CARLA TORRES BRACONI

**Proteoma do baculovírus *Anticarsia gemmatalis* múltiplo
nucleopoliedrovírus em linhagens celulares distintas e
comparação da proteína de envelope GP64 em variantes
geográficas**

Tese apresentada ao Programa de Pós-Graduação
em Microbiologia do Instituto de Ciências da
Universidade de São Paulo, para obtenção do
Título de Doutor em Ciências.

São Paulo
2013

CARLA TORRES BRACONI

**Proteoma do baculovírus *Anticarsia gemmatalis* múltiplo
nucleopoliedrovírus em linhagens celulares distintas e
comparação da proteína de envelope GP64 em variantes
geográficos**

Tese apresentada ao Programa de Pós-Graduação
em Microbiologia do Instituto de Ciências da
Universidade de São Paulo, para obtenção do
Título de Doutor em Ciências.

Área de concentração: Microbiologia

Orientador: Prof. Dr. Paolo Marinho de Andrade
Zanotto

Versão original

São Paulo
2013

DADOS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
Serviço de Biblioteca e Informação Biomédica do
Instituto de Ciências Biomédicas da Universidade de São Paulo

reprodução não autorizada pelo autor

Braconi, Carla Torres.

Proteoma do baculovírus *Anticarsia gemmatilis* múltiplo nucleopoliedrovírus em linhagens celulares distintas e comparação da proteína de envelope GP64 em variantes geográficos / Carla Torres Braconi. -- São Paulo, 2013.

Orientador: Prof. Dr. Paolo Marinho de Andrade Zanotto.

Tese (Doutorado) – Universidade de São Paulo. Instituto de Ciências Biomédicas. Departamento de Microbiologia. Área de concentração: Microbiologia. Linha de pesquisa: Vírus de inseto.

Versão do título para o inglês: The baculovirus *Anticarsia gemmatilis* multiple nucleopolyhedrovirus proteome and the comparison of multiple isolated envelope proteins GP64.

1. Baculovírus 2. Controle biológico 3. Proteoma 4. AgMNPV
5. GP64 I. Zanotto, Prof. Dr. Paolo Marinho Andrade II. Universidade de São Paulo. Instituto de Ciências Biomédicas. Programa de Pós-Graduação em Microbiologia III. Título.

ICB/SBIB0134/2013

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS

Candidato(a): Carla Torres Braconi.

Título da Tese: Proteoma do baculovírus *Anticarsia gemmatalis* múltiplo nucleopoliedrovirus em linhagens celulares distintas e comparação da proteína de envelope GP64 em variantes geográficos.

Orientador(a): Prof. Dr. Paolo Marinho de Andrade Zanotto.

A Comissão Julgadora dos trabalhos de Defesa da Tese de Doutorado, em sessão pública realizada a/...../....., considerou

☐ **Aprovado(a)**

☐ **Reprovado(a)**

Examinador(a): Assinatura:
Nome:
Instituição:

Examinador(a): Assinatura:
Nome:
Instituição:

Examinador(a): Assinatura:
Nome:
Instituição:

Examinador(a): Assinatura:
Nome:
Instituição:

Presidente: Assinatura:
Nome:
Instituição:

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS

Cidade Universitária "Armando de Salles Oliveira"
Av. Prof. Lineu Prestes, 2415 - CEP. 05508-000 São Paulo, SP - Brasil
Telefone : (55) (11) 3091-7733 - telefax : (55) (11) 3091-7438
e-mail: cep@icb.usp.br

Comissão de Ética em Pesquisa

CERTIFICADO DE ISENÇÃO

Certificamos que o Protocolo CEP-ICB Nº 359/10 referente ao projeto intitulado: *“Estudo das proteínas estruturais do nucleopaliedrovirus anticarsia gemmatalis (AgMNPV-2D) em linhagens de células de inseto distintas e comparação da GP64 em variantes geográficos”* sob a responsabilidade de **Carla Torres Braconi**, foi analisado na presente data pela CEUA - COMISSÃO DE ÉTICA NO USO DE ANIMAIS e pela CEPSh- COMISSÃO DE ÉTICA EM PESQUISA COM SERES HUMANOS, tendo sido deliberado que o referido projeto não envolve manipulação animal ou humana que justifique uma aprovação quanto aos princípios éticos exigidos por ambas as Comissões.

São Paulo, 25 de fevereiro de 2010.

PROF. DR. WOTHAN TAVARES DE LIMA
Coordenador da CEUA - ICB/USP

PROF. DR. PAOLO M.A. ZANOTTO
Vice-Coordenador da CEPsh - ICB/USP

*Ao meu marido Alfredo, pelo suporte incondicional,
incentivo e paciência nesta caminhada.*

*Aos meus pais e irmã que sempre me apoiaram, meus
exemplos de amor, força e perseverança.*

Amo muito vocês.

AGRADECIMENTOS

Ao meu orientador Paolo Zanotto, por me ensinar e me dar liberdade de criar e pesquisar nesses últimos anos da melhor e mais intensa forma possível. Obrigada pelo conhecimento, pelo exemplo de pesquisa e pela paixão a Ciência. Por tudo o que me ajudou a construir nesse tempo e pela oportunidade dessa grande experiência de vida. Serei sempre grata pelos anos que passei em seu laboratório e tenho muito orgulho de ser um aluno do LEMB.

A minha família, que me deu valores e uma ótima educação. Obrigada pelo apoio incondicional, força e por sempre acreditarem na minha caminhada. Eu sou muito feliz em ter vocês ao meu lado, Pai, Mãe e Martoca. Ao meu marido Alfredo, por ser este homem que eu admiro demais, que meu apoiou nos dias difíceis e que curtiu comigo meus dias felizes de cientista, por vezes sem nem saber do que eu estava falando. A todos vocês que entenderam o quanto minha formação era importante.

A minha segunda família, os amigos do LEMB que sem eles o dia a dia não teria sido o mesmo, por todos os anos de convivência de muitas risadas, cafés e muito aprendizado. A todos que fizeram e fazem toda a diferença até hoje: Amiga Camis, Amigo Frank, Atila, Caio, J Vela, J Blocolina, Ana V e Julian. Aos mais novos do grupo, Anderson, Dani, Noemi e Nicholas que também fizeram parte. A todos que passaram pelo LEMB e que fizeram diferença nestes últimos 8 anos.

Aos meus amigos de HH, que sempre compartilharam muito mais do que cerveja, mas ideias, amizade, carinho e muita conversa boa. Obrigada pelo suporte de sempre, atenção e compreensão durante todos estes anos. Enfim, por serem meus amigos de verdade para toda hora.

Aos amigos rivierenses e as amigas mackenzitas, que fizeram parte direta e indiretamente deste processo. Amigos queridos que tornam minha vida muito mais feliz.

Ao Prof. James E. Maruniak e a Alejandra G. Maruniak, pessoas queridas que me

proporcionaram três meses de uma experiência incrível. Pela presteza e pelos cuidados durante a temporada na Universidade da Flórida.

Ao grupo de massas do LNBio em Campinas coordenado pelo Dra. Adriana Franco Paes Leme, pela colaboração, pela utilização de seus equipamentos e por me ajudar durante os experimentos de espectrometria de massas. Obrigada especial a Bianca e a Romenia pelo suporte.

Ao doutorando Daniel M. P. Ardisson de Araújo e ao professor Bergmann Ribeiro muito obrigada pela colaboração, e pela presteza em me ajudar durante os experimentos de western blot na Universidade de Brasília.

Ao Instituto de Ciências Biomédicas e ao departamento de Microbiologia, que me proporcionaram a formação e o aprendizado. Um agradecimento especial, a prof. Heloiza Barbosa que me proporcionou uma experiência diária de vida.

A minha ex-orientadora de iniciação científica que fez parte dessa história.

A Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) que tornou tudo isso possível com o suporte financeiro.

Meus sinceros agradecimentos a todos que me ajudaram neste trabalho.

If we knew what it was we were doing, it would not be called research, would it?
Albert Einstein

“De tudo ficaram três coisas...
A certeza de que estamos começando...
A certeza de que é preciso continuar...
A certeza de que podemos ser interrompidos
antes de terminar...
Façamos da interrupção um caminho novo...
Da queda, um passo de dança...
Do medo, uma escada...
Do sonho, uma ponte...
Da procura, um encontro!”
Fernando Pessoa

RESUMO

Braconi CT. Proteoma do baculovírus *Anticarsia gemmatalis* múltiplo nucleopoliedrovírus em linhagens celulares distintas e comparação da proteína de envelope GP64 em variantes geográficas. [Tese (Doutorado em Microbiologia)]. São Paulo: Instituto de Ciências Biomédicas, Universidade de São Paulo; 2013.

A família *Baculoviridae* compreende um grande e diverso grupo de vírus patogênicos de insetos, com aproximadamente 700 espécies de hospedeiros. Durante o ciclo viral, dois fenótipos são produzidos: o ODV (*occlusion derived virus*) responsável pela infecção primária do intestino médio da larva de insetos e o BV (*Budded virus*) responsável pela infecção sistêmica. No Brasil, desde a década de 1980, o *Anticarsia gemmatalis* múltiplo nucleopoliedrovírus (AgMNPV) é utilizado como controle biológico viral no controle da lagarta-da-soja *Anticarsia gemmatalis*. Com o sequenciamento do genoma do AgMNPV-2D, foram confirmadas potencialmente 152 ORFs, com 26 ORFs que codificam para proteínas estruturais, entre elas a glicoproteína 64 (GP64), fundamental para infecção secundária. No entanto, dados a respeito das proteínas presentes nos fenótipos do AgMNPV-2D são escassos. Este estudo objetiva identificar as proteínas estruturais presentes no AgMNPV-2D por duas abordagens distintas de espectrometria de massas. Outro objetivo consistiu em comparar possíveis adaptações contextuais da *gp64* de diferentes isolados geográficos do AgMNPV por clonagem, sequenciamento por Sanger e sequenciamento de alta cobertura. Através deste procedimento, identificamos as substituições sinônimas e não sinônimas de *gp64* e vimos que ela não suporta a noção de isolamento geográfico dos AgMNPVs amostrados. Também identificamos 44 e 33 proteínas presentes nos fenótipos ODV e BV no AgMNPV-2D, respectivamente. Embora parte das proteínas já havia sido descrita, o que é coerente com outros proteomas da família *Baculoviridae*, seis proteínas foram identificadas pela primeira vez no ODV e sete delas associadas ao fenótipo BV. Contudo, não foi possível predizer função para as proteínas pobremente categorizadas. Além disso, 11 proteínas celulares foram identificadas no AgMNPV-2D, possivelmente necessárias para infecção viral. Este achado contribui para o melhor entendimento da morfogênese do AgMNPV e fatores associados à multiplicação viral.

Palavras-chave: AgMNPV. Baculovírus. Controle Biológico. Proteoma. *gp64*.

ABSTRACT

Braconi CT. The baculovirus *Anticarsia gemmatalis* multiple nucleopolyhedrovirus proteome and the comparison of multiple isolated envelope proteins GP64” [Thesis (Ph.D. in Microbiology)]. São Paulo: Instituto de Ciências Biomédicas, Universidade de São Paulo; 2013.

Baculoviridae are arthropod-specific viruses that commonly infect insect orders, with more than 700 known host species. During the viral life cycle, two distinct phenotypes are produced: the budded virus (BV) and, the occlusion-derived virus (ODV) for intra and across host spread, respectively. Since the 80's several countries have been using the *Anticarsia gemmatalis* multiple nucleopolyhedrovirus (AgMNPV) as a biological control agent against the velvet bean caterpillar (*A. gemmatalis*). Brazil coordinated the largest program using AgMNPV as a biological pesticide in soybean plantations. The genome of the AgMNPV-2D carries at least 152 potential genes, 26 of which possibly code for structural proteins. Proteomic studies were done on a few baculoviruses, with six ODV and two BV proteomes finished so far. Moreover, there is limited data on virion proteins carried by the AgMNPV-2D. In this study, the structural proteins of AgMNPV-2D were analyzed by two complementary mass spectrometry techniques. The additional objective was to compare the gene *gp64* of different geographical population of AgMNPV by cloning, Sanger sequencing and next generation sequencing methods. This analysis allowed us to observe the synonymous and non-synonymous substitutions of *gp64* and refuted the notion of a geographical isolation of the sampled AgMNPV. We also observed a total of 44 proteins associated to the ODV and 33 to the BV of AgMNPV-2D. Several proteins have been described as structural already, though our data showed six new proteins in the ODV and seven new proteins carried by the BV. Furthermore, 11 cellular proteins were also identified, which are possibly assorted during viral morphogenesis and carried with the virion. Unfortunately, it was not possible to predict gene function for poorly characterized genes. These findings may provide novel insights into AgMNPV biology and its host interaction, leading us to a better understanding about morphogenesis and also the associated factors of the viral multiplication.

Key words: AgMNPV. Baculovirus. Biological Control. Proteome. *gp64*.

LISTA DE ABREVIATURAS E SIGLAS

ACN- Acetonitrila

AcMNPV - *Autographa californica* múltiplo nucleopoliedrovírus

AgMNPV - *Anticarsia gemmatilis* múltiplo nucleopoliedrovírus

Blast - *Basic Local Alignment Search Tool*

BmNPV - *Bombyx mori* nucleopoliedrovírus

BSA - Albumina Sérica Bovina

BV - *budded virus*

ChChNPV - *Chrysodeixis chalcites* nucleopoliedrovírus

CpGV - *Cydia pomonella* granulovírus

2DE - Eletroforese bidimensional

DNA - *deoxyribonucleic acid* (ácido desoxirribonucleico)

DTT- Ditiotritol

EDTA - ácido etilenodiamino tetra-acético

EppoNPV - *Epiphyas postvittana* nucleopoliedrovírus

ESI-Q-TOF - *Electrospray quadrupole time of flight*

GV - granulovírus

GRN - *Gene regulatory network*

h – hora

HearNPV - *Helicoverpa armigera* nucleopoliedrovírus

HycuMNPV - *Hyphantria cunea* múltiplo nucleopoliedrovírus

HPLC - *High performance Liquid Chromatography* (Cromatografia líquida de alta performance)

IOA- Iodoacetamida

kb - kilobase

kDa- kilodalton

L - litro

LC - MS/MS - *High performance Liquid Chromatography coupled with tandem mass spectrometry* (Cromatografia líquida de alta performance acoplada a espectrometria de massas em sequência)

LdMNPV - *Lymantria dispar* múltiplo nucleopoliedrovírus

M - *multiple*

M - Molar

MALDI-Q-TOF (*Matrix assisted laser desorption/ionization quadrupole time of flight*)

min - minuto

ml - mililitro

mM miliMolar

MOI - *multiplicity of infection*

MS/MS - espectros de massas

NaCl - Cloreto de sódio

NPV - nucleopoliedrovírus

ODV- *occlusion derived virus*

Oligo - oligonucleotídeo
OpMNPV - *Orgyia pseudotsugata* multiple nucleopolyhedrovirus
ORF - *Open reading frame* (fase aberta de leitura)
pb - pares de base
PBS - *Phosphate buffer saline* (Tampão fosfato)
PCR- *Polymerase Chain Reaction* (reação em cadeia da polimerase)
p.i. - pós-infecção
pI - ponto isoelétrico
PM - membrana peritrófica
PrGV - *Pieris rapae* Granulovirus
r.p.m - rotações por minuto
RNA - *ribonucleic acid* (ácido ribonucleico)
s - segundo
S - *single*
SDS-PAGE - Eletroforese em gel de poliacrilamida contendo dodecil sulfato de sódio
SfMNPV - *Spodoptera frugiperda* múltiplo nucleopoliedrovirus
SF-9 - linhagem celular IPLB-SF-9
SFB - soro fetal bovino
U - unidade
UFL-AG - linhagem celular UFL-AG-286
 μ l – microlitro
 ω – diferença de *dN-dS*
 Γ – Variação taxa de substituição do modelo de tempo reversível GTR

LISTA DE FIGURAS

Figura 1 - Reconstrução filogenética dos baculovírus cujos genomas foram sequenciados...	19
Figura 2 - Os fenótipos virais produzidos durante o ciclo de vida dos baculovírus.....	21
Figura 3 - Ciclo biológico ilustrativo de um <i>Alphabaculovirus</i>	23
Figura 4 - Foto da lagarta <i>Anticarsia gemmatalis</i> praga da monocultura de soja e microscopia eletrônica de um poliedro do baculovírus AgMNPV.....	29
Figura 5 - Esquema do genoma do AgMNPV-2D.....	32
Figura 6 - Rede regulatória (GRN) baseada no transcriptoma do AgMNPV-2D a partir da infecção em células UFL-AG-286.....	34
Figura 7 - Estrutura ilustrativa dos fenótipos virais de um <i>Alphabaculovirus</i>	36
Figura 8 - Microscopia eletrônica e estrutura ilustrativa do nucleocapsídeo viral.....	38
Figura 9 - Reconstrução filogenética baseada nas sequências de <i>gp64</i> dos NPV I.....	39
Figura 10 - Esquema da estrutura da GP64 de AcMNPV realizada por cristalografia de raio X.....	40
Figura 11 - Esquema das fontes de ionização e os tipos de analisadores de massas utilizados nos espectrômetros de massas.....	42
Figura 12 - Esquema geral da metodologia descrita para o estudo das proteínas estruturais do AgMNPV-2D em duas linhagens celulares estudadas.....	46
Figura 13 - Esquema geral da metodologia descrita para a análise comparativa da ORF-124 (<i>gp64</i>) dos isolados geográficos do AgMNPV.....	58
Figura 14 - Foto ilustrativa dos dois fenótipos apos a ultracentrifugação em gradiente de sacarose.....	64
Figura 15 - Eletroforese bidimensional em gel de poliacrilamida 12% demonstrando as proteínas estruturais do fenótipo ODV purificado de células UFL-AG-286.....	67
Figura 16 - Eletroforese bidimensional em gel de poliacrilamida 12% com as proteínas estruturais identificadas por MALDI-Q-TOF no fenótipo ODV purificado de células UFL-AG-286.....	69
Figura 17 - SDS-PAGE 12% com o perfil de proteínas do fenótipo ODV do AgMNPV -2D corado com <i>Comassie Blue-R</i>	71
Figura 18 - Eletroforese SDS PAGE 12% demonstrando as proteínas estruturais do fenótipo BV isolado e purificado de células: UFL-AG-286 e IPLB – SF-9.....	73

Figura 19 - Esquema das proteínas identificadas por MS/MS no AgMNPV em linhagem UFL-AG-286.....	75
Figura 20 - Uma célula UFL-AG-286 representada durante a infecção por AgMNPV-2D mostrando as proteínas celulares adquiridas durante o ciclo de multiplicação viral.....	79
Figura 21 - Cinco <i>western blot</i> distintos das proteínas de AgMNPV-2D.....	80
Figure 22 - Proteínas encontradas no AgMNPV-2D.....	81
Figura 23 - Amplificação da ORF 124 por PCR da região, GP64, dos isolados geográficos de AgMNPV.....	82
Figura 24 - Relações filogenéticas entre <i>gp64</i> de isolados geográficos de AgMNPV mostrados na Tabela 1.....	83
Figura 25 - Análise de mutação sinônima e não sinônima na <i>gp64</i> entre os isolados geográficos.....	84
Figura 26 - Localização das mutações não sinônimas no modelo estrutural da proteína GP64 do AgMNPV-2D, baseada na estrutura cristalográfica de raio X da GP64 do AcMNPV.....	86
Figura 27 - Estrutura modular baseada na rede do AgMNPV-2D a partir da infecção em células UFL-AG-286 pelo método de clusterização do programa GLay.....	96

SUMÁRIO

1 INTRODUÇÃO	18
1.1 A Família <i>Baculoviridae</i>, taxonomia e morfologia	18
1.2 Fenótipos virais e ciclo de vida dos <i>Alphabaculovirus</i>	20
1.3 Regulação gênica e replicação dos baculovírus	23
1.3.1 Fase precoce e replicação do DNA viral	24
1.3.2 Fase tardia	25
1.3.3 Fase muito tardia	26
1.4 Evolução da família <i>Baculoviridae</i>	26
1.5 A família <i>Baculoviridae</i>: aplicação biológica e benefícios econômicos	28
1.6 O <i>Alphabaculovirus Anticarsia gemmatalis</i> múltiplo nucleopoliedrovírus	30
1.7 Estudo proteômico na família <i>Baculoviridae</i>	34
1.8 Glicoproteína 64 (GP64)	38
1.9 Métodos de Proteômica	40
2 OBJETIVOS	44
2.1 Objetivos Gerais	44
2.2 Objetivos Específicos	44
3 METODOLOGIA	45
3.1 Análise do proteoma do AgMNPV-2D por espectrometria de massas	45
3.1.1 Cultivo Celular	47
3.1.2 Infecção Viral	47
3.1.3 Titulação Viral	47
3.1.4 Amplificação e Produção Viral	48
3.1.5 Concentração viral e dissolução dos corpos de oclusão	48
3.1.6 Purificação das Partículas Virais	49
3.1.7 Quantificação proteica	50
3.1.8 Precipitação proteica	51
3.1.9 Eletroforese SDS-PAGE	51
3.1.10 Eletroforese Bidimensional (2DE)	52
3.1.11 Digestão proteica in gel	53
3.1.12 Espectrometria de massa e análise dos espectros de massas dos peptídeos tripticos	54
3.1.13 Análise de dados e mapeamento in silico no genoma do AgMNPV-2D	55
3.1.14 Western blot para confirmação dos dados obtidos por espectrometria de massas	56

3.2 Análise de diversidade genética da gp64 (ORF-124)	57
3.2.1 Análise comparativa da ORF-124	57
3.2.2 Multiplicação viral in vivo, purificação, extração do DNA viral total e pirosequenciamento de isolados geográficos de AgMNPV	59
3.2.3 Análise das sequências de gp64 (ORF-124) geradas por pirosequenciamento	60
3.2.4 Síntese de oligonucleotídeos para a ORF-124 e Reações de PCR	60
3.2.5 Preparo de bactérias Químio-competente e Clonagem do produto de PCR	61
3.2.6 Sequenciamento por Sanger e análise dos clones	61
3.2.7 Análise filogenética da gp64	62
4 RESULTADOS	63
4.1 Análise do proteoma estrutural do AgMNPV-2D em duas linhagens de células de insetos por espectrometria de massas	63
4.1.1 Produção e purificação viral	63
4.1.2 Quantificação Proteica	64
4.1.3 Identificação das proteínas do AgMNPV-2D por eletroforese bidimensional (2DE) e espectrometria de massas MALDI-Q-TOF	65
4.1.4 Identificação das proteínas do AgMNPV-2D por eletroforese SDS-PAGE e espectrometria de massas LC-MS/MS	70
4.1.5 Proteínas do Hospedeiro identificadas no AgMNPV-2D	78
4.1.6 Análises das proteínas por Western blot	79
4.2 Análise comparativa da ORF-124 (gp64)	81
4.2.1 Avaliação da diversidade da gp64	84
5 DISCUSSÃO	87
5.1 Proteoma do AgMNPV-2D	87
5.1.1 Proteínas presentes no fenótipo ODV e BV na linhagem UFL-AG-286	88
5.2 Fenótipo BV nas duas linhagens celulares estudadas	90
5.2.1 Semelhanças encontradas no fenótipo BV nas duas linhagens celulares estudadas	90
5.2.2 Diferenças encontradas no fenótipo BV nas duas linhagens celulares estudadas são congruentes em diferentes níveis de descrição	94
5.3 Proteínas celulares encontradas no proteoma do AgMNPV-2D	98
5.4 Análise comparativa da gp64	99
6 CONCLUSÕES	103
REFERÊNCIAS*	104

APÊNDICES	120
Apêndice A - Artigos Submetidos	121
Apêndice B - Triplicatas Géis Bidimensionais (2DE).....	187
Apêndice C - Reação de PCR e diversidade da GP64	190
ANEXO - Lista de publicações	193

1 INTRODUÇÃO

1.1 A Família *Baculoviridae*, taxonomia e morfologia

Vírus representam um importante desafio aos insetos e doze famílias já foram associadas a estes artrópodes (Erlandson, 2008). A família *Baculoviridae* compreende um grande e diverso grupo de vírus patogênicos a artrópodes, que infectam insetos das ordens; Lepidóptera, Himenóptera e Díptera. Aproximadamente 700 espécies de hospedeiros de baculovírus já foram reportados na literatura, majoritariamente isolados em Lepidópteras (O'Reilly et al., 1993; Tanada, Kaya, 1993). Estes vírus são estáveis no ambiente e ocorrem naturalmente em populações de insetos e apresentam alta especificidade de hospedeiros (Moscardi, 1999). Estas características tornaram a família *Baculoviridae* interessante para a aplicação segura como controle biológico de pragas presentes em áreas de manejo de agrícola, pastos e áreas florestais (Moscardi, 1999; Szewczyk et al., 2006; Tanada, Kaya, 1993). Além disso, com o amplo desenvolvimento tecnológico na década de 80, alguns baculovírus, como o AcMNPV e o BmNPV, tornaram-se úteis como vetores versáteis de expressão de genes heterólogos (Maeda, 1989). Os vetores baseados em baculovírus são um excelente sistema de produção de proteínas recombinantes em linhagens de células de insetos (Boyce, Bucher, 1996; Hofmann et al., 1995; Smith et al., 1983; Smith et al., 1985). Recentemente, alguns estudos os descreveram como vetores para transdução *in vitro*, *in vivo* e *ex vivo* em uma variedade de linhagens celulares de mamíferos tornando-os possíveis candidatos para o uso em terapia gênica (Airenne et al., 2013; Hu, 2006; Kost et al., 2005; Wang, Balasundaram, 2010).

A classificação taxonômica da família é baseada em análises filogenéticas e características moleculares de mais de setenta genomas completos. A comparação de seu conteúdo gênico permitiu a identificação de trinta e sete genes conservados entre todos genomas sequenciados, chamados de *core genes* (Garavaglia et al., 2012). Estes genes desempenham funções biológicas importantes durante o ciclo viral, como: transcrição, produção da partícula viral, infecção do intestino do inseto e estabelecimento da infecção (Herniou et al., 2003; Rohrmann, 2008; van Oers, Vlaskov, 2007). Ademais, o conteúdo e a organização gênica foram considerados para estabelecer as relações evolutivas entre os membros da família *Baculoviridae* (Herniou et al., 2001). Portanto, desde 2008, o Comitê Internacional de Taxonomia de vírus (ICTV) (www.ictvonline.org) divide esta família em

quatro gêneros: (i) os *Alphabaculovirus*, (NPVs específicos de lepidópteras), (ii) os *Betabaculovirus* (GVs específicos de lepidópteras), (iii) os *Gamabaculovirus* (NPVs específicos de himenópteras) e (iv) os *Deltabaculovirus* (NPVs específicos de dípteros) (figura 1).

Figura 1 - Reconstrução filogenética baseada nos 37 genes conservados de baculovírus sequenciados até o momento.

A relação entre os membros da família está demonstrada na divisão dos quatro gêneros na árvore: i) os ramos em azul são os *Alphabaculovirus*, (NPVs específicos de lepidópteras), (ii) os ramos em rosa são os *Betabaculovirus* (GVs específicos de lepidópteras), (iii) os ramos em verde são os *Gamabaculovirus* (NPVs específicos de himenópteras) e o ramo em laranja os *Deltabaculovirus* (NPVs específicos de dípteros). A seta em vermelho aponta o *Anticarsia gemmatalis* múltiplo nucleopoliedrovírus (AgMNPV).

Os baculovírus são envelopados e seu genoma é composto de uma molécula de DNA circular dupla fita covalentemente fechada, que varia de 80 a 180 Kb e codifica de 100 a 180 proteínas (Granados et al., 1986; Herniou et al., 2003; Slack, Arif, 2007). O material genético destes vírus está envolto por um capsídeo proteico (nucleocapsídeo) em forma de bastonete (em latim *baculum*), que forma o vírion, a partícula infectante (Summers, Smith, 1987). Os vírions dessa família podem ser oclusos em uma matriz para-cristalina denominada de corpo de oclusão (*Occlusion body*, OB) (Tanada, Kaya, 1993). Os baculovírus foram inicialmente divididos em dois grupos com base nas diferenças morfológicas encontradas nestes corpos de oclusão em: nucleopoliedrovírus (NPVs) e granulovírus (GVs). Os NPVs podem apresentar apenas um nucleocapsídeo (SNPV - *Single Nuclear Polydrosis Virus*) ou até dezenas (MNPV - *Multiple Nuclear Polydrosis Virus*) por corpo de oclusão (Bilimoria, 1991). Neste grupo os vírions são oclusos em corpos denominados poliedros, constituídos principalmente pela proteína poliedrina (28 kDa) (Bilimoria, 1991; Granados et al., 1986). Os NPVs específicos de lepidópteras ainda são subdivididos em grupos I e II baseado em estudos filogenéticos com o gene da poliedrina (*polh*) (Zanotto et al., 1992; Zanotto et al., 1993), genes conservados da família *Baculoviridae* (Herniou et al., 2003) e genomas completos (Oliveira et al., 2006b; Wolff et al., 2008; Zanotto, Krakauer, 2008). Os dois grupos requerem proteínas de envelope distintas para fusão do vírus com a célula hospedeira: a glicoproteína GP64 está presente no grupo I, enquanto a proteína F está no grupo II (Herniou et al., 2003; Monsma et al., 1996; Pearson et al., 2000). Alguns baculovírus apresentam as duas proteínas no envelope viral, porém apenas uma mantém função de fusão (Okano et al., 2006). Além disso, 11 genes adicionais parecem ser exclusivos dos NPVs (Herniou et al., 2001; Jehle et al., 2006). Os GVs, por outro lado, normalmente apresentam um nucleocapsídeo por corpo de oclusão com formato ovicilíndrico, denominados grânulo (Crook, 1991; Summers, Smith, 1987). Os grânulos são compostos principalmente pela proteína granulina (28 kDa), que apresenta 50% de similaridade com a sequência de aminoácidos da poliedrina (Akiyoshi et al., 1985; Crook, 1991). Ademais, os GVs foram somente isolados de Lepidópteras e alguns membros deste grupo causam infecção limitada ao intestino médio (Tanada, Kaya, 1993).

1.2 Fenótipos virais e ciclo de vida dos *Alphabaculovirus*

Os baculovírus possuem uma característica peculiar no ciclo infeccioso: a produção de dois fenótipos distintos em estrutura e função, mas de mesma composição gênica (Blissard,

1996). Como mencionado, os fenótipos desempenham funções diferentes durante o ciclo viral: o ODV (*occlusion derived virus*) é responsável pela transmissão horizontal entre os hospedeiros e pela infecção primária das células colunares do intestino médio da larva de insetos (Granados et al., 1986); e o BV (*Budded virus*) não ocluso, é responsável pela infecção sistêmica, *i.e.*, propagação da infecção célula a célula, infectando inicialmente os hemócitos na hemolinfa e subsequentemente espalhando a infecção para os demais tecidos do hospedeiro (Figura 2).

Figura 2 - Os fenótipos virais produzidos durante o ciclo de vida dos baculovírus.

À esquerda está representando o BV, com um único nucleocapsídeo envolto pela membrana celular e proteínas virais inseridas. No caso de um NPV I, a proteína GP64 forma peplômeros em uma das pontas da partícula de BV. Ao meio, está representado um ODV com diversos nucleocapsídeos envoltos pelo mesmo envelope. À direita está representado um corpo de oclusão, o poliedro, onde os ODVs dos NPVs estão embebidos e protegidos das adversidades do ambiente.

Fonte: adaptada de <http://commons.wikimedia.org/wiki/Baculovirus>. Com permissão.

O ciclo infeccioso de um *Alphabaculovirus* tem início quando o inseto ingere o poliedro presente no meio ambiente (Figura 3). Estes vírus evoluíram de forma a iniciar a infecção no intestino médio do inseto, onde o corpo de oclusão é dissolvido em pH alcalino (pH = 10,5 a 11,0) e os ODVs são liberados no lúmen intestinal (Dow, 1992). Essas partículas atravessam a membrana peritrófica (PM), barreira física de proteção do intestino médio do inseto, sendo que alguns baculovírus ainda possuem enhacinas capazes de degradar

componentes da PM (Wang, Granados, 1997). As proteínas responsáveis pela infectividade oral presentes no envelope lipoprotéico dos ODVs se ligam aos receptores das células colunares do intestino médio (Fang et al., 2006; Faulkner et al., 1997; King, Possee, 1992; Peng et al., 2012). Isso permite que os vírions consigam se fundir e penetrar no citoplasma, dando início à infecção primária. Uma vez dentro da célula, os nucleocapsídeos são direcionados ao núcleo através da polimerização de filamentos de actina que penetram pelos poros nucleares, onde são desnudados, liberando o DNA viral (Charlton, Volkman, 1991; Lanier, Volkman, 1998; Ohkawa et al., 2010). A partir desse momento, o vírus controla a maquinaria celular e inicia a sua replicação com a transcrição de alguns genes virais, seguida da síntese de DNA viral (Granados et al., 1986). Os primeiros nucleocapsídeos são produzidos no núcleo e migram para o citoplasma, em direção membrana à plasmática já modificada pela inserção de proteínas do envelope; glicoproteínas GP64 no caso do grupo I e proteínas F no caso do grupo II (Blissard, Rohrmann, 1990). Assim é produzido o fenótipo BV, que deixa a célula hospedeira por brotamento polarizado da membrana celular (O'Reilly et al., 1993). As proteínas presentes no envelope lipoproteico do BV (GP64 e/ou proteína F) formam peplômeros que se ligam aos receptores celulares, dando início à infecção célula a célula. A infecção sistêmica segue pelo tecido traqueal, células da hemolinfa, corpo gorduroso, túbulos de Malpighi e por fim células cerebrais da lagarta (Soares, Ribeiro, 2005). Na fase tardia da infecção há intenso acúmulo de nucleocapsídeos nos núcleos celulares, onde são envelopados em grupos e dão origem aos ODVs, que promovem grande aumento do volume nuclear e culminam na lise da célula (Granados et al., 1986). A nova progênie de poliedros é liberada no ambiente pelo excremento ou quando o inseto morre (Keddie et al., 1989). Alguns insetos infectados podem migrar para um lugar mais elevado na planta, o que facilita a ação de predadores e a dispersão dos corpos de oclusão, comportamento controlado pela expressão de dois genes virais, *ptp* e *egt* (Hoover et al., 2011; Kamita et al., 2005; Katsuma et al., 2012). Os poliedros presentes no solo e nas folhas são responsáveis pela disseminação dos baculovírus de inseto para inseto. Tais estruturas são muito estáveis e resistentes, protegendo os vírions de degradação e inativação no meio ambiente, principalmente por radiação ultravioleta (UV) (Coulibaly et al., 2009; Granados et al., 1986).

Figura 3 - Ciclo biológico ilustrativo de um *Alphabaculovirus*.

Os poliedros presentes nas plantações (P) são ingeridos pelas lagartas. No intestino médio, são dissolvidos e os ODVs são liberados, iniciando a infecção primária. Os nucleocapsídeos (NC) são encaminhados para o núcleo, onde há intensa replicação viral culminando na produção de BVs, responsáveis pela infecção secundária dos tecidos da lagarta. Nos estágios finais da infecção, os poliedros são acumulados nos núcleos das células infectadas e ocasionam lise. A nova progênie de poliedros (P) é liberada no ambiente, onde servirão de inóculo para infecções em novas lagartas.

Fonte: Adaptada com permissão de (King, Possee, 1992).

1.3 Regulação gênica e replicação dos baculovírus

A produção de dois fenótipos durante o ciclo de vida dos baculovírus é reflexo de um diferente padrão de regulação gênica que ocorre em cascata, onde genes expressos em uma fase temporal coordenam a expressão de genes da fase seguinte (O'Reilly et al., 1993). Nesta cascata, a expressão gênica viral é dividida em três fases: (i) precoce (*early* - E), composta de genes preferencialmente expressos de uma a seis horas pós infecção (p.i.), (ii) tardia (*late* - L), genes expressos de seis a dezoito horas p.i. e (iii) muito tardia (*very late* - VL), genes expressos de dezoito horas p.i. até o final da montagem viral (King, Possee, 1992).

1.3.1 Fase precoce e replicação do DNA viral

Esta fase se caracteriza pelo começo do ciclo viral e da manipulação do metabolismo celular. Os genes desta fase são transcritos pela RNA polimerase DNA-dependente II celular que reconhece os promotores precoces virais similares aos celulares (Herniou et al., 2003; King, Possee, 1992). Além disso, os genes desta fase são subdivididos em duas classes: imediata (*immediate-early* - IE) e precoce tardia (*delayed early* - DE). Na fase imediata, dois genes críticos são transcritos, *ie-1* e *ie-0*. A proteína IE-1 é acumulada muito precocemente no núcleo e interage diretamente com o DNA viral. IE-1 e IE-0 são responsáveis pela ativação transcricional de genes da fase seguinte. Ademais, a expressão de IE-1 está diretamente relacionada à replicação do DNA viral e à inibição precoce de apoptose na célula infectada (Choi, Guarino, 1995; Leisy et al., 1995; Schultz et al., 2009). Os genes expressos na fase DE estão relacionados à síntese do DNA viral, como *DNA pol*, *helicase* (*p143*), *lef-1* (*late expression factor-1*), *lef-2* e *lef-3*. Com os produtos destes genes traduzidos, a maquinaria de replicação é formada.

Acredita-se que os baculovírus utilizem diferentes regiões do genoma como pontos de origem de replicação ou sequências iniciadoras, como: (i) regiões homólogas repetitivas, (*hrs*), (ii) sequências não *hrs* e (iii) promotores precoces. As *hrs* são frequentemente palíndromos imperfeitos, com 30 pb, que ocorrem em várias regiões do genoma (Wu et al., 1999). A proteína IE-1 se liga as *hrs* facilitando a entrada da maquinaria de replicação do DNA viral (Kool et al., 1994; Okano et al., 2006). Sucintamente, a Helicase inicia a abertura da dupla fita de DNA com a provável ajuda de uma topoisomerase celular (Craig, Nash, 1983; Okano et al., 2006). Uma vez formada a forquilha de replicação, os DNAs fita simples são estabilizados pela LEF-3, que possibilita a formação do complexo de primases, LEF-1 e LEF-2. Este complexo sintetiza uma sequência curta de oligonucleotídeos que servirá de molde para a DNA polimerase continuar o processo no sentido 5'-3' da nova cadeia. No sentido contrário, são produzidos os fragmentos de Okasaki na replicação da fita 3'-5' que serão ligados pela DNA ligase do hospedeiro, uma vez que a DNA polimerase processa apenas no sentido 5'-3' (Okano et al., 2006). A Helicase (P143) e LEF-3 também atuam na inibição de apoptose na célula infectada (Schultz, Friesen, 2009). Além dessas, outras proteínas também estão envolvidas no processo de síntese de DNA, como a nuclease alcalina (Alk-Exo), conservada em todos os baculovírus. Esta proteína se associa ao DNA de fita simples, apresenta atividade exonuclease e endonuclease no sentido 5'-3', digere uma pequena região

do DNA recém duplicado e forma uma cauda de DNA fita simples na extremidade 3' de cada DNA, que pode se ligar por homologia em outras sequências e gerar fragmentos maiores (Mikhailov et al., 2003, 2004; Okano et al., 2006). Na fase DE ainda são expressos os genes relacionados à transcrição, como RNA polimerase viral, constituída do produto de quatro genes, *lef-4*, *lef-8*, *lef-9* e *p47* (Gross, Shuman, 1998), e alguns genes auxiliares necessários para modulação celular, como *pp31*.

1.3.2 Fase tardia

A fase tardia (*late - L*) se inicia ao final da replicação do DNA viral e os genes desta fase já são transcritos pela RNA polimerase viral que reconhece os motivos do início de transcrição A/G/T TAAG (Blissard, Rohrmann, 1990; Gross, Shuman, 1998). Nesta fase, são expressos os genes relacionados a estrutura da partícula viral, como: nucleocapsídeo e envelope viral (King, Possee, 1992). Os genes estruturais conservados em todos baculovírus são transcritos nesta fase: *gp41*, *p6.9*, *vp39*, *vp1054*, *p74*, *vp91*, *odv-e56* e *odv-e27*. As proteínas codificadas a partir destes genes estão diretamente envolvidas com o empacotamento do DNA sintetizado na fase anterior, para formação dos novos nucleocapsídeos que darão origem primeiramente ao fenótipo BV (Herniou et al., 2003). Nesta fase, também são transcritos os genes auxiliares e moduladores de respostas celulares, ocasionadas pela infecção viral. Entre estes, estão os genes que suprimem a resposta de morte programada da célula, a apoptose. Um bom exemplo são os genes *Iaps* (anti-apoptóticos), *p35* e *p49* (Clem, Miller, 1994; Clem, 2001). A família gênica *Iaps* foi primeiramente descrita em baculovírus, no granulovírus *Cydia pomonella* (CpGV) e no nucleopoliedrovírus *Orgyia pseudotsugata* (OpMNPV) (Birnbaum, 1994 #1123). A família gênica *Iaps* foi primeiramente descrita em baculovírus, no granulovírus *Cydia pomonella* (CpGV) e no nucleopoliedrovírus *Orgyia pseudotsugata* (OpMNPV) (Birnbaum et al., 1994). Estes genes apresentam um motivo conservado BIR de aproximadamente 70 aminoácidos no N-Terminal único desta família e um domínio *ring finger* de 40 aminoácidos no C-terminal, que transfere ubiquitina para proteínas celulares apoptóticas, encaminhando-as para via de degradação (Crook et al., 1993; Vaux, Silke, 2005). Nos baculovírus, já foram identificados por homologia cinco cópias ou genes parálogos de *Iaps* (*iaps-1 a 5*); contudo apenas uma das cópias apresenta função anti-apoptótica (Clem, 2007; Luque et al., 2001). Recentemente, demonstrou-se que a *iap-1* apresenta atividade contrária e estimula apoptose na fase tardia em HycuMNPV

(*Hyphantria cunea* multiple nucleopolyhedrovirus), AcMNPV, BmNPV e OpMNPV (*Orgyia pseudotsugata* multiple nucleopolyhedrovirus) (Ikeda et al., 2011). Além desta família gênica, na fase tardia os baculovírus também expressam moduladores de ciclo celular, como é o caso dos genes *odv-e27* e *pk1*. A *odv-e27* é homóloga à ciclina B de *Bombyx mori*. Logo, além do papel estrutural no nucleocapsídeo, a proteína age na fase G2 do ciclo celular, induzindo a célula a continuar nesta fase, o que permite a finalização da replicação viral (Baluchamy, Gopinathan, 2005). Uma função similar foi associada a *pk1* em AcMNPV e BmNPV, promovendo a parada no ciclo celular em fase G2 (Katsuma et al., 2007; Mishra et al., 2008; Reilly, Guarino, 1994).

1.3.3 Fase muito tardia

Na fase muito tardia, os vírions são ocluídos nos corpos de oclusão e dois genes estruturais são hiperexpressos: a poliedrina, predominante no corpo de oclusão e a P10, proteína de 10 kDa envolvida no processo de oclusão das partículas virais via associação ao citoesqueleto celular e na lise celular no final da infecção (Van Oers, Vlak, 1997). Estes genes podem ser deletados do genoma dos baculovírus sem afetar a produção dos fenótipos virais, especialmente o ODV. Ademais, os promotores destes genes muito tardios são amplamente utilizados em vetores de expressão de genes heterólogos. Alguns estudos demonstraram que a presença de uma sequência de 8 pares de bases antes do códon de início (ATG) determina a atividade máxima dos promotores de genes muito tardios (Rankin et al., 1988). A proteína VLF-1 interage para ativar a expressão de *poliedrina* e *p10* (Mistretta, Guarino, 2005; Yang, Miller, 1999). O gene *vlf-1* é um gene *core*, que codifica a uma proteína pertencente a família de proteínas λ integrase, que catalisam o rearranjo de DNA, presentes em uma variedade de vírus e organismos (Iamarino et al., 2012; Okano et al., 2006). Evidências sugerem que esta proteína pode estar envolvida com o empacotamento dos novos genomas nos baculovírus.

1.4 Evolução da família *Baculoviridae*

Como mencionado anteriormente, os estudos filogenéticos de genomas completos realizados com baculovírus possibilitaram a identificação do conteúdo gênico, sua conservação e apontaram as relações entre os membros desta família e, portanto, como eles

podem ter evoluído (Garcia-Maruniak et al., 2004; Herniou et al., 2003; Lauzon, 2006; Oliveira et al., 2006b; Wolff et al., 2008; Zanotto et al., 1993; Zanotto, Krakauer, 2008). A filogenia de baculovírus evidencia uma série de mecanismos de interação com o hospedeiro conservadas em diversos organismos evolutivamente relacionados, apresentando maior similaridade entre os baculovírus que infectam Lepidópteras (NPVs e GVs) do que entre os NPVs que infectam as ordens Himenóptera e Díptera (Herniou et al., 2003). Uma característica dos vírus de DNA que infectam células eucarióticas é a capacidade de inserir, deletar, duplicar, recombinar e/ou transpor parte do seu genoma com o genoma do hospedeiro. Esta troca do material genético pode possibilitar novas interações, como a ampliação de espectros de hospedeiros ou até ganho de função (Miele et al., 2011). Especula-se que durante a evolução os baculovírus devem ter infectado um ancestral comum de himenópteros, dípteros e lepidópteros e com a divergência destes, há aproximadamente 350 milhões de anos coevoluíram com seus respectivos hospedeiros (Herniou et al., 2004). Esta interação possibilitou que os baculovírus ganhassem funções específicas como a aquisição recente da proteína homóloga à Poli (ADP-ribose) polimerase (PARP) celular presente apenas no AgMNPV (Castro Oliveira et al., 2008; Oliveira et al., 2006b). Esta proteína pode possuir função similar as PARP-1 encontradas em outros organismos com um papel importante na manutenção da integridade do DNA (Smulson et al., 1994). Da mesma forma, em *Drosophila*, estão envolvidas na elongação do telômero, na estrutura da cromatina e transcrição de uma variedade de gene relacionados a imunidade, resistência a estresses, resposta a hormônio e ainda há indícios de sua relação com a inibição de retroelementos (Tulin et al., 2006).

Outros genes importantes do hospedeiro também foram sequestrados pelos baculovírus, como é o caso das enzimas catepsina e quitinase, que facilitam a degradação do inseto e o liquefazem na fase final do ciclo viral. A catepsina esta presente na maioria dos NPVs dos grupos I e II e em alguns genomas de GVs, e as catepsinas de AcMNPV e BmNPV apresentam similaridade de 60% de seus aminoácidos com a catepsina de *Bombyx mori* (Daimon et al., 2006). No entanto, estas enzimas não estão presentes no *Anticarsia gemmatilis* múltiplo nucleopoliedrovírus (AgMNPV). Durante a evolução dos baculovírus, outras duas proteínas foram adquiridas do hospedeiro ancestral, a EGT e PTP-1, ambas envolvidas com comportamento conhecido como ELA (*enhanced locomotory behavior*). Estas proteínas estimulam as lagartas infectadas a migrarem para lugares mais elevados de plantas ou árvores, facilitando a dispersão dos corpos de oclusão recém formados. A EGT,

ecdysteroid uridine 5'-diphosphate (UDP) – glucosyltransferase, foi descrita associada a ELA em lagartas *Lymantria dispar* infectadas com LdMNPV (*Lymantria dispar* múltiplo nucleopoliedrovírus) (Hoover et al., 2011). Já a PTP-1 é uma fosfatase de proteínas com resíduos de tirosina ou de serina/treonina conservada somente nos NPV do grupo I. Recentemente, foi associada à hiperatividade locomotora em *Bombyx mori* infectada com BmNPV (Katsuma et al., 2012).

