interactions et leur fixation sur l'ADN de manière coopérative. Le complexe spécifique ainsi formé (l'enhanceosome) va alors stimuler la transcription de manière synergique, c'est-à-dire plus importante que ne l'aurait fait l'addition des effets de chacun des facteurs pris séparément (Carey, 1998).

Dans les travaux présentés ci-après, nous allons voir que certaines régions activatrices du baculovirus *Ac*MNPV et les facteurs qu'elles fixent pourraient fonctionner comme un enhanceosome viral.

L' analyse détaillée d'une des régions homologues du baculovirus *Ac*MNPV, hr5, déjà connue pour être une région activatrice de la transcription des gènes du baculovirus (Carson *et al*., 1991; Leisy *et al*., 1995; Lu et Carstens, 1993; Nissen et Friesen, 1989), a permis de caractériser de nombreux motifs TRE et CRE, qui sont la cible de facteurs de transcription des familles AP1 et CREB/CREM/ATF (Ogliastro, 1994). Comme le génome du baculovirus ne contient pas de gènes codant ce type de facteurs, cette découverte suggère que des protéines de l' hôte pourraient interagir avec ces sites et être fonctionnellement importants pour l'activation transcriptionnelle des gènes du baculovirus. Nous avons donc mené une étude plus approfondie des sites CRE et TRE présents dans les hrs et des facteurs cellulaires qu'ils fixent, aboutissant à la proposition d'un modèle de type "enhanceosome" pour l'activation de la transcription baculovirale médiée par les régions hrs.

II- RAPPELS

Après avoir rappelé les principales caractéristiques des baculovirus et l'état des connaissances concernant la régulation de la transcription chez ce virus, nous étudierons plus en détail les régions hrs et leurs interactions (avérées ou putatives) avec différents facteurs de transcription viraux et cellulaires. Les caractéristiques générales des facteurs de transcription de la famille AP1 et CREB/CREM/ATF, dont les sites de fixation sont présents dans les hrs, ainsi que le détail du fonctionnement d'un enhanceosome seront également abordés, toutes ces notions étant indispensables à la fois pour poser le problème biologique et pour éclairer les résultats obtenus au cours du travail de thèse (exposés dans l'article 1).

1. Les baculovirus

1.1. Histoire et importance

La région d'Alès était au 19ème siècle un centre important de l'élevage du ver à soie, *Bombyx mori*. Dans ces élevages sévissait une maladies très contagieuse, appelée "grasserie", à laquelle le poète latin Vida faisait déjà allusion en 1527 dans son poème "De Bombyce" (traduction de Bonnafous):

"*Dès que d'une eau saline il ressent l'amertume,*
Un invisible fléau l'irrite et le consume;
Il se courbe, il se dresse, il se recourbe encor,
Languit, s'énerve et prend l'éclat trompeur de l'or;
La fibre se déchire, un mal impur l'affecte.
En vain Cypris naquit au sein des flots amers,
Cypris ne put dompter l'âcre saveur des mers.
Il est même, grands Dieux! oserai-je le dire?
Des hommes enivrés d'un coupable délire
Qui de cette eau funèbre arrosent les rameaux;
Garde donc, jour et nuit, tes frêles animaux."

Les symptômes consistent en effet en une liquéfaction totale de la larve avec émission d'un liquide blanc et gras (d'où le nom de la maladie, figure 4), conduisent à une mort rapide (une semaine en moyenne) des larves. Louis Pasteur (pas encore inventeur du vaccin antirabique),

viendra étudier la grasserie en 1865 à la magnanerie de St Christol (pas encore laboratoire de Pathologie Comparée), mais ne parviendra pas à identifier l'agent pathogène responsable, et pour cause: il s'agit d'un virus qui ne sera caractérisé que bien plus tard, avec l'avènement de la microscopie électronique. Sa forme en bâtonnet (baculus en latin) et la découverte que de nombreux autres virus, infectant essentiellement les arthropodes et plus particulièrement les lépidoptères, présentent les mêmes caractéristiques permettront alors leur regroupement dans la famille des baculovirus (Federici, 1997).

Les baculovirus sont parmi les plus étudiés des virus d'invertébrés en raison des deux intérêts majeurs qu'ils représentent: d'un point de vue agronomique tout d'abord, ils sont employés en tant qu'agents de lutte biologique contre un grand nombre de lépidoptères ravageurs (pour revue, (Wood et Granados, 1991). A l'heure actuelle, même si leur emploi reste marginal face aux pesticides chimiques, plusieurs pays en développement les utilisent avec succès, au Brésil notamment. Le deuxième intérêt des baculovirus est biotechnologique: en effet, c'est l'un des systèmes d'expression eucaryotes les plus utilisés, permettant l'expression en grandes quantités de protéines recombinantes (Lopez-Ferber *et al*., 1995) sans les inconvénients des systèmes mammifères (possibilités de contamination par des pathogènes transmissibles à l'Homme) ou végétaux (problèmes de glycosylation des protéines).

Enfin, le fait même que les baculovirus soient les virus d'invertébrés les mieux caractérisés au niveau moléculaire en font un modèle de choix pour l'étude des relations hôte/virus chez les invertébrés.

1.2. Caractéristiques et cycle viral

Les baculovirus sont des virus à ADN double brin circulaire de grande taille (90 à 180 kb) enveloppés par une membrane et inclus dans une matrice protéique cristalline, le corps d'inclusion. On distingue deux sous-familles, les granulovirus (un seul virion par corps d'inclusion, le granule, composé de granuline) et les polyédrovirus (plusieurs virions par corps d'inclusion, le polyèdre, composé de polyédrine) (Figure 5; Murphy *et al*., 1995). Parmi ces derniers, le baculovirus d'*Autographa californica* (*Ac*MNPV) est le vecteur d'expression le plus employé en biotechnologies, le virus le plus étudié et donc le principal modèle de la famille. Sa spécificité d'hôte étant large (Bishop *et al*., 1995; Payne, 1986), il est capable d'infecter plusieurs lignées cellulaires d'espèces de lépidoptères différentes, les plus utilisées étant *Sf*21 et l'un de ses clones dérivés, *Sf*9, issues de *Spodoptera frugiperda*.

Le cycle biologique d'*Ac*MNPV est schématisé sur la figure 6: les polyèdres présents sur la plante nourricière sont ingérés par la larve au cours de son alimentation. Dans le mésentéron, le

pH alcalin provoque la dissolution des polyèdres, libérant les virions qui vont passer à travers la membrane péritrophique (Derksen et Granados, 1988), pénétrer dans les cellules (par fusion de la membrane des virions et de la cellule; Granados *et al*., 1981) et aller jusqu'au noyau (1-4 hrs p.i.). Là, ils vont effectuer plusieurs cycles de réplication aboutissant à la formation de virions non occlus: c'est l'infection primaire. Ces virions vont alors être dirigés vers le pôle basal de la cellule épithéliale, bourgeonner vers l'espace intercellulaire et acquérir une enveloppe à partir de la membrane plasmique cellulaire. Ils vont ensuite envahir tous les autres tissus en passant par l'hémolymphe et le système trachéen (Barrett *et al*., 1998; Engelhard *et al*., 1994): c'est l'infection secondaire, qui va produire dans un premier temps des virions non occlus responsables de l'infection de cellule à cellule. Dans un deuxième temps (24-48 hrs p.i.), les virions vont rester dans la cellule où ils ont été produits, être enveloppés dans une membrane produite *de novo* puis inclus dans une matrice cristalline de polyédrine, formant le polyèdre (Lee *et al*., 1996). Dans ces phases tardives de l'infection, une autre protéine, p10, est également produite, ses amas fibrillaires cytoplasmiques et nucléaires favorisant la rupture du noyau et la lyse cellulaire (van Oers *et al*., 1993) (figure 7). Sous l'action de protéases et d'une chitinase virale (Hawtin *et al*., 1997), l'exosquelette de la larve est dégradé, et son contenu liquéfié disperse des millions de polyèdres sur la plante. Ceux-ci, très stables, protègent les virions occlus pendant de longues périodes et permettent la transmission horizontale du virus.

1.3. Contrôle général de l'expression des gènes

La régulation de l'expression des gènes chez les baculovirus se fait principalement au niveau transcriptionnel et par vagues successives, chacune étant contrôlée par les produits de la phase précédente. Comme le montre la figure 8, on distingue 4 phases principales: très précoce (immediate-early ou IE) qui débute presque immédiatement après l'entrée du virus dans la cellule, précoce retardée (delayed-early, ou DE), tardive (late, ou L) et très tardive (very late, VL).

1.3.1. La phase très précoce

A son entrée dans la cellule, le virus ne possède pas de machinerie transcriptionnelle propre. Il va donc utiliser celle de l'hôte pour transcrire 4 gènes codant pour des protéines activatrices de la transcription (IE0, IE1, IE2 et PE38; Friesen, 1997) essentielles pour le bon déroulement de la phase suivante. Les promoteurs de ces gènes viraux possèdent une structure très similaire à celle des promoteurs cellulaires dépendants de l'ARN polymérase II cellulaire, avec une TATA box à 28-31 pb en amont du site d'initiation de la transcription (Blissard *et al*., 1992; Guarino et Smith, 1992) et un motif d'initiation CAGT caractéristique des arthropodes (Cherbas et Cherbas, 1993;

Dickson et Friesen, 1991), ce qui leur permet d'être fonctionnels en l'absence de facteurs viraux. Ceci explique que le génome nu soit infectieux lorsqu'il est introduit dans la cellule, et permet d'utiliser ces promoteurs dans des constructions plasmidiques en cellules non infectées (Cartier *et al*., 1994; Jarvis *et al*., 1990; Ohresser *et al*., 1994).

Les 4 protéines produites au cours de cette phase sont toutes des activateurs transcriptionnels, mais, à part IE1 que nous développerons dans la section suivante, leurs propriétés et leurs rôles ne sont pas très bien caractérisés. Certaines d'entre elles sont capables de stimuler leur propre promoteur (IE1, IE2) (Yoo et Guarino, 1994) ou les autres promoteurs IE (IE0 transactive IE1, par exemple) (Kovacs *et al*., 1991).

1.3.2. La phase précoce retardée

Par définition, cette phase dure jusqu'au début de la réplication de l'ADN viral. Un grand nombre de gènes essentiels au contrôle de l'hôte (contrôle de la mue, blocage des réponses cellulaires antivirales) et à la réplication virale (ADN polymérase, hélicase) y sont exprimés. L'expression des gènes précoces est également dépendante de la machinerie transcriptionnelle et de l'ARN polymérase II cellulaire, mais nécessite en plus les produits de la phase précédente. Les promoteurs de ces gènes ont une structure semblable à celle des promoteurs cellulaires, mais la présence d'éléments de réponse aux transactivateurs très précoces est indispensable pour que la transcription soit efficace. Ces éléments de réponse peuvent être situés soit dans le promoteur (mais ils sont mal caractérisés) soit dans des régions activatrices distales (enhancers), beaucoup mieux connues (Friesen, 1997). C'est le cas par exemple des régions homologues (hrs) qui fixent IE0 et IE1, dont nous détaillerons le fonctionnement dans la prochaine section.

1.3.3. La phase tardive

Elle débute avec la réplication du virus, et conduit à la production de virions libres, ou "budded virus" (BVs). Elle se caractérise par l'expression d'une ARN polymérase virale résistante à l'α-amanitine (Beniya *et al*., 1996), à la différence de son homologue cellulaire. Cette enzyme va permettre la transcription des gènes tardifs, qui codent majoritairement des protéines de structure. Leur promoteur présente une structure particulière, avec un motif TAAG (la boîte de Rohrman) essentiel à la reconnaissance par l'ARN polymérase virale (pour revue, (Lu et Miller, 1997).

1.3.4. La phase très tardive

Elle coïncide avec l'apparition et la production des virions occlus, et se traduit par la très forte expression de deux protéines, la polyédrine (qui va former la matrice cristalline des polyèdres) et p10 (dont le(s) rôle(s) est mal connu). Les promoteurs de ces deux gènes possèdent tous deux la boîte de Rohrman, des séquences "suractivatrices" (burst sequences) situées dans la région non traduite en 5' du gène étant vraisemblablement responsables de la transcription très intense de ces gènes (Ooi *et al*., 1989).

2. Le couple IE1/hrs

Tous les baculovirus possèdent un gène ie1 plus ou moins conservé et beaucoup contiennent des régions présentant une structure particulière, les régions homologues (hrs). Chacun de ces deux éléments viraux présente individuellement une importance particulière pour le cycle du baculovirus, mais c'est surtout l'étude de leur interaction et de ses effets qui a permis de mettre en évidence l'importance du couple IE1/hrs.

2.1. Les protagonistes: présentation et structure

2.1.1. Les régions homologues (hrs)

Les hrs ont été trouvées chez un grand nombre de baculovirus, comme *Ac*MNPV (Ayres *et al*., 1994), *Op*MNPV (Ahrens *et al*., 1997), *Cf*MNPV (Xie *et al*., 1995), *Ag*MNPV (Garcia-Maruniak *et al*., 1996), *Ld*MNPV (Kuzio *et al*., 1999), *Bm*NPV (Gomi *et al*., 1999), *Px*GV (Hashimoto *et al*., 2000). Ce sont des séquences composées d'éléments répétés les uns à la suite des autres, présentes en plusieurs exemplaires fortement conservés (d'où leur nom) et réparties dans tout le génome (figure 9). Chacun des élément répété porte typiquement un palindrome imparfait qui constitue le site de fixation pour le transactivateur viral précoce IE1 (Friesen, 1997). Ces séquences contiennent également de nombreux sites de fixation pour des facteurs de transcription appartenant aux familles AP1 et CREB/CREM/ATF (figure 10), dont nous exposeront ultérieurement les principales caractéristiques (section 3.2).

2.1.2. La protéine Immediate Early 1 (IE1)

Des homologues du gène ie1 de *Ac*MNPV ont été identifiés chez tous les baculovirus étudiés jusqu'à présent, ce qui donne une indication sur son rôle central dans la biologie des baculovirus. Les transcrits ie1 sont détectés immédiatement (15 minutes) après l'entrée du virus dans la cellule et la protéine atteint son niveau maximal 6 heures après le début de l'infection (Chisholm et Henner, 1988). Son niveau reste élevé pendant toute la durée du cycle viral (Choi et Guarino,

1995b; Ohresser *et al*., 1994), y compris durant les phases tardives, ce qui suggère que le promoteur du gène répond aussi à l'ARN polymérase virale (Pullen et Friesen, 1995). La protéine localise dans le noyau, mais ses signaux d'adressage nucléaires n'ont pas été identifiés. Sa structure est modulaire, avec un domaine N-terminal acide impliqué dans l'activation transcriptionnelle et un domaine C-terminal permettant son oligomérisation (figure 11A). Le(s) domaine(s) impliqué(s) dans la fixation de IE1 sur l'ADN n'ont pas encore été déterminés (Kovacs *et al*., 1992; Rodems *et al*., 1997). Ces deux domaines sont fonctionnels séparément, ce qui est une caractéristique générale des protéines transactivatrices.

L'expression de IE1 en lysat de réticulocyte a montré que la forme monomérique est capable de se fixer sur un demi-palindrome en l'absence d'autres facteurs, viraux ou cellulaires (Choi et Guarino, 1995a). Par contre, la transactivation par IE1 nécessite la dimérisation préalable de la protéine puis la fixation du dimère sur le palindrome complet (figure 11B; Rodems et Friesen, 1995).

Notons que chez plusieurs baculovirus (*Ac*MNPV, *Op*MNPV, *Ld*MNPV...), une forme plus longue de IE1 est produite au début de l'infection: IE0. Le promoteur de ie0 est situé 4 kb en amont de l'ATG du gène ie1, et le transcrit primaire est épissé de telle sorte qu'un court exon en 5' est fusionné à la séquence codante de ie1. IE0 possède donc une extension de 54 acides aminés à l'extrémité N-terminale. Les transcrits ie0 sont détectés aussi précocément que les transcrits ie1, mais leur expression décroît rapidement 6 heures p.i. IE0 possédant le même domaine C-terminal que IE1, il est également capable de se fixer sur les hrs (Chisholm et Henner, 1988).

2.2. Les protagonistes: rôles connus

2.2.1. hrs

2.2.1.1. transactivation

Les régions hrs sont des activateurs transcriptionnels qui sont fonctionnels à la fois dans la cellule non infectée et dans le contexte de l'infection virale. Dans la cellule non infectée, les séquences hrs possèdent des propriétés typiques de régions activatrices en *cis* (enhancers): 1) elles sont capables d'activer un promoteur placé en aval quel que soit leur sens (direct ou inversé; Rodems et Friesen, 1993), et 2) l'activation est indépendante de la distance entre hr et promoteur (Rodems et Friesen, 1993). Cependant, dans le contexte de la cellule infectée (donc en présence de facteurs viraux), la transactivation médiée par les hrs est beaucoup plus efficace qu'en cellule non infectée. L'identification du transactivateur viral IE1 et de son site de

reconnaissance palindromique dans les hrs ont permis de préciser le mécanisme d'activation de la transcription des gènes viraux précoces (Choi et Guarino, 1995a; Guarino et Dong, 1991; Kovacs *et al*., 1992; Leisy *et al*., 1995; Rodems et Friesen, 1995): un grand nombre d'études ont ainsi montré que la plupart des promoteurs de ces gènes dépendants de l'ARN polymérase II sont fortement stimulés par le couple IE1/hr, comme p35 qui présente une activation de la transcription 3000 fois supérieure en présence de IE1 (Nissen et Friesen, 1989). Des promoteurs cellulaires comme celui de l'actine A3 de *B. mori* sont également activables par le système IE1/hrs (Lu *et al*., 1997). Le couple IE1/hrs semble donc être un système activateur puissant et universel de la transcription dépendante de l'ARN polymérase II cellulaire. Par contre, les promoteurs des gènes viraux tardifs, qui dépendent de l'ARN polymérase virale, ne semblent pas être activables par ce système.

2.2.1.2. réplication

A côté de ce rôle avéré dans la transcription, les hrs jouent probablement aussi un rôle fondamental dans la réplication du virus, en tant qu'origine de réplication: plusieurs travaux ont en effet montré que ces séquences sont indispensables à la réplication de plasmides in vitro (Lee et Krell, 1994; Leisy *et al*., 1995; Pearson et Rohrmann, 1995). Cependant, la délétion simultanée de tous les hrs conduisant à un virus non viable, aucune preuve directe n'a été apportée à ce jour. La délétion d'un seul hr (en l'occurence hr5 de *Ac*MNPV) n'a ainsi eu aucun effet détectable sur la réplication, suggérant une compensation par les autres hrs (Rodems et Friesen, 1993).

2.2.2. IE1

2.2.2.1. transactivation

Comme nous l'avons déjà abordé, IE1 est la principale protéine transactivatrice de l'expression des gènes précoces du baculovirus, par deux mécanismes distincts: (i) dans la transactivation dépendante des hrs, IE1 se fixe en dimère sur les palindromes et active la transcription des promoteurs contrôlés par l'ARN polymérase II, indépendamment de la distance hr-promoteur. Nous développerons plus en détail ce mécanisme dans la prochaine section. (ii) IE1 est également capable de transactiver directement son propre promoteur, certains promoteurs de gènes précoces (comme 39K, dnapol, gp64) mais aussi des promoteurs cellulaires insectes (comme le promoteur de l'actine A3 de *Bombyx*; Lu *et al*., 1996) et même mammifères

(Murges *et al*., 1997), en l'absence de séquence hr. Cependant, aucun site de fixation pour IE1 n'a été clairement caractérisé dans ces promoteurs (Friesen, 1997), ce qui pourrait suggérer que l'activité transactivatrice de IE1 passe par des interactions protéine-protéine qui restent à déterminer.

IE1 a aussi une activité de répression de la transcription sur quelques promoteurs, comme ie0 et ie2 (Carson *et al*., 1991; Kovacs *et al*., 1991). Il est possible que cette inhibition passe par la fixation de IE1 sur une séquence présentant des homologies avec le palindrome des hrs et présente entre la TATA box et le motif CAGT de ces promoteurs (Leisy *et al*., 1996).

Notons que IE0, dont nous avons décrit précédemment la structure, présente des propriétés similaires mais distinctes de IE1: cette protéine est également capable de transactiver via les hrs, mais dans au moins un cas (le gène 39K) est incapable de stimuler directement le promoteur comme le fait IE1 (Kovacs *et al*., 1991). Les fonctions de IE0 sont mal connus, mais il est possible qu'il joue un rôle important au tout début de l'infection, peut-être en dimérisant avec IE1.

2.2.2.2. réplication

Des études de réplication in vitro de plasmides en présence de cosmides exprimant différents gènes baculoviraux ont montré que IE1 est l'un des 5 gènes essentiels à la réplication du virus (Ahrens et Rohrmann, 1995; Lu et Miller, 1995). Par ailleurs, l'utilisation d'un baculovirus portant une mutation sensible à la température dans le gène ie1 (figure 11A) a montré que IE1 est nécessaire au déroulement normal de la réplication (Choi et Guarino, 1995c; Ribeiro *et al*., 1994). Cependant, son rôle exact n'a pas été élucidé: il est possible qu'il soit impliqué directement dans la formation du complexe de réplication par ses propriétés de fixation sur les hrs, qui sont elles-mêmes des origines de réplication putatives. Toutefois, on ne peut exclure que c'est son activité activatrice des gènes précoces eux-même impliqués dans la réplication qui le rend indispensable (Lu et Miller, 1997).

3. Implication de facteurs cellulaires dans la transactivation médiée par IE1/hrs

Nous avons vu dans la section précédente que IE1 transactive fortement les promoteurs dépendant de l'ARN polymérase II cellulaire. Directement ou indirectement, ce transactivateur viral doit donc interagir avec des facteurs de transcription généraux de l'hôte, mais aucun n'a pour l'instant été identifié. Friesen (1997) a proposé deux modèles impliquant de telles

interactions pour expliquer le mécanisme de la transactivation médiée par les hrs (figures 12A) ou indépendante des hrs (figure 12B).

Dans le premier modèle (transactivation dépendante des hrs), le dimère IE1 fixé sur le palindrome pourrait interagir avec des facteurs du complexe de préinitiation, comme les TAFs, ou des facteurs du complexe d'initiation basal (figure 12A). Alternativement, en association avec d'autres facteurs, il pourrait modifier la structure de la chromatine et rendre le promoteur ciblé plus accessible à la machinerie de transcription de l'hôte. Dans ce modèle, les hrs permettraient de réduire les concentrations de IE1 nécessaires à la transactivation, par le recrutement massif de ce facteur sur les nombreux sites palindromiques et donc par l'augmentation locale de sa concentration près du promoteur. L'activité transactivatrice de IE1 serait ainsi possible dès le début de l'infection, quand la quantité de protéine est encore très basse.

Dans le deuxième modèle (figure 12B), qui correspond à l'activation directe des promoteurs précoces, IE1 pourrait, sans fixation sur l'ADN, interagir avec des facteurs de transcription présents au niveau du promoteur.

3.1. Arguments en faveur de l'implication de facteurs cellulaires

A côté de ces facteurs généraux de la transcription qui interagissent sans doute avec IE1 au niveau du promoteur, toute une série d'expériences et de données très diverses convergent vers l'idée que des facteurs de l'hôte interviennent au niveau des régions homologues. Les points suivants recensent ces différents arguments en faveur d'une implication de facteurs de l'hôte dans la transactivation dépendante de IE1 médiée par les hrs.

3.1.1. Les enhancers de nombreux virus portent des sites de fixation pour des facteurs de l'hôte

La phase qui suit l'entrée du virus dans la cellule-hôte est cruciale: en effet, l'installation de l'infection nécessite l'expression rapide et abondante de protéines virales, alors que la population virale est très faible (Friesen, 1997). Pour résoudre ce problème, la plupart des virus ont développé une stratégie d'"optimisation" de la transcription, qui passe par le recrutement de facteurs de transcription et de cofacteurs de l'hôte sur leurs propres promoteurs et régions activatrices. Par exemple, chez le virus herpes simplex de type I, le gène LAT est impliqué dans le phénomène de latence/réactivation du virus. Son promoteur contient deux sites de fixation pour des facteurs de transcription de la famille CREB/CREM/ATF, CRE1 et CRE2, qui sont reconnus par au moins 4 protéines de l'hôte, contrôlant ainsi en partie la transcription du gène LAT (Millhouse *et al*., 1998).

Dans l'exemple du cytomégalovirus simien (SCMV), la région promotrice/activatrice proximale du gène très précoce IE94 a une structure qui ressemble beaucoup à celle des hrs du baculovirus, puisqu'elle est composée de 11 répétitions palindromiques, chacune portant un site CRE (Chang *et al.*, 1990). Comme les hrs, cette région est capable d'activer de nombreux promoteurs hétérologues placés en *cis*, par le biais de la fixation de facteurs de transcription de l'hôte (de type CREB/CREM/ATF) sur les CREs (pour revue, Sandford et Burns, 1996). Dans le virus, cette région est essentielle puisqu'elle permet aux temps très précoces de l'infection l'activation de la transcription du transactivateur IE94, dont les propriétés sont similaires à celles de IE1 chez les baculovirus.

3.1.2. De nombreux transactivateurs viraux interagissent avec des facteurs de l'hôte

Beaucoup de virus ont développé une stratégie complémentaire à la précédente pour augmenter la transcription de leurs gènes: comme nous l'avons vu dans deux exemples prototypiques (le baculovirus *Ac*MNPV et le cytomégalovirus simien), ils expriment immédiatement après leur entrée dans la cellule des protéines activatrices de la transcription qui vont stimuler les gènes viraux de la phase suivante, mais aussi certains gènes cellulaires dont l'expression est indispensable au bon déroulement de leur cycle (pour revue, (Op De Beeck et Caillet-Fauquet, 1997). Le mécanisme d'action de ces transactivateurs a été bien étudié chez les virus de mammifères, et semble assez homogène: typiquement, le facteur viral ne se fixe pas directement sur l'ADN, mais interagit plutôt avec des facteurs de transcription de l'hôte présents au niveau des régions activatrices ou des promoteurs cibles. Ainsi, les sites TRE présents dans le LTR du visnavirus fixent des facteurs cellulaires de type AP1 dont l'interaction avec le transactivateur précoce TAT est nécessaire à son effet activateur (Morse *et al.*, 1999). De manière similaire, le transactivateur TAX de HTLV1 interagit avec des facteurs cellulaires (CREB, ATF1) fixés sur trois sites CRE du LTR. Cette situation largement répandue chez les virus de mammifères est différente de celle du transactivateur baculoviral IE1, puisqu'il possède son propre site de fixation dans les hrs, et qu'il est capable de s'y fixer seul. Cependant, l'implication systématique de facteurs cellulaires dans les mécanismes de transactivation par le couple transactivateur viral-région activatrice observée chez les virus de mammifères suggère qu'il pourrait en être de même chez les baculovirus.

Par ailleurs, beaucoup de transactivateurs viraux interagissent avec CBP ou P300 (regroupées dans la terminologie CBP/P300; Hottiger et Nabel, 2000), deux protéines coactivatrices cellulaires dont la structure et les fonctions sont très similaires et qui sont impliquées dans de

nombreuses fonctions biologiques comme la différenciation ou la prolifération cellulaire. Au niveau transcriptionnel, CBP/P300 fonctionne comme une protéine intégratrice, interagissant avec une très grande variété de facteurs de transcription, spécifiques et généraux (Vo et Goodman, 2001). Elle constitue à ce titre un facteur clé des complexes transcriptionnels, et son ciblage systématique par les transactivateurs viraux permet d'augmenter ou de bloquer l'activation transcriptionnelle d'un grand nombre de gènes (pour revue, Hottiger et Nabel, 2000).

3.1.3. Les hrs activent la transcription même en l'absence de facteurs viraux

De nombreuses expériences ont montré que différents hrs sont capables de transactiver les promoteurs de gènes viraux précoces comme p35 (Nissen et Friesen, 1989) ou ie2 (Lu et Carstens, 1993) en l'absence de facteurs viraux. Ces travaux suggèrent donc que des facteurs cellulaires activateurs de la transcription se fixent sur les hrs (Lu *et al*., 1997; Kremer et Knebel-Mörsdorf, 1998).

3.1.4. Les hrs semblent fixer des facteurs de l'hôte in vitro

Plusieurs expériences de gel retard ont montré que des facteurs de l'hôte peuvent se fixer sur les hrs (Leisy *et al*., 1995), mais n'ont pas été caractérisés.

3.1.5. La région hr5 complète transactive mieux que la somme de ses palindromes

Des constructions portant soit la région hr5 entière soit ses palindromes les uns à la suite des autres sans régions flanquantes ont été testées en présence de IE1. Que ce soit dans des expériences de transfection ou en conditions d'infection, l'activation du gène rapporteur par la région hr complète a été plus forte que par le multimère de palindromes. Ce résultat suggère donc que les régions flanquant les palindromes contiennent des motifs importants pour la transactivation par IE1 (Rodems et Friesen, 1993).

3.1.6. Le domaine transactivateur de IE1 isolé interfère avec la protéine complète

L'activation transcriptionnelle dépendante de IE1 et médiée par les hrs est inhibée par l'expression simultanée du domaine N-terminal isolé de IE1, qui est pourtant incapable de se fixer sur l'ADN. Ce résultat pourrait s'expliquer par une compétition entre ce domaine transactivateur et la protéine IE1 intacte pour des facteurs de l'hôte nécessaires à la transactivation par IE1 (Kovacs *et al*., 1992).

3.1.7. Un facteur de l'hôte identifié se fixe sur les hrs

Habib et Hasnain (1996) ont montré qu'une protéine cellulaire de 38 kDa, hr1-BP, se fixe spécifiquement au niveau des palindromes portés par hr1. Ce facteur interagit avec le sillon mineur de l'ADN, une caractéristique partagée par les protéines architecturales de type HMG, dont l'une des propriétés est de courber l'ADN, rapprochant physiquement les régions activatrices du promoteur. Hr1-BP n'a cependant pas été complètement caractérisé, et son rôle exact dans la transactivation médiée par hr1 reste incertain.

3.1.8. Les hrs contrôlent en partie la spécificité d'hôte de la transactivation par IE1

La comparaison de l'efficacité transactivatrice du couple IE1/hr dans différentes lignées cellulaires montre des différences importantes. Ainsi, la région hr5 de *Ac*MNPV est plus efficace dans les cellules de *S. frugiperda* que dans celles de *O. pseudotsugata* (Theilmann et Stewart, 1991). Une autre étude en cellules de *L. dispar* a établi que la région hr4 de *Ld*MNPV est plus efficace que la région hr5 de *Ac*MNPV, et inversement en cellules *Sf*9 (Pearson et Rohrmann, 1997). Ces expériences suggèrent donc que la transactivation ne dépend pas seulement de IE1 mais aussi de facteurs spécifiques de l'hôte fixés par les hrs.

Ces différents résultats convergent donc vers l'idée que des sites de fixation pour des facteurs de l'hôte sont présents dans les hrs, et qu'ils sont fonctionnellement importants pour la transactivation médiée par IE1. Les motifs TRE et CRE identifiés dans les hrs (Ogliastro, 1994), qui fixent des facteurs de transcription de type AP1 et CREB/CREM/ATF, pourraient constituer d'excellents candidats, et nous allons à ce titre détailler leurs caractéristiques générales dans la section suivante.

3.2. Les facteurs de transcription AP1 et CREB/CREM/ATF

3.2.1. Structure et caractéristiques

3.2.1.1. la superfamille bZip

Ces facteurs de transcription font partie d'une grande famille de protéines ayant de l'affinité pour l'ADN et possédant tous une structure de type région basique-glissière de leucine (bZip; Struhl, 1989). Comme le montre l'exemple de CREB (figure 13 A), dont le complexe avec l'ADN a été récemment cristallisé (Schumacher *et al.*, 2000), cette structure est extrêmement

simple: la région basique est responsable de la fixation à l'ADN tandis que la glissière de leucine permet la dimérisation de la protéine. La figure illustre également le fait que l'interaction spécifique et stable entre ces facteurs et l'ADN nécessite la formation d'un dimère (Yamamoto *et al*., 1988), les régions basiques de chaque monomère passant à travers le sillon majeur de l'ADN et établissant avec lui des contacts directs et indirects.

3.2.1.2. les membres de la famille et leurs sites de fixation

La superfamille des protéines bZip se subdivise en familles plus restreintes en fonction de la capacité de dimérisation des facteurs et leur affinité pour un motif ADN particulier, dont par exemple les familles des facteurs GCN4, C/EBP, AP1 et CREB/CREM/ATF (pour revue, Meyer et Habener, 1993; Vallejo, 1994). Ce sont ces deux dernières familles qui vont nous intéresser plus particulièrement.

La famille AP1 regroupe des facteurs de transcription qui ont été au départ caractérisés par leur affinité pour le motif palindromique imparfait TGACTCA, appelé TRE (TPA response element; Angel *et al*., 1987; Lee *et al*., 1987b). Les plus connus sont les protéines Fos et Jun (Bohmann *et al*., 1987; Franza *et al*., 1988).

Les membre de la famille CREB/CREM/ATF ont été regroupés par leur capacité à se fixer sur un motif palindromique très semblable à TRE, TGACGTCA, appelé CRE (cAMP response element). Ce motif canonique a été identifié pour la première fois dans le promoteur du gène de la somatostatine humaine par la fixation de CREB (CRE binding protein) (Montminy et Bilezikjian, 1987). Les facteurs ATF (Activating Transcription Factors) ont été caractérisés à peu près simultanément par leur fixation sur le même motif présent dans les promoteurs des gènes adénoviraux E2, E3 et E4 (Lee *et al*., 1987a). Enfin, à côté de ces facteurs activateurs, plusieurs isoformes de CREM (CRE modulator), ont un effet inhibiteur sur la transcription dépendante des motifs CRE (Delmas *et al*., 1994).

En plus des sites TRE et CRE "canoniques" que nous venons de décrire, de nombreux variants ont été observées dans divers promoteurs et régions activatrices (tableau 1). Plusieurs études comparatives *in vitro* et in vivo ont montré que ces variants n'ont pas la même affinité ni la même spécificité pour les facteurs bZip, modulant l'effet transactivateur de ceux-ci (Benbrook et Jones, 1994; Craig *et al*., 2001; Drust *et al*., 1991).

Par ailleurs, comme nous l'avons déjà abordé plusieurs fois, des sites TRE et CRE canoniques ou variants ont été identifiés dans le génome d'un très grand nombre de virus, comme l'adénovirus (Fax *et al*., 2000), le virus herpes simplex de type 1 (Millhouse *et al*., 1998), les

cytomégalovirus simien et humain (Chang *et al*., 1990), HTLV1, le visnavirus (Morse *et al*., 1999), le virus de l'hépatite B (Williams et Andrisani, 1995).

Finalement, des revues récentes indiquent que la distinction entre les familles AP1 et CREB/CREM/ATF est en fait assez floue, et découle pour l'essentiel de l'histoire de la découverte des protéines qui les composent (Hai et Hartman, 2001). Trois arguments principaux appuient cette affirmation: 1) comme nous allons le détailler dans la section suivante, les membres d'une famille peuvent former des hétérodimères avec les membres de l'autre famille, 2) la plupart d'entre eux peuvent se fixer sur le site de fixation de l'autre famille, et 3) certaines protéines présentent des homologies de séquence plus fortes avec un membre de l'autre famille qu'avec les membres de la sienne propre. Ainsi, ATF3 est plus proche de Fos (famille AP1) que des membres de la famille CREB/CREM/ATF) (Hai et Hartman, 2001).

3.2.1.3. combinatoire des facteurs AP1 et CREB/CREM/ATF

Chez les mammifères, au moins 15 facteurs de transcription rattachés à ces deux familles de protéines bZip ont été décrits. Si la formation d'homodimères est possible pour certains d'entre eux (mais pas pour tous, comme fosB chez l'Homme; Ryseck *et al*., 1990), la plupart de ces facteurs ont la possibilité de s'hétérodimériser par le biais de leur glissière de leucine (Herdegen et Leah, 1998). Le tableau 2 recense quelques-uns de ces complexes observés in vitro et in vivo. Plusieurs études ont montré que la formation de ces hétérodimères peut modifier l'affinité des facteurs qui le composent pour le site TRE ou CRE. Pa exemple, si l'homodimère Jun a une forte affinité pour TRE, l'hétérodimère Jun/ATF-2 a beaucoup plus d''affinité pour CRE (Hai et Curran, 1991). Cet exemple illustre en fait la tendance des hétérodimères entre membres de la famille AP1 et CREB/CREM/ATF à avoir une affinité plus forte pour le motif CRE que pour TRE.

Tous ces éléments montrent donc que de très nombreuses combinaisons dimériques sont susceptibles de se former et de se fixer avec des affinités très variables sur les sites TRE et CRE. Ces résultats pourraient ainsi expliquer comment, à partir d'un répertoire assez limité de facteurs, la cellule peut réguler très finement et de façon adaptée au contexte cellulaire la transcription de ses gènes.

3.2.2. Activation de ces facteurs

Nous prendrons l'exemple de CREB chez les mammifères, car c'est le facteur de la superfamille bZip dont le mécanisme d'action a été le plus étudié.

CREB a été caractérisée initialement par sa capacité à se fixer sur l'élément CRE présent dans le promoteur du gène de la somatostatine et à activer la transcription en réponse à une augmentation du taux d'AMP cyclique dans la cellule (Andrisani *et al*., 1987). CREB est donc l'effecteur final de la voie de transduction du signal dépendante de l'AMPc (Montminy, 1997). Cependant, des études ont ensuite montré que CREB est également l'effecteur final d'autres voies de transduction du signal, comme la voie calcique (pour revue, Shaywitz et Greenberg, 1999). Dans tous les cas cependant, l'activation de CREB passe par sa phosphorylation au niveau d'une sérine en position 133 de la séquence primaire, portée par un domaine particulier, appelé KID (figure 14). Cette phosphorylation est effectuée par diverses protéines kinases, par exemple une PKA dans la voie AMPc et une CaMK dans la voie calcique (Shaywitz et Greenberg, 1999). Le principal effet de cette phosphorylation du domaine KID de CREB est de rendre possible son interaction avec le domaine KIX du coactivateur ubiquiste CBP/P300 (figure 15), qui va servir de pont moléculaire entre CREB et la machinerie de transcription basale. Par ailleurs, par son activité histone acétyltransférase (HAT), CBP/P300 va modifier la conformation de la chromatine et rendre plus accessible le promoteur-cible à la machinerie transcriptionnelle (Chan et La Thangue, 2001; Shaywitz et Greenberg, 1999).

A côté du domaine KID, responsable de l'activation de la transcription signal-inductible, d'autres domaines de CREB interagissent constitutivement avec des éléments de la machinerie transcriptionnelle basale, comme Q2 (figure 14) avec TFIID et TFIIB (Montminy, 1997). La figure 16 illustre un exemple d'activation de CREB par la voie AMPc, aboutissant à l'activation du gène de la somatostatine, avec (A) la transduction du signal dans la cellule, et (B) les mécanismes moléculaires au niveau du promoteur.

Cet exemple de CREB permet de donner une idée générale du mode de fonctionnement des facteurs de transcription bZip. Cependant, chacun des facteurs AP1 et CREB/CREM/ATF présente des particularités propres, qui peuvent varier d'un tissu à l'autre et que nous ne détaillerons pas. Mentionnons seulement une distinction apparaissant typiquement dans le système nerveux, entre les facteurs exprimés constitutivement (plutôt CREB/CREM/ATF) et inductibles (plutôt AP1) (Herdegen et Leah, 1998). Les facteurs constitutifs sont présents dans les cellules non stimulées, probablement déjà fixés sur leur élément de réponse dans les promoteurs, et vont être activés par un signal extracellulaire comme nous l'avons illustré pour CREB. Les facteurs inductibles, eux, ne sont exprimés que dans des phases précises de la vie de la cellule. Un exemple classique est celui de c-Fos, dont le promoteur contient deux sites CRE qui sont activables par les voies calcique et AMPc en réponse à un signal extracellulaire (Berkowitz *et al*., 1989; Fisch *et al*., 1989).

Si, chez les mammifères, l'expression (constitutive ou induite) et les modalités d'activation des facteurs AP1 et CREB/CREM/ATF sont assez bien connues, les études sont moins nombreuses chez les insectes. Ainsi, par comparaison avec la quinzaine de facteurs identifiés chez les mammifères, cinq seulement l'ont été chez la Drosophile: dCREB-A (Smolik *et al*., 1992), dCREB-2 (Usui *et al*., 1993), ATF4 (Hewes *et al*., 2000), dFos (Perkins *et al*., 1990) et dJun (Zhang *et al*., 1990). Parmi eux, dCREB-2 a été particulièrement étudiée en raison de son implication dans l'établissement de la mémoire à long terme (voir section suivante). Ces expériences ont montré qu'un de ses isoformes, dCREB2-a, est inductible par la voie AMPc et que cette induction est liée à la phosphorylation de la protéine, probablement sur la sérine 231 (qui serait ainsi l'homologue fonctionnel de la sérine 133 de la CREB mammifère) (Yin *et al*., 1995a; Yin *et al*., 1995b).

3.2.3. Rôles

Les facteurs de transcription des familles AP1 et CREB/CREM/ATF ont une répartition ubiquiste, en relation avec la multiplicité de leurs rôles dans l'activation de la transcription de gènes impliqués dans les grands processus cellulaires, comme la prolifération, la différenciation et les réponses adaptatives aux stimulations de l'environnement. Par exemple, au cours du développement, CREB est indispensable à l'établissement du répertoire complet de lymphocytes T chez la souris (Barton *et al*., 1996). Chez la même espèce, son haploinsuffisance conduit à des malformations comme le nanisme ou des myopathies cardiaques (Fentzke *et al*., 1998; Struthers *et al*., 1991), tandis que chez la drosophile, ATF4 est un facteur important dans le contrôle de la mue et de la métamorphose (Hewes *et al*., 2000). Ces facteurs sont également extrêmement importants dans le développement et la plasticité du système nerveux (Silva *et al*., 1998). Ainsi, une des fonctions les plus documentées de CREB à la fois chez les mammifères et les invertébrés (Aplysie, Drosophile) est son rôle dans la plasticité synaptique et l'établissement de la mémoire à long terme (Silva *et al*., 1998; Yin *et al*., 1995a; Yin *et al*., 1995b). Dans ce système, l'hétérodimère fos/jun apparaît également fondamental, intervenant en amont de CREB et favorisant à la fois le renforcement synaptique et l'apparition de nouvelles synapses (Sanyal *et al*., 2002).

Ces facteurs de transcription jouent également un rôle important pour un très grand nombre de virus, avec deux grands types de stratégies: 1) comme nous l'avons abordé dans les sections 3.1.1. et 3.1.2., certains virus détournent à leur profit les facteurs AP1 et CREB/CREM/ATF cellulaires pour augmenter la synthèse de leurs propres gènes ou pour moduler la synthèse de certains gènes de l'hôte. Cette stratégie implique la fixation de ces protéines de l'hôte dans les

promoteurs ou régions activatrices virales, et /ou leur ciblage par des protéines virales, et 2) certains virus expriment des homologues de ces facteurs, pour stimuler leurs propres promoteurs et/ou contrôler la transcription de l'hôte. Deux exemples classiques concernent Fos et Jun, qui ont été décrits pour la première fois comme les oncogènes de deux rétrovirus induisant des sarcomes respectivement chez le rat (FBJ-MSV;Curran et Teich, 1982) et le poulet (virus du sarcome aviaire 17; Maki *et al*., 1987).

4. Structure et fonction d'un enhanceosome

Les rappels bibliographiques des sections précédentes ont permis d'identifier les différents facteurs viraux et cellulaires, connus ou putatifs, qui interviennent dans la transactivation par les hrs: IE1, IE0 pour le virus, hr1-BP et des facteurs de type AP1 ou CREB/CREM/ATF pour la cellule. Une telle situation, où plusieurs facteurs de transcription différents possèdent des sites de reconnaissance proches les uns des autres sur une séquence activatrice de la transcription, nous ramène finalement à la notion d'enhanceosome, que nous allons brièvement développer maintenant.

Ces dernières années, deux modèles (ceux des enhancers du gène de l'interféron β et du gène TCRα) ont permis de mieux comprendre comment l'organisation de l'enhancer et l'assemblage coopératif des facteurs qui le fixent aboutissent à la formation d'un complexe nucléoprotéique, l'enhanceosome, ainsi que les modalités de la stimulation de la machinerie transcriptionnelle. Dans le modèle de l'interféron β, 4 facteurs de transcription se fixent en dimères (IRF1/IRF1, p50/p65 de NF-κB, ATF/cJun) sur leurs sites respectifs dans l'enhancer. Une protéine architecturale (HMG I) se fixe dans le même temps au niveau du sillon mineur de l'ADN (figure 17A), induisant une courbure de l'ADN qui facilite à la fois la fixation des autres facteurs, leurs interactions et leur rapprochement du complexe de préinitiation (figure 17B) (Kim et Maniatis, 1997; Merika et al., 1998). L'enhanceosome du gène TCR-α présente une structure très similaire, avec des facteurs de transcription (AML-1, Ets-1, et CREB/ATF) et une protéine architecturale (LEF-1) différents (figure 18) (Giese *et al*., 1995 ; Mayall *et al*., 1997).

Dans les deux cas, les protéines activatrices présentes dans l'enhanceosome vont interagir spécifiquement avec certains facteurs généraux de la machinerie transcriptionnelle et avec des coactivateurs comme CBP/P300, augmentant la stabilité du complexe de préinitiation. Réciproquement, ces interactions vont stabiliser l'assemblage des constituants de l'enhanceosome (figure 17B) (pour revue, Carey, 1998). On voit donc que la formation de l'enhanceosome et son interaction avec la machinerie de transcription sont des phénomènes

dynamiques, coopératifs et très spécifiques: chaque élément doit être à la bonne place pour que le "transcriptosome" final (enhanceosome + complexe de préinitiation) puisse s'établir. Par le recrutement plus rapide et la stabilisation de la machinerie transcriptionnelle sur le promoteur, l'enhanceosome va alors permettre d'augmenter la fréquence d'initiation de la transcription et donc l'expression du gène.

III- OBJECTIFS DE L'ÉTUDE

Si l'analyse bibliographique nous a fourni un certain nombre d'arguments indirects permettant d'étayer l'hypothèse de l' enhanceosome pour expliquer l'activation transcriptionnelle médiée par les hrs, la plupart des interactions restent à vérifier expérimentalement. Dans ce vaste programme, nous avons choisi d'étudier plus particulièrement les motifs de reconnaissance pour des facteurs de transcription identifiés dans les hrs. Les objectifs de notre travail sont donc:

1) identifier, recenser et analyser précisément la répartition des motifs de reconnaissance pour des facteurs de transcription dans les hrs du baculovirus *Ac*MNPV.

2) tester si des protéines cellulaires se fixent spécifiquement sur ces motifs.

3) identifier, ou à défaut caractériser ces protéines cellulaires.

4) tester si ces motifs sont fonctionnels par eux-même dans la cellule, en dehors du contexte viral

5) mesurer l'importance de ces sites pour la transactivation dépendante de IE1 et médiée par les hrs.

Pour répondre à ces cinq objectifs, nous avons effectué un certain nombre d'expériences dont les résultats sont présentés et discutés dans l'article suivant. Les résultats ne figurant pas dans cet article ainsi que les expériences en cours seront développés dans la section suivante.

B- RESULTATS

I- RÉSUMÉ DE L'ARTICLE 1

" CRE-like sites present in baculovirus homologous regions (hrs) specifically bind host factors and are required for the IE1-mediated transactivation".

Le baculovirus *Ac*MNPV contient 8 régions homologues (hrs) dispersées dans le génome, qui constituent des régions activatrices puissantes lorsqu'elles sont placées en *cis* de promoteurs (viraux ou hétérologues) dépendants de l'ARN polymérase II. Cette activation est fortement augmentée par la fixation du transactivateur viral très précoce IE1 sur des sites palindromiques portés par les hrs, mais son mécanisme est mal connu. Dans ce travail, nous montrons que les

hrs de *Ac*MNPV mais aussi d'autres baculovirus contiennent dans les régions interpalindromiques un grand nombre de motifs CRE, TRE et Ebox qui fixent des facteurs de transcription ubiquistes des familles bZip et bHLH. Des expériences de gel retard et d'UV-crosslinking indiquent que, à la différence des motifs Ebox, les sites CRE et dans une moindre mesure TRE fixent au moins 6 protéines de l'hôte. En transfection, l'activation de promoteurs par ces sites CRE montre qu'ils sont fonctionnels in vivo. De plus, des expériences de cotransfections et de transfection/infection montrent que les sites CRE non mutés sont requis pour la transactivation dépendante de IE1 dans le contexte de l'infection. Ces résultats apportent des éléments nouveaux à la connaissance du mécanisme de fonctionnement des hrs et de l'activation par IE1, et suggèrent un modèle dans lequel le recrutement de facteurs de transcription de l'hôte est indispensable au bon déroulement des phases précoces de l'infection baculovirale.

II- ARTICLE 1

CRE-like sites present in baculovirus homologous regions (hrs) specifically bind host factors and are required for the IE1-mediated transactivation

Igor Landais[1,$], Rachel Vincent[2,$], Lorenne Robert[1], Lionel Galibert[1], Gérard Devauchelle[1], Martine Duonor Cerutti[1] and Mylene Ogliastro[1,*]

[1] Laboratoire de Pathologie Comparée, UMR 5087, 30380 Saint Christol-les-Alès, France

[2] present address : Laboratoire de Pathologie Comparée, UMR 5087 - CP101, 34 095 Montpellier Cedex 05, France

$ these authors have contributed equally to this work

*Corresponding author. Phone (33) 466 78 37 41. Fax: (33) 466 52 46 99. Email: ogliastr@ensam.inra.fr

Running title: Requirement of CRE sites for the hr enhancer function

SUMMARY

The genome of the Autographa californica Multinucleocapsid Polyhedrosis virus (*Ac*MNPV) contains eight interspersed homologous regions (hrs) that function as potent enhancer sequences when *cis*-linked to both viral and heterologous RNA polymerase II-dependent promoters. This enhancement is strongly increased through the binding of the major immediate-early viral transregulator IE1 on 28-mer palindromic sites present in hrs, but the mechanism of this transactivation is unknown. In this work, we show that hrs of *Ac*MNPV and other baculoviruses carry in interpalindromic sequences a large number of CRE, TRE and E-box motifs, known to bind ubiquitous transcription factors of the bZip and bHLH families. EMSA and UV-crosslinking experiments indicate that, unlike E-box, CRE and at lesser extent TRE motifs bind at least 6 insect host factors. In transfection assays, activation of promoters *cis*-linked to CRE sites show that these motifs are functional *in vivo*. Furthermore, cotransfection and transfection/infection assays using constructs carrying both CRE sites and a 28-mer palindrome demonstrate that intact CRE sites are required for the IE1-dependent transactivation in the context of viral infection. Taken together, these data provide new informations on the mechanism of hrs and IE1-dependent activating functions and suggest a model in which the recruitment of host transcription factors contributes to the progression of the early phases of the infection.

INTRODUCTION

A common strategy of numerous animal DNA and RNA viruses consists in expressing potent transregulators that stimulate viral gene expression during the early times of infection, when viral DNA levels are the lowest. The best studied of these transregulators belong to mammalian viruses, for instance Adenovirus (1), HTLV-1[1] (2) and cytomegalovirus (3). Among invertebrate viruses, the *Ac*MNPV baculovirus uses the same strategy, expressing at least four immediate-early transregulators, IE0, IE1, IE2 and PE38 (4).

Baculoviruses are double-stranded, circular DNA viruses which gene expression is regulated in a cascade fashion, with immediate-early (IE), early (E) and late (L) genes successively (5). IE and E genes are transcribed under the control of the host RNA polymerase II (6) and many of them display a promoter structure that mimics RNA polymerase II-dependant promoters of higher eukaryotes (4). By contrast, late promoters are triggered by a virally-encoded RNA polymerase (7).

The 67-kDa IE1 protein is the most potent and the best studied IE transregulator. In transient assays, IE1 has been shown to transactivate all the early baculoviral genes studied so far by two distinct mechanisms : i) by direct stimulation of target promoters without clearly established DNA recognition sequence (4), and ii) through binding on specific recognition sequences located into potent enhancers of early genes, the hrs, leading to a strong transcriptional activation of these genes (8,9). *Ac*MNPV contains eight hrs ranging in size from 30 to 800 bp, interspersed throughout the circular genome. These well-conserved regions are characterized by repeated 28-mer imperfect palindromes, that constitute the IE1 binding sites, flanked by inter-palindromic regions containing direct and inverted repeats (10). Like IE1, such sequences have been observed in all the baculoviruses studied so far, highlighting the crucial importance of the IE1/hr couple in the biology of this virus family.

The modalities of IE1 binding on hr sequences have been extensively studied : *in vitro* studies have shown that monomeric IE1 is able to bind a half-palindrome in the absence of other factors (11) and further works have established that only the dimeric form of IE1 confers enhancer activity to hrs (12). However, several evidences strongly suggest that additional factors, not yet characterized, are involved in the IE1/hr transactivating system: i) in transient assays, different hrs have been shown to transactivate *cis*-linked promoters of viral genes like p35 (8), 39K (13), ie2 (9), p143 (14) in the absence of IE1 or other viral factors, meaning that host factors must interact with hrs (15,16), ii) enhancement by full-length hr5 is greater than the sum of its 28-mer in transient assays but also in the infection context (17), suggesting that flanking sequences increase enhancer activity perhaps through the binding of host factors, iii) the N-terminal trans-activating domain of IE1 is capable of dominantly interfere with wild-type IE1 for the hr-dependent enhancement, suggesting that this domain may interact with host factors important for the hr-

mediated enhancer activity (18), and iiii) a 38 kDa host protein, hr1-BP, has been shown to bind the core 28-mer palindrome with short, shuffled flanking sequences (19). However, the exact importance of this host factor is not clear, since neither its role in the enhancer function of hrs nor its potential interaction with IE1 have been well established.

Beside these results in the baculovirus system, some mammalian viruses display enhancer regions which structure and function resemble that of hrs of baculovirus : the 5' proximal enhancer domain of the Simian CMV IE94 promoter displays 11 palindromic repeats each carrying a CRE site (20). Like hrs, this region strongly activates heterologous genes in a wide variety of cell types and is functionally important for the virus, through the binding of host CREB and other transcription factors (for review, 21). Moreover, early transregulators expressed by mammalian viruses systematically interact with ubiquitous transcription factors and coregulators of the host, like CREB (22,23) and CBP (for review, 24).

All these data collectively suggest that host factors are involved in the hr-dependent transactivation of early viral genes by IE1. In this paper, we show for the first time that interpalindromic sequences of hrs contain a large number of conserved CRE motifs that specifically bind insect host factors. These sites are capable of enhancing a *cis*-linked reporter gene in the absence of viral factors and we further demonstrate that they are required for the IE1transactivation in the context of viral infection.

EXPERIMENTAL PROCEDURES

Plasmid constructions. (i) IE1 expression vectors. The pIE1 plasmid was constructed by cloning in pUC 18 the *Cla*I-*Xba*I fragment from the *Hind*III-G fragment of *Ac*MNPV, as previously described (25). pIE1i425 was constructed with the QuickChange mutagenesis kit using the pIE1 plasmid as template: 12 nt (GAGGACCTTCCA) were inserted at position 1275 of the ie1 coding sequence, leading to a 4 aa-insertion (EDLP) at position 425 of the IE1 protein (26).

(ii) CRE/EB reporter plasmids. The *Bombyx mori* actin promoter (prA_3) that has been shown to be responsive in *Sf*9 cells (15) was ligated into the pBLCAT2 expression vector (27) at the *Hind*III-*Bgl*II sites generating the prA_3CAT plasmid. Double stranded oligonucleotides containing 1 to 4 copies of the CRE/EB site (GAGA**TGATGTCA**TTTGTTTT) were cloned upstream of the prA_3 at the *Hind*III site, generating the 1xCRE-CAT, 2xCRE-CAT and 3xCRE-CAT plasmids (schemes on figure 4).

(iii) hr5, hr5/4 and hr5/4m reporter plasmids. The hr5 region was ex*cis*ed from the *Ac*MNPV genome by *Mlu*I digestion and cloned in pUC 19, generating the phr5 plasmid. phr5 was digested by *Hind*III and *Sac*I, blunt-ended and the resulting hr5 fragment was cloned at the *Hind*III site of prA_3CAT, generating hr5-CAT.

Oligonucleotides corresponding to both strands of the *Ac*MNPV hr5/4 restriction fragment (boxed in figure 1: *TTCTACTTGTAAAG*CACAATCAAAAAGA**TGATGTCA**TTTGTTTTTCAAAACTGAACTCG*CTTTACGAGTA GAA* ; palindromic hemi-sites are in italic and CRE-like motif in bold characters) were annealed, generating cohesive *Eco*RI ends. Double stranded oligonucleotides were self-ligated, blunt-ended and cloned at the *Hind*III site of prA_3CAT, generating the 1xhr5/4-CAT (no self-ligation) and 2xhr5/CAT (one self-ligation). Similar procedure was used to generate 1xhr5/4m-CAT and 2xhr5/4m-CAT, for which the CRE sites TGATGTCA were mutated to TGAT**CATG** (scheme on figure 4).

(iiii) pal reporter plasmids. A double-stranded oligonucleotide corresponding to the 28-mer palindrome (AGCTTCG*CTTTACGAGTAGAATTCTACTTGTAAAG*CA) was cloned at the HindIII site of prA_3CAT, immediately upstream of the prA_3, generating pal1-CAT. In pal2-CAT, the palindrome was placed 60 pb upstream of the prA_3, with a 60-mer junk sequence between the 28-mer and the prA_3 (AGCTTCG*CTTTACG AGTAGAATTCTACTTGTAAAG*CAGGCCACGAGACAGCTTACAGCACATCGGACCGGTGGCAGTACACAA GACCAACGAATCA). All the constructs described above were verified by sequencing.

Cells, transfections, infections and CAT assays - *Spodoptera frugiperda Sf*9 cells (28) (3.10^6 cells/25 cm^2 plate) were transfected using a total of 20 µg of indicated plasmids and 40 µl DOTAP (Roche) following the recommendations of the manufacturer. In transfection/infection assays, cells were infected 24h post-transfection with the C6 strain of *Ac*MNPVat a MOI of 10 as previously described (29). Cells were collected 48 h posttransfection and lysed in liquid nitrogen by 3 rounds of freeze and thaw in 150 µl of Tris-HCl 0.25M, pH 7.5. After centrifugation, total protein concentration was estimated by Bradford assay using BSA as a standard. An aliquot was used for western blot analysis (see below), while the remaining supernatant was further heated 20 minutes at 65°C, insoluble aggregates being precipitated by centrifugation. CAT activity was measured as previously described (30) using the same amount of total protein for each sample (see legend of figures 4 and 6 for quantities used).

Western blot analysis - Protein samples were electrophoresed on a 10% SDS-PAGE and transferred onto a nitrocellulose membrane (Optitran BA-S 85, Schleicher and Schuell). Membrane was revealed as previously described using an IE1 polyclonal antibody (25).

Preparation of *Sf*9 nuclear extracts- Cells were harvested in log-phase growth and nuclear extracts were prepared as described by Dignam (31). Extraction buffer was 0.4 M KCl final. Nuclear extracts were dialyzed 2H at 4°C against buffer D (20 mM HEPES pH 7.9, 20% glycerol, 0.1 M KCl, 0.2 mM EDTA, 1X antiprotease cocktail (Complete, Roche), 0.2 mM DTT) and fractionated on heparin column (HiTrap heparin HP column, Amersham Bioscience) following the recommendations of the manufacturer: briefly, dialyzed nuclear extracts were applied to a pre-equilibrated 1ml-heparin column, washed with 5 column volumes of start buffer (0.01M sodium phosphate, pH 7.0) and eluted using a total of 2 ml of elution buffer (start buffer with increasing concentration of NaCl, from 0 to 2M). Each fraction (200 µl) was dialyzed against buffer D.

Partial purification of IE1- Nuclear extracts were prepared from *Ac*MNPV-infected *Sf*9 cells ($2.4x10^8$ cells) collected 6 h post-infection, and were loaded onto a cation exchange column (HiTrap SP HP column, Amersham Bioscience). The procedure was similar to the heparin fractionation (see above) except the composition of the start buffer (50 mM Sodium acetate pH 5.6) and the elution buffer (start buffer with with increasing concentration of NaCl, from 0 to 1M). IE1-enriched fractions were detected by Western blot, dialysed and subjected to a second round of purification on heparin column, as decribed above. The fraction displaying the strongest quantity of IE1 (estimated by Western blot) was conserved.

Electrophoretic Mobility Shift Assay (EMSA)- (i) CRE/EB probe. 20 µg of each *Sf*9 heparin-fractionated nuclear extracts were incubated with 10 ng of double-stranded α-32P ATP-labelled CRE/EB oligonucleotides (sequence in figure 2) in 20 µl (final volume) of binding buffer (5 mM Tris-HCl pH 8.0, 0.5 mM dithiothreitol, 0.5 mM EDTA, 1% Ficoll, 25 mM NaCl, 0.1 µg/µl poly[dI-dC], Sigma). Reaction mixture was loaded on a 4% nondenaturing polyacrylamide gel and ran at 150 V in 0.5X TBE buffer. Gels were exposed to a Kodak Biomax MS film overnight at -80°C. All EMSAs were performed using the nuclear extract fraction that displayed the strongest CRE-binding activity (fraction 7), as determined in a preliminary EMSA (not shown). When indicated, double-stranded competitor oligonucleotides were added to the binding mixture.

(ii) Pal-CRE probe. Before use, double-stranded hr5/4 derived pal-CRE probe (AATCAAAAAGA**TGATGTCA**TTTGTTTTTCAAAACTGAACTCTCGG*CTTTACGAGTAGAATTCTACGTGT AAAA*CACAATCAAGA) was purified on a 4% polyacrylamide gel. Each binding reaction was performed using 2 µl of fraction 7 or/and 2 µl of partially purified IE1 in the same conditions than for CRE/EB EMSAs. The 61-bp pal competitor carries the leftmost 28-mer palindrome of hr5 as previously described (12).

UV-crosslinking- After 20 min incubation, indicated EMSA mixes were UV-irradiated at 254 nm for 30 minutes as previously described (32). Crosslinked protein-DNA complexes were resolved on a 10% denaturing SDS-PAGE. Gels were autoradiographed overnight at –80°C.

RESULTS

hr sequences contain a large number of binding sites for ubiquitous transcription factors

Baculovirus hrs have been shown to display a potent enhancer function even in the non-infected context, suggesting that binding of host factors on these sequences are involved in the enhancer mechanism. However, exhaustive analysis of hr sequences has never been done. We therefore performed a careful search for transcription factors binding motifs in *Ac*MNPV hrs. Analysis of the 8 hrs (hr1, hr1a, hr2, hr3, hr4a, hr4b, hr4c, hr5) revealed that they contain on both strands a large number of conserved motifs described in a wide variety of viral, prokaryotic and eukaryotic enhancer and promoter sequences. Figure 1 shows the localization of E-box (CANNTG), TRE (TGACTCA) and CRE-like (TGATGTCA) elements, that bind transcription factors of the bHLH (E-box) and bZip (TRE, CRE) families. TRE motifs have been shown to bind bZip factors of the AP1 family (like Fos and Jun), whereas CRE sites bind factors of the CREB/ATF family. Further analysis showed that CRE sites found in hrs differ from the canonical CRE of the somatostatin promoter (TGACGTCA) (33) by one nucleotide. This CRE variant has been found to naturally occur and to be functional in a wide variety of promoters, like in the promoter of the bovine and murine glycoprotein hormone α gene (34). Further analysis of the relative position of the E-box, TRE and CRE motifs in each hr showed that all but one E-box motifs overlap a TRE or a CRE-like site (figure 1). Such juxtaposition has been detected in proximal promoter region of numerous genes (35).

Interestingly, 50 of the 51 (98%) CRE motifs present in the *Ac*MNPV genome are concentrated within the hrs. This feature is not restricted to *Ac*MNPV, since we found in *Bm*NPV and *Ld*NPV genomes 64 and 43 identical CRE-like sites, 92% and 93% being concentrated in hrs, respectively. Two other baculovirus, *Op*NPV and *Px*GV, display the same repartition of putative TRE or CRE sites: 69% of the 35 *Op*NPV TRE-like motif (TGACGCA) and 100% of

the 92 *Px*GV less conserved CRE-like motifs (TGAGAGGG and TGACCAAA) localize in hrs. Thus, concentration of TRE/CRE motifs in hrs of at least 5 baculoviruses suggests that these sites may be important for the hr fonction.

Cellular proteins specifically bind hr CRE-like motifs *in vitro*

To test whether host factors bind these motifs, we performed EMSAs using a double-stranded oligonucleotide probe carrying the typical CRE/E-box motif of hr5/4, a prototypical hr fragment (boxed in figure 1). This probe was incubated with nuclear extracts from the *Sf*9 cell line and diverse competitors to assess the specificity of the binding. As shown on Figure 2A, two major shifted complexes are detected (lane 2) that are well competed by oligonucleotides carrying the intact CRE-like motif overlapped with E-box (lanes 3-4). Competitors carrying a truncated E-box motif competed as well (lanes 5-6), suggesting that host factors involved in both shifted complexes are not E-box specific. Further competitive assays using CRE motifs mutated in the core sequence failed to compete these complexes (lanes 7-12), confirming that host factors specifically bind the CRE-like motif. Figure 2B shows that limited deletions or mutations in the 5'-flanking sequence of the CRE site have no effect on the binding (lanes 2-7). However, optimal binding on the core sequence requires at least a 2 bp-flanking sequence (lanes 8-11). Lanes 12-13 further show that the specificity of the binding is on the CRE and not on the E-box site, since the E-box competitor by itself is unable to displace the shifted complexes.

Taken together, these experiments indicate that insect host factors specifically bind the CRE but not the E-box motifs carried by hrs.

CRE-like binding complexes comprise at least 6 host proteins ranging in size from 28 to 170 kDa

To determine the number and the molecular weight of host proteins that bind the CRE-like motif, we realized UV-crosslinking experiments using the same binding conditions than for the gel shift. After 30 minutes irradiation, four distinct labeled species were generated, migrating approximatively at 28, 50, 80 and 170 kDa (figure 3A). Shorter exposure of the region around 50 kDa showed that 3 proteins migrate between 40 and 55 kDa (figure 3B). In competition assay, the CRE oligonucleotide strongly displaced the 28-, 40-, 83- and 170-kDa products but was less efficient for the 45 and 55 kDa proteins (lane 2, figures 3 A and 3B). Competitors bearing mutations in the CRE motif failed to compete all the products, except the 83-kDa one (lanes 3-5 of figure 3A, lane 3 of figure 3B). Altogether, these results suggest that three insect proteins of 28-, 40- and 175-kDa are very specific to the CRE motif.

CRE-like sites are functional *in vivo*

The next purpose of this study was to test the functionality of the CRE-like motifs in vivo. We used the interesting property of many transcription factor binding sites that can function as enhancer elements when multimerized and placed upstream of a test promoter. These synthetic enhancers often mimic the activities of the natural enhancers from which they are derived (36). We thus cloned one to four copies of the CRE motif upstream of the CAT reporter gene driven by the A_3 promoter, generating the 1x, 2x, 3x and 4xCRE-CAT. In order to further investigate the activity of the CRE sites in the hr context, we cloned one or two copies of the prototypical hr5/4 fragment (boxed in figure 1) in the same prA_3CAT construct, leading to the 1xhr5/4-CAT and 2xhr5/4-CAT plasmids. The hr5/4 fragment contains one CRE site overlapped with an E-box motif, and a half palindrome at each end. Its dimerization leads to the reconstitution of a complete 28-mer palindrome flanked by two CRE/E-box sites and ended by two half-palindromes (shemes in figure 4). As a positive control, we cloned the complete hr5 sequence in prA_3CAT. We tested these constructs in transient assays and CAT expression was measured 48h post-transfection (figure 4). Lanes 1-5 show that the CAT activity increases with the number of CRE copies in an additive manner, demonstrating that CRE-like sites are functional by themselves in the *Sf*9 cell. CAT activity also increases with the number of hr5/4 fragments (lanes 6-7), leading to activation levels similar to constructs harboring only CRE sites (compare lanes 2 and 6; 3 and 7). Interestingly, the hr5 construct displays a CAT activity corresponding approximately to what is expected from the additive effect of its 9 CRE/TRE modules (lane 8), suggesting that the enhancer activity of hr5 results from the additivity of its modules.

These results collectively demonstrate that the CRE-like sites present in hrs are functional activator modules in *Sf*9 cells. They also suggest that the enhancer effect of hrs in non-infected cells is at least partially triggered by its CRE sites.

IE1 and host CRE-binding proteins bind together a hr5/4-derived sequence *in vitro*

Analysis of the hrs structure revealed that CRE sites are systematically located close to the 28-mer palindromic motifs that bind the viral transactivator IEI. We then wished to test whether host factors and IE1 could bind together a hr sequence. We performed EMSAs using the hr5/4-derived pal-CRE probe, that harbors both a 28-mer IE1 binding site and a CRE/EB site. This probe was incubated either with host nuclear extracts (figure 5, lanes 1-4), partially purified IE1 (lanes 5-7) or with a mix of both (lanes 8-12). Lanes 1-4 show that host factors specifically bind to the probe, since the shifted complex (noted I on the figure) is competed by the CRE and the pal-CRE

competitors (lane 2 and 4) but not by the mutated CRE one (lane 3). Lanes 5-7 indicate that the strong binding of IE1 to the probe (lane 5, complex II) is well competed by the pal-CRE competitor (lane 7) but only partially by pal, a shorter competitor corresponding to the leftmost palindrome of hr5 (lane 6). Lane 8 shows that a new, higher complex (noted III on the figure) appears when both host nuclear extracts and purified IE1 are incubated with the probe. This higher complex is partially competed by the CRE competitor (lane 9) but not by the mutated one (lane 10). Lane 11 indicates that pal strongly competes the higher complex and partially the IE1 one, while pal-CRE competes well both shifts.

Taken together, these results suggest that host factors and IE1 simultaneously bind their respective recognition site on the hr5/4 fragment *in vitro*.

Intact CRE-like sites are required for the hr-mediated transactivation by IE1 in the viral context

We next wished to test whether CRE-like motifs are important for the IE1-dependent transactivation. As in the precedent transfection experiment, we used the 2xhr5/4-CAT construct, that contains a 28-mer palindrome flanked by two CRE/E-box sites. We also tested 2xhr5/4m-CAT, an almost identical construct carrying mutated CRE sites previously shown to be unable to bind host factors *in vitro* (figure 2A, lanes 9-10). As controls, we used the pal1-CAT and pal2-CAT constructs, in which the 28-mer palindrome was cloned either immediately upstream of the prA_3 (pal1-CAT) or at the same position than the palindrome of 2xhr5/4-CAT (pal2-CAT, schemes in figure 4). In this latter construct, we replaced the flanking CRE-containing sequence by an irrelevant one. All these constructs were tested in transfection assays for their responsiveness to IE1.

We first performed cotransfection experiments, in which IE1 was expressed out of the viral context from an IE1 expression vector, pIE1. As a negative IE1 control, we used pIE1i425, a construct that encodes a mutated form of IE1 harboring a 4 aa insertion at position 425, causing loss of dimeric IE1 binding to DNA (26). We also verified by western blot that IE1 is expressed at equivalent level in each assay. Figure 6 shows that constructs carrying only the palindrome (lanes 5, 8) display a CAT activity higher than constructs containing the 2xhr5/4 or 2xhr5/4m fragments (lanes 12, 16). This result indicates that the palindrome alone is able to mediate the IE1 transactivation, and suggests a negative effect of the hr5/4 sequences flanking the palindrome. When the CAT construct lacks the palindrome (lane 2) or when mutated IE1 (IE1i425) is coexpressed (lanes 11, 15), only a basal CAT activity is detected, confirming that IE1 is responsible for the observed transactivation. Moreover, transactivation levels appear to be only weakly affected by the position of the *cis*-linked 28-mer relative to the prA_3 (compare lanes 5 and 8), as expected for an enhancer element.

These results show that, in non-infected cells, IE1 is able to transactivate with maximal efficiency the reporter gene *cis*-linked to a palindrome in the absence of any hr flanking sequence. However, the reduced CAT expression observed from the 2xhr5/4m compared to the 2xhr5/4 construct (compare lanes 12, 16) suggests that CRE sites may play a role in the native hr context.

To investigate the role of CRE sites in the infection context, we conducted transfection/infection assays: *Sf*9 cells were first transfected using CAT reporter constructs, and infected 24-h post-transfection with *Ac*MNPV, the virus driving the IE1 expression. Western blot indicated that IE1 is expressed at higher levels in infected cells than in transfected ones (figure 6).

Results of this transfection-infection assay strongly differed from the cotransfection assays, with a dramatic reduction of CAT activity for the constructs containing the 28-mer palindrome without hr flanking sequences (compare lanes 5 and 6, 8 and 9). By contrast, the 2xhr5/4-CAT construct, that carries both the 28-mer palindrome and native hr5/4 flanking sequences, enhanced CAT expression at levels similar to cotransfection assays (lanes 12, 13). These results suggest that flanking sequences are important for the IE1-dependent transactivation in the context of viral infection. Moreover, 2xhr5/4m-CAT, that bears the same flanking sequences than 2xhr5/4-CAT except mutations in the CRE-like sites, failed to transactivate the CAT expression (compare lanes 13 and 17, 16 and 17). This indicates that interpalindromic CRE-like sites are required for the IE1 transactivation in infected cells.

In summary, these data reveal that CRE-like sites located in sequences flanking the 28-mer IE1-binding sites are required for the IE1-dependent transactivation mediated by hrs, principally in the infection context.

DISCUSSION

The IE1-dependent transactivation mediated by hrs has been the subject of many studies (reviewed in 37), but the precise mechanism remains to be elucidated: it has been hypothesized that IE1, through its dimerization, may stabilize the transcription initiation complex (12,38), but the involvement of host proteins in this process has been poorly studied. In this paper, we reveal the importance of new hr motifs that specifically bind insect host factors *in vitro* and are crucial for the IE1-dependent transactivation *in vivo*.

CRE binding properties- Among the 3 transcription factor recognition sites we identified in hrs, a strong and specific binding on the CRE motif was detected using insect nuclear extracts. CRE-binding factors exhibited a lower affinity for the TRE motif (data not shown) while no binding was detected on the E-box motif. We thus decided to focus our analysis on the CRE motif.

UV-crosslinking experiments indicated that at least 3 different *S. frugiperda* proteins (the 28, 40 and 170 kDa labeled products) specifically bind the CRE motif. Size comparison with the five bZip factors described in Drosophila, including dCREB-A (39), dCREB-2 (40), cryptocephal (crc, homolog to mammalian ATF4) (41), dFos (42) and dJun (43) indicates that 2 lepidoptera proteins could correspond to Drosophila factors (*i.e.* the 28 and 40 kDa products). The remaining 170 kDa protein displays a size unusual for bZip factors, but it may be related to two CRE-binding proteins of 120 kDa (44) and 180 kDa (45) detected in mammalian extracts. Interestingly, the 180 kDa protein was identified using a CRE probe belonging to the LTR of HTLV-1 and shown to be indispensable to the transactivation by Tax (45).

Several works have shown that members of the CREB/ATF family form selective cross-family heterodimers with members of the AP1 family (46). This heterodimer formation often leads to altered DNA binding specificity and transcriptional activities, and is thought to increase the combinatorial possibilities of regulation by CREB/ATF and AP1 factors through CRE and TRE sites (for review, 47). From our results, it is likely that heterodimeric rather than homodimeric complexes bind CRE sites, since six proteins highly variable in size are detected, whereas only two complexes close to each other appear in EMSAs. Moreover, these CRE-specific complexes retain affinity for the TRE motif (data not shown), a feature generally shared by bZip cross-family heterodimeric complexes (46). It remains to determine which heterodimers actually form and which of them are functional *in vitro* and *in vivo*.

CRE-dependent transcriptional activation in uninfected cells- The fact that multimerization of CRE motifs led to a proportional increase of CAT activity indicates that these elements are functional *in vivo*, probably through their binding by host activating factors. For the hr5 complete sequence, we observed a CAT activity stronger than expected from the additive effect of its CRE and TRE modules, suggesting that they could act in a cooperative way. Alternatively, we cannot exclude that other hr motifs are involved, like the C/EBP binding motifs previously identified in hr1 (48). Interestingly, as a member of the bZip family, C/EBP has been shown to heterodimerize with CREB/ATF factors (47), increasing the combinatorial repertoire of potential heterodimers that may bind hr sequences.

Expression of IE1 in uninfected cells led to a strong transactivation of reporter constructs harboring the IE1 palindrome alone. By comparison, constructs carrying the palindrome in the native hr context displayed a moderately reduced CAT activity, suggesting that flanking sequences may have a negative effect on the IE1-dependent transactivation. The construct containing flanking sequences with intact CRE sites seemed to be less affected than the one containing mutated CREs, suggesting that CRE motifs partially counter this negative effect. Interestingly, such a negative cooperativity has been previously detected in the cholecystokinin promoter (35),

between juxtaposed CRE and E-box sites. Further experiments are needed to investigate this putative negative E-box effect in hrs.

Importance of CRE sites in the viral context- Testing the same constructs in the context of viral infection, we observed different results compared to the non-infected conditions (*i.e.* in transfection assays), as illustrated by the general decrease of the IE1-dependent transactivation. Interestingly, such contrasted results have been previously reported in several baculovirus studies: Rodems *et al.* (17) and others (16,49) showed an important reduction of the IE1-dependent stimulation mediated by hrs in viral infections compared to plasmid transfection. In this "repressive" viral context, we demonstrate that the presence of intact CRE sites flanking the IE1-binding palindrome is essential for the IE1-dependent transactivation. This result is highly consistent with several works performed in the human cytomegalovirus model, indicating that sequence elements not necessary for gene activation in transient assays can play critical role in regulating gene expression in the context of the viral genome (50,51).

In addition to the requirement of intact CRE sites *in vivo*, we show that these sites specifically bind host factors *in vitro*. Taken together, these results strongly suggest that, in the context of infection, transactivation by IE1 requires the binding of host factors on intact flanking CRE sites. This hypothesis is further strenghtened by EMSAs suggesting the simultaneous binding of IE1 and host factors on a hr fragment.

To explain the differences observed between transfection and infection assays, several authors (16,17) suggested that lower levels of IE1 in infected cells might account for the observed discrepancy. However, our experiments did not support this hypothesis, since we detected about 30-fold more IE1 protein in infected versus transfected cells. However, since (i) the *Ac*MNPV genome carries more than 35 IE1 binding sites in hrs and in addition probably recruits IE1 in a large number of early genes promoters, (ii) each infected *Sf*9 cell approximately contained 10 *Ac*MNPV copies (we infected with a M.O.I of 10), it is likely that, despites its 30-fold excess, almost all the IE1 molecules are recruited by viral DNA in infected cells. Consequently, remaining free IE1 may not be sufficient to transactivate by itself the reporter construct, a situation that is probably similar to the immediate-early phase of the infection, when IE1 concentrations are the lowest (4,52). In these conditions, we hypothesize that binding of transcription factors on CRE/TRE sites close to the 28-mer palindrome could potentiate the binding and/or the transactivating effect of IE1.

As an alternative hypothesis, we suggest that the CRE-dependent transactivation may not be mediated by IE1 but by a factor present in infected but not in transfected cells. IE0, a spliced form of IE1, may be a good candidate, since (i) it displays an additional N-terminal domain (53) which may be important to its transactivation function (54), *e.g.* by interacting with host factors, (ii) it peaks earlier in the infection (55), at times when viral products are

the lowest and consequently when recruiting host transcription factors could further potentiate the transactivator (52), and (iii) unlike IE1, IE0 requires *cis*-linkage of an hr sequence to transactivate the 39K promoter (55), consistently with the concentration of CRE/TRE sites in hrs. Future experiments will help to delineate the respective role of IE1 and IE0 in hr transactivation.

Additional putative roles of CRE sites- Several works pointed out that hrs confer a certain host-specificity to the IE1-dependent transactivation. For example, *Ac*MNPV hr5 is more efficient for the *Ac*MNPV IE1-dependent transactivation in *Sf*9 cells than in *Op* cells (56), similar results being observed in Ld cells (54). These data suggest that binding of species-specific host factors on hr sequences may control for a part the level of IE1-dependent transactivation, that is highly consistent with our results. It is then likely that optimal IE1-dependent transactivation triggered by the precise interplay between host-specific CRE-binding factors and IE1 on hrs is crucial for a successful infection.

Multiple CRE/TRE sites concentred in hrs could also be a powerful viral strategy to control host cell transcription: massive titration of ubiquitous and essential host transcription factors from the CREB/ATF and AP1 families by the virus genome could contribute to the transcriptional shutt-off observed in host cells during infection. Finally, hrs and IE1 are essential elements of the viral replication (57). This origin of replication function of hrs in association with IE1 is far to be fully understood but, as suggested by Leisy *et al*. (13) and by analogy with other DNA viruses, host factors are probably involved in this process as well. It would therefore be interesting to investigate the implication of the CRE/TRE sites in the replicative function of hrs.

hr activating complex- Three proteins have been previously shown to specifically interact with hr sequences: the potent viral transactivators IE0 and IE1 bind the 28-mer palindromic motif, while the host factor hr1-BP binds the core palindrome with short flanking sequences (19). In this work, we demonstrate that novel host proteins bind with high specificity the CRE motif present in hrs, and that this motif is required for the IE1-dependent transactivation. Furthermore, *in vitro* studies suggest that host factors and IE1 bind simultaneously on hr sequence. Figure 7 summarizes all these interactions and proposes an enhanceosome model (58) for the IE1 activating complex that may form on a prototypical hr region. In this model, the coactivator CBP connects the enhanceosome and the preinitiation complex, as supported by our results suggesting that CBP interacts with IE1 *in vitro* (data not shown). Such interaction is not surprising, since CBP has been shown to be targeted by a large number of viral transregulators in mammalian viruses (reviewed in 59).

In conclusion, we demonstrate in this paper that the IE1-dependent transactivation requires structurally and functionally important CRE motifs present within hrs. Several host factors specifically bind these viral motifs *in*

vitro, but they remain to be identified. Future studies will help to delineate the precise interplay with other proteins of the hr-binding complex, like IE1 and CBP, and will provide a better understanding of the baculovirus-host interactions during the crucial immediate-early phase of the infection.

AKNOWLEDGEMENTS

This work was supported by the Institut National de la Recherche Agronomique (INRA), the Centre National de la Recherche Scientifique (CNRS) and the Ministère de la Recherche et de la Technologie (I. L. grant). We particularly thank V. Ollendorf for his very constructive remarks, M. Lopez-Ferber for helpfull discussions, and P. Fournier and N. Volkoff for their critical reading of the manuscript.