1.5 A família *Baculoviridae*: aplicação biológica e benefícios econômicos

Doenças de insetos são conhecidas desde o século XVI, quando pela primeira vez lagartas do bicho-da-seda foram descritas liquefeitas (Benz, 1986). Contudo, somente no século XX a família *Baculoviridae* foi identificada e caracterizada (Miller, 1996). Em seguida, diversos estudos sobre baculovírus demonstraram que estes vírus não replicam em vertebrados, plantas e microrganismos. Aproximadamente cinquenta produtos baseados nesta família viral foram comercialmente utilizados em plantações no mundo inteiro (Moscardi, 1999). Desde 1975, países como os Estados Unidos da América utilizam vírus desta família no controle de pragas, como a aplicação de *Helicoverpa zea* NPV (HzNPV) registrada como GemStar™ em plantações de algodão no controle de pragas dos gêneros *Heliothis* e *Helicoverpa* (Couch, Ignoffo, 1981; Moscardi, 1999). Outro baculovírus importante foi o *Helicoverpa armigera* SNPV (HaSNPV), utilizado em plantações de algodão, soja e tomate infestadas pela lagarta *Helicoverpa armigera* em países como Austrália, China e Índia (Erlandson, 2008; Moscardi, 1999; Sun, Peng, 2007; Yang et al., 2012). Na China, os NPVs contra a *H. armigera* já foram utilizados em mais de 100 mil hectares/ano (Sun, Peng, 2007; Yang et al., 2012). Diversos bioinseticidas virais foram baseados no baculovírus *Cydia pomonella* granulovírus (CpGV), aplicados em pomares de maçã e pera contra a *Cydia pomonella* em diversos países da Europa, Argentina, Canadá e Rússia (Erlandson, 2008; Moscardi, 1999). No Brasil, alguns membros da família *Baculoviridae* também foram utilizados, como o *Spodoptera frugiperda* múltiplo nucleopoliedrovírus (SfMNPV) em mais de 20 mil hectares/ano em plantações de milho (Moscardi, 1999; Wolff et al., 2008) e *Anticarsia gemmatilis* múltiplo nucleopoliedrovírus (AgMNPV) em aproximadamente 2 milhões de hectares de plantações de soja em 2003/2004. Os incentivos para a utilização comercial deste baculovírus ocorrem desde a década de 80, quando a Empresa Brasileira de Pesquisa e Agropecuária (EMBRAPA) iniciou o maior programa de controle biológico contra

a lagarta *Anticarsia gemmatalis* (figura 4) (Moscardi, 1999; Moscardi et al., 2011). Este baculovírus também foi utilizado em outros países: Argentina, Paraguai e México (Moscardi, 1999; Moscardi et al., 2011). Como já mencionado, estes vírus também foram aplicados em áreas florestais. Um bom exemplo foi a formulação baseada em *Lymantria dispar* MNPV (LdMNPV) em florestas temperadas contra a ordem Lepidóptera. Ademais, estes ecossistemas apresentam maior estabilidade que áreas de monocultura agrícolas, o que permite que os membros da família *Baculoviridae* ocorram naturalmente e permaneçam no meio ambiente por mais tempo (Erlandson, 2008; Moscardi, 1999; Szewczyk, 2009).

Figura 4 - Foto da lagarta *Anticarsia gemmatalis*, praga da monocultura de soja, e microscopia eletrônica de um poliedro do baculovírus AgMNPV.

À esquerda, a lagarta *Anticarsia gemmatalis* se alimentando e promovendo danos à plantação de soja. À direita, uma foto de microscopia eletrônica ilustrando a formação dos poliedros (AgMNPV); ainda é possível observar a presença de vários nucleocapsídeos ODVs em seu interior.

Fonte: Adaptada do site da Embrapa (www.embrapa.br) e (Szewczyk et al., 2006). Com permissão.

A segunda utilização dos vírus desta família, veio com a descoberta de promotores fortes presentes nos genes estruturais muito tardios (*p10* e *poliedrina*) que são genes não essenciais. Em seguida, estes vírus passaram a ter grande relevância na biologia molecular como ótimas ferramentas para clonagem e expressão de genes heterólogos (Hofmann et al., 1995; Szewczyk et al., 2006). As primeiras proteínas expressas no sistema baculovírus foram o β -interferon (Smith et al., 1983) e a interleucina-2 (Smith et al., 1985), ambos produzidos em AcMNPV. O baculovírus melhor caracterizado até o momento é o AcMNPV, considerado o protótipo da família. Porém, tanto o AcMNPV quanto o BmNPV tornaram-se úteis como vetores versáteis de expressão de genes heterólogos (Maeda, 1989), por meio dos quais é possível se produzir títulos de proteína mais altos que em outros sistemas eucarióticos de expressão (Brown et al., 1991). Convenientemente, os genomas de baculovírus são

grandes e permitem a inserção de grandes fragmentos de DNA de aproximadamente 20 Kb. Por este motivo, consegue-se expressar múltiplos genes heterólogos, bem como possibilitam a inserção de sequências sinais específicas no gene “engenheirado”, fazendo com que a proteína expressa seja exportada para fora da célula (Brown et al., 1991; French et al., 1990). Este sistema ainda permite a expressão de glicoproteínas de mamíferos com modificações pós-traducionais do tipo N-glicosilação. Nestes casos, é necessário utilizar uma linhagem de inseto geneticamente modificada que apresenta a capacidade de produzir os N-glicanos com ácido siálico na cadeia terminal, uma vez que a glicosilação de insetos é mais simples que a de mamíferos (Harrison, Jarvis, 2006). No entanto, os baculovírus são de fácil manipulação em laboratório com diversas culturas de células bem estabelecidas.

A utilização mais recente estabelece os vírus da família como possíveis candidatos a vetores de expressão para o uso em terapia gênica. Estudos da década de 80 baseados em AcMNPV demonstraram que 35 linhagens de vertebrados foram infectadas por baculovírus, sem evidência de replicação mas com a presença de nucleocapsídeo no núcleo celular (Volkman, Goldsmith, 1983). Desde de então, o fenótipo BV passou a ser utilizado como ferramenta biotecnológica para transdução em diversos tipos celulares, incluindo células primárias, células progenitoras e células tronco apresentando baixa citotoxicidade (Airenne et al., 2013; Boyce, Bucher, 1996; Hu, 2006; Kost et al., 2005). Atualmente, estudos *ex vivo* e *in vivo* baseados em vetores de baculovírus para terapia gênica estão sendo realizados com sucesso em uma variedade de tumores (Kim et al., 2010; Wang, Balasundaram, 2010). Estes vírus ainda são candidatos promissores como adjuvantes vacinais, com imunidade protetora *in vivo* (Madhan, 2010).

1.6 O *Alphabaculovirus Anticarsia gemmatalis* múltiplo nucleopoliedrovírus

Como já mencionado, o baculovírus melhor caracterizado da família é o AcMNPV. Este foi o primeiro baculovírus a ter seu genoma completamente sequenciado (Ayres et al., 1994) e a ter sua composição proteica estrutural identificada (Braunagel, 2003). Embora a maioria dos estudos sejam realizados em AcMNPV, o baculovírus mais utilizado como agente de controle biológico foi o *Anticarsia gemmatalis* múltiplo nucleopoliedrovírus (AgMNPV), originalmente isolado da lagarta-da-soja *Anticarsia gemmatalis* (Lepidoptera: Noctuidae) em Campinas, em 1972 (Allen, Knell, 1977). Desde a década de 1980, alguns estudos foram realizados a fim de se identificar e distinguir os diferentes genótipos dos vírus coletados em

1979, determinado o isolado 2D como protótipo da espécie AgMNPV, que representava cerca de 40% dos isolados obtidos (Maruniak et al., 1999). O estabelecimento da linhagem celular UFL-AG-286 derivada da lagarta *Anticarsia gemmatalis* facilitou o cultivo *in vitro* deste baculovírus (Sieburth, Maruniak, 1988). Em 1989, o mapa físico do genoma deste vírus foi resolvido através de análises de restrição (Maruniak, 1989), o gene da poliedrina e sua sequência promotora foram identificadas (Zanotto et al., 1992). Posteriormente, as *hrs* do AgMNPV foram também elucidadas (Garcia-Maruniak et al., 1996). No ano de 1997, quatro linhagens celulares de Lepidópteros distintos foram verificadas quanto à suscetibilidade ao AgMNPV. Na linhagem UFL-AG-286 a infecção foi a mais produtiva, seguida da infecção em IPLB-SF-21 (derivada da lagarta *Spodoptera frugiperda*), sendo menos eficiente na linhagem IPRI-CF-124T (derivada de *Choristoneura fumiferana*) e abortiva na linhagem BM-5 (derivada de *Bombyx mori*) (Castro et al., 1997). Em 2001, foi construído o vírus recombinante (vAgGalA2) contendo o gene da β -galactosidase (β -gal) sob controle do promotor da poliedrina. Esta reconstrução mostrou capacidade de atingir altos níveis de expressão de proteínas heterólogas (Ribeiro et al., 2001). Em 2005, as larvas de *Anticarsia gemmatalis* infectadas com o AgMNPV recombinante (vAgEGTA-lacZ), foram monitoradas temporalmente durante a evolução das infecções primárias e secundárias nos diferentes tecidos através da expressão do gene lac-, em vinte quatro horas pós infecção todos os tecidos estavam completamente infectados pelo vAgEGTA-lacZ (Soares, Ribeiro, 2005). Em 2006, o sequenciamento de seu genoma foi concluído (132.242 bp e 40.427 resíduos) possibilitando a identificação de 152 ORFs (Open Reading Frames) com no mínimo 50 aminoácidos, ou seja, fases de leitura abertas ou potencialmente codificando genes (Figura 5) (Oliveira et al., 2006b). Dentre elas, foram identificadas três novas ORFs que não ocorrem em nenhum outro baculovírus e, portanto, podem ser exclusivas deste vírus. Além dessas, no genoma do AgMNPV-2D foram identificados 26 genes responsáveis pela expressão de proteínas estruturais, sendo que 22 são conservados em todos os baculovírus sequenciados até o momento: PIF-2 (*Ag12*), VP1054 (*Ag55*), VLF-1 (*Ag76*), GP41 (*Ag79*), VP91/P95 (*Ag82*), VP39 (*Ag86*), P33 (*Ag89*), ODV-E25 (*Ag91*), Helicase (P143) (*Ag92*), 38K (*Ag94*), P6.9 (*Ag96*), P40 (*Ag97*), P48 (*Ag99*), ODV-EC43 (*Ag107*), PIF-3 (*Ag114*), PIF-1 (*Ag117*), Alk-Exo (*Ag129*), P74 (*Ag134*), P49 (*Ag138*), ODV-E18 (*Ag139*), ODV-E27 (*Ag140*), ODV-E56 (*Ag144*) e quatro estão presentes em todos baculovírus que infectam Lepidópteras: 25K FP (*Ag61*), ODV-E66 (PIF-5) (*Ag49*), PK1 (*Ag151*) e Poliedrina (*Ag1*) (Garavaglia et al., 2012; Herniou et al., 2003).

Figura 5 - Esquema do genoma do AgMNPV-2D

As setas representam as 152 ORFs anotadas no genoma do AgMNPV-2D com 132.239 pb. Em vermelho estão indicadas as ORFs na fita positiva e em azul as ORFs na fita negativa. Em laranja foram representadas as *hrs*, sequências repetitivas. A linha preta no interior do ciclo, significa o conteúdo GC (em%).

Fonte: Adaptada de (Oliveira et al., 2006b). Com permissão.

No Brasil, o AgMNPV foi amplamente utilizado, proporcionando ao país consideráveis benefícios econômicos e ambientais (Moscardi, 1999; Moscardi et al., 2011). Porém, em 2007 o programa de controle biológico reduziu drasticamente a aplicação deste baculovírus para 300.000 ha/ano nas plantações de soja. Alguns fatores foram determinantes para a redução do programa, incluindo: a resistência dos agropecuaristas em usar métodos naturais, o tempo de morte das lagartas, que pode chegar a cinco dias e portanto, continuavam a danificar a plantação durante esse tempo, a manutenção constante de um isolado mais

virulento no campo, a necessidade do monitoramento das larvas de AgMNPV na plantação e a comum dupla ou e até tripla infestação com outras espécies de insetos nas plantações de soja, como a lagarta falsa medideira, *Pseudoplusia includens*, que não é infectada pelo AgMNPV, e a ocorrência de lagartas do gênero *Helicoverpa* (Moscardi et al., 2011) (www.embrapa.br). A introdução destas novas espécies foi ocasionada por um processo acumulativo de práticas inadequadas de cultivo em áreas muito extensas e adjacentes, associadas a um manejo inapropriado de agrotóxico, tornando o agroecossistema mais suscetível aos insetos-praga (Moscardi et al., 2011).

Recentemente, a análise de transcrição gênica global do AgMNPV-2D foi finalizada em duas linhagens distintas de Lepidópteras empregando a ferramenta de PCR em tempo real. Neste estudo, foram observados transcritos para 149 ORFs do AgMNPV-2D nas duas linhagens utilizadas (Oliveira et al., 2013). O padrão de expressão observado em UFL-AG-286 foi utilizado para montagem da rede regulatória (GRN) do AgMNPV-2D e foram levados em conta os aspectos da dinâmica temporal e informação da concentração dos RNA mensageiros produzidos durante o ciclo viral (Oliveira et al., 2013). A rede de expressão possibilitou identificar os genes com alta conectividade e grande relevância para integridade da GRN: *gp64* (Ag124) e *egt* (Ag 18) (Figura 6) (Oliveira et al., 2013). Além disso, foi possível identificar cinco comunidades (I-V) distintas, sendo que os genes que codificam para proteínas estruturais se concentraram densamente na comunidade I (Figura 6). Contudo, a análise de expressão gênica não significa que todas as proteínas serão traduzidas e que seus produtos serão funcionais (Tyers, Mann, 2003). Assim, o estudo com as proteínas que compõe a estrutura do AgMNPV, bem como aquelas que auxiliam a infecção, permite um maior entendimento da morfologia, multiplicação e possíveis proteínas celulares associadas a este baculovírus.

Figura 6 - Rede regulatória (GRN) baseada no transcriptoma do AgMNPV-2D a partir da infecção em células UFL-AG-286

A rede de expressão do AgMNPV foi representada com as 149 ORFs e seus componentes diretamente ligados (setas). O números se referem às ORFs (nós) e o tamanho de cada nó é proporcional ao valor de *betweenness centrality* (BC) e ao coeficiente de *stress* (SC). Os números em vermelho representam os 37 genes *core*. Estes dois parâmetro indicam a importância de cada ORF na rede do AgMNPV (programa Cytoscape v2.8.0). Além disso, foi possível identificar 5 comunidades (I-V) pelo método de clusterização do programa GLayer, com os genes que codificam a proteínas estruturais densamente concentrados na comunidade I.

Fonte: Adaptada do trabalho de (Oliveira et al., 2013). Com permissão.

1.7 Estudo proteômico na família *Baculoviridae*

As proteínas virais são reguladas principalmente por modificações pós-traducionais que promovem a interação direta vírus-célula (Steen, Mann, 2004; Zhang et al., 2010b). Geralmente, os estudos de proteomas de vírus ajudam a identificar suas proteínas estruturais e as proteínas celulares que podem ser levadas junto da partícula como: actina, β tubulina, fator de elongação 1 α e 2, HSP70 e 90 (Maxwell, Frappier, 2007). Além disso, proteomas virais

permitem relacionar novas proteínas que não foram previamente inferidas com os dados de sequenciamento e análise de transcriptoma. Neste sentido, o primeiro estudo utilizando abordagens proteômicas foi realizado no fenótipo ODV do AcMNPV e possibilitou a identificação de 44 proteínas neste fenótipo (Braunagel, 2003). Desde 2003, outros proteomas de ODV foram descritos, como: *Culex nigripalpus* nucleopoliedrovírus (CuniNPV) (Perera et al., 2007), *Helicoverpa armigera* nucleopoliedrovírus (HearNPV) (Deng et al., 2007), *Bombyx mori* nucleopoliedrovírus (BmNPV) (Liu et al., 2008), *Chrysodeixis chalcites* nucleopoliedrovírus (ChChNPV) (Xu et al., 2011) e o *Pieris rapae* granulovírus (Wang et al., 2011). Em 2010, Wang e colaboradores (Wang et al., 2010) descreveram pela primeira vez o proteoma do fenótipo BV do AcMNPV e identificaram 34 proteínas estruturais, entre elas 13 inéditas. Além disso, 11 proteínas celulares foram encontradas na estrutura do BV (Wang et al., 2010). O estudo mais recente e completo da família foi realizado em HearNPV. Empregando-se cinco abordagens distintas de proteômica foi possível identificar e quantificar todas as proteínas que compõem os dois fenótipos virais, inclusive as celulares (Hou et al., 2013). Além disso, todas as modificações pós-traducionais, como N-glicosilação, fosforilação e acetilação do N-terminal de todas as proteínas de BV e ODV foram analisadas (Hou et al., 2013).

Estes trabalhos identificaram um padrão de distribuição das proteínas encontradas nos dois fenótipos da família *Baculoviridae*. Entre as identificadas, oito estão presentes no envelope do ODV: P74 (*per os infectivity factors*, PIF0), PIF1, PIF2, PIF3, PIF4, PIF5 (ODV-E66), ODV-E25 e ODV-E56 (Figura 7) (Faulkner et al., 1997; Peng et al., 2012; Slack, Arif, 2007). Entretanto, apenas cinco destas proteínas são responsáveis pela formação de um complexo estável, formado pelas proteínas conservadas em todos os baculovírus: P74 (PIF0), PIF1, PIF2, PIF3 e PIF4 juntamente com a VP91/P95. Este complexo promove o contato inicial do vírus com as células epiteliais do intestino médio (Peng et al., 2012). Um estudo recente em SfMNPV também propõe a inclusão da proteína ODV-EC43 como parte das PIFs, renomeando-a como PIF-6 (Simon et al., 2012). Ademais, foi proposto o envolvimento desta proteína no complexo citado acima (Peng et al., 2012). As proteínas ODV-E25 e PIF5 (ODV-E66) possuem motivos transmembrana no N-terminal e ODV-E56 possui motivo transmembrana no C-terminal (Braunagel et al., 1996a; Russell, Rohrmann, 1993). As três proteínas mencionadas também fazem parte do envelope do ODV, e embora não estejam diretamente relacionadas à ligação do vírus com a célula podem apresentar ação sinérgica com o complexo de PIFs facilitando a infecção do ODV (Peng et al., 2012). A proteína de

envelope, ODV-E18, é compartilhada também pelo fenótipo BV, com papel importante durante a sua produção (McCarthy et al., 2008). Outra proteína, FP25K de envelope está diretamente envolvida com o tráfico de proteínas no envelope do ODV (Braunagel et al., 2004). A GP41 está presente entre o envelope e o capsídeo em uma estrutura chamada de tegumento e é essencial para a saída do nucleocapsídeo do núcleo (Figura 7) (Olszewski, Miller, 1997a; Pan et al., 2005).

Figura 7 - Estrutura ilustrativa dos fenótipos virais de um *Alphabaculovirus*

O BV está representado à esquerda, demonstrando o DNA fita dupla condensado pela proteína p6.9, que auxilia no empacotamento do material genético no nucleocapsídeo, formado predominantemente pela proteína VP39. No ápice do envelope bilipídico estão os peplômeros formados pela proteína GP64. À direita estrutura está representado o ODV e sua estrutura. Os nucleocapsídeos inseridos no tegumento, ocluídos no mesmo envelope. Além disso, estão representadas todas possíveis proteínas estruturais presentes no envelope viral. Fonte: Adaptada de Slack & Arif, 2007 (Slack, Arif, 2007). Com permissão.

As metodologias proteômicas possibilitaram a identificação de proteínas associadas ao nucleocapsídeo do BV e do ODV. Estas estruturas polares são formadas por uma garra na base e na outra extremidade o ápice (Figura 8). Os nucleocapsídeos são montados em uma estrutura eletrodensa mencionada acima, o estroma virogênico (Fraser, 1986). Estes apresentam similaridade nos dois fenótipos virais com muitas proteínas importantes compartilhadas: P6.9, VLF-1, VP39, P40, P49, VP80, PP78/83, ODV-E27, VP39, VP1054, P33 e 38K (Figura 8). A P6.9 é uma proteína pequena e básica, rica em resíduos de arginina e serina/treonina, codificada por um dos genes *core*. A arginina tem alta afinidade com o grupo fosfato do DNA, e nos baculovírus a p6.9 está diretamente ligada a condensação do DNA

viral (Wilson et al., 1987). A proteína VLF-1 mencionada anteriormente, também está presente no nucleocapsídeo e é fundamental para produção do mesmo nos dois fenótipos virais (Li et al., 2005). A VP39 é a principal proteína do nucleocapsídeo de BV/ODV, conservada em todos os baculovírus, e está intimamente relacionada com a reorganização da actina celular (Charlton, Volkman, 1991; Lanier, Volkman, 1998; Ohkawa et al., 2010). Ensaio realizado com duplo híbrido, demonstraram que a P40 (BV/ODV C42), PP78/83 e ODV-E27 estão relacionadas a base do nucleocapsídeo (Braunagel et al., 2001). Além dessas três proteínas, a VP80 também está presente nesta estrutura e seus homólogos são encontrados em todos os baculovírus que infectam Lepidópteros (Marek et al., 2011). A proteína p33 forma um complexo estável com a proteína supressora de tumor do p53 hospedeiro e parece estar intimamente ligada a apoptose (Figura 8) (Prihod'ko et al., 1999). A p33 também é necessária para formação de ODV e produção de BV (Wu, Passarelli, 2010). Já a proteína VP1054 possui homólogos em todos os baculovírus sequenciados até o momento e sua homologia com PUR α celulares (Marek et al., 2013) foi recentemente demonstrada em AcMNPV. Estas proteínas estão envolvidas em múltiplas funções celulares; uma delas é de se ligar a DNA fita simples e a RNA, preferencialmente em sequências repetitivas compostas de GGN, onde “N” pode ser qualquer nucleotídeo (Marek et al., 2013), ressaltando sua importância para formação dos nucleocapsídeos (Marek et al., 2013; Olszewski, Miller, 1997b).

Figura 8 - Microscopia eletrônica e estrutura ilustrativa do nucleocapsídeo viral

O nucleocapsídeo desnudado (em **A**) e agrupado (em **B**) foi representado com as proteínas estruturais e suas prováveis localizações em **C**. Esta estrutura é muito semelhante nos dois fenótipos virais. O diâmetro do nucleocapsídeo varia entre 40-70 nm diâmetro e 250-400 nm em comprimento. O tamanho do nucleocapsídeo pode variar conforme o tamanho do genoma viral.

Fonte: **A** e **B**, Zanotto (1985). **C**, adaptada de Slack & Arif, (Slack, Arif). Com permissão.

1.8 Glicoproteína 64 (GP64)

Como já mencionado anteriormente, a glicoproteína 64 de 60 kDa (GP64) é a maior proteína estrutural exclusiva do fenótipo BV presente somente nos nucleopoliedrovírus do Grupo I (Figura9) (Volkman, Goldsmith, 1984; Zanotto et al., 1993). Esta glicoproteína está envolvida na fusão do fenótipo BV com a célula hospedeira, fundamental para infecção secundária (Oomens, Blissard, 1999). Uma vez no interior da célula, regula a diminuição do pH, necessária para a liberação do nucleocapsídeo no citosol (Blissard, Wenz, 1992; Kingsley et al., 1999; Markovic et al., 1998). No grupo NPV II, a proteína F possui função similar à GP64, contudo esta tem origem evolutiva distinta (Jkel et al., 2000; Lung et al., 2002). Esta glicoproteína ainda é utilizada em vetores de terapia gênica, uma vez que tem capacidade de ligar-se a vários receptores celulares, entre eles uma variedade de receptores de células de mamíferos (Airenne et al., 2013; Hu, 2006; Kost et al., 2005; Wang, Balasundaram, 2010). Ainda é possível incorporar sequências ao N-terminal da proteína, tornando-a altamente imunogênica (Kost et al., 2005).

Figura 9 - Reconstrução filogenética baseada nas sequências de *gp64* dos NPV I

Nesta árvore filogenética foram utilizadas sequências de aminoácidos de 13 linhagens distintas de GP64. Um ortólogo de GP64 compartilhado pelo vírus de RNA o Thogoto (*Ortomyxoviridae*), que é vetorado por ácaros (Morse et al., 1992), foi usado como grupo externo.

A superfamília da GP64 presente nos baculovírus, que inclui também a glicoproteína (GP75) do Toghoto vírus (Família *Ortomyxoviridae*) (Morse et al., 1992), foi classificada como penetrina viral da classe III ou γ -penetrinas (Garry, Garry, 2008). Como característica estrutural comum, as penetrinas da classe III possuem alças de fusão internas, domínios de α -hélice estendidos e domínios centrais e uma pequena cauda citoplasmática na região carboxi-terminal transmembrânica (Kadlec et al., 2008). A GP64 do AcMNPV teve sua estrutura resolvida por cristalografia por raio X. Sua estrutura é alongada, em forma de um trímero composta por cinco domínios (I-V) (Figura 10 a) formados principalmente por folhas β , com o centro das α -hélices envolvido na trimerização (Figura 9b e c). Nesta estrutura, ainda foi possível mapear os resíduos conservados entre GP64 e a GP75 do Toghoto vírus envolvidos na fusão vírus-célula. Estes resíduos estão presentes na alça 1 (leucina 82) e alça 2

(fenilalanina 153 e alanina 154) do domínio I (Kadlec et al., 2008). Este achado corroborou o que havia sido descrito na estrutura da glicoproteína do vírus da estomatite vesicular (VSV) (Roche et al., 2006) e na estrutura da glicoproteína gB do herpesvírus-1 (Heldwein et al., 2006), também *penetrinas* virais da classe III. Posteriormente, a atividade da GP64 indicou que os domínios funcionais determinados no AcMNPV compartilhados entre os NPVs I, ectodomínio, o domínio transmembranico e, a cauda citoplasmática, eram surpreendentemente distintos no AgMNPV-2D (Li, Blissard, 2009). Assim, estudar em maior detalhe a GP64 no AgMNPV é fundamental não só para o entendimento da infecção viral, mas também para futuras aplicações deste baculovírus.

Figura 10 - Esquema da estrutura da GP64 de AcMNPV realizada por cristalografia de raio X

(a) Esquema dos cinco domínios resolvidos para GP64: I subdividido em Ia e Ib (resíduos 60-217), II (resíduos 44-59 e 218-271), III (resíduos 22-39 e 398-373), IV (resíduos 374-407) e V (resíduos 413-460). Nesta proteína, a sequência sinal está no resíduo 1-20 e cauda citoplasmática na região C-terminal entre os resíduos 461-482. (b e c) Representação do trímero da glicoproteína GP64.

Fonte: Adaptada de Kadlec et al., 2008 (Kadlec et al., 2008). Com permissão.

1.9 Métodos de Proteômica

As primeiras abordagens proteômicas tinham como objetivo mapear as proteínas obtidas de amostras biológicas (*E. coli*, *Mus musculus*, *Cavia porcellus*) e determinar o perfil proteico por géis bidimensionais (2DE) (Klose, 1975; O'Farrell, 1975; Scheele, 1975). Esta metodologia e os géis unidimensionais (SDS-PAGE) possibilitaram a separação e visualização das proteínas, mas a ausência de uma tecnologia sensível na época não permitia a identificação das mesmas (Graves, Haystead, 2002). Neste sentido, a primeira metodologia

descritiva foi a Degradação de Edman, que possibilitou o sequenciamento de proteínas a partir de amostras separadas por 2DE e a criação do primeiro banco de dados de proteínas (Aebersold et al., 1987; Aebersold et al., 1988; Edman, 1949). Desde a década de 1990, o método de detecção por Degradação de Edman foi amplamente substituído pela espectrometria de massas (MS) (Aebersold, Mann, 2003; Amster, 1996). Assim, a espectrometria de massas tornou-se padrão ouro para proteômica, permitindo que a massa molecular (relação massa/carga (m/z)) de uma proteína e os peptídeos oriundos de sua fragmentação fossem identificados com alta sensibilidade, em poucos segundos (Aebersold, Mann, 2003; Graves, Haystead, 2002; Steen, Mann, 2004; Tyers, Mann, 2003). Atualmente, as metodologias disponíveis possibilitam uma variedade de aplicações, como caracterizar todas proteínas expressas em um sistema, identificar todas suas isoformas e modificações pós-traducionais, as interações proteína-proteína, as prováveis localizações e/ou compartimentalização, a quantificação das proteínas presentes e também a descrição estrutural de cada proteína (Aebersold, Mann, 2003; Graves, Haystead, 2002; Steen, Mann, 2004; Tyers, Mann, 2003; Wiese et al., 2007). Basicamente, um espectrômetro de massas é composto por uma fonte de íons, um analisador de massa, um detector e um sistema de aquisição de dados (Figura 11) (Cantú, 2008). As duas fontes de ionização mais utilizadas em um espectrômetro podem ser dos tipos: MALDI (*matrix assisted laser desorption/ionization-time of flight*) e ESI (*Electrospray*) (Aebersold, Mann, 2003; Cantú, 2008; Fenn et al., 1989; Karas, Hillenkamp, 1988).

As amostras proteicas menos complexas são normalmente utilizadas em espectrômetros de massas com fonte do tipo MALDI. Por outro lado, amostras biológicas complexas são mais indicadas aos espectrômetros de massa com fonte do tipo ESI, que ainda permitem integrar uma cromatografia líquida de alta performance (HPLC). Assim, os peptídeos digeridos dos géis SDS-PAGE ou 2DE podem ser eluídos em ordem de hidrofobicidade e diretamente injetadas no espectrômetro de massas (Figura 11) (Aebersold, Mann, 2003; Steen, Mann, 2004). Nesta etapa, os peptídeos são ionizados com um alto potencial elétrico de forma a preservar a estrutura polipeptídica, e as espécies são transferidas para fase gasosa (Cantú, 2008; Steen, Mann, 2004). Estes peptídeos vaporizados entram no sistema a vácuo e são encaminhados para os analisadores de massas: quadrupólos (Q), *ion-traps*, *time-of-flight* (TOF), *orbitrap*, entre outros (Figura 11) (Aebersold, Mann, 2003; Cantú, 2008). Os analisadores de massas ainda podem ser acoplados em sequência, o que faz com que os íons separados (íons percussores-MS) pela relação massa/carga no primeiro analisador

sejam fragmentados na próxima etapa. Os íons mais intensos e mais informativos da etapa anterior são submetidos a dissociação por colisão (CID) de gás inerte (argônio, nitrogênio ou hélio) e o espectro final (MS/MS) é detectado.

Figura 11 - Esquema de um espectrômetro de massas: fontes de ionização e analisadores de massas

Os espectrômetro de massas são compostos por uma fonte de íons, um analisador de massa e um detector. Na figura foram representadas os dois tipos de fontes de ionização: MALDI e ESI. Logo após a ionização, os íons são encaminhados para os analisadores de massa: quadrupólos (Q), *ion-traps*, *time-of-flight* (TOF), *orbitrap*, entre outros. Além disso, estão representado dois analisadores de massas acoplados em sequência, MALDI-TOF/TOF (*time-of-flight*) e ESI-Q-TOF (Quadrupolo/*time-of-flight*). Ao final, os espectros de massas (MS/MS) gerados são detectados por eletromultiplicadoras, que são os detectores mais comuns. Fonte: Adaptada de (Aebersold, Mann, 2003; Steen, Mann, 2004). Com permissão.

Posteriormente, estes espectros (MS/MS) que contém a relação m/z e a sequência de peptídeos são utilizados para análises de busca de proteínas em banco de dados pelos *softwares* Mascot, Masslynx, PeptideSearch, Sequest entre outros (Cantú, 2008). O *software* Mascot, por exemplo, faz a busca dos espectros de massas obtidos (experimental) contra as os fragmentos teóricos preditos para todos os peptídeos depositados nos bancos de dados. O cálculo envolve a probabilidade das m/z de um fragmento prático e teórico coincidirem de maneira aleatória, expresso como o negativo do logaritmo deste número, que é representado em *score*. Quanto maior este valor, maior a confiabilidade do resultado gerado (Cantú, 2008). Os dados gerados pelo Mascot ainda podem ser validados por outro *software*, o Scaffold.

Neste *software*, dentre os algoritmos desenvolvidos, há a validação das proteínas encontradas pelo Mascot com valor expresso em porcentagem, e a cobertura de quantos espectros são gerados para cada proteína (Keller et al., 2002; Nesvizhskii et al., 2003).

As metodologias de proteômica foram amplamente utilizadas para explorar a composição proteica dos vírus, isto é, a descrição de seus proteomas estruturais. Apesar da importância do AgMNPV-2D, nenhum estudo foi realizado para obtenção e caracterização das proteínas estruturais de seus fenótipos, BV e ODV.

2 OBJETIVOS

2.1 Objetivos Gerais

Este trabalho teve como objetivo central identificar as proteínas presentes nos dois fenótipos do AgMNPV-2D em duas linhagens de células de insetos distintos. Com os resultados obtidos, outro objetivo relevante foi proposto, analisar *in silico* as possíveis diferenças na ORF-124, *gp64*, em populações selvagens de *Anticarsia gemmatalis* múltiplo nucleopoliedrovírus coletadas em diferentes regiões da América do Sul, no período de 1980 a 1990. Estes resultados poderão incrementar o conhecimento sobre a morfologia deste baculovírus, assim como revelar proteínas novas que auxiliam na interação vírus-hospedeiro.

2.2 Objetivos Específicos

- (i) Identificar diferenças qualitativas de proteínas expressas nos dois fenótipos virais (ODV e BV) do AgMNPV-2D, em duas linhagens distintas de células de lepidópteros;
- (ii) Identificar as proteínas celulares presentes nos dois fenótipos virais;
- (iii) Identificar os possíveis polimorfismos em sequências de *gp64* obtidas por pirosequenciamento em 17 populações de *Anticarsia gemmatalis* múltiplo nucleopoliedrovírus (AgMNPV);
- (iv) Amplificar, clonar e sequenciar a *gp64* de 17 populações selvagens amostradas em diferentes regiões geográficas da América do Sul, no período de 1980 a 1990.

3 METODOLOGIA

3.1 Análise do proteoma do AgMNPV-2D por espectrometria de massas

O baculovírus escolhido para o desenvolvimento deste estudo foi o AgMNPV isolado 2D, por ser o genótipo mais frequente originário da população selvagem em 1972 (Allen, Knell, 1977). Em nosso laboratório, este isolado teve o genoma sequenciado e 152 ORFs foram identificadas (Oliveira et al., 2006b). Para identificar o conjunto de proteínas estruturais presentes nos dois fenótipos virais, BV e ODV, foram utilizadas: *(i)* duas linhagens celulares originárias de insetos distintos, *(ii)* dois métodos de separação de proteínas e *(iii)* duas abordagens proteômicas por espectrometria de massas. Um esquema da metodologia utilizada está representado a seguir (Figura 12).

Figura 12 - Esquema geral da metodologia descrita para o estudo das proteínas estruturais do AgMNPV-2D em duas linhagens celulares estudadas.

3.1.1 Cultivo Celular

As linhagens celulares UFL-AG-286, derivada da lagarta *Anticarsia gemmatalis* (Sieburth, Maruniak, 1988), e IPLB-SF-9, clone da IPLB-SF-21 proveniente de ovário da lagarta *Spodoptera frugiperda* (Vaughn et al., 1977) foram mantidas em nosso laboratório por meio de cultivo e congelamento em nitrogênio líquido. Durante o cultivo, a linhagem UFL-AG-286 foi mantida em meio Grace (Gibco BRL), suplementada com 10% de soro fetal bovino (SFB, Invitrogen) e a linhagem IPLB-SF-9 foi mantida em meio SF900 (Gibco BRL). Ambas linhagens permaneceram à 28 °C (Celsius), com dois repiques semanais. As duas linhagens foram utilizadas por serem consideradas permissivas e semipermissivas, respectivamente.

3.1.2 Infecção Viral

As metodologias para infecção, amplificação dos estoques virais e a titulação de baculovírus foram utilizadas de acordo com o procedimento descrito por (King, Possee, 1992). Primeiramente, foi utilizado um inóculo viral de AgMNPV-2D com titulação de 10^6 unidades infecciosas/ml. O estoque viral foi realizado em linhagem UFL-AG-286, em cerca de $3,6 \times 10^6$ células/ml. Após essa etapa, as células em meio Grace foram transferidas para garrafa de cultivo de 150 cm², onde permaneceram por aproximadamente 2 h a 28 °C. As células que não aderiram a garrafa foram descartadas e o inóculo viral foi adicionado em 10 ml de meio de cultivo, visando atingir a multiplicidade de infecção desejada. A garrafa foi mantida por 1 h a 28 °C para adsorção viral. Após este período, o meio de cultura foi substituído pelo meio de cultivo novo (13 ml) suplementado com 10% de soro fetal bovino (SFB). As células infectadas foram incubadas a 28 °C por cinco dias. Por fim, o sobrenadante de cultivo foi centrifugado a 1.000 g por 10 min (minutos) em temperatura ambiente para retirada das células. Do sobrenadante (estoque viral), foi retirada uma alíquota para a titulação viral. O estoque viral titulado foi armazenado a 4 °C.

3.1.3 Titulação Viral

O título viral foi determinado por *end-point dilution* em placa de 96 cavidades, sendo que em cada uma destas foram adicionados 4×10^4 células UFL-AG-286 em meio Grace.

Esta placa permaneceu em estufa por 2 h a 28 °C para adesão celular. Neste período foi preparada uma alíquota do estoque viral de forma serial na base 10 (10^{-1} até 10^{-8}). Após as 2 h as células que não aderiram foram descartadas e o meio foi substituído por 100 µl da diluição viral de cada base, em triplicata. A placa permaneceu por 1 h a 28 °C para adsorção viral e posteriormente o meio foi trocado por meio Grace suplementado com 10% de SFB. A placa foi incubada a 28 °C por seis dias e após este período o efeito citopático foi observado em microscópio ótico pela presença de corpos de oclusão no interior do núcleo da célula. O título viral foi determinado pelo método de TCID₅₀ (Reed, Muench, 1938), dose viral capaz de matar 50% das células.

3.1.4 Amplificação e Produção Viral

O estoque viral foi amplificado seguindo o mesmo procedimento da infecção, porém nesta etapa foram utilizadas garrafas de cultivo de 300 cm² com infecções de 1 MOI (multiplicidade de infecção). Esta infecção garantiu que as células continuassem crescendo e que novos vírus se formassem e infectassem outras células. Após 72 h.p.i de infecção foram observados corpos de oclusão no interior do núcleo celular e o sobrenadante foi centrifugado a 1.000 g por 10 min à temperatura ambiente, para retirada dos restos celulares. O estoque viral foi armazenado a 4 °C e novamente titulado. A produção viral seguiu o mesmo padrão da amplificação do estoque. Mas, para a padronização dos experimentos, foram testados 0,1 MOI, 0,5 MOI, 1 MOI, 3 MOI e 5 MOI (nas duas linhagens celulares). As culturas infectadas, UFL-AG-286 e IPLB-SF-9, originadas destas infecções foram centrifugadas a 1.000 g por 10 min, e desta vez o sobrenadante foi coletado após 48 h.p.i para análise das proteínas de BV, e os corpos de oclusão (OB) foram coletados as 72 h.p.i para análise das proteínas do ODV presentes no precipitado. Ambos os fenótipo virais foram armazenados a 4 °C para os experimentos descritos a seguir.

3.1.5 Concentração viral e dissolução dos corpos de oclusão

Inicialmente foram concentrados os vírions oriundos dos sobrenadantes de infecção por AgMNPV-2D que continham o fenótipo BV. Cada sobrenadante foi ultracentrifugado em colchão de sacarose 25% (w/w), 5 mM NaCl e 10 mM EDTA. Para isso, os vírions foram aplicados cuidadosamente sob 3 ml de sacarose preparados em seis tubos de policarbonato

($V_{\text{total}} = 38\text{ml}$). Todos os tubos foram adequadamente pesados em balança analítica e ultracentrifugados (Beckman, rotor SW28) a 24.000 r.p.m (rotação por minuto) por 1 h e 15 min a 4 °C. Já para precipitados da seção 3.1.4, o procedimento de concentração foi diferente, uma vez que os OBs se encontravam no interior da célula. Assim, para promover a ruptura das células íntegras cada precipitado (OBs + células) foi tratado com SDS 0,5%, liberando os corpos de oclusão que contem o fenótipo ODV. Para cada 10^7 células/ml foi utilizado 1 ml da solução SDS 0,5%. Os OBs foram centrifugados a 5.500 g por 10 min em temperatura ambiente. O sobrenadante foi descartado e o precipitado foi ressuspensionado em 1 ml de solução NaCl 0,5 M e novamente centrifugado. Em seguida, o sobrenadante foi descartado e os corpos de oclusão com os ODVs foram ressuspensionados em solução salina de fosfato (PBS) pH 6,2, contados em Câmara de Neubauer e armazenados em tubos de 2 ml a -20 °C.

Para facilitar a identificação dos polipeptídios estruturais do fenótipo ODV embebidos nos corpos de oclusão, utilizamos a metodologia de dissolução de poliedros, visando diminuir a contaminação por poliedrina. Para isso, os vírions foram dissolvidos em solução de carbonato de sódio 1 M, conforme metodologia descrita por (Miller, Dawes, 1978). Desta forma, foi adicionado 55 µl de carbonato de sódio 1 M em 500 µl de isolados distintos de ODV em UFL-AG-286, por trinta minutos à temperatura ambiente. Posteriormente, foi adicionado 60 µl de solução Tris-HCl com pH=7,6 e centrifugado por dez minutos a 5.500 g. Ao final da centrifugação, os vírions estavam no sobrenadante e foram transferidos para um novo tubo.

3.1.6 Purificação das Partículas Virais

Os dois fenótipos virais foram purificados por ultracentrifugação em gradiente de sacarose conforme metodologia descrita em (O'Reilly et al., 1993). Para a purificação de BV foram utilizados 2 ml do concentrado descrito acima (seção 3.1.5). Este volume foi aplicado cuidadosamente sobre gradiente linear de sacarose (25% - 60% (w/w)) previamente preparado em tubo de 17 ml (rotor SW32.1 Ti) e em seguida ultra centrifugado (Beckman), a 26.500 r.p.m por 3 h a 4 °C. Os BVs foram visualizados através da formação de uma banda branca que se forma entre 47% e 49% no gradiente de sacarose. Esta banda foi coletada com volume mínimo para minimizar a contaminação por sacarose e, posteriormente diluída em 15 ml de PBS distribuídos em tubos SW32.1 Ti e ultracentrifugado novamente a 24.000 r.p.m por 1h e meia a 4 °C. O sobrenadante foi descartado e o precipitado ressuspensionado em 1 ml de

solução de PBS pH 6,2 e inibidor de protease em uma relação 1:100 (GE Healthcare). Já para a purificação dos ODVs foram utilizados 2 ml do vírus obtido da dissolução no seção 3.1.6, descrito anteriormente. Os ODVs foram colocados sobre o gradiente de sacarose (40% - 65%) previamente preparados em tubos SW28 e ultracentrifugados (Beckman, rotor SW28) a 26.500 r.p.m por 3 h em temperatura ambiente, resultando na formação de uma nítida banda branca (54% - 56%) contendo os ODVs. Essa banda foi em seguida coletada e diluída em 25 ml de água Milli-Q, e centrifugada a 5.500 g por 10 min para garantir máxima remoção de sacarose da amostra viral. O sobrenadante foi descartado e o precipitado foi ressuspensionado em 2 ml de água Milli-Q com inibidor de protease 1:100 (GE Healthcare), a fim de se garantir a integridade das proteínas, e posteriormente estocado a -20 °C.

3.1.7 Quantificação proteica

A concentração proteica medida para cada fenótipo viral proveniente das duas linhagens celulares foi feita pelo método colorimétrico de Bradford (Bradford, 1976). Esse método é utilizado para determinar a concentração de proteínas em solução e consiste na formação de um complexo entre o corante *Comassie Brilliant Blue* G-250 e as proteínas em solução. O valor de absorbância é proporcional a quantidade de proteína presente em solução, medido em relação a uma concentração conhecida. Este método é utilizado especialmente para quantificar as concentrações de proteínas usadas em gel eletroforese unidimensional. Para a construção da curva padrão, utilizou-se a albumina sérica bovina (BSA), que forma um complexo corante-BSA, com coeficiente de extinção constante numa faixa de 10 vezes a sua concentração diluída, permitindo portanto quantificar as proteínas dos dois fenótipos nas duas linhagens. A curva padrão de BSA foi calculada diretamente pelo *software* do espectrofotômetro NanoDrop (Thermo Scientific). Desta forma foi possível calcular a concentração de proteínas da amostra ($\mu\text{g}/\mu\text{l}$). Foram preparadas soluções para composição da curva padrão, a partir do estoque de BSA (1,0 mg/ml) diluído em água Milli-Q em concentrações de 5, 10, 20, 30 e 40 $\mu\text{g}/\mu\text{l}$ ($V = 500 \mu\text{l}$), em triplicatas. Dessas soluções foram retiradas alíquotas de 50 μl . Cada isolado viral proveniente das duas linhagens celulares, com concentração desconhecida de proteína foi preparado com fator de diluição 1:2, também em triplicata. Em seguida, foi adicionado 1 ml de solução de Bradford (1:4) com tempo de incubação de 5 min. Para a leitura de absorbância, foi utilizado o espectrofotômetro NanoDrop, já programado para a leitura com Bradford em 595 nm.

A concentração proteica para as eletroforeses bidimensionais (2D) foi determinada pelo 2D Quant-Kit (Amersham Biosciences). Inicialmente, foi construída uma curva padrão com a albumina sérica bovina (BSA) com concentração de 1,0 mg/ml e foram preparadas soluções com 0, 10, 20, 30, 40 e 50 µg/µl de BSA em triplicata. Cada solução foi precipitada e foi adicionado 100 µl de solução de cobre e 400 µl de água MiliQ ($V = 500 \mu\text{l}$). Posteriormente, foi adicionado 1ml de reagente colorimétrico em cada tubo incubado por 20 minutos abrigado da luz. Em seguida, foi realizada a leitura de absorbância em 480 nm no espectrofotômetro NanoDrop e a curva padrão foi montada. O mesmo protocolo foi realizado com os isolados virais purificados seguindo orientação do fabricante. Os valores de absorbância foram utilizados na curva padrão para determinar a concentração proteica das amostras a serem submetidas a eletroforese bidimensional (2DE). As concentrações de proteína foram calculadas considerando a leitura de absorção (OD) da proteína corrigida para o branco, o fator de diluição da proteína e o volume utilizado no procedimento.

3.1.8 Precipitação proteica

Para focalização isoeétrica adequada, as proteínas virais foram precipitadas previamente com acetona com objetivando-se eliminar possíveis interferentes que poderiam prejudicar as próximas etapas. Deste modo, foi utilizada uma alíquota de 300 µl de PBS (pH=6,2) de cada amostra previamente quantificada com cinco vezes o volume de acetona gelada, que foi incubado no gelo por 30 min. Em seguida, cada amostra foi centrifugada por cinco minutos à 21.000 g, à 4 °C. O sobrenadante foi descartado e o precipitado foi mantido à temperatura ambiente até a evaporação total da acetona. Ao final, cada precipitado foi ressuspensionado em 150 µl de tampão redutor (0,5 M Tris-HCl pH 6,8, β - mercaptoetanol, 0,05% (w/v) solução azul de bromofenol, glicerol e solução dodecil sulfato de sódio (SDS) 10%) para as eletroforeses SDS-PAGE e em 500 µl de solução de rehidratação (0,5% de tampão uréia 7M e tiuréia 2M, 4% Chaps, 1% DTT, 1% de solução IPG (pH 4-7 ou pH 3-10) e azul de bromofenol) para as eletroforeses bidimensionais (2DE).