BIBLIOGRAPHY

1. Jones, N. (1995) *Curr Top Microbiol Immunol* **199**, 59-80.
2. Armstrong, A. P., Franklin, A. A., Uittenbogaard, M. N., Giebler, H. A., and Nyborg, J. K. (1993) *Proc Natl Acad Sci U S A* **90**, 7303-7307.
3. Klucher, K. M., Sommer, M., Kadonaga, J. T., and Spector, D. H. (1993) *Mol Cell Biol* **13**, 1238-1250.
4. Friesen, P. D. (1997) in *The Baculoviruses* (Miller, L. K., ed), pp. 141-170, Plenum Press, New York
5. Blissard, G. W., and Rohrmann, G. F. (1990) *Annu Rev Entomol* **35**, 127-155
6. Huh, N. E., and Weaver, R. F. (1990) *J Gen Virol* **71**, 195-201.
7. Blissard, G. W. (1996) *Cytotechnology* **20**, 73-93
8. Nissen, M. S., and Friesen, P. D. (1989) *J Virol* **63**, 493-503.
9. Carson, D. D., Summers, M. D., and Guarino, L. A. (1991) *J Virol* **65**, 945-951.
10. Ayres, M. D., Howard, S. C., Kuzio, J., Lopez-Ferber, M., and Possee, R. D. (1994) *Virology* **202**, 586-605.
11. Choi, J., and Guarino, L. A. (1995) *J Virol* **69**, 4548-4551.
12. Rodems, S. M., and Friesen, P. D. (1995) *J Virol* **69**, 5368-5375
13. Leisy, D. J., Rasmussen, C., Kim, H. T., and Rohrmann, G. F. (1995) *Virology* **208**, 742-752.
14. Lu, A., and Carstens, E. B. (1993) *Virology* **195**, 710-718.
15. Lu, M., Farrell, P. J., Johnson, R., and Iatrou, K. (1997) *J Biol Chem* **272**, 30724-30728
16. Kremer, A., and Knebel-Mörsdorf. (1998) *Virology* **249**, 336-351
17. Rodems, S. M., and Friesen, P. D. (1993) *J Virol* **67**, 5776-5785.

18. Kovacs, G. R., Choi, J., Guarino, L. A., and Summers, M. D. (1992) *Journal of Virology* **66**, 7429-7437

19. Habib, S., and Hasnain, S. E. (1996) *J Biol Chem* **271**, 28250-28258

20. Chang, Y. N., Crawford, S., Stall, J., Rawlins, D. R., Jeang, K. T., and Hayward, G. S. (1990) *J Virol* **64**, 264-277.

21. Sandford, G. R., and Burns, W. H. (1996) *Virology* **222**, 310-317.

22. Bantignies, F., Rousset, R., Desbois, C., and Jalinot, P. (1996) *Mol Cell Biol* **16**, 2174-2182.

23. Fax, P., Lipinski, K. S., Esche, H., and Brockmann, D. (2000) *J Biol Chem* **275**, 8911-8920.

24. Hottiger, M. O., and Nabel, G. J. (2000) *Trends Microbiol* **8**, 560-565.

25. Ohresser, M., Morin, N., Cerutti, M., and Delsert, C. (1994) *J Virol* **68**, 2589-2597.

26. Rodems, S. M., Pullen, S. S., and Friesen, P. D. (1997) *Journal of Virology* **71**, 9270-9277

27. Luckow, B., and Schutz, G. (1989) *Nucleic Acids Res* **17**, 8451-8462.

28. Vaughn, J. L., Goodwin, R. H., Tompkins, G. J., and McCawley, P. (1977) *In Vitro* **13**, 213-217.

29. Chaabihi, H., Ogliastro, M. H., Martin, M., Giraud, C., Devauchelle, G., and Cerutti, M. (1993) *J Virol* **67**, 2664-2671.

30. Gorman, C. M., Merlino, G. T., Willingham, M. C., Pastan, I., and Howard, B. H. (1982) *Proc Natl Acad Sci U S A* **79**, 6777-6781.

31. Dignam, J. (1990) *Methods Enzymol* **182**, 194-203

32. Chodosh, L. A., Carthew, R. W., and Sharp, P. A. (1986) *Mol Cell Biol* **6**, 4723-4733.

33. Andrisani, O. M., Hayes, T. E., Roos, B., and Dixon, J. E. (1987) *Nucleic Acids Res* **15**, 5715-5728.

34. Fenstermaker, R. A., Farmerie, T. A., Clay, C. M., Hamernik, D. L., and Nilson, J. H. (1990) *Mol Endocrinol* **4**, 1480-1487.

35. Rourke, I. J., Hansen, T. V., Nerlov, C., Rehfeld, J. F., and Nielsen, F. C. (1999) *FEBS Lett* **448**, 15-18.

36. Tjian, R., and Maniatis, T. (1994) *Cell* **77**, 5-8.

37. Miller, L. K. (1997) *The baculoviruses* (Fraenkel-Conrat, H., and Wagner, R. R., Eds.), Plenum Press, New York

38. Leisy, D. J., and Rohrmann, G. F. (2000) *Virology* **274**, 196-202.

39. Smolik, S. M., Rose, R. E., and Goodman, R. H. (1992) *Mol Cell Biol* **12**, 4123-4131.

40. Usui, T., Smolik, S. M., and Goodman, R. H. (1993) *DNA Cell Biol* **12**, 589-595.

41. Hewes, R. S., Schaefer, A. M., and Taghert, P. H. (2000) *Genetics* **155**, 1711-1723.

42. Perkins, K. K., Admon, A., Patel, N., and Tjian, R. (1990) *Genes Dev* **4**, 822-834.

43. Zhang, K., Chaillet, J. R., Perkins, L. A., Halazonetis, T. D., and Perrimon, N. (1990) *Proc Natl Acad Sci U S A* **87**, 6281-6285.

44. Andrisani, O., and Dixon, J. E. (1990) *J Biol Chem* **265**, 3212-3218.

45. Jeang, K. T., Boros, I., Brady, J., Radonovich, M., and Khoury, G. (1988) *J Virol* **62**, 4499-4509.

46. Hai, T., and Curran, T. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 3720-3724

47. Hai, T., and Hartman, M. G. (2001) *Gene* **273**, 1-11.

48. Habib, S., Pandey, S., Chatterji, U., Burma, S., Ahmad, R., Jain, A., and Hasnain, S. E. (1996) *DNA Cell Biol* **15**, 737-747

49. Lu, M., Johnson, R. R., and Iatrou, K. (1996) *Virology* **218**, 103-113

50. Kerry, J. A., Priddy, M. A., Jervey, T. Y., Kohler, C. P., Staley, T. L., Vanson, C. D., Jones, T. R., Iskenderian, A. C., Anders, D. G., and Stenberg, R. M. (1996) *J Virol* **70**, 373-382.

51. Kerry, J. A., Priddy, M. A., Staley, T. L., Jones, T. R., and Stenberg, R. M. (1997) *J Virol* **71**, 2120-2126.

52. Choi, J., and Guarino, L. A. (1995) *Virology* **209**, 99-107.

53. Chisholm, G. E., and Henner, D. J. (1988) *Journal of Virology* **62**, 3193-3200

54. Pearson, G., and Rohrmann, G. (1997) *Virology* **235**, 153-165

55. Kovacs, G. R., Guarino, L. A., and Summers, M. D. (1991) *J Virol* **65**, 5281-5288.

56. Theilmann, D. A., and Stewart, S. (1991) *Virology* **180**, 492-508.

57. Kool, M., Voeten, J. T., Goldbach, R. W., Tramper, J., and Vlak, J. M. (1993) *J Gen Virol* **74**, 2661-2668.

58. Carey, M. (1998) *Cell* **92**, 5-8

59. Vo, N., and Goodman, R. H. (2001) *J Biol Chem* **276**, 13505-13508.

[1]FOOTNOTE

The abbreviations used are: *Ac*, *Autographa californica*; ATF, activating transcription factor; bHLH, basic helix-loop-helix; *Bm*, *Bombyx mori*; bZip, basic region-leucine zipper; CAT, chloramphenicol acetyl transferase; CBP, CREB binding protein; C/EBP, CAAT box-enhancer binding protein; CMV, cytomegalovirus; CRE, cAMP response element; hr, homologous region; HTLV1, human T-cell leukemia virus 1; IE, immediate-early; *Ld*, *Lymantria dispar*; LTR, long terminal repeat; MNPV, multinuclear polyhedrosis virus; *Op*, *Orgya pseudotsugata*; *Px*GV, *Plutella xylostella* granulovirus; *Sf*, *Spodoptera frugiperda*; TRE, TPA response element.

FIGURES LEGENDS

Figure 1. *Ac*MNPV hr sequences contain a large number of CRE, TRE and E-box motifs

Schematic representation of the eight *Ac*MNPV hrs : IE1-binding 28-mer palindromes (double arrows) and interpalindromic fragments (numbered /1 to /N) are indicated above and below each hr, respectively. Boxes and circles in interpalindromic sequences correspond to consensus CRE-like, TRE and E-box motifs. The hr5/4 fragment is boxed. Position of each hr in the *Ac*MNPV genome, in bp : hr1 (1-440), hr1a (7747-7864), hr2 (26293-26961), hr2a (48679-48708), hr3 (70468-71133), hr4a (93456-93605), hr4b (97396-97881), hr4c (102545-102648), hr5 (117479-117987).

Figure 2. Host factors specifically bind the CRE-like motif

EMSAs were performed using the CRE/EB probe, *Sf*9 nuclear extracts and competitors as indicated on the figure. (A) Competitions with mutated CRE oligonucleotides. Binding was carried out in the absence (lane 1) or in the presence of nuclear extracts (lanes 2-12). As in (B), names and sequences of competitors added in the binding reaction (lanes 3-12) are indicated below the figure, with CRE and E-box motifs boxed. Mutations in the CRE motif are in bold characters. Triangles above the lanes indicate increasing amounts (100- and 200-fold excess) of competitor. (B) Competitions with 5' end-truncated CRE oligonucleotides.

Figure 3. At least six host factors bind the CRE motif

(A) The CRE/EB probe was incubated alone (lane 1) or with *Sf*9 nuclear extracts (lanes 2-6) in conditions identical to EMSAs, before being UV-irradiated for 30 min. Binding complexes were separated on SDS-PAGE. Competitions (100-fold excess of competitor) were performed using homologous CRE/EB (lane 3) or mutated CREm1, CREm2, CREm3 oligonucleotides (lanes 4, 5, 6). (B) Detail of the 50 kDa-complex, shown after shorter exposure of (A). The lanes in (B) correspond to those in (A). Arrowheads indicate each of the 6 CRE-binding products. Molecular weights (MW, in kDa) are indicated on the left of (A). For oligonucleotide sequences, see figure 2.

Figure 4. CRE sites are functional *in vivo*

35 µg of total soluble proteins from *Sf*9 cells transfected using 20 µg prA_3CAT-derived plasmids were assayed for CAT activity, except hr5-CAT for which a ten fold dilution was used to be in the linear range of the enzymatic activity. CAT activity was represented as -fold enhancement over the activity of the basal prA_3CAT plasmid. Each

construct is schematized on the left of the figure: broken arrow and black CAT box indicate the prA_3CAT promoter-reporter gene construct, CRE boxes represent CRE motifs and single arrows represent half-sites (hs) of the 28-mer palindrome (P, double arrow). 1xhr5/4-CAT and 2xhr5/4-CAT schemes represent the native monomeric or dimerized hr5/4 fragment cloned upstream of prA_3CAT. Complete hr5 sequence (schematized in figure 1) is indicated by a hr5 box. A representative experiment among three independent ones is shown.

Figure 5. CRE-binding factors and IE1 bind together the pal-CRE probe

EMSAs were performed using the pal-CRE probe and either 2 µl *Sf*9 nuclear extracts (lanes 1-4), 2 µl partially purified IE1 (lanes 5-7) or a combination of both (lanes 8-12). Shifted complexes I-III are indicated by arrows. Competitors (100-fold excess) are indicated below the figure. For sequences, see figure 2 and experimental procedures.

Figure 6. CRE sites are required for the IE1-dependent transactivation in the viral context

25 µg of total soluble proteins from cotransfected or transfected/infected *Sf*9 cells were assayed for CAT activity. **Cotransfections** (lanes 1, 2, 4, 5, 7, 8, 10-12, 14-16): *Sf*9 cells were co-transfected using 10 µg prA_3CAT-derived constructs and 10 µg of either pUC 19 or the expression vectors pIE1 or pIE1i425. **Transfection/infections** (lanes 3, 6, 9, 13, 17) : *Sf*9 cells were co-transfected using 10 µg prA_3CAT-derived constructs and 10 µg of pUC 19, before being infected with *Ac*MNPV (MOI: 10) 24h post-transfection. Each reporter construct is schematized on the left of the figure, with the same representation as in figure 4, except mutated CRE motifs that are represented by CREm boxes. CAT activity is represented by % conversion, which indicates the percentage of acetylated forms of chloramphenicol. Each assay was performed in triplicate. Relative quantities of IE1 expressed in cotransection and transfection-infection assays were estimated by western blot (w.b.) using 100 µg and 10 µg total soluble proteins, respectively.

Figure 7. Proposition of an enhanceosome model for the IE1-dependent transactivation mediated by hrs

Viral and host transcription factors bind specifically to conserved and functionnaly important motifs in hrs : IE-1 on palindromes and host factors on CRE sites. Another host factor, hr1-BP, binds the palindrome and short flanking sequences in the minor groove of DNA, suggesting that it may be an architectural proteins (19). In this respect, hr1-BP may bend DNA and bring together the enhancer-binding transcription factors and the pre-initiation complex. In

addition, preliminary data suggest that IE1 may interact with CBP, a crucial coactivator that could also recruit some additional proteins in the enhanceosome complex. "?" indicates interactions remaining to be precised or established.

C- RESULTATS COMPLEMENTAIRES ET PERSPECTIVES

Les travaux présentés dans cet article montrent que des protéines cellulaires fixent spécifiquement les sites CRE situés dans les régions hrs du baculovirus *Ac*MNPV. Ces sites, fonctionnels dans la cellule non infectée, sont indispensables dans le contexte de l'infection virale pour la transactivation dépendante de IE1. Ces résultats, ainsi que les différentes données déjà publiées sur la transactivation dépendante de IE1 et médiée par les hrs, permettent de proposer un modèle de type enhanceosome rendant compte de cette transactivation. Cependant, l'article ne présente pas tous les résultats obtenus, ni les expérimentations en cours : c'est l'objet de ce chapitre.

I- AFFINITÉ DES FACTEURS DE L'HÔTE POUR LE MOTIF TRE

Dans la discussion de l'article 1, nous indiquons que les facteurs cellulaires qui fixent le motif CRE ont une moindre affinité pour le motif TRE. Ce point est intéressant, car plusieurs études chez les mammifères ont montré que les hétérodimères formés d'un membre de la famille AP1 (ayant par définition une plus forte affinité pour le site TRE) et d'un membre de la famille CREB/ATF (fixant le site CRE) ont en général plus d'affinité pour le site CRE que pour le site TRE (Hai et Curran, 1991). Dans notre cas, il pourrait donc s'agir d'un argument indirect en faveur de la fixation sur le site CRE d'hétérodimères plutôt que d' homodimères.

Pour comparer leur affinité relative, nous avons réalisé plusieurs expériences de gel retard en utilisant soit une sonde TRE/CRE (figure 19A), soit la sonde CRE/EB (figure 19B) et divers compétiteurs. Ces expériences montrent tout d'abord que des complexes semblables se forment sur les deux sondes, suggérant que les mêmes facteurs de l'insecte se fixent sur les deux motifs. Ce résultat n'est pas surprenant, puisque les deux motifs ne diffèrent que par un seul nucléotide. Cependant, la comparaison entre les compétitions avec le motif CRE et le motif TRE suggèrent que ces facteurs ont plus d'affinité pour CRE: en effet, dans tous les cas, les compétiteurs CRE déplacent très fortement le complexe retardé (puits 3, 6, 8, 10), alors que dans les mêmes conditions les compétiteurs TRE sont moins efficace (puits 4, 12, 13). La comparaison entre les puits 12 et 13 suggère également que la longueur des régions flanquantes sont importantes pour la fixation des facteurs cellulaires sur le site TRE: en effet, le compétiteur TRE/EB, qui porte 8 nucléotides en 3' du motif TRE, est plus efficace que le compétiteur TRE, qui n'en porte que 3.

La compétition modérée observée avec le fragment CRE/TRE, qui porte pourtant un site CRE intact, pourrait être dûe au fait qu'il a été ajouté en moindre excès que le compétiteur CRE: en

effet, nous avons utilisé dans tous les cas 1µg de compétiteur, alors que la masse molaire de CRE/TRE est environ le double de celle du fragment CRE.

II- PURIFICATION DES PROTÉINES FIXANT LE MOTIF CRE

L'un des objectifs initiaux de ce travail de thèse était l'identification des protéines cellulaires fixant les motifs CRE. La détermination de leur nombre et leur taille approximative, obtenue par la technique d'UV-crosslinking, est présentée dans l'article 1. Nous avons en parallèle commencé à les purifier, pour pouvoir les identifier par microséquençage.

1. Résultats

1.1. Chromatographie d'affinité avec le motif CRE

Les expériences de chromatographie d'affinité ont donné des résultats peu probants, en ce sens que 1) malgré des lavages répétés, un grand nombre de protéines sont obtenues à l'issue de l'élution de la colonne, probablement à cause de leur interaction non spécifique avec la biotine, la streptavidine et/ou les billes magnétiques, et 2) la comparaison entre les résultats obtenus avec la sonde CRE et la sonde CRE mutée ne montre pas de différences significatives. Après plusieurs essais infructueux dans différentes conditions de temps d'incubation et de stringence des lavages, nous avons opté pour une stratégie de séparation des protéines par leurs propriétés physico-chimiques, sans utiliser leur affinité pour CRE.

1.2. Séparation sur colonnes successives

Malgré sa lourdeur, cette stratégie présente plusieurs avantages: elle permet de partir de quantités beaucoup plus importantes d'extraits nucléaires au départ, ce qui peut être crucial si les facteurs recherchés sont présents en faible quantité dans la cellule. Par ailleurs, nous disposons grâce au gel retard d'un moyen très fiable et sensible pour repérer les fractions contenant les facteurs recherchés.

Nous avons utilisé 1 litre de culture cellulaire, soit $1,5x10^9$ cellules, dont les extraits nucléaires ont été passés sur une colonne de gel filtration pour un premier fractionnement par la taille. La figure 20 présente le gel retard effectué à partir des fractions obtenues (nommées GF1 à GF19). On observe deux complexes retardés distincts provenant de fractions différentes (GF9-10 et GF12-13), suggérant une première séparation des facteurs fixant CRE. Nous avons poursuivi la purification des fractions présentant le signal le plus intense (fractions GF12-13) par passage sur une colonne d'héparine, un substrat utilisé généralement pour la purification des protéines ayant

de l'affinité pour l'ADN. Le gel retard a permis de repérer les fraction présentant une activité de fixation sur les sites CRE (figure 21, fractions H8-18). Pour repérer les protéines impliquées dans ces complexes retardés, nous avons ensuite effectué des expériences d'UV-crosslinking avec les compétiteurs CRE/EB3 et CREmut, qui sont présentées sur la figure 22A. Les résultats obtenus par autoradiographie du gel complètent et précisent la figure 3 de l'article 1, dans laquelle on distinguait 6 protéines de 28, 40, 45, 55, 83 et 170 kDa : le complexe retardé correspondant à la fraction GF10 est composé des 4 protéines dont le PM est le plus élevé (45, 55, 83 et 170 kDA), et suggère que le produit de 83 kDa pourrait être constitué de plusieurs protéines de PM proches. Dans le cas de la fraction GF13, les protéines impliquées semblent être celles de petit PM (majoritairement le complexe autour de 50 kDa déjà observé dans la figure 3A de l'article, et une faible quantité de la protéine de 28 kDa). Enfin, la fraction H11 semble contenir les mêmes facteurs, mais en proportion inverse: le facteur de 28 kDa est majoritaire.

Après autoradiographie, le gel a été coloré au nitrate d'argent (figure 22B). Sa superposition avec l'image précédente a permis de préciser la localisation des bandes colorées à l'argent susceptibles de correspondre aux protéines marquées (figures 22A et B). Cependant, les signaux diffus de l'autoradiographie et le grand nombre de bandes révélées à l'argent rendent cette méthode d'identification peu fiable. Nous avons donc effectué un troisième fractionnement en passant les fractions H9-15 sur une colonne échangeuse de cations (EC). Deux fractions seulement (EC3, 4) ont permis la formation d'un complexe en gel retard (figure 23A), et le crosslinking aux UV a révélé (par une exposition courte au film autoradiographique) qu'il contient deux protéines de PM très proche, autour de 28 kDa (figure 23B). La coloration à l'argent des protéines contenues dans la fraction 4 a permis de détecter deux bandes isolées mais d'intensité extrêmement faible à cette taille (non visibles sur la figure 23C) qui correspondent probablement aux protéines que nous recherchons. Dans le but de microséquencer ces produits, nous avons déposé la totalité du reliquat des fractions 4 et 5 (après précipitation au TCA) sur un gel polyacrylamide, que nous avons coloré au bleu de Coomassie après électrophorèse. Nous avons ainsi pu repérer les deux protéines, mais leur quantité s'est révélée insuffisante pour le microséquençage. Un nouveau cycle de fractionnement à partir de 1,8 l de cellules est en cours.

2. Discussion et perspectives

La purification en quantités suffisantes de ces facteurs de 28 kDa est donc l'objectif majeur de la suite de ce travail, et devrait permettre leur identification par microséquençage. Pour les autres protéines, les conditions de purification restent à déterminer. Le cas des facteurs autour de

50 kDa risque d'être assez délicat, étant donnée la diminution puis la disparition de leur signal en crosslinking aux UV au cours des étapes de purification successives (comparer les fractions GF13, H11 et EC2 des figures 22A et 23B). Cette diminution est peut-être dûe au fait que les facteurs impliqués sont élués dans des fraction différentes, alors qu'ils ne peuvent se fixer qu'en hétérodimères sur la sonde. Il est également possible qu'ils perdent leurs propriétés de fixation sur l'ADN au cours des étapes successives d'extraction et de fractionnement, en raison d'une stabilité réduite.

L'obtention de séquences peptidiques correspondant à ces protéines par microséquençage devrait permettre de remonter aux gènes, soit par PCR soit par criblage des banques d'ADN complémentaires de *S. frugiperda* disponibles au laboratoire, par l'utilisation d'amorces et de sondes dégénérées déterminées à partir de ces séquences peptidiques.

Une approche différente pour identifier les facteurs fixant le motif CRE pourrait être menée en parallèle: la glissière de leucine étant un domaine conservé chez les facteurs AP1 et CREB/ATF (voir par exemple la figure 13B), les séquences disponibles chez la drosophile pourraient permettre de désigner des sondes pour cribler les banques d'ADNc de *S. frugiperda*. Cette méthode présenterait l'avantage de donner directement accès à la séquence de gènes codant pour des protéines possédant un domaine bZip, et donc d'identifier et surtout d'exprimer ces facteur. Cela dit, étant donné leur probable faible niveau d'expression, il est très possible que ces gènes ne soient pas représentés dans les banque d'ADNc. A moyen terme, le criblage de la banque de BACs de *S. frugiperda*, en cours de caractérisation, devrait permettre de contourner ce problème potentiel.

III- VOIE AMPC ET ACTIVATION TRANSCRIPTIONNELLE MÉDIÉE PAR LES SITES CRE.

Nous n'avons pas pu tester si les protéines cellulaires fixant les sites CRE sont phosphorylées ou non. Par contre, nous avons voulu voir si la voie de transduction du signal dépendante de l'AMPc, qui aboutit chez les mammifères à la phosphorylation et à l'activation de CREB, joue un rôle dans l'activation transcriptionnelle par les sites CREs en transfection dans la cellule *Sf*9.

Pour cela, nous avons transfecté des cellules avec les plasmides 3xCRE-CAT et 3xCREm-CAT (constructions schématisées dans la figure 4 de l'article 1) en présence soit d'un activateur (Forskoline) soit d'un inhibiteur (H89) de la voie AMPc (la figure 24 montre le mode d'action de ces deux molécules) puis mesuré l'activité CAT 48H post transfection. Ces expériences préliminaires n'ont pas donné de résultat clair, les effets des deux produits sur l'activité CAT

variant d'un essai à l'autre, ce qui suggère que la perturbation du taux d'AMPc intracellulaire et de l'activité de la protéine kinase A n'a pas d'effet notable sur l'activation transcriptionnelle par les sites CRE. Ces observations concordent avec les résultats exposés dans la thèse de Myriam Julien (1999), montrant des taux très élevés d'AMPc constitutif (c'est-à-dire sans induction particulière) dans la cellule *Sf*9.

Toutes ces données suggèrent donc que la voie AMPc n'est pas fonctionnelle dans cette lignée, comme cela a été montré dans une lignée de carcinome embryonnaire humain, F9 (Masson *et al*., 1992). Ces points restent cependant à confirmer, ainsi que les modalités d'activation des facteurs se fixant sur les sites CRE, dans la cellule *Sf*9 non infectée et surtout dans le contexte de l'infection.

IV- PURIFICATION DE IE1 EXPRIMÉ EN SYSTÈME HOMOLOGUE/HÉTÉROLOGUE

Lors d'expériences préliminaires en gel retard avec la sonde CRE-pal pour montrer la fixation simultanée de IE1 et de facteurs cellulaires sur les hrs, nous nous sommes aperçus que les résultats obtenus à partir d'extraits nucléaires bruts issus de cellules infectées ou transfectées avec pIE1 étaient inexploitables, car 1) très "smeareux", probablement à cause du trop grand nombre de facteurs (spécifiques et/ou parasites) se fixant sur la sonde, et 2) il était impossible de dissocier les retards de migration dûs à IE1 et aux facteurs cellulaires. Pour résoudre ce double problème, nous avons décidé de travailler d'une part sur des extraits nucléaires de cellules *Sf*9 fractionnés (c'est l'objet de la partie précédente), et d'autre part sur la protéine IE1 purifiée produite dans un système hétérologue. Pour cela, nous avons tout d'abord développé une stratégie d'expression de la protéine en bactéries, simple à mettre en oeuvre, avec ajout d'un "tag" à l'extrémité N-terminale pour faciliter sa purification. Dans une première expérience, IE1 a été élué de la colonne d'affinité anti-FLAG par compétition avec un excès de peptide FLAG. La figure 25 montre que IE1 est fortement exprimé dans le système bactérien, et que sa purification semble efficace (puits 9 et 10). Nous avons ensuite testé IE1 purifié en gel retard avec différentes sondes portant le palindrome (CRE-pal, pal) mais aucun complexe n'a pu être détecté. Pour éliminer la possibilité d'un problème lié à la présence du tag, nous avons effectué de nouvelles purifications, en éluant cette fois IE1 par clivage du FLAG à l'entérokinase. Malheureusement, le gel retard a encore une fois donné un résultat négatif. L'absence d'activité de fixation sur le palindrome de cette protéine, par ailleurs bien exprimée et détectée par l'anticorps polyclonal anti IE1, n'est pas liée à la nécessité d'autres facteurs, puisque Choi et

Guarino (1995a) ont montré que IE1 produit en lysat de réticulocyte peut se fixer seul sur son site. Il reste donc la possibilité d'un problème lié à un mauvais repliement de la protéine ou des modifications post-traductionnelles inadéquates dans le système bactérien.

Nous avons donc dans un deuxième temps adopté une stratégie d'expression transitoire de IE1 " flagué " dans la cellule *Sf*9, par transfection: deux clones porteurs de l'épitope FLAG ont été transfectés, et l'expression de FLAG-IE1 vérifiée (et avérée) par western blot. La protéine a été purifiée selon la même méthode que précédemment, avec élution compétitive par le peptide flag ou clivage par l'entérokinase, puis utilisée en gel retard. Comme pour l'expression en bactéries, aucun complexe retardé n'a pu être détecté, sans que nous parvenions à expliquer pourquoi. Les conditions du gel retard ne sont pas en cause, puisque IE1 partiellement purifié à partir de cellules infectées est capable de se fixer sur le même site dans les mêmes conditions. Par ailleurs, Olson *et al*. (2001) ont montré que le peptide FLAG porté par IE1 (exprimé en lysat de réticulocyte) n'empêche pas sa fixation sur le palindrome en gel retard. Il est possible que des mutations dans la séquence codante de IE1 soient apparues lors de la construction des plasmides (notamment l'étape de mutagenèse dirigée), la protéine mutée ne pouvant alors plus se fixer sur l'ADN: nous n'avons en effet vérifié par séquençage que la région qui porte la séquence codante du FLAG dans le plasmide pFLAG-IE1. Une autre hypothèse pourrait être que l'expression de FLAG-IE1 en système hétérologue et/ou les conditions de sa purification rendent cette protéine incapable de se fixer sur l'ADN dans nos conditions de gel retard. Ceci pourrait être dû à une mauvaise maturation post-traductionnelle ou à la dégradation partielle de la protéine au cours des étapes de purification.

Ces approches s'étant finalement révélées décevantes, il sera probablement plus efficace dans la suite du travail de purifier IE1 à partir de cellules infectées, comme nous avons déjà commencé à le faire. Cette méthode, bien qu'assez lourde, présente l'avantage de permettre l'obtention de la protéine "naturelle" produite lors de l'infection, et pourrait permettre la co-purification de protéines cellulaires ou virales interagissant avec elle. En parallèle, le système d'expression de flag-IE1 en lysat de réticulocyte, qui produit une protéine active en gel retard indépendamment du contexte de la cellule d'insecte, est disponible depuis peu (Olson *et al*., 2001). Son utilisation devrait permettre de tester beaucoup plus précisément les interactions protéine-protéine et protéine-ADN entre IE1, ses partenaires cellulaires et les hrs in vitro.

V- INTERACTION ENTRE IE1 ET CBP

Comme nous l'avons abordé plusieurs fois dans l'introduction et dans l'article 1, CBP/P300 est une protéine coactivatrice ubiquiste et fondamentale, universellement impliquée par la multitude de ses interactions avec de nombreux facteurs de transcription dans la formation, la stabilité et l'activité des complexes transcriptionnels dans la cellule. De plus, CBP/P300 est la cible de nombreuses protéines transactivatrices virales, comme par exemple E1A de l'adénovirus, E6 du virus du papillome humain, Zta de l'EBV, Tat du HIV, Tax de HTLV1...(pour revue Hottiger et Nabel, 2000; Vo et Goodman, 2001). Au vu de ces interactions, il était intéressant de tester si le transactivateur majeur du baculovirus, IE1, est capable d'interagir avec CBP. Comme l'homologue de ce coactivateur n'a pas été identifié chez *S. frugiperda*, nous avons exprimé en baculovirus la CBP humaine marquée en N-terminal par l'épitope FLAG (Kurokawa *et al*., 1998), tenant compte du fait que les domaines d'interaction protéine-protéine de CBP sont assez bien conservés d'une espèce à l'autre. Des expériences de coimmunoprécipitations ont alors été effectuées: la CBP humaine exprimée en baculovirus a été purifiée sur des billes d'agarose couplées à l'anticorps anti-FLAG, puis après plusieurs lavages intensifs, la présence de IE1 révélée en western blot à l'aide d'un anticorps polyclonal (Ohresser *et al*., 1994). La figure 26 montre que la quantité de IE1 copurifié est nettement plus importante lorsque CBP est exprimée que quand elle ne l'est pas, suggérant que IE1 interagit avec la CBP humaine. Cependant, des études in vitro plus exhaustives sont requises, notamment par l'utilisation de témoins négatifs supplémentaires, comme par exemple des protéines "flaguées" autres que CBP.

Par ailleurs, des expériences de gel retard sont en cours pour tester précisément l'interaction CBP-IE1, avec utilisation de protéines IE1 et CBP purifiées, d'un anticorps anti-flag pour générer des "supershifts" et de différentes sondes portant le site de fixation de IE1. Ces expériences seront par ailleurs doublées d'essais de crosslink aux UV pour vérifier la taille des facteurs impliqués.

Les outils dont nous disposons devraient également permettre de tester l'importance de CBP dans notre modèle *in vivo*, par exemple par la mesure de l'effet de la surexpression de la CBP humaine dans des expériences de transfection ou de transfection-infection avec les constructions 2xhr5/4-CAT.

En parallèle, la CBP endogène devra être identifiée pour valider complètement notre modèle en cellule d'insecte, et pourrait permettre ultérieurement de caractériser les interactions que CBP

établit avec les autres facteurs de transcription (généraux et hr-spécifiques) impliqués dans la transactivation médiée par les hrs.

D- CONCLUSION

Cette première partie des travaux de thèse nous a permis à la fois de progresser dans la connaissance de la régulation de la transcription chez le baculovirus *Ac*MNPV, de caractériser de nouvelles interactions entre ce virus et sa cellule-hôte, *Sf*9, et de mettre en évidence l'existence chez *S. frugiperda* de nouvelles protéines cellulaires qui sont probablement des facteurs de transcription de la famille AP1 et CREB/CREM/ATF. Par ailleurs, des arguments indirects suggèrent qu'une autre protéine de l'hôte, le coactivateur CBP, est la cible du facteur viral IE1. Pour identifier ces protéines, nous avons mené en parallèle deux approches, dont la première a été décrite dans la section précédente ; elle est biochimique et consiste à purifier suffisamment ces facteurs pour pouvoir les microséquençer.

La deuxième stratégie a été initiée dans le but d'accéder plus facilement à la séquence codante des gènes exprimés dans la cellule *Sf*9, et a abouti à la construction d'une banque d'ADN complémentaire. Son exploitation, par criblage et par séquençage à grande échelle, fait l'objet de la deuxième partie de cette thèse.

Mise en place et exploitation d'une banque d'ADN complémentaires de cellules *Sf*9

A- INTRODUCTION

Spodoptera frugiperda (figure 27 A, B, C) est un lépidoptère important, tant au niveau économique que pour son intérêt scientifique: d'un point de vue agronomique, c'est un ravageur des cultures polyphage, qui provoque des dégâts considérables (figure 27 E, F) en Amérique du sud (Ching'oma et Pitre, 1999), par exemple au Brésil où cette espèce est responsable de pertes avoisinant les 285 millions de dollars sur coton, maïs, riz et millet (source: EMBRAPA, www.cnpms.embrapa.br). De plus, *S. litura* et *S. littoralis*, deux espèces très proches, provoquent elles aussi des dégâts en Amérique du Nord et Centrale, en Afrique, en Inde et dans les serres en Europe (Anderson *et al*., 2001). A côté de cet aspect agronomique, *S. frugiperda* a un autre intérêt économique en tant qu'organisme source des lignées cellulaires *Sf*21 et *Sf*9 (figure 27D) (Vaughn *et al*., 1977). Ces cellules sont les hôtes de laboratoire du baculovirus *Ac*MNPV, très utilisé en tant que vecteur d'expression de protéines hétérologues. On peut donc les considérer comme des usines biologiques pour la production de protéines recombinantes. Enfin, d'un point de vue scientifique, et comme nous l'avons abordé dans le chapitre précédent, le système baculovirus-cellule d'insecte est le modèle le plus avancé pour l'étude fondamentale et appliquée des interactions moléculaires entre virus et cellule-hôte chez les invertébrés.

Cependant, le fait que relativement peu de gènes aient été caractérisés chez *S. frugiperda* constitue un frein considérable au développement des travaux sur cette espèce. Nous avons donc initié un programme d'étude génomique, qui a débuté par la construction d'une banque d'ADN complémentaires de cellules *Sf*9. Nous avons ensuite défini plusieurs objectifs concernant son exploitation en relation avec les thématiques de l'équipe et du laboratoire:

1) criblage de la banque par sonde nucléique, pour rechercher la séquence codante de deux gènes mis en évidence au cours d'études transcriptionnelles menées sur le baculovirus et sa cellule-hôte : hsp90 et CBP.

2) séquençage à grande échelle de cette banque, débouchant sur l'établissement et la caractérisation globale d'une banque d'ESTs *Sf*9, cette étude étant axée sur la mise en évidence de caractéristiques spécifiques des lépidoptères.

B- EXPLOITATION DE LA BANQUE D'ADNC PAR CRIBLAGE AVEC UNE SONDE NUCLEIQUE

I- HSP90

L' identification du gène codant pour la protéine HSP90 de *Spodoptera frugiperda* est le résultat d'une étude initiée par Martine Cerutti et Mylène Ogliastro, et poursuivie par Jean

Michel Pommet sur un baculovirus recombinant exprimant le gène du récepteur à l'acide rétinoïque humain, isoforme gamma sous contrôle du promoteur de la polyédrine (AcRAR ; (Ogliastro, 1994; Pommet, 2000). Les premiers travaux ont montré que ce virus présente la particularité de surexprimer la protéine très tardive P10 (Ogliastro, 1994). Une analyse plus précise de ce virus a également permis de montrer un retard de mortalité notable des cellules infectées, ce qui est paradoxal puisqu'une des fonctions suggérées de la protéine P10 est d'accélérer la désintégration du noyau cellulaire (Pommet, 2000). Pour tenter d'identifier les gènes viraux et cellulaires impliqués dans le retard de mortalité observé, une banque différentielle d'ADNc a été construite à partir d'ARNm de cellules infectées par le virus sauvage et par *Ac*RAR. Parmi les clones obtenus, l'un d'entre eux portait un fragment de 279 pb avec une séquence présentant de très fortes similarités avec la séquence du gène conservé *hsp90*. L'utilisation de ce fragment comme sonde en Northern blot a ensuite montré que le gène *hsp90* de *Spodoptera frugiperda* est effectivement surexprimé lors de l'infection par le virus AcRAR, un résultat confirmé sur la protéine par utilisation d'un anticorps anti-HSP90 (Pommet, 2000). Par ailleurs, des cinétiques de la quantité d'ARNm ont montré que l'expression de ce gène de l'hôte n'est pas affectée par l'extinction générale des synthèses cellulaires induite par le virus sauvage dès le début de l'infection : autrement dit, hsp90 est exprimé à des niveaux élevés tout au long de l'infection, y compris pendant les phases très tardives, un résultat inattendu qui renforce l'idée que la protéine est impliquée dans le retard de mortalité de la cellule infectée par *Ac*RAR.

Pour aller plus loin dans l'analyse de la régulation et des effets de HSP90 au cours de l'infection, la construction d'un vecteur d'expression (plasmide, baculovirus recombinant) était une étape incontournable. Nous avions donc besoin de la séquence codante complète du gène. Pour cela, nous avons criblé la banque d'ADNc *Sf*9 avec la sonde déjà utilisée en Northern blot, ce qui a conduit à l'obtention de plusieurs clones de phages. Leur conversion en plasmide et le séquençage de l'un d'entre eux a finalement permis d' obtenir la séquence complète de l' ADNc du gène *hsp90* de *Spodoptera frugiperda*. Une analyse in silico de la banque d'ESTs de *Bombyx mori* accessible sur internet (www.ab.a.u-tokyo.ac.jp/silkbase) a en outre conduit à la reconstitution de la séquence chez cet autre lépidoptère. En collaboration avec le laboratoire japonais qui a établi cette banque, nous avons poursuivi la caractérisation parallèle de *hsp90* chez ces deux lépidoptères, aboutissant à la publication d'un article dans la revue Gene (Landais *et al*., 2001; en annexe).

II- CBP

Comme nous l'avons vu dans la première partie de ce mémoire, il est possible que la protéine CBP/P300 soit la cible du transactivateur viral IE1 et interagisse avec certains autres facteurs de transcription impliqués dans la transactivation dépendante de IE1. Par ailleurs, en tant que coactivateur général de la transcription dépendante de l'ARN polymérase II, c'est une protéine clé de la cellule qui établit (tout au moins chez les mammifères) des contacts avec une multitude de protéines cellulaires. L'identification de la protéine CBP/P300 endogène chez *S. frugiperda* nous serait donc très utile, non seulement pour mieux caractériser les mécanismes de la transactivation virale, mais également pour accéder aux protéines cellulaires avec lesquelles elle interagit.

CBP/P300 est une grosse protéine d'environ 300 kDa, dont la séquence complète est disponible chez un petit nombre d'espèces seulement : l'Homme, le rat, la souris, la drosophile et Caenorhabditis (Goodman et Smolik, 2000; Vo et Goodman, 2001). Sa séquence n'est pas très conservée, sauf au niveau de certains domaines fonctionnels cractéristiques de la protéine (figure 15). Pour identifier la CBP de *S. frugiperda*, nous avons tout d'abord criblé la banque d'ADNc *Sf*9 avec la séquence codante de la CBP humaine (Lan Xu, communication personnelle), postulant que l'hybridation avec les domaines conservés serait suffisante pour repérer les clones positifs. Malgré plusieurs tentatives en variant la stringence de l'hybridation et des lavages, cette stratégie a échoué, sans que nous puissions en déterminer la cause : soit la banque ne contient aucun clone correspondant à ce gène, soit l'homologie de séquence entre la CBP humaine et celle de *S. frugiperda* est trop faible. Pour tenter de contourner ce problème potentiel, nous avons développé une seconde approche basée sur l'utilisation des domaines les plus conservées du gène CBP: tout d'abord, à partir de l'alignement global des séquences nucléotidiques de l'Homme, de la souris et de la drosophile nous avons sélectionné deux régions très conservées entre ces trois espèces. Nous avons ensuite désigné des oligonucléotides dans ces domaines (figure 28), avec lesquels des PCR ont été réalisées sur l'ADN de cellules *Sf*9. Cependant, aucun produit d'amplification n'a pu être obtenu malgé l'utilisation de conditions de PCR variées. D'autres approches doivent donc être élaborées pour identifier la CBP de *S. frugiperda* ; par exemple, le criblage de la banque avec une sonde issue de la séquence de la drosophile pourrait donner de meilleurs résultats qu'avec celle de l'Homme, le gène étant probablement plus conservé entre les deux insectes.

C- EXPLOITATION DE LA BANQUE PAR SEQUENÇAGE A GRANDE ECHELLE

I- MISE EN PLACE DE LA BANQUE D'ESTS ET CHOIX DE L'ANALYSE DES GÈNES DE PROTÉINES RIBOSOMALES

Le laboratoire a établi par l'intermédiaire de Philippe Fournier une collaboration avec le docteur K. Mita (National Institute of Agrobiological Sciences, Japon) pour le séquençage de 10000 clones de la banque d'ADNc de *Sf*9, qui a abouti à la mise en place d'une banque d'ESTs. Cette banque contient actuellement 5900 séquences, qui ont été complètement annotées par Gérard Devauchelle, permettant d'identifier 1687 cadres ouverts de lecture différents (ORFs). Ceux-ci ont été regroupés en 7 catégories (table 1 de l'article suivant), la plus représentée en termes d'ESTs étant la famille des gènes de protéines ribosomales (30% du nombre total d'ESTs), qui contient 76 membres. La redondance moyenne du nombre d'ESTs par gène étant très élevée dans cette famille (21), nous avons pu vérifier et/ou compléter dans la plupart des cas la séquence de l'ADNc, aboutissant à l'obtention de la séquence complète de 67 ADNc de protéines ribosomales chez *Spodoptera frugiperda*. Nous avons choisi d'approfondir l'analyse de cette famille, car elle présente des caractéristiques intéressantes dans le cadre de l'analyse du génome des lépidoptères: 1) les gènes qui la composent sont en général très conservés d'une espèce à l'autre, ce qui facilite les alignements et comparaisons de séquences, 2) la famille est très diversifiée, avec environ 80 gènes ne présentant aucune homologie de séquence entre eux, augmentant les chances que certains d'entre eux présentent des particularités propres aux lépidoptères ou aux insectes, 3) c'est une famille de gènes fonctionnellement très importante, puisqu'en plus d'être des constituants essentiels du ribosome, les protéines ribosomales peuvent avoir d'autres fonctions, dites "extraribosomales" (Wool, 1996), et 4) la séquence de tous les gènes de protéines ribosomales sont disponibles chez la Drosophile, permettant une comparaison globale de cette famille de gènes chez les insectes. De plus, même si elles demandent un gros travail de mise en forme, les mêmes séquences peuvent être obtenues chez l'Anophèle, dont le génome vient d'être publié, et chez *B. mori*, à partir des banques d'ESTs. Enfin, malgré l'importance de la famille, celle-ci n'a paradoxalement jamais été décrite globalement chez les insectes. C'est l'objet de l'article suivant, qui sera présenté après quelques rappels sur les protéines ribosomales.

II- RAPPELS SUR LES PROTÉINES RIBOSOMALES

1. Caractéristiques, nombre et nomenclature

Il existe deux types de ribosomes dans la cellule eucaryote : le ribosome "cytoplamique", présent sur le réticulum endoplamique ou libre dans le cytoplasme, et le ribosome mitochondrial (mitoribosome) présent dans la mitochondrie (Suzuki *et al*., 2001). Dans la suite du travail, nous allons étudier exclusivement le ribosome cytoplasmique de *S. frugiperda*.

Les protéines ribosomales sont en général des protéines basiques et de petite taille (Warner, 1999). Elles ont été initialement séparées par des méthodes biochimiques (gels 2D) à partir de ribosomes entiers, ce qui a conduit à l'identification progressive de la plupart d'entre elles. Cependant, ces méthodes se sont révélées hasardeuses pour l'estimation du nombre de protéines ribosomales et leur comparaison entre espèces, et ont abouti à la publication de nomenclatures très différentes d'une espèce à l'autre (pour revue, Mager *et al*., 1997). Les progrès du séquençage ont ensuite permis de repérer les orthologues chez les différents organismes, aboutissant à une nomenclature plus ou moins unifiée, présentée par Swiss-Prot dans une base spécialisée (www.expasy.ch/cgi-bin/lists?ribosomp.txt/). Pour notre part, en accord avec d'autres équipes travaillant sur les protéines ribosomales de Lépidoptères (K. Gordon et D. Heckel, en Australie) nous avons choisi de suivre la nomenclature utilisée chez le rat (Wool *et al*., 1995), qui est la plus complète et la première à avoir été publiée chez les eucaryotes. Cette classification correspond à celle de Swiss-Prot, sauf pour une protéine ribosomale : la L36A correspond à la L44 chez le rat.

On compte environ 55 protéines ribosomales chez les procaryotes (Ramakrishnan et White, 1998) et 80 chez les eucaryotes (Doudna et Rath, 2002), ces derniers possédant des ribosome de plus grande taille et plus complexes. Si leur nombre est à peu près constant chez les différentes espèces où elles ont été recensées (la levure en compte 78, une de moins que l'Homme et le rat), chez les insectes par contre aucune donnée n'a encore été publiée. Au niveau des gènes codant ces protéines ribosomales, la situation est plus contrastée: chez l'Homme par exemple, chaque protéine ribosomale est codée par un seul gène fonctionnel, mais une multitude de pseudogènes non exprimés sont présents dans le génome (Kenmochi *et al*., 1998). Chez la levure, on compte 137 gènes fonctionnels, résultant d'une ancienne duplication du génome (Warner, 1999). Des études pour localiser ces gènes dans le génome de ces deux espèces ont montré qu'ils se répartissent sur tous les chromosomes, contrairement aux ARN ribosomaux.

2. Origine

Il est probable que le ribosome primordial était constitué uniquement d'ARN, et que les protéines ribosomales ne sont venues l'"habiller" qu'ensuite (Wool, 1996). Celles-ci ont deux origines possibles : soit elles préexistaient dans la cellule avec une fonction définie et ont été recrutées ensuite par le ribosome, soit elles sont apparues spécifiquement pour remplir une fonction dans le ribosome.

3. Expression

Le ribosome eucaryote est l'un des organites les plus complexes de la cellule, avec deux sous-unités constituées de 80 protéines et 4 molécules d'ARN. Son assemblage a lieu dans le noyau, au niveau du nucléole, lors d'un processus très complexe que nous ne détaillerons pas ici (pour revue, Jakel et Gorlich, 1998; Scheer et Hock, 1999). Ce mécanisme implique que les protéines ribosomales, traduites dans le cytoplasme comme des protéines "classiques", passent dans le noyau pour être incorporées dans le ribosome. Pour cela, elles possèdent un signal de localisation nucléaire particulier et empruntent une voie d'import nucléaire spécifique (Rout *et al*., 1997).

Etant donné que le ribosome contient un exemplaire de chaque protéine ribosomale, leur synthèse doit être rigoureusement coordonnée, car l'assemblage du ribosome ne peut être menée à bien que si tous les éléments le composant sont disponibles au bon moment. Cette situation est parfaitement illustrée par les mutants Minute chez la drosophile, chez lesquels le phénotype anormal (petite taille, retard de développement, fertilité réduite...) est lié à la production insuffisante d'une seule protéine ribosomale (Saeboe-Larssen *et al*., 1998). Plusieurs études ont montré que les promoteurs de la plupart des gènes de protéines ribosomales contiennent des sites de fixation pour des facteurs de transcription impliqués dans la coordination de l'expression (Li *et al*., 1999; Yoshihama *et al*., 2002), mais le mécanisme reste mal connu.

4. Rôles

Comme leur nom l'indique, les protéines ribosomales sont des constituants importants du ribosome et jouent donc un rôle dans la synthèse des protéines. Leur rôle est essentiellement structural, et permet à la fois de conformer correctement et de stabiliser la structure tertiaire des ARN ribosomaux. La cristallisation récente de plusieurs ribosomes procaryotes (Ramakrishnan, 2002) a permis d'affiner considérablement leur agencement au sein du ribosome et leur fonction.

Par exemple, le support de l'activité peptidyl transferase de la grande sous-unité du ribosome a longtemps été débattu: est-il lié uniquement à l'ARN ribosomal ou bien implique-t-il des protéines? Les travaux de Nissen *et al*. (2000) ont clairement répondu à la question, en montrant que le site enzymatique ne contient aucune protéine ribosomale: l'ARN ribosomal seul possède l'activité catalytique, et le ribosome est bien un ribozyme.

Un nombre important de travaux publiés ces dernières années ont montré que certaines protéines ribosomales ont aussi une fonction extraribosomale. S20 par exemple est impliquée dans la transcription chez la levure, tandis que S3 joue un rôle dans la réparation de l'ADN chez la Drosophile et l'Homme. L'existence de ces fonctions extraribosomales et le fait qu'elles sont pour la plupart liées à la capacité d'interagir avec les acides nucléiques suggère que ces protéines préexistaient à leur recrutement dans le ribosome, avec une activité liée à l'ADN ou à l'ARN (Wool, 1996).

5. Biais de codons

Une des caractéristiques majeures des protéines ribosomales est qu'elles sont présentent en un seul exemplaire dans le ribosome (Howe et Hershey, 1982). Comme nous l'avons vu plus haut, leur expression est coordonnée et aboutit à la synthèse de quantités équivalentes de chacune d'entre elle dans la cellule, du moins théoriquement. Cette propriété a trouvé une application intéressante dans le domaine de l'analyse de l'évolution des génomes, dans le cadre de l'étude du biais de codons.

Le code génétique étant redondant, la plupart des acides aminés sont codés par plusieurs codons. Or l'analyse des séquences codantes de nombreux gènes a montré que, dans la plupart des cas, certains codons sont plus représentés que d'autres: c'est la préférence (ou biais) de codon, par contraste avec une situation où tous les codons seraient représentés de façon équivalente (pas de biais). D'autres travaux ont ensuite établi que les codons correspondant aux ARN de transfert les plus abondants dans la cellule sont sélectionnés préférentiellement, l'augmentation du taux de synthèse des protéines qui en résulte procurant un avantage sélectif à l'organisme (Akashi, 2001; Moriyama et Powell, 1997).

Par ailleurs, l'intensité du biais peut varier considérablement d'un gène à l'autre au sein de la même espèce ou bien pour le même gène chez des espèces différentes (figure 29): certains gènes sont extrêmement biaisés (à la limite, chaque acide aminé est codé par un seul codon), d'autres pas du tout (tous les codons sont représentés de la même façon). Pour expliquer ce phénomène, plusieurs hypothèses ont été testées, et ont abouti à la mise en évidence de plusieurs paramètres qui agissent sur l'intensité du biais: le niveau d'expression du gène est probablement le facteur le

plus déterminant, car plus un gène est exprimé, plus son coût de production est élevé pour la cellule; un fort biais de codon, en augmentant l'efficacité de la traduction, va minimiser ce coût et donc être sélectionné (Duret et Mouchiroud, 1999; Gouy et Gautier, 1982).

Parmi les autres paramètres, plusieurs auteurs ont suggéré que la longueur de la séquence codante pouvait avoir une influence sur le niveau de biais: plus la protéine à synthétiser est longue, plus sa synthèse coûte cher à la cellule, mais aussi plus les risques d'erreurs de traduction augmentent. Le biais de codon pourrait diminuer ces risques d'erreurs, par l'utilisation de codons pour lesquel l'appariement codon-anticodon est optimal (Akashi, 1994; Comeron et Kreitman, 1998). Cependant, l'analyse de la corrélation entre longueur des séquences codantes et intensité du biais de codon ont donné des résultats contrastés : chez les organismes unicellulaires, comme les bactéries ou la levure, une telle corrélation a été établie (Coghlan et Wolfe, 2000; Eyre-Walker, 1996). Par contre, chez trois espèces pluricellulaires (*Caenorhabditis elegans*, *Arabidopsis thaliana* et *Drosophila melanogaster*), c'est l'effet inverse qui a été observé: une corrélation négative entre taille de la protéine et biais (Duret et Mouchiroud, 1999; Moriyama et Powell, 1998). Pour expliquer cette différence, plusieurs raisons ont été avancées, dont notamment l'influence du niveau d'expression, qui pourrait masquer l'effet de la taille (Coghlan et Wolfe, 2000). Pour éliminer cet effet parasite, l'utilisation d'un ensemble de gènes dont le niveau d'expression est équivalent est donc nécessaire. Dans ce cadre, la famille des gènes de protéines ribosomales présente plusieurs intérêts: elle est composée de 80 gènes indépendants, de taille variable et dont le niveau d'expression est comparable. Moriyama et Powell (1998) ont utilisé ces propriétés pour tester la corrélation entre taille des séquences codantes et biais chez *E. coli*, *S. cerevisiae* et *D. melanogaster*. Ils ont trouvé une corrélation positive et significative chez la bactérie et la levure, mais chez la drosophile, l'analyse a porté sur un trop petit nombre de gènes (25) pour être significative. Une corrélation positive chez les organismes pluricellulaires reste donc à établir. Dans ce cadre, l'article présenté maintenant expose les résultats obtenus sur la quasi-totalité des gènes de protéines ribosomales chez deux insectes, *S. frugiperda* et *D. melanogaster*.

III- RÉSULTATS

1. Résumé en français de l'article

"Analysis of the ribosomal protein gene family from *Spodoptera frugiperda* (Lepidoptera) reveals insect-specific features and unexpectedly low codon usage bias"

Dans cet article, nous présentons l'ensemble des gènes de protéines ribosomales d'insectes à partir de données provenant d'une banque d'ESTs de *Spodoptera frugiperda* complétées par des informations issues de *D. melanogaster*, *B. mori* et quelques autres insectes. L'analyse du biais de codons montre que les séquences codantes des gènes de protéines ribosomales sont en moyenne beaucoup moins biaisées chez les Lépidoptères que chez les autres organismes. L'analyse d'autres gènes donne des résultats similaires, ce qui suggère que ce faible biais est une caractéristique générale chez les Lépidoptères. L'étude de la corrélation entre taille des séquences codantes de protéine ribosomale et biais de codon montre également des résultats contrastés, puisque, à la différence de la drosophile et d'autres organismes multicellulaires, un corrélation faible mais significative a pu être établie chez *Spodoptera*. Une comparaison systématique de chaque séquence de *Spodoptera* avec d'autres espèces a ensuite été effectuée. Elle a permis de montrer l'existence d'une insertion de 40 et 10 acides aminés dans les protéines ribosomales L35A et L36, qui est dans les deux cas restreinte aux insectes étudiés. L'analyse de ces insertions suggère qu'elles n'ont pas une forte valeur sélective, et qu'elles pourraient constituer des marqueurs phylogénétiques intéressants. Nous avons enfin confronté la signature de chaque protéine ribosomale de *Spodoptera frugiperda* avec celle proposée par Prosite. Pour 9 d'entre elles, nous avons trouvé une différence restreinte soit aux insectes, soit aux lépidoptères testés. Finalement, tous ces résultats suggèrent que la famille des protéines ribosomales est un bon outil pour explorer les particularités de la classe des insectes et de l'ordre des Lépidoptères.

2. Article 2

Analysis of the ribosomal protein genes family from *Spodoptera frugiperda* (Lepidoptera) reveals insect-specific features and unexpectedly low codon usage bias

Landais I.[1], Ogliastro M.[1], Mita K.[2], Nohata J.[2], Lopez-Ferber M.[1], Duonor-Cerutti M.[1], Shimada T.[3], Fournier P.[1] and Devauchelle G.[1]

1. Laboratoire de Pathologie Comparée, INRA/CNRS/Université Montpellier II, UMR 5087, 30380 Saint Christol les Alès, France.

2. Laboratory of Insect Genome, National Institute of Agrobiological Sciences, Tsukuba, Ibaraki 305-8634, Japan.

3. Laboratory of Insect Genetics and Bioscience, Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences,
The University of Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113-8657, Japan.

Corresponding author : Gérard DEVAUCHELLE

Tel : +33 4 66 78 37 03, Fax : +33 4 66 52 46 99, email : devauche@ales.inra.fr

Running title : insect ribosomal proteins

Keywords : EST library, housekeeping gene, insect phylogeny, L35A, L36, protein signature, *Sf*9.

ABSTRACT

In this report, we present the presumably complete set of insect ribosomal protein (rp) genes using data from an EST library of *Spodoptera frugiperda* (Lepidoptera) complemented with informations from other Insect species, especially *D. melanogaster* and *B. mori*. Analysis of codon usage bias showed that rp coding sequences (cds) are in average surprisingly much less biased in Lepidoptera compared to other organisms studied so far. We found similar result using other genes, suggesting that it could be a general feature in Lepidoptera. Study of the correlation between rp cds length and level of codon usage bias also displayed contrasted results since, unlike *Drosphila* and other multicellular organisms, a weak but significant positive correlation was detected in *Spodoptera*. We next performed a systematic comparative study of each *Spodoptera* RP sequence, allowing us to detect in L35A and L36 an insertion of 40 and 10 aa, respectively, which were both restricted to the insect species tested. Analysis of these insertions suggested that they probably do not have a strong selective value and may be good phylogenetic markers for Lepidoptera. For each RP, we finally systematically confronted the *Spodoptera* signature to the canonical signature proposed by Prosite, or we suggested new RP signatures when they were missing. We found discrepancies for 9 RPs, which appeared to be either shared by all the insect species tested or restricted to Lepidoptera. These data collectively suggest that the well-conserved RP family is a good comparative tool to study the Insecta class and the Lepidoptera order.

The rp cDNA sequences described in this report have been submitted to Genbank under the accession nos. AF391091-AF391092, AF395586-AF395587, AF400182-AF400225 and AF429973.

The following persons have especially contributed to this work : F. Sehnal for providing the G. mellonella EST sequences, G. Cerutti and M. Bouton for bioinformatic design.

INTRODUCTION

The Lepidoptera order, with about 150000 described species, is one of the 4 major orders among insects in term of number of species. In addition, from an agronomical point of view, Lepidoptera are major pests of cultivated plants (pre-and post-harvest), since about 1/5 of the total agricultural production in the world is estimated to be damaged by Lepidoptera species. Paradoxically, contrasting with their considerable economical importance, Lepidoptera species have not been extensively studied, and more particularly little is known about the specific features of Lepidoptera compared to other Insecta. Then, in order to promote collaborative researches on Lepidoptera species, an international consortium on Lepidoptera genomics was launched in 2001 (International Lepidopteran Genome Project, at http://www.ab.a.u-tokyo.ac.jp/lep-genome/) with the aim 1) to push efforts towards the sequencing of the entire genome of the silkworm *B. mori*, and possibly of a Lepidoptera pest in a second step, 2) to increase the amounts of gene expression data by EST sequencing into various Lepidoptera models including *Spodoptera frugiperda*, *Mamestra brassicae*, *Helicoverpa armigera*, *Heliothis virescens*, *Galleria mellonella*, *Manduca sexta*, *Plutella xylostella*, *Pieris rapae* and *Antheraea assama*, and 3) to increase the exchange of informations between labs by creating an international Lepidoptera database.

In this frame, our lab is contributing by establishing a genomic resource node on *Spodoptera frugiperda*, a major pest in South America (Ching'oma and Pitre, 1999) and furthermore closely related to other species causing damages in North and Central Africa, India and in Glasshouses in Europe (Anderson et al., 2001).

Besides this agronomical aspect, *S. frugiperda* also displays a special interest in biotechnologies since it is the source of the *Sf*9 cell line (Vaughn et al., 1977), widely used for the production of recombinant proteins in the baculovirus expression system (Luckow, 1991).

Then, as a first genomic approach to study this important species, we decided to construct a cDNA library from *Sf*9 cultured cells, from which 6111 ESTs were sequenced. Global analysis of 5245 fully annotated ESTs showed that the most abundant family of ESTs belonged to the ribosomal protein (RP) genes family, allowing us to characterize the almost complete set of rp cds *of S. frugiperda*.

The RP family contains about 55 proteins in prokaryotes (Ramakrishnan and White, 1998) and 80 in eukaryotes (Doudna and Rath, 2002) which are generally well conserved, small and highly basic (Warner, 1999). Each one is present once in the ribosome, except two which are present

twice. The main function of RPs is structural, essentially to maintain the correct folding and three-dimensional structure of rRNAs (Noller et al., 1992).

RPs constitute a set of highly expressed proteins, detected both at mRNA level (they represent 18 of the 30 most abundant mRNAs in yeast; Velculescu et al., 1997) and at protein level (they constitute 5-10% of cellular proteins in mammalian cells; Kenmochi et al., 1998). Each of them is necessary for the building of the ribosome, the lack or the insufficient synthesis of a single RP causing the loss or a reduced production of ribosome. These troubles lead to the embryonic death or developmental disorders, well characterized in *Drosophila* by the minute phenotypes (for a review, see Warner and Nierras, 1998) and suspected in Human to be responsible for several genetic disorders (Kenmochi et al., 1998). Interestingly, an increasing number of studies have shown that, in addition to their role into the ribosome, numerous RPs have extraribosomal functions (Wool, 1996), opening new ways to investigate RPs and their probable involvement into numerous pathological disorders. Another medical interest of RPs comes from the fact that the ribosome is a target of many important antibiotics (Ramakrishnan, 2002).

However, despite the importance of the RP family, only a few number of eukaryote species have their entire set of rp genes described, *e.g.* yeast (Mager et al., 1997), human (Uechi et al., 2001) and arabidopsis (Barakat et al., 2001).

In this paper, we present the presumably complete set of insect rp genes using data from the Lepidoptera *S. frugiperda* complemented with information from other Insect species, especially *D. melanogaster* and *B. mori*. This materials allowed us to perform 1) an analysis of the codon usage bias occuring in *Spodoptera* and *Drosophila* rp genes, taking advantage of the fact that we disposed of the almost complete set in both species, and 2) a comparative sequence analysis either between insects and other organisms or, at lower level, between Lepidoptera and other insects. We especially emphasized both studies on features which appeared to be restricted either to Insecta or Lepidoptera, in order to test whether such a highly expressed, conserved family of genes could be a good tool to insvestigate the features specific for the insect or the Lepidoptera genome.

RESULTS

Global analysis of the EST library and of the rp genes family in *S. frugiperda*

5245 ESTs have been sequenced and fully annotated from the *Sf*9 cDNA library, allowing us to identify 1687 independant identified or putative genes (table 1). We found a great proportion of underrepresented ones (17% of the ESTs, corresponding to 45% of the genes) that belong to the

"low homology" and "no hits found" families. Surprisingly, these ORFs did not display homologies with identified or putative genes from other organisms, even among the hypothetical proteins described through the entire genome of *D. melanogaster*. Conversely, as expected in such a proliferative cell line, 3 major families of housekeeping genes (ribosomal, mitochondrial and proteasome families) were overrepresented : they make up 55% of the ESTs but correspond to only 13.4% of the genes.

One-third of the total number of ESTs belonged to the rp genes family and represented 76 independant rp genes. Moreover, the high redundancy of ESTs in this family (average redundancy : 21) allowed us to reconstitute and accurate 67 complete cDNA sequences, which were submitted to Genbank (table 2). We failed to obtain the 9 complete remaining sequences, which are quoted as fragments in the same table.

By comparison with the Swiss-Prot nomenclature, 3 additional RPs which were absent from our library (L4, SA and S9) are supposed to be present in *S. frugiperda*, because they have been characterized in a wide set of eukaryotic species including at least one insect (*i.e. D. melanogaster*). We then estimated that the rp family comprises 79 genes in *S. frugiperda*, like in *Drosophila* and Human (Uechi et al., 2001) and one more than in yeast (Mager et al., 1997).

Since almost all of the rp genes were represented by several ESTs in the library, we performed a systematic alignment of these sequences in order to search for the existence of allelic variations or isoforms. We did not detected such events, consistently with the situation reported in Human where only one gene encodes each RP (Dudov and Perry, 1984).

Expression of rp genes in *Sf*9 cells

We aimed to perform a codon usage bias analysis with taking advantage of the assumed equimolar ratio of expression of each RP in the cell (see next section). However, while this feature has been early experimentally verified in *E. coli* (Howe and Hershey, 1982), an increasing number of works have pointed out the overexpression of several RPs in various conditions, including normal tissues or cell growth conditions (Bortoluzzi et al., 2001). Consequently, a preliminary step for our codon usage bias study was to check the expression level of these RPs. We first examined the number of clones belonging to each rp gene in our EST database, postulating that the number of clones of a given transcript was roughly representative of the level of transcriptional expression of the corresponding gene.

This analysis allowed us to detect important variations in the number of ESTs belonging to each rp gene, from 1 to 152 (table 2), suggesting their unequal expression in the *Sf*9 cell line.

However, such a result needed to be verified by an independant and more direct quantification of mRNAs, since the library construction potentially introduces strong bias in the gene representativity. We then performed a Northern blot experiment in order to quantify the mRNA level of 4 RPs displaying various number of ESTs.

Total RNAs from the *Sf*9 cell line (*i.e.* the same material from which the cDNA library was established) was hybridized with 4 RP probes corresponding to L23, L37A, S23 and S2 (represented by 152, 51, 19 and 4 clones, respectively) and each hybridization band was counted.

The result (table 3) showed variations that did not exceeded twofold between the 4 RP mRNA levels, in contradiction with the great variations observed in the EST library (38-fold variation between L23 and S2). In addition, the relative level of expression between RPs given by the two experiments was different : apart from L23, which was the most represented RP in the EST library with the strongest hybridization signal on the Northern blot, the other RPs did not displayed the same rank in the two estimation techniques: L37A was represented by 51 clones in the library with the weakest hybridization signal, while S2 had a stronger hybridization signal but was represented only by 4 clones.

In conclusion, despite the small number of RPs tested, this experiment strongly suggested that an important cloning bias occured in our library and consequently that the EST redundancy probably did not reflect, even roughly, the level of transcriptional expression of the corresponding gene. This result also suggested that RP mRNAs are expressed at roughly equivalent levels in the *Sf*9 cell line, strengthening the hypothesis of equimolar expression of each member of the RP family.

RP codon usage bias

Eyre-Walker (1996) and Moriyama and Powell (1998) used the assumed one-to-one ratio of RPs to avoid the influence of gene expression level on codon usage bias and then to study more precisely the effect of cds length, both in prokaryotes and eukaryotes. They concluded that, for RPs, codon usage bias is positively correlated to cds length in unicellular organisms (*E. coli* and *S. cerevisiae*). However, in *D. melanogaster*, the correlation was positive but not significant ($r : -0.13$, $p>0.5$). The authors suggested that it may be due to the small sample size, since only 25 RP sequences were available to date for this species.

Then, as we disposed of 67 complete RP sequences from *S. frugiperda*, we aimed to investigate more extensively the effect of cds length on codon usage bias in an eukaryotic, multicellular

species. Furthermore, as all the *D. melanogaster* RP coding sequences were available from GADFly (http://www.fruitfly.org/annot/), it seemed interesting to compare data belonging to these two insect species.

We used the "Effective Number of Codons" (Wright, 1990) to measure the bias level, since this method is independent of a reference table and displays similar results compared to other methods like CAI (Sharp and Li, 1987; Moriyama and Powell, 1998). ENC values vary from 20 (for maximally biased gene) to 61 (no bias).

We first calculated the average %GC of rp cds, which appeared to be stronger than the GC content of coding sequences in both *S. frugiperda* (calculated from 100 complete cds belonging to our library) and *D. melanogaster* (52.2%/47.8% and 57.4%/53%, respectively). This analysis also showed that the GC content was higher in *D. melanogaster* than in *S. frugiperda*. Further analysis on the synonymous sites of rp cds showed that, like in *Drosophila*, G and especially C are favored at this position in *Spodoptera*. This preference was however weaker in the latter species (table 4). We next calculated the ENC value for 64 *S. frugiperda* rp cds, and we found that the average ENC was surprisingly high (49.1) for such a highly expressed family of genes (Powell and Moriyama, 1997). By comparison, average ENC for rp cds in *D. Melanogaster*, *S. cerevisiae* and *E. coli* were 35.0, 28.8 and 35.3, respectively (Moriyama and Powell, 1998). However, as the *Drosophila* value was calculated from only 25 RP sequences in this work, we recalculated the average ENC using the entire set of rp cds of this species, and we obtained a similar value (37.1). In order to test whether the result found in *S. frugiperda* was specific for this species, we calculated the average ENC value from 20 RPs in another Lepidoptera, *B. mori*, and in *C. elegans* (Nematoda). We then compared the results with the values obtained from the same 20 RPs in *D. melanogaster* and *S. frugiperda*. Since we found very similar results in *S. frugiperda* and *D. melanogaster* between ENC values calculated from these 20 RPs or from the entire set of RPs (49.2/49.1 and 36.2/37.1, respectively), we assumed that the average ENC values calculated for 20 *B. mori* and *C. elegans* rp cds were also representative. The result was clear, with an average ENC value even higher in *B. mori* than in *S. frugiperda* (57.4 and 49.2, respectively) compared *to D. melanogaster* (36.2) and *C. elegans* (35.7), suggesting that rp cds are unusually unbiased in Lepidotera species.

In order to verify whether this low codon usage bias was restricted to the rp genes family, we next compared the ENC values for 23 other homologous genes cds in *S. frugiperda* and *D. melanogaster* (table 5). In average, the codon usage bias was significantly stronger *in D. melanogaster* (42.7) than in *S. frugiperda* (50.3). Furthermore, calculation of the average ENC for the 11 genes cds displaying the strongest bias in *D. melanogaster* (*i.e.* ENC <40, average

value : 34.8) showed that *Spodoptera* genes cds displayed an almost unchanged average bias (48.7).

This results suggested that the low codon usage bias is not restricted to the rp gene family, but is probably a general feature in *Spodoptera*.

We next calculated the correlation coefficient between rp cds length and bias. Three conditions were tested, with all 64 *S. frugiperda* rp cds or excluding those <60 (figure 1A) or <100 aa, as recommended by Moriyama and Powell (1998). In all cases, we found a weak but significant positive correlation between cds length and codon usage bias (r : -0.28, $p<0.03$; r : -0.34, $p<0.01$; r : -0.27, $p<0.04$, respectively), consistent with previous results in monocellular organisms (Eyre-Walker, 1996; Coghlan et Wolfe, 2000). However, using 75 RP sequences from *D. melanogaster*, the same analysis displayed an opposite result : a weak, significant negative correlation (r : 0.30, $p<0.05$) when excluding RPs < 100aa (figure 1B), not significant when including RPs <100aa (r : 0.20, $p<0.1$). Moreover, when we discarded S2 from both analysis (this rp gene is much longer than others and not biased), we did not found any correlation between codon usage bias and gene length (r : 0.10 and 0.01, respectively).

RP signatures

We aligned each of the 67 complete *S. frugiperda* RP sequences with those of other eukaryotes and prokaryotes species. In most cases, no peculiarity was noticed when analyzing these alignments, since RPs are generally well conseved. However, for 2 RPs, the alignment revealed a feature which seemed to be specific for the Insecta : in the L36 sequence, we found a 10 aa insertion close to the N terminus (figure 2), while in L35A the N terminus displayed a stretch of 50 additional aa compared to other sequences. These 2 “insertions” displayed neither homologies with known sequences nor identified motifs.

In the absence of sequences corresponding to other classes of arthropods and close embranchments, it was not possible to determine when these insertions appeared in the evolution. We then performed a phylogenetic analysis restricted to the only 5 Insecta sequences available for each RP. At the aa level, both insertion sequences displayed the same pattern : they were highly similar among Lepidoptera but strongly diverged from Diptera (figure 2). A strong divergence was also detected between the 2 Diptera species. By contrast, the flanking sequences seemed to be more conserved. In order to investigate more precisely this point, we next calculated neighbor-joining trees using either the entire sequences deleted for the insertion (figure 3A, 3C) or the insertions only (figure 3B, 3D). We used aa sequences for L35A, as

preliminary works indicated that L35A was saturating at nucleotide level (data not shown). By contrast, we used nucleotide sequences for L36 since this RP was well conserved at aa level. In this analysis, both insertions displayed similar features : 1) Lepidoptera sequences are highly similar but strongly diverge from the Diptera ones (figures 3B, 3D), and 2) both insertions have a higher substitution rate than the rest of the sequence, as shown by the strong stretch of the branch connecting Lepidoptera and Diptera (compare figure 3A and 3B, 3C and 3D). This suggests that the insertions are less constraint than the rest of the RP sequence. In L35A for example, only the tree calculated from the N terminal extension correctly group the Lepidoptera species (the Noctuoidea *S.frugiperda* and *T.ni* apart from the Bombycoidea *B. mori*), with a good robustness (bootstap score 986), the rest of the sequence being probably too conserved to be a good marker at this phylogenetic level. By contrast, the tree calculated from the L36 insertion sequence displayed the same branchments and a correct robustness as the trees calculated from the entire sequence, despites the fact that the insertion is very short (30 bp). Interestingly, these L36 trees group with high bootstrap scores the Bombycoidea *B. mori* and the Pyraloidea *G. mellonella* apart from the Noctuoidea *S. frugiperda*, in contrast with the classically admitted groupment of these species : Bombycoidea and Noctuoidea superfamilies belong to the Macrolepidotptera, a subset of Ditrysia which does not include the Pyraloidea (Minet, 1991).

We next performed a systematic comparative analysis for each RP signature. As a reference, we used the signatures proposed by the Prosite documentation of the Swissprot database. Among the 67 *S. frugiperda* RP sequences, 49 displayed the canonical signature. In a second set of 9 RPs, no signature was proposed by Prosite. Table 6 shows suggested signatures for these RPs, determined from the alignment of the complete set of sequences present in databases. All these signatures were tested using PHI-BLAST for their ability to display only RPs in the BLAST result. They appeared to be highly specific, with pattern significance at least <1e-10.

Into the last set of 9 RP, comparison between the Prosite signature and the *S. frugiperda* RP sequence revealed one or more aa discrepancies. In order to confirm this result, we added in the analysis more Lepidoptera and other Insect species, which allowed to form 4 distinct groups (table 7 A, B, C, D), wherein discrepancies appeared to be (A) "insect-specific", *i.e.* located at the same residue in all the insect sequences tested (RP L10A, L24, S24), (B) "Lepidoptera-specific", *i.e.* present in Lepidoptera but not in Diptera species (L13, L31, L36), (C) present in all the Lepidoptera but not in all the Diptera species tested (RP S20, S21) and (D) numerous in all the sequences tested (RPS30). One can note that, in this last case, the signature is not located into the RP "sensu stricto" sequence (which is well conserved between organisms) but in the

fused N-terminal ubiquitin-like sequence (Fau protein), which is not well conserved (Michiels et al., 1993).

DISCUSSION

Global analysis of the *Sf*9 library

The construction of a cDNA library from the *Sf*9 cell line and the subsequent sequencing of 6000 ESTs were initially performed to better characterize the housekeeping genes of *S. frugiperda*. As expected, we found a high proportion of ESTs belonging to this category, which allowed us to identify almost all the rp genes in this Lepidoptera species. Comparison with the rp gene family of *D. melanogaster* and *H. sapiens* confirmed that insects and human possess the same number of rp genes (79). Furthermore, for each of them, absence of sequence variations between ESTs suggested that RPs are encoded by a single gene in *Spodoptera*, like in Human (Dudov and Perry, 1984). This hypothesis is strengthened by the fact that Adams *et al*. (2001) found in the *Drosophila* genome only 128 predicted ORFs corresponding to rp genes, including mitochondrial (mt) ones. If we assume that the total number of distinct rp genes encoded by the nuclear genome is about 150 in eukaryotes (~80 for the ribosome; Uechi *et al*., 2001, and 70 for the mitoribosome; Suzuki *et al*., 2001), it is more likely that each rp gene is represented by a unique copy in the *Drosophila* genome. Moreover, from the same data, it is probable that about 20 mt rp genes have not yet been identified. By contrast with the situation observed in these species, in yeast most of the RPs are encoded by two genes yielding nearly identical proteins, but this situation probably results from an ancient duplication of the yeast genome (Wolfe and Shields, 1997).

The EST library also displayed a high number of weakly represented, putative genes which did not display any homology with neither known nor hypothetical protein genes of other organisms, even in *D. melanogaster* which genome has been entirely sequenced. The presence of a high number of unknown genes in our EST library is similar to the situation observed in *B. mori* EST databases (Shimada, personal communication, and SilkBase, data not shown), suggesting that a significant proportion of putative genes characterized from the *S. frugiperda* and *B. mori* EST libraries may be specific to Lepidoptera. This hypothesis has been recently strengthened by the finding of several homologs in the two species (data not shown). Further in-depth EST sequencing using more Lepidoptera species will be needed to ascertain the specificity of these genes.

rp gene expression level

Except two, each RP is present once in the ribosome, as demonstrated by biochemical and structural studies (Ramakrishnan and White, 1998). Consequently, it has been assumed that each RP is expressed at equimolar ratio in the cell. This hypothesis has been experimentally verified early by Howe and Hershey (1982) in *E. coli* and some works extrapolated this experimental result not only in prokaryotes but also in eukaryotes, especially for codon bias analysis (Eyre-Walker, 1996; Moriyama and Powell, 1998). However, a great number of reports have highlighted the overexpression of several RPs in precise conditions, *e.g*. cancer or specific tissues (Pogue-Geile et al., 1991; Wang et al., 2001; Yang and Sehnal, 1998). These findings have been correlated with the discovery of an increasing number of extraribosomal functions for RPs (Wool, 1996). In this way, in the absence of exhaustive studies of direct quantification of RPs in eukaryotes, it is possible that the level of expression of all individual RPs is not equal, even in normal tissue or cell growth conditions (Bortoluzzi et al., 2001).

In order to test this hypothesis, we examined the number of clones belonging to each RP in our EST database, postulating that 1) mRNA level correlates well with protein concentration, as previously shown in yeast (Futcher et al., 1999) and 2) in a cDNA library, the number of clones of a given transcript is roughly representative of the level of transcriptional expression of the corresponding gene. Other authors used these postulates to estimate gene expression from the number of corresponding clones in cDNA libraries (Bortoluzzi et al., 2001; Duret and Mouchiroud, 1999; Moriyama and Powell, 1998; Okano et al., 2001).

We found great variations in the number of ESTs corresponding to each rp gene, suggesting that they were not expressed at similar levels in the *Sf*9 cell line. In order to verify this feature, we chose 4 RPs for which the number of ESTs was very different, and we estimated their mRNA level by Northern blot, a technique classically performed when an accurate estimation of mRNA concentration is needed (Velculescu et al., 1997; Zhang et al., 1997; Galitski et al., 1999).

This experiment showed limited variations between mRNA levels for the 4 rp genes tested, in contradiction with the EST redundancy results. This suggested that 1) our library is biased and not reliable to estimate rp gene expression, and 2) rp genes are expressed at comparable levels in the *Sf*9 cell line.

The fact that the EST library is probably biased is not surprising, considering that the high number of steps involved in the EST library construction (from the initial RNA extraction to the final plamid library sequencing) facilitates the occurence of bias. However, these conclusions should be modulated, since only a few number of genes were tested. We cannot exclude that the

EST library may be relevant to estimate expression level in other EST categories. DNA microarray experiments using materials and data from the *Sf*9 cDNA and EST libraries are currently in preparation and should give us a larger and more precise quantification of gene expression in this cell line.

We then conclude that the use of quantitative data from EST libraries to estimate the expression level of mRNAs or proteins should be preceded by experiments testing the lack of bias, *i.e.* by establishing a correct correlation between the number of EST clones and the "real" level of gene expression using dedicated techniques like SAGE or DNA microarrays (Velculescu et al., 1997; Wodicka et al., 1997; see also the next paragraph).

average codon usage bias in rps

Compositional bias in genomic DNA has been first investigated by Sueoka (1962) in prokaryotes and concerned the G+C content of bacterial genomes. Since then, with the gain of exponential increase of sequences available in databases, numerous studies have been published, showing new DNA biases and suggesting different models to explain them. Actually, a wide consensus has progressively emerged upon the fact that translational selection occurs at silent sites (Akashi and Eyre-Walker, 1998; Dunn et al., 2001). Numerous reports have shown that synonymous codons which bind the most common tRNAs are preferentially selected by natural selection (Akashi, 2001; Moriyama and Powell, 1997). Consequently, the codon usage bias is thought to maximize the rate of protein synthesis and then to improve the fitness of the organism. Further analysis showed that important variations occur in the codon usage bias level between genes cds: some of them are not biased (*i.e.* all the codons are equally used for each aa) while others are extremely biased. In order to explain these differences, numerous works focused on the potential correlations between codon usage bias and diverse parameters : gene expression level, gene length, recombination rate (Akashi et al., 1998; Comeron et al., 1999), synonymous substitution rate (Dunn et al., 2001; Moriyama and Gojobori, 1992). These studies showed that the level of codon usage bias is strongly correlated with the level of gene expression in all the species tested (Duret and Mouchiroud, 1999; Gouy and Gautier, 1982). This is consistent with the hypothesis that stronger selection on highly expressed genes maximizes translational efficiency by the use of "major" codons (Moriyama and Powell, 1998). However, in two Lepidoptera species, we found that rp genes, which are among the most highly expressed genes in the eukaryotic cell (Bortoluzzi et al., 2001; Velculescu et al., 1997), displayed in average a very low codon usage bias. This result is puzzling, since in all the

organisms studied so far, like *E. coli*, *S. cerevisiae* and *D. melanogaster* (Eyre-Walker, 1996; Moriyama and Powell, 1998), and in *C. elegans* (this work), rp genes cds were shown to be extremely biased (*i.e.* ENC < 35; Powell and Moriyama, 1997). Further analysis on synonymous sites showed the same favored nucleotides (G and especially C) in *Spodoptera* and *Drosophila* rp genes, which seems to be a general feature in this last species (Powell and Moriyama, 1997).