3.1.9 Eletroforese SDS-PAGE

A análise em gel de poliacrilamida contendo SDS foi utilizada para caracterizar as proteínas virais. O método baseia-se na separação das proteínas pela sua sequência primária

ou massa molecular. Para a eletroforese SDS-PAGE foi utilizada a cuba Hoefer SE 600 Ruby (GE Healthcare). Os géis SDS-PAGE possuíam 1,0 mm de espessura (18 x 16 cm). Para o gel de empilhamento foram utilizados géis de 4% de poliacrilamida (0,5 M Tris-HCl pH 6,8, SDS 10%, acrilamida/bis 30%, 10% persulfato de amônia, 5 µl N,N,N',N'-Tetramethyl-1,2-diaminomethane-TEMED) para concentrar as proteínas. Para a separação das proteínas foram utilizados géis de poliacrilamida 12% preparados com 1,5 M Tris-HCl pH= 8,8, SDS 10%, acrilamida/bis 30%, 10% persulfato de amônia, 5 µl N,N,N',N'-Tetramethyl-1,2-diaminomethane-TEMED. Cada fenótipo BV e ODV purificado foi ressuspendido em tampão redutor e foi aquecido a 95°C por 5 min. O padrão de massas moleculares utilizado 10 a 250 kDa (Fermentas) foi aplicado ao gel, juntamente com as amostras e a eletroforese ocorreu a 110 V em tampão de corrida. Após a eletroforese, o gel foi corado com *Comassie Blue Brilliant R-250* por 1 hora e descorado em solução etanol 20% e ácido acético glacial 5% em água. Os géis foram digitalizados no ImageScanner (Ge Healthcare) para que as informações qualitativas sobre as proteínas estruturais de cada fenótipo viral pudessem ser identificadas.

3.1.10 Eletroforese Bidimensional (2DE)

Nesta etapa as proteínas foram separadas de acordo com seu ponto isoelétrico (*pI*) na primeira dimensão, e na segunda dimensão de acordo com sua massa molecular por meio de eletroforese em gel SDS-PAGE (O'Farrell, 1975). Para a separação na primeira dimensão iniciamos as triplicatas experimentais com 135 µg do fenótipo ODV (UFL-AG-286) em 250 µl de solução de reidratação (descrita na seção 3.1.9). Primeiramente, a hidratação ocorreu na fita IPG com gradiente de pH pré-fixado (fita de 13cm com pH linear de 3 - 10 e fita de 13cm com pH linear de 4 - 7) posicionada sobre a amostra à temperatura ambiente. Após 16 horas foi iniciada a focalização isoelétrica no aparelho Ethan IPGphor III (GE Healthacare) com as seguintes etapas: (S1) 500V por 500Vhr, (G1) 1000V por 750 Vhr, (G2) 8000V por 8000Vhr e (S2) 8000V por 8000Vhr. Ao final da focalização isoelétrica, cada fita IPG foi equilibrada em soluções de equilíbrio: A (Tris-HCl 50 mM pH 8,8, uréia 6 M, SDS 2%, glicerol 30%, diotiotreitól (DTT) 65 mM e azul de bromofenol 0,002%) por 15 min sob agitação; e B (Tris-HCl 50 mM pH 8,8, uréia 6 M, SDS 2%, glicerol 30%, iodoacetamida (IOA) 100 mM e azul de bromofenol e 0,002%) durante 15 min sob agitação. Após o equilíbrio, cada fita IPG foi cuidadosamente posicionada sobre um gel SDS-PAGE 12%, o padrão de massas moleculares

de 10 a 250 kDa (Fermentas) foi aplicado e cada fita IPG foi imobilizada com 0,5% agarose e azul de bromofenol 0,002%, para a separação das proteínas na segunda dimensão. Na eletroforese SDS-PAGE, utilizamos novamente a cuba Hoefer SE 600Ruby (GE Healthcare), agora acoplada ao sistema de banho circulante a fim de se manter tampão de corrida (Tris base, Glicina e SDS, pH 8,3) a 20 °C, uma vez que as triplicatas foram corridas juntas. A corrente inicial foi de 25 mA/gel por 30 min e posteriormente aumentada para 35 mA/gel até que o azul de bromofenol atingisse o final do gel. Para a detecção dos *spots*, os géis bidimensionais foram fixados com solução etanol 40% e ácido acético 15% em água por 30 min, corados com *Comassie Blue Brilliant R-250* por 1 hora e descorado em solução etanol 20%, ácido acético glacial 5% em água. Para cada fenótipo foram realizadas as triplicatas do gel 2DE. As imagens dos géis foram digitalizadas no ImageScanner (Ge Healthcare) e analisadas no *software* Image Master 2D Platinum 7.0 (Ge Healthcare) para a identificação e comparação dos *spots*.

3.1.11 Digestão proteica in gel

As bandas foram manualmente recortadas do gel SDS-PAGE (seção 3.1.10) com ajuda de um estilete e cada banda foi transferida para um tubo 1,5 ml e os *spots* identificados nas triplicatas dos géis 2DE (seção 3.1.10) foram retirados manualmente com ajuda de uma ponteira e também transferidos para um tubo 1,5 ml. Inicialmente, as bandas e os *spots* extraídos foram submetidos a um procedimento de descoloração em metanol 50% / ácido acético 2,5% por 3 horas. Em seguida, esta solução foi retirada e as bandas e os *spots* foram desidratados por duas vezes com 200 µl de acetonitrila (ACN) 100% por 5 min, intercalado com secagem a vácuo em *Speed-vac* por 2 min. Seguimos para a redução das proteínas com 30 µl em DTT 10 mM em Bicarbonato de amônio (NH_4HCO_3) 50 mM por 30 min e a alquilação com 30 µl de IOA 50 mM em NH_4HCO_3 50 mM por 30 min no escuro. Após, as bandas e os *spots*, foram reidratados com 200 µl NH_4HCO_3 100 mM por 10 min, desidratados com 200 µl de acetonitrila 100% por 5 min e novamente foram reidratados com NH_4HCO_3 100 mM. Todas as amostras foram evaporadas por 5 minutos a vácuo em *Speed-vac*. Neste período, a solução de tripsina foi preparada em gelo, 20 µg da Tripsina Gold Mass Spectrometry Grade (Promega) em 1ml de NH_4HCO_3 50 mM, concentração final de 20 ng/µl. Em seguida, todas as bandas e *spots* foram submetidos a hidrólise triptica. Para os *spots* e bandas menos intensos foram adicionados 0,3 µg e 0,7 µg da solução de tripsina,

respectivamente. Já para os *spots* e bandas mais intensos foram utilizados 0,5 µg e 0,9 µg da solução de tripsina, respectivamente. Os tubos permaneceram no gelo por 30 min. Posteriormente, todos foram centrifugados, o excesso de solução de tripsina foi retirado e foram adicionados 5 a 20 µl de bicarbonato de amônio 50 mM. A hidrólise trípica foi incubada por 16 horas a 37 °C. No dia seguinte, os peptídeos foram extraídos do gel utilizando 30 µl de ácido fórmico 5% e este passo foi repetido para garantir o máximo de extração dos peptídeos. O volume final de cada banda e *spot* foi transferido para um novo tubo e concentrado em *Speed-vac* por 3 horas, até restar apenas 1 µl por tubo. As amostras foram armazenadas a -20 °C até a realização das análises por espectrometria de massas.

3.1.12 Espectrometria de massa e análise dos espectros de massas dos peptídeos tríplicos

Para análise das amostras utilizamos o espectrômetro de massas (MS) no grupo de Espectrometria de massas localizado no LNBio (Laboratório Nacional de Biociências) em Campinas - SP coordenado pela Dra. Adriana Paes Leme e o MS do centro de proteômica do *Interdisciplinary Center for Biotechnology Research* da Universidade da Flórida - Flórida (USA) em colaboração com o professor James E. Maruniak (*Associate Professor of Virology*). O espectrômetro de massas (MS) é constituído por uma fonte de íons, um analisador de massas, um detector e um sistema de aquisição de dados. Para análises dos *spots* obtidos dos géis 2DE foi utilizado um espectrômetro de massas com fonte de ionização MALDI (*matrix assisted laser desorption/ionization*) acoplada ao analisador de voo em série do tipo Q-TOF (*quadrupole-time of flight*) Premier (Micromass MS-Technologies). As amostras foram ressuspendidas em 5 µl de ácido de Trifluoroacético (TFA) 0,1%, dos quais 2 µl foram aplicados na placa de MALDI. Foi utilizado ainda 2 µl de PEG como amostra referência e 2 µl matriz de α -ciano 8,6 mg/ml para ambos. A análise no espectrômetro foi realizada com potencial de aceleração de 20kV, no modo íon positivo refletor, calibrada para ionização dos peptídeos por 60 segundos e os 20 íons mais intensos foram encaminhados para o Q-TOF. Para aquisição dependente de informação, o *scanner* foi programado para m/z de 400 a 1800 e os 4 íons mais intensos foram adquiridos para a dissociação induzido por colisão (CID). Os espectros de massas (MS/MS) de cada ciclo foram adquiridos.

Para identificação das proteínas em LC-MS/MS utilizamos o espectrômetro de massa quadrupolo híbrido TOF (QSTAR elite, Applied Biosystems, Framingham, MA) com a fonte de ionização ESI (*Electrospray*). As proteínas previamente digeridas foram injetadas no

capilar (LC Packing PepMap) e dessanilizadas por 5 minutos com fluxo de 3 µl/min de 0,1% v/v ácido acético. Após este processo, as amostras foram injetadas por nanofluxo nas colunas empacotadas C18 (LC Packing PepMap) para entrada no HPLC. A eluição do gradiente começou com 3% do solvente B (0,1% v/v ácido acético, 96,9% v/v acetonitrila em 3% v/v H₂O) e 97% do solvente A (1% v/v ácido acético, 3% v/v acetonitrila e 96,9% v/v H₂O) e terminou com 60% do solvente A, 40% do solvente B por 30 min para separação das proteínas. A focalização e a ionização foram calibradas para 225 V e 2400 V, respectivamente. Para aquisição dependente de informação, o *scanner* foi programado para m/z de 400 a 1800 e os 4 íons mais intensos foram adquiridos para a dissociação induzido por colisão (CID). Os MS/MS de cada ciclo foram adquiridos.

3.1.13 Análise de dados e mapeamento in silico no genoma do AgMNPV-2D

Os espectros de massas (MS/MS) de cada amostra foram adquiridos e analisados pelo programa Mascot (Matrix Science, London, UK; *version* 2.2.2), contra o genoma do AgMNPV-2D e um banco de Lepidóptera montado a partir do NCBI (<http://www.ncbi.nlm.nih.gov/>). Os parâmetros do Mascot foram estabelecidos com tolerância de fragmentação de massa dos íons de 0,05 Da até 0,50 Da. As modificações variáveis foram especificadas: iodoacetamidação da Cys, a desaminação da Asn e Gln e oxidação da Met. Além do Mascot, também utilizamos o programa Scaffold (*version* Scaffold-03-00-04, Proteome Software Inc., Portland, OR) para validar a identificação dos peptídeos e proteínas a partir dos dados de MS/MS. Para a identificação dos peptídeos foi estabelecido o valor de probabilidade acima de 95,0% em relação ao algoritmo Peptide Prophet (Keller et al., 2002) e para confirmação das proteínas valores de probabilidade acima de 99% com no mínimo 2 espectros únicos de identificação em relação ao algoritmo Peptide Prophet (Nesvizhskii et al., 2003). Para a predição de função teórica e localização das proteínas celulares identificadas foi utilizado o banco de proteínas Uniprot (<http://www.uniprot.org/>) e também o banco de dados de proteínas para a procura de cristais resolvidos (<http://www.rcsb.org/>).

3.1.14 Western blot para confirmação dos dados obtidos por espectrometria de massas

Nesta etapa duas amostras de cada fenótipo foram testadas com anticorpos policlonais contra para as proteínas do AgMNPV-2D: PTP, ODV-E26, PIF-3, LEF-1 e GP64. Os anticorpos foram produzidos em camundongos pelo aluno Daniel Mendes e gentilmente cedidos pelo Prof. Bergmann Moraes Ribeiro (Universidade de Brasília - UnB). Para confirmar os dados obtidos por espectrometria de massas, os *western blots* foram realizados em minigéis de poliacrilamida, preparados com géis de empilhamento 5% e gel concentrador 12% (conforme descrito no seção 3.1.10). Cada fenótipo BV e ODV purificado foi ressuspendido em tampão redutor aquecido a 95 °C por 5 min. O padrão de massas moleculares de 10 a 250 kDa (Fermentas) foi aplicado ao gel juntamente com as amostras e a eletroforese ocorreu a 110 V em tampão de corrida. Após a eletroforese, cada gel foi incubado com 2 papéis filtro e 01 membrana de nitrocelulose (Biorad) do tamanho do gel em tampão de transferência Bjerrum e Schaffer-Nielsen (48 mM de Tris; 39 mM de glicina; 20% metanol, pH 9,2) por 15 min. As proteínas de cada gel foram transferidas a 15 V e 300 mA por 15 min (Aparelho Transblot SD semi-dry transfer cell - Biorad). Cada membrana foi lavada em solução de bloqueio PBS/leite (PBS 1X com leite em pó desnatado 5%) por no mínimo 1 hora. As membranas foram lavadas por 15 min em tampão PBS/Tween (PBS 1X, Tween 0,05%) sob agitação leve, substituindo o tampão a cada 5 min. As membranas foram incubadas separadamente com solução contendo o anticorpo primário: PTP, ODV-E26, PIF-3, LEF-1 e GP64 (1:500) diluídos em PBS/leite (PBS 1X; leite em pó desnatado 0,5%). Após 1 hora, cada membrana foi novamente lavada por 15 min em tampão PBS/Tween (PBS 1X, Tween 0,05%) sob agitação leve, substituindo o tampão a cada 5 min. Este tampão foi trocado pela solução contendo o anticorpo secundário (anti-mouse Promega diluição 1:10.000) diluído em PBS/leite (PBS 1X; leite em pó desnatado 0,5%). Após 1 hora foi repetida a lavagem com tampão PBS/Tween (PBS 1X, Tween 0,05%). Cada membrana foi incubada com tampão da fosfatase alcalina (NaCl 100 mM, MgCl₂ 5 mM e Tris-HCl 100 mM – pH 9.0) acrescido de NBT/BCIP (substrato para fosfatase alcalina/SIGMA), até o aparecimento das bandas reativas. Ao final, a reação foi interrompida com sucessivas lavagens em água destilada.

3.2 Análise de diversidade genética da *gp64* (ORF-124)

3.2.1 Análise comparativa da ORF-124

A análise comparativa da ORF-124 (*gp64*) foi realizada a partir de dezessete variantes geográficas do AgMNPV coletados desde de 1980 pelo Prof. Dr. Flávio Moscardi da EMBRAPA soja (Empresa Brasileira de Pesquisa Agropecuária), que nos cedeu gentilmente. O AgMNPV-2D é o genótipo que corresponde a 40% da população selvagem e foi o protótipo estudado em nosso laboratório. No entanto, com a colaboração do Prof. Flávio Moscardi foi possível identificar a diversidade na ORF-124 nos diferentes isolados geográficos de AgMNPV (Tabela 1) por meio de dois métodos: (i) sequenciamento da região da *gp64* por sequenciamento de segunda geração e (ii) amplificação da *gp64*, clonagem e sequenciamento por Sanger. Um esquema da metodologia utilizada está representado abaixo (Figura 13).

Figura 13 - Esquema geral da metodologia descrita para a análise comparativa da ORF-124 (*gp64*) dos isolados geográficos do AgMNPV

Multiplicação viral in vivo, Purificação e Extração do DNA viral total (item 3.2.2)

Pirosequenciamento isolados geográficos AgMNPV (item 3.2.3)

Análise sequências *gp64* pelo *software* Geneious 6.1 (item 3.2.3)

PCR *gp64*, Bactérias Químio-competente, Clonagem (itens 3.2.4 a 3.2.5)

Sequenciamento por Sanger e análise sequências *gp64* pelo *software* Codon Code Aligner (item 3.2.6)

Tabela 1 - Isolados geográficos de *Anticarsia gemmatalis* múltiplo nucleopoliedrovírus

Isolado	Localidade/data de isolamento
Embrapa-Soja - VPN26 VPN <i>Anticarsia gemmatalis</i>	Dourados - Ms 07/02/84
Embrapa-Soja - VPN27 VPN <i>Anticarsia gemmatalis</i>	Argentina - 1982
Embrapa-Soja - VPN28 VPN <i>Anticarsia gemmatalis</i>	Uruguai - 02/84
Embrapa-Soja - VPN29 VPN <i>Anticarsia gemmatalis</i>	Ponta Grossa - Pr 1985
Embrapa-Soja - VPN30 VPN <i>Anticarsia gemmatalis</i>	Uruguai - 17/11/82
Embrapa-Soja - VPN31 VPN <i>Anticarsia gemmatalis</i>	Sertanópolis-Pr - 1986
Embrapa-Soja - VPN32 VPN <i>Anticarsia gemmatalis</i>	Dourados-Ms - 17/11/82
Embrapa-Soja - VPN33 VPN <i>Anticarsia gemmatalis</i>	Pelotas-Rs - 1984
Embrapa-Soja - VPN34 VPN <i>Anticarsia gemmatalis</i>	Distrito Warta, Londrina - Pr 1990
Embrapa-Soja - VPN35 VPN <i>Anticarsia gemmatalis</i>	Rio Grande do Sul - 1984
Embrapa-Soja - VPN36 VPN <i>Anticarsia gemmatalis</i>	Passo Fundo - Rs - 11/82
Embrapa-Soja - VPN37 VPN <i>Anticarsia gemmatalis</i>	Campo Mourão - Pr 11/82
Embrapa-Soja - VPN38 VPN <i>Anticarsia gemmatalis</i>	Uruguai - 17/11/82
Embrapa-Soja - VPN39 VPN <i>Anticarsia gemmatalis</i>	Sertanópolis-Pr - 1987
Embrapa-Soja - VPN40 VPN <i>Anticarsia gemmatalis</i>	Londrina –Pr - Sem data
Embrapa-Soja - VPN42 VPN <i>Anticarsia gemmatalis</i>	Isolado de laboratório - 1987
Embrapa-Soja - VPN43 VPN <i>Anticarsia gemmatalis</i>	Prop. Luis Pesce, Londrina. 1980

3.2.2 Multiplicação viral in vivo, purificação, extração do DNA viral total e pirosequenciamento de isolados geográficos de AgMNPV

Os isolados geográficos de AgMNPV foram amplificados por meio de infecções *per os* em larvas de *A. gemmatalis* (O'Reilly et al., 1993). As partículas virais contidas em corpos de oclusão foram purificadas por ultracentrifugação e a extração do DNA genômico foi realizada pelo método Fenol:Clorofórmio:Álcool isoamílico (25:24:1). O DNA de cada amostra foi pirosequenciado em plataforma *GS FLX Titanium®* (454 Life Sciences, Roche) em colaboração com o Prof. Manfred Weidmann, da University Medical Center, Gottingen, na Alemanha. Os conjuntos de leituras obtidos para cada uma das 17 amostras foram mapeados contra o genoma de AgMNPV-2D e seus os genomas foram montados como objetivo do mestrado do aluno Anderson F. Brito.

3.2.3 Análise das sequências de *gp64* (ORF-124) geradas por pirosequenciamento

As leituras obtidas pelo pirosequenciamento relacionadas a *gp64* (ORF124) dos 17 isolados geográficos foram segregadas para as análises da diversidade específica deste gene. Todas as leituras foram analisadas utilizando o *software* Geneious R6.1 (Biomatters, <http://www.geneious.com>). Os parâmetros foram escolhidos visando incluir o maior número de variantes biologicamente significantes, baseado em trabalhos prévios que utilizaram dados da plataforma Roche/454. No Geneious, a sequência da *gp64* do genoma do AgMNPV-2D foi utilizada como referência para o alinhamento e para comparação da variabilidade intrapopulacional deste gene. As sequências de baixa qualidade foram retiradas da análise e os parâmetros utilizados foram: qualidade de base *phred* igual a 20 cobertura mínima de 100 leituras por sítio, e penalidade para regiões homopoliméricas (redução de qualidade de 30% para cada base adicional). Seguimos para as análises de variabilidade ou polimorfismos (SNPs) da *gp64*, com os parâmetros de variação de cobertura mínima igual a 100 e variação mínima de frequência igual a 0,1%. Assim, apenas variantes que estivessem presentes em frequência > x% (aqui, x= 0,1%) foram considerados.

3.2.4 Síntese de oligonucleotídeos para a ORF-124 e Reações de PCR

Os oligonucleotídeos para a amplificação da *gp64* foram desenhados a partir da sequência genômica do vírus AgMNPV-2D (Oliveira et al., 2006b), utilizando o OLIGO 6.8 Primer Analysis Software (Molecular Biology Insights, Inc., Cascade, Co). Os oligos sintetizados (Apêndice C, tabela C2) foram utilizados para a avaliação da variabilidade do gene da *gp64* nos 17 isolados geográficos. Cada reação foi realizada no volume final de 50 µl, com: 1 µl de DNA (~ 2 nanogramas), 10 pmol de cada oligo, 0,4 mM de DNTP, 2,0 mM de MgCl₂, tampão 1X e 1,5 U de Taq polimerase *High Fidelity* (Invitrogen) (Apêndice C, tabela C1). Foi utilizado um ciclo inicial de 95 °C por 2 minutos seguido de 35 ciclos de 95 °C por 30 segundos, 55 °C ou 57 °C por 30 segundos e 68 °C por 50 segundos, com uma extensão final de 72 °C por 10 minutos. A amostra foi verificada em gel de agarose 1%, corado com brometo de etídeo.

3.2.5 Preparo de bactérias Químio-competente e Clonagem do produto de PCR

Quinhentos mililitros de meio líquido TB (triptona: 11,8g, extrato de levedura: 23,6g, K_2HPO_4 : 9,4g, KH_2PO_4 : 2,2g, glicerol: 4ml) foram inoculados com 5 ml de uma cultura de *E. coli* DH5 α crescida durante 12 horas. A cultura foi incubada sob agitação constante a 37 °C até atingir a D.O.₆₀₀ 0,7 (fim da fase log). As células foram lavadas com 200 ml (1 vez) de H₂O mili-Q estéril gelada e 200 ml (2 vezes) com glicerol 10% para garantir a eliminação de sais. As alíquotas (50 μ l) foram mantidas a temperatura de -80 °C. Antes de sua utilização foi testada a eficiência da preparação com o vetor PUC. Para realizar a clonagem do gene da *gp64* amplificados (seção 3.2.4) foi utilizado o kit TOPO TA (Invitrogen), seguindo as instruções do fabricante. Após a ligação, 5 μ l da reação foi transformado em bactérias DH5 α químio-competentes por choque térmico (42 °C, 30s). Em seguida, foi adicionado 250 μ l de meio SOC e incubou-se a 37 °C por 1h. Após este período, 50 μ l de cada transformação foram plaqueados em meio LB – Km sólido (50 μ g/ μ l) e as placas foram incubadas a 37 °C por 16h. As colônias de cada placa foram inoculadas em meio LB-Km líquido (50 μ g/ μ l), e incubadas a 37 °C por 16h. Para a extração dos plasmídeos foi utilizado o kit *GeneJet Plasmid Miniprep* (Thermo Scientific), seguindo as instruções do fabricante.

3.2.6 Sequenciamento por Sanger e análise dos clones

O sequenciamento dos clones foi realizado em sequenciador automático ABI-PRISM 3100 (Applied Biosystems). Este método garante que o fragmento clonado, posteriormente sequenciado é uma fita única de DNA, permitindo o mapeamento das mutações ligadas necessário para a construção da filogenia. Para isso, foi utilizando oligonucleotídeos específicos para o vetor de clonagem (M13 *Foward* e M13 Reverso) (Apêndice C, tabela C3) e aproximadamente 150 ng de plasmídeo em cada reação. As reações foram realizadas em volume final de 10 μ l nas seguintes condições: 1,0 μ l de Big Dye Terminator v3.1 Applied Biosystems, CA, USA), 3,0 μ l de tampão de sequenciamento (5X), 1,0 μ l de oligo (3.2 ou 30 pmol), DNA e H₂O MilliQ para completar o volume final para 10 μ l de reação. As amostras foram submetidas a ciclagem em termociclador de 96°C por 2min, seguido de 35 ciclos de 96 °C por 30s, 52 °C por 30s e 60 °C por 4min. Para a remoção de dideoxinucleotídeos não incorporados, foi realizada a precipitação com 40 μ l de isopropanol 65% aos 10 μ l da reação.

Após homogeneização por vórtex e breve centrifugação, a reação foi mantida por 15 min em temperatura ambiente abrigada da luz. Após este período a placa foi centrifugada por 30 min a 3.300 g. O sobrenadante foi retirado, acrescentou-se 150 µl de etanol 60% seguido de centrifugação a 3.300 g por 10 min. Após a retirada do sobrenadante o *pellet* foi seco a 96 °C por 3 min e ressuspendido em 10 µl de Hi-Di Formamida (Applied Biosystems). O DNA foi desnaturado por 3 min a 96 °C, seguido de 5 min no gelo e em seguida submetido eletroforese capilar em sequenciador automático 3100. As sequências foram obtidas através do programa Sequencing Analysis v3.7 (Applied Biosystems) e os *contigs* para os diferentes clones foram montados utilizando o programa Codon-Code-Aligner (<http://www.codoncode.com/aligner/>). A análise de polimorfismos também foi realizada contra a *gp64* do genoma do AgMNPV-2D utilizando o *software* Geneious R6.1 (Biomatters, <http://www.geneious.com>). O alinhamento dos clones foi realizado no *software* Se-Al v2.0a11.

3.2.7 Análise filogenética da *gp64*

Árvores filogenéticas foram geradas pelo programa GARLI v2.0 (Zwickl, 2006), que emprega um algoritmo genético estocástico para estimar simultaneamente a melhor topologia, comprimentos de ramos e parâmetros do modelo de substituição empregado, que maximizam o valor logarítmico da Verossimilhança (lnL). Dez replicações independentes foram executadas sob o modelo de substituição *general time reversible* (GTR) com taxa de variação gama e uma proporção de sítios invariantes (GTR+Γ+I). A replicação com lnL mais alto foi selecionada. Para avaliar o suporte estatístico dos ramos das árvores, foram empregadas mil amostragens não paramétricas de *bootstrap*. Para inferir as pressões de seleção agindo sobre a *gp64* (aminoácidos sob seleção), foi utilizado o programa HyPhy v0.99 (Pond, Muse, 2005), que estimou a diferença entre as taxas de substituição não sinônimas (*dN*) e sinônimas (*dS*) por sítio codante, com algoritmo de contagem ancestral de única verossimilhança (SLAC) assumindo um nível de confiança de 95% ($\alpha = 0,05$). Valores da diferença ω (onde ω representa *dN-dS*) maiores que zero são indicativos de seleção positiva ($\omega > 0$), enquanto valores menores que zero indicam seleção negativa ou purificadora ($\omega < 0$).

4 RESULTADOS

4.1 Análise do proteoma estrutural do AgMNPV-2D em duas linhagens de células de insetos por espectrometria de massas

4.1.1 Produção e purificação viral

Quando infectadas, as duas linhagens celulares apresentaram efeitos citopáticos visualizados pela presença de poliedros no núcleo celular, perda do formato celular original e algumas células “mortas” no meio de cultivo. O estoque viral escolhido apresentou $1 \times 10^{7,15}$ TCID₅₀/ml. Ambas as linhagens celulares foram infectadas partindo deste mesmo inóculo. Diferentes MOIs foram testadas (0,1 MOI, 0,5 MOI, 1 MOI, 3MOI e 5 MOI). Para a linhagem UFL-AG-286, foi estabelecido que as infecções de 1 MOI produziram partículas virais suficiente para os experimentos. Excepcionalmente, a linhagem IPLB-SF-9 requereu 5 MOI para se atingir a mesma concentração viral obtida em UFL-AG-286. Como descrito na seção de métodos, o fenótipo ODV foi coletado 72 h.p.i e o fenótipo BV foi coletado 48 h.p.i. nas duas linhagens. Um dos maiores desafios deste estudo foi garantir a obtenção de partículas virais puras sem contaminantes celulares ou proteína de soro bovina. Por este motivo, os vírus isolados de cada infecção de 1 MOI e 5 MOI dos cultivos de UFL-AG-286 e IPLB-SF-9, respectivamente, foram previamente centrifugados, concentrados em seus diferentes fenótipos e ultracentrifugados separadamente em gradiente de sacarose formando uma banda bem definida. No fenótipo BV, a banda se formou entre 47 - 49% do gradiente (Figura 14), enquanto que no fenótipo ODV a banda se formou entre 54 - 56% (Figura 14). A fim de se garantir a pureza do material, os fenótipos obtidos foram submetidos a uma segunda ultracentrifugação em gradiente de sacarose. Apesar de possíveis perdas durante este processo, a purificação das amostras assegurou que os dados obtidos durante as análises por espectrometria de massas eram provenientes das partículas virais dos dois fenótipos, e não de contaminantes do meio de cultura.

Figura 14 - Foto ilustrativa dos dois fenótipos após a ultracentrifugação em gradiente de sacarose.

Resultado da ultracentrifugação em gradiente de sacarose, a esquerda foi representado o fenótipo ODV com a banda branca formada entre 54 - 56%, enquanto que no fenótipo BV (direita) a banda se formou entre 47 - 49% do gradiente.

4.1.2 Quantificação Proteica

As partículas virais BV e ODV isoladas de UFL-AG-286 e IPLB-SF-9 foram quantificadas pelo método de Bradford. A concentração média do fenótipo ODV produzidos em 72 h.p.i em UFL-AG-286 foi de 150 $\mu\text{g}/\mu\text{l}$. Durante as quantificações, estas amostras foram diluídas na razão 1:2, porque algumas apresentaram concentração acima da estabelecida na curva padrão. Todavia, para o mesmo fenótipo produzido em IPLB-SF-9 foram realizadas inúmeras infecções sem sucesso e não obtivemos a concentração desejada. Assim, nesta linhagem, o fenótipo ODV não foi considerado para análise. A concentração média do fenótipo BVs produzidos em 48 h.p.i na linhagem infectada com 1 MOI UFL-AG-286 foi de 50 $\mu\text{g}/\mu\text{l}$ e para o BV produzido em IPLB-SF-9 infectada com 5 MOI foi de 30 $\mu\text{g}/\mu\text{l}$. Todas as amostras quantificadas foram congeladas e posteriormente utilizadas para eletroforese SDS-PAGE. Como descrito na seção de métodos, na quantificação das amostras submetidas para a eletroforese bidimensional (2DE) foi utilizado o 2D Quant-Kit (Amersham Biosciences). Todos os isolados virais foram precipitados em acetona gelada e ressuspendidos em 70 μl de solução de reidratação para as eletroforeses 2DE. Apenas uma alíquota de 25 μl foi quantificada em duplicata pelo 2D Quant Kit. Damos início à quantificação do fenótipo ODV produzido em UFL-AG-286 e a quantidade média do fenótipo ODV utilizado nas triplicatas dos géis 2DE foi de 500 μg . A quantidade média das partículas de BVs produzidas na linhagem UFL-AG-286 utilizando o 2D Quant Kit (Amersham Biosciences) foi 360 μg . Somente os vírus isolados da linhagem UFL-AG-286 foram

quantificadas pelo o 2D Quant Kit (Amersham Biosciences) congeladas e posteriormente utilizadas para a eletroforeses 2DE.

4.1.3 Identificação das proteínas do AgMNPV-2D por eletroforese bidimensional (2DE) e espectrometria de massas MALDI-Q-TOF

Com os fenótipos purificados e quantificados, a próxima etapa foi a determinação teórica da faixa de pH e massa molecular das proteínas. Para isso, os dados genômicos do AgMNPV (Oliveira et al., 2006b) foram relacionados aos dados do UniProt (<http://www.uniprot.org/>). Assim, pudemos ter uma ideia inicial dos pontos isoelétricos e optar pela fita IPG com pH pré-fixado que melhor se enquadraria para a focalização isoelétrica das proteínas estruturais dos dois fenótipos estudados. A Tabela 2 lista as ORFs do AgMNPV que codificam proteínas estruturais, com as respectivas massas moleculares e pontos isoelétricos (*pI*). Com estes dados, optamos por utilizar as fitas IPG com pH linear de 3-10 de 13 cm, uma vez que a grande maioria das proteínas estavam dentro desta faixa de pH. Porém, para melhorar a separação das proteínas com *pI* em pH fisiológico, também optamos por usar fitas IPG com pH linear de 4-7 de 13 cm. Para a segunda dimensão em gel de poliacrilamida, a concentração utilizada foi de 12%, uma vez que as massas moleculares teóricas variaram entre 7 e 141 kDa (Tabela 2). Outro fator importante para as eletroforeses 2DE foi a retirada de sais e outros compostos que frequentemente interferem na focalização. Por esta razão, as proteínas foram precipitadas com acetona gelada e ressuspensas diretamente em tampão de hidratação.

Tabela 2 - ORFs que codificam as 38 proteínas estruturais do AgMNPV-2D (Oliveira et al., 2006a) e predição das massas das proteína (kDa) e de seus pontos isoelétricos (*pI*).

ORF	Gene	Proteína (kDa)	<i>pI</i>
1	<i>polh</i>	29	6,1
12	<i>pif-2</i>	43,5	5,1
17	<i>odv-e26</i>	24	10,6
19	<i>lef-1</i>	28	7,8
22	<i>proteína F</i>	74	7,9
49	<i>odv-e66</i>	76	5,5
55	<i>vp1054</i>	43	8,1
61	<i>25k,fp</i>	27	9,7
65	<i>dna pol</i>	114	8
66	<i>desmoplaquina</i>	101	4,9
67	<i>lef-3</i>	43	5,2
77	<i>vlf-1</i>	43,5	9,8
79	<i>gp41</i>	42,1	7,1
82	<i>vp91/p95</i>	94	4,4
85	<i>cg30</i>	28	8,7
86	<i>vp39</i>	38	6,5
89	<i>p33</i>	30	8
91	<i>odv-e25</i>	25	7,1
92	<i>Helicase</i>	141	7
96	<i>p6.9</i>	7	12,8
97	<i>p40</i>	40	7,1
98	<i>p12</i>	12	9,5
100	<i>p87</i>	78	5,5
107	<i>odv-ec43</i>	45	7,8
114	<i>pif-3</i>	22	4,1
117	<i>pif</i>	58	5,5
124	<i>gp64</i>	58	5,8
125	<i>p24</i>	21	5,5
128	<i>p25</i>	25	9,9
129	<i>alk-exo</i>	47	9
133	<i>p10</i>	10	4
134	<i>p74</i>	74	4,9
138	<i>p49</i>	56	8,8
139	<i>odv-e18</i>	9,2	9,6
140	<i>odv-e27</i>	35	7,9
143	<i>ie-1</i>	68	6,2
144	<i>odv e-56</i>	40	5,6
152	<i>p78/83</i>	70	5,9

Uma vez determinada a faixa de pH e a concentração de poliacrilamida, iniciamos as eletroforese 2DE com o fenótipo ODV quantificado e isolado de UFL-AG-286. Alguns dos isolados foram testados e visando análise das proteínas no espectrômetro de massas MALDI-Q-TOF ficou determinada a concentração de 135µg/225µl para este fenótipo. A figura 15 representa dois géis 2DE do fenótipo ODV realizado em triplicatas experimentais (Apêndice B, Figura B1) analisadas pelo software Platinum 7.0 (Ge Healthcare). O primeiro gel exemplifica as proteínas estruturais do ODV nas fitas IPG com pH variando de 3 - 10 de 13 cm com oitenta e cinco *spots* detectados nos três géis 2DE analisados. Na figura 15 foi

possível observar que não houve problema de condutividade e arrastes horizontais, com exceção de algumas proteínas que focalizaram na borda das fitas em pH 10, além da p 6.9 e BV/ODV-E26, que apresentam *pI* maior que 10 (tabela 2). Além disso, a mesma amostra de ODV foi focalizada em fitas IPG com pH linear de 4 - 7 de 13 cm (Figura 15 abaixo). No segundo gel foi possível observar somente 16 *spots* presentes nos três géis analisados. No total, cento e um *spots* foram identificados em ordem crescente de intensidade, retirados manualmente das triplicatas, reduzidos, alquilados, digeridos com tripsina e congelados a -20°C. No Laboratório de Espectrometria de massas do LNBio os peptídeos obtidos da digestão foram ressuspensos em ácido de Trifluoroacético (TFA) 0,1% e exportados para placa de MALDI, seguida de espectrometria de massa automatizada no MALDI-Q-TOF.

Figura 15 - Eletroforese bidimensional em gel de poliacrilamida 12% demonstrando as proteínas estruturais do fenótipo ODV purificado de células UFL-AG-286.

A concentração aplicada por fita foi 135µg/225µl. As amostras foram hidratadas e focalização isoeletrica da fita de 13 cm ocorreu em 5h e 15min e ao final foram acumulados 17.500 Vhr. As fitas foram equilibradas e em seguida, posicionadas sobre o gel de poliacrilamida 12%. Foram aplicados 10 µl do marcador de peso molecular (Pm) para proteína de 11-240 kDa (Fermentas) em condições de corrida de 110V durante 6h. As imagens foram capturadas no ImagerScanner (GE Healthcare).

Os espectros de massas obtidos nesta corrida foram analisados no servidor Mascot (Mascot Distiller v.2.3.2.0, 2009 (Matrix Science Ltd.)) e as proteínas foram identificadas nos dois bancos de dados citados em material e métodos. Os primeiros dados acerca da composição proteica do fenótipo ODV do AgMNPV-2D estão listados no Apêndice A, tabela S1. Dentre os 85 *spots* na eletroforese 2DE com pH linear de 3-10, somente 21 proteínas puderam ser identificadas a partir destes *spots* (Figura 16). Como pode ser observado, duas proteínas foram coradas intensamente no centro do gel: (1) Poliedrina (*ag1*) e (10) CG30 (*ag85*) (Apêndice A, Tabela S1). Outras 19 proteínas também foram identificadas na estrutura do ODV: 41.5 kDa (*ag15*), EGT (*ag18*), 38 kDa (*ag20*), IAP-1 (*ag40*), FP25K (*ag61*), DNAPol (*ag65*), LEF-3 (*ag 67*), GP41 (*ag79*), VP39 (*ag86*), ODV-E25 (*ag91*), P40 (*ag97*), PNK/PNL (*ag103*), PTP (*ag107*), P24 (*ag125*), Alk-Exo (*ag129*), P10 (*ag133*), P49 (*ag138*), ODV-E27 (*ag140*) e PK-1 (*ag151*) (Apêndice A, tabela S1). As proteínas mais abundantes deste fenótipo foram identificadas mais de uma vez em diferentes *spots*: VP39, ODV-E25, P49 e ODV-E27. Além disso, três proteínas foram encontradas exclusivamente utilizando esta abordagem, 2DE e MALDI-Q-TOF: DNAPol (*ag65*), CG30 (*ag85*) e Alk-Exo (*ag129*). A proteína DNAPol é essencial para a replicação do DNA viral; CG30 sem função conhecida para esta proteína e a Alk-Exo é homóloga a uma nuclease alcalina. Infelizmente, nesta abordagem mais da metade dos espectros de massas mostraram baixa cobertura ou baixa pontuação contra o banco teórico da AgMNPV-2D (Apêndice B, Figura B2 e tabela B1). Nesta etapa, não foi possível identificar algumas proteínas conservadas em todos os baculovírus: GP41, ODV-E56, P6.9, P74, VP91/P95 e VP1054.

Figura 16 - Eletroforese bidimensional em gel de poliacrilamida 12% com as proteínas estruturais identificadas por MALDI-Q-TOF no fenótipo ODV purificado de células UFL-AG-286.

Os números 1-21 indicam as proteínas identificadas por 2 DE e MALDI-Q-TOF. Foram aplicados 10 µl do marcador de peso molecular (M) para proteína de 11-250 kDa (Fermentas) em condições de corrida de 110V durante 6h.

Com o fenótipo BV cultivado na linhagem UFL-AG-286, obtivemos as seguintes concentrações com o 2D Quant Kit: 75 µg/ 225 µl e 120 µg/ 225 µl. Seguimos com a hidratação das fitas com pH linear 3-10 e focalização isoeletrica (Ethan IPG phor II System-GE Healthacare), para as eletroforeses 2DE. Porém, tanto para as triplicatas de 2DE com 75 µg/225µl quanto para as triplicatas de 2DE 120 µg/225 µl, não observávamos um perfil proteico claro para o fenótipo BV na linhagem UFL-AG-286. Este fato nos fez reavaliar a eletroforese 2DE. As condições de tampões usadas no decorrer dos experimentos de 2DE não possibilitam condições ideais de solubilização de todas proteínas presentes nas partículas virais, principalmente as proteínas de membrana, que eram perdidas durante a fase de preparação das amostras. Portanto, para minimizar as limitações da eletroforese 2DE e da sensibilidade de um espectrômetro de massas com a fonte de ionização MALDI, seguimos com outra abordagem para identificação das proteínas do AgMNPV-2D.

4.1.4 Identificação das proteínas do AgMNPV-2D por eletroforese SDS-PAGE e espectrometria de massas LC-MS/MS

Nesta etapa, utilizamos os fenótipos virais, ODV e BV isolados de UFL-AG-286 e o fenótipo BV isolados de IPLB-SF-9, previamente purificados por gradiente de sacarose e quantificados pelo método de Bradford. As partículas virais tiveram suas proteínas separadas em gel de poliacrilamida 12% (SDS-PAGE). Em seguida, todas as bandas foram digeridas com tripsina e submetidas a cromatografia líquida de alta performance (HPLC) acoplada a um espectrômetro de massas do tipo ESI-Q-TOF. Inicialmente, utilizamos as triplicatas obtidas do fenótipo ODV de UFL-AG-286 realizadas nas eletroforeses 2DE. Na figura 17, pode-se observar o perfil das proteínas estruturais em gel SDS-PAGE 12% e suas respectivas massas moleculares, bem como a alta expressão da poliedrina, proteína com 28 kDa. Além disso, os géis SDS-PAGE indicaram a presença de trinta e uma bandas no fenótipo ODV. Cada banda identificada foi numerada à direita do gel (O1 a O31). Todas as bandas foram retiradas manualmente com a ajuda de um estile e as proteínas presentes nas mesmas foram reduzidas, alquiladas e digeridas com tripsina. Os peptídeos obtidos da digestão foram congelados a -20 °C.

Figura 17 - SDS-PAGE 12% com o perfil de proteínas do fenótipo ODV do AgMNPV-2D

As triplicatas experimentais de ODV foram aplicadas (1, 2 e 3) com aproximadamente 150 µg de proteínas por poço. À direita estão indicadas as bandas (com os números O1-O31) que foram retiradas do gel, digeridas com tripsina e analisadas por LC MS/MS (Apêndice A, tabela S1). Foram aplicados 10 µl do marcador de peso molecular (M) para proteína de 10-250 kDa (Fermentas), em condições de corrida de 110 V durante 6 h. As imagens foram capturadas no Imagescanner (GE Healthcare).

No LNBio, os peptídeos obtidos da digestão foram ressuspensos em 12 µl de ácido de fórmico 0,1%, centrifugados e transferidos para um novo tubo. Trinta e um tubos contendo os peptídeos do fenótipo ODV foram injetados separadamente no HPLC. Assim, os diferentes peptídeos presentes em cada tubo puderam ser eluídos em ordem de hidrofobicidade, para entrada no espectrômetro de massas. Na injeção final do HPLC, o líquido foi vaporizado e os peptídeos ionizados com a fonte ESI (*Electrospray*). Para determinação dos espectros de massas continuamos utilizando os analisadores de massa: quadrupólos e TOF em *tandem*. Os MS/MS obtidos nesta corrida também foram comparados com o servidor Mascot (Mascot Distiller v.2.3.2.0, 2009 (Matrix Science Ltd.) contra os dois bancos de dados (o genoma do AgMNPV-2D e o banco de proteínas Lepidoptera montado a partir do NCBI), utilizando os parâmetros citados em material e métodos. Trinta e nove proteínas foram identificadas e os dados acerca da composição proteica do fenótipo ODV do AgMNPV-2D estão resumidos no Apêndice A, tabela S1. Quase metade das proteínas

relacionadas ao ODV foram confirmadas pelas duas abordagens proteômicas utilizadas. Além disso, 21 proteínas foram apenas identificadas quando utilizamos o LC-MS/MS: PTP 1 (*ag9*), PIF-2 (*ag12*), ODV-E26 (*ag17*), PP31 (*ag25*), ODV-E66 (*ag49*), VP1054 (*ag55*), Desmoplaquina (*ag66*), VLF-1 (*ag76*), VP91/P95 (*ag82*), p33 (*ag89*), Helicase (*ag92*), p6.9 (*ag96*), P48 (*ag99*), VP80/P87 (*ag100*), PIF-1 (*ag117*), PP34/PEP (*ag127*), P74 (*ag134*), ODV-E18 (*ag139*), IE-1 (*ag143*), ODV-E56 (*ag144*) e P78/83 (*ag152*) (Apêndice A, Tabela S1). No total, 42 proteínas foram identificadas na estrutura do fenótipo ODV do AgMNPV-2D, sendo que 13 delas foram associadas ao nucleossomo viral, 15 relacionadas ao nucleocapsídeo e 10 ao envelope viral (Figura 19). Duas proteínas, a poliedrina e a P10 foram observadas como principais componentes do poliedro ou corpo oclusão. Entretanto, duas proteínas hipotéticas, AG15 e a AG20 foram identificadas pelas duas abordagens pela primeira vez no fenótipo ODV. Estas são proteínas pouco caracterizadas em baculovírus, com as análises não encontramos domínios conhecidos depositados em bancos de dados e podem ser alvos interessantes para estudos futuros. Além disso, nossos dados revelaram cinco proteínas importantes envolvidas na regulação de ciclo celular do hospedeiro, como: PTP, EGT, PP31 IAP-1 e PK1 (Apêndice A, Figura 1A e tabela S2). Os dados obtidos pelo servidor Mascot em relação ao banco de proteínas do hospedeiros serão apresentados na próxima seção.

Demos continuidade às eletroforeses SDS-PAGE com as duplicatas obtidas do fenótipo BV isolado de UFL-AG-286 e do fenótipo BV isolado de IPLB-SF-9, obtidos de uma infecção de 5 MOI. Conforme descrito em métodos, todas as amostras quantificadas por Bradford, precipitadas em acetona e ressuspensas em tampão redutor para as eletroforeses SDS-PAGE 12%. O fenótipo BV isolado de UFL-AG-286 e de IPLB-SF-9 foram aplicados em gel SDS-PAGE 12% (Figura 18). Como pode ser observado, as massas moleculares das proteínas variaram entre 7 e 140 kDa, com a presença majoritária de duas proteínas de 60 kDa e 40 kDa, provavelmente a GP64 (glicoproteína responsável pela entrada do vírus nas células) e a VP39 (principal proteína do nucleocapsídeo). Além delas, a menor proteína detectada, com aproximadamente 7 kDa, deve corresponder à proteína básica de 6.9kDa, que possui função de ligar-se ao DNA viral, análoga a histonas. Além disso, aproximadamente 30 bandas foram visualizadas no fenótipo BV, mas a intensidade das bandas variou quando comparamos as duas linhagens (Figura 18). O fenótipo BV foi coletado 48 horas pós infecção nas duas linhagens. Porém, para este tempo de infecção, o isolado da linhagem IPLB-SF-9 não produziu a mesma quantidade de partícula viral, *i.e*, a mesma concentração proteica

obtida na linhagem UFL-AG-286 (linhagem do hospedeiro natural) mesmo quando partimos de uma multiplicidade de infecção 5 vezes maior (MOI = 5). Os números a direita do gel SDS-PAGE (Figura 18) representam as bandas que foram cortadas e digeridas com Tripsina e congeladas a -20 °C. No total, 36 bandas foram enviadas e submetidas para análise em LC-MS/MS com fonte de ionização ESI (*Electrospray*) no centro de Proteômica (*Interdisciplinary Center for Biotechnology*) da Universidade da Florida (UF-USA) em colaboração com o Professor James E. Maruniak e a Dra. Alejandra Garcia-Maruniak.

Figura 18 - Eletroforese SDS PAGE 12% demonstrando as proteínas estruturais do fenótipo BV isolado e purificado de células UFL-AG-286 e IPLB- SF-9.