In addition, when we calculated in *S. frugiperda* the ENC value of several other genes which homologs in *D. melanogaster* are strongly biased, we found low bias values as well. It is likely that, in Lepidoptera, genes that are highly expressed or having highly biased homologs in other organisms remain weakly biased. The more convincing hypothesis to explain this surprising feature in Lepidoptera could be an equivalent ratio (or slight differences) between tRNAs corresponding to the same aa, since such a situation should strongly reduce the selection between codons (Moriyama and Powell, 1997; Shields et al., 1988). However, this hypothesis currently cannot be verified since few informations are available on the tRNA composition in Lepidoptera. The only published data concern the $tRNA_1$gly and tRNA ala species, which are overexpressed during the 5^{th} larval instar in *B. mori*, in order to respond to the overexpression of the fibroin, especially rich in glycine and alanine residues (Garel, 1976). However, these works focused on a single tRNA species for both aa (Sprague et al., 1977; Srinivasan and Gopinathan, 2001) and did not displayed the relative concentrations of isoacceptors.

As an alternative hypothesis, Duret and Mouchiroud (1999) suggested that low levels of codon usage bias could be related to small population size, since random drift may overcome fitness differences among synonymous codons in small populations. However, in *S. frugiperda* and *B. mori*, population sizes are probably sufficiently important to overcome this effect.

Another hypothesis could be a lower recombination rate in Lepidoptera genomes compared to other organisms, since codon bias has been shown to be lower in regions of low recombination (Comeron et al., 1999; Kliman and Hey, 1994) as a consequence of the "Hill-Robertson" effect (Hill and Robertson, 1966). In contradiction with this hypothesis, Yasukochi (1998) estimated from genetic maps that crossing-over rates are higher in *B. mori* than in *D. melanogaster*. However, this indirect estimation need to be crosschecked by additional studies on recombination rates in Lepidoptera.

codon usage bias and cds length

Extensive studies attempting to correlate cds length and codon usage bias showed that the more convincing model to explain this effect is the translational accuracy model (Akashi, 1994;

Comeron and Kreitman, 1998) : selection to reduce translational misincorporations should be higher in longer cds because the cost of producing dysfunctional peptides will be proportional to their length. However, these studies displayed contrasted results : a significant positive correlation was found in monocellular organisms, *i.e*. prokaryotes and yeast (Coghlan and Wolfe, 2000; Eyre-Walker, 1996) but not in multicellular ones (Duret and Mouchiroud, 1999; Moriyama and Powell, 1998). Different mechanisms have been invoked to explain that result (for review, see Akashi, 2001), one of them being that the strong influence of gene expression level on codon usage bias could mask the effect of cds length (Coghlan and Wolfe, 2000).

To reduce this interference, an approach consisted in keeping constant the gene expression level in the analysis. In this context, the RP family displayed 3 interesting features: 1) as discussed before, each RP is present at equimolar ratio into the ribosome and we assume that they are approximately equally expressed in the cell, 2) it is a large family of about 75-80 proteins with sizes ranging from 25 to about 900 aa, and 3) since they are highly expressed, most of them have been characterized in various organisms.

The use of the almost complete set of *S. frugiperda* rp cds allowed us to show for the first time a positive, weak but significant correlation between cds length and codon usage bias in a multicellular organism. This result is convenient with the predictions of the translational efficiency model previously verified in unicellular organisms (Coghlan and Wolfe, 2000; Eyre-Walker, 1996). Moriyama and Powell (1998) also found a positive correlation between codon bias and cds length using 25 *D. melanogaster* rps, but the result was not significant. Authors suggested that it could be due to the small sample size. We performed a similar analysis using the entire set of *D. melanogaster* rp genes, and surprisingly we found an opposite result : codon usage bias was weakly negatively correlated to cds length when using all the set of rp genes, not correlated when removing S2, an rp gene which is much longer and less biased than the others. Interestingly, this result corresponded to the general tendancy observed from global bias analysis in all the multicellular organisms studied so far, *i.e. D. melanogaster*, *C. elegans* and *A. thaliana* (Duret and Mouchiroud, 1999; Moriyama and Powell, 1998) in which longer genes cds statistically display lower codon usage bias. In order to explain the opposite results found in unicellular (positive correlation between codon bias and gene length) and multicellular organisms (negative correlation), these authors suggested that they probably display different evolutionary constraints. However, the fact that in *S.frugiperda*, the correlation between codon usage bias and cds length is positive like in unicellular organisms, in contrast with results from another insect, raises new interrogations about the influence of cds length on codon usage bias in multicellular organisms : First, as discussed by Coghlan and Wolfe (2000), it is possible that

the results previously obtained in multicellular organisms were biased by an inadequate estimation of gene expression level. This situation already occured in *S. cerevisiae*, for which the publication of several independant mRNA expression data (Velculescu et al., 1997; Wodicka et al., 1997) allowed Coghlan and Wolfe (2000) to demonstrate the global positive correlation between codon usage bias and gene length in this species. They used sets of genes displaying "objective" equivalent mRNA levels, while in previous studies gene expression levels were estimated from the number of EST clones in databases, that is probably not a reliable estimation method, as discussed above. Secondly, it remains that the opposite results observed in our work between *S. frugiperda* and *D. melanogaster* cannot be explained by such hypothesis, since we used the same set of equally expressed genes in the two species.

Actually, the discrepancy may be simply related to the strong difference in average bias of rp genes cds between *D. melanogaster* and *S. frugiperda* : in *Drosophila*, since most of the rp cds are extremely biased due to their high level of expression, it is possible that the additional effect of cds length is too weak to be detected. In other words, in this species, the rp genes family may not be appropriate to study the correlation between codon usage bias and cds length, due to the extreme bias displayed by its members. By contrast, *in S. frugiperda*, since the bias observed in rp cds is globally low, the weak effect of cds length could not be overcome. In this hypothesis, the positive correlation detected in *S. frugiperda* rp genes may reflect a general tendancy of codon usage bias to be stronger in longer cds, fitting with the translational accuracy model, not only in unicellular but also in multicellular organisms. However, as shown by the weak correlation coefficient values observed in this study, the effect of cds length is probably very faint and may be detectable only when other parameters acting on codon usage bias (mainly the gene expression level) are kept constant. More independant, reliable and wide gene expression data (SAGE, DNA microarrays) from multicellular organisms are needed to test with a better reliability the effect of cds length on codon usage bias.

L35A and L36 insertions

Initially, we calculated the trees corresponding to the L35A and L36 rp sequences with the aim to "vizualize" the substitution rates of the insertions compared to the rest of the sequence. These two rp genes appeared to be informative at different level (aa for L35A, nucleotide for L36), but we found that in both cases, substitution rates were significantly higher in the insertions than in the rest of the sequence. Then, as the L35A and L36 insertions 1) seem to be restricted to a few number of phyla, perhaps only to Insecta, 2) have evolved faster than the rest of the RP

sequence, with an aa sequence that is totally different between Diptera and Lepidoptera, 3) take place in RPs which are quite well phylogenetically conserved, they may have a weaker selective value. The fact that in both cases insertions are colinear in all the insect sequences tested (same position, same length) suggests that they have evolved from a common insect ancestor rather than from separate insertion events in Diptera and Lepidoptera. The strong divergence observed not only between Lepidoptera and Diptera but also between the two Diptera species (contrasting with the high conservation among Lepidoptera) is convenient with the known phylogenetic relationships of these species : the Brachycera group (from which belong *Drosophila*) probably splitted from the lower Diptera (from which belong *Anopheles*) during the upper triassic (208-245 million years ago; Yeates and Wiegemann, 1999), perhaps only 50 million years after the existence of the dipteran stem lineage (250 million years) and the split with the lepidopteran lineage (Pashley and Ke, 1992; Regier et al., 1998; Yeates and Wiegemann, 1999). It is then likely that the split between Brachycera and lower Diptera is almost as ancient as between Diptera and Lepidoptera, which may explain the strong divergence of the insertion sequences between Diptera species. By contrast, in the Lepidoptera order, superfamilies of the clade Ditrysia arose much more recently, in the late cretaceous to early tertiary (50-80 million years; Pashley and Ke, 1992; Regier et al.,1998), which is also convenient with the conservation of the insertions observed in Lepidoptera species. It is thus possible that these insertions retain a phylogenetic signal at the Lepidoptera level : For L35A, only the tree calculated from the 5' extension displayed a correct and robust groupment of species, suggesting that the rest of the sequence was too conserved to be resolutive. Interestingly, for L36, both trees grouped with a good robustness the Bombycoidea and the Pyraloidea species apart from the Noctuoidea species. This result does not fit exactly with the morphology-based phylogeny of these species, as Bombycidea and Noctuoidea superfamilies are grouped into a derived subset of Ditrysia, the Macrolepidoptera, from which pyraloids are excluded (Minet, 1991). However, Pyraloidea are thought to be closely related to the Macrolepidoptera (Minet, 1991; Nielsen, 1989; Regier et al., 1998), and moreover the relationships between the ditrysian superfamilies, as a "heretofore intractable problem (Nielsen, 1989) are far from being precisely solved. Crosscheckings with information from other genes resolutive at the superfamily level in Lepidoptera should be useful to validate the phylogenetic interest of these rp genes. However, to date, despites the increasing number of molecular markers used to resolve the Lepidoptera phylogeny, no one is sufficiently informative at the superfamily level : Elongation factor-1a (EF-1a), Phosphoenolpyruvate carboxykinase (PEPCK) and Dopa decarboxylase (DDC) are resolutive at higher levels, whereas Period seems to be adapted to lower levels (for a review, see Regier et al., 1998).

As a concluding remark, we note that, in RPs, such insertions specific for a group of species belonging to the same phylum are not restricted to Insecta: RP L29 displays a C-terminal extension of at least 40 aa which seems to be restricted to mammalians.

RP signatures

We focused our comparative analysis of RP signatures on the class of Insecta. Among 67 *S. frugiperda* RPs tested, 9 carried a signature displaying one or more discrepancies with the canonical signature proposed by Swiss-Prot. Addition of more insect species showed that these discrepancies were not randomly distributed into the signature sequence, but restricted to the same residues among species. Moreover, we found that several of them were present in all the insect species tested while some others were restricted to Lepidoptera. A third categorie showed an intermediate distribution, with discrepancies shared by all the Lepidoptera and one Diptera species (*i.e. A. gambiae*), while the other Diptera (*D. melanogaster*) displayed the classical signature.

As discussed above for the L35A and L36 insertions, more RP sequences belonging to other Insecta and Arthropoda species are needed to mark out the extension of the "Insecta-specific" discrepancies. At lower taxonomic level, "Lepidoptera-specific" discrepancies, which are absent from the Diptera species, are perhaps present in other insect orders, or conversely restricted to only a subset of the Lepidoptera order.

Nevertheless, despite the few number of species tested, this signature analysis highlights the fact that, even among the highly conserved RPs, one can find features which seem to be specific for insects or Lepidoptera. It thus confirms that the ribosomal proteins family is a valuable tool to investigate this class and this order.

More generally, this study pointed out the fact that proposed signatures cannot be considered as representative of a protein, even well-conserved, until a sufficient quantity of species chosen in all the kingdoms are included into the signature design. The new signatures that we propose for 9 RPs in this paper do not escape to this rule, and are susceptible of evolving with the increase of RP sequencing data from species belonging to the great number of embranchments and classes not yet represented in databases.

Summary

In this paper, we performed a comparative analysis of insect rp genes using sequences belonging to our *S. frugiperda* cDNA library, the *D. melanogaster* public database Flybase, the *B. mori*

SilkBase and other insects when available. Both sequence comparisons and codon usage bias analysis revealed unexpected discrepancies, not only between insects and other organisms, but also inside the Insecta class. These differences highlighed the fact that, despite the high evolutive conservation of the rp genes, this family display several specificities which seem to be restricted to the Insecta class or to the Lepidoptera order. Thus, we conclude that the RP family is an interesting tool to study these organisms, and especially the specificities of Lepidoptera compared to other insects.

METHODS

S. frugiperda cDNA library and sequencing

One microgram of polyA+ mRNAs from the *S. frugiperda Sf*9 cell line was used to construct a phage λTriplex2 cDNA library (Clontech), as previously described (Landais et al., 2001). The titer of the unamplified library was 4.10^6 pfu/ml, with 98% recombinant clones, as measured by white/blue screening. After amplification, the library displayed a titer of $2.5.10^{10}$ pfu/ml.

0.1 µl of the amplified λ library was converted to a plasmid library using the cre/lox recombinase system in *E. coli* strain BM25.8 at 31°C, following the recommendations of the manufacturer. 10000 clones were harvested and put in collection, from which 40 were subjected to a PCR-based estimation of insert size. The average insert size was of ±950 bp (data not shown).

To date, 6111 clones have been sequenced using the 5'λ sequencing primer designed by the manufacturer (Clontech), as previously described (Okano et al., 2001). The resulting EST library contains to date 5245 EST and is expected to contain about 8000 EST at the end of the sequencing work.

Annotation of the sequences

Each sequence was systematically compared with other ESTs of the library in order to identify identical or overlapping sequences. This step allowed to cluster the ESTs belonging to the same gene and significantly reduced the number of sequences to be annotated.

ESTs belonging to the same cluster were then aligned in order 1) to accurately determine the nucleotide sequence and 2) to reconstitute (when possible) the full-length cDNA sequence.

EST sequences were automatically subjected to a BLASTX search (Altschul et al., 1997), which allowed to unambiguously identify most of the ESTs belonging to conserved genes but was less reliable for short (*i.e.* <300 bp) and not conserved sequences.

We defined 5 categories of identified ORFs (table 1), *i.e.* displaying at least E values $<10^{-4}$ in the BLAST result. We distributed remaining ORFs (*i.e.* not previously identified by automatic Blast) into 2 additional categories : "low homology" (10^{-1}>E value>10^{-4}), which typically displayed short sequences homologies (*e.g.* motifs), and "no hits found" (E value>10^{-1}) for which the well-characterized ORF did not displayed any homology with sequences available in public databases.

Among the identified EST sequences, those belonging to rp genes were translated into aa sequence using DNA Strider 1.2 and further characterized using the Ribosome Swiss-Prot Protein Database (http://www.expasy.ch/cgi-bin/lists?ribosomp.txt). As nomenclature of rp genes is complex and differs between organisms, we systematically referred to the Swiss-Prot nomenclature, which also corresponds to the human rp nomenclature, except one (L44 : L36A; Kenmochi et al., 1998).

RP signatures

The Prosite Database (http://www.expasy.ch/prosite/) displays for most of the RPs a documentation file (ProDoc) which proposes an amino-acid (aa) signature. We systematically compared this signature with the *S. frugiperda* RP aa sequence. When discrepancies were detected, we verified their existence in a wide variety of organisms, from plants to mammals (accession numbers in table 2). In order to increase the number of Insect species, we searched for rp genes sequences using BlastX in the *B. mori* EST database (SilkBase : http:// www.ab.a.u-tokyo.ac.jp/silkbase/), the *A. gambiae* genomic database (http://www. ensembl.org/Anopheles_gambiae/) and Gadfly at Flybase (http://www.fruitfly.org/annot/). *Galleria mellonella* sequences were kindly provided by F. Senhal from his personal EST database.

When no signature was proposed by Prosite, we suggested one deduced from conserved aa displayed by the alignment of all available sequences present in public databases. The alignments were performed using Multalin (Corpet, 1988) and each signature was tested using PHI-BLAST (Altschul *et al.*, 1997).

Phylogenetic analysis

We used ClustalX (Thompson et al., 1997) to generate alignments of the L35A and L36 sequences. For L36, nucleotide sequences were subsequently manually aligned in order to match

the protein alignment. Bootstrapped, unrooted neighbor-joining trees were calculated using the same program. Trees were drawn by Treeview (PAGE, 1996).

Codon usage bias analysis

Codon bias was measured for 64 *S. frugiperda*, 75 *D. melanogaster* and 20 *B. mori* RP complete coding sequences (cds) using the "Effective number of codons" (ENC) method (Wright, 1990). As recommended by the author, ENC values greater than 61 were revised to 61. For each rp gene, the correlation coefficient r between the ENC value and the cds length (expressed by the number of codons) was calculated and its probability was deduced from the associated table.

Northern blot

Total RNAs were extracted using Trizol reagent (Gibco) from *Sf*9 cells in log phase. 20 μg was electrophoresed and transferred onto a nylon membrane (Hybond N+) by capillary blotting following standard procedures and UV-fixed. Membrane was prehybridized 3 h at 42°C in FPH buffer (50%formamide, 5X Denhardt, 0.1% SDS, 5X SSC, 0.1 mg/ml denatured herring sperm DNA).

325 bp-fragments of RP L23, L37A, S2 and S23 were PCR-amplified from corresponding clones of the plasmid library using oligos L23for (GCATCTCGTTGGGTCTCCCTG) and L23 rev (GCCAGAGATCCGCGCACTCC), L37Afor (GTCAACCGGAGAGTTAGACAGG) and L37Arev (CCTTAAGCTAACGTCAGATG), S2for (CGGAGGTTTTGGTTCCCGTG) and S2rev (GTCTCCGATAGCCACGAAGG), S23 for (GTAGCCGATTGGTGTCGTGC) and S23rev (CTTCGTCGTTCTCTTCAATGTGG). 60 ng of each fragment was labelled by random priming (Megaprime labelling kit, Roche) using α^{32}P dCTP and purified on Chroma-spin 30 columns (Clontech). The four probes ($1.2x10^6$ cpm each) were combined and hybridized with the blot in FPH buffer. After a 16h-hybridization at 42°C, membranes were rinsed twice in 2X SSC, 0.1%SDS, twice in 0.2X SSC, 0.1% SDS and exposed for 16 hours. Scan and quantification of hybridization signal intensities were perfomed using a Storm 820 and ImageQuant (Amersham Pharmacia Biotech). We performed this experiment in duplicate.

ACKNOWLEDGMENTS

We are grateful to F. Sehnal for providing the *G. mellonella* RP sequences and J. Y. Rasplus for helpful discussions about phylogeny. We also thank all the persons who have participated to the

establishment of the *Spodoptera* EST database and the SilkBase. This work was supported by the Centre National de la Recherche Scientifique (CNRS), the Institut National de la Recheche Agronomique (INRA), the University of Montpellier and by a grant from the French Ministère de la Recherche et de la Technologie (MRT). The construction of the SilkBase was supported partly by a grant-in-aid from the Japan Ministry of Education, Science, Sport and Culture, and by the "Program for Promotion of Basic Research Activities for Innovation Bioscience" (PROBRAIN).

REFERENCES

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F. et al. 2000. The genome sequence of Drosophila melanogaster. *Science* **287:** 2185-2195.

Akashi, H. 1994. Synonymous codon usage in Drosophila melanogaster: natural selection and translational accuracy. *Genetics* **136:** 927-935.

Akashi, H., and Eyre-Walker, A. 1998. Translational selection and molecular evolution. *Curr Opin Genet Dev* **8:** 688-693.

Akashi, H., Kliman, R.M., and Eyre-Walker, A. 1998. Mutation pressure, natural selection, and the evolution of base composition in Drosophila. *Genetica* **103:** 49-60.

Akashi, H. 2001. Gene expression and molecular evolution. *Curr Opin Genet Dev* **11:** 660-666.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389-3402.

Anderson, P., Jonsson, M., and Morte, U. 2001. Variation in damage to cotton affecting larval feeding preference of *Spodoptera* littoralis. *Entomologia Experimentalis et Applicata* **101:** 191-198.

Barakat, A., Szick-Miranda, K., Chang, I.F., Guyot, R., Blanc, G., Cooke, R., Delseny, M., and Bailey-Serres, J. 2001. The organization of cytoplasmic ribosomal protein genes in the Arabidopsis genome. *Plant Physiol* **127:** 398-415.

Bortoluzzi, S., d'Alessi, F., Romualdi, C., and Danieli, G.A. 2001. Differential expression of genes coding for ribosomal proteins in different human tissues. *Bioinformatics* **17:** 1152-1157.

Ching'oma, G.P., and Pitre, H.N. 1999. Fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae), larval development and moth fecundity on sorghum at various stages of maturity. *Ceiba* **40:** 63-67.

Coghlan, A., and Wolfe, K.H. 2000. Relationship of codon bias to mRNA concentration and protein length in Saccharomyces cerevisiae. *Yeast* **16:** 1131-1145.

Comeron, J.M., and Kreitman, M. 1998. The correlation between synonymous and nonsynonymous substitutions in Drosophila: mutation, selection or relaxed constraints? *Genetics* **150:** 767-775.

Comeron, J.M., Kreitman, M., and M. Aguade. 1999. Natural selection on synonymous sites is correlated with gene length and recombination in Drosophila. *Genetics* **151:** 239-249.

Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16:** 10881-10890.

Doudna, J.A., and Rath, V.L. 2002. Structure and function of the eukaryotic ribosome: the next frontier. *Cell* **109:** 153-156.

Dudov, K.P., and Perry, R.P. 1984. The gene family encoding the mouse ribosomal protein L32 contains a uniquely expressed intron-containing gene and an unmutated processed gene. *Cell* **37:** 457-468.

Dunn, K.A., Bielawski, J.P., and Yang, Z. 2001. Substitution rates in Drosophila nuclear genes: implications for translational selection. *Genetics* **157:** 295-305.

Duret, L., and Mouchiroud, D. 1999. Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, and Arabidopsis. *Proc Natl Acad Sci U S A* **96:** 4482-4487.

Eyre-Walker, A. 1996. Synonymous codon bias is related to gene length in Escherichia coli: selection for translational accuracy? *Mol Biol Evol* **13:** 864-872.

Futcher, B., Latter, G.I., Monardo, P., McLaughlin, C.S. , and Garrels, J.I. 1999. A sampling of the yeast proteome. *Mol Cell Biol* **19:** 7357-7368.

Galitski, T., Saldanha, A.J., Styles, C.A., Lander, E.S. , and Fink, G.R. 1999. Ploidy regulation of gene expression. *Science* **285:** 251-254.

Garel, J. 1976. Quantitative adaptation of isoacceptor tRNAs to mRNA codons of alanine, glycine and serine. *Nature* **260:** 805-806.

Gouy, M., and Gautier, C. 1982. Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res* **10:** 7055-7074.

Hill, W.G., and Robertson, A. 1966. The effect of linkage on limits to artificial selection. *Genet Res* **8:** 269-294.

Howe, J.G., and Hershey, J.W. 1982. Immunochemical analysis of molecular forms of protein synthesis initiation factors in crude cell lysates of Escherichia coli. *Arch Biochem Biophys* **214:** 446-451.

Kenmochi, N., Kawaguchi, T., Rozen, S., Davis, E., Goodman, N., Hudson, T.J., Tanaka, T., and PAGE, D.C. 1998. A map of 75 human ribosomal protein genes. *Genome Res* **8:** 509-523.

Kliman, R.M., and Hey, J. 1994. The effects of mutation and natural selection on codon bias in the genes of Drosophila. *Genetics* **137:** 1049-1056.

Landais, I., Pommet, J., Mita, K., Nohata, J., Gimenez, S., Fournier, P., Devauchelle, G., Duonor-Cerutti, M., and Ogliastro, M. 2001. Characterization of the cDNA encoding the 90 kDa heat-shock protein in the Lepidoptera Bombyx mori and *Spodoptera frugiperda*. *Gene* **271:** 223-231.

Luckow, V.A. (1991). Cloning and expression of heterologous genes in insect cells with baculovirus vectors. *In* "Recombinant DNA Technology and Applications" (ed. B.R.K.a.H.C.S. Prokop A.), pp. 97-152. Mc Graw-Hill, N.Y.

Mager, W.H., Planta, R.J., Ballesta, J.G., Lee, J.C., Mizuta, K., Suzuki, K., Warner, J.R., and Woolford, J. 1997. A new nomenclature for the cytoplasmic ribosomal proteins of Saccharomyces cerevisiae. *Nucleic Acids Res* **25:** 4872-4875.

Michiels, L., Van der Rauwelaert, E., Van Hasselt, F., Kas, K., and Merregaert, J. 1993. fau cDNA encodes a ubiquitin-like-S30 fusion protein and is expressed as an antisense sequence in the Finkel-Biskis-Reilly murine sarcoma virus. *Oncogene* **8:** 2537-2546.

Minet, J. 1991. Tentative reconstruction of the ditrysian phylogeny (Lepidoptera: Glossata). *Entomologica Scandinavica* **22:** 69-95.

Moriyama, E.N., and Gojobori, T. 1992. Rates of synonymous substitution and base composition of nuclear genes in Drosophila. *Genetics* **130:** 855-864.

Moriyama, E.N., and Powell, J.R. 1997. Codon usage bias and tRNA abundance in Drosophila. *J Mol Evol* **45:** 514-523.

---. 1998. Gene length and codon usage bias in Drosophila melanogaster, Saccharomyces cerevisiae and Escherichia coli. *Nucleic Acids Res* **26:** 3188-3193.

Nielsen, E.S. 1989. Phylogeny of major lepidopteran groups. *In* "The hierarchy of life" (eds. K.B. Fernholm and H Jörnvall), pp. 281-294. Elsevier, Amsterdam.

Noller, H.F., Hoffarth, V., and Zimniak, L. 1992. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* **256:** 1416-1419.

Okano, K., Shimada, T., Mita, K., and Maeda, S. 2001. Comparative expressed-sequence-tag analysis of differential gene expression profiles in BmNPV-infected BmN cells. *Virology* **282:** 348-356.

Page, R.D. 1996. TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12:** 357-358.

Pashley, D.P., and Ke, L.D. 1992. Sequence evolution in mitochondrial ribosomal and ND-1 genes in Lepidoptera: implications for phylogenetic analyses. *Mol Biol Evol* **9:** 1061-1075

Pogue-Geile, K., Geiser, J.R., Shu, M., Miller, C., Wool, I.G., Meisler, A.I., and Pipas, J.M. 1991. Ribosomal protein genes are overexpressed in colorectal cancer: isolation of a cDNA clone encoding the human S3 ribosomal protein. *Mol Cell Biol* **11:** 3842-3849.

Powell, J.R., and Moriyama, E.N. 1997. Evolution of codon usage bias in Drosophila. *Proc Natl Acad Sci U S A* **94:** 7784-7790.

Ramakrishnan, V., and White, S.W. 1998. Ribosomal protein structures: insights into the architecture, machinery and evolution of the ribosome. *Trends Biochem Sci* **23:** 208-212.

Ramakrishnan, V. 2002. Ribosome structure and the mechanism of translation. *Cell* **108:** 557-572.

Regier, J.C., Fang, Q.Q., Mitter, C., Peigler, R.S., Friedlander, T.P., and Solis, M.A. 1998. Evolution and phylogenetic utility of the period gene in Lepidoptera. *Mol Biol Evol* **15:** 1172-1182.

Sharp, P.M., and Li, W.H. 1987. The Codon Adaptation Index - a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* **15:** 1281-1295.

Shields, D.C., Sharp, P.M., Higgins, D.G., and Wright, F. 1988. "Silent" sites in Drosophila genes are not neutral: evidence of selection among synonymous codons. *Mol Biol Evol* **5:** 704-716.

Sprague, K.U., Hagenbuchle, O., and Zuniga, M.C. 1977. The nucleotide sequence of two silk gland alanine tRNAs: implications for fibroin synthesis and for initiator tRNA structure. *Cell* **11:** 561-570.

Srinivasan, L., and Gopinathan, K.P. 2001. Differential expression of individual gene copies from within a tRNA multigene family in the mulberry silkworm Bombyx mori. *Insect Mol Biol* **10:** 523-530.

Sueoka, N. 1962. On the genetic basis of variation and heterogeneity of DNA base composition. *Proc Natl Acad Sci U S A* **34:** 95-114.

Suzuki, T., Terasaki, M., Takemoto-Hori, C., Hanada, T., Ueda, T., Wada, A., and Watanabe, K. 2001. Structural compensation for the deficit of rRNA with proteins in the mammalian mitochondrial ribosome. Systematic analysis of protein components of the large ribosomal subunit from mammalian mitochondria. *J Biol Chem* **276:** 21724-21736.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25:** 4876-4882.

Uechi, T., Tanaka, T., and Kenmochi, N. 2001. A complete map of the human ribosomal protein genes: assignment of 80 genes to the cytogenetic map and implications for human disorders. *Genomics* **72:** 223-230.

Vaughn, J.L., Goodwin, R.H., Tompkins, G.J., and McCawley, P. 1977. The establishment of two cell lines from the insect *Spodoptera frugiperda* (Lepidoptera; Noctuidae). *In Vitro* **13:** 213-217.

Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett Jr, D.E., Hieter, P., Vogelstein, B., and Kinzler, K.W. 1997. Characterization of the yeast transcriptome. *Cell* **88:** 243-251.

Wang, Q., Yang, C., Zhou, J., Wang, X., Wu, M., and Liu, Z. 2001. Cloning and characterization of full-length human ribosomal protein L15 cDNA which was overexpressed in esophageal cancer. *Gene* **263:** 205-209.

Warner, J.R., and Nierras, C.R. 1998. Trapping human ribosomal protein genes. *Genome Res* **8:** 419-421.

Warner, J.R. 1999. The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci* **24:** 437-440.

Wodicka, L., Dong, H., Mittmann, M., Ho, M.H., and Lockhart, D.J. 1997. Genome-wide expression monitoring in Saccharomyces cerevisiae. *Nat Biotechnol* **15:** 1359-1367.

Wolfe, K.H., and Shields, D.C. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387:** 708-713.

Wool, I.G. 1996. Extraribosomal functions of ribosomal proteins. *Trends Biochem Sci* **21:** 164-165.

Wright, F. 1990. The 'effective number of codons' used in a gene. *Gene* **87:** 23-29.

Yang, C., and Sehnal, F. 1998. Ribosomal protein L5 of Bombyx mori: cDNA sequence and tissue differences in the L5 mRNA content. *Insect Mol Biol* **7:** 301-306.

Yasukochi, Y. 1998. A dense genetic map of the silkworm, Bombyx mori, covering all chromosomes based on 1018 molecular markers. *Genetics* **150:** 1513-1525.

Yeates, D.K., and Wiegemann, B.M. 1999. Congruency and controversy: toward a higher-level phylogeny of Diptera. *Annu Rev Entomol* **44:** 397-428

Zhang, L., Zhou, W., Velculescu, V.E., Kern, S.E., Hruban, R.H., Hamilton, S.R., Vogelstein, B., and Kinzler, K.W. 1997. Gene expression profiles in normal and cancer cells. *Science* **276:** 1268-1272.

WEB SITE REFERENCES

http://www.ab.a.u-tokyo.ac.jp/lep-genome/; the International Lepidoptera Genome Project.

http://www.expasy.ch/cgi-bin/lists?ribosomp.txt/; the Ribosome Swiss-Prot Protein Database.

http://www.expasy.ch/prosite/; the Prosite Database.

http://www.ab.a.u-tokyo.ac.jp/silkbase/; the *B. mori* EST database (SilkBase).

http://www.ensembl.org/Anopheles_gambiae/; the International *Anopheles* genome project at ENSEMBL.

http://www.fruitfly.org/annot/; GADfly : the Berkeley Genome Annotation Database of *Drosophila*.

LEGENDS OF FIGURES

Figure 1 : Correlation between RP length (expressed in number of aa) and codon usage bias (ENC) using (A) 63 *S. frugiperda* RPs having a minimal size of 60 aa and (B) 67 *D. melanogaster* RPs having a minimal size of 100 aa. Each RP is represented by a dot on the diagrams. The arrow in the *D. melanogaster* plot indicates the spot corresponding to S2, a ribosomal protein much longer than the others. Correlation coefficients and associated probability are indicated at the bottom of each diagram.

[a] the correlation coefficient was calculated with discarding S2 from the analysis.

Figure 2 : Alignment of *S. frugiperda* RP L36 aa sequence with 12 other L36 aa sequences, including 4 insects. Conserved residues shared by at least 9 species are shaded in grey. The 10 aa-insertion in Insect species is shaded in black, with conserved residues in white, bold letters. Residues conserved only between Diptera species are in grey, bold letters. For complete names of species and accession numbers, see table 2.

Figure 3 : Phylogenetic analysis of RP L35A aa sequences (A, B) and L36 nucleotide sequences (C, D) belonging to 5 insect species (3 Lepidoptera and 2 Diptera species). Unrooted neighbor-joining trees were constructed using L35A and L36 sequences deleted from the insertions (A and C), or using only the insertions (B and D). The scales at the bottom of the figure correspond to 10% of divergence between sequences. The number of bootstrap trees (of 1000) in which the same branches occur is indicated near the nodes. For the complete names of species, see table 2.

IV- PERSPECTIVES

1. Les caractéristiques particulières des protéines ribosomales chez les insectes

1.1. Analyse des insertions dans la séquence des gènes L35A et L36

Les protéines ribosomales sont probablement des protéines très anciennes, dont la fonction structurale dans le ribosome, très fixée, a conduit à une forte conservation de la séquence (Wool, 1996). Dans ce contexte, la découverte que deux d'entre elles, L35A et L36, présentent une insertion spécifique des insectes était inattendue. La question principale quant à ces insertions concerne leur répartition phylogénétique, car nous ne disposons pour ces deux protéines que des séquences d'un nombre restreint d'espèces. Par ailleurs, même s'il est plus probable que ces insertions étaient présentes chez l'ancêtre commun des diptères et des lépidoptères (elles sont colinéaires chez tous les insectes testés, c'est-à-dire avec la même taille et la même position dans la protéine), on ne peut exclure définitivement l'hypothèse qu'elles résultent d'évènements d'insertion indépendants chez les différentes espèces.

La réponse à ces questions passe tout d'abord par le séquençage de ces deux gènes chez un plus grand nombre d'espèces, en se concentrant sur les invertébrés, et plus particulièrement sur les arthropodes. Au niveau méthodologique, l'obtention de fragments PCR pour le séquençage va être assez délicat dans le cas de L35A, car l'insertion consiste en une extension N-terminale de la séquence codante, ce qui ne va pas faciliter le dessin de l'amorce en 5' du gène. Pour L36 par contre, le dessin des amorces devrait être beaucoup plus facile, car les régions flanquant l'insertion sont bien conservées chez toutes les espèces testées. De plus, la faible taille de l'insertion (30 pb) devrait permettre un séquençage aisé des fragments amplifiés pour un coût modéré, permettant de tester un plus grand nombre d'espèces.

Le point suivant concerne l'utilisation de ces insertions comme marqueurs phylogénétiques. Le peu de séquences disponibles pour l'instant ne permet pas de conclure formellement à leur utilité, mais suggère tout de même qu'elles pourraient être intéressantes pour analyser la classe des insectes. Le fait que la divergence de séquences entre les cinq espèces d'insectes testées semble grossièrement être corrélée avec leur distance phylogénétique estimée mérite d'être exploré plus avant. Cependant, pour L36, la faible taille de l'insertion risque de grever son intérêt phylogénétique.

1.2. Biais de codons

Le faible biais de codons suggéré par l'analyse des gènes de protéines ribosomales chez deux lépidoptères (*Spodoptera* et *Bombyx*) doit être confirmé par un plus grand nombre de gènes chez

un plus grand nombre de Lépidoptères. Si cette caractéristique se confirme, au moins chez *Spodoptera frugiperda*, l'étape suivante sera de déterminer sa cause. L'hypothèse la plus probable (une concentration équivalente des ARNt) pourra être testée facilement par Northern blot, en utilisant des sondes spécifiques pour chaque isoaccepteur.

En ce qui concerne la relation entre taille des gènes et biais de codons chez les organismes pluricellulaires, la corrélation positive observée dans la famille des gènes de protéines ribosomales chez *Spodoptera* va à l'encontre de ce qui a été publié chez *Caenorhabditis*, *Drosophila* et *Arabidopsis*. Nous suggérons dans l'article qu'il pourrait s'agir d'un artefact lié à une mauvaise estimation du niveau d'expression des gènes utilisés dans ces analyses. Dans ce cadre, un nouvelle analyse de la corrélation entre taille des gènes et biais de codons pourrait être menée en utilisant des méthodes de mesure du niveau d'expression plus fiables, comme les microarrays ou bien la méthode SAGE. Nous pensons qu'une telle étude pourrait donner ds résultats différents de ceux déjà publiés chez les organismes pluricellulaires, comme cela a déjà été le cas pour la levure (Coghlan et Wolfe, 2000). Dans cette hypothèse, le modèle proposé en introduction pour expliquer cette corrélation serait applicable aussi bien aux organismes unique pluricellulaires.

1.3. Les gènes lépidoptères: des cibles pour de nouvelles molécules?

La banque d'ESTs *Sf*9 présente un contraste frappant entre des séquences peu nombreuses et représentées par un très grand nombre d'ESTs, et d'autres, beaucoup plus nombreuses et représentées par un petit nombre d'ESTs. Parmi ces dernières, une forte proportion ne présente aucune homologie avec les séquences présentes dans les banques internationales, pas même chez la drosophile, dont le génome est entièrement séquencé. Il est donc possible que ces séquences correspondent à des gènes qui ne sont présents que chez les lépidoptères. Pour confirmer cette hypothèse, une comparaison exhaustive avec les banques d'ESTs de *B. mori* devrait permettre d'identifier plusieurs gènes candidats. Ces gènes, potentiellement spécifiques des Lépidoptères, devront ensuite être étudiés plus en détail, par exemple par la détermination de leur pattern d'expression. Les filtres haute densité en cours de préparation pour les gènes de *S. frugiperda* seront sans doute très utiles pour cette analyse.

La caractérisation de gènes spécifiques des lépidoptères et l'identification de leur(s) fonction(s) est l'un des enjeux majeurs du programme international pour la génomique des lépidoptères: ils pourraient en effet permettre à plus ou moins long terme la mise au point de nouvelles stratégies de contrôle des lépidoptères ravageurs, par exemple en utilisant ces gènes comme cibles de nouvelles molécules "lépidoptéricides", moins dommageables pour le reste de l'entomofaune.

Dans le même ordre d'idée, les insertions observées dans la séquence des protéines ribosomales L35A et L36 pourraient éventuellement elles aussi constituer des cibles intéressantes. Cependant, cela suppose au préalable une étude structurale approfondie pour déterminer la place, l'importance et l'accessibilité de ces insertions dans la protéine et dans le ribosome.

2. Des outils pour le programme génomique de *S. frugiperda*

Les gènes de protéines ribosomales ont été utilisées comme outil pour caractériser la banque d'ESTs *Sf*9: ainsi que le suggèrent les résultats présentés dans l'article, le nombre d'ESTs correspondant à un gène de protéine ribosomale donné n'est généralement pas corrélé avec la quantitié d'ARNm dans la cellule *Sf*9. Ce biais de représentativité, testé sur quatre gènes seulement, nécessite d'être vérifié plus précisément. L'utilisation de filtres à haute densité, en cours de réalisation, devraient permettre de préciser son étendue.

D- CONCLUSION

La stratégie de séquençage à grande échelle de la banque d'ADNc de *Sf*9 a donné beaucoup d'informations à la fois sur les gènes exprimés dans la cellule *Sf*9 et sur la fiabilité de la banque au niveau de la représentativité des gènes. L'analyse approfondie de la famille des gènes de protéines ribosomales a en outre permis de mettre en évidence plusieurs caractéristiques spécifiques des insectes et des lépidoptères. Ces résultats, inattendus étant donnée la grande conservation générale de cette famille de gènes, montre que les protéines ribosomales peuvent être de bons outils pour étudier les insectes et même les lépidoptères.

Cette première analyse génomique du lépidoptère *S. frugiperda* a donc porté ses fruits, en répondant aux objectifs définis au moment de la création des banques d'ADNc et d'ESTs de *Sf*9. Cela dit, l'intérêt de ce programme d'étude génomique de *S. frugiperda* ne se cantonne pas aux thématiques du laboratoire de Pathologie Comparée : il s'intègre également dans le " Programme International sur le génome des Lépidoptères " (International Lepidoptera genome Project, www. ab.a.u-tokyo.ac.jp/lep-genome), lancé en juin 2001. Ce projet vise à promouvoir la coopération internationale pour le séquençage du génome complet de *B. mori* et la génomique comparative d'autres lépidopères économiquement importants. Il devrait permettre d'approfondir les connaissances fondamentales et appliqués sur les Lépidoptères, un ordre d'insectes qui compte plus de 160000 espèces dont un grand nombre sont des ravageurs importants des cultures, pré- et post-récolte. Le plan d'action de ce projet consiste en trois points : 1) le séquençage complet du génome de *B. mori*, qui servira de référence. Ce choix s'explique par la grande diversité de mutants, de cartes physiques et génétiques, de banques de BACs et d'ESTs

disponibles pour cette espèce, 2) la constitution de cartes physiques basées sur les BACs et de banques d'ESTs pour les autres lépidoptères, et 3) le développement d'outils post-génomiques, comme la transgenèse, les microarrays, l'analyse de mutants... pour lier informations génomiques et biologie des Lépidoptères. Dans le cadre de ce programme, nos résultats sur les particularités génomiques des lépidoptères devraient rapidement être relayés et étoffés par les données émanant d'équipes participant elles aussi au projet, mais sur d'autres espèces.

Conclusions et perspectives générales

L'étude fonctionnelle des relations moléculaires entre le baculovirus et la cellule aux temps précoces de l'infection constitue la première partie de ce mémoire. Elle a permis de progresser dans la compréhension des mécanismes de la transcription, non seulement pour le virus mais aussi pour le lépidoptère du fait de la similitude de leurs promoteurs. Ainsi, nous avons tiré parti du baculovirus et de la connaissance de son génome pour identifier et commencer à caractériser chez son hôte des facteurs probablement largement impliqués dans l'activation de la transcription: nous avons maintenant une idée (bien sûr très approximative) de leur nombre, leur masse moléculaire, leur activité dans le contexte de la cellule saine et infectée et les méthodes de purification qui permettront d'aller plus loin dans leur étude.

Le modèle baculovirus-cellule hôte est probablement l'un des plus avancés chez les invertébrés pour ce qui concerne les interactions moléculaires virus/cellule. L'implication de protéines de l'hôte dans l'activation de la transcription par le couple IE1/hrs apporte un niveau de complexité supplémentaire à ce mécanisme très étudié, que nous avons intégré dans un modèle de type enhanceosome. Des études supplémentaires vont être nécessaires pour le valider, passant par l'identification, la caractérisation et l'étude des modalités d'assemblage des différentes molécules composant l'enhanceosome. Sa dynamique et les interactions qu'il établit avec le complexe transcriptionnel basal devront également être déterminées. Tous ces travaux vont probablement permettre la découverte de nouveaux gènes, contribuant ainsi à l'enrichissement de la connaissance des modalités de la transcription chez le baculovirus et les lépidoptères, mais aussi du répertoire de gènes identifiés et caractérisés chez *Spodoptera frugiperda*.

Dans la deuxième partie du travail de thèse, nous avons suivi une approche complémentaire permettant de développer l'accès, jusqu'ici très limité, à ces gènes. Pour cela, une banque d'ADNc de cellules *Sf*9 a été construite, dont nous avons commencé la caractérisation par criblage et par séquençage à grande échelle. Nous avons choisi d'étudier plus particulièrement quelques-uns des gènes nouvellement identifiés (hsp90, les gènes de protéines ribosomales) en fonction de leur importance majeure dans la biologie de la cellule et de leur commodité d'étude. Leur caractérisation a fourni des informations multiples, permettant d'utiliser ces gènes à la fois comme outils (pour la génomique, la biologie moléculaire) et, à un niveau plus fondamental, comme marqueurs de la spécificité des insectes et/ou des lépidoptères.

Ce programme de génomique a été renforcé par la construction de nouvelles banques d'ADNc à partir de tissus différenciés (en cours de réalisation), donnant potentiellement accès à des gènes trop faiblement exprimés dans la cellule *Sf*9 (certains facteurs de transcription, par exemple). Le séquençage de 35000 clones devrait ainsi donner une vue globale du transcriptome de *Spodoptera frugiperda*, et sa comparaison avec celui d'autres lépidoptères aboutir à

l'identification de gènes spécifiques des lépidoptères. En parallèle, une banque génomique est en passe d'être achevée, et devrait donner accès à l'ensemble des gènes, de leur promoteur et des régions régulatrices.

Ces outils génomiques en cours d'élaboration vont nécessiter chez *S. frugiperda* le développement parallèle d'outils post-génomiques, pour faire le lien entre le gène et sa fonction dans la cellule. La mise au point de la modification génétique (par transposon par exemple) chez cette espèce est dans ce cadre un enjeu important, mais des techniques alternatives de contrôle de l'expression des gènes (endogènes ou hétérologues), moins lourdes à mettre en oeuvre et potentiellement très performantes sont déjà à l'étude, notamment l'ARN interférence (ARNi) et le système densovirus.

Pour conclure, l'analyse transcriptionnelle des étapes précoces de l'infection baculovirale menée au cours de cette thèse n'aura pas bénéficié des outils génomiques mis en place pour en faciliter l'étude. Cependant, il est probable que ceux-ci vont se révéler fort utiles dans le cadre des travaux encore à accomplir, notamment pour l'étude de l'enhanceosome viral et de la multitude des facteurs cellulaires qu'il contient ou avec lesquels il interagit.

Annexe : caractérisation du gène HSP90 chez deux lépidoptères

A- RESUME EN FRANÇAIS DE L'ARTICLE

"Characterization of the cDNA encoding the 90 kDa heat-shock protein in the Lepidoptera *Bombyx mori* and *Spodoptera frugiperda*".

Cet article présente pour la première fois la séquence complète de l'ADNc de *hsp90* chez deux Lépidoptères. La séquence de *B. mori* a été reconstituée à partir de 15 clones partiels provenant de banques d'ESTs de différents tissus et lignées cellulaires, montrant l'expression ubiquitaire du gène. L'ADNc de *S. frugiperda* a été isolé à partir d'une banque établie à partir de la lignée cellulaire *Sf*9. Les deux ADNc codent pour des protéines très homologues présentant la signature classique de HSP90. Leur taille est de 716 aa (*B. mori*) et 717 aa (*S. frugiperda*), avec une masse moléculaire calculée de 83 kDa, similaire à celle *de D. melanogaster*. Contrairement aux vertébrés, *hsp90* est un gène unique chez les deux lépidoptères. Le séquençage de la région génomique correspondante montre que, à la différence des diptères, il n'y a pas d'intron dans ce gène chez les lépidoptères. Une analyse phylogénétique basée sur les séquences en acides aminées des deux lépidoptères et de 23 autres espèces montre une bonne correspondance avec la phylogénie connue de ces espèces, à la fois à haut et bas niveau taxonomique. L'analyse transcriptionnelle effectuée chez *S. frugiperda* montre que l'induction du gène *hsp90* ne survient que 14°C au-dessus des conditions physiologiques de croissance.

B- ARTICLE 3

Bibliographie

Adams M. D., Celniker S. E., Holt R. A., Evans C. A., Gocayne J. D., Amanatides P. G., Scherer S. E., Li P. W., Hoskins R. A., Galle R. F. *et al.* 2000. The genome sequence of *Drosophila melanogaster*. Science **287:**2185-2195.

Ahrens C. H. et Rohrmann G. F. 1995. Replication of *Orgyia pseudotsugata* baculovirus DNA: lef-2 and ie-1 are essential and ie-2, p34, and *Op*-iap are stimulatory genes. Virology **212:**650-662.

Ahrens C. H., Russell R. L., Funk C. J., Evans J. T., Harwood S. H. et Rohrmann G. F. 1997. The sequence of the *Orgyia pseudotsugata* multinucleocapsid nuclear polyhedrosis virus genome. Virology **229:**381-399.

Akashi H. 1994. Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. Genetics **136:**927-935.

Akashi H. et Eyre-Walker A. 1998. Translational selection and molecular evolution. Curr Opin Genet Dev **8:**688-693.

Akashi H., Kliman R. M. et Eyre-Walker A. 1998. Mutation pressure, natural selection, and the evolution of base composition in *Drosophila*. Genetica **103:**49-60.

Akashi H. 2001. Gene expression and molecular evolution. Curr Opin Genet Dev **11:**660-666.

Altschul S. F., Madden T. L., Schaffer A. A., Zhang J., Zhang Z., Miller W. et Lipman D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:**3389-3402.

Anderson P., Jonsson M. et Morte U. 2001. Variation in damage to cotton affecting larval feeding preference of *Spodoptera littoralis*. Entomologia Experimentalis et Applicata **101:**191-198.

Andrisani O. et Dixon J. E. 1990. Identification and purification of a novel 120-kDa protein that recognizes the cAMP-responsive element. J Biol Chem **265:**3212-3218.

Andrisani O. M., Hayes T. E., Roos B. et Dixon J. E. 1987. Identification of the promoter sequences involved in the cell specific expression of the rat somatostatin gene. Nucleic Acids Res **15:**5715-5728.

Angel P., Imagawa M., Chiu R., Stein B., Imbra R. J., Rahmsdorf H. J., Jonat C., Herrlich P. et Karin M. 1987. Phorbol ester-inducible genes contain a common cis element recognized by a TPA-modulated trans-acting factor. Cell **49:**729-739.

Armstrong A. P., Franklin A. A., Uittenbogaard M. N., Giebler H. A. et Nyborg J. K. 1993. Pleiotropic effect of the human T-cell leukemia virus Tax protein on the DNA binding activity of eukaryotic transcription factors. Proc Natl Acad Sci U S A **90:**7303-7307.

Ausubel F. M., Brent R., Kingston R. E., Moore D. D., Seidman J. G., Smith J. A. et Struhl K. 1996. Current protocols in molecular biology. Greene Publishing Associates and Wiley Interscience.

Ayres M. D., Howard S. C., Kuzio J., Lopez-Ferber M. et Possee R. D. 1994. The complete DNA sequence of Autographa californica nuclear polyhedrosis virus. Virology **202:**586-605.

Bantignies F., Rousset R., Desbois C. et Jalinot P. 1996. Genetic characterization of transactivation of the human T-cell leukemia virus type 1 promoter: Binding of Tax to Tax-responsive element 1 is mediated by the cyclic AMP-responsive members of the CREB/ATF family of transcription factors. Mol Cell Biol **16:**2174-2182.

Barakat A., Szick-Miranda K., Chang I. F., Guyot R., Blanc G., Cooke R., Delseny M. et Bailey-Serres J. 2001. The organization of cytoplasmic ribosomal protein genes in the Arabidopsis genome. Plant Physiol **127:**398-415.

Barrett J. W., Brownwright A. J., Primavera M. J. et Palli S. R. 1998. Studies of the nucleopolyhedrovirus infection process in insects by using the green fluorescence protein as a reporter. J Virol **72:**3377-3382.

Barton K., Muthusamy N., Chanyangam M., Fischer C., Clendenin C. et Leiden J. M. 1996. Defective thymocyte proliferation and IL-2 production in transgenic mice expressing a dominant-negative form of CREB. Nature **379:**81-85.

Benbrook D. M. et Jones N. C. 1994. Different binding specificities and transactivation of variant CRE's by CREB complexes. Nucleic Acids Res **22:**1463-1469.

Beniya H., Funk C. J., Rohrmann G. F. et 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.

Berkowitz L. A., Riabowol K. T. et Gilman M. Z. 1989. Multiple sequence elements of a single functional class are required for cyclic AMP responsiveness of the mouse c-fos promoter. Mol Cell Biol **9:**4272-4281.

Bishop D. H. L., Hirst M. L., Possee R. D. et Cory J. S. 1995. Genetic engineering of microbes: virus insecticides - a case study, p. 249-277. *In* Darby G. K., Hunter P. A. et Russell A. D. (ed.), Fifty years of microbials. Cambridge University Press, Cambridge.

Blissard G. W. et Rohrmann G. F. 1990. Baculovirus diversity and molecular biology. Annu Rev Entomol **35:**127-155.

Blissard G. W., Kogan P. H., Wei R. et 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.

Blissard G. W. 1996. Baculovirus--insect cell interactions. Cytotechnology **20:**73-93.

Bohmann D., Bos T. J., Admon A., Nishimura T., Vogt P. K. et Tjian R. 1987. Human proto-oncogene c-jun encodes a DNA binding protein with structural and functional properties of transcription factor AP-1. Science **238:**1386-1392.

Bortoluzzi S., d'Alessi F., Romualdi C. et Danieli G. A. 2001. Differential expression of genes coding for ribosomal proteins in different human tissues. Bioinformatics **17:**1152-1157.

Carey M. 1998. The enhanceosome and transcriptional synergy. Cell **92:**5-8.

Carson D. D., Summers M. D. et Guarino L. A. 1991. Transient expression of the *Autographa californica* nuclear polyhedrosis virus immediate-early gene, IE-N, is regulated by three viral elements. J Virol **65:**945-951.

Cartier J. L., Hershberger P. A. et Friesen P. D. 1994. Suppression of apoptosis in insect cells stably transfected with baculovirus p35: dominant interference by N-terminal sequences p35(1- 76). J Virol **68:**7728-7737.

Chaabihi H., Ogliastro M. H., Martin M., Giraud C., Devauchelle G. et Cerutti M. 1993. Competition between baculovirus polyhedrin and p10 gene expression during infection of insect cells. J Virol **67:**2664-2671.

Chan H. M. et La Thangue N. B. 2001. p300/CBP proteins: HATs for transcriptional bridges and scaffolds. J Cell Sci **114:**2363-2373.

Chang Y. N., Crawford S., Stall J., Rawlins D. R., Jeang K. T. et Hayward G. S. 1990. The palindromic series I repeats in the simian cytomegalovirus major immediate-early promoter behave as both strong basal enhancers and cyclic AMP response elements. J Virol **64:**264-277.

Cherbas L. et Cherbas P. 1993. The arthropod initiator: the capsite consensus plays an important role in transcription. Insect Biochem Mol Biol **23:**81-90.

Ching'oma G. P. et Pitre H. N. 1999. Fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae), larval development and moth fecundity on sorghum at various stages of maturity. Ceiba **40:**63-67.

Chisholm G. E. et Henner D. J. 1988. Multiple early transcripts and splicing of the *Autographa californica* Nuclear Polyhedrosis Virus IE-1 gene. Journal of Virology **62:**3193-3200.

Chodosh L. A., Carthew R. W. et Sharp P. A. 1986. A single polypeptide possesses the binding and transcription activities of the adenovirus major late transcription factor. Mol Cell Biol **6:**4723-4733.

Chodosh L. A. 1996. UV crosslinking of proteins to nucleic acids, p. 12.15.11-18. *In* Ausubel F. M., Brent R., Kingston R. E., Moore D. D., Seidman J. G., Smith J. A. et Struhl K. (ed.), Current Protocols in Molecular Biology, vol. 2. Greene Publishing Associates and Wiley Interscience.

Choi J. et Guarino L. A. 1995a. The baculovirus transactivator IE1 binds to viral enhancer elements in the absence of insect cell factors. J Virol **69:**4548-4551.

Choi J. et Guarino L. A. 1995b. Expression of the IE1 transactivator *of Autographa californica* nuclear polyhedrosis virus during viral infection. Virology **209:**99-107.

Choi J. et Guarino L. A. 1995c. A temperature-sensitive IE1 protein of *Autographa californica* nuclear polyhedrosis virus has altered transactivation and DNA binding activities. Virology **209:**90-98.

Coghlan A. et Wolfe K. H. 2000. Relationship of codon bias to mRNA concentration and protein length in *Saccharomyces cerevisiae*. Yeast **16:**1131-1145.

Comeron J. M. et Kreitman M. 1998. The correlation between synonymous and nonsynonymous substitutions in *Drosophila*: mutation, selection or relaxed constraints? Genetics **150:**767-775.

Comeron J. M., Kreitman M. et Aguade M. 1999. Natural selection on synonymous sites is correlated with gene length and recombination in *Drosophila*. Genetics **151:**239-249.

Corpet F. 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res **16:**10881-10890.

Craig J. C., Schumacher M. A., Mansoor S. E., Farrens D. L., Brennan R. G. et Goodman R. H. 2001. Consensus and variant cAMP-regulated enhancers have distinct CREB-binding properties. J Biol Chem **276:**11719-11728.

Curran T. et Teich N. M. 1982. Identification of a 39,000-dalton protein in cells transformed by the FBJ murine osteosarcoma virus. Virology **116:**221-235.

Delmas V., Molina C. A., Lalli E., de Groot R., Foulkes N. S., Masquilier D. et Sassone-Corsi P. 1994. Complexity and versatility of the transcriptional response to cAMP. Rev Physiol Biochem Pharmacol **124:**1-28.

Derksen A. C. et Granados R. R. 1988. Alteration of a lepidopteran peritrophic membrane by baculoviruses and enhancement of viral infectivity. Virology **167:**242-250.

Dickson J. A. et Friesen P. D. 1991. Identification of upstream promoter elements mediating early transcription from the 35,000-molecular-weight protein gene of *Autographa californica* nuclear polyhedrosis virus. J Virol **65:**4006-4016.

Dignam J. 1990. Preparation of extracts from higher eukaryotes. Methods Enzymol **182:**194-203.

Doudna J. A. et Rath V. L. 2002. Structure and function of the eukaryotic ribosome: the next frontier. Cell **109:**153-156.

Drust D. S., Troccoli N. M. et Jameson J. L. 1991. Binding specificity of cyclic adenosine 3',5'-monophosphate-responsive element (CRE)-binding proteins and activating transcription factors to naturally occurring CRE sequence variants. Mol Endocrinol **5:**1541-1551.

Dudov K. P. et Perry R. P. 1984. The gene family encoding the mouse ribosomal protein L32 contains a uniquely expressed intron-containing gene and an unmutated processed gene. Cell **37:**457-468.

Dunn K. A., Bielawski J. P. et Yang Z. 2001. Substitution rates in *Drosophila* nuclear genes: implications for translational selection. Genetics **157:**295-305.

Duret L. et Mouchiroud D. 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. Proc Natl Acad Sci U S A **96:**4482-4487.

Engelhard E. K., Kam-Morgan L. N., Washburn J. O. et Volkman L. E. 1994. The insect tracheal system: a conduit for the systemic spread of *Autographa californica* M nuclear polyhedrosis virus. Proc Natl Acad Sci U S A **91:**3224-3227.

Eyre-Walker A. 1996. Synonymous codon bias is related to gene length in *Escherichia coli*: selection for translational accuracy? Mol Biol Evol **13:**864-872.

Fax P., Lipinski K. S., Esche H. et Brockmann D. 2000. cAMP-independent activation of the adenovirus type 12 E2 promoter correlates with the recruitment of CREB-1/ATF-1, E1A(12S), and CBP to the E2-CRE. J Biol Chem **275:**8911-8920.

Federici B. A. 1997. Baculovirus pathogenesis, p. 33-59. *In* Miller L. K. (ed.), The Baculoviruses. Plenum Press, New York.

Fenstermaker R. A., Farmerie T. A., Clay C. M., Hamernik D. L. et Nilson J. H. 1990. Different combinations of regulatory elements may account for expression of the glycoprotein hormone alpha-subunit gene in primate and horse placenta. Mol Endocrinol **4:**1480-1487.

Fentzke R. C., Korcarz C. E., Lang R. M., Lin H. et Leiden J. M. 1998. Dilated cardiomyopathy in transgenic mice expressing a dominant- negative CREB transcription factor in the heart. J Clin Invest **101:**2415-2426.

Fisch T. M., Prywes R., Simon M. C. et Roeder R. G. 1989. Multiple sequence elements in the c-fos promoter mediate induction by cAMP. Genes Dev **3:**198-211.

Franza B. R., Jr., Rauscher F. J., 3rd, Josephs S. F. et Curran T. 1988. The Fos complex and Fos-related antigens recognize sequence elements that contain AP-1 binding sites. Science **239:**1150-1153.

Friesen P. D. 1997. Regulation of Baculovirus early gene expression, p. 141-170. *In* Miller L. K. (ed.), The Baculoviruses. Plenum Press, New York.

Futcher B., Latter G. I., Monardo P., McLaughlin C. S. et Garrels J. I. 1999. A sampling of the yeast proteome. Mol Cell Biol **19:**7357-7368.

Galitski T., Saldanha A. J., Styles C. A., Lander E. S. et Fink G. R. 1999. Ploidy regulation of gene expression. Science **285:**251-254.

Garcia-Maruniak A., Pavan O. H. et Maruniak J. E. 1996. A variable region of Anticarsia gemmatalis nuclear polyhedrosis virus contains tandemly repeated DNA sequences. Virus Res **41:**123-132.

Garel J. 1976. Quantitative adaptation of isoacceptor tRNAs to mRNA codons of alanine, glycine and serine. Nature **260:**805-806.

Giese K., Kingsley C., Kirshner J. R. et Grosschedl R. 1995. Assembly and function of a TCR alpha enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein-protein interactions. Genes Dev **9:**995-1008.

Gomi S., Majima K. et Maeda S. 1999. Sequence analysis of the genome of *Bombyx mori* nucleopolyhedrovirus. J Gen Virol **80:**1323-1337.

Goodman R. H. et Smolik S. 2000. CBP/p300 in cell growth, transformation, and development. Genes Dev **14:**1553-1577.

Gorman C. M., Merlino G. T., Willingham M. C., Pastan I. et Howard B. H. 1982. The Rous sarcoma virus long terminal repeat is a strong promoter when introduced into a variety of eukaryotic cells by DNA-mediated transfection. Proc Natl Acad Sci U S A **79:**6777-6781.

Gouy M. et Gautier C. 1982. Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res **10:**7055-7074.

Granados R. R., Lawler K. A. et Burand J. P. 1981. Replication of *Heliothis zea* baculovirus in an insect cell line. Intervirology **16:**71-79.

Guarino L. A. et Dong W. 1991. Expression of an enhancer-binding protein in insect cells transfected with the *Autographa californica* nuclear polyhedrosis virus IE1 gene. J Virol **65:**3676-3680.

Guarino L. A. et Smith M. 1992. Regulation of delayed-early gene transcription by dual TATA boxes. J Virol **66:**3733-3739.

Habib S. et Hasnain S. E. 1996. A 38-kDa host factor interacts with functionally important motifs within the *Autographa californica* multinucleocapsid nuclear polyhedrosis virus homologous region (hr1) DNA sequence. J Biol Chem **271:**28250-28258.

Habib S., Pandey S., Chatterji U., Burma S., Ahmad R., Jain A. et Hasnain S. E. 1996. Bifunctionality of the *Ac*MNPV homologous region sequence (hr1): enhancer and ori functions have different sequence requirements. DNA Cell Biol **15:**737-747.

Hai T. et Curran T. 1991. Cross-family dimerization of transcription factors Fos/Jun and ATF/CREB alters DNA binding specificity. Proc. Natl. Acad. Sci. USA **88:**3720-3724.

Hai T. et Hartman M. G. 2001. The molecular biology and nomenclature of the activating transcription factor/cAMP responsive element binding family of transcription factors: activating transcription factor proteins and homeostasis. Gene **273:**1-11.

Hashimoto Y., Hayakawa T., Ueno Y., Fujita T., Sano Y. et Matsumoto T. 2000. Sequence analysis of the *Plutella xylostella* granulovirus genome. Virology **275:**358-372.

Hawtin R. E., Zarkowska T., Arnold K., Thomas C. J., Gooday G. W., King L. A., Kuzio J. A. et Possee R. D. 1997. Liquefaction of *Autographa californica* nucleopolyhedrovirus-infected insects is dependent on the integrity of virus-encoded chitinase and cathepsin genes. Virology **238:**243-253.

Herdegen T. et Leah J. D. 1998. Inducible and constitutive transcription factors in the mammalian nervous system: control of gene expression by Jun, Fos and Krox, and CREB/ATF proteins. Brain Res Brain Res Rev **28:**370-490.

Hewes R. S., Schaefer A. M. et Taghert P. H. 2000. The cryptocephal gene (ATF4) encodes multiple basic-leucine zipper proteins controlling molting and metamorphosis in Drosophila. Genetics **155:**1711-1723.

Hill W. G. et Robertson A. 1966. The effect of linkage on limits to artificial selection. Genet Res **8:**269-294.

Hottiger M. O. et Nabel G. J. 2000. Viral replication and the coactivators p300 and CBP. Trends Microbiol **8:**560-565.

Howe J. G. et Hershey J. W. 1982. Immunochemical analysis of molecular forms of protein synthesis initiation factors in crude cell lysates of *Escherichia coli*. Arch Biochem Biophys **214:**446-451.

Huh N. E. et 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.

Jakel S. et Gorlich D. 1998. Importin beta, transportin, RanBP5 and RanBP7 mediate nuclear import of ribosomal proteins in mammalian cells. Embo J **17:**4491-4502.

Jarvis D. L., Fleming J. A., Kovacs G. R., Summers M. D. et Guarino L. A. 1990. Use of early baculovirus promoters for continuous expression and efficient processing of foreign gene products in stably transformed lepidopteran cells. Biotechnology (N Y) **8:**950-955.

Jeang K. T., Boros I., Brady J., Radonovich M. et Khoury G. 1988. Characterization of cellular factors that interact with the human T- cell leukemia virus type I p40x-responsive 21-base-pair sequence. J Virol **62:**4499-4509.

Jones N. 1995. Transcriptional modulation by the adenovirus E1A gene. Curr Top Microbiol Immunol **199:**59-80.

Julien M. 1999. Etude de mutants CFTR dans le système d'expression Baculovirus / cellules d'insectes. Thèse: Biologie cellulaire, 260 p. Montpellier II.

Kenmochi N., Kawaguchi T., Rozen S., Davis E., Goodman N., Hudson T. J., Tanaka T. et Page D. C. 1998. A map of 75 human ribosomal protein genes. Genome Res **8:**509-523.

Kerry J. A., Priddy M. A., Jervey T. Y., Kohler C. P., Staley T. L., Vanson C. D., Jones T. R., Iskenderian A. C., Anders D. G. et Stenberg R. M. 1996. Multiple regulatory events influence human cytomegalovirus DNA polymerase (UL54) expression during viral infection. J Virol **70:**373-382.

Kerry J. A., Priddy M. A., Staley T. L., Jones T. R. et Stenberg R. M. 1997. The role of ATF in regulating the human cytomegalovirus DNA polymerase (UL54) promoter during viral infection. J Virol **71:**2120-2126.

Kim T. K. et Maniatis T. 1997. The mechanism of transcriptional synergy of an in vitro assembled interferon-beta enhanceosome. Mol Cell **1:**119-129.

Kliman R. M. et Hey J. 1994. The effects of mutation and natural selection on codon bias in the genes of *Drosophila*. Genetics **137:**1049-1056.

Klucher K. M., Sommer M., Kadonaga J. T. et Spector D. H. 1993. In vivo and in vitro analysis of transcriptional activation mediated by the human cytomegalovirus major immediate-early proteins. Mol Cell Biol **13:**1238-1250.

Kool M., van den Berg P. M., Tramper J., Goldbach R. W. et Vlak J. M. 1993a. Location of two putative origins of DNA replication of *Autographa californica* nuclear polyhedrosis virus. Virology **192:**94-101.

Kool M. et Vlak J. M. 1993. The structural and functional organization of the *Autographa californica* nuclear polyhedrosis virus genome. Arch Virol **130:**1-16.

Kool M., Voeten J. T., Goldbach R. W., Tramper J. et Vlak J. M. 1993b. Identification of seven putative origins of *Autographa californica* multiple nucleocapsid nuclear polyhedrosis virus DNA replication. J Gen Virol **74:**2661-2668.

Kovacs G. R., Guarino L. A. et Summers M. D. 1991. Novel regulatory properties of the IE1 and IE0 transactivators encoded by the baculovirus *Autographa californica* multicapsid nuclear polyhedrosis virus. J Virol **65:**5281-5288.

Kovacs G. R., Choi J., Guarino L. A. et Summers M. D. 1992. Functional dissection of the *Autographa californica* nuclear polyhedrosis virus immediate-early 1 transcriptional regulatory protein. Journal of Virology **66:**7429-7437.

Kremer A. et Knebel-Mörsdorf. 1998. The early baculovirus he65 promoter : on the mechanism of transcriptional activation by IE1. Virology **249:**336-351.

Kurokawa R., Kalafus D., Ogliastro M. H., Kioussi C., Xu L., Torchia J., Rosenfeld M. G. et Glass C. K. 1998. Differential use of CREB binding protein-coactivator complexes. Science **279:**700-703.

Kuzio J., Pearson M. N., Harwood S. H., Funk C. J., Evans J. T., Slavicek J. M. et Rohrmann G. F. 1999. Sequence and analysis of the genome of a baculovirus pathogenic for *Lymantria dispar*. Virology **253:**17-34.

Laemmli U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:**680-685.

Landais I., Pommet J., Mita K., Nohata J., Gimenez S., Fournier P., Devauchelle G., Duonor-Cerutti M. et Ogliastro M. 2001. Characterization of the cDNA encoding the 90 kDa heat-shock protein in the Lepidoptera *Bombyx mori* and *Spodoptera frugiperda*. Gene **271:**223-231.

Lee H. et Krell P. J. 1994. Reiterated DNA fragments in defective genomes of *Autographa californica* nuclear polyhedrosis virus are competent for *Ac*MNPV-dependent DNA replication. Virology **202:**418-429.

Lee K. A., Hai T. Y., SivaRaman L., Thimmappaya B., Hurst H. C., Jones N. C. et Green M. R. 1987a. A cellular protein, activating transcription factor, activates transcription of multiple E1A-inducible adenovirus early promoters. Proc Natl Acad Sci U S A **84:**8355-8359.

Lee S. Y., Poloumienko A., Belfry S., Qu X., Chen W., MacAfee N., Morin B., Lucarotti C. et Krause M. 1996. A common pathway for p10 and calyx proteins in progressive stages of polyhedron envelope assembly in *Ac*MNPV-infected *Spodoptera frugiperda* larvae. Arch Virol **141:**1247-1258.

Lee W., Mitchell P. et Tjian R. 1987b. Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. Cell **49:**741-752.

Leisy D. J., Rasmussen C., Kim H. T. et Rohrmann G. F. 1995. The *Autographa californica* nuclear polyhedrosis virus homologous region 1a: identical sequences are essential for DNA replication activity and transcriptional enhancer function. Virology **208:**742-752.

Leisy D. J., Rasmussen C., Kim H. T. et Rohrmann G. 1996. Presented at the Annual meeting of the American Society of Virology.

Leisy D. J. et Rohrmann G. F. 2000. The *Autographa californica* nucleopolyhedrovirus IE-1 protein complex has two modes of specific DNA binding. Virology **274:**196-202.

Li B., Nierras C. R. et Warner J. R. 1999. Transcriptional elements involved in the repression of ribosomal protein synthesis. Molecular and cellular biology **19:**5393-5404.

Lopez-Ferber M., Sisk W. P. et Possee R. D. 1995. Baculovirus transfer vectors. Methods Mol Biol **39:**25-63.

Lu A. et Carstens E. B. 1993. Immediate-early baculovirus genes transactivate the p143 gene promoter of *Autographa californica* nuclear polyhedrosis virus. Virology **195:**710-718.

Lu A. et Miller L. K. 1995. The roles of eighteen baculovirus late expression factor genes in transcription and DNA replication. J Virol **69:**975-982.

Lu A. et Miller L. K. 1997. Regulation of Baculovirus late and very late gene expression, p. 193-216. *In* Miller L. K. (ed.), The Baculoviruses. Plenum Press, New York.

Lu M., Johnson R. R. et Iatrou K. 1996. Trans-activation of a cell housekeeping gene promoter by the IE1 gene product of baculoviruses. Virology **218:**103-113.

Lu M., Farrell P. J., Johnson R. et Iatrou K. 1997. A baculovirus (*Bombyx mori* nuclear polyhedrosis virus) repeat element functions as a powerful constitutive enhancer in transfected insect cells. J Biol Chem **272:**30724-30728.

Luckow B. et Schutz G. 1989. Cell-type specificity of regulatory elements identified by linker scanning mutagenesis in the promoter of the chicken lysozyme gene. Nucleic Acids Res **17:**8451-8462.

Luckow V. A. 1991. Cloning and expression of heterologous genes in insect cells with baculovirus vectors, p. 97-152. *In* Prokop A. B. R. K. a. H. C. S. (ed.), Recombinant DNA Technology and Applications. Mc Graw-Hill, N.Y.

Mager W. H., Planta R. J., Ballesta J. G., Lee J. C., Mizuta K., Suzuki K., Warner J. R. et Woolford J. 1997. A new nomenclature for the cytoplasmic ribosomal proteins of Saccharomyces cerevisiae. Nucleic Acids Res **25:**4872-4875.

Maki Y., Bos T. J., Davis C., Starbuck M. et Vogt P. K. 1987. Avian sarcoma virus 17 carries the jun oncogene. Proc Natl Acad Sci U S A **84:**2848-2852.

Masson N., Ellis M., Goodbourn S. et Lee K. A. 1992. Cyclic AMP response element-binding protein and the catalytic subunit of protein kinase A are present in F9 embryonal carcinoma cells but are unable to activate the somatostatin promoter. Mol Cell Biol **12:**1096-1106.

Mayall T. P., Sheridan P. L., Montminy M. R. et Jones K. A. 1997. Distinct roles for P-CREB and LEF-1 in TCR alpha enhancer assembly and activation on chromatin templates in vitro. Genes Dev **11:**887-899.

Merika M., Williams A. J., Chen G., Collins T. et Thanos D. 1998. Recruitment of CBP/p300 by the IFN beta enhanceosome is required for synergistic activation of transcription. Mol Cell **1:**277-287.

Meyer T. E. et Habener J. F. 1993. Cyclic adenosine 3',5'-monophosphate response element binding protein (CREB) and related transcription-activating deoxyribonucleic acid- binding proteins. Endocr Rev **14:**269-290.

Michiels L., Van der Rauwelaert E., Van Hasselt F., Kas K. et Merregaert J. 1993. fau cDNA encodes a ubiquitin-like-S30 fusion protein and is expressed as an antisense sequence in the Finkel-Biskis-Reilly murine sarcoma virus. Oncogene **8:**2537-2546.

Miller L. K. 1997. The baculoviruses. Plenum Press, New York.

Millhouse S., Kenny J. J., Quinn P. G., Lee V. et Wigdahl B. 1998. ATF/CREB elements in the herpes simplex virus type 1 latency-associated transcript promoter interact with members of the ATF/CREB and AP-1 transcription factor families. J Biomed Sci **5:**451-464.

Minet J. 1991. Tentative reconstruction of the ditrysian phylogeny (Lepidoptera: Glossata). Entomologica Scandinavica **22:**69-95.

Montminy M. 1997. Transcriptional regulation by cyclic AMP. Annu Rev Biochem **66:**807-822.

Montminy M. R. et Bilezikjian L. M. 1987. Binding of a nuclear protein to the cyclic-AMP response element of the somatostatin gene. Nature **328:**175-178.

Moriyama E. N. et Gojobori T. 1992. Rates of synonymous substitution and base composition of nuclear genes in *Drosophila*. Genetics **130:**855-864.

Moriyama E. N. et Powell J. R. 1997. Codon usage bias and tRNA abundance in *Drosophila*. J Mol Evol **45:**514-523.

Moriyama E. N. et Powell J. R. 1998. Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*. Nucleic Acids Res **26:**3188-3193.

Morse B. A., Carruth L. M. et Clements J. E. 1999. Targeting of the visna virus tat protein to AP-1 sites: interactions with the bZIP domains of fos and jun in vitro and in vivo. J Virol **73:**37-45.

Murges D., Kremer A. et Knebel-Morsdorf D. 1997. Baculovirus transactivator IE1 is functional in mammalian cells. J Gen Virol **78:**1507-1510.

Murphy F. A., Fauquet C. M., Bishop D. H. L., Ghabrial S. A., Jarvis A. W., Martelli G. P., Mayo M. A. et Summers M. D. 1995. Virus Taxonomy. Springer Verlag, Vienne.

Nielsen E. S. 1989. Phylogeny of major lepidopteran groups., p. 281-294. *In* B Fernholm K. B., and H Jörnvall (ed.), The hierarchy of life. Elsevier, Amsterdam.

Nissen M. S. et Friesen P. D. 1989. Molecular analysis of the transcriptional regulatory region of an early baculovirus gene. J Virol **63:**493-503.

Nissen P., Hansen J., Ban N., Moore P. B. et Steitz T. A. 2000. The structural basis of ribosome activity in peptide bond synthesis. Science **289:**920-930.

Noller H. F., Hoffarth V. et Zimniak L. 1992. Unusual resistance of peptidyl transferase to protein extraction procedures. Science **256:**1416-1419.

Ogliastro M. H. 1994. Activité transcriptionnelle de la région p10 du baculovirus d'*Autographa californica* (*Ac*MNPV) en cellules infectées et non infectées
Mise en évidence d'un transactivateur. Thèse: Sciences de la Vie, 132 p. Paris 6.

Ohresser M., Morin N., Cerutti M. et Delsert C. 1994. Temporal regulation of a complex and unconventional promoter by viral products. J Virol **68:**2589-2597.

Okano K., Shimada T., Mita K. et Maeda S. 2001. Comparative expressed-sequence-tag analysis of differential gene expression profiles in *Bm*NPV-infected BmN cells. Virology **282:**348-356.

Olson V. A., Wetter J. A. et Friesen P. D. 2001. Oligomerization mediated by a helix-loop-helix-like domain of baculovirus IE1 is required for early promoter transactivation. J Virol **75:**6042-6051.

Ooi B. G., Rankin C. et Miller L. K. 1989. Downstream sequences augment transcription from the essential initiation site of a baculovirus polyhedrin gene. J Mol Biol **210:**721-736.

Op De Beeck A. et Caillet-Fauquet P. 1997. Viruses and the cell cycle. Progress in cell cycle research **3:**1-19.

Page R. D. 1996. TreeView: an application to display phylogenetic trees on personal computers. Comput Appl Biosci **12:**357-358.

Payne C. C. 1986. Insect pathogenic viruses as pest control agents, p. 183-200. *In* Franz J. M. (ed.), Biological and plant health protection. Fisher-Verlag, Stuttgart.

Pearson G. et Rohrmann G. 1997. Splicing is required for transactivation by the immediate early gene 1 of the *Lymantria dispar* Multinucleocapsid Nuclear Polyhedrosis Virus. Virology **235:**153-165.

Pearson M. N. et Rohrmann G. F. 1995. *Lymantria dispar* nuclear polyhedrosis virus homologous regions: characterization of their ability to function as replication origins. J Virol **69:**213-221.

Perkins K. K., Admon A., Patel N. et Tjian R. 1990. The *Drosophila* Fos-related AP-1 protein is a developmentally regulated transcription factor. Genes Dev **4:**822-834.

Pogue-Geile K., Geiser J. R., Shu M., Miller C., Wool I. G., Meisler A. I. et Pipas J. M. 1991. Ribosomal protein genes are overexpressed in colorectal cancer: isolation of a cDNA clone encoding the human S3 ribosomal protein. Mol Cell Biol **11:**3842-3849.

Pommet J. 2000. Effets pléiotropes d'un récepteur à l'acide rétinoïque humain dans un système Baculovirus/cellules d'insecte. Thèse: Biologie cellulaire, 198 p. Montpellier II.

Powell J. R. et Moriyama E. N. 1997. Evolution of codon usage bias in *Drosophila*. Proc Natl Acad Sci U S A **94:**7784-7790.

Pullen S. S. et Friesen P. D. 1995. 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.

Ramakrishnan V. et White S. W. 1998. Ribosomal protein structures: insights into the architecture, machinery and evolution of the ribosome. Trends Biochem Sci **23:**208-212.

Ramakrishnan V. 2002. Ribosome structure and the mechanism of translation. Cell **108:**557-572.

Regier J. C., Fang Q. Q., Mitter C., Peigler R. S., Friedlander T. P. et Solis M. A. 1998. Evolution and phylogenetic utility of the period gene in Lepidoptera. Mol Biol Evol **15:**1172-1182.

Ribeiro B. M., Hutchinson K. et Miller L. K. 1994. A mutant baculovirus with a temperature-sensitive IE-1 transregulatory protein. J Virol **68:**1075-1084.

Rodems S. M. et Friesen P. D. 1993. The hr5 transcriptional enhancer stimulates early expression from the *Autographa californica* nuclear polyhedrosis virus genome but is not required for virus replication. J Virol **67:**5776-5785.

Rodems S. M. et Friesen P. D. 1995. Transcriptional enhancer activity of hr5 requires dual-palindrome half sites that mediate binding of a dimeric form of the baculovirus transregulator IE1. J Virol **69:**5368-5375.

Rodems S. M., Pullen S. S. et Friesen P. D. 1997. DNA-dependent transregulation by IE1 of *Autographa californica* Nuclear Polyhedrosis Virus: IE1 domains required for transactivation and DNA binding. Journal of Virology **71:**9270-9277.

Rourke I. J., Hansen T. V., Nerlov C., Rehfeld J. F. et Nielsen F. C. 1999. Negative cooperativity between juxtaposed E-box and cAMP/TPA responsive elements in the cholecystokinin gene promoter. FEBS Lett **448:**15-18.

Rout M. P., Blobel G. et Aitchison J. D. 1997. A distinct nuclear import pathway used by ribosomal proteins. Cell **89:**715-725.

Ryseck R. P., Kovary K. et Bravo R. 1990. Integrity of FOS B leucine zipper is essential for its interaction with JUN proteins. Oncogene **5:**1091-1093.

Saeboe-Larssen S., Lyamouri M., Merriam J., Oksvold M. P. et Lambertsson A. 1998. Ribosomal protein insufficiency and the minute syndrome in *Drosophila*: a dose-response relationship. Genetics **148:**1215-1224.

Sandford G. R. et Burns W. H. 1996. Rat cytomegalovirus has a unique immediate early gene enhancer. Virology **222:**310-317.

Sanyal S., Sandstrom D. J., Hoeffer C. A. et Ramaswami M. 2002. AP-1 functions upstream of CREB to control synaptic plasticity in *Drosophila*. Nature **416:**870-874.

Scheer U. et Hock R. 1999. Structure and function of the nucleolus. Curr Opin Cell Biol **11:**385-390.

Schumacher M. A., Goodman R. H. et Brennan R. G. 2000. The structure of a CREB bZIP.somatostatin CRE complex reveals the basis for selective dimerization and divalent cation-enhanced DNA binding. J Biol Chem **275:**35242-35247.

Sharp P. M. et Li W. H. 1987. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res **15:**1281-1295.

Shaywitz A. J. et Greenberg M. E. 1999. CREB: a stimulus-induced transcription factor activated by a diverse array of extracellular signals. Annu Rev Biochem **68:**821-861.

Shields D. C., Sharp P. M., Higgins D. G. et Wright F. 1988. "Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. Mol Biol Evol **5:**704-716.

Silva A. J., Kogan J. H., Frankland P. W. et Kida S. 1998. CREB and memory. Annu Rev Neurosci **21:**127-148.

Smolik S. M., Rose R. E. et Goodman R. H. 1992. A cyclic AMP-responsive element-binding transcriptional activator in *Drosophila melanogaster*, dCREB-A, is a member of the leucine zipper family. Mol Cell Biol **12:**4123-4131.

Sprague K. U., Hagenbuchle O. et Zuniga M. C. 1977. The nucleotide sequence of two silk gland alanine tRNAs: implications for fibroin synthesis and for initiator tRNA structure. Cell **11:**561-570.

Srinivasan L. et Gopinathan K. P. 2001. Differential expression of individual gene copies from within a tRNA multigene family in the mulberry silkworm *Bombyx mori*. Insect Mol Biol **10:**523-530.