Os números à direita do gel na linhagem UFL-AG-286 (B1- B19) estão identificados no Apêndice A, tabela S2. Os números à direita do gel da linhagem IPLB-SF-9 infectada com 5 MOI (S1-S17) estão identificados na tabela 3. O marcador molecular (M) utilizado foi de 19-118 kDa (Fermentas) em condições de corrida de 120 V durante 6 h. O gel foi corado por 1 hora com *Comassie Blue* R-250 e descorado em solução (10% Ácido acético/40% Etanol) por 3 horas.

Inicialmente, quatro dos peptídeos digeridos foram testados por ambos os tempos: 1 hora e 30 minutos, visando a melhor separação na coluna cromatográfica e identificação no

espectrômetro de massa híbrido-Q-TOF (QSTAR elite, Applied Biosystems, Framingham, MA). Após os testes, resolvemos utilizar 1 hora de análise para identificação das proteínas do fenótipo BV. Os dados obtidos por MS/MS de cada amostra foram analisados utilizando o programa Mascot (Matrix Science, London, UK; *version* 2.2.2) contra o genoma do AgMNPV-2D e o banco de insetos montado a partir do NCBI (<http://www.ncbi.nlm.nih.gov>). Em seguida, os dados obtidos pelo Mascot foram validados no programa Scaffold (versão Scaffold-03-00-04, Proteome Software Inc., Portland, OR). Dado que para o fenótipo BV utilizamos apenas uma abordagem proteômica, só foram aceitas as proteínas que no programa Scaffold apresentaram confiabilidade de 99% (Nesvizhskii et al., 2003) com no mínimo dois fragmentos de peptídeos únicos ou dois espectros únicos (Keller et al., 2002). As proteínas estruturais identificadas do fenótipo BV na linhagem UFL-AG 286 estão apresentadas na figura 19 (Apêndice A, tabela S2). Os dados são muito consistentes quando comparados com os dois proteomas deste mesmo fenótipo do AcMNPV (protótipo da família) isolado em célula IPLB-SF-9 (Wang et al., 2010) e HearNPV (Hou et al., 2013).

Como esperávamos, algumas proteínas foram identificadas em mais de uma banda: proteína F, PP31, V-Ubiquitina, Desmoplaquina, VLF-1, VP39, AG112, AG113, GP64 e ME53. Acreditava-se que algumas destas proteínas pertencessem apenas ao fenótipo ODV mas também foram comuns ao BV as seguintes proteínas: PP31, Desmoplaquina, IAP-2, AG73, AG113, PEP/PP34, ODV-E18 e ODV-E56 (Apêndice A, tabela S2). Curiosamente, a proteína V-Ubiquitina de 9 kDa foi identificada somente nas bandas 7, 8 e 14. Muito provavelmente, esta proteína de membrana se complexa com proteínas de maior massa molecular. De maneira similar ao encontrado para o fenótipo ODV, foram detectadas proteínas hipotéticas sem função conhecida descrita para o fenótipo BV: AG15, AG52, AG73, AG112, AG113 e AG130 (Apêndice A, Figura 1B e tabela S2). Interessantemente, nossos dados revelaram seis novas proteínas no BV do AgMNPV-2D: AG15, LEF-3, P33, Helicase, PNK/PNL e AG130.

No total, identificamos 33 proteínas no fenótipo BV na linhagem UFL-AG-286, das quais sete foram associadas ao nucleossomo, 13 ao nucleocapsídeo e sete ao envelope do fenótipo BV (Figura 19) (Apêndice A, Figura 1B e tabela S2). Porém, não foi possível encontrar a P6.9 através desta metodologia e o mesmo foi observado no proteoma do AcMNPV (Wang et al., 2010). Em teoria, esta proteína básica de 52 aa com 22 resíduos de arginina (R), após a digestão seria clivada em um peptídeo de 11 aa e nove peptídeos de 2 a 5 aa. Porém, utilizando o LC MS/MS não foi possível identificar o peptídeo maior e os

peptídeos pequenos não são considerados íons precursores para este espectrômetro. Além disso, todos os espectros de massas gerados no QSTAR elite (Applied Biosystems, Framingham, MA) foram analisados contra um banco de dados contaminante fornecido pela Applied Biosystems (QSTAR elite, Applied Biosystems, Framingham, MA) e não foi encontrada a proteína albumina de soro bovino, utilizada para suplementar o crescimento celular durante o cultivo da linhagem UFL-AG-286.

Figura 19 - Esquema das proteínas identificadas por MS/MS no AgMNPV em linhagem UFL-AG-286.

Proteínas identificadas em BV e ODV de AgMNPV-2D a partir de uma infecção em UFL-AG-286. As linhas cheias indicam proteínas compartilhadas pelos dois fenótipos e as linhas tracejadas indicam proteínas encontradas apenas em um fenótipo. **(A)** Esquema de um poliedro (ODV) de dentro para fora: 10 proteínas do envelope (4 presentes no BV), 13 proteínas do nucleossomo (7 compartilhadas no BV) e 15 proteínas do nucleocapsídeo (10 compartilhadas no BV). EGT, AG15 e AG20 foram identificadas no ODV porém sem localização conhecida. **(B)** A partícula de BV representada com as proteínas identificadas: 3 proteínas únicas de envelope, 7 proteínas do nucleossomo compartilhadas e 3 proteínas exclusivas do nucleocapsídeo. AG15, AG52, AG73, AG112, AG113 e AG130 foram identificadas no BV porém sem localização conhecida.

Ao analisarmos os peptídeos digeridos do fenótipo BV isolado da linhagem IPLB-SF-9, confirmamos com significância 25 proteínas estruturais contra o genoma do AgMNPV-2D (Tabela 3) no programa Mascot (Matrix Science, London, UK; version 2.2.2). Resultado corroborado pelo programa Scaffold (version Scaffold-03-00-04, Proteome Software Inc., Portland, OR). Para confirmar as proteínas encontradas em IPLB-SF-9, a duplicata das proteínas do fenótipo BV nesta linhagem foi novamente analisada no centro de Proteômica da Universidade da Flórida. Entre as proteínas identificadas, estão a glicoproteína GP64, proteína F, PP31, V-Ubiquitina, Desmoplaquina, VLF-1, VP39, AG112, GP64 e ME53 (Tabela 3). Entretanto, algumas proteínas analisadas em conjunto pelo Mascot e pelo Scaffold mostraram baixos valores de similaridade contra o genoma do AgMNPV-2D ou apenas um único espectro de massa, e por este motivo foram desconsideradas. Na linhagem PLB-SF-9, não foram encontradas as proteínas AG52, IAP-2, AG73, P33, 38K, P40, P48 e AG113, que haviam sido identificadas no BV na linhagem UFL-AG-286.

Tabela 3 - Proteínas identificadas no fenótipo BV AgMNPV isolado da linhagem IPLB-SF-9 por LC MS/MS

AgMNPV ORF	AcMNPV ORF	Banda	Proteína	kDa	Função	Cov.	Espec.	NCBI
15	18	S10	AG15	41	Não essencial (Wang et al., 2007)	4	3	gi 116326084
22	23	S4	proteína F	75	Não essencial em BV NPV I (Westenberg et al., 2007)	15	10	gi 116326092
25	36	S12	PP31	30	Fosfoproteína envolvida em regulação gênica (Guarino et al., 2002)	20	4	gi 116326094
26	35	S2	V Ubiquitina	9	Ubiquitina viral (Guarino, 1990; Reilly, Guarino, 1996)	12	2	gi 116326095
55	54	S9	VP1054	43	Formação do nucleocapsídeo (Olszewski, Miller, 1997a)	9	2	gi 116326124
61	61	S16	FP 25 K	27	Proteína de BV/ODV (Harrison et al., 1996)	40	8	gi 116326221
66	66	S2	Desmoplaq.	101	Processo enovelar o DNA (Ke et al., 2008)	17	13	gi 116326134
67	67	S9	LEF-3	43	Proteína de ligação a ssDNA (Hang et al., 1995)	20	6	gi 116326135
76	77	S10	VLF-1	44	Proteína de nucleocapsídeoBV/ODV (Li et al., 2005)	16	5	gi 116326144
86	89	S8/S10/S11	VP39	38	BV/ODV nucleocapsídeo e organizadora de actina (Charlton, Volkman, 1991)	26	5	gi 116326154
91	94	S13	ODV-E25	26	Proteína de envelope (Slack, Arif, 2007)	32	7	gi 116326159
92	95	S1	Helicase	141	Replicação de DNA (Kool et al., 1994)	10	3	gi 116326160
100	104	S5	VP80/P87	78	Capsídeo (Lu, Carstens, 1991)e produção de BV (Tang et al., 2008)	7	2	gi 116326168
103	86	S3/S5	PNK/PNL	78	Proteína com 2 domínios: quinase e T4 RNA ligase (Durantel et al., 1998)	3	2	gi 116326171
107	109	S15	ODV-EC43	20	BV/ ODV proteína de envelope (Fang et al., 2003)	8	2	gi 116326175
112	-	S10	AG 112	41	Sem função conhecida	20	7	gi 116326180
124	128	S2/S7	GP64	57	Essencial para infecção secundária em BV (Oomens, Blissard, 1999)	6	2	gi 116326192
125	129	S16	P24	21	Associada ao nucleocapsídeo (Wolgamot et al., 1993)	20	8	gi 116326193
127	131	S11	PP34/PEP	33	Fosfoproteína (Whitt, 1988)	27	7	gi 116326195
130	-	S14	AG 130	24	Sem função conhecida	18	4	gi 116326198
136	139	S7	ME 53	52	Envolvida em replicação de DNA e produção de BV (Xi et al., 2007)	18	6	gi 116326204
138	142	S6	P49	56	BV/ODV nucleocapsídeo (McCarthy et al., 2008)	9	4	gi 116326213
139	143	S17	ODV-E18	9	BV/ ODV proteína de envelope (Braunagel et al., 1996a)	40	4	gi 116326207
140	144	S11	ODV-E27	34	BV/ ODV proteína de envelope e capsídeo (Braunagel et al., 1996a)	23	7	gi 116326208
144	148	S10	ODV-E56	40	BV/ ODV associada ao envelope (Braunagel et al., 1996b)	9	2	gi 116326212

* As linhas cinzas mostram as ORFs conservadas em todos baculovírus sequenciados até o momento.

* Cov. significa cobertura em porcentagem da proteínas utilizando o *software Scaffold*.* Os números do *GenBank* encontrados no genoma do AgMNPV-2D.

4.1.5 Proteínas do Hospedeiro identificadas no AgMNPV-2D

Para avaliar a presença de proteínas celulares nos fenótipos do AgMNPV-2D, todos os espectros de massas gerados por MALDI-Q-TOF do fenótipo ODV e LC MS/MS do fenótipo ODV e BV foram utilizados no servidor Mascot (www.matrixscience.com) mas desta vez, comparados com o banco não redundante de insetos montado a partir do NCBI (<http://www.ncbi.nlm.nih.gov>). Na figura 20 (Apêndice A, tabela S3) foram relacionadas 11 proteínas celulares identificadas no fenótipo ODV: actina, α tubulina, fator alongação 1 α , fator alongação Tu, Histona 4, proteína ribossomal S9, translocase ADP/ATP, ATP sintase e proteína de ligação nuclear GTP α e as chaperonas HSP90 e HSP70 (Figura 20). Já no fenótipo BV, encontramos apenas as proteínas actina e a chaperona HSP 90 (Figura 20). Todas as proteínas do hospedeiro encontradas foram classificadas por localização de acordo com o NCBI e o Uniprot (<http://www.uniprot.org/>): citoesqueleto, ligação ao DNA, membrana interna da mitocôndria e citoplasmática/núcleo (Apêndice A, tabela S3). Resultados preliminares de Hibridização Subtrativa Supressiva (HSS) demonstraram que algumas destas proteínas já apresentavam aumento no nível de expressão de RNA mensageiro durante a infecção por AgMNPV-2D, como foi o caso da chaperona HSP90 (Oliveira et al., 2013). Muitas proteínas de hospedeiro identificadas no nosso estudo já haviam sido descritas em outros vírus envelopados, como o vírus Vaccinia (VV), Herpes vírus (AIHV-1, HSV e KSHV), vírus Epstein-Barr (EBV), citomegalovírus humano (HCMV), vírus respiratório sincicial (RSV), vírus influenza (IV) e vírus da imunodeficiência humana (HIV-1). Além disso, outras proteínas também haviam sido descritas em outros baculovírus (AcMNPV e HearNPV): actina, α tubulina, fator de alongação 1 α , HSP 70 e 90.

Figura 20 - Uma célula UFL-AG-286 representada durante a infecção por AgMNPV-2D mostrando as proteínas celulares adquiridas durante o ciclo de multiplicação viral.

4.1.6 Análises das proteínas por Western blot

Durante os experimentos de espectrometria de massas algumas proteínas demonstraram baixa cobertura e pontuação contra o banco de dados e AgMNPV-2D (dado não apresentado). Portanto, anticorpos policlonais foram utilizados em experimentos de imunomarcação contra as seguintes proteínas: GP64, LEF-1, ODV-E26, PIF-3 e PTP. Os dados obtidos referem-se a dois fenótipos virais purificados de UFL-AG-286, e estão apresentados na figura 21 (Apêndice A, figura 2). A imunomarcação confirmou a presença das proteínas LEF-1, ODV-E26, PIF-3 e PTP somente no ODV (Figura 21). Como é possível observar na figura 21, estas quatro proteínas não foram identificadas no fenótipo BV, corroborando os dados obtidos por espectrometria de massas. A proteína GP64 foi identificada no fenótipo BV como esperado. As análises do proteoma mostraram espectros de baixa cobertura contra a GP64 em ODV, sugerindo a presença desta proteína neste fenótipo. Estes resultados foram posteriormente confirmados pela imunomarcação usando anticorpos policlonais anti-GP64 figura 21.

Figura 21 - Cinco *Western blot* distintos das proteínas de AgMNPV-2D

O extraído total de proteínas BV e ODB purificadas de duas infecções distintas de AgMNPV-2D foram purificados para analisados por *Western blot*. Neste experimentos foram utilizadas células infectadas (I) e células saudáveis (H) como controle positivo e negativo, respectivamente. Os extratos de ODV (1-2) foram aplicados nos cinco géis e os extratos, assim como BV (1-2). À direita foram representados os anticorpos policlonais utilizados: LEF-1, ODV-E26, PIF-3, PTP e GP64. À esquerda o tamanho de cada proteína esperada (kDa). O padrão de peso molecular está representado com M (11-250 kDa).

Sumarizamos os dados obtidos nesta etapa do estudo representando todas as proteínas que compõe as partículas virais (BV e ODV) do AgMNPV-2D obtidas na linhagem UFL-AG286 (Figura 22).

Figura 22 - Proteínas encontradas no AgMNPV-2D

No diagrama, foi representado o total de proteínas que compõe as partículas virais (BV e ODV) do AgMNPV-2D, identificadas com todas abordagens utilizadas: 2DE e MALDI-Q-TOF, SDS-PAGE e LC MS/MS e imunomarcção. Além disso, as proteínas celulares de UFL-AG-286 incorporadas durante a infecção foram representas em cada fenótipo.

4.2 Análise comparativa da ORF-124 (*gp64*)

De maneira distinta do que se conhece sobre o fenótipo ODV de outros baculovírus, encontramos GP64 em AgMNPV neste fenótipo, este achado foi confirmado por espectrometria de massas e imunomarcção. Uma vez que estes experimentos foram realizados apenas com o isolado AGMNPV-2D, optamos por avaliar a diversidade genética deste gene em diferentes isolados disponibilizados pela Embrapa Soja usando sequenciamento de alta cobertura, assim como descrito na seção de métodos. Em seguida, os dados de leitura da ORF 124 (*gp64*) foram analisados separadamente contra o genoma do AgMNPV-2D (104602<106101), que codifica para a glicoproteína GP64 de 499 aminoácidos (Oliveira et al., 2006b). As sequências foram analisadas pelo *software* Geneious para comparar a diversidade genética encontrada na *gp64*. A cobertura média de leitura da ORF124 variou de 500 a 1000 *reads*, portanto, o nível de cobertura de leitura sítio a sítio foi considerado satisfatório para identificar as bases que variaram ao longo do gene. Como o tamanho dos

fragmentos sequenciados pelo método de pirosequenciamento não permite estabelecer a ligação (“*genetic linkage*”) entre as mutações encontradas em diferentes *reads*, a diversidade genética da ORF 124 (1499 pb) também foi avaliada por PCR, clonagem e sequenciamento tradicional (Sanger) (Figura 23), utilizando oligonucleotídeos específicos (Apêndice C, Tabela C3).

Figura 23 - Amplificação da ORF 124 por PCR da região, GP64, dos isolados geográficos de AgMNPV

No gel de agarose 1% a letra M, representa o peso molecular 1KB plus (Invitrogen), ao lado o controle negativo, os números representam os 17 isolados geográficos e o 2D para o AgMNPV-2D, que amplificaram uma banda de 1650 kb. Somente a amostra 29 não apresentou a banda durante a amplificação.

Os fragmentos amplificados foram purificados, sequenciados e os eletroferogramas foram analisados com o auxílio do programa CodonCode Aligner v 1.2.1 (CodonCode Corporation, Dedham, MA, US). Nas clonagens dos isolados geográficos foram utilizados os fragmentos amplificados (Figura 23). Todos os clones foram sequenciados e alinhados em relação à sequência da ORF 124 do AgMNPV-2D. No total, foram sequenciados cinco clones da amostra 26, cinco clones da amostra 27, seis clones da amostra 28, três clones da amostra 29, 15 clones da amostra 30, dois clones da amostra 31 e 10 clones da amostra 32. Todas as sequências foram montadas em *contigs* com o auxílio do programa CodonCode e alinhadas contra o *gp64* do AgMNPV. Para tentar compreender a distribuição espacial dos isolados e avaliar a existência de agrupamentos regionais destas amostras, as relações filogenéticas entre diferentes sequências de cada amostra foram inferidas pelo método de máxima verossimilhança incluindo a região codificadora completa do *gp64* (1499 nucleotídeos).

Figura 24 - Relações filogenéticas entre *gp64* de isolados geográficos de AgMNPV mostrados na Tabela 1

A filogenia foi realizada com os clones dos isolados de AgMNPV sequenciados por Sanger. Abaixo foi representada a localidade de coleta de cada amostragem.

A genealogia inferida a partir deste gene nos possibilitou afirmar que não há agrupamento temporal/espacial de amostras de acordo com isolados para estas populações de AgMNPV, exceto para o isolado 30 (Figura 24).

4.2.1 Avaliação da diversidade da *gp64*

No *software* Geneious, também foram avaliadas as substituições e/ou polimorfismos (SNPs) na região do *gp64* de AgMNPV. Para as análises de variabilidade, consideramos todos os *reads* obtidos no pirosequenciamento (apenas retiramos da análise o NPV 42 devido a baixa cobertura dos *reads*) e todos as sequências dos clones obtidos por Sanger. A figura 25 demonstra o número de sítios variáveis neste gene, considerando todos os isolados juntos ou seja, variabilidade interpopulacional. No total, foram observadas vinte mutações e as frequências de cada uma encontrada por pirossequenciamento foi anotada na tabela C4 no Apêndice C. Em relação ao tipo de mutação, sinônima e não sinônima, a maior parte das mutações detectadas no *gp64* foi do tipo sinônima, 3 vezes mais do que as não sinônimas, totalizando 15 (Figura 25).

Figura 25 - Análise de mutações sinônimas e não sinônimas no *gp64*, entre os isolados geográficos

Os retângulos cinzas representam as mutações sinônimas e os vermelhos não sinônimas detectadas na *gp64* dos isolados geográficos de AgMNPV comparados com o AgMNPV-2D. Os números representam os códons que mudaram neste gene. A *gp64* foi dividida em regiões N-terminal (1-651pb), central (652-1382pb) e C-terminal (1383-1497). Foram identificadas três mutações do tipo não sinônimas no N-terminal, uma do tipo não sinônimas na região central e uma no C-terminal.

Além disso, observamos um número maior dos dois tipos de mutações na região N-terminal (1-651 pb) do *gp64* entre os isolados, tanto em sequências de pirosequenciamento quanto Sanger. Já o número de mutações sinônimas foi prevalente na região central (652-1382pb), com apenas uma substituição não sinônima (701 pb). Por outro lado, na região C-terminal (1-651 pb) foi encontrada apenas uma mutação não sinônima na posição 1479 pb da

gp64. Esse excesso de sinônimas sobre não sinônimas revela uma seleção negativa moderada, ou conservadora, com um valor de dN/dS global de 0,33. Ainda utilizamos a reconstrução filogenética da *gp64* para análise de seleção códon a códon, que não mostrou nenhum códon sob seleção positiva e sete códons sob seleção negativa (Tabela 4).

Tabela 4 - Códon negativamente selecionados detectados na *gp64*

Posição do aa	ω	<i>p</i> -valor
12	-3,734029	0,00001
62	-5,000000	0,004115
79	-5,000000	0,004951
212	-0,000444	0,002973
223	-3,366007	0,027846
265	-3,537871	0,024537
350	-6,000000	0,001372

A posição dos aminoácidos foi referida ao isolado AgMNPV-2D (número de acesso GI 116326192). Sítios negativamente selecionados ($\omega < 0$) foram identificados pela diferença significativa entre as taxas de mutação não sinônima (dN) e sinônima (dS) por sítio, utilizando a contagem ancestral de única verossimilhança implementada no HyPhy.

Ainda tentando entender a presença de GP64 em ODV no AgMNPV-2D, levantamos a hipótese de que as mutações não sinônimas na GP64 poderiam estar presentes em domínios importantes, como alças de fusão ou sítios de glicosilação. Assim, as mutações encontradas pelos dois métodos de sequenciamento foram posicionadas no modelo estrutural da GP64 de AgMNPV baseada na estrutura da GP64 de AcMNPV (Figura 26).

Figura 26 - Localização das mutações não sinônimas no modelo estrutural da proteína GP64 do AgMNPV-2D, baseada na estrutura cristalográfica de raio X da GP64 do AcMNPV

O modelo da GP64 do AgMNPV-2D foi colorido de acordo com os domínios da GP64 do AcMNPV: domínio I (verde), domínio II (vermelho), domínio III (azul), domínio IV (roxo) e domínio V (laranja). A posição do aminoácido foi referida ao primeiro códon da sequência da *gp64* do isolado AgMNPV-2D, cujo número de acesso é GI 116326192. Neste modelo, os círculos cinzas representam os sítios de glicosilação nos resíduos 159, 197, 354, 384 e 426. As mutações não sinônimas foram representadas com círculos vermelhos, apenas 3 foram posicionadas, nos resíduos 50, 96 e 234. As outras duas mutações caíram em sítios não resolvidos pela estrutura de GP64.

5 DISCUSSÃO

5.1 Proteoma do AgMNPV-2D

Diversos estudos comparativos realizados durante a década de 1980 testaram a susceptibilidade de linhagens derivadas de insetos a membros da família *Baculoviridae*. Estes ensaios *in vitro* demonstraram a capacidade dos baculovírus de infectar e replicar, e ao final os títulos virais eram quantificados para cada linhagem (McClintock et al., 1986; McIntosh et al., 1985; Rice, Miller, 1986; Wickham et al., 1992). As diferenças observadas possibilitaram categorizar as linhagens derivadas de insetos em permissíveis, semipermissíveis ou não-permissíveis à infecção para cada baculovírus. Nas linhagens permissíveis, nas quais a síntese de DNA e proteínas é completa, ocorre produção de progênie e lise celular. Nas linhagens semipermissíveis ocorre limitada replicação viral e nas não-permissíveis o vírus entra na célula mas não há replicação (McClintock et al., 1986). Os mecanismos que tornam uma célula permissível ou abortiva à infecção variam entre os diferentes baculovírus e são desconhecidos nas linhagens celulares (Bilimoria, 1992; McClintock et al., 1986; Rice, Miller, 1986). Para o AgMNPV, entre as quatro linhagens de Lepidópteras avaliadas por Castro e colaboradores (1997), a linhagem UFL-AG-286 (da lagarta *Anticarsia gemmatalis*) (Sieburth, Maruniak, 1988) foi a mais produtiva, o que revela a melhor adaptação do AgMNPV ao seu hospedeiro natural (refletida pela maior permissividade), seguida da infecção em IPLB-SF-21 (derivada da lagarta *Spodoptera frugiperda*) (Castro et al., 1997). Além das diferenças relatadas na cinética de infecção deste baculovírus, análises de transcrição gênica global do AgMNPV-2D realizadas em UFL-AG-286 e em IPLB-SF-9, derivada da linhagem IPL-SF-21, evidenciaram diferenças significativas na expressão dos transcritos de 149 ORFs. O padrão de expressão foi mais precoce na linhagem UFL-AG-286 comparada à linhagem IPLB-SF-9 (Oliveira et al., 2013).

Como consequência natural dos resultados obtidos na análise da expressão gênica, malhas regulatórias gênicas (*gene regulatory network* - GRN) do AgMNPV-2D foram inferidas e mostraram-se muito distintas nas duas linhagens, o que pode ser reflexo de diferenças na interação com hospedeiros distintos. Todos estes estudos comparativos nos levaram a refletir se as diferenças qualitativas e quantitativas de cinética e do transcriptoma viral poderiam estar representadas no proteoma do AgMNPV-2D. Por este motivo, nosso

estudo comparou o proteoma do AgMNPV-2D nas duas linhagens mencionadas (UFL-AG-286 e IPLB-SF-9).

5.1.1 Proteínas presentes no fenótipo ODV e BV na linhagem UFL-AG-286

Os primeiros estudos com as proteínas de baculovírus foram realizados por marcação com anticorpos e muitas delas eram preditas com base em dados obtidos por genômica e transcriptômica. Como mencionado anteriormente, os primeiros proteomas da família *Baculoviridae* foram iniciados nos últimos 10 anos e as metodologias de espectrometria de massas possibilitaram a elucidação dos proteoma de seis ODV e dois BV. Estes estudos proteômicos identificaram as proteínas estruturais de cada fenótipo e também as proteínas auxiliares e celulares que são incorporadas durante a infecção viral. Um dos maiores desafios dos proteomas virais consiste em garantir partículas virais livres de contaminantes celulares. Dessa forma, neste trabalho, as coletas das cinéticas virais foram realizadas a 48 h.p.i. para o fenótipo BV e a 72 h.p.i para o fenótipo ODV, com posterior etapa de purificação. Com isso, apresentamos pela primeira vez o perfil das proteínas estruturais do *Anticarsia gemmatilis* múltiplo nucleopoliedrovírus (AgMNPV-2D), isolado de uma infecção 1 MOI em linhagem UFL-AG-286 derivada da lagarta *Anticarsia gemmatilis* (Sieburth, Maruniak, 1988), considerada seu hospedeiro natural (Moscardi, 1999; Moscardi et al., 2011).

Conforme demonstrado nos resultados, as abordagens proteômicas propostas possibilitaram identificar 44 proteínas no fenótipo ODV do AgMNPV-2D (Braconi et al.). Trabalhos semelhantes realizados na família *Baculoviridae* corroboraram nossos resultados (Braunagel, 2003; Deng et al., 2007; Liu et al., 2008; Perera et al., 2007; Wang et al., 2011; Xu et al., 2011). No entanto, seis proteínas virais foram exclusivamente do proteoma do AgMNPV, como: PTP-1, EGT, AG20, PP31, IAP-1 e PK1 (Apêndice A, seção discussão) (Braconi et al.). No proteoma do ODV foram realizadas duas metodologias de espectrometria de massas, MALDI-Q-TOF e LC MS/MS (ESI-Q-TOF). Os resultados obtidos de géis 2DE seguidos de análise por espectrômetro de massas com a fonte MALDI apresentou pouca sensibilidade. Isto se deve ao fato de MALDI ionizar preferencialmente a proteína mais abundante da amostra e não gerar espectros de massas para outras proteínas que podem estar presentes na mesma amostra (Aebersold, Mann, 2003; Cantú, 2008; Steen, Mann, 2004). No entanto, três proteínas foram detectadas somente quando utilizamos esta abordagem: DNAPol (*ag65*), CG30 (*ag85*) e Alk-Exo (*ag129*) (Apêndice A, seção discussão) (Braconi et al.).

Embora os corpos de oclusão (OBs) tenham sido tratados com solução alcalina e os ODVs purificados mais de uma vez, a poliedrina foi encontrada em grande quantidade como contaminante, fato comum entre os estudos proteômicos da família *Baculoviridae*. Estudos semelhantes demonstraram também as limitações da separação de misturas proteicas utilizando eletroforese bidimensional (2DE) (Liu et al., 2008; Zhang et al., 2010a). A presença de proteínas de difícil solubilização, como proteínas ácidas ou básicas, assim como proteínas muito hidrofóbicas podem dificultar a focalização das mesmas na primeira dimensão (Steen, Mann, 2004). Por este motivo, a eletroforese SDS-PAGE e o espectrômetro de massa com uma fonte de ionização ESI (*electrospray*) foram utilizados acoplados a uma cromatografia líquida de alta performance (HPLC).

As análises dos espectros de massas de um LC MS/MS, geraram dados mais robustos e com maior cobertura das proteínas contra os bancos de dados (Apêndice A, material suplementar: tabelas S1 e S2) (Braconi et al.). Através desta metodologia, também foram relacionadas 33 proteínas ao fenótipo BV do AgMNPV-2D. Ademais, identificamos seis proteínas que se acreditava pertencerem apenas ao fenótipo ODV, PP31, Desmoplaquina, ODV-E25, PEP/PP34, ODV-E18, ODV-E56, este achado foi corroborado pelos outros dois proteomas realizados com o fenótipo BV (Apêndice A, seção discussão). Os dados obtidos para *Autographa californica* múltiplo nucleopoliedrovírus (AcMNPV) (Wang et al., 2010) e *Helicoverpa armigera* nucleopoliedrovírus (HearNPV) (Hou et al., 2013) são similares aos encontrados no AgMNPV- 2D (Braconi et al.). No entanto, seis proteínas foram descritas apenas no BV do AgMNPV-2D: AG15, LEF-3, P33, Helicase, PNK/PNL e AG130 (Apêndice A, seção discussão) (Braconi et al.).

Com o estudo das proteínas do AgMNPV-2D, foi possível identificar uma tendência dos dois fenótipos virais de levar no vírion proteínas moduladoras de resposta celular, como proteínas anti-apoptóticas, proteínas reguladoras de ciclo celular e proteínas envolvidas na redução de metamorfose (muda) e comportamento de motilidade na lagarta (Apêndice A, seção discussão) (Braconi et al.). Embora ainda seja muito importante o melhor entendimento da interação vírus célula, poucos experimentos foram realizados no AgMNPV para tentar elucidar as mudanças no hospedeiro em resposta à infecção viral.

5.2 Fenótipo BV nas duas linhagens celulares estudadas

As cinéticas de infecção do AgMNPV-2D tanto em IPLB-SF-9 quanto UFL-AG-286 para o fenótipo BV foram coletadas 48 h.p.i. A figura 18 mostra claramente a diferença na concentração de proteínas obtidas para o fenótipo BV nas duas linhagens. Acreditamos que estas diferenças quantitativas são provenientes do padrão de infecção distinto nas duas linhagens: UFL-AG-286 permissível e IPLB-SF-9 semipermissível. No estudo comparativo realizado por Castro e colaboradores (Castro et al.), nas células UFL-AG-286 infectadas com AgMNPV-2D foram observados efeitos citopáticos após 48 h.p.i. com perda da morfologia celular, 80-90% das células apresentando sinais de hipertrofia nuclear devido a presença de poliedros. Por outro lado, em IPLB-SF-21 (IPLB-SF-9 derivada da linhagem IPL-SF-21), no mesmo período pós-infecção, apenas 40-50% das células apresentam os mesmos efeitos citopáticos (Castro et al., 1997). Além disso, na linhagem UFL-AG-286, trinta e três proteínas foram identificadas para o fenótipo BV, enquanto em IPLB-SF-9 confirmamos apenas vinte e cinco. Crucialmente, os dados sugerem que o “atraso” observado nos dados de transcriptoma do AgMNPV-2D em IPLB-SF-9 pode significar que a etapa crucial de ativação da expressão de genes mais tardios esteja prejudicada, como os que codificam para montagem e oclusão viral, e que este fator pode ser extrapolado para o “atraso” da tradução das proteínas virais na linhagem IPLB-SF-9. Este fato pode explicar o atraso na morfogênese viral nesta linhagem, bem como a ausência das proteínas: AG52, IAP-2, AG73, P33, 38K, P40, P48 e AG113.

5.2.1 Semelhanças encontradas no fenótipo BV nas duas linhagens celulares estudadas

Como esperado, a maioria das proteínas encontradas no fenótipo BV nas duas linhagens celulares estudadas foram propostas como sendo presentes no nucleocapsídeo e são essenciais para sua formação: VP39, P49, ODV-E27, FP25K, VP80/P87, P24, VP1054, VLF-1, Desmoplaquina e PEP/PP34 (Tabela 3 e Apêndice A, tabela S2). Dados similares foram relatados nos dois estudos proteômicos com fenótipo BV do grupo NPVs, AcMNPV e HearNPV (Hou et al., 2013; Wang et al., 2010).

A proteína estrutural VP39 é o maior e mais abundante componente do capsídeo, acreditamos ser este o motivo de estar presente em mais de uma banda no gel de poliacrilamida 12% do fenótipo BV nas duas linhagens (Tabela 3 e Apêndice A, tabela S2).

Além de seu papel estrutural, esta proteína ajuda no transporte dos nucleocapsídeos para o núcleo no início da infecção. Posteriormente, os nucleocapsídeos recém formados de BV são encaminhados para a membrana celular, através da reorganização da actina celular que será discutido em maior detalhe na próxima seção (Charlton, Volkman, 1991; Lanier, Volkman, 1998; Ohkawa et al., 2010). Já P49 e a ODV-E27, proteínas conservadas em todos os baculovírus sequenciados até o momento, são consideradas componentes estruturais da base do nucleocapsídeo (Braunagel et al., 2001; Slack, Arif, 2007).

A proteína ODV-E27 (ORF140) possui grande homologia com a ciclina B da lagarta *Bombyx mori* e é co-precipitada com a quinase dependente de ciclina celular 2 (cdK2). A presença do complexo ODV-E27/cdK2 faz com que o ciclo celular seja mantido em fase G2 de divisão celular, até que o baculovírus complete o ciclo de infecção (Baluchamy, Gopinathan, 2005; Braunagel et al., 1998). Também é interessante ressaltar que, durante a análise de expressão temporal, os primeiros transcritos da ORF140 foram detectados a 11 h.p.i na UFL-AG-286, enquanto só foram detectados 48 h.p.i na linhagem IPLB-SF-9 (Oliveira et al., 2013). Este fato corrobora a nossa hipótese de que pode haver um “atraso” na tradução da ODV-E27 na linhagem celular IPLB-SF-9, que desempenha papel importante na modulação de ciclo celular.

Outra proteína associada ao nucleocapsídeo de BV encontrada nas duas linhagens estudadas foi a FP25K. Mutantes desta proteína em AcMNPV apresentam redução do número de corpos de oclusão, mas sem decaimento da formação de nucleocapsídeos e com aumento na produção de BV (Harrison, Summers, 1995). A VP80/P87 e P24 também foram relacionadas aos nucleocapsídeos de BV nas duas linhagens. Esta proteína é essencial para produção de BV e a maturação dos nucleocapsídeos (Tang et al., 2008), como reforçado pela imunomarcagem e microscopia confocal em EGFP fusionado a VP80 de AcMNPV (Marek et al., 2011). Além da forte interação com as proteínas p78/83 e VLF-1, bem como sinais de colonização de VP80 com DNA viral recém replicado no estroma virôgenico para formação de novos nucleocapsídeos (Marek et al., 2011). Os homólogos de *vp80* são encontrados apenas em *Alphabaculovirus* (Li et al., 1997; Lu, Carstens, 1992). Ademais, aVP80/P87 interage diretamente com outra proteína essencial para montagem de BV, a 38K, o que sugere que esta proteína deva ser essencial também para o BV do AgMNPV-2D (Wu et al., 2006). Entretanto, os espectros de massas desta proteína só foram encontrados no fenótipo BV na linhagem UFL-AG-286. Em AcMNPV, mutantes de 38K são inviáveis, o que sugere que esta proteína também tenha papel crucial no fenótipo BV do AgMNPV-2D (Hefferon, 2003; Wu

et al., 2006). Por outro lado, P24 é conservada em *Alpha e Betabaculovirus*, embora sua função seja pouca descrita. A imunoprecipitação em OpMNPV e AcMNPV demonstrou a presença dela nos nucleocapsídeos de BV e ODV (Wolgamot et al., 1993).

Também confirmamos a presença das proteínas VP1054, Desmoplaquina, VLF-1 e PTP/PP34 nos nucleocapsídeos de BV nas duas linhagens de Lepidópteras. Foi interessante encontrar as proteínas VP1054, Desmoplaquina e VLF-1, que estão relacionadas ao empacotamento do DNA viral no nucleocapsídeo. A VP1054 é uma proteína de 43 kDa codificada por um gene *core* que se associa à VP39 na organização de seus monômeros em anéis para formação do nucleocapsídeo (Olszewski, Miller, 1997b) e é provável que a interação entre VP80/P87, 38K e VP1054 forme a maquinaria que promove a encapsidação do DNA viral (Wu et al., 2008). Em AcMNPV, foi recentemente demonstrado que a ausência de VP1054 reduz drasticamente a produção viral, e que ela é homóloga e estruturalmente semelhante à proteína PUR α celular (Marek et al., 2013). Estas proteínas foram envolvidas com múltiplos papéis celulares, como se ligar a DNA fita simples e RNA, preferencialmente em sequências repetitivas compostas de GGN, onde “N” pode ser qualquer nucleotídeo (Marek et al., 2013). Provavelmente, durante sua evolução os baculovírus “sequestraram” *vp1054* (PUR α) de um hospedeiro ancestral e ela se tornou essencial para a produção da progênie viral (Marek et al., 2013). Como já mencionado, VLF-1 é conservada e possui papel importante na ativação da expressão dos genes muitos tardios em ODV. Portanto, esperamos que o fato desta proteína estar presente no fenótipo BV está mais relacionado a sua atividade de ligação ao DNA pós replicação, exercendo seu papel de recombinase em sítios específicos e processando o DNA viral para o empacotamento dos novos nucleocapsídeos (Vanarsdall et al., 2006).

Com os dados obtidos no fenótipo BV nas duas linhagens, encontramos a proteína conservada envolvida com o enovelamento do DNA viral, a Desmoplaquina. Devido a sua homologia com proteínas envolvidas com ciclo celular, como topoisomerases I, miosina e proteína centromérica (Rohrmann, 2011), é provável que esta proteína também esteja envolvida no empacotamento do DNA viral no AgMNPV-2D. Como já discutido, a proteína PEP/PP34 de 32 kDa, exerce um papel importante na estrutura e proteção do fenótipo ODV no meio ambiente (Apêndice A, seção discussão) (Braconi et al.). No entanto, encontramos esta fosfoproteína também no BV, corroborado pelos dados obtidos no BV do AcMNPV e do HearNPV (Hou et al., 2013; Wang et al., 2010).

Além das proteínas anteriormente descritas, também identificamos a PP31 e a ME53 em mais de um espectro de massas no BV isolado em UFL-AG-286 e em apenas um espectro no BV de IPLB-SF-9. A PP31 é uma fosfoproteína de 30 kDa exclusiva do fenótipo BV envolvida na regulação da expressão de genes tardios (Guarino et al., 2002; Schultz, Friesen, 2009; Yamagishi et al., 2007), que não é essencial para a infectividade deste fenótipo (Wang et al., 2010; Yamagishi et al., 2007). Já a proteína ME53 é conservada em *Alpha e Betabaculovirus* e experimentos *knockout* sugeriram inicialmente que estivesse envolvida na replicação de DNA (Xi et al., 2007). Mais recentemente, em AcMNPV, a ME53 foi envolvida na produção de BV, sem afetar a replicação do DNA viral (de Jong et al., 2009).

Com os espectros de massas obtidos, também observamos proteínas de envelope compartilhadas no fenótipo BV entre as duas linhagens, e não foi surpresa encontrar a glicoproteína GP64 (Tabela 3 e Apêndice A, tabela S2). Curiosamente, também encontramos a proteína F, que é utilizada pelos nucleopoliedrovírus do grupo II e pelos granulovírus e possui função similar a GP64. Os únicos baculovírus que não possuem homólogos da proteína F são NeleNPV (Lauzon et al., 2004) e NeseNPV (Garcia-Maruniak et al., 2004). No grupo NPV I, a proteína F aparentemente perdeu sua função original mas ainda foi relacionadas ao aumento da infectividade oral no fenótipo ODV (Lung et al., 2003).

A proteína ODV-E18 de 9kDa, também foi encontrada no BV derivado da linhagem IPLB-SF-9, sendo essencial para produção de BV (McCarthy et al., 2008). Estudos de imunoprecipitação e expressão gênica em nucleopoliedrovírus demonstraram que as ORFs que codificam para ODV-E18 e a ODV-E27, são transcritas como um RNA mensageiro bicistrônico (Braunagel et al., 1996b; Katsuma et al., 2011). Interessantemente, no genoma do AgMNPV-2D, a ORF 139 (ODV-E18) e a ORF 140 (ODV-E27) estão localizadas na fita positiva a 15 pares de base de distância. Quando traduzidas, estas duas proteínas formam um heterodímero chamado de ODV-E35, que acredita-se ser capaz de estabilizar a ligação da base do nucleocapsídeo com o envelope viral (Braunagel et al., 1996b).

Outra proteína compartilhada é a ODV-E25, que apresenta um sinal transmembrânico N-terminal e foi encontrada no fenótipo BV nas duas linhagens, mas é muito mais abundante no fenótipo ODV. Assim como a ODV-E25, a ODV-E56 apresenta um forte sinal transmembrânico (mas C-terminal) necessário para translocação desta proteína do núcleo para o envelope. Ela está presente no fenótipo BV e não é essencial para formação da partícula viral como no fenótipo ODV (Braunagel et al., 1996a).

Como mencionado em resultados, a proteína V-Ubiquitina de baixo peso molecular (9 kDa) também foi identificada no BV nas duas linhagens estudadas. Proteínas Ubiquitinas estão amplamente distribuídas nos eucariotos, procariotos e nos vírus. Elas se ligam covalentemente à proteínas celulares e mudam a estabilidade, localização e atividade das mesmas. Sua atividade resulta na degradação da proteína alvo (Welchman et al., 2005). No fenótipo BV, a V-ubiquitina se encontra covalentemente ligada à parte interna da membrana, mas não é essencial para sua replicação (Guarino, 1990; Reilly, Guarino, 1996).

5.2.2 Diferenças encontradas no fenótipo BV nas duas linhagens celulares estudadas são congruentes em diferentes níveis de descrição

O ciclo dos baculovírus é regulado por fatores virais e celulares (Chen et al., 2013; Lu, Miller, 1995; Morris, Miller, 1992; Oliveira et al., 2013). Como já foi mencionado, as linhagens utilizadas neste estudo são de origens distintas. A UFL-AG-286 foi originada a partir de tecido embrionário e IPLB-SF-9 de ovariano, o que pode ser um fator importante para as diferenças encontradas no proteoma do BV do AgMNPV-2D. Podemos levantar a hipótese que a linhagem IPLB-SF-9 pode expressar fatores celulares que controlam melhor a infecção viral. Por outro lado, a ausência de espectros de massas para as proteínas AG52, IAP-2, AG73, P33, 38K, P40, P48 e AG113 na linhagem celular IPLB-SF-9 também nos fez questionar se fatores virais podem ser determinantes para esta linhagem suportar limitada replicação do AgMNPV-2D. Possivelmente, o fato do gênero *Spodoptera* não ser o hospedeiro natural do AgMNPV explicaria o comportamento semipermissivo desta interação. Por outro lado, a eficiente replicação do AgMNPV em *Anticarsia gemmatilis*, da qual vem sendo isolado há décadas ao longo das Américas, sugere que este seja seu hospedeiro natural ou pelo menos preferencial; ao qual ele é melhor adaptado. Estes argumentos explicariam a ausência de espectros de massas para as proteínas AG52, IAP-2, AG73, P33, 38K, P40, P48 e AG113 na interação semipermissiva da linhagem celular IPLB-SF-9. Ademais, podemos questionar quais fatores podem ser determinantes para esta marcante diferença, levantando uma série de hipóteses: (i) a linhagem IPLB-SF-9 pode expressar fatores celulares que controlam melhor a infecção viral; (ii) fatores celulares ou virais necessários não estariam presentes para facilitar a multiplicação viral; (iii) fatores celulares e virais, mesmo que presentes, não teriam uma interação satisfatória para a multiplicação viral ou; (iv) alguma combinação entre os itens acima. Um aspecto interessante é que a rede do AgMNPV-2D

inferida durante a infecção do IPLB-SF-9 apresentou menor modularidade (Q) que a GRN em UFL-AG-286 (Oliveira et al., 2013). Crucialmente, a modularidade da GRN em UFL-AG-286 está associada a redundância funcional (Oliveira et al., 2013) e consequentemente a um possível aumento de robustez. Neste sentido, a proposição que o AgMNPV é melhor adaptado ao sistema *Anticarsia* e que a infecção em *Spodoptera* seria efetivamente uma GRN sob perturbação, explicaria a não detecção na espectrometria de massas dos nós circulados em vermelho na figura 27 (AG52, IAP-2, AG73, P33, 38K, P40, P48 e AG113). É importante ressaltar que a maioria (8 de 9) dos nós deletados da GRN em IPLB-SF-9 ocorrem na comunidade I (Oliveira et al., 2013). Este arranjo tem uma baixa probabilidade ($p = 0,00174$) de ser ao acaso, estimada depois de 100.000 amostragens aleatórias de Monte Carlo, levando-se em conta o tamanho de cada comunidade, *e.g.* 56 nós na comunidade I. Este achado surpreendente sugere uma funcionalidade para a modularidade da GRN. Pode-se alegar que a semipermissividade observada é associada a um efeito restrito na GRN o que corrobora a noção de que a compartimentalização associada a redundância aumenta a robustez do sistema. Formalmente, isto foi revelado em inferências feitas em níveis de descrição aninhados, *i.e.*, no contexto da biologia molecular e de sistemas, o que torna esta congruência interessante e meritória de estudos subsequentes.

Figura 27 - Estrutura modular baseada na rede do AgMNPV-2D a partir da infecção em células UFL-AG-286 pelo método de clusterização do programa GLay

A rede de expressão do AgMNPV foi representada com as 149 ORFs e seus componentes diretamente ligados (setas). Ademais foram representadas 5 comunidades (I-V ordenadas por tamanho e os genes distintos que compõe estas comunidades representados numericamente e por função nos retângulos coloridos. Os genes que codificam as proteínas estruturais estão densamente concentrados na comunidade I e os círculos vermelhos indicam os genes que não foram identificados no fenótipo BV na linhagem celular IPLB-SF-9.

Fonte: Adaptada do trabalho de (Oliveira et al., 2013). Com permissão.

Em relação às proteínas codificadas pelas ORF52 e ORF73, pouco foi descrito sobre sua função e papel. Para a ORF113, sabe-se que a proteína de 45 kDa codificada foi encontrada no nucleocapsídeo do ODV de AcMNPV e que está presente apenas nos NPVs do Grupo I (Braunagel, 2003), mas pouco se sabe sobre sua função.