Struhl K. 1989. Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins. Trends Biochem Sci **14:**137-140.

Struthers R. S., Vale W. W., Arias C., Sawchenko P. E. et Montminy M. R. 1991. Somatotroph hypoplasia and dwarfism in transgenic mice expressing a non- phosphorylatable CREB mutant. Nature **350:**622-624.

Sueoka N. 1962. On the genetic basis of variation and heterogeneity of DNA base composition. Proc Natl Acad Sci U S A **34:**95-114.

Suzuki T., Terasaki M., Takemoto-Hori C., Hanada T., Ueda T., Wada A. et Watanabe K. 2001. Structural compensation for the deficit of rRNA with proteins in the mammalian mitochondrial ribosome. Systematic analysis of protein components of the large ribosomal subunit from mammalian mitochondria. J Biol Chem **276:**21724-21736.

Theilmann D. A. et Stewart S. 1991. Identification and characterization of the IE-1 gene of Orgyia pseudotsugata multicapsid nuclear polyhedrosis virus. Virology **180:**492-508.

Thompson J. D., Gibson T. J., Plewniak F., Jeanmougin F. et Higgins D. G. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res **25:**4876-4882.

Tjian R. et Maniatis T. 1994. Transcriptional activation: a complex puzzle with few easy pieces. Cell **77:**5-8.

Uechi T., Tanaka T. et Kenmochi N. 2001. A complete map of the human ribosomal protein genes: assignment of 80 genes to the cytogenetic map and implications for human disorders. Genomics **72:**223-230.

Usui T., Smolik S. M. et Goodman R. H. 1993. Isolation of *Drosophila* CREB-B: a novel CRE-binding protein. DNA Cell Biol **12:**589-595.

Vallejo M. 1994. Transcriptional control of gene expression by cAMP-response element binding proteins. J Neuroendocrinol **6:**587-596.

van Oers M. M., Flipsen J. T., Reusken C. B., Sliwinsky E. L., Goldbach R. W. et Vlak J. M. 1993. Functional domains of the p10 protein of *Autographa californica* nuclear polyhedrosis virus. J Gen Virol **74:**563-574.

Vaughn J. L., Goodwin R. H., Tompkins G. J. et McCawley P. 1977. The establishment of two cell lines from the insect *Spodoptera frugiperda* (Lepidoptera; Noctuidae). In Vitro **13:**213-217.

Velculescu V. E., Zhang L., Zhou W., Vogelstein J., Basrai M. A., Bassett D. E., Jr., Hieter P., Vogelstein B. et Kinzler K. W. 1997. Characterization of the yeast transcriptome. Cell **88:**243-251.

Vo N. et Goodman R. H. 2001. CREB-binding protein and p300 in transcriptional regulation. J Biol Chem **276:**13505-13508.

Wang Q., Yang C., Zhou J., Wang X., Wu M. et Liu Z. 2001. Cloning and characterization of full-length human ribosomal protein L15 cDNA which was overexpressed in esophageal cancer. Gene **263:**205-209.

Warner J. R. et Nierras C. R. 1998. Trapping human ribosomal protein genes. Genome Res **8:**419-421.

Warner J. R. 1999. The economics of ribosome biosynthesis in yeast. Trends Biochem Sci **24:**437-440.

Williams J. S. et Andrisani O. M. 1995. The hepatitis B virus X protein targets the basic region-leucine zipper domain of CREB. Proc Natl Acad Sci U S A **92:**3819-3823.

Wodicka L., Dong H., Mittmann M., Ho M. H. et Lockhart D. J. 1997. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. Nat Biotechnol **15:**1359-1367.

Wolfe K. H. et Shields D. C. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. Nature **387:**708-713.

Wood H. A. et Granados R. R. 1991. Genetically engineered baculoviruses as agents for pest control. Annu Rev Microbiol **45:**69-87.

Wool I. G., Chan Y. L. et Gluck A. 1995. Structure and evolution of mammalian ribosomal proteins. Biochem Cell Biol **73:**933-947.

Wool I. G. 1996. Extraribosomal functions of ribosomal proteins. Trends Biochem Sci **21:**164-165.

Wright F. 1990. The 'effective number of codons' used in a gene. Gene **87:**23-29.

Xie W. D., Arif B., Dobos P. et Krell P. J. 1995. Identification and analysis of a putative origin of DNA replication in the *Choristoneura fumiferana* multinucleocapsid nuclear polyhedrosis virus genome. Virology **209:**409-419.

Yamamoto K. K., Gonzalez G. A., Biggs W. H., 3rd et Montminy M. R. 1988. Phosphorylation-induced binding and transcriptional efficacy of nuclear factor CREB. Nature **334:**494-498.

Yang C. et Sehnal F. 1998. Ribosomal protein L5 of *Bombyx mori*: cDNA sequence and tissue differences in the L5 mRNA content. Insect Mol Biol **7:**301-306.

Yasukochi Y. 1998. A dense genetic map of the silkworm, *Bombyx mori*, covering all chromosomes based on 1018 molecular markers. Genetics **150:**1513-1525.

Yeates D. K. et Wiegemann B. M. 1999. Congruency and controversy: toward a higher-level phylogeny of Diptera. Annual Review of Entomology **44:**397-428.

Yin J. C., Del Vecchio M., Zhou H. et Tully T. 1995a. CREB as a memory modulator: induced expression of a dCREB2 activator isoform enhances long-term memory in *Drosophila*. Cell **81:**107-115.

Yin J. C., Wallach J. S., Wilder E. L., Klingensmith J., Dang D., Perrimon N., Zhou H., Tully T. et Quinn W. G. 1995b. A *Drosophila* CREB/CREM homolog encodes multiple isoforms, including a cyclic AMP-dependent protein kinase-responsive transcriptional activator and antagonist. Mol Cell Biol **15:**5123-5130.

Yoo S. et Guarino L. A. 1994. The *Autographa californica* nuclear polyhedrosis virus ie2 gene encodes a transcriptional regulator. Virology **202:**746-753.

Yoshihama M., Uechi T., Asakawa S., Kawasaki K., Kato S., Higa S., Maeda N., Minoshima S., Tanaka T., Shimizu N. et Kenmochi N. 2002. The human ribosomal protein genes: sequencing and comparative analysis of 73 genes. Genome Res **12:**379-390.

Zhang K., Chaillet J. R., Perkins L. A., Halazonetis T. D. et Perrimon N. 1990. *Drosophila* homolog of the mammalian jun oncogene is expressed during embryonic development and activates transcription in mammalian cells. Proc Natl Acad Sci U S A **87:**6281-6285.

Zhang L., Zhou W., Velculescu V. E., Kern S. E., Hruban R. H., Hamilton S. R., Vogelstein B. et Kinzler K. W. 1997. Gene expression profiles in normal and cancer cells. Science **276:**1268-1272.

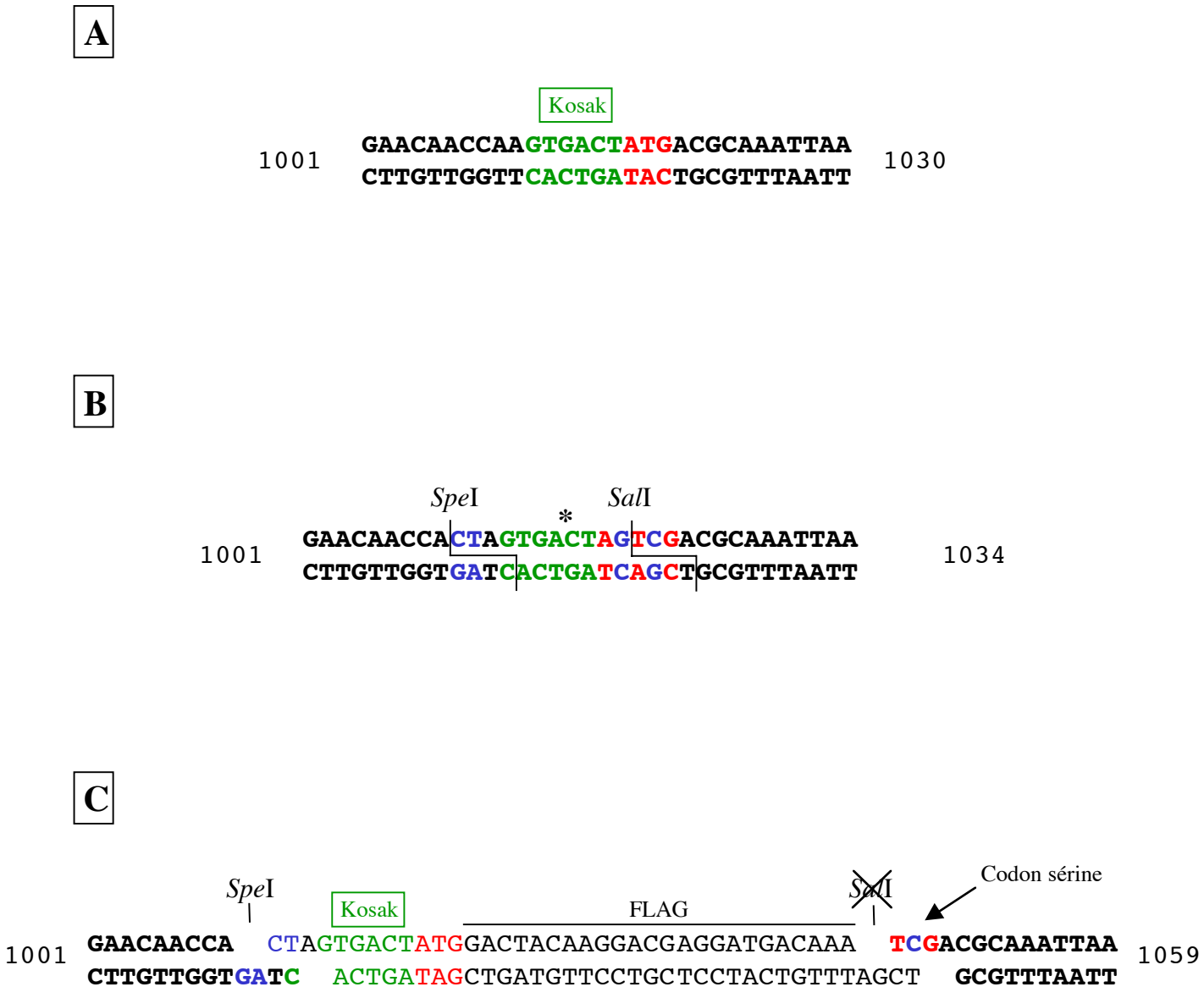


Figure 1. Construction du plasmide pIE1-SpeSal et clonage de la séquence codante du peptide FLAG en fusion avec celle de ie1. A) Séquence du plasmide pIE1 natif autour du codon initiateur de la traduction (en rouge), avec la séquence de Kosak (en vert). Les chiffres indiquent la position du fragment dans le plasmide. B) La même région du plasmide pIE1-SpeSal: 4 nucléotides (en bleu) ont été insérés, générant les sites de restriction *Spe*I et *Sal*I, dont la coupure est matérialisée par des traits. L'astérisque indique la position du deuxième site *Spe*I créé par ces insertions. C) Construction du plasmide pFLAG-IE1, par clonage dans pIE1-SpeSal de la séquence codante du peptide FLAG avec une séquence de Kosak et un ATG. Le site SpeI est conservé, Le site SalI perdu. La flèche indique le codon surnuméraire (sérine) généré par le clonage.

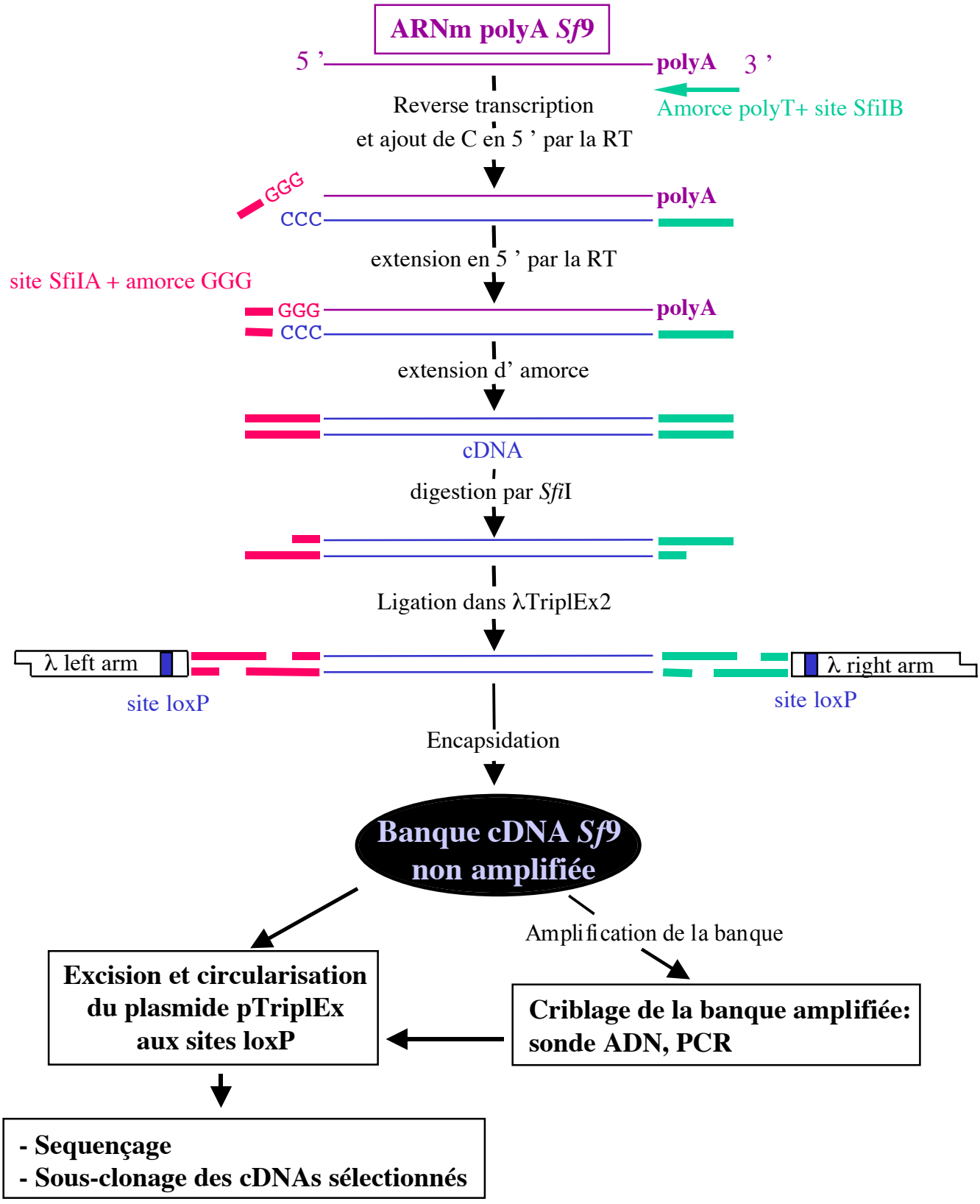
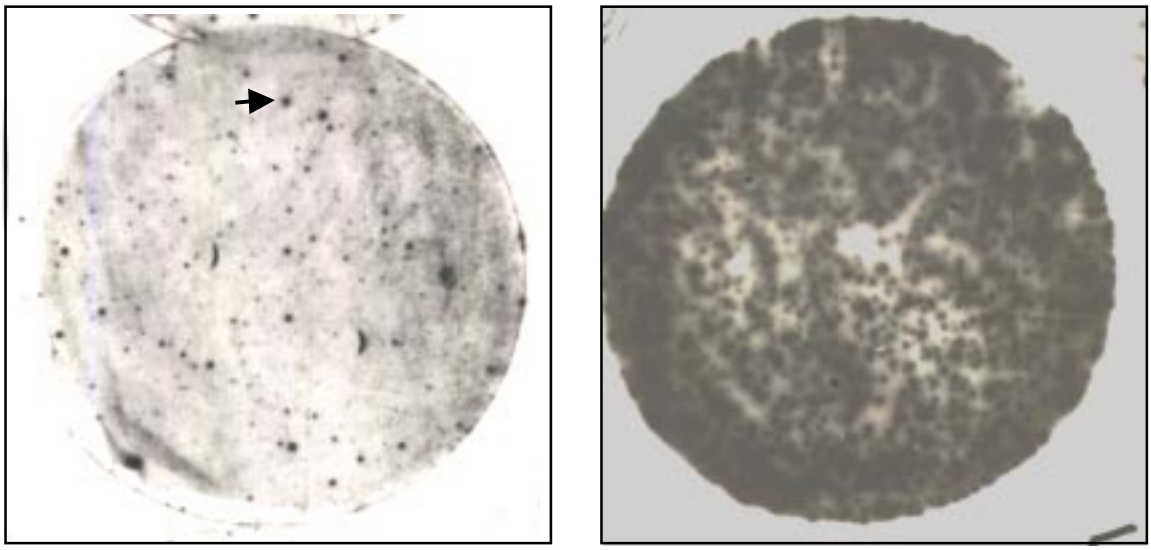


Figure 2. Construction et applications de la banque cDNA *Sf*9 (kit Clontech)



A B

Figure 3. Purification d' un clone issu de la banque cDNA Sf9. A) hybridation d 'une membrane initiale (premier tour de criblage, environ 50000 plages par membrane) avec une sonde hsp90. Chaque tache circulaire correspond à une plage de lyse dont les phages hybrident avec la sonde. La plage de lyse correspondant au signal d'hybridation indiqué par la flèche a été carottée et soumise à plusieurs cycles supplémentaires de purification. B) Troisième cycle de purification : toutes les plages de lyse présentent un signal positif. Plusieurs d 'entre elles vont être carottées et les phages qu'elles contiennent convertis en plasmide pour séquençage.

Figure 4. Larve de Spodoptera frugiperda au dernier stade de l'infection par le baculovirus AcMNPV. Le liquide qu'elle exsude contient une quantité énorme de polyèdres qui vont assurer la dispersion et la transmission horizontale du virus. C'est la consistance et la couleur de ce liquide qui ont donné son nom à la maladie: la grasserie.

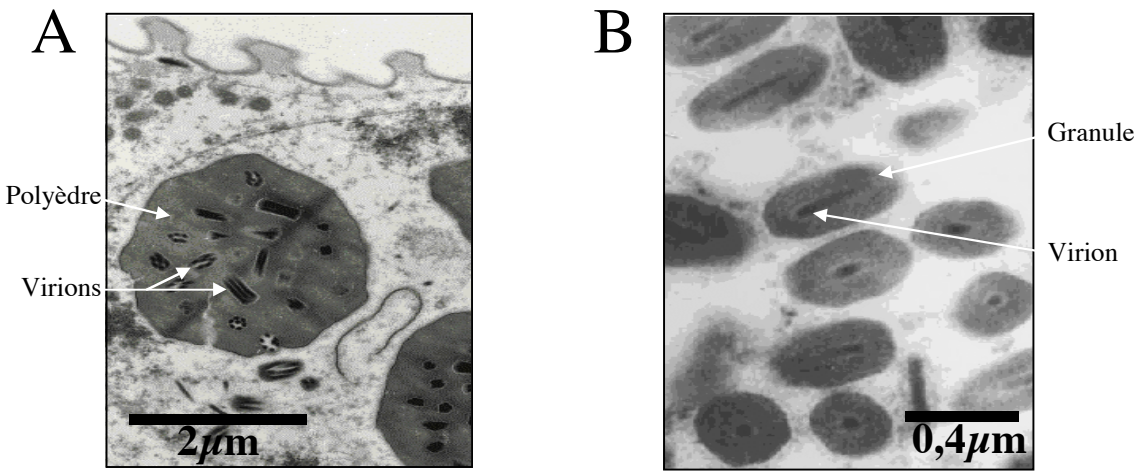


Figure 5. Les deux genres de la famille Baculoviridae. A) virus de polyédroses nucléaires ou nucleopolyhedrovirus. Les corps d'inclusion (les polyèdres, composés de polyédrine) contiennent plusieurs virions. B) Virus de granuloses ou granulovirus. Les corps d'inclusion (les granules, composés de granuline) sont plus petits et ne contiennent qu'un seul virion (d'après M. Lopez-Ferber).

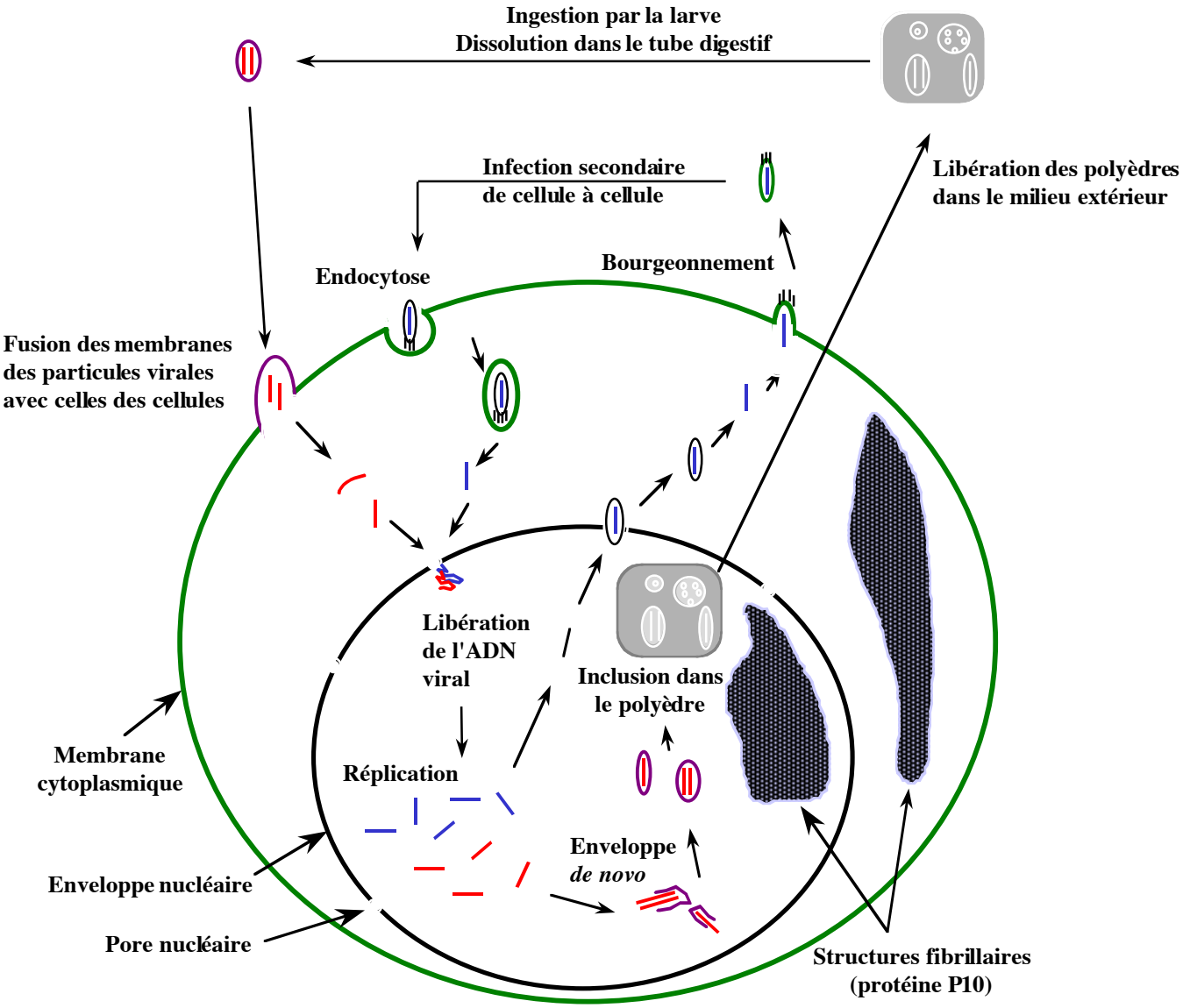


Figure 6. Cycle de multiplication d'un baculovirus. Les virions occlus (ainsi que leur ADN) sont de couleur rouge, les virions non occlus (et leur ADN) sont en bleu. La membrane plasmique et l'enveloppe des virions non occlus sont en vert, celle des virions occlus en violet. Pour les détails du cycle viral, voir le texte.

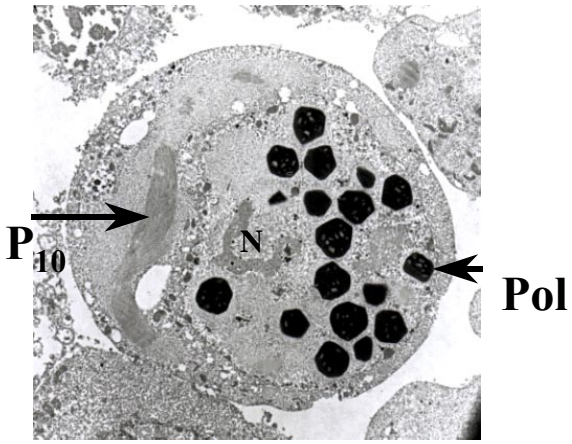


Figure 7. Photographie en microscopie électronique d'une cellule Sf9 infectée par le baculovirus AcMNPV (48h p.i.).
N, noyau; P_{10}, protéine P_{10} formant des amas fibrillaires; Pol, protéine polyédrine formant des polyèdres dans lesquels on distingue des virions.

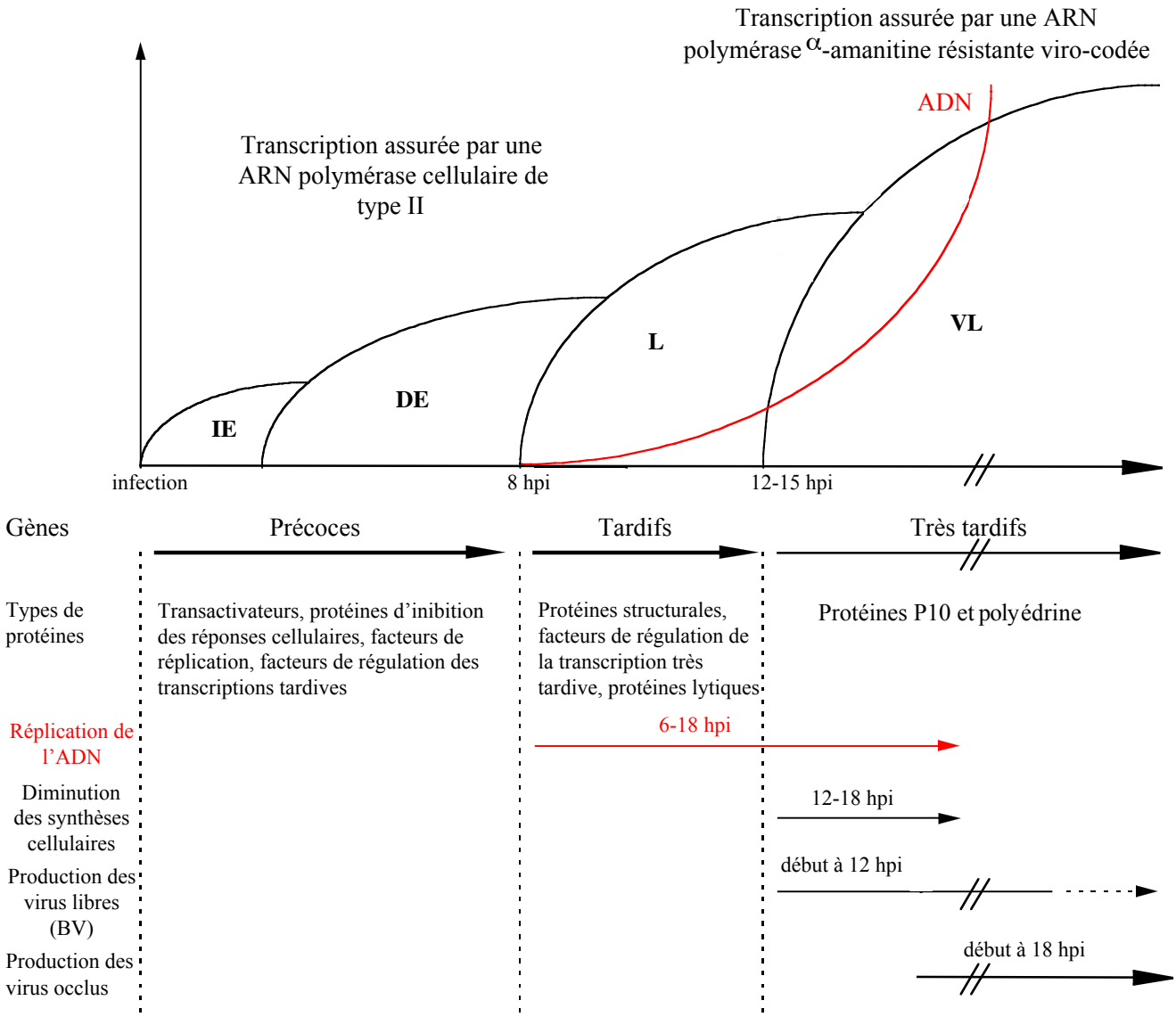


Figure 8. Représentation schématique de l 'expression en cascade des gènes d 'un baculovirus (4 phases: IE, immediate early, DE, delayed early, L, late, VL, very late) et des principaux évènements de l 'infection virale (réplication, contrôle des synthèses cellulaires, production des deux types de virions). hpi, heures post-infection; BV, budded virion (d 'après M. Lopez-Ferber).

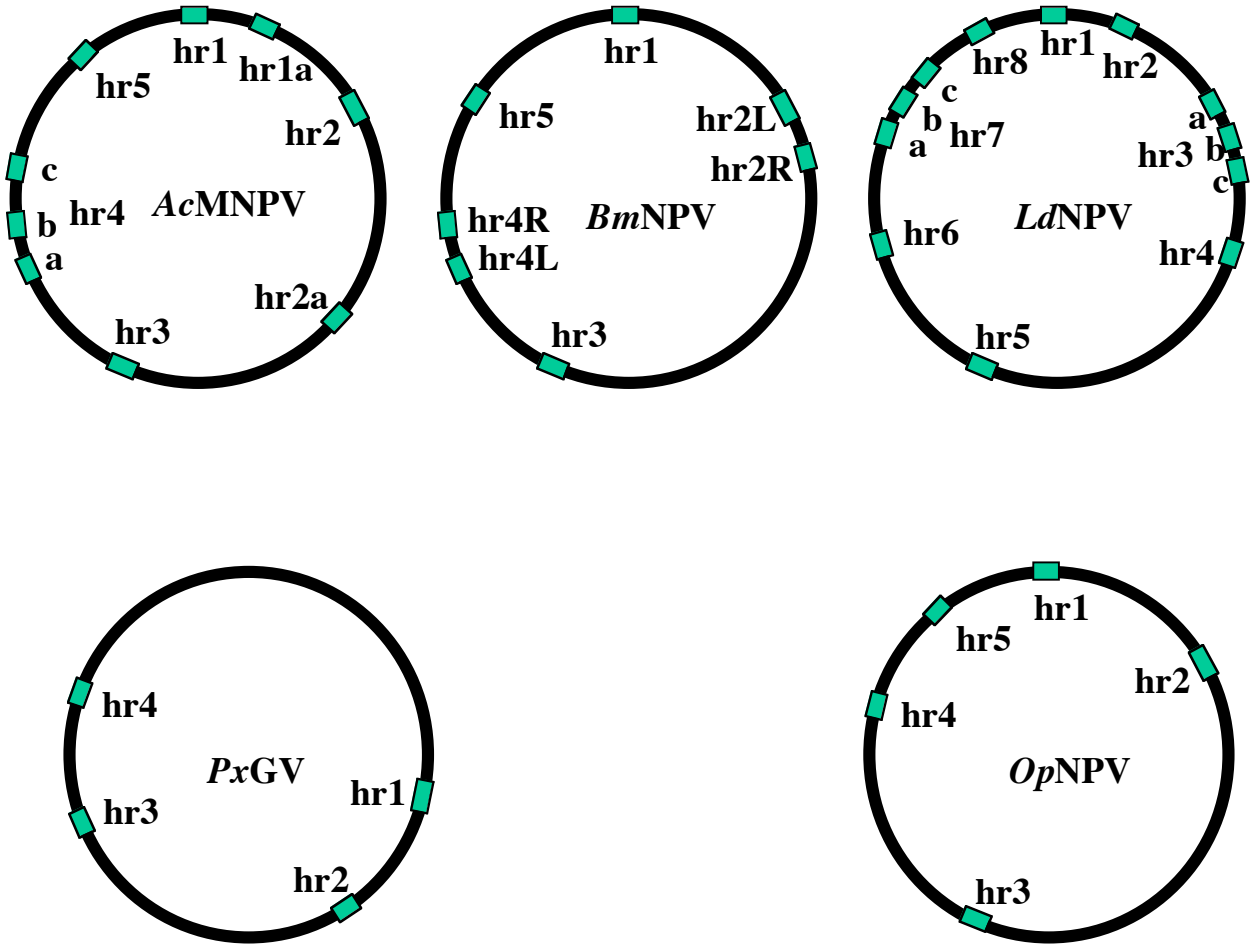


Figure 9. Représentation schématique des régions homologues (hr) présentes dans le génome de 5 baculovirus de lépidoptères. GV, granulovirus; MNPV, multinucléopolyhedrosis virus; *Ac*, *Autographa californica*; *Bm*, *Bombyx mori*; *Ld*, *Lymantria dispar*; *Px*, *Plutella xylostella*; *Op*, *Orgya pseudotsugata*.

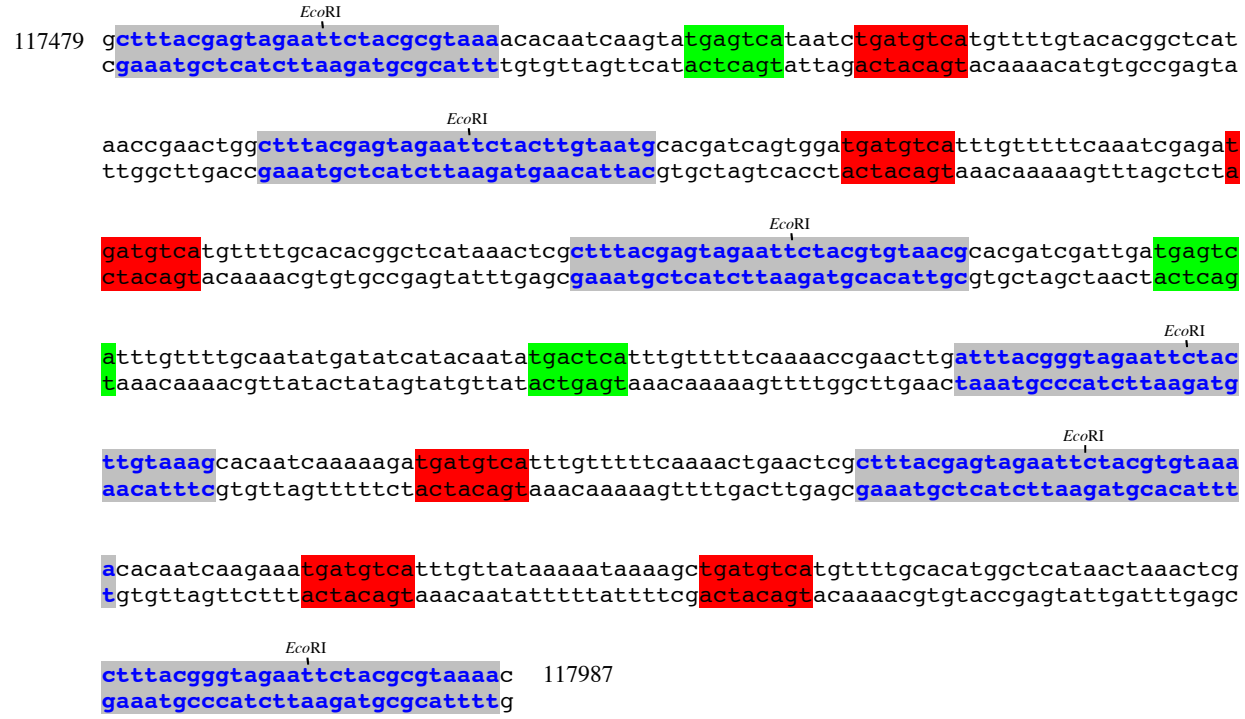
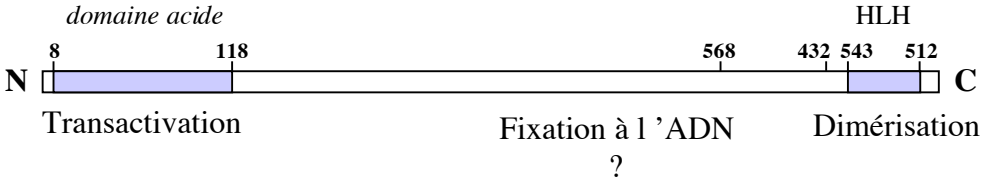


Figure 10. Visualisation des sites de fixation pour des facteurs de transcription dans la séquence de la région homologue hr5: 1) les palindromes imparfaits fixant le transactivateur viral IE1 sont en bleu surligné gris, chacun étant bissecté par un site *Eco*RI, 2) les sites TRE (TGA(C/G)TCA sont surlignés en vert, et 3) les sites CRE (TGATGTCA) sont surlignés en rouge (Ogliastro M., 1995). Les chiffres de part et d'autre de la séquence correspondent à la position de hr5 dans le génome du baculovirus *Ac*MNPV, souche C6 (Ayres et al., 1994).

A



B

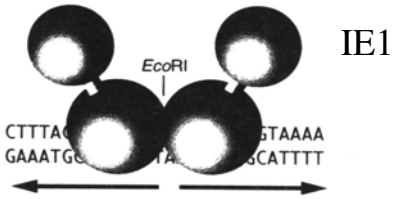


Figure 11. A) Représentation schématique des domaines fonctionnels de IE1. La protéine (582 acides aminés) comporte une région impliquée dans l 'activation transcriptionnelle (domaine riche en résidus acides, aa 8-118) et une région importante pour la dimérisation, qui a une structure de type hélice-boucle-hélice (HLH). Le domaine de fixation à l 'ADN n 'a pas encore été déterminé. Les résidus 432 et 512, dont la mutation rend IE1 sensible à la température, sont indiqués. N et C, extrémités amino- et carboxy-terminales (d 'après Rodems *et al.*, 1997). B) modèle de la fixation de IE1 sur son site de reconnaissance palindromique dans les hrs: chaque monomère peut interagir avec un demi-palindrome, mais l'activité transactivatrice de IE1 nécessite sa fixation sous forme dimérique. *Eco*RI, site enzymatique bissectant la séquence palindromique (d 'après Friesen, 1997).

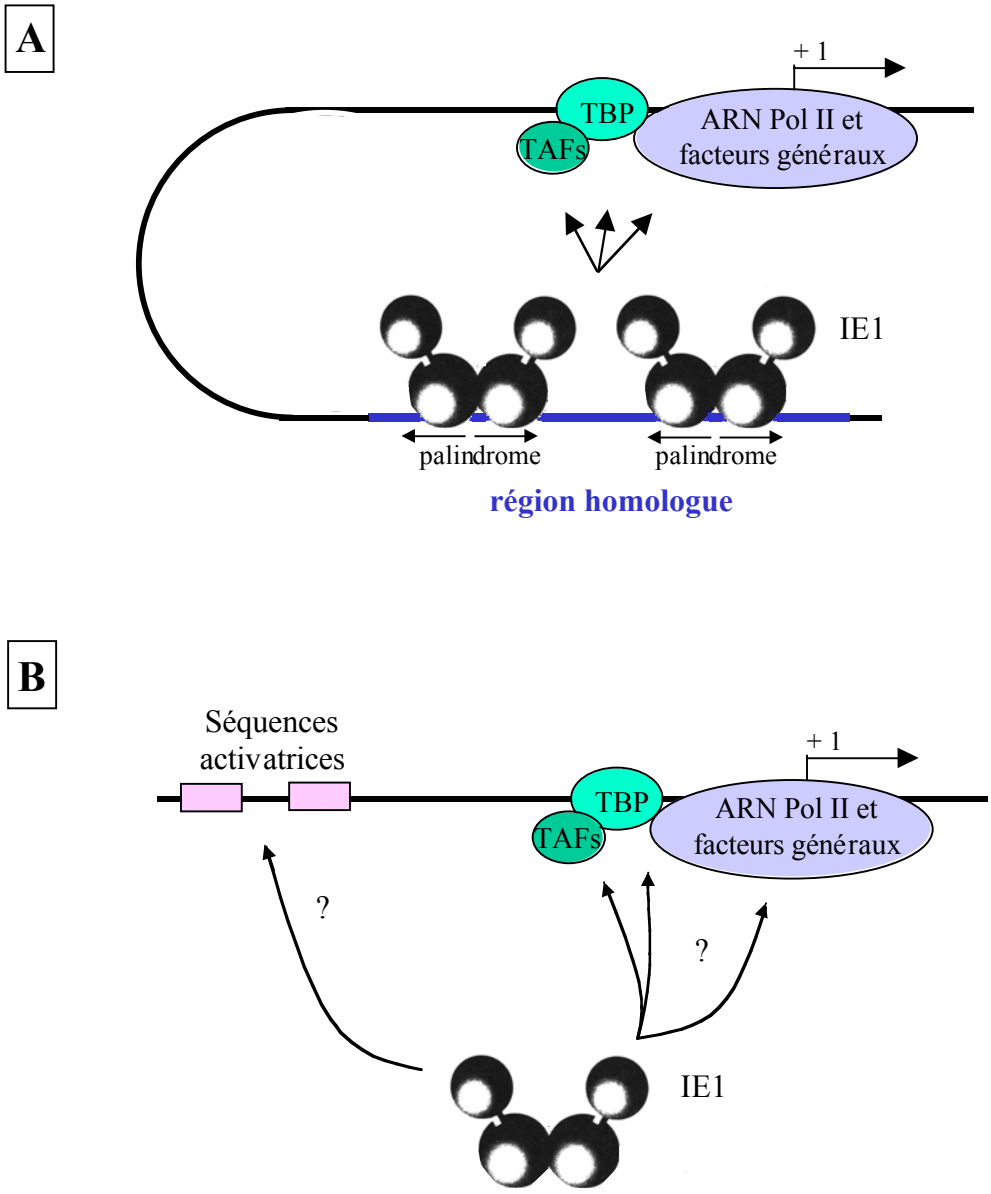
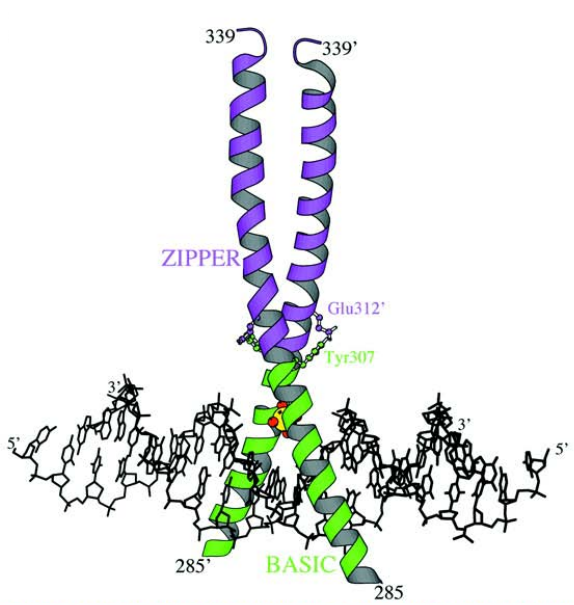


Figure 12. Modèles proposés pour l'activation transcriptionnelle par IE1. A) Transactivation dépendante des hrs: la fixation du dimère IE1 sur les palindromes présents dans une région homologue distale permettrait l'activation de la machinerie transcriptionnelle sur le promoteur cible, en facilitant l'assemblage du complexe d'initiation ou en stimulant des facteurs de l'hôte associés à la TBP (TAFs). B) Transactivation indépendante des hrs: en l 'absence de fixation à l 'ADN, IE1 pourrait stimuler directement le complexe d'initiation de la transcription, en interagissant avec des facteurs de transcription fixés sur des séquences activatrices en amont du promoteur ou bien avec les facteurs généraux au niveau du promoteur (d 'après Friesen, 1997).