Uma das principais estratégias utilizadas por diversos organismos para conter uma infecção viral é ativar o mecanismo de morte programada na célula pela ativação de enzimas da família das caspases, promovendo a apoptose (Clem, 2007; Ikeda et al., 2011; Luque et al., 2001). Neste sentido, no proteoma do AgMNPV-2D identificamos pela primeira vez membros da família IAPs, IAP-1 no ODV, IAP-2 no BV e também proteína anti-apoptótica p33 (Apêndice A, seção discussão) (Braconi et al.). No entanto, a ausência das proteínas IAP-2 e p33 no BV isolado de IPLB-SF-9 nos chamou a atenção. Acreditamos que a ausência dessas proteínas desde o começo da infecção pode facilitar a entrada da célula em apoptose. Sabe-se que no baculovírus EppoMNPV (*Epiphyas postvittana* nucleopoliedrovírus) IAP-2 apresenta atividade anti-apoptótica (Maguire et al., 2000), mas em AgMNPV-2D esta proteína perdeu o domínio BIR e somente com trabalhos funcionais poderemos verificar se ainda possui ação anti-apoptótica. Já experimentos com p33 mostraram que ela se complexa com a proteína celular p53, envolvida em apoptose e regulação de ciclo celular (Prihod'ko et al., 1999) e a existência de uma proteína homóloga a p53 em insetos sugere que p33 se ligue a ela para inibir apoptose e permitir a progressão do ciclo celular (Rebollar et al., 2006). Além disso, ensaios de gene *knockout* com *p33* em AcMNPV mostraram que sua ausência é letal para a produção de BV (Wu, Passarelli, 2010).

Também não foi possível identificar as proteínas P40 e P48, envolvidas com a morfogênese e a produção de BV (Vanarsdall et al., 2007; Yuan et al., 2008). Outro fator que deve ser levado em consideração é que, apesar da espectrometria de massas constituir uma técnica muito sensível para a identificação de proteínas, alguns experimentos ainda devem ser confirmados com os dados obtidos de BV na linhagem IPLB-SF-9. Alguns fatores podem contribuir para a falha na identificação de proteínas: (i) baixa concentração de uma proteína específica, (ii) degradação proteica, (iii) perda durante o processo de digestão e (iv) falta de sensibilidade do LC MS/MS a peptídeos pequenos (Aebersold, Mann, 2003; Steen, Mann, 2004; Wang et al., 2010; Wang et al., 2011).

5.3 Proteínas celulares encontradas no proteoma do AgMNPV-2D

Neste estudo, ficou clara a necessidade de se obter amostras livres de contaminantes celulares, para que os dados obtidos por espectrometria de massas de proteínas celulares no AgMNPV-2D pudessem refletir a interação de proteínas vírus-célula no nosso modelo experimental. Supreendentemente, foram identificadas 11 proteínas celulares no fenótipo ODV e duas delas foram compartilhadas pelo fenótipo BV (Figura 20). Entre elas estão proteínas com diferentes funções e localizações: citoesqueleto (actina, α -tubulina, EF Tu, HSP90, RPS9 e EF1 $_{\alpha}$), ligação com DNA (H4), mitocôndria (ANT1 e ATP-synt abN) e nuclear (RAN e HSP70) (Apêndice A, tabela S3 em material suplementar). O proteoma quantitativo da linhagem *Spodoptera frugiperda* foi realizado, comparando as proteínas celulares obtidas durante uma infecção por AcMNPV com as proteínas de uma célula não infectada (Carinhas et al., 2011). Nossos resultados de proteínas celulares estão de acordo com o estudo realizado na linhagem IPLB-SF-9, que identificou muitas proteínas de ribossomos e de citoesqueleto alteradas durante a infecção (Carinhas et al., 2011).

Além disso, as chaperonas *heat shock* 70 e 90 também foram encontradas “up” reguladas, o que corroborou tanto nossos dados obtidos por hibridização subtrativa (HSS) de genes celulares (Oliveira et al., 2013) quanto o proteoma. Na estrutura do ODV, encontramos HSP70 e 90 e na estrutura do BV a HSP90, nas duas linhagens estudadas. Estas duas proteínas desempenham múltiplos papéis na célula, entre eles regulação de ciclo celular, diferenciação e principalmente dobramento de proteínas (Mayer, Bukau, 2005). A expressão de chaperonas pode aumentar consideravelmente quando o organismo está sob diferentes condições de patologia, como infecções virais (Carinhas et al., 2011; Mayer, Bukau, 2005). Recentemente, demonstrou-se que a HSP90 do vírus sincicial respiratório (RSV) é utilizada para a formação e transmissão da partícula viral, sugerindo um papel importante também para a maturação viral (Radhakrishnan et al., 2010). Diante disso, acreditamos que esta proteína tem uma função análoga em AgMNPV-2D.

Os únicos estudos com a família *Baculoviridae* que buscaram proteínas celulares presentes em seus vírions foram o BV do AcMNPV (Wang et al., 2010) e o BV e o ODV de HearNPV (Hou et al., 2013). Como demonstrado em nossos resultados, identificamos actina do hospedeiro compartilhada nos dois fenótipos estudados (Figura 20). Achado corroborado pelos proteomas BV de AcMNPV (Wang et al., 2010) e de HearNPV (Hou et al., 2013). Esta proteína presente no citoesqueleto celular é utilizada em diferentes estágios do ciclo dos

baculovírus. No início da infecção, as proteínas do nucleocapsídeo viral interagem com a actina de forma a possibilitar o transporte do nucleocapsídeo para o núcleo (Charlton, Volkman, 1991; Lanier, Volkman, 1998; Volkman, 1988). Já na fase tardia, a actina está envolvida com a formação dos novos capsídeos. E o estudo de Marek e colaboradores (2011) mostra claramente a distribuição e sobreposição de VP80 com filamentos de actina no estroma virogênico. A actina também está envolvida com o transporte dos nucleocapsídeos para região periférica da membrana nuclear e com o brotamento dos mesmos do núcleo, seguido pelo brotamento dos novos vírus do citoplasma (Marek et al., 2011; Ohkawa et al., 2010).

De uma forma geral, as proteínas celulares encontradas no fenótipo BV foram similares aos proteomas do mesmo fenótipo de AcMNPV e de HearNPV. Contraditoriamente, no BV de AcMNPV e de HearNPV foram associadas 11 e 101 proteínas celulares neste fenótipo, respectivamente. No entanto, o tempo de coleta da cinética viral utilizada nestes estudos foi de 72 h.p.i. (Hou et al., 2013; Wang et al., 2010), enquanto para o AgMNPV-2D os BVs foram coletados em 48 h.p.i.. Acreditamos que as diferenças encontradas se devam à expressão de proteínas celulares consideravelmente aumentada no tempo de coleta de 72 h.p.i., tentando conter a infecção viral, e à diferença de sensibilidade nas metodologias utilizadas. Até o momento, não é sabido se estas proteínas são ocluídas ao acaso durante a montagem viral ou se há seleção para que isto aconteça. Curiosamente, encontramos um padrão de proteínas celulares semelhantes ao descrito para outros vírus envelopados associadas ao AgMNPV-2D. Possivelmente, os vírus não relacionados filogeneticamente podem ter as mesmas necessidades no início da infecção e possivelmente precisam do mesmo conjunto de proteínas celulares exclusivas para os vírus envelopados (Apêndice A, seção discussão) (Braconi et al.).

5.4 Análise comparativa da *gp64*

Com os dados provenientes dos diversos projetos de sequenciamento concluídos até o momento, nosso grupo de pesquisa tem conduzido análises de sequências gênicas por meio de uma abordagem comparativa (Oliveira et al., 2006b; Wolff et al., 2008; Zanutto, Krakauer, 2008). Informações obtidas de genomas de organismos evolutivamente próximos e distantes estão possibilitando um melhor entendimento a cerca da evolução dos baculovírus em particular (Oliveira et al., 2006b; Wolff et al., 2008) e de diversas famílias de vírus de DNA

em geral (Zanotto, Krakauer, 2008). A reconstrução filogenética de proteínas tem sido útil para se inferir a função e o modo de interação de diversas outras proteínas (Herniou et al., 2003). Neste sentido, as filogenias de *gp64* demonstram a existência de dois subgrupos dentro de NPV I e suportam a noção que a GP64 é uma sinapomorfia do NPV, provavelmente uma aquisição recente deste grupo (Herniou et al., 2003). A GP64 do AgMNPV é mais relacionada com as dos vírus do subgrupo I, o que é corroborado pelas árvores genômicas (Oliveira et al., 2006b; Wolff et al., 2008). Ademais, com os dados da malha regulatória gênica (*gene regulatory network* - GRN) do AgMNPV-2D em UFL-AG-286 observa-se que *gp64* é um nó fundamental, apresentando um dos valores mais elevados de *betweenness centrality* (BC) (Oliveira et al., 2013). Este valor significa a importância de cada nó em uma rede e o número de vezes que esta via é utilizada. É muito importante considerar que altos valores de BC encontrados correlacionam com baixo valor de diversidade polimórfica não só para *gp64* (Oliveira et al., 2013). Dada a importância de *gp64* no AgMNPV e a presença de GP64 no proteoma de ODV, sequenciamos a *gp64* dos diferentes isolados geográficos que pudessem nos dar indícios de regiões evolutivamente importantes e/ou aquisição de novas funções.

Conforme apresentado na seção de resultados, os dados de *gp64* do AgMNPV obtidos por técnicas de clonagem, sequenciamento por Sanger e com alta cobertura, demonstram a baixa variabilidade interpopulacional neste gene em todos isolados geográficos estudados. Este achado é corroborado pelo genoma de baculovírus ser DNA fita dupla, cuja taxa de substituição por sítio por ano estimada é em torno de 10^{-6} a 10^{-8} (Duffy et al., 2008). Outro resultado interessante foi o grande número de substituições sinônimas em relação as não-sinônimas nesta população. No entanto, observamos seleção negativa com significância estatística apenas para os códons 12, 62, 79, 212, 265 e 350 da GP64 (Tabela 4), o que indica que esta proteína de AgMNPV esteja sob seleção purificadora. As mutações ocorrem de maneira aleatória nos genomas dos organismos e a seleção apenas atuará na fixação das mesmas em nível populacional ao longo do tempo. Não obstante, o que chama a atenção nos dados apresentados na reconstrução filogenética de *gp64* (Figura 24), que com exceção do agrupamento formado pelas linhagens do isolado 30, as de outros isolados apresentaram-se dispersas pela árvore. Alternativamente, este padrão pode ser sugestivo de diferentes eventos de migração entre as localidades representadas no mapa (Figura 24).

Observamos também o maior número dos dois tipos de substituições, sinônimas e não sinônimas, na região N-terminal (1-651 pb) do gene da *gp64*. O aparecimento de

substituições não-sinônimas nas mesmas posições entre isolados de AgMNPV obtidos de diferentes tempos amostrais e localidades, corrobora a importância deste achado. Acreditamos com base na estrutura resolvida da GP64 do AcMNPV que a maior concentração de mutações se deva ao fato desta região interagir com diferentes proteínas virais e intracelulares (Kadlec et al., 2008). Na estrutura da GP64 do AcMNPV foi mapeada a alça de fusão no domínio I entre os aminoácidos 121 e 160 próximo ao C-terminal (Kadlec et al., 2008) e não ao N-terminal como havia sido previamente proposto (Zhou, Blissard, 2008). A alça de fusão é responsável pela ligação da proteína aos receptores celulares e mutações pontuais na alça 2 dos resíduos 153 (fenilalanina), 154 (alanina) e 156 (histidina) de GP64 do AcMNPV inibem a fusão do vírus na linhagem IPLB-SF-9, sem formação de sincício (Kadlec et al., 2008; Zhou, Blissard, 2008). Ademais, também não é observada fusão viral quando há mutação nos resíduos localizados na alça 1 do domínio I, S81L e D82F, quando dois resíduos polares são substituídos por resíduos hidrofóbicos (Kadlec et al., 2008). Outro trabalho reportou que quando certos sítios de glicosilação desta proteína são eliminados a ligação do vírus com a célula diminui, sugerindo que glicosilação para GP64 pode ter papel importante na ligação vírus-célula (Jarvis et al., 1998), por este motivo checamos a localização dos sítios de glicosilação (Figura 26). Todas as mutações encontradas na GP64 do AgMNPV não estão relacionadas aos aminoácidos de fusão e sítios de glicosilação. Além disso, sabe-se que o domínio III está presente na região central de GP64 (Figura 26), formando o centro do trímero com 11 resíduos hidrofóbicos (Kadlec et al., 2008). Como esperado, as regiões que desempenham funções estruturais importantes na GP64 do AgMNPV (*e.g.* trimerização, e domínio V em laranja na figura 26) e/ou interagem com moléculas bastante conservadas (*e.g.* região transmembrânica, domínio III em azul e IV em roxo na figura 26) sofreram pressão purificadora, registrada pelas mutações sinônimas, que se presume serem fruto de seleção negativa. Portanto, acreditamos que trabalhos funcionais são necessários a fim de prever o impacto dos polimorfismos na interação proteína-proteína e principalmente determinar o papel de GP64 no ODV de AgMNPV.

O estudo de vírus tem revelado e elucidado uma série de mecanismos celulares, muitos dos quais se apresentam de maneira conservada em diversos organismos evolutivamente relacionados (Volkman, 2007). Os resultados obtidos a partir do proteoma do fenótipo ODV nos forneceram subsídios para levantar a hipótese que a presença de GP64 contribui para uma infecção viral mais efetiva deste fenótipo, assim como se observa no fenótipo BV. Esta proteína estrutural foi recentemente adquirida pelo NPV Grupo I e pode

apresentar uma nova propriedade biológica com funções mais precoces no ciclo viral ou ser necessária para interagir com fatores celulares no fenótipo ODV (Slack, Arif, 2007). Neste estudo, acreditamos ter gerado dados que corroboram a categorização de AgMNPV em dois fenótipos virais (Figuras 19 e 22), bem como a associação de suas proteínas com o hospedeiro, contribuindo para um melhor entendimento acerca da biologia do AgMNPV. Os achados de GP64 do AgMNPV nos revelam que esta proteína pode ter um novo papel no ODV, que deve ser confirmado com investigação funcional, mostrando a interação desta proteína com outras. A fim de se explorar seu potencial para futuras aplicações como apresentação de antígeno, alteração de tropismo celular e *gene delivery*. Portanto, a maior compreensão dos processos de infecção no nível molecular, bioquímico e celular permitirá expandir o espectro de hospedeiros.

6 CONCLUSÕES

Com os dados obtidos no proteoma do AgMNPV-2D e análise comparativa do gene *gp64* dos isolados de AgMNPV as principais conclusões deste trabalho foram:

- Ao todo foram identificadas 44 proteínas estruturais presentes no ODV isolado da linhagem UFL-AG-286, entre elas seis proteínas não foram previamente descritas neste fenótipo ou em outros baculovírus;
- No total foram identificadas 33 proteínas estruturais no BV. Entre elas, sete novas proteínas foram associadas a este fenótipo, previamente desconhecidas nos dois proteomas disponíveis da família *Baculoviridae*;
- No proteoma, foram encontradas proteínas associadas à modulação e regulação de ciclo celular, bem como ao controle de atividade sistêmica do hospedeiro, nos dois fenótipos virais. Possivelmente para facilitar a transmissão viral;
- Pudemos observar onze proteínas celulares na partícula viral ODV, proteínas similares são “sequestradas” por outros vírus envelopados não relacionados filogeneticamente;
- Oito proteínas estavam ausentes no proteoma do fenótipo BV obtido da linhagem semipermissiva IPLB-SF-9 comparada à linhagem permissiva UFL-AG-286. Coerente com os dados obtidos de expressão gênica e inferências topológicas da GRN viral;
- Em relação às sequências de *gp64*, observamos um perfil similar de substituições tanto com os dados obtidos por Sanger quanto por sequenciamento de alta cobertura entre os isolados geográficos;
- Não foi identificado nenhum sítio sob seleção positiva na *gp64* dos isolados geográficos, forte indicativo de que este gene esteja sob pressão purificadora;
- As sequências obtidas por Sanger para *gp64* dos isolados geográficos sugerem diferentes eventos de migração entre as localidades amostradas;
- Nenhuma relação entre os códons substituídos no gene *gp64* na população evidencia mudança de sítio de fusão, glicosilação e estrutura proteica.

REFERÊNCIAS*

- Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*. 2003;422(6928):198-207.
- Aebersold RH, Leavitt J, Saavedra RA, Hood LE, Kent SB. Internal amino acid sequence analysis of proteins separated by one- or two-dimensional gel electrophoresis after in situ protease digestion on nitrocellulose. *Proc Natl Acad Sci U S A*. 1987;84(20):6970-4.
- Aebersold RH, Pipes G, Hood LE, Kent SB. N-terminal and internal sequence determination of microgram amounts of proteins separated by isoelectric focusing in immobilized pH gradients. *Electrophoresis*. 1988;9(9):520-30.
- Airenne KJ, Hu YC, Kost TA, Smith RH, Kotin RM, Ono C, et al. Baculovirus: an insect-derived vector for diverse gene transfer applications. *Mol Ther*. 2013;21(4):739-49.
- Akiyoshi D, Chakerian R, Rohrmann GF, Nesson MH, Beaudreau GS. Cloning and sequencing of the granulin gene from the *Trichoplusia ni* granulosis virus. *Virology*. 1985;141(2):328-32.
- Allen GE, Knell JD. A nuclear polyhedrosis virus of *Anticarsia gemmatilis*: I, ultrastructure, replication, and pathogenicity. *Fl Entomol*. 1977(60):233-40.
- Amster IJ. Fourier transform mass spectrometry. *Journal of mass spectrometry*. 1996;31(12):1325-37.
- Ayres MD, Howard SC, Kuzio J, Lopez-Ferber M, Possee RD. The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. *Virology*. 1994;202(2):586-605.
- Baluchamy S, Gopinathan KP. Characterization of a cyclin homolog from *Bombyx mori* nucleopolyhedrovirus. *Virus research*. 2005;108(1):69-81.
- Benz GA. Introduction: historical perspective. In: Granados RR, and Federici, B.A, editor. *The Biology of Baculoviruses*. Florida: CRC; 1986. p. 1- 35.
- Bilimoria SL. The biology of nuclear polyhedrosis viruses. In: Kurstak E, editor. *Viruses of Invertebrates*. New York: Marcel Dekker, INC; 1991. p. 1-72.
- Bilimoria SL, Demirbag, Z., NG, H., & Reinisch, A. J. Abortive cell culture infections of nuclear polyhedrosis viruses as model systems for host specificity. *Pesquisa Agropecuária Brasileira*. 1992;Brasília 27:123-41.
- Birnbaum MJ, Clem RJ, Miller LK. An apoptosis-inhibiting gene from a nuclear polyhedrosis virus encoding a polypeptide with Cys/His sequence motifs. *J Virol*. 1994;68(4):2521-8.

* De acordo com: International Committee of Medical Journal Editors. [Internet]. Uniform requirements for manuscripts submitted to Biomedical Journal: sample references. [updated 2011 Jul 15]. Available from: <http://www.icmje.org>

Blissard GW, Rohrmann GF. Baculovirus diversity and molecular biology. *Annu Rev Entomol.* 1990;35:127-55.

Blissard GW, Wenz JR. Baculovirus gp64 envelope glycoprotein is sufficient to mediate pH-dependent membrane fusion. *J Virol.* 1992;66(11):6829-35.

Blissard GW. Baculovirus--insect cell interactions. *Cytotechnology.* 1996;20(1-3):73-93.

Boyce FM, Bucher NL. Baculovirus-mediated gene transfer into mammalian cells. *Proc Natl Acad Sci U S A.* 1996;93(6):2348-52.

Braconi CT, Ardisson-Araújo DM, Leme AP, Oliveira JV, Pauletti B, Garcia-Maruniak A, et al. Cell modulation factors are carried along structural proteins in the baculovirus AgMNPV-2D budded and occluded virions. *Proteomics.* 2013;[submetido].

Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem.* 1976;72:248-54.

Braunagel SC, Elton DM, Ma H, Summers MD. Identification and analysis of an *Autographa californica* nuclear polyhedrosis virus structural protein of the occlusion-derived virus envelope: ODV-E56. *Virology.* 1996a;217(1):97-110.

Braunagel SC, He H, Ramamurthy P, Summers MD. Transcription, translation, and cellular localization of three *Autographa californica* nuclear polyhedrosis virus structural proteins: ODV-E18, ODV-E35, and ODV-EC27. *Virology.* 1996b;222(1):100-14.

Braunagel SC, Parr R, Belyavskiy M, Summers MD. *Autographa californica* nucleopolyhedrovirus infection results in Sf9 cell cycle arrest at G2/M phase. *Virology.* 1998;244(1):195-211.

Braunagel SC, Guidry PA, Rosas-Acosta G, Engelking L, Summers MD. Identification of BV/ODV-C42, an *Autographa californica* nucleopolyhedrovirus orf101-encoded structural protein detected in infected-cell complexes with ODV-EC27 and p78/83. *J Virol.* 2001;75(24):12331-8.

Braunagel SC. Determination of the protein composition of the occlusion-derived virus of *Autographa californica* nucleopolyhedrovirus. *Proceedings of the National Academy of Sciences.* 2003;100(17):9797-802.

Braunagel SC, Williamson ST, Saksena S, Zhong Z, Russell WK, Russell DH, et al. Trafficking of ODV-E66 is mediated via a sorting motif and other viral proteins: facilitated trafficking to the inner nuclear membrane. *Proc Natl Acad Sci U S A.* 2004;101(22):8372-7.

Brown CS, Van Lent JW, Vlak JM, Spaan WJ. Assembly of empty capsids by using baculovirus recombinants expressing human parvovirus B19 structural proteins. *J Virol.* 1991;65(5):2702-6.

Cantú MD, Carrilho, E., Wulff, N. A., & Palma, M. S. . Sequenciamento de peptídeos usando espectrometria de massas: um guia prático. *Química Nova.* 2008;31(3):669-75.

- Carinhas N, Robitaille AM, Moes S, Carrondo MJ, Jenoe P, Oliveira R, et al. Quantitative proteomics of *Spodoptera frugiperda* cells during growth and baculovirus infection. *PLoS ONE*. 2011;6(10):e26444.
- Castro ME, Souza ML, Araujo S, Bilimoria SL. Replication of *Anticarsia gemmatalis* nuclear polyhedrosis virus in four lepidopteran cell lines. *J Invertebr Pathol*. 1997;69(1):40-5.
- Castro Oliveira JV, Melo FL, Romano CM, Iamarino A, Rizzi TS, Yeda FP, et al. Structural and phylogenetic relationship of ORF 31 from the *Anticarsia gemmatalis* MNPV to poly (ADP-ribose) polymerases (PARP). *Virus Genes*. 2008;37(2):177-84.
- Charlton CA, Volkman LE. Sequential rearrangement and nuclear polymerization of actin in baculovirus-infected *Spodoptera frugiperda* cells. *J Virol*. 1991;65(3):1219-27.
- Chen YR, Zhong S, Fei Z, Hashimoto Y, Xiang JZ, Zhang S, et al. The transcriptome of the baculovirus *Autographa californica* multiple nucleopolyhedrovirus in *Trichoplusia ni* cells. *J Virol*. 2013;87(11):6391-405.
- Choi J, Guarino LA. A temperature-sensitive IE1 protein of *Autographa californica* nuclear polyhedrosis virus has altered transactivation and DNA binding activities. *Virology*. 1995;209(1):90-8.
- Clem RJ, Miller LK. Control of programmed cell death by the baculovirus genes p35 and iap. *Mol Cell Biol*. 1994;14(8):5212-22.
- Clem RJ. Baculoviruses and apoptosis: the good, the bad, and the ugly. *Cell Death Differ*. 2001;8(2):137-43.
- Clem RJ. Baculoviruses and apoptosis: a diversity of genes and responses. *Curr Drug Targets*. 2007;8(10):1069-74.
- Couch TL, Ignoffo CM. Formulation of insect pathogens. In: Burges HD, editor. *Microbial control of pests and plant diseases 1970-80* 1981. p. 621-34.
- Coulibaly F, Chiu E, Gutmann S, Rajendran C, Haebel PW, Ikeda K, et al. The atomic structure of baculovirus polyhedra reveals the independent emergence of infectious crystals in DNA and RNA viruses. *Proc Natl Acad Sci U S A*. 2009;106(52):22205-10.
- Craig NL, Nash HA. The mechanism of phage lambda site-specific recombination: site-specific breakage of DNA by Int topoisomerase. *Cell*. 1983;35(3 Pt 2):795-803.
- Crook D. Baculoviridae: Sub-group b: Comparative aspects of granulosis viruses. . In: E K, editor. *Viruses of invertebrates*. New York: Marcel Dekker; 1991. p. 73-110.
- Crook NE, Clem RJ, Miller LK. An apoptosis-inhibiting baculovirus gene with a zinc finger-like motif. *J Virol*. 1993;67(4):2168-74.
- Daimon T, Katsuma S, Kang W, Shimada T. Comparative studies of *Bombyx mori* nucleopolyhedrovirus chitinase and its host ortholog, BmChi-h. *Biochemical and biophysical research communications*. 2006;345(2):825-33.

- de Jong J, Arif BM, Theilmann DA, Krell PJ. Autographa californica multiple nucleopolyhedrovirus me53 (ac140) is a nonessential gene required for efficient budded-virus production. *J Virol*. 2009;83(15):7440-8.
- Deng F, Wang R, Fang M, Jiang Y, Xu X, Wang H, et al. Proteomics Analysis of Helicoverpa armigera Single Nucleocapsid Nucleopolyhedrovirus Identified Two New Occlusion-Derived Virus-Associated Proteins, HA44 and HA100. *Journal of Virology*. 2007;81(17):9377-85.
- Dow JA. pH GRADIENTS IN LEPIDOPTERAN MIDGUT. *J Exp Biol*. 1992;172(Pt 1):355-75.
- Duffy S, Shackelton LA, Holmes EC. Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet*. 2008;9(4):267-76.
- Durantel D, Croizier L, Ayres MD, Croizier G, Possee RD, Lopez-Ferber M. The pnk/pnl gene (ORF 86) of Autographa californica nucleopolyhedrovirus is a non-essential, immediate early gene. *J Gen Virol*. 1998;79 (Pt 3):629-37.
- Edman P. A method for the determination of amino acid sequence in peptides. *Arch Biochem*. 1949;22(3):475.
- Erlandson M. Insect Pest Control by Viruses. In: Editors-in-Chief:↯↯ BWJM, Regenmortel MHVv, editors. *Encyclopedia of Virology (Third Edition)*. Oxford: Academic Press; 2008. p. 125-33.
- Fang M, Wang H, Yuan L, Chen X, Vlak JM, Hu Z. Open reading frame 94 of Helicoverpa armigera single nucleocapsid nucleopolyhedrovirus encodes a novel conserved occlusion-derived virion protein, ODV-EC43. *J Gen Virol*. 2003;84(Pt 11):3021-7.
- Fang M, Nie Y, Wang Q, Deng F, Wang R, Wang H, et al. Open reading frame 132 of Helicoverpa armigera nucleopolyhedrovirus encodes a functional per os infectivity factor (PIF-2). *J Gen Virol*. 2006;87(Pt 9):2563-9.
- Faulkner P, Kuzio J, Williams GV, Wilson JA. Analysis of p74, a PDV envelope protein of Autographa californica nucleopolyhedrovirus required for occlusion body infectivity in vivo. *J Gen Virol*. 1997;78 (Pt 12):3091-100.
- Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. *Science*. 1989;246(4926):64-71.
- Fraser MJ. Ultrastructural observations of virion maturation in Autographa californica nuclear polyhedrosis virus infected Spodoptera frugiperda cell cultures. *Journal of ultrastructure and molecular structure research*. 1986;95(1):189-95.
- French TJ, Marshall JJ, Roy P. Assembly of double-shelled, viruslike particles of bluetongue virus by the simultaneous expression of four structural proteins. *J Virol*. 1990;64(12):5695-700.
- Garavaglia MJ, Miele SA, Iserte JA, Belaich MN, Ghiringhelli PD. The ac53, ac78, ac101, and ac103 genes are newly discovered core genes in the family Baculoviridae. *J Virol*. 2012;86(22):12069-79.

- Garcia-Maruniak A, Pavan OH, Maruniak JE. A variable region of *Anticarsia gemmatilis* nuclear polyhedrosis virus contains tandemly repeated DNA sequences. *Virus Res.* 1996;41(2):123-32.
- Garcia-Maruniak A, Maruniak JE, Zanutto PMA, Doumbouya AE, Liu JC, Merritt TM, et al. Sequence Analysis of the Genome of the *Neodiprion sertifer* Nucleopolyhedrovirus. *Journal of Virology.* 2004;78(13):7036-51.
- Garry CE, Garry RF. Proteomics computational analyses suggest that baculovirus GP64 superfamily proteins are class III penetrenes. *Virology Journal.* 2008;5(1):28.
- Granados RR, Derksen AC, Dwyer KG. Replication of the *Trichoplusia ni* granulosis and nuclear polyhedrosis viruses in cell cultures. *Virology.* 1986;152(2):472-6.
- Graves PR, Haystead TAJ. Molecular Biologist's Guide to Proteomics. *Microbiology and Molecular Biology Reviews.* 2002;66(1):39-63.
- Gross CH, Shuman S. Characterization of a baculovirus-encoded RNA 5'-triphosphatase. *J Virol.* 1998;72(9):7057-63.
- Guarino LA. Identification of a viral gene encoding a ubiquitin-like protein. *Proc Natl Acad Sci U S A.* 1990;87(1):409-13.
- Guarino LA, Mistretta TA, Dong W. DNA binding activity of the baculovirus late expression factor PP31. *Virus Res.* 2002;90(1-2):187-95.
- Hang X, Dong W, Guarino LA. The *lef-3* gene of *Autographa californica* nuclear polyhedrosis virus encodes a single-stranded DNA-binding protein. *J Virol.* 1995;69(6):3924-8.
- Harrison RL, Summers MD. Mutations in the *Autographa californica* multinucleocapsid nuclear polyhedrosis virus 25 kDa protein gene result in reduced virion occlusion, altered intranuclear envelopment and enhanced virus production. *J Gen Virol.* 1995;76 (Pt 6):1451-9.
- Harrison RL, Jarvis DL, Summers MD. The role of the AcMNPV 25K gene, "FP25," in baculovirus *polh* and *p10* expression. *Virology.* 1996;226(1):34-46.
- Harrison RL, Jarvis DL. Protein N-glycosylation in the baculovirus-insect cell expression system and engineering of insect cells to produce "mammalianized" recombinant glycoproteins. *Advances in virus research.* 2006;68:159-91.
- Hefferon KL. ORF98 of *Autographa californica* nucleopolyhedrovirus is an auxiliary factor in late gene expression. *Can J Microbiol.* 2003;49(3):157-63.
- Heldwein EE, Lou H, Bender FC, Cohen GH, Eisenberg RJ, Harrison SC. Crystal structure of glycoprotein B from herpes simplex virus 1. *Science.* 2006;313(5784):217-20.
- Herniou EA, Luque T, Chen X, Vlcek JM, Winstanley D, Cory JS, et al. Use of whole genome sequence data to infer baculovirus phylogeny. *J Virol.* 2001;75(17):8117-26.
- Herniou EA, Olszewski JA, Cory JS, O'Reilly DR. The genome sequence and evolution of baculoviruses. *Annu Rev Entomol.* 2003;48:211-34.

- Herniou EA, Olszewski JA, O'Reilly DR, Cory JS. Ancient coevolution of baculoviruses and their insect hosts. *J Virol.* 2004;78(7):3244-51.
- Hofmann C, Sandig V, Jennings G, Rudolph M, Schlag P, Strauss M. Efficient gene transfer into human hepatocytes by baculovirus vectors. *Proc Natl Acad Sci U S A.* 1995;92(22):10099-103.
- Hoover K, Grove M, Gardner M, Hughes DP, McNeil J, Slavicek J. A gene for an extended phenotype. *Science.* 2011;333(6048):1401.
- Hou D, Zhang L, Deng F, Fang W, Wang R, Liu X, et al. Comparative Proteomics Reveal Fundamental Structural and Functional Differences between the Two Progeny Phenotypes of a Baculovirus. *J Virol.* 2013;87(2):829-39.
- Hu YC. Baculovirus vectors for gene therapy. *Advances in virus research.* 2006;68:287-320.
- Iamarino A, de Melo FL, Braconi CT, Zanotto PM. BF integrase genes of HIV-1 circulating in Sao Paulo, Brazil, with a recurrent recombination region. *PLoS ONE.* 2012;7(4):e34324.
- Ikeda M, Yamada H, Ito H, Kobayashi M. Baculovirus IAP1 induces caspase-dependent apoptosis in insect cells. *J Gen Virol.* 2011;92(Pt 11):2654-63.
- Jarvis DL, Wills L, Burow G, Bohlmeier DA. Mutational analysis of the N-linked glycans on Autographa californica nucleopolyhedrovirus gp64. *J Virol.* 1998;72(12):9459-69.
- Jehle JA, Lange M, Wang H, Hu Z, Wang Y, Hauschild R. Molecular identification and phylogenetic analysis of baculoviruses from Lepidoptera. *Virology.* 2006;346(1):180-93.
- Jkel I, Westenberg M, Goldbach RW, Blissard GW, Vlak JM, Zuidema D. A novel baculovirus envelope fusion protein with a proprotein convertase cleavage site. *Virology.* 2000;275(1):30-41.
- Kadlec J, Loureiro S, Abrescia NG, Stuart DI, Jones IM. The postfusion structure of baculovirus gp64 supports a unified view of viral fusion machines. *Nat Struct Mol Biol.* 2008;15(10):1024-30.
- Kamita SG, Nagasaka K, Chua JW, Shimada T, Mita K, Kobayashi M, et al. A baculovirus-encoded protein tyrosine phosphatase gene induces enhanced locomotory activity in a lepidopteran host. *Proc Natl Acad Sci U S A.* 2005;102(7):2584-9.
- Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem.* 1988;60(20):2299-301.
- Katsuma S, Mita K, Shimada T. ERK- and JNK-dependent signaling pathways contribute to Bombyx mori nucleopolyhedrovirus infection. *J Virol.* 2007;81(24):13700-9.
- Katsuma S, Kang W, Shin-i T, Ohishi K, Kadota K, Kohara Y, et al. Mass identification of transcriptional units expressed from the Bombyx mori nucleopolyhedrovirus genome. *J Gen Virol.* 2011;92(Pt 1):200-3.

Katsuma S, Koyano Y, Kang W, Kokusho R, Kamita SG, Shimada T. The baculovirus uses a captured host phosphatase to induce enhanced locomotory activity in host caterpillars. *PLoS Pathog.* 2012;8(4):e1002644.

Ke J, Wang J, Deng R, Wang X. Autographa californica multiple nucleopolyhedrovirus ac66 is required for the efficient egress of nucleocapsids from the nucleus, general synthesis of preoccluded virions and occlusion body formation. *Virology.* 2008;374(2):421-31.

Keddie BA, Aponte GW, Volkman LE. The pathway of infection of Autographa californica nuclear polyhedrosis virus in an insect host. *Science.* 1989;243(4899):1728-30.

Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem.* 2002;74(20):5383-92.

Kim YK, Kwon JT, Choi JY, Jiang HL, Arote R, Jere D, et al. Suppression of tumor growth in xenograft model mice by programmed cell death 4 gene delivery using folate-PEG-baculovirus. *Cancer Gene Ther.* 2010;17(11):751-60.

King LA, Possee RD. The Baculovirus Expression System - A laboratory guide. New York, NY (USA): Chapman & Hall; 1992.

Kingsley DH, Behbahani A, Rashtian A, Blissard GW, Zimmerberg J. A discrete stage of baculovirus GP64-mediated membrane fusion. *Mol Biol Cell.* 1999;10(12):4191-200.

Klose J. Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik.* 1975;26(3):231-43.

Kool M, Ahrens CH, Goldbach RW, Rohrmann GF, Vlak JM. Identification of genes involved in DNA replication of the Autographa californica baculovirus. *Proc Natl Acad Sci U S A.* 1994;91(23):11212-6.

Kost TA, Condreay JP, Jarvis DL. Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat Biotechnol.* 2005;23(5):567-75.

Lanier LM, Volkman LE. Actin binding and nucleation by Autographa californica M nucleopolyhedrovirus. *Virology.* 1998;243(1):167-77.

Lauzon HAM, Lucarotti CJ, Krell PJ, Feng Q, Retnakaran A, Arif BM. Sequence and Organization of the Neodiprion lecontei Nucleopolyhedrovirus Genome. *Journal of Virology.* 2004;78(13):7023-35.

Lauzon HAM. Genomic comparison of Neodiprion sertifer and Neodiprion lecontei nucleopolyhedroviruses and identification of potential hymenopteran baculovirus-specific open reading frames. *Journal of General Virology.* 2006;87(6):1477-89.

Leisy DJ, Rasmussen C, Kim HT, Rohrmann GF. The Autographa californica nuclear polyhedrosis virus homologous region 1a: identical sequences are essential for DNA replication activity and transcriptional enhancer function. *Virology.* 1995;208(2):742-52.

- Li X, Pang A, Lauzon HA, Sohi SS, Arif BM. The gene encoding the capsid protein P82 of the *Choristoneura fumiferana* multicapsid nucleopolyhedrovirus: sequencing, transcription and characterization by immunoblot analysis. *J Gen Virol*. 1997;78 (Pt 10):2665-73.
- Li Y, Wang J, Deng R, Zhang Q, Yang K, Wang X. vlf-1 deletion brought AcMNPV to defect in nucleocapsid formation. *Virus Genes*. 2005;31(3):275-84.
- Li Z, Blissard GW. The Pre-Transmembrane Domain of the *Autographa californica* Multicapsid Nucleopolyhedrovirus GP64 Protein Is Critical for Membrane Fusion and Virus Infectivity. *Journal of Virology*. 2009;83(21):10993-1004.
- Liu X, Chen K, Cai K, Yao Q. Determination of protein composition and host-derived proteins of *Bombyx mori* nucleopolyhedrovirus by 2-dimensional electrophoresis and mass spectrometry. *Intervirology*. 2008;51(5):369-76.
- Lu A, Carstens EB. Nucleotide sequence of a gene essential for viral DNA replication in the baculovirus *Autographa californica* nuclear polyhedrosis virus. *Virology*. 1991;181(1):336-47.
- Lu A, Carstens EB. Nucleotide sequence and transcriptional analysis of the p80 gene of *Autographa californica* nuclear polyhedrosis virus: a homologue of the *Orgyia pseudotsugata* nuclear polyhedrosis virus capsid-associated gene. *Virology*. 1992;190(1):201-9.
- Lu A, Miller LK. The roles of eighteen baculovirus late expression factor genes in transcription and DNA replication. *J Virol*. 1995;69(2):975-82.
- Lung O, Westenberg M, Vlak JM, Zuidema D, Blissard GW. Pseudotyping *Autographa californica* multicapsid nucleopolyhedrovirus (AcMNPV): F proteins from group II NPVs are functionally analogous to AcMNPV GP64. *J Virol*. 2002;76(11):5729-36.
- Lung OY, Cruz-Alvarez M, Blissard GW. Ac23, an Envelope Fusion Protein Homolog in the Baculovirus *Autographa californica* Multicapsid Nucleopolyhedrovirus, Is a Viral Pathogenicity Factor. *Journal of Virology*. 2003;77(1):328-39.
- Luque T, Finch R, Crook N, O'Reilly DR, Winstanley D. The complete sequence of the *Cydia pomonella* granulovirus genome. *J Gen Virol*. 2001;82(Pt 10):2531-47.
- Madhan S, Prabakaran, M., & Kwang, J. . Baculovirus as vaccine vectors. . *Current gene therapy*. 2010;10(3):201-13.
- Maeda S. Expression of foreign genes in insects using baculovirus vectors. *Annu Rev Entomol*. 1989;34:351-72.
- Maguire T, Harrison P, Hyink O, Kalmakoff J, Ward VK. The inhibitors of apoptosis of *Epiphyas postvittana* nucleopolyhedrovirus. *J Gen Virol*. 2000;81(Pt 11):2803-11.
- Marek M, Merten OW, Galibert L, Vlak JM, van Oers MM. Baculovirus VP80 protein and the F-actin cytoskeleton interact and connect the viral replication factory with the nuclear periphery. *J Virol*. 2011;85(11):5350-62.

- Marek M, Romier C, Galibert L, Merten OW, van Oers MM. Baculovirus VP1054 Is an Acquired Cellular PURLpha, a Nucleic Acid-Binding Protein Specific for GGN Repeats. *J Virol*. 2013;87(15):8465-80.
- Markovic I, Pulyaeva H, Sokoloff A, Chernomordik LV. Membrane fusion mediated by baculovirus gp64 involves assembly of stable gp64 trimers into multiprotein aggregates. *J Cell Biol*. 1998;143(5):1155-66.
- Maruniak JE. Molecular biology of *Anticarsia gemmatilis* baculovirus. *Memorias do Instituto Oswaldo Cruz*. 1989(84):107-11.
- Maruniak JE, Garcia-Maruniak A, Souza ML, Zanutto PM, Moscardi F. Physical maps and virulence of *Anticarsia gemmatilis* nucleopolyhedrovirus genomic variants. *Arch Virol*. 1999;144(10):1991-2006.
- Maxwell KL, Frappier L. Viral Proteomics. *Microbiology and Molecular Biology Reviews*. 2007;71(2):398-411.
- Mayer MP, Bukau B. Hsp70 chaperones: cellular functions and molecular mechanism. *Cell Mol Life Sci*. 2005;62(6):670-84.
- McCarthy CB, Dai X, Donly C, Theilmann DA. *Autographa californica* multiple nucleopolyhedrovirus ac142, a core gene that is essential for BV production and ODV envelopment. *Virology*. 2008;372(2):325-39.
- McClintock JT, Dougherty EM, Weiner RM. Semipermissive Replication of a Nuclear Polyhedrosis Virus of *Autographa californica* in a Gypsy Moth Cell Line. *J Virol*. 1986;57(1):197-204.
- McIntosh AH, Ignoffo CM, Andrews PL. In vitro host range of five baculoviruses in lepidopteran cell lines. *Intervirology*. 1985;23(3):150-6.
- Miele SAB, Garavaglia MJ, Belaich MN, Ghiringhelli PD. Baculovirus: Molecular Insights on Their Diversity and Conservation. *International Journal of Evolutionary Biology*. 2011;2011:1-15.
- Mikhailov VS, Okano K, Rohrmann GF. Baculovirus alkaline nuclease possesses a 5'→3' exonuclease activity and associates with the DNA-binding protein LEF-3. *J Virol*. 2003;77(4):2436-44.
- Mikhailov VS, Okano K, Rohrmann GF. Specificity of the endonuclease activity of the baculovirus alkaline nuclease for single-stranded DNA. *J Biol Chem*. 2004;279(15):14734-45.
- Miller LK, Dawes KP. Restriction endonuclease analysis to distinguish two closely related nuclear polyhedrosis viruses: *Autographa californica* MNPV and *Trichoplusia ni* MNPV. *Appl Environ Microbiol*. 1978;35(6):1206-10.
- Miller LK. Insect viruses. In: Fields BN, Knipe DM, Howley PM, Chanock RM, editors. *Fields virology*. Philadelphia: Lippincott-Raven; 1996. p. 533-56.

- Mishra G, Chadha P, Das RH. Serine/threonine kinase (pk-1) is a component of *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV) very late gene transcription complex and it phosphorylates a 102 kDa polypeptide of the complex. *Virus Res.* 2008;137(1):147-9.
- Mistretta TA, Guarino LA. Transcriptional activity of baculovirus very late factor 1. *J Virol.* 2005;79(3):1958-60.
- Monsma SA, Oomens AG, Blissard GW. The GP64 envelope fusion protein is an essential baculovirus protein required for cell-to-cell transmission of infection. *J Virol.* 1996;70(7):4607-16.
- Morris TD, Miller LK. Promoter influence on baculovirus-mediated gene expression in permissive and nonpermissive insect cell lines. *J Virol.* 1992;66(12):7397-405.
- Morse MA, Marriott AC, Nuttall PA. The glycoprotein of Thogoto virus (a tick-borne orthomyxo-like virus) is related to the baculovirus glycoprotein GP64. *Virology.* 1992;186(2):640-6.
- Moscardi F. Assessment of the application of baculoviruses for control of Lepidoptera. *Annu Rev Entomol.* 1999;44:257-89.
- Moscardi F, Lobo de Souza M, M.E.B. C, Szewczyk B. Baculovirus Pesticides: Present State and Future Perspectives. In: Ahmad I, Ahmad F, Pichtel J, editors. *Microbes and Microbial Technology.* New York: Springer Science; 2011. p. 415-46.
- Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem.* 2003;75(17):4646-58.
- O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem.* 1975;250(10):4007-21.
- O'Reilly DR, Miller LK, Luckow VA. *Baculovirus Expression Vectors: A Laboratory Manual.* Oxford University Press U, editor. New York 1993.
- Ohkawa T, Volkman LE, Welch MD. Actin-based motility drives baculovirus transit to the nucleus and cell surface. *The Journal of cell biology.* 2010;190(2):187-95.
- Okano K, Vanarsdall AL, Mikhailov VS, Rohrmann GF. Conserved molecular systems of the Baculoviridae. *Virology.* 2006;344(1):77-87.
- Oliveira JVC, Wolff JL, Garcia-Maruniak A, Ribeiro BM, de Castro ME, de Souza ML, et al. Genome of the most widely used viral biopesticide: *Anticarsia gemmatilis* multiple nucleopolyhedrovirus. *Journal of General Virology.* 2006a;87(Pt 11):3233-50.
- Oliveira JVC, Wolff JLC, Garcia-Maruniak A, Ribeiro BM, de Castro MEB, de Souza ML, et al. Genome of the most widely used viral biopesticide: *Anticarsia gemmatilis* multiple nucleopolyhedrovirus. *Journal of General Virology.* 2006b;87(11):3233-50.
- Oliveira JVC, Brito A, Braconi CT, Freire CCM, Iamarino A, Zanotto PM. Modularity and evolutionary constraints in a baculovirus gene regulatory network *BMC Systems Biology.* [article]. 2013;[em publicação].

- Olszewski J, Miller LK. A role for baculovirus GP41 in budded virus production. *Virology*. 1997a;233(2):292-301.
- Olszewski J, Miller LK. Identification and characterization of a baculovirus structural protein, VP1054, required for nucleocapsid formation. *J Virol*. 1997b;71(7):5040-50.
- Oomens AG, Blissard GW. Requirement for GP64 to drive efficient budding of *Autographa californica* multicapsid nucleopolyhedrovirus. *Virology*. 1999;254(2):297-314.
- Pan L, Li Z, Gong Y, Yu M, Yang K, Pang Y. Characterization of gp41 gene of *Spodoptera litura* multicapsid nucleopolyhedrovirus. *Virus Res*. 2005;110(1-2):73-9.
- Pearson MN, Groten C, Rohrmann GF. Identification of the *lymantria dispar* nucleopolyhedrovirus envelope fusion protein provides evidence for a phylogenetic division of the *Baculoviridae*. *J Virol*. 2000;74(13):6126-31.
- Peng K, van Lent JW, Boeren S, Fang M, Theilmann DA, Erlandson MA, et al. Characterization of novel components of the baculovirus per os infectivity factor complex. *J Virol*. 2012;86(9):4981-8.
- Perera O, Green TB, Stevens SM, White S, Becnel JJ. Proteins Associated with *Culex nigripalpus* Nucleopolyhedrovirus Occluded Virions. *Journal of Virology*. 2007;81(9):4585-90.
- Pond SLK, Muse SV. HyPhy: hypothesis testing using phylogenies. *Statistical methods in molecular evolution*: Springer; 2005. p. 125-81.
- Prikhod'ko GG, Wang Y, Freulich E, Prives C, Miller LK. Baculovirus p33 binds human p53 and enhances p53-mediated apoptosis. *Journal of Virology*. 1999;73(2):1227-34.
- Radhakrishnan A, Yeo D, Brown G, Myaing MZ, Iyer LR, Fleck R, et al. Protein analysis of purified respiratory syncytial virus particles reveals an important role for heat shock protein 90 in virus particle assembly. *Mol Cell Proteomics*. 2010;9(9):1829-48.
- Rankin C, Ooi BG, Miller LK. Eight base pairs encompassing the transcriptional start point are the major determinant for baculovirus polyhedrin gene expression. *Gene*. 1988;70(1):39-49.
- Rebollar E, Valadez-Graham V, Vazquez M, Reynaud E, Zurita M. Role of the p53 homologue from *Drosophila melanogaster* in the maintenance of histone H3 acetylation and response to UV-light irradiation. *FEBS Lett*. 2006;580(2):642-8.
- Reed LJ, Muench H. A simple method of estimating fifty percent endpoints. *American Journal of Epidemiology*. 1938;27((3)):493-7.
- Reilly LM, Guarino LA. The pk-1 gene of *Autographa californica* multinucleocapsid nuclear polyhedrosis virus encodes a protein kinase. *J Gen Virol*. 1994;75 (Pt 11):2999-3006.
- Reilly LM, Guarino LA. The viral ubiquitin gene of *Autographa californica* nuclear polyhedrosis virus is not essential for viral replication. *Virology*. 1996;218(1):243-7.

- Ribeiro BM, Gatti CD, Costa MH, Moscardi F, Maruniak JE, Possee RD, et al. Construction of a recombinant *Anticarsia gemmatalis* nucleopolyhedrovirus (AgMNPV-2D) harbouring the beta-galactosidase gene. *Arch Virol*. 2001;146(7):1355-67.
- Rice WC, Miller LK. Baculovirus transcription in the presence of inhibitors and in nonpermissive *Drosophila* cells. *Virus Res*. 1986;6(2):155-72.
- Roche S, Bressanelli S, Rey FA, Gaudin Y. Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G. *Science*. 2006;313(5784):187-91.
- Rohrmann G. Structural proteins of baculovirus occlusion bodies and virions. In: GF R, editor. *Baculovirus Molecular Biology* 2011. p. 1-25.
- Rohrmann GF. *Baculovirus Molecular Biology*. National Library of Medicine (US), National Center for Biotechnology Information, 2008. Disponível em: <<http://www.ncbi.nlm.nih.gov/books/NBK1736/> 11/16/2011].
- Russell RL, Rohrmann GF. A 25-kDa protein is associated with the envelopes of occluded baculovirus virions. *Virology*. 1993;195(2):532-40.
- Scheele GA. Two-dimensional gel analysis of soluble proteins. Characterization of guinea pig exocrine pancreatic proteins. *J Biol Chem*. 1975;250(14):5375-85.
- Schultz KL, Friesen PD. Baculovirus DNA replication-specific expression factors trigger apoptosis and shutoff of host protein synthesis during infection. *J Virol*. 2009;83(21):11123-32.
- Schultz KL, Wetter JA, Fiore DC, Friesen PD. Transactivator IE1 is required for baculovirus early replication events that trigger apoptosis in permissive and nonpermissive cells. *J Virol*. 2009;83(1):262-72.
- Sieburth P, Maruniak JE. Growth characteristics of a continuous cell line from the velvetbean caterpillar, *Anticarsia gemmatalis* hübner (Lepidoptera: Noctuidae). *In Vitro Cell Dev Biol Plant*. 1988(24):195-8.
- Simon O, Palma L, Williams T, Lopez-Ferber M, Caballero P. Analysis of a naturally-occurring deletion mutant of *Spodoptera frugiperda* multiple nucleopolyhedrovirus reveals sf58 as a new per os infectivity factor of lepidopteran-infecting baculoviruses. *J Invertebr Pathol*. 2012;109(1):117-26.
- Slack J, Arif BM. The baculoviruses occlusion-derived virus: virion structure and function. *Advances in virus research*. 2007;69:99-165.
- Smith GE, Summers MD, Fraser MJ. Production of human beta interferon in insect cells infected with a baculovirus expression vector. *Mol Cell Biol*. 1983;3(12):2156-65.
- Smith GE, Ju G, Ericson BL, Moschera J, Lahm HW, Chizzonite R, et al. Modification and secretion of human interleukin 2 produced in insect cells by a baculovirus expression vector. *Proc Natl Acad Sci U S A*. 1985;82(24):8404-8.

- Smulson M, Istock N, Ding R, Cherney B. Deletion mutants of poly(ADP-ribose) polymerase support a model of cyclic association and dissociation of enzyme from DNA ends during DNA repair. *Biochemistry*. 1994;33(20):6186-91.
- Soares JS, Ribeiro BM. Pathology of *Anticarsia gemmatilis* larvae infected by two recombinant *A. gemmatilis* multicapsid nucleopolyhedroviruses. *Research in microbiology*. 2005;156(2):263-9.
- Steen H, Mann M. The abc's (and xyz's) of peptide sequencing. *Nature Reviews Molecular Cell Biology*. 2004;5(9):699-711.
- Summers MD, Smith GE. A manual of methods for baculovirus vectors and insect cell culture procedures. *Texas Agric Exper Sta Bull*. 1987(1555):1-57.
- Sun XL, Peng HY. Recent advances in biological control of pest insects by using viruses in China. *Virologica Sinica* 2007. p. 158-62.
- Szewczyk B, Hoyoscarvajal L, Paluszek M, Skrzecz I, Lobodesouza M. Baculoviruses — re-emerging biopesticides. *Biotechnology Advances*. 2006;24(2):143-60.
- Szewczyk B, Rabalski, L., Krol, E., Sihler, W., & Souza, M. D. . Baculovirus biopesticides—a safe alternative to chemical protection of plants. *Journal of Biopesticides*. 2009;2(2):209-16.
- Tanada Y, Kaya HK. *Insect Pathology*. Press SDA, editor. San Diego 1993.
- Tang XD, Xu YP, Yu LL, Lang GJ, Tian CH, Zhao JF, et al. Characterization of a *Bombyx mori* nucleopolyhedrovirus with Bmvp80 disruption. *Virus Res*. 2008;138(1-2):81-8.
- Tulin A, Naumova NM, Menon AK, Spradling AC. *Drosophila* poly(ADP-ribose) glycohydrolase mediates chromatin structure and SIR2-dependent silencing. *Genetics*. 2006;172(1):363-71.
- Tyers M, Mann M. From genomics to proteomics. *Nature*. 2003;422(6928):193-7.
- Van Oers MM, Vlask JM. The baculovirus 10-kDa protein. *J Invertebr Pathol*. 1997;70(1):1-17.
- van Oers MM, Vlask JM. Baculovirus genomics. *Curr Drug Targets*. 2007;8(10):1051-68.
- Vanarsdall AL, Okano K, Rohrmann GF. Characterization of the role of very late expression factor 1 in baculovirus capsid structure and DNA processing. *J Virol*. 2006;80(4):1724-33.
- Vanarsdall AL, Pearson MN, Rohrmann GF. Characterization of baculovirus constructs lacking either the Ac 101, Ac 142, or the Ac 144 open reading frame. *Virology*. 2007;367(1):187-95.
- Vaughn JL, Goodwin RH, Tompkins GJ, McCawley P. The establishment of two cell lines from the insect *Spodoptera frugiperda* (Lepidoptera; Noctuidae). *In vitro*. 1977;13(4):213-7.
- Vaux DL, Silke J. IAPs, RINGs and ubiquitylation. *Nat Rev Mol Cell Biol*. 2005;6(4):287-97.

- Volkman LE, Goldsmith PA. In vitro survey of *Autographa californica* nuclear polyhedrosis virus interaction with nontarget vertebrate host cells. *Applied and environmental microbiology*. 1983;45(3):1085-93.
- Volkman LE, Goldsmith PA. Budded *Autographa californica* NPV 64K protein: Further biochemical analysis and effects of postimmunoprecipitation sample preparation conditions. *Virology*. 1984;139(2):295-302.
- Volkman LE. *Autographa californica* MNPV nucleocapsid assembly: inhibition by cytochalasin D. *Virology*. 1988;163(2):547-53.
- Volkman LE. Baculovirus infectivity and the actin cytoskeleton. *Curr Drug Targets*. 2007;8(10):1075-83.
- Wang P, Granados RR. An intestinal mucin is the target substrate for a baculovirus enhancin. *Proc Natl Acad Sci U S A*. 1997;94(13):6977-82.
- Wang R, Deng F, Hou D, Zhao Y, Guo L, Wang H, et al. Proteomics of the *Autographa californica* Nucleopolyhedrovirus Budded Virions. *Journal of Virology*. 2010;84(14):7233-42.
- Wang S, Balasundaram G. Potential cancer gene therapy by baculoviral transduction. *Curr Gene Ther*. 2010;10(3):214-25.
- Wang XF, Zhang BQ, Xu HJ, Cui YJ, Xu YP, Zhang MJ, et al. ODV-associated proteins of the *Pieris rapae* granulovirus. *J Proteome Res*. 2011;10(6):2817-27.
- Wang Y, Wu W, Li Z, Yuan M, Feng G, Yu Q, et al. ac18 is not essential for the propagation of *Autographa californica* multiple nucleopolyhedrovirus. *Virology*. 2007;367(1):71-81.
- Welchman RL, Gordon C, Mayer RJ. Ubiquitin and ubiquitin-like proteins as multifunctional signals. *Nat Rev Mol Cell Biol*. 2005;6(8):599-609.
- Westenberg M, Uijtdewilligen P, Vlak JM. Baculovirus envelope fusion proteins F and GP64 exploit distinct receptors to gain entry into cultured insect cells. *J Gen Virol*. 2007;88(Pt 12):3302-6.
- Whitt M. A phosphorylated 34-kDa protein and a subpopulation of polyhedrin are thiol linked to the carbohydrate layer surrounding a baculovirus occlusion body. *Virology*. 1988;163(1):33-42.
- Wickham TJ, Shuler ML, Hammer DA, Granados RR, Wood HA. Equilibrium and kinetic analysis of *Autographa californica* nuclear polyhedrosis virus attachment to different insect cell lines. *J Gen Virol*. 1992;73 (Pt 12):3185-94.
- Wiese S, Reidegeld KA, Meyer HE, Warscheid B. Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics*. 2007;7(3):340-50.
- Wilson ME, Mainprize TH, Friesen PD, Miller LK. Location, transcription, and sequence of a baculovirus gene encoding a small arginine-rich polypeptide. *J Virol*. 1987;61(3):661-6.
- Wolff JLC, Valicente FH, Martins R, Oliveira JVD, Zanotto PMdA. Analysis of the genome of *Spodoptera frugiperda* nucleopolyhedrovirus (SfMNPV-19) and of the high genomic

heterogeneity in group II nucleopolyhedroviruses. *Journal of General Virology*. 2008;89(5):1202-11.

Wolgast GM, Gross CH, Russell RL, Rohrmann GF. Immunocytochemical characterization of p24, a baculovirus capsid-associated protein. *J Gen Virol*. 1993;74 (Pt 1):103-7.

Wu W, Lin T, Pan L, Yu M, Li Z, Pang Y, et al. *Autographa californica* multiple nucleopolyhedrovirus nucleocapsid assembly is interrupted upon deletion of the 38K gene. *J Virol*. 2006;80(23):11475-85.

Wu W, Liang H, Kan J, Liu C, Yuan M, Liang C, et al. *Autographa californica* multiple nucleopolyhedrovirus 38K is a novel nucleocapsid protein that interacts with VP1054, VP39, VP80, and itself. *J Virol*. 2008;82(24):12356-64.

Wu W, Passarelli AL. *Autographa californica* multiple nucleopolyhedrovirus Ac92 (ORF92, P33) is required for budded virus production and multiply enveloped occlusion-derived virus formation. *J Virol*. 2010;84(23):12351-61.

Wu Y, Liu G, Carstens EB. Replication, integration, and packaging of plasmid DNA following cotransfection with baculovirus viral DNA. *J Virol*. 1999;73(7):5473-80.

Xi Q, Wang J, Deng R, Wang X. Characterization of AcMNPV with a deletion of me53 gene. *Virus Genes*. 2007;34(2):223-32.

Xu F, Ince IA, Boeren S, Vlak JM, van Oers MM. Protein composition of the occlusion derived virus of *Chrysodeixis chalcites* nucleopolyhedrovirus. *Virus Res*. 2011;158(1-2):1-7.

Yamagishi J, Burnett ED, Harwood SH, Blissard GW. The AcMNPV pp31 gene is not essential for productive AcMNPV replication or late gene transcription but appears to increase levels of most viral transcripts. *Virology*. 2007;365(1):34-47.

Yang MM, Li ML, Zhang Y, Wang YZ, Qu LJ, Wang QH, et al. Baculoviruses and insect pests control in China. *African Journal of Microbiology Research*. 2012;6(2):214-8.

Yang S, Miller LK. Activation of baculovirus very late promoters by interaction with very late factor 1. *J Virol*. 1999;73(4):3404-9.

Yuan M, Wu W, Liu C, Wang Y, Hu Z, Yang K, et al. A highly conserved baculovirus gene p48 (ac103) is essential for BV production and ODV envelopment. *Virology*. 2008;379(1):87-96.

Zanotto PM, Sampaio MJ, Johnson DW, Rocha TL, Maruniak JE. The *Anticarsia gemmatilis* nuclear polyhedrosis virus polyhedrin gene region: sequence analysis, gene product and structural comparisons. *J Gen Virol*. 1992;73 (Pt 5):1049-56.

Zanotto PM, Kessing BD, Maruniak JE. Phylogenetic interrelationships among baculoviruses: evolutionary rates and host associations. *J Invertebr Pathol*. 1993;62(2):147-64.

Zanotto PMA, Krakauer DC. Complete genome viral phylogenies suggests the concerted evolution of regulatory cores and accessory satellites. *PLoS ONE*. 2008;3(10):e3500.

Zhang L, Jia X, Zhang X, Sun J, Peng X, Qi T, et al. Proteomic analysis of PBMCs: characterization of potential HIV-associated proteins. *Proteome Science*. 2010a;8(1):12.

Zhang L, Jia X, Zhang X, Sun J, Peng X, Qi T, et al. Proteomic analysis of PBMCs: characterization of potential HIV-associated proteins. *Proteome Sci*. 2010b;8:12.

Zhou J, Blissard GW. Identification of a GP64 Subdomain Involved in Receptor Binding by Budded Virions of the Baculovirus *Autographica californica* Multicapsid Nucleopolyhedrovirus. *Journal of Virology*. 2008;82(9):4449-60.

Zwickl DJ. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion [Ph.D. thesis]. Austin: The University of Texas; 2006.

APÊNDICES

Apêndice A - Artigos Submetidos

Cell modulation factors are carried along structural proteins in the baculovirus AgMNPV-2D budded and occluded virions

Journal:	PROTEOMICS
Manuscript ID:	pmic.201300312
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	24-Jul-2013
Complete List of Authors:	Braconi, Carla T.; Biomedical Sciences Institute - University of São Paulo, Department of Microbiology Ardisson-Araújo, Daniel; University of Brasília, Cell Biology Department Leme, Adriana; Brazilian Biosciences National Laboratory - LNBio- CNPEM, Laboratory of Mass Spectrometry Oliveira, Juliana; Biomedical Sciences Institute - University of São Paulo, Department of Microbiology Pauletti, Bianca; Brazilian Biosciences National Laboratory - LNBio- CNPEM, Laboratory of Mass Spectrometry Garcia-Maruniak, Alejandra; University of Florida, Entomology and Nematology Department Ribeiro, Bergmann; University of Brasília, Cell Biology Department Maruniak, James; University of Florida, Entomology and Nematology Department Zanotto, Paolo; Biomedical Sciences Institute - University of São Paulo, Department of Microbiology
Keywords (free text entry):	baculovirus, AgMNPV-2D , proteomics, occlusion-derived virus (ODV), Budded Virus (BV)

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

**Cell modulation factors are carried along structural proteins in the
baculovirus AgMNPV-2D budded and occluded virions**

Carla Torres Braconi¹, Daniel Mendes Pereira Ardisson-Araújo², Adriana
Franco Paes Leme³, Juliana Velasco de Castro Oliveira¹, Bianca Alves Pauletti³,
Alejandra Garcia-Maruniak⁴, Bergmann Morais Ribeiro², James E. Maruniak⁴ &
Paolo Marinho de Andrade Zanotto^{1§}

¹ Laboratory of Molecular Evolution and Bioinformatics, Department of Microbiology,
Biomedical Sciences Institute - ICB II, University of São Paulo - USP, Brazil
² Cell Biology Department, University of Brasília, Brasília, DF, CEP 70910-970, Brazil
³ Laboratory of Mass Spectrometry, Brazilian Biosciences National Laboratory - LNBio-
CNPEM, Campinas - SP, Brazil
⁴ Insect Viruses Laboratory, Entomology and Nematology Department, University of Florida,
Gainesville, FL, USA

§Corresponding author: Laboratory of Molecular Evolution and Bioinformatics, Department
of Microbiology, Institute of Biomedical Sciences – ICB II, University of São Paulo – USP,
Av. Prof. Lineu Prestes, 1374, São Paulo, SP, Brasil, CEP: 05508-900.
Phone: +55 11 3091-8453
FAX: +55 11 3091-7354

e-mail address:
P.M.A.Z: *pzanotto@usp.br*

Running title: AgMNPV-2D proteome
Categories: Proteomics
Keywords: AgMNPV-2D / baculovirus / ODV / BV/ proteomics

Abstract

Baculoviruses infect mainly insects, producing two distinct phenotypes during the viral life cycle: the budded virus (BV) and, the occlusion-derived virus (ODV) for intra and host spread, respectively. Since the 80's several countries have been using the multiple nucleopolyhedrovirus *Anticarsia gemmatalis* (AgMNPV) as a biological control agent against the velvet bean caterpillar (*A. gemmatalis*). The genome of the AgMNPV-2D carries at least 152 potential genes, with twenty-four that possibly code for structural proteins. Proteomic studies were done on a few baculoviruses, with six ODV and two BV proteomes finished so far. Moreover, there are limited data on virion proteins carried by the AgMNPV-2D. Herein, the structural proteins of AgMNPV-2D were analyzed by MALDI-Q-TOF and LC MS/MS. A total of 44 proteins were associated to the ODV and 33 to the BV of AgMNPV-2D. Several proteins have been described as structural proteins already, though our data showed six new proteins in the AgMNPV-2D ODV and seven new proteins carried by the BV. Eleven cellular proteins were also identified, which are possibly assorted during viral morphogenesis and carried with the virion. These findings may provide novel insights into AgMNPV biology and its host interaction.

Introduction

Two distinct phenotypes occur during the life cycle of baculoviruses: the budded virus (BV) responsible for the intra-host systemic infection, and the occlusion-derived virus (ODV), that is essential for inter-host, oral primary infection in the environment [1]. In the end of the viral cycle, the ODV form is occluded in a paracrystalline protein matrix that forms the occlusion bodies (OB) [2-4]. *Baculoviridae* are arthropod-specific viruses and, due to their high specificity to insects, they became widely used as biological control agents of pests in agriculture and forestry as an alternative to harmful pesticides [5-8]. Baculoviruses have a double-stranded DNA immersed in a rod-shaped nucleocapsid, coding for 100 to 180 proteins [9, 10]. The *Baculoviridae* are divided into two groups based on the OBs morphologically: nucleopolyhedrovirus (NPV) and granulovirus (GV). The lepidopteran NPVs are subdivided into group I and II, following phylogenetic studies done with the polyhedrin gene (*polh*) [11], a larger set of conserved genes [9, 12] or complete genomes [13] and, also the requirement of distinct fusion proteins to enter in host cells [14, 15]. Moreover, there are additional 11 genes that appear to be exclusive of NPV [16, 17]. Baculovirus are grouped in four genera following the

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

63 insectan order they belong to: *Alphabaculovirus* (lepidopteran nucleopolyhedrovirus),
64 *Betabaculovirus* (leptodopteran granulovirus), *Gammabaculovirus* (hymenopteran
65 nucleopolyhedrovirus) and *Deltabaculovirus* (diptera nucleopolyhedrovirus) [18].

66 Proteomic tools and mass spectrometry (MS) have been essential for studying
67 proteins encoded by or associated with viruses in general and baculoviruses in
68 particular. The first study that applied this methodology identified forty-four
69 structural proteins on the ODV phenotype of *Autographa californica multiple*
70 *nucleopolyhedrovirus* (AcMNPV) [19]. Likewise, the protein complement of the
71 ODV from several other baculoviruses were also studied: *Culex nigripalpus*
72 *nucleopolyhedrovirus* (CuniNPV) [20], *Helicoperva armigera nucleopolyhedrovirus*
73 (HearNPV) [21, 22], *Bombyx mori nucleopolyhedrovirus* (BmNPV) [23],
74 *Chrysodeixis chalcites nucleopolyhedrovirus* (ChChNPV) [24] and the *Pieris rapae*
75 Granulovirus (PrGV) [25]. Thus far, the proteins in the BV phenotype of both
76 AcMNPV and HearNPV were identified [22, 26]. However, there are limited protein
77 data on the ODV and the BV of the *Anticarsia gemmatalis multiple*
78 *nucleopolyhedrovirus* (AgMNPV), an *Alphabaculovirus*.

79 Many Latin American countries use the AgMNPV as a biological control agent
80 against the velvet bean caterpillar, *Anticarsia gammatalis* (Lepidoptera: Noctuidae),
81 which is an important pest of soybean crops [7, 27]. The AgMNPV isolate 2D
82 (AgMNPV-2D) was chosen as prototype for basic studies [28-31], since it was the
83 most frequent genotype originally isolated in a sampled from a wild population in
84 Peru in 1972 [32]. The complete genome sequence (132,239 bp) of AgMNPV-2D
85 revealed 152 open reading frames (ORFs), with twenty-four ORFs that possibly
86 encode for structural proteins [13]. Lately, the gene expression profile of AgMNPV-
87 2D was determined in two different insect cell lines, showing that at least 149 ORFs
88 are transcribed during the viral replication cycle [33]. However, the analysis of gene
89 transcription or translation does not mean that all mRNAs produce functional proteins
90 [34]. Crucially, proteins are regulated by post-translation modifications in order to
91 allow for virus-cell interactions [35, 36]. In the present study, structural and
92 associated proteins in the ODV and BV of the AgMNPV-2D were analyzed by two
93 complementary mass spectrometry techniques. Peptides were identified by
94 correlation of tandem mass spectra against the coding capacity of the AgMNPV-2D

genome. A total of 44 proteins were identified in the ODV and 33 in the BV of the AgMNPV-2D. This data provides insights into the AgMNPV system, leading to a more detailed understanding of viral multiplication, morphogenesis and possible associations to cell-encoded proteins.

Materials and Methods

Cell culture, virus infection, ODV and BV purification and

quantification. The UFL-AG-286 cells [37] used for growing AgMNPV-2D were maintained at 28°C in Grace's medium (Invitrogen) supplemented with 10% (v/v) Fetal Bovine Serum (Invitrogen). Cells were infected with AgMNPV-2D at a multiplicity of infection (MOI) of 1. After 48 hours post infection (h.p.i), cell debris were removed by centrifugation and the supernatant with BV was centrifuged on a first round on 3 ml of 25% (wt/vol) sucrose cushion in 5 mM of NaCl and 10 mM EDTA. The virion pellet was resuspended in 0.1xTE and then purified in a continuous (25% to 60%) sucrose gradient and the band containing BV was diluted in 0.1xTE and then centrifuged at 80,000 g for 75 min. ODVs were collected at 72 hours post infection (h.p.i) and cell debris were also removed by centrifugation. Pellets with the occlusion bodies (ODV) were purified from infected cells by first using a 0.1 % SDS. Subsequently, the ODVs were released by alkaline treatment (pH=10,9) and separated by a second round of purification, the ODV were purified by a continuous (40% - 65%) sucrose gradient centrifugation and white bands containing ODVs were detected and collected. Both pellets within BV and ODV virions were resuspended in 0.1xTE with protease inhibitor (GE Healthcare (formerly Amersham Bioscience), USA) to prevent the possible degradation of virion components.

Bidimensional gels (2-DE), SDS-PAGE and In-Gel digestion. The 2-DE gels were performed with purified ODVs using the IPGphor system (GE Healthcare), USA). Briefly, 500 µg proteins were precipitated with cold acetone and pellets were solubilized in a 690 µl of rehydration solution containing 7M Urea, 2M thiourea, 2% Chaps, 1% DTT, 0.5% Buffer IPG, 0.002% bromophenol blue followed by a

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

hydration step of 3 IPG dry strips (pH 3-10 L (180 × 30 × 0.5 mm) overnight. Isoelectric focusing (IEF) was conducted in the IPGphor system GE (formerly Amersham Bioscience, USA). Focusing was programmed in a gradient mode with 4 steps and a total of 17.250 KVhr was accumulated. For the second dimension, each IPG strip was equilibrated with buffer and then, strips were transferred onto 12% uniform SDS-PAGE. All 2-DE gels were done in triplicate, digitalized with ImageScanner III (formerly Amersham Bioscience) and spots analyses were normalized by Imagemaster 2D Platinum software version 7.0 (GE (formerly Amersham Bioscience). Alternatively, purified ODV and BV were separated by 12% SDS- PAGE, fixed and stained with Comassie Brilliant Blue R-350 (GE (formerly Amersham Bioscience), USA). Spots and bands from SDS-PAGE gels were manually excised and then destained, reduced, alkylated and *in-gel* digested with trypsin modified sequencing grade (Promega, Madison, WI) performed according to [38].

Protein identification by Mass spectrometry. We used two complementary mass spectrometry techniques for the tryptic *in-gel* digested peptides. First, the matrix-assisted laser desorption/ionization time of flight (MALDI-Q-TOF) Premier mass spectrometer (Waters) for ODV spots enzymatically-digested from 2-DE gels triplicate. For MALDI Q-TOF analyses, each ODV tryptic peptide mixture was spotted onto the sample plate, mixed, dried, and analyzed in delayed extraction and reflectron modes using polyethylene glycol oligomers as external calibrant. Second, the tryptic peptide mixture obtained from the ODV and BV 12 % SDS-PAGE were analyzed by electrospray ionization (ESI) Waters® nanoACQUITY UPLC® System (Waters, Milford, MA) and a hybrid quadrupole-TOF mass spectrometer QSTAR Elite (Applied Biosystems, Framingham, MA). For Waters® nanoACQUITY UPLC® System, the peptides (4.5 µl) were separated using a C18 column (100 µm x 100 mm) RP-nanoUPLC (nanoAcquity, Waters) coupled with a Q-Tof Ultima mass spectrometer (Waters) with nanoelectrospray source, at a flow rate of 0.6 µl/min. The gradient was 2-90% acetonitrile in 0.1% formic acid over 45 min for each digested band from SDS-PAGE. The nanoelectrospray voltage was set to 3.5 kV with a cone

157 voltage of 30 V. The instrument was operated in the ‘top three’ mode in a data-
158 dependent analysis, in which one MS spectrum is acquired followed by MS/MS of the
159 top three most-intense peaks detected. For the QSTAR elite hybrid quadrupole-TOF
160 mass spectrometer, some of the enzymatically-digested samples from ODV and all
161 tryptic peptide mixture from BV were injected into a capillary trap (LC Packings
162 PepMap) and desalted for 5 min at a flow rate 3 μ l/min of 0.1% v/v acetic acid.
163 Samples were loaded onto an LC Packing® C18 Pep Map nanoflow HPLC column.
164 The elution gradient of the HPLC column started at 3% solvent B, 97% solvent A and
165 finished at 60% solvent A, 40% solvent B for 30 min for protein identification.
166 Solvent A consisted of 0.1% v/v acetic acid, 3% v/v ACN, and 96.9% v/v H₂O.
167 Solvent B consisted of 0.1% v/v acetic acid, 96.9% v/v ACN, and 3% v/v H₂O. LC-
168 MS/MS analyses were carried out on a hybrid quadrupole-TOF mass spectrometer
169 (QSTAR elite, Applied Biosystems, Framingham, MA). The focusing potential and
170 ion spray voltages were set to 225 V and 2400 V, respectively. The information-
171 dependent acquisition (IDA) mode of operation was employed in which a survey scan
172 from m/z 400–1800 was acquired followed by collision-induced dissociation (CID) of
173 the four most intense ions. Survey and MS/MS spectra for each IDA cycle were
174 accumulated for 1 and 3 s, respectively. All peptides were analyzed at the Mass
175 Spectrometry Laboratory at Brazilian Biosciences National Laboratory, CNPEM-
176 ABTLuS, Campinas, SP, Brazil and at the proteomic facility of Interdisciplinary
177 Center for Biotechnology Research (ICBR) at University of Florida (UF), USA.

179 **Data analysis and Protein search algorithm.** The spectra were acquired using
180 software MassLynx v.4.1 and the raw data files were converted by the software
181 Mascot Distiller v.2.3.2.0, 2009 (Matrix Science Ltd.). All data were searched against
182 our AgMNPV-2D ORF database consisting of 152 open reading frames (ORF) coding
183 for 40.427 residues and, an non redundant insect protein database obtained from
184 NCBI databank (<http://www.ncbi.nlm.nih.gov/>) using the MASCOT 2.0 online
185 software (MatrixScience, Boston, MA, USA). Mascot software was set with a
186 fragment ion mass tolerance of 0.1 Da and a parent ion tolerance of 0.1 Da. Then,
187 iodoacetamide derivative of Cys, deamidation of Asn and Gln, oxidation of Met, were

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

188 specified in Mascot as variable modifications. During the analyses, our samples were
189 checked against a contaminant database (supplied by Applied Biosystems). The BV
190 proteins were further confirmed with Scaffold software (version Scaffold-03-00-04,
191 Proteome Software Inc., Portland, OR) to validate MS/MS based peptide and protein
192 identification. Peptide were validated if they had greater than 99.0% probability
193 estimated by the Peptide Prophet algorithm [39] and if they consisted of at least two
194 unique peptides [40].

196 **Antisera generation and Western blot analyses.** Our MS results were further
197 confirmed by immunoblotting. For that, we designed polyclonal antibodies against
198 PTP, LEF-1, ODV-E26 and PIF-3. Briefly, these four genes from AgMNPV-2D were
199 amplified by PCR and the Gateway® Cloning Technology (Life Technologies™) kit
200 was used for cloning the PCR products in *E. Coli* DH5-a according to manufacturer's
201 instruction. The resulting constructs were sequenced (Macrogen, Korea) and
202 confirmed to encode AgMNPV-2D proteins amino-terminal 6xHis tagged. *E.*
203 *coli* BL21-AI were then used to express the proteins after induction with 0.2 % of L-
204 arabinose for 4 h at 37°C. The bacteria were collected by centrifugation, and used for
205 protein purification according to [41]. Polyclonal antibodies were raised in rats
206 against the isolated proteins by standard methods. Western blot analyses were
207 individually performed with the four antisera (1:500 dilution) produced. Proteins were
208 heated (100 °C) for 5 min in protein loading buffer (0.25 M Tris-Cl, pH 6.8, 4% SDS,
209 20% glycerol, 10% 2-mercaptoethanol, and 0.02% bromophenol blue) and analyzed
210 by SDS-12% PAGE using the Mini Protean Tetra Cell apparatus (BioRad) following
211 the manufacturer's instructions. The immobilized proteins were transferred onto
212 nitrocellulose membranes (Sigma) using the Trans-Blot® SD – Semi Dry Transfer
213 Cell (BioRad). The membranes were then blocked in 1x PBS Buffer (137 mM NaCl,
214 2.7 mM KCl, 10 mM Na₂HPO₄, 2 mM KH₂PO₄, pH 7.4) containing 5% skimmed
215 milk powder for 16 h at 4 °C, washed three times with PBS tween (0.05%) and probed
216 with the desired antiserum followed by incubation with the alkaline phosphatase-
217 conjugated anti-mouse/rat secondary antibody (Promega). Blots were developed using
218 the NBT/BCIP (Sigma) substrate dissolved in alkaline phosphatase buffer.

Results

ODV-associated proteins from MS/MS analyses. The bidimensional (2-DE) triplicates of purified ODV revealed 85 well-resolved spots ranging from 7 to 142 kDa (Figure S1). Among these protein spots, only twenty-one proteins could be identified by MALDI-Q-TOF analyses with Mascot Distiller v.2.3.2.0, 2009 (Matrix Science Ltd.) (Fig. S1 and Table S1). All data were based on the comparison between amino acid sequences deduced from tandem mass spectra and the theoretical ORF database of AgMNPV-2D (Tab. S1). Two proteins were stained more intensely in the center of the gel (Fig. S1). Spots number 1 and 10 were identified as POLH (*ag1*) the CG30 (*ag85*), respectively (Tab. S1). Nineteen proteins were also identified to be in the structure of ODV: 41.5 kDa (*ag15*), EGT (*ag18*), 38 kDa (*ag20*), IAP-1 (*ag40*), FP25K (*ag61*), DNAPol (*ag65*), LEF-3 (*ag67*), GP41 (*ag79*), VP39 (*ag86*), ODV-E25 (*ag91*), P40 (*ag97*), PNK/PNL (*ag103*), PTP (*ag107*), P24 (*ag125*), Alk-Exo (*ag129*), P10 (*ag133*), P49 (*ag138*), ODV-E27 (*ag140*) and PK-1 (*ag151*) (Tab. S1). Furthermore, many different spots were identified as products of the same ORFs, such as, POLH, VP39, ODV-E25, P49 and ODV-E27. Although, OBs were treated with alkaline solution and purified twice in sucrose gradient, we still found a large amount of polyhedrin as contaminant. This was expected since this protein is the main component of the matrix of OB, where ODVs are embedded [42, 43]. Three proteins were determined using only 2-DE and MALDI-Q-TOF methodologies: DNAPol (*ag65*), CG30 (*ag85*) and Alk-Exo (*ag129*). Unfortunately, more than half of the spots showed low matching values for sequence coverage or low score against the theoretical ORF database of AgMNPV-2D using only this MS method. Additionally, thirty-one tryptic peptides from ODV SDS-PAGE gel (Figure S2A) were further separated by high-performance liquid chromatography coupled on a hybrid quadrupole-TOF mass spectrometer. The MS/MS peaks were analyzed by Mascot Distiller v.2.3.2.0, 2009 (Matrix Science Ltd.) and checked against theoretical ORF database of AgMNPV-2D. Data analyses allowed the identification of 39 associated proteins on ODV of AgMNPV-2D (Tab. S1). Almost half of these proteins were confirmed using both MS methodologies (Tab. S1). Moreover, twenty one proteins

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

were only identified by LC-MS/MS: PTP 1 (*ag9*), PIF-2 (*ag12*), ODV-E26 (*ag17*), PP31 (*ag25*), ODV-E66 (*ag49*), VP1054 (*ag55*), Desmoplakin (*ag66*), VLF-1 (*ag76*), VP91/P95 (*ag82*), P33 (*ag89*), Helicase (*ag92*), P6.9 (*ag96*), P48 (*ag99*), VP80/P87 (*ag100*), PIF-1 (*ag117*), PP34/PEP (*ag127*), P74 (*ag134*), ODV-E18 (*ag139*), IE-1 (*ag143*), ODV-E56 (*ag144*) and P78/83 (*ag152*) (Tab. S1). In sum, 42 proteins were established as ODV component, with thirteen proteins associated to the nucleosome, fourteen proteins to the nucleocapsid structure and other ten proteins were associated to the ODV envelope (Figure 1A). Two proteins were correlated to the polyhedra; polyhedrin and P10 and we further identified EGT, AG15 and AG20, yet with unknown function in ODV (Fig. 1A and Tab. S1). Nearly all proteins were reported in other proteomics studies of *Baculoviridae* [19-21, 24, 25], except PTP, EGT, AG20, PP31 IAP-1 and PK1, which were reported for the first time as s proteins of ODV in AgMNPV-2D.

BV associated proteins LC-MS/MS analyses. Having established the ODV proteins, we digested *in gel* nineteen bands from SDS-PAGE of BV virions (Figure S2B). All samples were loaded into the HPLC column coupled on the QSTAR elite hybrid quadrupole-TOF mass spectrometer. We determined thirty-three structural proteins on BV of AgMNPV-2D with more than 99% confidence and 2 unique peptides (Fig.1B and Table S2). As expected, some proteins were present in more than one band: F-like protein, PP31, V-ubiquitin, Desmoplakin, VLF-1, VP39, AG112, AG113, GP64 and ME53. Except for V-ubiquitin, most of them were found in bands with the expected molecular weight (Tab. S2). Furthermore, six proteins were previously thought to be present only in ODV but were also common to BV structure: PP31, Desmoplakin, ODV-E25, PEP/PP34, ODV-E18 and ODV-E56. This observation supports what was observed in other BV proteomes of NPVs [22, 26]. Seven of the thirty-three proteins were associated to nucleosome of BV particle, thirteen proteins were associated to nucleocapsid and seven proteins were associated to BV envelope (Fig. 1B). Similarly to the ODV, we detected proteins with unknown function on the BV: AG15, AG52, AG73, AG112, AG113 and AG130 (Fig. 1B and Table S2). Remarkably, six new structural proteins were detected in the BV of the

AgMNPV-2D: AG15, LEF-3, P33, Helicase, PNK/PNL and AG130. We did not detect the basic protein P6.9 (7 kDa) in the BV proteome. It is possible that P6.9 was cleaved into one peptide of 11aa length and another small oligopeptide fragments were generated during the cleavage, as was also seen for other baculovirus proteomes [26]. Nevertheless it was also not detectable by electrospray ionization used during LC MS/MS. No bovine serum albumin was found (used for supplementing cell growth) against the contaminant database supplied by Applied.

Western blot analyses. We carried out immunoblotting using polyclonal antibodies against: PTP, LEF-1, ODV-E26 and PIF-3 to further confirm the proteome results (Figure 2 and Figure S3). During the MS/MS experiments, LEF-1 and PIF-3 from ODV showed low score and coverage against the AgMNPV-2D ORF database (data not shown). However, the western blot analyses confirmed both proteins in two distinct ODV purified of AgMNPV-2D. We also confirmed ODV-E26 and PTP present only in ODV (Fig. 2 and Fig. S3). Therefore, the polyclonal antibodies were tested against two distinct BV purified from AgMNPV-2D in UFL-AG-286 and all four proteins were not detected, supporting the MS/MS results.

Proteins of AgMNPV-2D and other baculovirus studies. Twenty-two viral proteins were shared by both phenotypes of AgMNPV-2D (Figure 3). This was expected, since most of common proteins are present in the nucleocapsid and are essential for its formation, like: VP39, P40, P49, ODV-E27, FP25K, VP80/P87, P24, VP1054, PTP, Desmoplakin VLF-1 and ODV-EC43. VP39 is the major capsid protein as well as the most abundant protein in the nucleocapsid [44]. P40, P49 and ODV-E27 are highly conserved in all baculoviruses sequenced so far and, appear to be organized at the base of nucleocapsid [45]. Furthermore, ODV-E27, with an important role on G2 phase, had high similarity to cyclin B of the *Bombyx mori*. ODV-E27 was shown to play a role in host cell replication arrest until the end of virus packing [46]. The FP25K was associated to both, the BV and the ODV nucleocapsids. This protein is directly involved in trafficking proteins into envelope on ODV [47]. The VP80/P87 and P24 were previously associated to BV and ODV as capsid structural proteins. Likewise, VP1054 is essential for nucleocapsid assembly in BV

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

and ODV and organization of VP39 monomers [48]. Desmoplakin is suggested to act at the supercoiling process of DNA and appears to be necessary for an efficient egress of virion from the cellular nucleus [49]. VLF-1 is required for the production of nucleocapsid and regulates very late gene expression [50]. The ODV-EC43 orthologous are present in alpha- and betabaculoviruses and this protein was suggested to be a PIF6 in the SfMNPV [51]. We also identified several BV/ODV common structural proteins present in the envelope, like: ODV-E25, ODV-E18, PEP and ODV-E56. Most of these proteins were previously thought to be only in ODV envelope. The ODV-E18 was also found in BV and it plays an important role during its production [52]. Both ODV-E25 and ODV-E56 have a transmembrane *motif* on their N-terminal and C-terminal, respectively [53, 54]. These two proteins are more abundant in the ODV and are not essential for the production of BV [53, 54]. We also detected PEP, a phosphoprotein of OB envelope [55] in both phenotypes. This data corroborated with AcMNPV BV proteome [26]. Furthermore, the proteins of AgMNPV-2D were searched for homologues among the six published ODV proteome and sequenced genomes (AcMNPV[19], CuniNPV [20], HearNPV [21, 22], BmNPV [23], ChchNPV [24] and PrGV [25]). Almost half of ODV proteome was encoded by baculovirus core genes (Table S1). Similarly as in the case of the ODV, twelve BV proteins were highly conserved (core genes) (Table S2) [22, 26].

Proteins found only in the AgMNPV-2D proteome. Our data revealed six proteins in the AgMNPV-2D ODV that were not found so far in other ODV proteome studies: PTP, EGT, AG20, PP31, IAP-1 and PK1. PTP is a protein tyrosine phosphatase produced only by group I NPV that preferentially uses RNA as substrate. PTP and ecdysteroid uridine 5'-diphosphate (UDP)-glucosyltransferase (EGT) have been described in hyper-mobility in *Bombyx mori* infected with BmNPV [56] and *Lymantria dispar* infected by LdMNPV (*Lymantria dispar* multiple nucleopolyedrovirus) [57], respectively. We also identified a 38kDa protein coded by the AG20 ORF using both MS approaches, implying the abundance of this protein in the AgMNPV-2D ODV. The PP31 was also found in the ODV and is critical for late gene expression [58]. This phosphoprotein was also only found in the AcMNPV BV proteome so far [26]. IAP-1 was identified by MALDI and ESI ionization, this protein

was initially classified as anti-apoptotic, but it is possible that it has an opposite function in the AgMNPV, as in other Group I NPV. The last unique protein identified in ODV was the protein kinase, PK1, a serine/threonine kinase required for *polh* transcription in AcMNPV [59] that possibly phosphorylates the p6.9 basic protein (Zanotto, P.MA. unpublished results). On the other hand, proteome of AgMNPV revealed six new structural proteins in BV composition: AG15, LEF-3, P33, Helicase, PNK/PNL and AG130. Both AG15 and AG130 were detected by MS analysis, producing two proteins with 41 kDa and 30 kDa, respectively. On the other hand, LEF-3, P33 and Helicase are part of nucleosome of BV binding with viral DNA, and its plays a important role in DNA replication [60, 61]. The PNK/PKL has two described domains: kinase and T4 RNA ligase [62].

Host proteins associated to AgMNPV-2D. The MS/MS data from proteomic analyses, MALDI-Q-TOF and ESI-Q-TOF MS/MS, supplied additional information of host proteins inside the BV and ODV of AgMNPV-2D by using a non-redundant insect databases from NCBI databank (<http://www.ncbi.nlm.nih.gov/>). Eleven cellular proteins were identified, such as: actin, α -tubulin, elongation factor 1 α , elongation factor Tu, chaperone proteins like (HSP) 70 and 90, ribosomal protein S9, ADP/ATP translocase, ATP synthase, Histone H4-like protein and GTP binding nuclear protein ran (Fig. 3 and Table S3). However, we could identify only actin and HSP 90 in the BV (Fig.3). The associated host proteins were divided functionality, based on NCBI and Uniprot (<http://www.uniprot.org/>) as: mitochondrion inner membrane, cytoplasm and cytoplasm/nucleus. Our preliminary, unpublished data on RNA suppression subtractive hybridization (SSH) showed that most of these cellular proteins had their mRNA level altered during AgMNPV-2D infection [33]. Importantly, these virion-associated host proteins have been identified previously in other enveloped viruses (Vaccinia virus (VV) [63], Herpesviruses (AIHV-1 [64], HSV [65] and KSHV [66]), Epstein-Barr virus (EBV) [67], human cytomegalovirus (HCMV) [68], respiratory syncytial virus (RSV) [69], influenza virus (IV)[70] and HIV-1 [71]) and some have already been identified in other baculovirus[22, 26], like:

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

375 Actin, α tubulin, HSP 70 and 90, elongation factor 1 α , Histone H4-like protein and
376 GTP binding nuclear protein ran.

377

378 **Discussion**

379

380 By using different proteomic methodologies we identified the structural proteins of
381 the AgMNPV-2D isolate. Our complete proteomic studies and immunoblotting
382 experiments confirmed 44 and 33 proteins present in the ODV and BV, respectively
383 (Fig. 3). Additionally, ten host proteins possibly carried by ODV and two with the
384 BV phenotypes were also detected.

385

386 **Complete MS/MS analyses of AgMNPV.** We combined 2-DE and MALDI
387 (Tab. S1) with SDS-PAGE and LCMS/MS to obtain our proteomic data (Table S1
388 and S2). Combining these two methods was also shown to yield more information on
389 other baculovirus proteomes [19-26]. Among the methodologies used, the mass
390 spectrometry with the matrix-assisted laser desorption/ionization time of flight
391 (MALDI) was less sensitive even when using peptides separated by 2-dimensional
392 electrophoresis (2-DE). Indeed, only 21 proteins were detected in the ODV using this
393 technique. However, when we used 2-DE and MALDI-Q-TOF three proteins were
394 exclusively documented as ODV components: DNAPol (*ag65*), CG30 (*ag85*) and
395 Alk-Exo (*agl29*). It was not surprising to find DNA pol (*ag65*) in the ODV, since
396 baculovirus encodes a replication system independent of that from the host cell. DNA
397 pol is well conserved in all baculovirus genomes and it was also found as an ODV
398 component in the AcMNPV, HearNPV, PrGV [19, 21, 25]. In addition, CG30 (*ag85*)
399 and Alk-Exo (*agl29*) were described only in the AcMNPV ODV proteome [19]. In
400 baculovirus, CG30 has no clear role but, since it harbors both a zinc-finger-like and a
401 leucine zipper *motif*, it could be involved in gene regulation [72]. The Alk-Exo
402 (*agl29*) is a homologue of an alkaline nuclease and it has 5' to 3'
403 exonuclease/endonuclease activity and is involved in DNA recombination and
404 replication [73, 74]. Critically, during the MALDI analyses, it became also evident
405 the limitation of 2-DE electrophoresis for insoluble proteins, especially membrane

proteins. To overcome these limitations, SDS-PAGE was used for both ODV and BV, within a LC MS/MS analyses. By using a stringent search against AgMNPV database, we confirmed the previously detected twenty-one proteins and we also detected twenty-one additional proteins in the ODV, to a total of forty-two. Moreover, thirty-three structural proteins were confirmed in the BV of the AgMNPV-2D.

Western blot of AgMNPV proteins. In the ODV proteome we obtained a low score for both PIF-3 and LEF-1 against AgMNPV-2D database and, in the BV proteome we did not detect the P6.9 basic protein (6.9 kDa). Regardless of the sensitivity of mass spectrometry, the failure in detecting some proteins was expected, since: (i) proteins can degrade, (ii) low abundance of specific proteins, (iii) low MS sensitivity or even, (iv) another abundant protein at a band or *spot* that can hamper the detection of camouflaged peptide during MS analyses, which will only focus on the most abundant protein species [19, 24-26, 35]. Therefore, we selected some proteins, such as LEF-1 and PIF-3, which had low score and prepared polyclonal antibodies against them. By doing so we confirmed that our MS results on the ODV and made unequivocal that these were in fact structural proteins.