A



B

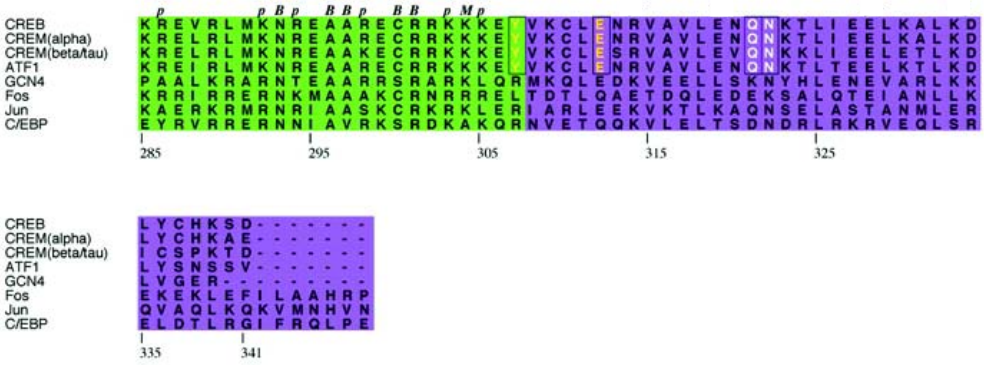


Figure 13. A) Structuration dans l'espace des régions bZIP du dimère CREB fixé sur l 'ADN. Les régions basiques sont colorées en vert et les glissières de leucine en magenta. Les chaînes latérales des résidus Tyrosine 307 et Glutamine 312, qui établissent des liaisons hydrogène entre les deux hélices, sont indiquées. L 'ion magnésium hexahydraté, qui occupe la cavité entre les régions basiques et l 'ADN, est également schématisé.
B) Alignement des séquences protéiques de la région bZIP de quelques membres de la superfamille des facteurs de transcription bZIP. ***p***, résidus établissant des contacts avec les phosphates de l 'ADN; ***B***, résidus établissant des contacts spécifiques avec les bases de l 'ADN; ***M*** , résidus en contact avec le magnésium hexahydraté. Les résidus Tyr 307 et Glu 312 sont colorés en jaune; en blanc, autres résidus établissant des contacts interhélices.

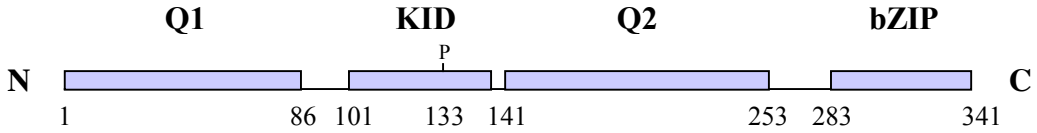


Figure 14. Position (en acides aminés) des principaux domaines de la protéine CREB humaine: Q1 et Q2, régions riches en glutamines; KID, domaine inductible par kinase; bZIP, domaine basique à glissière de leucine; P, site de phosphorylation par la protéine kinase A (sur la sérine 133). N, C, extrémités amino- et carboxy-terminales.

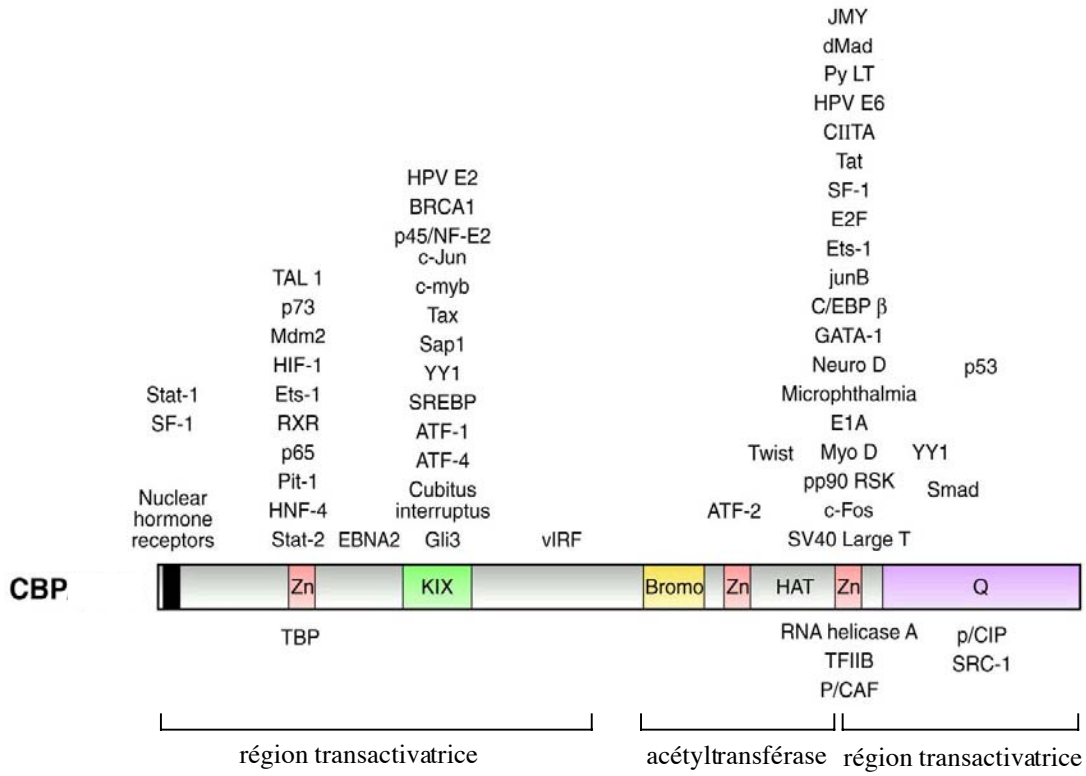


Figure 15. Organisation modulaire de CBP et interactions avec d'autres protéines. CBP possède 3 structures en doigt de zinc (Zn) riches en cystéines et histidines, un domaine d'interaction avec CREB (KIX), un bromodomaine (Bromo) probablement impliqué dans la reconnaissance de résidus acétylés, un domaine HAT (histone acétyltransférase) et une région transactivatrice riche en glutamines. La protéine est donc composée de deux grands domaines transactivateurs de part et d 'autre d'une région centrale à activité acétyltransférase. Au-dessus du schéma sont mentionnés quelques activateurs transcriptionnels (cellulaires et viraux) connus pour interagir avec CBP; les facteurs de transcription généraux et les protéines à activité HAT possédant la même propriété sont indiqués sous la figure (d 'après Vo et Goiodman, 2001; Chan et La Thangue, 2001).

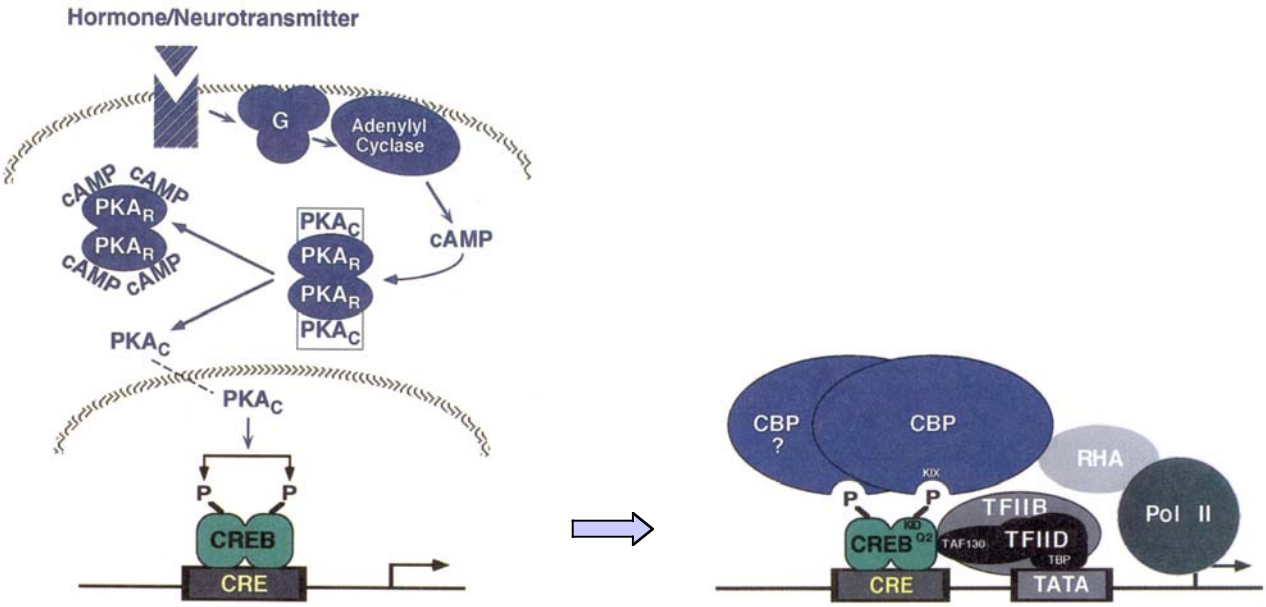


Figure 16. Activation de la transcription par la voie de transduction du signal AMPc/PKA, avec A) phosphorylation de CREB en réponse au signal extracellulaire, et B) recrutement de CBP sur CREB phosphorylé et activation de la transcription.

A: Les protéines G activées par un signal extracellulaire (ici une hormone ou un neurotransmetteur qui se fixe sur son récepteur membranaire) vont activer l'adénylate cyclase, permettant l'augmentation de la concentration intracellulaire d'AMPc. Celui-ci va se fixer sur les sous-unités régulatrices de la protéine kinase A (PKA_R), libérant les sous-unités catalytiques (PKA_C) actives qui vont passer dans le noyau et phosphoryler CREB au niveau du domaine KID. CREB est déjà fixée sur le site CRE présent dans le promoteur.

B: La forme phosphorylée de CREB est capable de recruter CBP au niveau de son domaine KIX. Cette protéine coactivatrice va interagir avec plusieurs facteurs de transcription généraux et avec l'ARN polymérase II, via l'ARN hélicase A (RHA). Par son domaine Q2, CREB interagit également avec plusieurs facteurs généraux, comme TAF130 de TFIID, et TFIIB. CREB et CBP vont ainsi faciliter la formation et stabiliser le complexe d'initiation (Shaywitz et Greenberg, 1999).

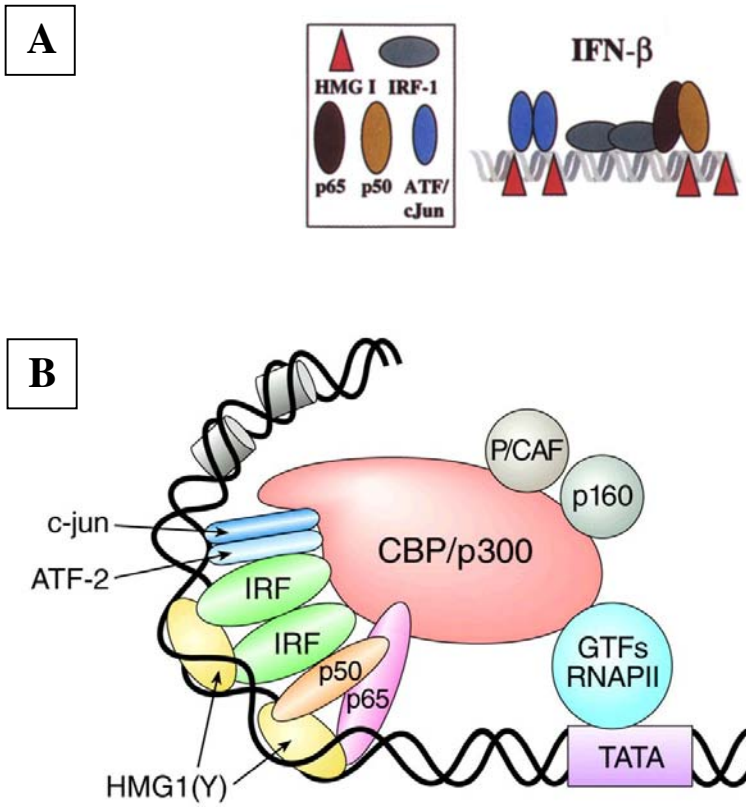


Figure 17. enhanceosome formé sur le promoteur du gène de l'interféron β. A) Disposition linéaire des facteurs de transcription (IRF1, p50, p65, cJun, ATF2) et protéines architecturales (HMG-1) fixés sur le promoteur. B) l'assemblage et l'interaction de ces facteurs de transcription crée une surface d'interaction stéréospécifique qui permet le recrutement de CBP/p300 et de la machinerie transcriptionnelle basale. GTF, facteurs de transcription généraux; RNAPII, ARN polymérase II; P/CAF et p160, coactivateurs.

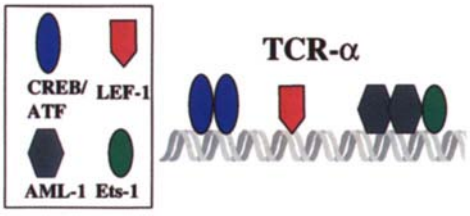
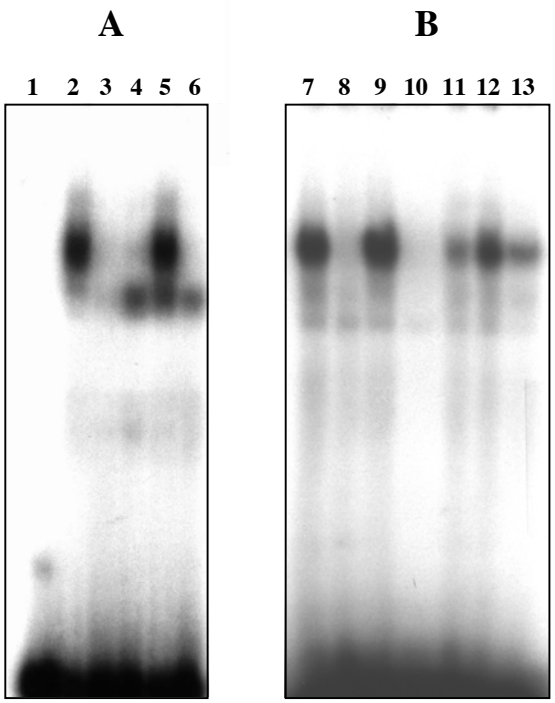


Figure 18. Autre exemple d'enhanceosome, formé sur le promoteur du gène TCRα: disposition linéaire des facteurs de transcription (CREB, ATF, AML-1, Ets-1) et protéines architecturales (LEF-1) fixés sur le promoteur.



1 pas d 'extraits nucléaires
compétiteur:
2, 7 pas de compétiteur
3, 10 CRE: GRGA **TGATGTCA** TTTGTTTT
4, 13 TRE/EB: ATTGA **TGAGTCA** TTTGTTTT
5, 9 EBOX : **GATGTCA** TTTGTTTT
6, 8 CRE/EB: GRGA **TGATGTCA** TTT
11 TRE/CRE: CAAGTA **TGAGTCA** TAAGC
TGATGTCA TGTTTTGC
12 TRE: ATTGA **TGAGTCA** TTT

Figure 19. Expériences de gel retard avec des extraits nucléaires *Sf*9 et A) la sonde TRE/CRE, B) la sonde CRE/EB. Les compétiteurs utilisés (x100 en excès) et leur séquence sont indiqués sous la figure. Les motifs TRE et CRE sont indiqués en gras.

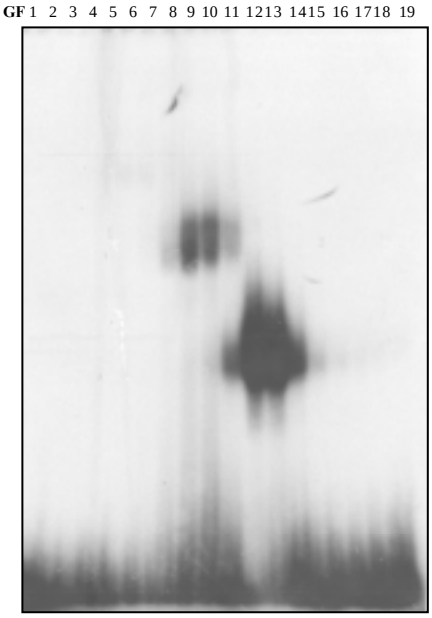


Figure 20. Expérience de gel retard réalisée avec la sonde CRE/EB et les fractions issues du passage d'extraits nucléaires *Sf*9 sur colonne de gel filtration (GF).

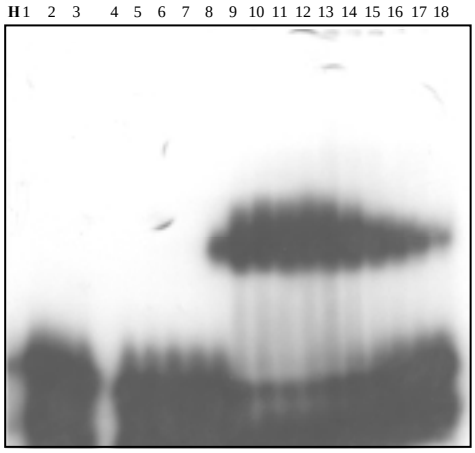


Figure 21. Les fractions 12-13 issues de la gel filtration précédente ont été passées sur colonne d'héparine (H), et les fractions résultantes testées en gel retard avec la sonde CRE/EB. Puits 1-3, « flow-through »; Puits 4-18, élution de la colonne avec une concentration croissante (0 à 2M) de NaCl.

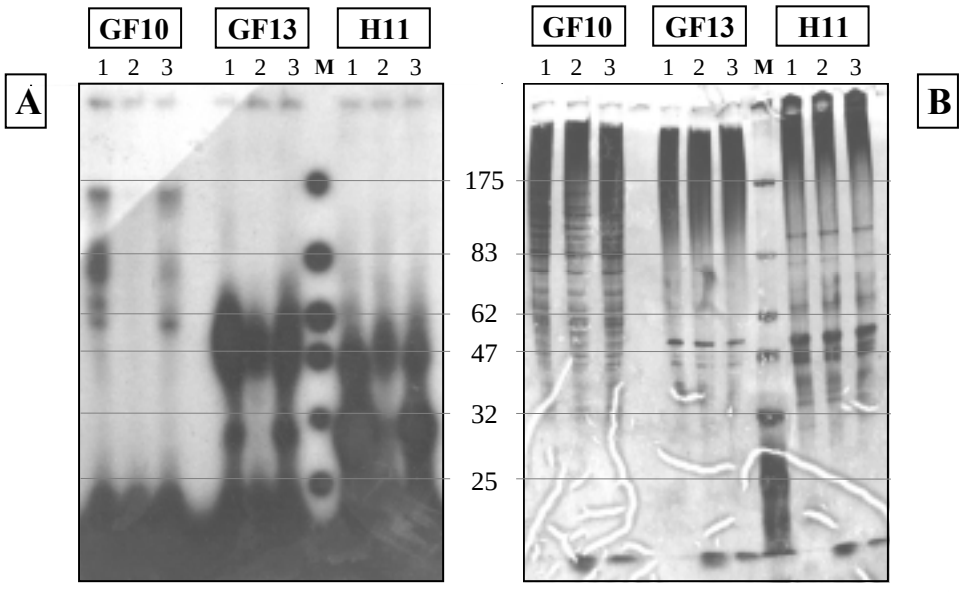
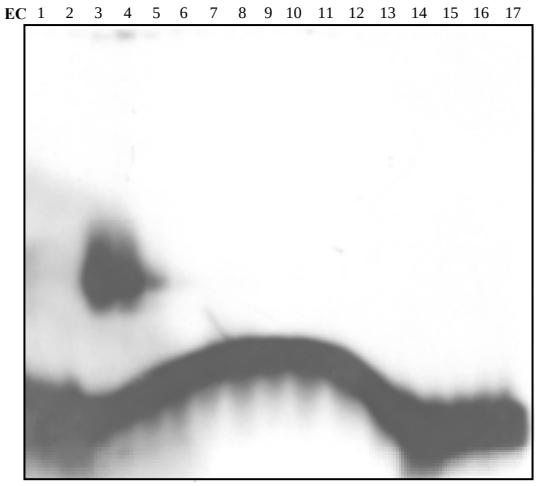
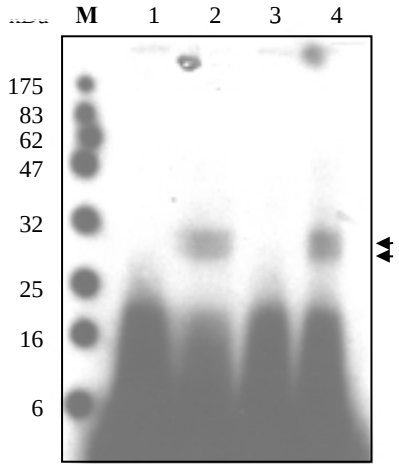


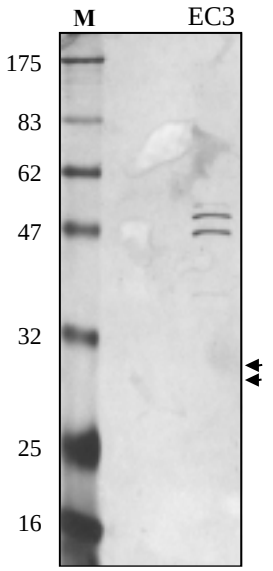
Figure 22. A) Expériences d 'UV-crosslinking avec la sonde CRE/EB et différentes fractions issues des colonnes de gel filtration (GF10, GF13) et d 'héparine (H11), et B) coloration au nitrate d 'argent du même gel.
1, pas de compétiteur; 2, compétiteur CRE/EB; 3, compétiteur CREm3 (pour les séquences, voir la figure 2 de l 'article1). M, marqueur de taille (les poids moléculaires sont mentionnés entre les deux panneaux de la figure).
Les lignes horizontales ont été placées pour faciliter la comparaison entre les deux images.



A



B



C

Figure 23. A) Les fractions H9-15 issues de la colonne d'héparine (voir figure 20) ont été passées sur colonne échangeuse de cations (EC), et les fractions résultantes testées en gel retard avec la sonde CRE/EB. Les protéines ont été éluées de la colonne avec une concentration croissante (0 à 1M) de NaCl.
B) Expérience d'UV crosslinking avec la sonde CRE/EB et la fraction EC3: 1, sonde seule; 2, pas de compétiteur; 3, compétiteur CRE/EB; 4, compétiteur CREm3 (pour les séquences, voir figure 2 de l 'article 1). Les flèches indiquent les deux protéines marqués. M: marqueur de taille (kDa).
C) coloration au nitrate d'argent de la fraction EC3 déposée sur gel SDS-PAGE 10%. Les flèches indiquent l 'emplacement de deux bandes très faiblement colorées (visibles à l 'œil, non visibles sur la figure) qui pourraient correspondre aux produits marqués visibles dans la figure précédente.

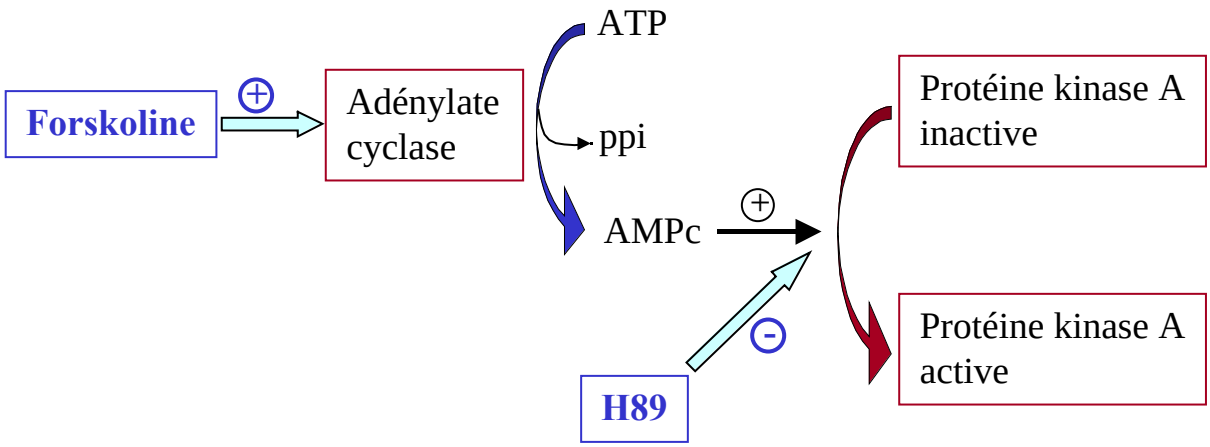


Figure 24. Activation (⊕) et inhibition (⊖) de la voie AMPc de transduction du signal par la forskoline et H89 (voir figure 15 pour le schéma de la voie AMPc complète).

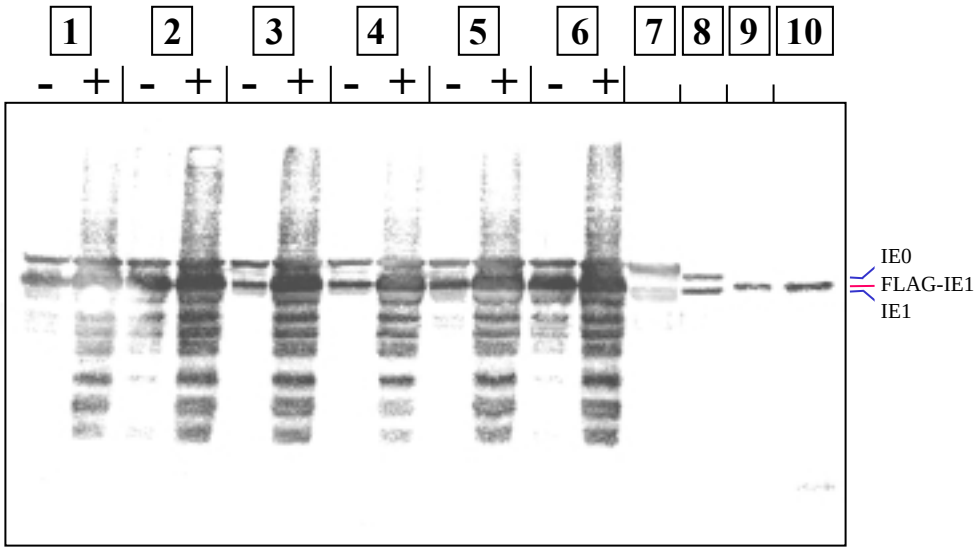


Figure 25. Expression de FLAG-IE1 en système bactérien. Le western blot a été effectué avec un anticorps polyclonal anti-IE1 (révélation à la phosphatase alcaline). 1 à 6, extraits protéiques totaux de 6 clones bactériens Xl1-Blue transformés par un plasmide d'expression pFLAG-IE1 non induits (-) ou induits (+) à l'IPTG; 7, extraits protéiques totaux de bactéries Xl1-Blue non induites; 8, extraits nucléaires de cellules *Sf*9 infectées par le virus AcMNPV; 8 et 9, purification des extraits 3+ et 6+ sur colonne d'affinité anti-FLAG. A droite de la figure, les traits bleus indiquent les bandes correspondant à IE1 et IE0 produits dans les cellules *Sf*9 infectées et le trait rouge la bande correspondant à FLAG-IE1 produit en système bactérien.

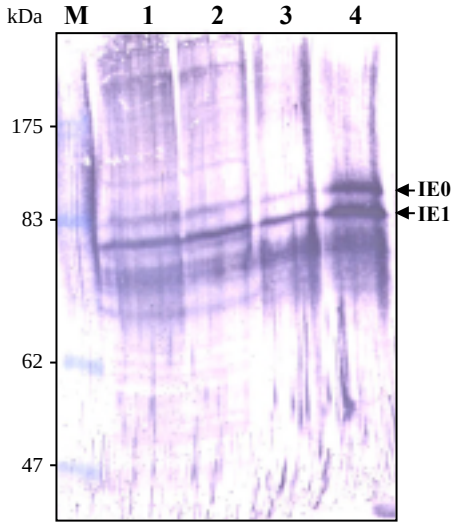


Figure 26. Co-immunoprécipitation IE1-CBP. 1,2x10^7 cellules Sf9 ont été infectées par un baculovirus *Ac*MNPV sauvage (wt, puits 1 et 3) ou recombinant exprimant la CBP humaine flaguée (*Ac*CBP, puits 2 et 4). Les extraits bruts (récupérés 48 hpi) ont été immunoprécipités avec des billes d'agarose couplées à un anticorps anti-FLAG. Après plusieurs lavages, les protéines fixées sur les billes ont été éluées et déposées sur gel SDS-PAGE. Le Western blot a été effectué avec un anticorps polyclonal anti-IE1 (révélation à la phosphatase alcaline). 1 et 2, extraits bruts wt et *Ac*CBP (10 µl); 3 et 4, immunoprécipité wt et *Ac*-CBP (à partir de 90 µl d 'extraits bruts). Les flèches à droite de la figure indiquent les bandes correspondant à IE0 et IE1. M, marqueur de tailles (en kDa).

A

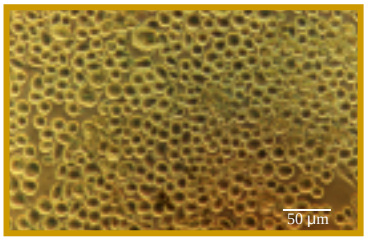
♂

B

♀

C

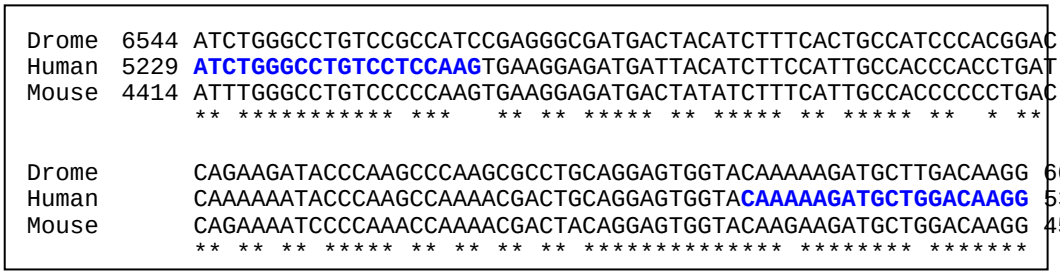
D



E

F

Figure 27. Le Lépidoptère Spodoptera frugiperda. A) et B) imago mâle (♂) et femelle (♀), C) larve; D) culture de cellules *Sf*9 sur boîte; E) et F), dégâts occasionnés par la larve sur maïs (feuilles et épis).



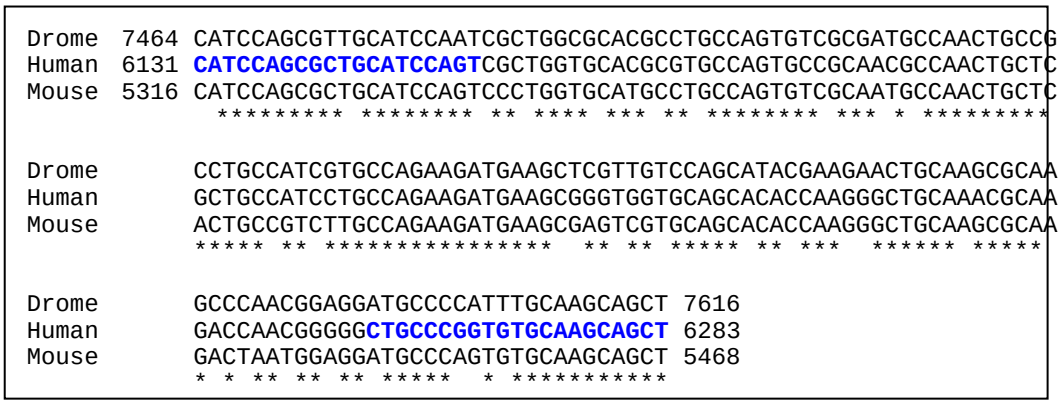


Figure 28. Alignement des séquences nucléotidiques de deux régions conservées du gène CBP provenant de trois espèces différentes: D. melanogaster (Drome), H. sapiens (Human) et Mus musculus (Mouse). Les chiffres indiqués au début et à la fin de chaque séquence correspondent à la position de la région dans le cDNA. Les nucléotides en caractères gras précisent la localisation des oligonucléotides (nommés CBP-Hum1 et CBP-Hum2) utilisés en PCR (voir texte).
Les numéros d'accession Genbank sont les suivants: Drome #U88570, Human U47751, Mouse S66385

A

gène hsp90 de *Bombyx mori*. 717 codons, ENC: 55

TTT	phe	F	6	TCT	ser	S	16	TAT	tyr	Y	11	TGT	cys	C	2
TTC	phe	F	17	TCC	ser	S	11	TAC	tyr	Y	16	TGC	cys	C	4
TTA	leu	L	7	TCA	ser	S	6	TAA	OCH	Z	1	TGA	OPA	Z	-
TTG	leu	L	10	TCG	ser	S	4	TAG	AMB	Z	-	TGG	trp	W	4
CTT	leu	L	19	CCT	pro	P	8	CAT	his	H	6	CGT	arg	R	9
CTC	leu	L	6	CCC	pro	P	4	CAC	his	H	7	CGC	arg	R	5
CTA	leu	L	5	CCA	pro	P	6	**CAA**	**gln**	**Q**	**9**	CGA	arg	R	3
CTG	leu	L	13	CCG	pro	P	4	**CAG**	**gln**	**Q**	**13**	CGG	arg	R	1
ATT	ile	I	18	ACT	thr	T	8	AAT	asn	N	9	AGT	ser	S	3
ATC	ile	I	27	ACC	thr	T	9	AAC	asn	N	22	AGC	ser	S	3
ATA	ile	I	-	ACA	thr	T	10	AAA	lys	K	34	AGA	arg	R	7
ATG	met	M	20	ACG	thr	T	4	AAG	lys	K	41	AGG	arg	R	6
GTT	val	V	16	GCT	ala	A	16	GAT	asp	D	29	GGT	gly	G	16
GTC	val	V	10	GCC	ala	A	7	GAC	asp	D	25	GGC	gly	G	9
GTA	val	V	9	GCA	ala	A	9	GAA	glu	E	51	GGA	gly	G	9
GTG	val	V	12	GCG	ala	A	6	GAG	glu	E	39	GGG	gly	G	-

B

gène hsp90 de Drosophila melanogaster. 718 codons, ENC: 32

TTT	phe	F	2	TCT	ser	S	9	TAT	tyr	Y	4	TGT	cys	C	-
TTC	phe	F	27	TCC	ser	S	22	TAC	tyr	Y	19	TGC	cys	C	5
TTA	leu	L	-	TCA	ser	S	1	TAA	OCH	Z	1	TGA	OPA	Z	-
TTG	leu	L	6	TCG	ser	S	7	TAG	AMB	Z	-	TGG	trp	W	4
CTT	leu	L	2	CCT	pro	P	3	CAT	his	H	1	CGT	arg	R	10
CTC	leu	L	7	CCC	pro	P	14	CAC	his	H	9	CGC	arg	R	16
CTA	leu	L	1	CCA	pro	P	3	**CAA**	**gln**	**Q**	**-**	CGA	arg	R	-
CTG	leu	L	45	CCG	pro	P	-	**CAG**	**gln**	**Q**	**23**	CGG	arg	R	-
ATT	ile	I	11	ACT	thr	T	9	AAT	asn	N	1	AGT	ser	S	-
ATC	ile	I	32	ACC	thr	T	22	AAC	asn	N	31	AGC	ser	S	11
ATA	ile	I	-	ACA	thr	T	3	AAA	lys	K	2	AGA	arg	R	-
ATG	met	M	20	ACG	thr	T	2	AAG	lys	K	75	AGG	arg	R	-
GTT	val	V	3	GCT	ala	A	11	GAT	asp	D	31	GGT	gly	G	12
GTC	val	V	21	GCC	ala	A	28	GAC	asp	D	33	GGC	gly	G	15
GTA	val	V	1	GCA	ala	A	2	GAA	glu	E	3	GGA	gly	G	6
GTG	val	V	19	GCG	ala	A	-	GAG	glu	E	73	GGG	gly	G	-

Figure 29. Tables d'usage des codons pour le gène hsp90 chez deux espèces d'insectes: A) *B. mori* (Lépidoptère) et B) *D. melanogaster* (diptère). Bien que la séquence en acides aminés de HSP90 soit très conservée entre ces deux organismes (Landais et al., 2001), la préférence de codons est très différente: chez *Bombyx*, tous les codons codant pour un acide aminé donné sont utilisés de manière globalement équivalente, alors que chez la drosophile l'un d'entre eux est fréquemment surreprésenté: comparer par exemple la glutamine (gln, surligné en bleu).
Le biais de codons global a été mesuré par la méthode de l'ENC (Effective Number of Codons; Wright, 1990), qui varie entre 20 (biais extrême) et 61 (pas de biais).

gène ou virus	séquence
Somatostatine	TTGGC**TGACGTCA**GAGAGAGAG
Glycoprotéine α (rat)	AAAC**TGA**T**GTCA**TGGTAA
PEPCK	GCCCC**T**T**ACGTCA**GAGGCGAGC
VIP	TACTG**TGACGTC**TTTCAGAGCA
Enképhaline	GGGCCTG**CGTCA**GC
fos	CCCAG**TGACGT**AGGA
Aromatase	TATGC**ACGTCA**CCCA
α Crystalline	ACCAGACT**GTCA**TCCC
BLV	GACAGA**GACGTCA**GCTG
HTLV-I	**TGACGTC**T
HTLV-II	GGCCC**TGACGTC**CCTCC

Tableau 1. Motifs CRE variants dans les promoteurs et les régions activatrices de gènes humains (sauf quand indiqué) et de virus (en italique). Les nucléotides correspondant au motif canonique sont en gras et surlignés en bleu.

Famille	Facteur	c-Jun	JunB	JunD	c-Fos	FosB	Fra-1	Fra-2	CREB	ATF2
AP1	c-Jun	+ ↑								
	JunB	+ ↑	+ ↑							
	JunD	+ ↑	+ ↑	+ ↑						
	c-Fos	+ ↑	+ ↑	+ ↑	-					
	FosB	+ ↑	+ ↑	+ ↑	-	-				
	Fra-1	+ ↓	+	+ ↑	-	-	-			
	Fra-2	+ ↓	+	+ ↑	-	-	-	-		
CREB/ATF	CREB	-			-				+ ↑	-
	CREM	-			-				+ ↓	
	ATF1	-			-				+ ↑	
	ATF2	+ ↑			±				-	+
	ATF3	+	+	+	-		-		-	+
	ATF4	+			+	+			-	
	ATF6									-
	ATFa	+	+	+	-					

+ : La dimérisation a été observée.
- : La dimérisation n'a pas été observée.
↑ : la dimérisation augmente l' affinité pour l'ADN.
↓ : la dimérisation diminue l'affinité pour l'ADN.

Tableau 2. Possibilités de dimérisation entre les facteurs de transcription des familles AP1 et CREB/ATF, observées in vitro et/ou in vivo chez les mammifères.