Previously undetected structural and cell modulation proteins found in the AgMNPV-2D. Major efforts have been done to identify structural and auxiliary proteins of baculovirus that directly interact with host factors, since they may be keys to understanding mechanisms of host-virus interaction. Interestingly, our proteome data of AgMNPV helped elucidate the structural proteins from both phenotypes and also brought new insights on six new proteins that do modulate and regulate cellular and systemic activity and that are carried by the ODV: PTP-1, EGT, AG20, PP31, IAP-1 and PK1. During baculovirus evolution the PTP-1 and EGT proteins appear to have been acquired from ancestral hosts [57]. The PTP-1 is a protein tyrosine phosphatase present only in Group I NPV and was recently described in hyper-mobility of *Bombyx mori* infected with BmNPV [56]. Likewise, EGT (ecdysteroid uridine 5'-diphosphate (UDP) - glucosyltransferase) is involved in the reduction of systemic ecdysone titers delaying molting and metamorphosis, allowing

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

the virus to better exploit larval tissues [5]. The EGT was also show to cause the ‘tree-top disease’ behavior in the caterpillar *Lymantria dispar* infected by LdMNPV (*Lymantria dispar* multiple nucleopolyedrovirus) [57]. Moreover, at the first day post-infection the *Lymantria dispar* larva still displayed the tree-top disease behavior, even with the EGT gene knocked out in the LdMNPV [57]. This suggests that other factors, such as PTP-1 carried inside the ODV particles, may have a role in this virus-induced host behavioral change that facilitates predation and consequently, increased viral dispersal. We further identified the 38kDa protein code by *Ag20* ORF using both MS methodologies. Interestingly, its homologues are conserved in all alphabaculovirus but with an unclear role [75]. Curiously, we found the phosphoprotein PP31 in both phenotypes, but in the AcMNPV it was only present in the BV [26]. This DNA binding protein was confirmed as a critical factor for late gene expression in experiments with RNA silencing during AcMNPV infection experiments [58, 76]. Additionally, we consistently observed IAP-1 using both mass spectrometry methodologies. This protein has been classified as anti-apoptotic (*iap* genes) conserved in all NPVs from Group I sequenced so far [77]. Nevertheless, IAP-1 was recently described as an important pro-apoptotic protein. It induces caspase-dependent apoptosis in late infection by HycuMNPV (*Hyphantria cunea* multiple nucleopolyhedrovirus), AcMNPV, BmNPV, and OpMNPV (*Orgyia pseudotsugata* multiple nucleopolyhedrovirus) [78]. As we detected this protein in the ODV of the AgMNPV, it may have an important role, probably shutting off the host protein synthesis, therefore facilitating viral replication and propagation of ODVs in the late stage of infection of Group I NPVs. The last protein identified in the ODV was the protein kinase one (PK1), a serine/threonine kinase that was associated with the BV of the HearNPV [22]. Mutations associated to *pk1* in AcMNPV and BmNPV blocked the very late gene expression, hence its essentiality for baculovirus replication [59, 79, 80]. It may also play an important role in G2 arrest during the cell-division cycle [81], like ODV-E27. In the AgMNPV, PK1 possibly phosphorylates the p6.9 basic protein (Zanotto P.M.A unpublished results), but PK1 as an ODV structural protein suggests the need for phosphorylation of host or viral proteins since the onset of the infection process.

Furthermore, seven new proteins were identified as part of the BV: AG15,

LEF-3, P33, P48, Helicase, PNK/PNL and AG130. Both AG15 and AG130 were detected by MS analysis, producing two proteins with 41 kDa and 30 kDa, respectively. Experiments carried out with the AcMNPV did not show a clear role for AG15 [82] and AG130 has no function described yet. On the other hand, LEF-3, P33 and Helicase (P143) are part of nucleosome of BV binding with viral DNA, and very important for DNA replication [60, 61]. Interestingly, LEF-3 and Helicase were also tested with RNA silencing with the AcMNPV and the absence of both proteins had a strong negative result at the early phase of apoptosis signaling [76]. The P33 protein also targets the host protein P53, inhibiting apoptosis or progression of insect cell cycle [83]. Gene knockout experiment with the P33 (Ac92) in the AcMNPV system were shown to be lethal, hampering BV production after DNA replication [84]. Like ODV, it seems that the BV carries along proteins that manipulate apoptosis assisting viral replication, which are important for BV production in the AgMNPV-2D. The PNK/PKL and P48 were identified in both the BV and ODV of the AgMNPV-2D. PNK/PKL, which may be part of an RNA repair pathway, has two domains: a RNA ligase with polynucleotide 5'-kinase activity and, a second domain with a polynucleotide 3'-phosphatase activity [85]. Homologues with unclear functions were found in five other baculovirus genomes including the AgMNPV, three of which are closely related to the AcMNPV [62]. The P48 protein has orthologues in alpha, beta and gammabaculovirus and during knockout experiments it was also shown to reduce BV assembly [86].

Host associated proteins in both phenotypes. Most of the host proteins associated with the AgMNPV-2D were previously reported in the virion structure of other enveloped viruses such as, Vaccinia virus (VV) [63], Herpesviruses (AIHV-1 [64], HSV [65] and KSHV [66]), Epstein-Barr virus (EBV) [67], human cytomegalovirus (HCMV) [68], respiratory syncytial virus (RSV) [69], influenza virus (IV) [70] and HIV-1 [71]. These were studies carried out with several mammalian viruses infecting mammalian cell lines. Nevertheless, similar cellular proteins were also detected in proteomics that compared uninfected with AcMNPV infected *S. frugiperda* (SF-9) cells [87]. Our results on proteins of AgMNPV-infected cells associated with the ODV and BV are in line with the SF-9 quantitative

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

502 proteomics [87]. We found several virion-associated host proteins, differing in
503 function and source, ranging from cytoskeleton structure (e.g., actin, α -tubulin, EF
504 Tu, HSP90, RPS9 and EF1 α), DNA binding (e.g., H4), mitochondria (e.g., ANT1,
505 and ATP-synth abN) and nucleus (e.g., RAN and HSP70). We also detected
506 chaperone proteins like HSP70 and HSP90, in abundance during our MS analyses.
507 These two proteins were identified in the ODV but we only found HSP90 in the BV.
508 Moreover, HSPs were found to play a significant role in protein metabolism of several
509 viruses [87]. It was not surprising to find the ribosomal protein S9, since it was the
510 most abundant proteins in a previous AcMNPV infection studies on SF-9 cells [87].
511 Cellular actin from the cytoskeleton was found to be abundant in our MS analysis.
512 We expected to find it in both progeny forms of the AgMNPV-2D, since it is required
513 for the movement of the nucleocapsid from the cytoplasm to the nucleus at the early
514 stages of infection [88, 89].
515 It is quite interesting to notice that phylogenetically unrelated enveloped viruses do
516 appear to sequester, a similar set of host proteins in their virions. This functional
517 convergence may indicate a fundamental, general role of specific cell-functions in the
518 early stages of the virus life cycle and possibly, new roles for envelope-associated
519 functions. In conclusion, we present the structural proteins from ODV and BV of
520 AgMNPV-2D, an important member of *Baculoviridae* family. This work also
521 brought new insights into the virion-associated host proteins, which will underpin
522 subsequent investigations on function and on the acting mechanisms of these proteins.
523 Furthermore, we reported thirteen proteins that were not previously known as
524 components of these phenotypes. We also first described proteins that could have pro-
525 apoptotic activity and/or be involved in the host behavior manipulation during
526 infection, helping viral transmission. Notwithstanding, functions for most of the
527 proteins described herein for the AgMNPV-2D were based on what is known from
528 other baculoviruses. Crucially, to a certain extent the present study furthers our
529 understanding of the baculoviruses in general. Additional experiments with
530 AgMNPV are needed to better define their proteins proper localization, function and
531 their molecular interactions.

Acknowledgments

We thank the Mass Spectrometry Laboratory at Brazilian Biosciences National Laboratory, CNPEM-ABTLuS, Campinas, SP, Brazil for their valuable support with the mass spectrometry analysis and Carolyn Diaz for her assistance with the mass spectrometry analysis at the proteome facility at University of Florida (UF), USA. We also would like to thanks Prof. James E. Maruniak and Dr. Alejandra G. Maruniak for their support during the experiments at University of Florida. CTB holds FAPESP scholarship (09/16740-8) and PMAZ holds a CNPq-PQ scholarship. This work was supported financially by FAPESP (Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP N° 2008/10713-6). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of interest. The authors declare that they have no conflict of interesting.

Figure Legends.

Figure 1. AgMNPV total structural proteins. BV and ODV proteins produced during UFL-AG-286 infection by AgMNPV-2D found by MS/MS analyses. Dashed lines represent proteins found only on one phenotype and full lines indicate proteins shared by both phenotypes. **A)** Schematics of a polyhedra (ODV) showing from the outside to the inside: 10 envelope proteins (4 also present in the BV), 13 nucleosome proteins (7 also shared with the BV) and 15 nucleocapsid proteins (10 also shared with the BV). EGT, AG15 and AG20 were isolated from the ODV but with unknown localization. **B)** On the right a BV particle with its proteins: 3 unique envelope proteins, 7 shared nucleosome proteins and 3 unique nucleocapsid proteins. AG15, AG20, AG52, AG112, AG113 and AG130 were isolated from the BV but with unknown localization.

Figure 2. Summary of 4 distinct western blots on AgMNPV-2D proteins. Total protein extracts from purified BV and ODV from two distinct AgMNPV-2D infections (1 and 2) are shown. Virus-infected cells (I) and Healthy cells (H) were applied in the central lanes as positive and negative controls, respectively. ODV extracts from infections 1 and 2 are shown in the first two lane in the left. BV extracts from infections 1 and 2 are shown in the last two lanes in the

right. All 4 proteins were only observed in the ODV phenotype supporting our MS/MS findings. The polyclonal antibodies used in each immunoblotting are shown on the right and the size of each protein expected is shown to the right. Blue, pre-stained molecular mass markers are indicated as M.

Figure 3. Venn diagram showing the phenotypic associations among all the AgMNPV-2D proteins. Total protein composition in the two phenotypes from AgMNPV-2D as obtained from 2-DE and MALDI-Q-TOF, SDS-PAGE and LC MS/MS and western blotting analyses. Twenty-two proteins were shared by both phenotypes.

Supplementary Figure 1. 2-DE gel of occlusion derived virus (ODV) of AgMNPV stained with Comassie Blue-R. Map of ODV proteins (1 - 21) identified using 2-DE and MALDI-Q-TOF. Prestained molecular mass marker (M) (Fermentas) is shown on the left (11-250 kDa) and IPG strip pH 3-10 L is shown above the gel.

Supplementary Figure 2. SDS-PAGE profiles of Occlusion Derived Virus (ODV) and Budded Virus (BV) of AgMNPV stained with Comassie Blue-R. A. The ODV proteins O1-O31 bands are indicated on the right and B. the BV proteins B1-B19 bands are indicated on the right. All bands were *in-gel* digested for LC MS/MS analyzes. Prestained molecular mass marker (M) (Fermentas) is shown on the left of each gel (15-250 kDa) and (19-118 kDa).

Supplementary Figure 3. Complete western blots on AgMNPV-2D proteins. Total protein extracts from purified BV and ODV from two distinct AgMNPV-2D infections (1 and 2) are shown. Virus-infected cells (I) and Healthy cells (H) were applied in the central lanes as positive and negative controls, respectively. ODV extracts from infections 1 and 2 are shown in the first two lane in the left. BV extracts from infections 1 and 2 are shown in the last two lanes in the right. All 4 proteins were only observed in the ODV phenotype supporting our MS/MS findings. The polyclonal antibodies used in each immunoblotting are shown on the right and the size of each protein expected is shown to the right. Blue, pre-stained molecular mass markers are indicated as M.

609 **Reference**

- 610 [1] Keddie, B. A., Aponte, G. W., Volkman, L. E., The pathway of infection of
611 Autographa californica nuclear polyhedrosis virus in an insect host. *Science*
612 1989, 243, 1728-1730.
- 613 [2] Blissard, G. W., Rohrmann, G. F., Baculovirus diversity and molecular biology.
614 *Annu Rev Entomol* 1990, 35, 127-155.
- 615 [3] Castro, M. E., Souza, M. L., Araujo, S., Bilimoria, S. L., Replication of Anticarsia
616 gemmatalis nuclear polyhedrosis virus in four lepidopteran cell lines. *J Invertebr*
617 *Pathol* 1997, 69, 40-45.
- 618 [4] King, L. A., Possee, R. D., *The Baculovirus Expression System - A laboratory*
619 *guide*, Chapman & Hall, New York, NY (USA) 1992.
- 620 [5] O'Reilly, D. R., Miller, L. K., Luckow, V. A., *Baculovirus Expression Vectors: A*
621 *Laboratory Manual*, New York 1993.
- 622 [6] Tanada, Y., Kaya, H. K., *Insect Pathology*, San Diego 1993.
- 623 [7] Moscardi, F., Assessment of the application of baculoviruses for control of
624 Lepidoptera. *Annu Rev Entomol* 1999, 44, 257-289.
- 625 [8] Szewczyk, B., Hoyoscarvajal, L., Paluszek, M., Skrzecz, I., Lobodesouza, M.,
626 Baculoviruses — re-emerging biopesticides. *Biotechnology Advances* 2006, 24,
627 143-160.
- 628 [9] Herniou, E. A., Olszewski, J. A., Cory, J. S., O'Reilly, D. R., The genome sequence
629 and evolution of baculoviruses. *Annu Rev Entomol* 2003, 48, 211-234.
- 630 [10] Slack, J., Arif, B. M., The baculoviruses occlusion-derived virus: virion
631 structure and function. *Advances in virus research* 2007, 69, 99-165.
- 632 [11] Zanotto, P. M., Kessing, B. D., Maruniak, J. E., Phylogenetic interrelationships
633 among baculoviruses: evolutionary rates and host associations. *J Invertebr Pathol*
634 1993, 62, 147-164.
- 635 [12] Garavaglia, M. J., Miele, S. A., Iserle, J. A., Belaich, M. N., Ghiringhelli, P. D., The
636 ac53, ac78, ac101, and ac103 genes are newly discovered core genes in the
637 family Baculoviridae. *J Virol* 2012, 86, 12069-12079.
- 638 [13] Oliveira, J. V. d. C., Wolff, J. L. C., Garcia-Maruniak, A., Ribeiro, B. M., *et al.*,
639 Genome of the most widely used viral biopesticide: Anticarsia gemmatalis
640 multiple nucleopolyhedrovirus. *Journal of General Virology* 2006, 87, 3233-3250.

- [14] Monsma, S. A., Oomens, A. G., Blissard, G. W., The GP64 envelope fusion protein is an essential baculovirus protein required for cell-to-cell transmission of infection. *J Virol* 1996, 70, 4607-4616.
- [15] Pearson, M. N., Groten, C., Rohrmann, G. F., Identification of the lymantria dispar nucleopolyhedrovirus envelope fusion protein provides evidence for a phylogenetic division of the Baculoviridae. *J Virol* 2000, 74, 6126-6131.
- [16] Herniou, E. A., Luque, T., Chen, X., Vlak, J. M., *et al.*, Use of whole genome sequence data to infer baculovirus phylogeny. *J Virol* 2001, 75, 8117-8126.
- [17] Jehle, J. A., Lange, M., Wang, H., Hu, Z., *et al.*, Molecular identification and phylogenetic analysis of baculoviruses from Lepidoptera. *Virology* 2006, 346, 180-193.
- [18] Jehle, J. A., Blissard, G. W., Bonning, B. C., Cory, J. S., *et al.*, On the classification and nomenclature of baculoviruses: a proposal for revision. *Arch Virol* 2006, 151, 1257-1266.
- [19] Braunagel, S. C., Determination of the protein composition of the occlusion-derived virus of *Autographa californica* nucleopolyhedrovirus. *Proceedings of the National Academy of Sciences* 2003, 100, 9797-9802.
- [20] Perera, O., Green, T. B., Stevens, S. M., White, S., Becnel, J. J., Proteins Associated with *Culex nigripalpus* Nucleopolyhedrovirus Occluded Virions. *Journal of Virology* 2007, 81, 4585-4590.
- [21] Deng, F., Wang, R., Fang, M., Jiang, Y., *et al.*, Proteomics Analysis of *Helicoverpa armigera* Single Nucleocapsid Nucleopolyhedrovirus Identified Two New Occlusion-Derived Virus-Associated Proteins, HA44 and HA100. *Journal of Virology* 2007, 81, 9377-9385.
- [22] Hou, D., Zhang, L., Deng, F., Fang, W., *et al.*, Comparative Proteomics Reveal Fundamental Structural and Functional Differences between the Two Progeny Phenotypes of a Baculovirus. *J Virol* 2013, 87, 829-839.
- [23] Liu, C., Li, Z., Wu, W., Li, L., *et al.*, *Autographa californica* multiple nucleopolyhedrovirus ac53 plays a role in nucleocapsid assembly. *Virology* 2008, 382, 59-68.
- [24] Xu, F., Ince, I. A., Boeren, S., Vlak, J. M., van Oers, M. M., Protein composition of the occlusion derived virus of *Chrysodeixis chalcites* nucleopolyhedrovirus. *Virus Res* 2011, 158, 1-7.
- [25] Wang, X. F., Zhang, B. Q., Xu, H. J., Cui, Y. J., *et al.*, ODV-associated proteins of the *Pieris rapae* granulovirus. *J Proteome Res* 2011, 10, 2817-2827.

- [26] Wang, R., Deng, F., Hou, D., Zhao, Y., *et al.*, Proteomics of the *Autographa californica* Nucleopolyhedrovirus Budded Virions. *Journal of Virology* 2010, 84, 7233-7242.
- [27] Moscardi, F., Lobo de Souza, M., M.E.B., C., Szewczyk, B., in: Ahmad, I., Ahmad, F., Pichtel, J. (Eds.), *Microbes and Microbial Technology*, Springer Science, New York 2011, pp. 415-446.
- [28] Maruniak, J. E., Molecular biology of *Anticarsia gemmatilis* baculovirus. *Memorias do Instituto Oswaldo Cruz* 1989, 107-111.
- [29] Zanotto, P. M., Sampaio, M. J., Johnson, D. W., Rocha, T. L., Maruniak, J. E., The *Anticarsia gemmatilis* nuclear polyhedrosis virus polyhedrin gene region: sequence analysis, gene product and structural comparisons. *J Gen Virol* 1992, 73 (Pt 5), 1049-1056.
- [30] Maruniak, J. E., Garcia-Maruniak, A., Souza, M. L., Zanotto, P. M., Moscardi, F., Physical maps and virulence of *Anticarsia gemmatilis* nucleopolyhedrovirus genomic variants. *Arch Virol* 1999, 144, 1991-2006.
- [31] Ribeiro, B. M., Gatti, C. D., Costa, M. H., Moscardi, F., *et al.*, Construction of a recombinant *Anticarsia gemmatilis* nucleopolyhedrovirus (AgMNPV-2D) harbouring the beta-galactosidase gene. *Arch Virol* 2001, 146, 1355-1367.
- [32] Allen, G. E., Knell, J. D., A nuclear polyhedrosis virus of *Anticarsia gemmatilis*: I, ultrastructure, replication, and pathogenicity. *Fl. Entomol.* 1977, 233-240.
- [33] Oliveira, J. V. d. C., Brito, A., Braconi, C. T., Freire, C. C. M., *et al.*, University of Sao Paulo 2013, pp. 1-38.
- [34] Tyers, M., Mann, M., From genomics to proteomics. *Nature* 2003, 422, 193-197.
- [35] Steen, H., Mann, M., The abc's (and xyz's) of peptide sequencing. *Nature Reviews Molecular Cell Biology* 2004, 5, 699-711.
- [36] Zhang, L., Jia, X., Zhang, X., Sun, J., *et al.*, Proteomic analysis of PBMCs: characterization of potential HIV-associated proteins. *Proteome Sci* 2010, 8, 12.
- [37] Sieburth, P., Maruniak, J. E., Growth characteristics of a continuous cell line from the velvetbean caterpillar, *Anticarsia gemmatilis* hübner (Lepidoptera: Noctuidae). *In Vitro Cel. Dev. Biol. Plant* 1988, 195-198.
- [38] Leme, A. F. P., Escalante, T., Pereira, J. G. C., Oliveira, A. K., *et al.*, High resolution analysis of snake venom metalloproteinase (SVMP) peptide bond cleavage specificity using proteome based peptide libraries and mass spectrometry. *Journal of Proteomics* 2011, 74, 401-410.

- [39] Keller, A., Nesvizhskii, A. I., Kolker, E., Aebersold, R., Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* 2002, **74**, 5383-5392.
- [40] Nesvizhskii, A. I., Keller, A., Kolker, E., Aebersold, R., A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* 2003, **75**, 4646-4658.
- [41] Anandarao, R., Swaminathan, S., Fernando, S., Jana, A. M., Khanna, N., Recombinant multiepitope protein for early detection of dengue infections. *Clin Vaccine Immunol* 2006, **13**, 59-67.
- [42] Hooft van Iddekinge, B. J., Smith, G. E., Summers, M. D., Nucleotide sequence of the polyhedrin gene of Autographa californica nuclear polyhedrosis virus. *Virology* 1983, **131**, 561-565.
- [43] Coulibaly, F., Chiu, E., Gutmann, S., Rajendran, C., *et al.*, The atomic structure of baculovirus polyhedra reveals the independent emergence of infectious crystals in DNA and RNA viruses. *Proc Natl Acad Sci U S A* 2009, **106**, 22205-22210.
- [44] Charlton, C. A., Volkman, L. E., Sequential rearrangement and nuclear polymerization of actin in baculovirus-infected *Spodoptera frugiperda* cells. *J Virol* 1991, **65**, 1219-1227.
- [45] Braunagel, S. C., Guidry, P. A., Rosas-Acosta, G., Engelking, L., Summers, M. D., Identification of BV/ODV-C42, an Autographa californica nucleopolyhedrovirus orf101-encoded structural protein detected in infected-cell complexes with ODV-EC27 and p78/83. *J Virol* 2001, **75**, 12331-12338.
- [46] Baluchamy, S., Gopinathan, K. P., Characterization of a cyclin homolog from *Bombyx mori* nucleopolyhedrovirus. *Virus research* 2005, **108**, 69-81.
- [47] Braunagel, S. C., Williamson, S. T., Saksena, S., Zhong, Z., *et al.*, Trafficking of ODV-E66 is mediated via a sorting motif and other viral proteins: facilitated trafficking to the inner nuclear membrane. *Proc Natl Acad Sci U S A* 2004, **101**, 8372-8377.
- [48] Olszewski, J., Miller, L. K., Identification and characterization of a baculovirus structural protein, VP1054, required for nucleocapsid formation. *J Virol* 1997, **71**, 5040-5050.
- [49] Ke, J., Wang, J., Deng, R., Wang, X., Autographa californica multiple nucleopolyhedrovirus ac66 is required for the efficient egress of nucleocapsids from the nucleus, general synthesis of preoccluded virions and occlusion body formation. *Virology* 2008, **374**, 421-431.

- 748 [50] Vanarsdall, A. L., Okano, K., Rohrmann, G. F., Characterization of the role of
749 very late expression factor 1 in baculovirus capsid structure and DNA processing.
750 *J Virol* 2006, *80*, 1724-1733.
- 751 [51] Simon, O., Palma, L., Williams, T., Lopez-Ferber, M., Caballero, P., Analysis of
752 a naturally-occurring deletion mutant of *Spodoptera frugiperda* multiple
753 nucleopolyhedrovirus reveals sf58 as a new per os infectivity factor of
754 lepidopteran-infecting baculoviruses. *J Invertebr Pathol* 2012, *109*, 117-126.
- 755 [52] McCarthy, C. B., Dai, X., Donly, C., Theilmann, D. A., *Autographa californica*
756 multiple nucleopolyhedrovirus ac142, a core gene that is essential for BV
757 production and ODV envelopment. *Virology* 2008, *372*, 325-339.
- 758 [53] Russell, R. L., Rohrmann, G. F., A 25-kDa protein is associated with the
759 envelopes of occluded baculovirus virions. *Virology* 1993, *195*, 532-540.
- 760 [54] Braunagel, S. C., Elton, D. M., Ma, H., Summers, M. D., Identification and
761 analysis of an *Autographa californica* nuclear polyhedrosis virus structural
762 protein of the occlusion-derived virus envelope: ODV-E56. *Virology* 1996, *217*,
763 97-110.
- 764 [55] Whitt, M., A phosphorylated 34-kDa protein and a subpopulation of
765 polyhedrin are thiol linked to the carbohydrate layer surrounding a baculovirus
766 occlusion body. *Virology* 1988, *163*, 33-42.
- 767 [56] Katsuma, S., Koyano, Y., Kang, W., Kokusho, R., *et al.*, The baculovirus uses a
768 captured host phosphatase to induce enhanced locomotory activity in host
769 caterpillars. *PLoS Pathog* 2012, *8*, e1002644.
- 770 [57] Hoover, K., Grove, M., Gardner, M., Hughes, D. P., *et al.*, A gene for an
771 extended phenotype. *Science* 2011, *333*, 1401.
- 772 [58] Guarino, L. A., Mistretta, T. A., Dong, W., DNA binding activity of the
773 baculovirus late expression factor PP31. *Virus Res* 2002, *90*, 187-195.
- 774 [59] Mishra, G., Chadha, P., Das, R. H., Serine/threonine kinase (pk-1) is a
775 component of *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV)
776 very late gene transcription complex and it phosphorylates a 102 kDa
777 polypeptide of the complex. *Virus Res* 2008, *137*, 147-149.
- 778 [60] Kool, M., Ahrens, C. H., Goldbach, R. W., Rohrmann, G. F., Vlak, J. M.,
779 Identification of genes involved in DNA replication of the *Autographa californica*
780 baculovirus. *Proc Natl Acad Sci U S A* 1994, *91*, 11212-11216.
- 781 [61] Hang, X., Dong, W., Guarino, L. A., The lef-3 gene of *Autographa californica*
782 nuclear polyhedrosis virus encodes a single-stranded DNA-binding protein. *J*
783 *Virol* 1995, *69*, 3924-3928.

- [62] Durantel, D., Croizier, L., Ayres, M. D., Croizier, G., *et al.*, The *pnk/pnl* gene (ORF 86) of *Autographa californica* nucleopolyhedrovirus is a non-essential, immediate early gene. *J Gen Virol* 1998, 79 (Pt 3), 629-637.
- [63] Resch, W., Hixson, K. K., Moore, R. J., Lipton, M. S., Moss, B., Protein composition of the vaccinia virus mature virion. *Virology* 2007, 358, 233-247.
- [64] Dry, I., Haig, D. M., Inglis, N. F., Imrie, L., *et al.*, Proteomic analysis of pathogenic and attenuated alcelaphine herpesvirus 1. *J Virol* 2008, 82, 5390-5397.
- [65] Loret, S., Guay, G., Lippe, R., Comprehensive characterization of extracellular herpes simplex virus type 1 virions. *J Virol* 2008, 82, 8605-8618.
- [66] Zhu, F. X., Chong, J. M., Wu, L., Yuan, Y., Virion proteins of Kaposi's sarcoma-associated herpesvirus. *J Virol* 2005, 79, 800-811.
- [67] Johannsen, E., Luftig, M., Chase, M. R., Weicksel, S., *et al.*, Proteins of purified Epstein-Barr virus. *Proc Natl Acad Sci U S A* 2004, 101, 16286-16291.
- [68] Varnum, S. M., Streblow, D. N., Monroe, M. E., Smith, P., *et al.*, Identification of proteins in human cytomegalovirus (HCMV) particles: the HCMV proteome. *J Virol* 2004, 78, 10960-10966.
- [69] Radhakrishnan, A., Yeo, D., Brown, G., Myaing, M. Z., *et al.*, Protein analysis of purified respiratory syncytial virus particles reveals an important role for heat shock protein 90 in virus particle assembly. *Mol Cell Proteomics* 2010, 9, 1829-1848.
- [70] Shaw, M. L., Stone, K. L., Colangelo, C. M., Gulcicek, E. E., Palese, P., Cellular proteins in influenza virus particles. *PLoS Pathog* 2008, 4, e1000085.
- [71] Chertova, E., Chertov, O., Coren, L. V., Roser, J. D., *et al.*, Proteomic and biochemical analysis of purified human immunodeficiency virus type 1 produced from infected monocyte-derived macrophages. *J Virol* 2006, 80, 9039-9052.
- [72] Passarelli, A. L., Miller, L. K., In vivo and in vitro analyses of recombinant baculoviruses lacking a functional *cg30* gene. *J Virol* 1994, 68, 1186-1190.
- [73] Mikhailov, V. S., Okano, K., Rohrmann, G. F., Baculovirus alkaline nuclease possesses a 5'-->3' exonuclease activity and associates with the DNA-binding protein LEF-3. *J Virol* 2003, 77, 2436-2444.
- [74] Mikhailov, V. S., Okano, K., Rohrmann, G. F., Specificity of the endonuclease activity of the baculovirus alkaline nuclease for single-stranded DNA. *J Biol Chem* 2004, 279, 14734-14745.

- [75] Miele, S. A. B., Garavaglia, M. J., Belaich, M. N., Ghiringhelli, P. D., Baculovirus: Molecular Insights on Their Diversity and Conservation. *International Journal of Evolutionary Biology* 2011, 2011, 1-15.
- [76] Schultz, K. L., Friesen, P. D., Baculovirus DNA replication-specific expression factors trigger apoptosis and shutoff of host protein synthesis during infection. *J Virol* 2009, 83, 11123-11132.
- [77] Clem, R. J., Baculoviruses and apoptosis: a diversity of genes and responses. *Curr Drug Targets* 2007, 8, 1069-1074.
- [78] Ikeda, M., Yamada, H., Ito, H., Kobayashi, M., Baculovirus IAP1 induces caspase-dependent apoptosis in insect cells. *J Gen Virol* 2011, 92, 2654-2663.
- [79] Reilly, L. M., Guarino, L. A., The pk-1 gene of *Autographa californica* multinucleocapsid nuclear polyhedrosis virus encodes a protein kinase. *J Gen Virol* 1994, 75 (Pt 11), 2999-3006.
- [80] Katsuma, S., Mita, K., Shimada, T., ERK- and JNK-dependent signaling pathways contribute to *Bombyx mori* nucleopolyhedrovirus infection. *J Virol* 2007, 81, 13700-13709.
- [81] Ikeda, M., Kobayashi, M., Cell-cycle perturbation in Sf9 cells infected with *Autographa californica* nucleopolyhedrovirus. *Virology* 1999, 258, 176-188.
- [82] Wang, Y., Wu, W., Li, Z., Yuan, M., *et al.*, ac18 is not essential for the propagation of *Autographa californica* multiple nucleopolyhedrovirus. *Virology* 2007, 367, 71-81.
- [83] Prikhod'ko, G. G., Wang, Y., Freulich, E., Prives, C., Miller, L. K., Baculovirus p33 binds human p53 and enhances p53-mediated apoptosis. *Journal of Virology* 1999, 73, 1227-1234.
- [84] Wu, W., Passarelli, A. L., *Autographa californica* multiple nucleopolyhedrovirus Ac92 (ORF92, P33) is required for budded virus production and multiply enveloped occlusion-derived virus formation. *J Virol* 2010, 84, 12351-12361.
- [85] Martins, A., Shuman, S., Characterization of a baculovirus enzyme with RNA ligase, polynucleotide 5'-kinase, and polynucleotide 3'-phosphatase activities. *J Biol Chem* 2004, 279, 18220-18231.
- [86] Yuan, M., Wu, W., Liu, C., Wang, Y., *et al.*, A highly conserved baculovirus gene p48 (ac103) is essential for BV production and ODV envelopment. *Virology* 2008, 379, 87-96.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

[87] Carinhas, N., Robitaille, A. M., Moes, S., Carrondo, M. J., *et al.*, Quantitative proteomics of *Spodoptera frugiperda* cells during growth and baculovirus infection. *PLoS ONE* 2011, 6, e26444.

[88] Ohkawa, T., Volkman, L. E., Welch, M. D., Actin-based motility drives baculovirus transit to the nucleus and cell surface. *The Journal of cell biology* 2010, 190, 187-195.

[89] Gandhi, K. M., Ohkawa, T., Welch, M. D., Volkman, L. E., Nuclear localization of actin requires AC102 in *Autographa californica* multiple nucleopolyhedrovirus-infected cells. *J Gen Virol* 2012, 93, 1795-1803.

AgMNPV total structural proteins. BV and ODV proteins produced during UFL-AG-286 infection by AgMNPV-2D found by MS/MS analyses. Dashed lines represent proteins found only on one phenotype and full lines indicate proteins shared by both phenotypes. A) Schematics of a polyhedra (ODV) showing from the outside to the inside: 10 envelope proteins (4 also present in the BV), 13 nucleosome proteins (7 also shared with the BV) and 15 nucleocapsid proteins (10 also shared with the BV). EGT, AG15 and AG20 were isolated from the ODV but with unknown localization. B) On the right a BV particle with its proteins: 3 unique envelope proteins, 7 shared nucleosome proteins and 3 unique nucleocapsid proteins. AG15, AG20, AG52, AG112, AG113 and AG130 were isolated from the BV but with unknown localization.

133x104mm (300 x 300 DPI)

Summary of 4 distinct western blots on AgMNPV-2D proteins. Total protein extracts from purified BV and ODV from two distinct AgMNPV-2D infections (1 and 2) are shown. Virus-infected cells (I) and Healthy cells (H) were applied in the central lanes as positive and negative controls, respectively. ODV extracts from infections 1 and 2 are shown in the first two lane in the left. BV extracts from infections 1 and 2 are shown in the last two lanes in the right. All 4 proteins were only observed in the ODV phenotype supporting our MS/MS findings. The polyclonal antibodies used in each immunoblotting are shown on the right and the size of each protein expected is shown to the right. Blue, pre-stained molecular mass markers are indicated as M.

52x27mm (300 x 300 DPI)

Venn diagram showing the phenotypic associations among all the AgMNPV-2D proteins. Total protein composition in the two phenotypes from AgMNPV-2D as obtained from 2-DE and MALDI-Q-TOF, SDS-PAGE and LC MS/MS and western blotting analyses. Twenty-two proteins were shared by both phenotypes.

107x68mm (300 x 300 DPI)

Supplementary material.

Figure S1. 2-DE gel of occlusion derived virus (ODV) of AgMNPV stained with Comassie Blue-R

Figure S2. SDS-PAGE profiles of Occlusion Derived Virus (ODV) and Budded Virus (BV) of AgMNPV stained with Comassie Blue-R

Figure S3. Complete western blots on AgMNPV-2D proteins

Table S1. List of AgMNPV ODV proteins identified using mass spectrometry

ODV PROTEOME					MALDI-Q-TOF				LC MS/MS		Function	Access
AgMNPV ORF	AcMNPV ORF	Spots 2DE	Bands SDS	Protein	size (kDa)	pI	Cov. (%)	Score	Cov. (%)	Score		
1	8	1	O20/O21/O22/O23	Polyhedrin	28.9	6,11	11	98	58	2060	OB main structure protein [1]	gi 1163
9	1		O30	PTP	20,1	8,26			37	395	BV/ODV protein and RNA phosphatase [2]	gi 1163
12	22		O16	PIF-2	44,3	5,30			18	428	ODV protein essential for oral infectivity [3]	gi 1163
15	18	2	O17	AG15	41,4	7,55	10	149	58	2697	Not essential protein [4]	gi 1163
17	16		O25	ODV-E26	24,0	10,22			10	234	BV/ODV envelope [5]	gi 1163
18	15	3	O5/O11/O12/O13	EGT	54,8	6,73	20	391	34	2155	This study, Enzyme that prevents insect molting [6]	gi 1163
20	13	4	O18	AG20	38,0	8,63	24	278	64	2071	Unknown function	gi 1163
25	36		O22/O29	PP31	30,4	9,52			28	389	Phosphoprotein involved in late gene expression [7]	gi 1163
40	27	5	O19	IAP-1	33,0	5,07	21	215	20	245	Induces apoptosis in NPV I [8]	gi 1163
49	46		O8/O9	ODV-E66	76,2	5,59			32	3209	ODV envelope protein [9]	gi 1163
55	54		O16	VP1054	43,8	8,54			32	499	Required for nucleocapsid formation [10]	gi 1163
61	61	6	O29/O30	25K FP	27,4	9,48	13	90	62	1422	BV/ODV protein and mutations reduces polyhedra [11]	gi 1163
65	65	7		DNA pol.	116,43	8,51	10	150	-	-	DNA replication [12]	gi 1163
66	66		O3/O4	Desmoplakin	101,3	5,06			35	2034	Supercoiling process of DNA [13]	gi 1163
67	67	8	O15	LEF-3	43,8	5,37	18	155	32	561	Protein that binds with ssDNA [14]	gi 1163
76	77		O17/O21	VLF-1	43,7	9,52			25	569	BV/ODV nucleocapsids and essential for its formation [15]	gi 1163
79	80	9	O18/O19	GP41	42,5	7,04	27	472	39	679	Tegument protein [16]	gi 1163
82	83		O5/O8	VP91/P95	95,3	4,62			30	1489	Viral capsid protein and envelope of ODV [17]	gi 1163
85	88	10		CG30	28,78	9,00	7	170	-	-	Zinc-finger like motif involved in gene regulation [18]	gi 1163
86	89	11	O19/O20/O21/O22	VP39	38,9	6,54	12	114	77	6582	BV/ODV major capsid protein and organization of host actin [19]	gi 1163
89	92		O22	P33	30,9	8,17	43	330	42	920	BV production and ODV formation [20]	gi 1163
91	94	12	O27/O28/O29	ODV-E25	26,1	6,84			38	709	ODV envelope protein [21]	gi 1163
92	95		O1/O2	Helicase	142,2	7,10			5	110	DNA replication [22]	gi 1163
96	100	13	O31	P6.9	6,7	12,36			16	102	Major DNA-binding protein [23]	gi 1163
97	101		O19	P40	40,7	7,01	10	115	30	888	Nuclear entry of P78/83 and BV/ODV morphogenesis [24]	gi 1163
99	103		O14/O16	P48	49,0	8,39			40	1121	BV production and ODV envelopment [25]	gi 1163
100	104		O6	VP80/P87	78,5	5,66			16	456	Capsid-associated [26] and BV production [27]	gi 1163
103	86	14	O7/O8	PNK/PNL	79,6	8,65	25	293	52	2516	Protein with 2 domains: kinase and T4 RNA ligase [28]	gi 1163
107	109	15	O15	ODV-EC43	45,9	8,18	33	291	42	1191	BV/ ODV envelope protein [29]	gi 1163
17	119	16	O10	PIF-1	60,2	5,61			16	148	ODV protein essential for oral infectivity in insect larvae [30]	gi 1163
125	129		O22/O24	P24	21,3	5,63	30	208	41	1549	BV/ODV nucleocapsid protein [31]	gi 1163
127	131		O20/O21	PEP/PP34	32,9	4,87			50	3802	Phosphoprotein of OB envelope [32]	gi 1163
129	133	17		Alk-Exo	47,94	9,12	10	145	-	-	Alkaline nuclease [33]	gi 1163
133	137	18	O30/O31	P10	10,3	4,25	33	208	74	1430	Releases polyhedra from cell nucleus [34]	gi 1163
134	138		O7/O8	P74	73,7	5,05			10	152	ODV envelope protein essential for oral infectivity [35]	gi 1163
138	142	19	O14/O15	P49	56,4	8,75	13	144	69	4699	BV/ODV nucleocapsids [36]	gi 1163
139	143		O30/O31	ODV-E18	9,3	8,31			40	1631	ODV envelope protein [37]	gi 1163
140	144	20	O19	ODV-E27	34,4	7,70	12	73	45	1498	ODV envelope protein and capsid structures [37]	gi 1163
143	147		O9/O10	IE1	68,2	6,27			26	1196	Transactivator, inducing apoptotic activity [38]	gi 1163
144	148		O10/O16	ODV-E56	40,1	5,77			31	1833	ODV envelope associated [39]	gi 1163
151	10	21	O22	PK1	31,6	8,87	12	59	26	729	Required for very late gene expression [40, 41]	gi 1163
152	9		O5/O6	P78/83	69,8	6			12	162	Nucleocapsid base [21]	gi 1163

* Shade lines are the ORFs conserved among all sequenced baculovirus genomes * Cov. means coverage in percentage of the proteins using mass spectrometry

* gi numbers from GenBank using the genome of AgMNPV

Table S2. List of AgMNPV BV proteins identified using mass spectrometry

BV PROTEOME				LC MS/MS			Access (#)
AgMNPV ORF	AcMNPV ORF	Band SDS	Protein	Size (kDa)	Seq.(%)	Spectra MS/MS	
15	18	B8	AG15	41	14	5	gi116326084
22	23	B15/B16	F-like	75	26	13	gi116326092
25	36	B3/B5/B6/B7	PP31	30	38	5	gi116326094
26	35	B7/B8/B14	V-ubiquitin	9	37	2	gi116326095
52	51	B6	AG52	37	13	4	gi116326121
55	54	B9	VP1054	43	13	4	gi116326124
61	61	B5	25K FP	27	47	12	gi116326221
66	66	B1/B2	Desmoplakin	101	15	12	gi116326134
67	67	B10	LEF-3	43	36	14	gi116326135
70	71	B16	IAP-2	27	17	4	gi116326138
73	74	B18	AG 73	20	14	2	gi116326141
76	77	B10/B11/B12	VLF-1	44	30	11	gi116326144
86	89	B14/B15/B16	VP39	38	49	14	gi116326154
89	92	B16	P33	30	14	4	gi116326157
91	94	B18	ODV E-25	26	36	7	gi116326159
92	95	B2	Helicase	141	16	17	gi116326160
94	98	B13	38K	34	11	3	gi116326208
97	101	B11	P40	40	8	3	gi116326165
99	103	B10/B11	P48	48	16	6	gi116326167
100	104	B4	VP80/P87	78	12	8	gi116326168
103	86	B5	PNK/PNL	78	40	27	gi116326171
107	109	B11	ODV-EC43	45	9	4	gi116326175
112	-	B9/B10/B13	AG112	41	42	18	gi116326180
113	114	B10/B11	AG 113	48	21	9	gi116326181
124	128	B7/B8/B9/B10	GP64	57	29	16	gi116326192
125	129	B17	P24	21	11	6	gi116326193
127	131	B15	PEP/PP34	33	37	9	gi116326195
130	-	B18	AG 130	24	7	2	gi116326198
136	139	B8/B9/B10	ME53	52	26	11	gi116326204
138	142	B8	P49	56	20	9	gi116326213
139	143	B19	ODV-E18	9	40	3	gi116326207
140	144	B13	ODV-E27	34	20	6	gi116326208
144	148	B11	ODV-E56	40	16	4	gi116326212

* Shade lines are the ORF's conserved among all sequenced baculovirus genomes * Cov. means coverage in percentage of the proteins using mass spectrometry

* gi numbers from GenBank using the genome of AgMNPV

Table S3. Best matches of MS/MS data of host proteins from MALDI-Q-TOF and ESI-Q-TOF analyses

Classification	Protein	Abbreviation	Species	Other virus(as) reported	NCBI GI
Cytoskeleton	Actin	Actin	<i>Heliothis virescens</i>	AcMNPV BV [20], AIHV-1 [49], EBV [50], HCMV [51], HearNPV BV [52], HIV [53], HSV [54], IV [55] KSHV [56], MMLV [57], RSV [58], VV [59]	14010639
	Alpha-tubulin	α tubulin	<i>Bombyx mori</i>	EBV [50], HCMV [51], HearNPV BV [52], HIV [53], HSV [54], IV [55] KSHV [56], RSV [58], VV [59]	19773422
	Elongation factor Tu	EF Tu	<i>Bombyx mori</i>		87248117
	Heat shock 90	HSP 90	<i>Spodoptera frugiperda</i>	EBV [50], HCMV [51], HearNPV BV, HIV [53], KSHV [56], RSV [58]	12005809
DNA binding	Ribosomal protein S9	RPS9	<i>Manduca sexta</i>		268306404
	Elongation factor 1 alpha	EF1_ α	<i>Helicoverpa armigera</i> <i>Bombyx mori</i>	EBV [50], HCMV [51], HearNPV BV[52], HIV [53], KSHV [56], RSV [58], VV [59]	217274
	Histone H4	H4	<i>Bombyx mori</i>	AIHV-1 [49], HearNPV BV [52], HIV [53], HIV [53], MMLV [57]	223972586
	ADP/ATP translocase	ANT1	<i>Helicoverpa armigera</i>		151384885
Mitochondrion inner membrane	ATP synthase	ATP-synt abN	<i>Bombyx mori</i>		87248463
	GTP binding nuclear protein ran	RAN	<i>Helicoverpa armigera</i>	HearNPV BV [52], HIV [53], HIV [53]	224831916
Cytoplasm/ Nucleus	Heat shock 70 kDa	HSP 70	<i>Bombyx mori</i>	EBV [50], HCMV [51], HearNPV [52], HIV [53], HSV [54], KSHV [56], MMLV [57], RSV [58], VV [59]	3426021

*gi numbers from GenBank using an non redundant indexed insect database from NCBI *Classification and function were determined using NCBI and Uniprot (<http://www.uniprot.org/>)

References

- [1] Whitford, M., Faulkner, P., A structural polypeptide of the baculovirus *Autographa californica* nuclear polyhedrosis virus contains O-linked N-acetylglucosamine. *J Virol* 1992, 66, 3324-3329.
- [2] Takagi, T., Taylor, G. S., Kusakabe, T., Charbonneau, H., Buratowski, S., A protein tyrosine phosphatase-like protein from baculovirus has RNA 5'-triphosphatase and diphosphatase activities. *Proc Natl Acad Sci U S A* 1998, 95, 9808-9812.
- [3] Pijlman, G. P., Identification of pif-2, a third conserved baculovirus gene required for per os infection of insects. *Journal of General Virology* 2003, 84, 2041-2049.
- [4] Wang, Y., Wu, W., Li, Z., Yuan, M., *et al.*, ac18 is not essential for the propagation of *Autographa californica* multiple nucleopolyhedrovirus. *Virology* 2007, 367, 71-81.
- [5] Beniya, H., Braunagel, S. C., Summers, M. D., *Autographa californica* nuclear polyhedrosis virus: subcellular localization and protein trafficking of BV/ODV-E26 to intranuclear membranes and viral envelopes. *Virology* 1998, 240, 64-75.
- [6] Hoover, K., Grove, M., Gardner, M., Hughes, D. P., *et al.*, A gene for an extended phenotype. *Science* 2011, 333, 1401.
- [7] Guarino, L. A., Mistretta, T. A., Dong, W., DNA binding activity of the baculovirus late expression factor PP31. *Virus Res* 2002, 90, 187-195.
- [8] Ikeda, M., Yamada, H., Ito, H., Kobayashi, M., Baculovirus IAP1 induces caspase-dependent apoptosis in insect cells. *J Gen Virol* 2011, 92, 2654-2663.
- [9] Hong, T., N-terminal sequences from *Autographa californica* nuclear polyhedrosis virus envelope proteins ODV-E66 and ODV-E25 are sufficient to direct reporter proteins to the nuclear envelope, intranuclear microvesicles and the envelope of occlusion derived virus. *Proceedings of the National Academy of Sciences* 1997, 94, 4050-4055.
- [10] Olszewski, J., Miller, L. K., Identification and characterization of a baculovirus structural protein, VP1054, required for nucleocapsid formation. *J Virol* 1997, 71, 5040-5050.
- [11] Harrison, R. L., Jarvis, D. L., Summers, M. D., The role of the AcMNPV 25K gene, "FP25," in baculovirus polh and p10 expression. *Virology* 1996, 226, 34-46.
- [12] Vanarsdall, A. L., Okano, K., Rohrmann, G. F., Characterization of the replication of a baculovirus mutant lacking the DNA polymerase gene. *Virology* 2005, 331, 175-180.
- [13] Ke, J., Wang, J., Deng, R., Wang, X., *Autographa californica* multiple nucleopolyhedrovirus ac66 is required for the efficient egress of nucleocapsids from the nucleus, general synthesis of preoccluded virions and occlusion body formation. *Virology* 2008, 374, 421-431.
- [14] Hang, X., Dong, W., Guarino, L. A., The lef-3 gene of *Autographa californica* nuclear polyhedrosis virus encodes a single-stranded DNA-binding protein. *J Virol* 1995, 69, 3924-3928.
- [15] Li, Y., Wang, J., Deng, R., Zhang, Q., *et al.*, vlf-1 deletion brought AcMNPV to defect in nucleocapsid formation. *Virus Genes* 2005, 31, 275-284.
- [16] Pan, L., Li, Z., Gong, Y., Yu, M., *et al.*, Characterization of gp41 gene of *Spodoptera litura* multicapsid nucleopolyhedrovirus. *Virus Res* 2005, 110, 73-79.

- [17] Russell, R. L., Rohrmann, G. F., Characterization of P91, a protein associated with virions of an *Orgyia pseudotsugata* baculovirus. *Virology* 1997, 233, 210-223.
- [18] Passarelli, A. L., Miller, L. K., In vivo and in vitro analyses of recombinant baculoviruses lacking a functional cg30 gene. *J Virol* 1994, 68, 1186-1190.
- [19] Charlton, C. A., Volkman, L. E., Sequential rearrangement and nuclear polymerization of actin in baculovirus-infected *Spodoptera frugiperda* cells. *J Virol* 1991, 65, 1219-1227.
- [20] Wu, W., Passarelli, A. L., Autographa californica multiple nucleopolyhedrovirus Ac92 (ORF92, P33) is required for budded virus production and multiply enveloped occlusion-derived virus formation. *J Virol* 2010, 84, 12351-12361.
- [21] Slack, J., Arif, B. M., The baculoviruses occlusion-derived virus: virion structure and function. *Advances in virus research* 2007, 69, 99-165.
- [22] Kool, M., Ahrens, C. H., Goldbach, R. W., Rohrmann, G. F., Vlak, J. M., Identification of genes involved in DNA replication of the Autographa californica baculovirus. *Proc Natl Acad Sci U S A* 1994, 91, 11212-11216.
- [23] Kelly, D. C., Brown, D. A., Ayres, M. D., Allen, C. J., Walker, I. O., Properties of the Major Nucleocapsid Protein of *Heliothis zea* Singly Enveloped Nuclear Polyhedrosis Virus. *Journal General Virology* 1983, 64, 399-408.
- [24] Vanarsdall, A. L., Pearson, M. N., Rohrmann, G. F., Characterization of baculovirus constructs lacking either the Ac 101, Ac 142, or the Ac 144 open reading frame. *Virology* 2007, 367, 187-195.
- [25] Yuan, M., Wu, W., Liu, C., Wang, Y., *et al.*, A highly conserved baculovirus gene p48 (ac103) is essential for BV production and ODV envelopment. *Virology* 2008, 379, 87-96.
- [26] Lu, A., Carstens, E. B., Nucleotide sequence of a gene essential for viral DNA replication in the baculovirus Autographa californica nuclear polyhedrosis virus. *Virology* 1991, 181, 336-347.
- [27] Tang, X. D., Xu, Y. P., Yu, L. L., Lang, G. J., *et al.*, Characterization of a Bombyx mori nucleopolyhedrovirus with Bmvp80 disruption. *Virus Res* 2008, 138, 81-88.
- [28] Durantel, D., Croizier, L., Ayres, M. D., Croizier, G., *et al.*, The pnk/pnl gene (ORF 86) of Autographa californica nucleopolyhedrovirus is a non-essential, immediate early gene. *J Gen Virol* 1998, 79 (Pt 3), 629-637.
- [29] Fang, M., Wang, H., Yuan, L., Chen, X., *et al.*, Open reading frame 94 of *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus encodes a novel conserved occlusion-derived virion protein, ODV-EC43. *J Gen Virol* 2003, 84, 3021-3027.
- [30] Song, J., Wang, R., Deng, F., Wang, H., Hu, Z., Functional studies of per os infectivity factors of *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus. *J Gen Virol* 2008, 89, 2331-2338.
- [31] Wolgamot, G. M., Gross, C. H., Russell, R. L., Rohrmann, G. F., Immunocytochemical characterization of p24, a baculovirus capsid-associated protein. *J Gen Virol* 1993, 74 (Pt 1), 103-107.
- [32] Whitt, M., A phosphorylated 34-kDa protein and a subpopulation of polyhedrin are thiol linked to the carbohydrate layer surrounding a baculovirus occlusion body. *Virology* 1988, 163, 33-42.
- [33] Li, L., Rohrmann, G. F., Characterization of a baculovirus alkaline nuclease. *J Virol* 2000, 74, 6401-6407.

- [34] Van Oers, M. M., Vlak, J. M., The baculovirus 10-kDa protein. *J Invertebr Pathol* 1997, 70, 1-17.
- [35] Faulkner, P., Kuzio, J., Williams, G. V., Wilson, J. A., Analysis of p74, a PDV envelope protein of *Autographa californica* nucleopolyhedrovirus required for occlusion body infectivity in vivo. *J Gen Virol* 1997, 78 (Pt 12), 3091-3100.
- [36] McCarthy, C. B., Dai, X., Donly, C., Theilmann, D. A., *Autographa californica* multiple nucleopolyhedrovirus ac142, a core gene that is essential for BV production and ODV envelopment. *Virology* 2008, 372, 325-339.
- [37] Braunagel, S. C., He, H., Ramamurthy, P., Summers, M. D., Transcription, translation, and cellular localization of three *Autographa californica* nuclear polyhedrosis virus structural proteins: ODV-E18, ODV-E35, and ODV-EC27. *Virology* 1996, 222, 100-114.
- [38] Schultz, K. L., Wetter, J. A., Fiore, D. C., Friesen, P. D., Transactivator IE1 is required for baculovirus early replication events that trigger apoptosis in permissive and nonpermissive cells. *J Virol* 2009, 83, 262-272.
- [39] Braunagel, S. C., Elton, D. M., Ma, H., Summers, M. D., Identification and analysis of an *Autographa californica* nuclear polyhedrosis virus structural protein of the occlusion-derived virus envelope: ODV-E56. *Virology* 1996, 217, 97-110.
- [40] Katsuma, S., Mita, K., Shimada, T., ERK- and JNK-dependent signaling pathways contribute to *Bombyx mori* nucleopolyhedrovirus infection. *J Virol* 2007, 81, 13700-13709.
- [41] Mishra, G., Chadha, P., Das, R. H., Serine/threonine kinase (pk-1) is a component of *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV) very late gene transcription complex and it phosphorylates a 102 kDa polypeptide of the complex. *Virus Res* 2008, 137, 147-149.
- [42] Westenberg, M., Uijtdewilligen, P., Vlak, J. M., Baculovirus envelope fusion proteins F and GP64 exploit distinct receptors to gain entry into cultured insect cells. *J Gen Virol* 2007, 88, 3302-3306.
- [43] Guarino, L. A., Identification of a viral gene encoding a ubiquitin-like protein. *Proc Natl Acad Sci U S A* 1990, 87, 409-413.
- [44] Reilly, L. M., Guarino, L. A., The viral ubiquitin gene of *Autographa californica* nuclear polyhedrosis virus is not essential for viral replication. *Virology* 1996, 218, 243-247.
- [45] Clem, R. J., Baculoviruses and apoptosis: a diversity of genes and responses. *Curr Drug Targets* 2007, 8, 1069-1074.
- [46] Wu, W., Lin, T., Pan, L., Yu, M., *et al.*, *Autographa californica* multiple nucleopolyhedrovirus nucleocapsid assembly is interrupted upon deletion of the 38K gene. *J Virol* 2006, 80, 11475-11485.
- [47] Oomens, A. G., Blissard, G. W., Requirement for GP64 to drive efficient budding of *Autographa californica* multicapsid nucleopolyhedrovirus. *Virology* 1999, 254, 297-314.
- [48] Xi, Q., Wang, J., Deng, R., Wang, X., Characterization of AcMNPV with a deletion of me53 gene. *Virus Genes* 2007, 34, 223-232.
- [49] Dry, I., Haig, D. M., Inglis, N. F., Imrie, L., *et al.*, Proteomic analysis of pathogenic and attenuated alcelaphine herpesvirus 1. *J Virol* 2008, 82, 5390-5397.
- [50] Johannsen, E., Luftig, M., Chase, M. R., Weicksel, S., *et al.*, Proteins of purified Epstein-Barr virus. *Proc Natl Acad Sci U S A* 2004, 101, 16286-16291.

- [51] Varnum, S. M., Streblow, D. N., Monroe, M. E., Smith, P., *et al.*, Identification of proteins in human cytomegalovirus (HCMV) particles: the HCMV proteome. *J Virol* 2004, 78, 10960-10966.
- [52] Hou, D., Zhang, L., Deng, F., Fang, W., *et al.*, Comparative Proteomics Reveal Fundamental Structural and Functional Differences between the Two Progeny Phenotypes of a Baculovirus. *J Virol* 2013, 87, 829-839.
- [53] Chertova, E., Chertov, O., Coren, L. V., Roser, J. D., *et al.*, Proteomic and biochemical analysis of purified human immunodeficiency virus type 1 produced from infected monocyte-derived macrophages. *J Virol* 2006, 80, 9039-9052.
- [54] Loret, S., Guay, G., Lippe, R., Comprehensive characterization of extracellular herpes simplex virus type 1 virions. *J Virol* 2008, 82, 8605-8618.
- [55] Shaw, M. L., Stone, K. L., Colangelo, C. M., Gulcicek, E. E., Palese, P., Cellular proteins in influenza virus particles. *PLoS Pathog* 2008, 4, e1000085.
- [56] Zhu, F. X., Chong, J. M., Wu, L., Yuan, Y., Virion proteins of Kaposi's sarcoma-associated herpesvirus. *J Virol* 2005, 79, 800-811.
- [57] Segura, M. M., Garnier, A., Di Falco, M. R., Whissell, G., *et al.*, Identification of host proteins associated with retroviral vector particles by proteomic analysis of highly purified vector preparations. *J Virol* 2008, 82, 1107-1117.
- [58] Radhakrishnan, A., Yeo, D., Brown, G., Myaing, M. Z., *et al.*, Protein analysis of purified respiratory syncytial virus particles reveals an important role for heat shock protein 90 in virus particle assembly. *Mol Cell Proteomics* 2010, 9, 1829-1848.
- [59] Resch, W., Hixson, K. K., Moore, R. J., Lipton, M. S., Moss, B., Protein composition of the vaccinia virus mature virion. *Virology* 2007, 358, 233-247.

Modularity and evolutionary constraints in a baculovirus gene regulatory network

Juliana Velasco de Castro Oliveira¹
Email: juliana.velasco@bioetanol.org.br

Anderson Fernandes de Brito¹
Email: andersonfbrito@gmail.com

Carla Torres Braconi¹
Email: cabraconi@usp.br

Caio César de Melo Freire¹
Email: caiomfreire@gmail.com

Atila Iamarino¹
Email: atila@usp.br

Paolo Marinho de Andrade Zanotto^{1,2,*}
* Corresponding author
Email: pzanotto@usp.br

¹ Department of Microbiology, Institute of Biomedical Sciences – ICB II, Laboratory of Molecular Evolution and Bioinformatics, University of São Paulo – USP, São Paulo, Brasil

² Departamento de Microbiologia, Instituto de Ciências Biomédicas - ICB II, Laboratório de Evolução Molecular e Bioinformática, Universidade de São Paulo - USP, Av. Prof. Lineu Prestes, 1374, São Paulo CEP: 05508-900, Brasil

Abstract

Background

The structure of regulatory networks remains an open question in our understanding of complex biological systems. Interactions during complete viral life cycles present unique opportunities to understand how host-parasite network take shape and behave. The *Anticarsia gemmatalis* multiple nucleopolyhedrovirus (AgMNPV) is a large double-stranded DNA virus, whose genome encodes for 152 open reading frames (ORFs). Here we present the analysis of the ordered cascade of the AgMNPV gene expression.

Results

We observed an earlier onset of the expression than previously reported for other baculoviruses, especially for genes involved in DNA replication. Most ORFs were expressed at higher levels in a more permissive host cell line. Genes with more than one copy in the genome had distinct expression profiles, which could indicate the acquisition of new functionalities. The transcription gene regulatory network (GRN) for 149 ORFs had a

modular topology comprising five communities of highly interconnected nodes that separated key genes that are functionally related on different communities, possibly maximizing redundancy and GRN robustness by compartmentalization of important functions. Core conserved functions showed expression synchronicity, distinct GRN features and significantly less genetic diversity, consistent with evolutionary constraints imposed in key elements of biological systems. This reduced genetic diversity also had a positive correlation with the importance of the gene in our estimated GRN, supporting a relationship between phylogenetic data of baculovirus genes and network features inferred from expression data. We also observed that gene arrangement in overlapping transcripts was conserved among related baculoviruses, suggesting a principle of genome organization.

Conclusions

Albeit with a reduced number of nodes (149), the AgMNPV GRN had a topology and key characteristics similar to those observed in complex cellular organisms, which indicates that modularity may be a general feature of biological gene regulatory networks.

Keywords

Baculovirus, Transcriptome, Real-time PCR, Gene regulatory network, Overlapping transcripts, Modularity

Background

Cellular pathways and gene regulatory networks (GRNs) are complex systems that emerge under natural selection and bare distinct evolutionary proprieties such as modularity. Modularity is an effective mechanism for keeping perturbations confined while preserving the complete system [1]. It is widely observed in various organisms and is possibly a fundamental biological design principle [2-4]. Emergent proprieties in networks also impose constraints on individual genes, mostly due to epistasis, leading to gene conservation. Important enzymes and other essential proteins – as reported by flux balance analysis [5,6] – tend to vary less than those under lower functional load [7], indicating that the flow of matter through metabolic networks expounds an evolutionary constraint imposed on components of any given pathway. These findings were confirmed for yeast, in which highly connected enzymes evolve more slowly than less connected ones [8]. Viruses intertwine their metabolic functions with those of the host cell, which allows infected cell to be understood as superorganisms, comprising the minimal essential condition for viral replication and biomagnification [9]; yet how viruses organize their gene network remains an open question.

Baculoviruses provide a rich environment where to investigate these relations, since these viruses have many completely sequenced genomes, each one encoding around one hundred plus open reading frames (ORFs) with transcription strategies and defined overall temporal regulation. The *Baculoviridae* constitute a family of invertebrate viruses that infect mainly insects of the order Lepidoptera, with large circular, covalently closed, double-stranded DNA genome [10,11]. The analysis of fifty-seven complete baculovirus genomes has shown 37 genes probably shared among all of them [12-17], most of which are involved in essential processes, such as replication, transcription and oral infectivity. The *Anticarsia gemmatalis* multiple nucleopolyhedrovirus (AgMNPV) has been used in Brazil and other countries to control the velvet bean caterpillar *Anticarsia gemmatalis* (Lepidoptera: Noctuidae), an

important pest of soybean crops [18,19]. The prototype AgMNPV isolate 2D (AgMNPV-2D) [20,21] has a genome of 132,239 bp, which may encode 152 ORFs [22].

Gene expression and replication cycle of baculovirus appear to be regulated by viral factors and the cellular milieu, where distinct gene classes are thought to be trans-activated (directly or indirectly) by closely related transcription complexes, causing expression to be controlled by distinct promoters, activated in a temporal concerted fashion during the infection cycle [10]. Three main transcriptional phases can be distinguished: *early* (immediate early and delayed early), *late* and *very late* [10]. The early phase precedes the onset of viral DNA replication and includes transcription of genes involved in host modulation, viral DNA replication and regulation of delayed early and late gene expression. The late and very late phases follow DNA replication and include the expression of genes required for virus assembly and occlusion (structural genes) [23].

An interesting feature of baculovirus transcription is that some ORFs can be transcribed in tandem, spanning up to seven units [24-26]. Specific genomic regions encode a wide variety of transcripts with different lengths, the so-called ‘overlapping transcripts’ [27-29]. Several tandem transcripts were mapped to different genomic regions of *Autographa californica* nucleopolyhedrovirus (AcMNPV) [27,30] and *Bombyx mori* nucleopolyhedrosis virus (BmNPV) genomes [31]. The expression profiles of six AgMNPV genes were investigated: *egt* [32]; *p10* [33]; *v-trex* [34]; *helicase* [35]; *iap-3* [36] and *p74* [37]. Nevertheless, there was no comprehensive study done so far on its complete gene content transcription and regulation during the replication cycle. Therefore, it is of interest to address the role of gene expression regulation in shaping baculovirus genome organization and evolution. Here we analyzed transcriptional organization of AgMNPV compared to randomized datasets in two permissive cell lines with different infection kinetics [38]. Our analysis focused on: (i) the viral transcriptome profile, (ii) the structural properties of the 5 gene regulatory network (GRN), (iii) genomic arrangement of transcripts of distinct viral cycle phases and, (iv) association of transcripts to distinct temporal promoter types.

Results

Viral genes are expressed sooner and at higher amounts in its permissive cell line

We performed a series of quantitative real-time PCR experiments to determine the expression profile of AgMNPV-2D 152 predicted ORFs [22], following infection of two permissive cell lines with different infection kinetics [38] at 0, 1, 3, 5, 7, 9, 11, 13, 24 and 48h post infection (p.i.). UFL-AG-286 was isolated from the natural host of AgMNPV [39], and IPLB-SF-9 was established from a different Noctuidae genus [40]. All pairs of amplicons for each gene were successfully amplified, sequenced and verified (see Methods). Except for ORFs 64 and 83, all remaining 149 ORFs were expressed in both cell lines (ORF 135 was excluded from the analysis due to nonspecific amplification). To help inspection, the average values (log 10) from three independent experiments are shown in Additional file 1: Table S1 (UFL-AG-286 cell line) and Additional file 2: Table S2 (IPLB-SF-9 cell line). Several ORFs reached significant levels of expression at different time points, reinforcing the expectation of a temporal structure in the viral cycle. At 7 h p.i., most genes were detected in both cells. Nevertheless, genes were expressed earlier in UFL-AG-286 cells than previously reported in literature for most baculovirus, including AgMNPV genes previously studied using Northern

blot and RT-PCR [32,33,35,36]. We compared the patterns obtained from the two cell lines considering the magnitude and the first detection (post-infection) of significant gene expression. Temporal differences in both cell lines were complex, with several ORFs being expressed at different relative times in each cell line. For instance, certain genes were expressed much later in the IPLB-SF-9 cell line (such as *odv-e27* (ORF 140)) and other genes were expressed in the same time post-infection (such as *lef-1* (ORF 19)), or with a delay of 1 h such as *cg30* (ORF 85), which was expressed at 0 h in UFL-AG-286 and 1 h later in IPLB-SF-9. We observed that most genes were more expressed in the UFL-AG-286 cell line than in the IPLB-SF-9 (Figure 1), reaching in some instances more than fifty-fold increase (such as, *39k/pp31* (ORF 25)), with the exception of *pe38* (ORF 149) that was expressed at higher levels in IPLB-SF-9 cells. Furthermore, differences in expression intensity between cell lines decreased during infection, to the extent that at 48 h p.i., *ag4*, *ag20*, *ag44* and some others were more expressed in IPLB-SF-9 than in UFL-AG-286. In sum, the relative temporal pattern for several genes changed depending on the cell line (Additional file 2: Tables S2 and Additional file 3: S3).

Figure 1 Relative differences of mRNA expression during infections in UFL-AG-286 and IPLB-SF-9 cell lines. Median values of mRNA expression of genes associated from left to right: (i) DNA replication, (ii) transcripton, (iii) structural proteins, (iv) genes with unknown function and, (v) early genes. The asterisk (*) indicates gene groups that had significantly difference between cell lines ($p < 0,05$).

The inferred gene regulatory network has non-random proprieties

Since we had three independent replicates of infection experiments lasting 48 hours each, we used a single 144 hrs time series to infer a putative AgMNPV-2D gene regulatory network (GRN) from 149 ORFs unfolded during infection in UFL-AG-286 and likewise, a single 144 hrs time series for IPLB-SF-9. Under the assumption of a recurrent temporal regulation program for viral gene expression, this approach should maximize true gene-to-gene cross-correlation signal present in the data, while mimicking three successive infection cycles in a row. Ranked lists of directed node interactions were generated with GENIE3 from expression data of AgMNPV-2D infecting UFL-AG-286, infecting IPLB-SF-9 and the respective randomized data (UFL-AG-286rand and IPLB-SF-9rand). One thousand links were necessary on average to assemble a single connected component with 149 nodes for all GRN (including those generated from random datasets). Nevertheless, the complete network for AgMNPV-2D on UFL-AG-286 needed 1120 edges to include all nodes into a single component, which was in agreement with similar but random GRNs, which had an average of 1030 edges. Preliminary exercises with random GRNs revealed that the data generated from our complete lists of directed node interactions generated with GENIE3 lacked essential complex features of known biological networks such as modularity. Since we needed more realistic null models to better discern features of real networks from the data, we decided to keep only the relevant links that would be sufficient to generate a single component network while generating randomized networks by shuffling the real data time points. This procedure obliterated the time structure of the real data but kept its value range and sampling density.

The network architecture displays increasing modularity

The ‘most important’ links were pruned using the maximum neighborhood component (MNC) and the density of maximum neighborhood component (DMNC) algorithms with *Hubba*, down to 765 (UFL-AG-286), 991 (IPLB-SF-9), 713 (UFL-AG-286rand) and 727

nodes (IPLB-SF-9rand), still maintaining a single large component yielding a ratio of around 5 edges per node. A model of the reduced AgMNPV-2D GRN in UFL-AG-286 cells (Figure 2) depicts few communities of highly interconnected nodes organized around a single large component, where the importance of links caused nodes to come closer together (force-directed layout). We inspected the topological features and attributes of this AgMNPV-2D GRN with the Network Analysis plug-in in Cytoscape and observed that the real data GRN had a trend for the increase of average clustering coefficient (ACC) distribution as the number of neighbors of a node increased, while a negative dependence tendency was observed on the scrambled expression data (Additional file 4: Figure S1). This trend was more pronounced in the infection in UFL-AG-286 cells (Additional file 4: Figure S1A and S1B). The average connectivity distribution of all neighbors of a node was also significantly different in real and scrambled data (Additional file 4: Figure S1). As the number of neighbors increased, real data GRN had a larger number of neighbors with higher ANC with positive increase of ANC (Additional file 4: Figure S1). We also observed that the GRN obtained infecting UFL-AG-286 differed from all others by having two nodes with stress centrality values one order of magnitude greater than that observed for both real and scrambled data: (i) *egt* (ORF 18) with SC = 1686 and, (ii) *gp64* (ORF 124) with SC = 1551, followed by *ag75* and *ag147*. Visual inspection of the force-directed GRN layout reveals that these four nodes appear to interconnect adjacent modules in the GRN (Figure 2). The GRN in IPLB-SF-9 shared several basic features with that unfolding in UFL-AG-286. However, for the sake of brevity details on the reduced modularity (possibly due to reorganization of the GRN under perturbation), will be dealt with elsewhere. Our results expounded a genuine modular architecture not caused by chance in the AgMNPV-2D GRN and they also indicated that this modularity was more pronounced in the infection in UFL-AG-286.

Figure 2 Gene regulatory network (GRN) of AgMNPV-2D infecting UFL-AG-286 cells generated with Cytoscape v2.8.0. A single component network shows the directed links (arrow edges) among 149 ORFs. Nodes (ORFs) have sizes proportional to their betweenness centrality (BC) and darkness proportional to stress coefficient (SC), both of which indicate node importance in the GRN. Five communities (I-V) were identified with the GLay clustering method, which indicate modules of highly interconnected nodes. Numbers correspond to ORFs in the AgMNPV-2D genome (see Additional file 1: Table S1) and the 37 genes shared among all baculoviruses are colored in red.

It is of utter importance to emphasize that node linkages in our GRNs depict the relationships *among regulatory elements* inferred from gene expression profile (transcriptional data). Hence, it has to be made clear that the networks we inferred convey the implied associations among transcription regulation elements *per se*, and not necessarily gene product function relationships (such as in an interactome). Fittingly, genetic regulatory functions are best understood when viewed in terms of *intergenic linkages of diverse modalities*, and these non-linear functions are not visible at the level of any individual gene [41].

The gene communities (modules) have redundant functions

We used the GLay clustering method in Cytoscape to detect ‘gene communities’ (modules) in the GRN. The modularity (Q) in UFL-AG-286 was 0.56 and we found 5 communities adding to a total of 662 edges. This value ($Q = 0.56$) was significant since it was higher than the 99% upper value for the confidence interval ($Q = 0.50$) obtained for the other GRNs ($Q = 0.46$ for UFL-AG-286rand, 0.41 for IPLB-SF-9 and $Q = 0.46$ for IPLB-SF-9rand). We then observed that each one of the five modules was populated differently when we ordered the ORFs in six

functional categories [42,43]: virion and capsid (structural), DNA replication, transcription, host modulation or auxiliary (Figure 3). Community I gathered most of genes associated with structural functions (23 of 36; $\chi^2 = 14.002$, $p = 0.0002$; d.f. = 1) while, more than 85% of its nodes (48 of 56; $\chi^2 = 27.572$, $p < 0.0001$; d.f. = 1) are conserved in almost all sequenced group I alphabaculoviruses (including 13 shared genes among all baculoviruses). Community II was the most densely interconnected module ($\chi^2 = 7.030$, $p = 0.008$; d.f. = 1), having one third of its genes (15 of 45 nodes; $\chi^2 = 16.659$, $p < 0.0001$; d.f. = 1) also conserved among group I. This community had the larger number of nodes associated with DNA replication (7 of 12; $\chi^2 = 4.900$, $p = 0.0269$; d.f. = 1) and had slightly more (albeit statistically insignificant) members with auxiliary functions (13 of 30; $\chi^2 = 3.073$, $p = 0.0796$; d.f. = 1). Community III was homogeneously populated with nodes from different functional categories. Community IV consisted mainly of genes without assigned functions (8 of 14; $\chi^2 = 4.689$, $p = 0.0304$; d.f. = 1). Community V had only four ORFs and one of them was the *helicase*. Interestingly, different functional classes were dispersed among different node communities, which is suggestive of redundancy.

Figure 3 Modular structure of AgMNPV-2D network unveiled with GLay clustering method. The five different communities are ordered by their size and are populated with some particular distinctions. The comparative amount of gene function in each community is represented by colored rectangles. Community **I** is densely populated with subsets of nodes associated with structural functions; community **II** has subsets of nodes more related to DNA replication and auxiliary functions; community **III** is homogeneously populated with nodes from different functional categories; community **IV** includes diverse genes without assigned functions; and community **V** is the smallest, where *helicase* (ORF 92) is located. Nodes represent ORFs of AgMNPV-2D and have colors according to the functional classification based on prior knowledge about baculovirus genes. Nodes also have different size representing the importance (betweenness centrality) of the ORF in the complete GRN (see Figure 2).

Gene function impacts on expression level variation and on GRN topology

We further investigated how the GRN architecture related to other features of nodes that were not considered during the inferential process. This is because the GRN we inferred is based on RNA expression through time and should reflect, to a greater extent, mostly the temporal program of gene expression. Therefore, given that promoter elements are known to have considerable temporal specificity, we first searched promoter motifs upstream of the initiation codon of each ORF and mapped as attributes in the five main communities (Figure 4). Within this region, 50 (33.56%) ORFs had a late motif, 23 (15.44%) contained an early motif, 42 (28.19%) had both early and late, and 34 (22.81%) lacked any recognized motif (Figure 4). The average clustering coefficient of early genes was higher for 20 genes with early promoters (ACC = 0.2195) than for 50 genes with only typical promoter elements (ACC = 0.1768), although not statistically different at a significance level of $\alpha \geq 0.05$. Community I had mainly nodes associated with late motifs (45 of 56 nodes with *late* and *early & late* motifs; $\chi^2 = 12.311$; $p = 0.0005$; d.f. = 1) and a remarkable and significant absence of ORFs with upstream unique early promoters (5 of 56 nodes; $\chi^2 = 40.898$, $p < 0.0001$; d.f. = 1). Community II had a large number of genes with unknown promoters (15 of 45; $\chi^2 = 4.047$; $p = 0.0442$; d.f. = 1).

Figure 4 Map of promoter types of AgMNPV-2D in UFL-AG-286 network. Nodes represent ORFs of AgMNPV-2D and have colors according to the different promoter motifs

(see text). The comparative amount of promoter motifs in each community is represented by colored rectangles. Community **I** is mainly populated by genes with late promoters motifs; community **II** is densely populated by ORFs expressed from unknown promoters; community **III**, **IV** and **V** have diverse promoter types. Nodes also have different size representing the importance (betweenness centrality) of the ORF in the complete GRN (see Figure 2).

Given the apparent assortativity (that is, the preferential attachment of nodes by shared properties) of both gene function and promoter types, we then compared expression profiles among genes with correlated functions, such as: (i) replication: *dnapol*, *lef-1*, *lef-2* and *helicase*; (ii) transcription: *lef-4*, *lef-5*, *lef-8*, *lef-9*, *p47* and *vlf-1*; (iii) structural proteins: *pif-1*, *pif-2*, *gp41*, *odv-e56*, *odv-ec27*, *p74*, *p6.9*, *vp91*, *vp39*, *vp1054* and *p33*; (iv) core genes with unknown function; (v) some immediate-early genes that do not belong to the 37 conserved genes (*cg30*, *ie-0*, *ie-1*, *ie-2*, *me53* and *pe38*); and carried out the statistical analysis (ANOVA) among groups followed by Tukey's multiple comparison test, assuming a significance level at $p < 0.05$. We noticed that early genes and those associated with DNA replication had less difference in their temporal and quantitative profiles between the two cell lines (Figure 1). This result suggested that core conserved functions do appear to have more invariance in the temporal structure of their expression program. Crucially, this temporal consistency was also associated with a trend on the association of a key GRN topologic feature and gene variability, since core functions (replication and capsid) had a higher median betweenness centrality ($BC = 0.0019$, Additional file 5: Figure S2) with a large variance and a significantly lower genetic diversity ($\theta = 0.42 \pm 0.035$), while satellite (all the others) functions had lower median betweenness centrality ($BC = 0.0013$) with large variance and a significantly higher genetic diversity ($\square = 0.47 \pm 0.044$, Additional file 5: Figure S2).

The spatial organization of regions of overlapping transcription (ROT) is important and conserved

To better infer the relation between time expression and gene position, we built a syntenic map for the complete genome of thirteen Group I alphabaculoviruses (Additional file 6: Figure S3 and Additional file 6: Table S3) with twenty-two highly conserved 'regions of overlapping transcription' (ROT) defined as multiply transcribed genomic loci encoding tandem RNA transcripts of different lengths. These ROTs had almost the same overall order and relative orientation in all genomes, even when genomic reorganization events, such as inversions and gene deletions took place. When compared to other genomic regions, for all 13 genomes investigated, we found in ROTs a higher frequency of ORFs with very short or absent intergenic spacers (overlapping genes) ($\chi^2 = 65.041$, $p < 0.00001$; d.f. = 1). If ROTs entail the synthesis of overlapping transcripts in alphabaculoviruses, it is possible that approximately 40% of the AgMNPV-2D genes (61 of 152) could be transcribed in tandem.

Given that some adjacent co-transcribed genes belonging to ROTs grouped in the same community, as expected if these genes shared similar transcription regulation, we investigated if the potential expression of tandem RNA transcripts could influence the regulatory association of nodes in the GRN. We observed that, at a coarse-grain, the bulk of physical distances among AgMNPV-2D genes, expressed in map units, did not show extensive positive correlation with Euclidean distances among expression profiles of each individual ORF (Additional file 7: Figure S4), indicating that the association of gene expression profile is not explained by the tandem organization of ROTs alone. Nevertheless, fifteen of the 37 shared genes among all baculoviruses [12] are within ROTs, shown as asterisks over each physical map in Additional file 6: Figure S3. Moreover, among 56 nodes

from community I (Figure 3), 20 belonged to ROTs number 1, 13, 14, 15, 17 and 19 (Additional file 6: Figure S3 and Figure 5A). Thus, one third of genes potentially expressed in tandem were neighbors in the GRN. Furthermore, the two-fold co-occurrence of five community I members in tandem, as observed in ROTs 17, 19 to 21 (in bold in Figure 5) was estimated to be low probability events ($p = 0.001$ each) during Monte Carlo simulations (Additional file 8: Table S4). Among all 45 nodes in the community II, 14 represent genes arranged in pairs throughout the genome (Figure 5A). Nonetheless, just one pair of genes from ROT 2 is probably transcribed in tandem, while ORFs were scattered all over the genome in the communities III, IV and V (Figure 5A), as would be expected by chance (Additional file 8: Table S4). We also found that early, late and unknown promoter motifs (Figure 5B) appear to be interleaved with some degree of heterogeneous distribution in the genome. For example a cluster of 6 ORFs with late motifs in ROTs 10 to 12 (in bold in Figure 5) had a low chance probability ($p = 0.000735$) (Additional file 9: Table S5), while the chance of observing a triple tandem of early and late motifs was higher, ($p = 0.0211$). Nevertheless, small ORF sets containing similar promoter motifs were arranged contiguously, and most of them coincide with the 22 ROTs we found, which have mostly ORFs with late or early and late promoter motifs. We also investigated how genes involved in the same biological process or with similar functions did cluster in the genome. As shown in Figure 5C, the majority of genes from the same functional category were not adjacent but rather interspersed sense-wise. Nonetheless, the occurrence of tree genes coding for virion proteins in the positive strand in ROTs 19 to 21 had a low chance probability ($p = 0.00216$), while the co-occurrence of 3 genes encoding transcription associated products in the positive strand in ROT 5 (in bold in Figure 5) had a even lower chance probability ($p = 0.00087$) (Additional file 10 Table S6).

Figure 5 Physical and functional organization of the AgMNPV-2D genome. The outer and the inner ring represent ORFs in positive and negative sense, respectively. The outer numbering means the AgMNPV-2D ORF numbers (Additional file 1: Table S1), and the inner numbering indicates the loci of the 22 regions of overlapping transcription (ROTs)(for further information, see Additional file 3: Table S3). ROTs with co-occurrence events are shown in bold. ORFs colored in black represent those whose expression profile was not included in the present work (ORFs 64, 83 and 135). **A.** Genomic map relating the physical position of genes to their distribution among the five regulatory modules (communities) of the GRN. **B.** Genomic map showing the gene promoter motifs arrangement in the AgMNPV genome. **C.** Genomic map presenting the genome organization according to gene functions.

Discussion

Expression of structural and regulatory genes vary more on less permissive cells

The use of two different infection kinetics allowed us to observe contrasting aspects of the transcriptional program of AgMNPV-2D *in vitro*, providing comprehensive, time-related information on the transcription dynamic and quantitative information on the expression of 149 of its 152 predicted ORFs. The ‘sooner and higher’ expression pattern in UFL-AG-286 compared to IPLB-SF-9 could be explained by the different states of adaptation of the virus to distinct cellular environments, since the infection in IPLB-SF-9 (a less permissive cell line) could be understood as a ‘perturbation’ of the AgMNPV-2D GRN. Nevertheless, we have not noticed significant difference in most genes related with DNA replication, such as *dnapol*

(ORF 65), *lef-1* (ORF 19) and *lef-2* (ORF 3) that were expressed at the same time and at similar levels in both cell lines. While replication-associated genes had similar expression, most structural genes had a significant delay in IPLB-SF-9 and notable differences in the magnitude of expression (Additional file 1: Table S1 and Additional file 3: Table S3). Hence, the expression profiles of the structural genes found in IPLB-SF-9 could help explain the delay in viral morphogenesis routinely observed in our laboratory that was also reported for IPLB-SF-21 cells [38].

We consistently observed earlier expression for most of the genes than expected based on the observed for other baculoviruses. For example, the *p6.9* gene (ORF 96) was reported as late, expressed at 12 h p.i. in the AcMNPV [44]. P6.9 is a small protein (6.9 kDa), basic, arginine-rich and it is involved in the condensation of viral DNA. It is not surprising to find this gene expressed earlier in the AgMNPV-2D viral cycle, since DNA replication begins at 2 h p.i. [38] and P6.9 is known to be necessary to package and stabilize the newly synthesized viral genome [45]. In addition, Iwanaga and colleagues [46] also found *p6.9* mRNA and other transcripts earlier in infection, which corroborates our results and earlier observations on other DNA viruses, such as FV-3 and CIV [47-49]. Crucially, the difference we observed could be due to the use of real-time PCR, which is much more efficient and requires fewer targets than hybridization techniques (such as Northern blot), detecting low amounts of RNA initially present in each sampling time. This precocity could simply reflect differences between baculoviruses, given the earlier onset of DNA replication in AgMNPV-2D (2 h p.i.) compared to others, such as AcMNPV (5 h p.i.) [50], SfMNPV (8 h p.i.) [51] and LdMNPV (18 h p.i.) [52].

Furthermore, we also noticed that genes with more than one copy in the genome (such as *iaps*, *pifs* and *ptps*) had distinct expression profiles. For example: (i) *iap-3* (ORF 34) was expressed earlier in UFL-AG-286 cells; (ii) *iap-2* (ORF 70) simultaneously, whereas; (iii) *iap-1* (ORF 40) showed expression only later, in both cell lines, being more expressed than *iap-2* and *iap-3*. Gene duplication and redundancy are important mechanisms that allow adaptive evolution of genomes, and represent 8-20% of the gene content in eukaryotes [53]. We would argue that this allows one of the copies of a gene to evolve with lesser functional constraint [54], which eventually could help it acquiring a new functionality (such as, *iap* and *bro* genes).

Conserved genes have higher centrality in a modular AgMNPV-2D GRN

By comparing with our randomized models the inferred GRN in UFL-AG-286 cells, where five communities of nodes were identified, it is notable that it had a significantly different modular structure. This was made clear by the trend in both the average connectivity and the average coefficient distribution values for nodes as function of the number of neighbors they had (Additional file 4: Figure S1). By looking at the AgMNPV-2D GRN as a single component we observed that nodes with highest importance, both in terms of BC and SC, included relevant host-interaction functions such as *gp64* (ORF 124) and *egt* (ORF 18). The *gp64* codes for an envelope glycoprotein paramount for effective membrane fusion during the secondary infection that is highly conserved in the group I nucleopolyhedroviruses (NPVs) [55]. Interestingly, the expression of the *egt* gene in *Lymantria dispar* multiple nucleopolyhedrovirus (LdMNPV) was shown to be the genetic basis of climbing behavior and tree top disease in silkworm infected with baculovirus, which facilitates preferential predation of sick larvae while maximizing viral spread in the environment [56].

A relevant property of our inferential procedure was that there was a positive correlation of reduced genetic diversity (θ) and importance of the gene in the GRN, as expected for key components of metabolic networks [7,8]. Core conserved functions also had more temporal consistency of expression and distinct topologic features in the GRN. These coherent results support our modeling effort, since they constitute *independent evidence* for a relationship among phylogenetic data of baculovirus genes and complex GRN attributes inferred from expression data alone.

The observed modular heterogeneity that separates key genes that are functionally related on different communities could help accommodate redundancy and shield important functions, therefore maximizing GRN robustness by compartmentalization [57]. For example, genes possibly associated with anti-apoptotic function (*iap* genes) [58] were well distributed among three communities (I, II and IV), while *dnapol* and *helicase* nested in different communities (III and V). Likewise, assigning several different functions to a given module would also provide robustness by increasing functional redundancy [59]. Nevertheless, some functional associations were evident. Community I was highly populated by structural proteins associated with capsid and virion formation and promoters with *early* & *late* motifs (Figure 4). Interestingly, in the AgMNPV-2D 48 of 56 nodes in community I (13 of which are core genes) are conserved in almost all group I alphabaculoviruses sequenced so far. This feature was also observed in community II, in which one third of the nodes consisted of highly conserved genes, many of which were genes associated with DNA replication. The clustering of genes in community II with unidentified regulatory motifs could entail promoter motifs not yet described [60], which is made more compelling giving the high connectivity among its members. Spatially or chemically isolated modules composed of several cellular components and carrying discrete functions are considered fundamental building blocks of cellular organization, but their presence in highly integrated biochemical networks lacks quantitative support. In a key comparative study, the metabolic networks of 43 distinct organisms were shown to have modular organization and in *E. coli*, the hierarchical modularity had a good correlation with metabolic functionality [61]. Spatially or chemically isolated modules composed of several cellular components and carrying discrete functions are considered fundamental building blocks of cellular organization, but their incidence and generality needs to be better investigated across a wider spectrum of organismic complexity, such as in viruses.

Unfortunately, at this time we are not able to wire cell-coded nodes in the AgMNPV-2D GRN. Nevertheless, we would argue that they should have not only a crucial role in the viral cycle, but also should be connected to the real viral GRN [9]. However, viral GRNs encompass both the cellular and viral encoded functions necessary to complete the viral multiplication cycle [9]. Accordingly, preliminary data on RNA suppression subtractive hybridization (SSH) experiments (Oliveira *et al.*, in preparation) indicated that at least 50 cellular distinct genes were down regulated during infection at 24 h.p.i. (Additional file 11: Table S7). By using both Gene Ontology (<http://www.geneontology.org/>) and EGene (<http://www.coccidia.icb.usp.br/egene/>) programs while using the eggNOG2.0 database (<http://eggno2.embl.de/>), we found sequences coding mainly for proteins related to metabolism or protein modification (33.9%), ion transport and energy (12.0%), nucleic acids metabolism (11.3%) and several other functions at a lower frequency. But importantly, we found cellular genes at lower frequency, which are of direct relevance to viral host interaction such as: (i) the significantly hypo-expressed anterior fat body protein (AFP), (ii) defensin and, (iii) heat shock 90 (*hsp90*) (Additional file 11: Table S7). At this juncture, it is hard to assess the extent at which the hypo-expression of cellular genes does not simply reflect

cellular degradation due to viral disruption of cellular homeostasis, or is due to interactions with viral gene products. Notwithstanding, these preliminary results could imply that at least 50 additional cellular genes should be wired to the viral GRN for a proper description of the viral life-cycle.

The conserved genome architecture of Group I baculovirus has implications for gene expression

The presence of highly conserved ROTs in closely related baculoviruses, observed in our sinteny analyses, suggests a complex pattern of gene expression regulation, possibly also present in other baculoviruses [28,31]. The intergenic spacers among ORFs in ROTs are either very short or absent, and overlapping genes are overwhelmingly located in these genomic regions ($\chi^2 = 65.041$, $p < 0.00001$; d.f. = 1). This would provide a competitive advantage, given that keeping genomes short would speed up genome replication at the cost of maximizing interdependence among superimposed genes [62]. Interestingly, we also found evidence that the virus tries to keep similar attributes dispersed around the genome, mainly avoiding their tandem organization (Additional file 9: Tables S5, Additional file 10: Table S6 and Additional file 11: Table S7) with a few notable exceptions. A simple explanation for this observation would be that by avoiding clusters of similar attributes, the genome maximizes robustness of the viral GRN. Moreover, whilst ROTs are somehow conserved among group I NPV, our findings suggest that physical proximity is not a reliable predictor of temporal patterns of gene expression, since in the AgMNPV-2D there was no correlation between the position of genes in the genome and their temporal expression (Additional file 7 Figure S4). These results are in line with the possibility that tandem transcripts from ROTs in baculovirus are not necessarily translated at once [24-26]. Moreover, long mRNAs of several genes have been observed late in infection [27,28,31,63].

The region comprising from ORF 129 (*p24*) to 133 (*alk-exo*) in AcMNPV encodes for diverse tandem transcripts [27], which share the same T-rich 3' end termination site [64]. These ORFs are transcribed as monocistrons at the onset of infection (up until around 2 h p.i.) and later are transcribed in tandem [27]. The synthesis of long transcripts has been related to regulatory functions, such as promoter occlusion [65] and RNA interference (RNAi) [66,67]. Promoter occlusion is caused by RNA polymerase transcriptional complex activity during the transcription of upstream genes, blocking the monocistronic expression of downstream genes [65]. The expression of antisense sequences in tandem RNAs has been suggested as a mechanism of down-regulation [68,69]. Moreover, it has been assumed that overlapping transcripts should have a functional relevance for regulation of gene expression [27-31]. Possibly, monocistronic expression is sufficiently high as to hamper mRNA detection from tandem transcripts, which would help explaining the lack of correlation shown in Additional file 7 Figure S4. Fittingly, it has been suggested that regulatory elements functioning as 'internal ribosome entry sites' (IRES) are functional in the IPLB-SF-9 cell line [70]. Furthermore, the capsid protein VP1054 is probably synthesized by IRES-mediated translation [71], since its monocistronic transcript was not found in AcMNPV and BmNPV gene expression programs [31].

Conclusions

The transcriptome of the AgMNPV-2D indicated that conserved viral functions showed expression synchronicity, higher betweenness centrality and reduced genetic diversity,

consistent with evolutionary constraints present in complex biological systems [7,8]. Most ORFs were expressed at ‘sooner and higher’ levels in a more permissive cell line. In 13 group I alphabaculovirus genomes, we found 22 highly conserved regions of overlapping transcription, whose extended 3’ UTR could play a role in gene expression regulation [66]. The inferred GRN had a modular topology, comprising five communities of highly interconnected nodes, consisting of genes with different functions, promoter motifs, and physical location in the genome. This modular heterogeneity was suggestive of architectural redundancy that could promote robustness [57,59]. The fact that a similar architecture was observed on a simpler GRN, such the one we report for the AgMNPV-2D, tends to support the idea that hierarchical modularity may be an extensive and generic self-similarity feature of system-level biological organization [61].

Methods

Primer design and validation

Specific primer pairs for each ORF annotated in the AgMNPV-2D genome [22] were designed using Oligo Analysis Software v. 6.8 (Molecular Biology Insights) (Additional file 12: Table S8). Each pair was tested by PCR amplification and sequencing. Consensus sequences were verified by BLAST searching to check cross-hybridization against each other and the AgMNPV-2D genome. After this initial validation step, each pair of primers was further tested in real-time PCR reactions to ensure its specificity as determined by a unimodal-melting curve. After the specific amplification was confirmed, primer pairs were also validated by five points dilution (10^{-2} , 10^{-4} , 10^{-6} , 10^{-8} and 10^{-10} or 10^{-3} , 10^{-4} , 10^{-5} , 10^{-6} and 10^{-7}) amplification by real-time PCR. Reaction efficiency of 0.9 and the correlation coefficient (r^2) values of 0.99 were assumed as minimal quality standards for data acquisition and melting curves were also analyzed to confirm specificity. ORF 135 (*bro-h*) was not included in the study because there were unspecific products for three alternative primer pairs designed inside and upstream of its coding region.

Cell lines and infection kinetic: virus infection, RNA extraction and reverse transcription

The UFL-AG-286 [39] and IPLB-SF-9 [72] cell lines were maintained in Grace medium (Gibco) supplemented with 10% fetal bovine serum (Gibco). These permissible cell lines were used since AgMNPV has different infection kinetics in them [38]. UFL-AG-286 was derived from embryonic tissue of the lepidopteron *Anticarsia gemmatalis* [39], which is natural host of AgMNPV [19,21,73-76], while IPLB-SF-9 is a clone of IPLB-SF-21 cell line, established from pupal ovaries of *Spodoptera frugiperda*, another Noctuidae species from a different genus [40]. The virus strain AgMNPV-2D [77] was multiplied only in UFL-AG-286 cells, and both cell lines were used in the kinetic infections. Cultures of 10^5 UFL-AG-286 e IPLB-SF-9 cells were infected at a multiplicity of infection (MOI) of 10 and incubated at 28°C for 1h to allow infection synchronization. Unattached virions were then removed and fresh culture medium was added to the infected cells. Mock-infected samples were treated in the same manner as virus-infected cells, but with fresh culture medium instead of viral suspension. All experiments were done in triplicate. Total RNA was extracted from infected and mock cells at 0, 1, 3, 5, 7, 9, 11, 13, 24 and 48h p.i. using RNeasy Plus Kit (Qiagen) according to the manufacturer’s recommendations. RNA samples were subjected to DNase I treatments (DNA-free, Ambion) according to the manufacturer’s protocols. 100 ng of each

RNA sample was reverse transcribed using both SuperScript III Reverse (Invitrogen) and High-Capacity cDNA Archive kit (Applied Biosystems) using oligo (dT) and random primers (respectively) to allow maximum conversion of mRNA into cDNA. The cDNAs were mixed and used in the real-time PCR reactions.

Real-time PCR

We used real-time PCR to quantify mRNAs concentration during infection in cell cultures. This was done because of its accurate qualitative and quantitative measurements of the amount of transcripts with high sensibility and reproducibility [78,79], outperforming microarray in these regards [80]. Reactions were performed along with the points of the standard curve and the cDNA samples from kinetic infections, to keep the same conditions for standards and experimental samples. Reactions mixes contained 1.0 µl of cDNA from each time point (or point dilution), 7.5 µl of iQ SYBR Green Supermix kit (Bio-Rad) and 0.5 µM of each primer, in a final volume of 15 µl. The cycling conditions were: 2 min at 96°C, followed by 40 cycles of 30 s at 96°C, 30 s at 50°C /52°C /54°C /57°C (depending on the T_m of primer pair), and 40 s at 72°C. Melting curve analysis was performed increasing the temperature of the last cycle (72°C) until reaching 96°C, 1°C per cycle, 5 s at each cycle. Amplification, detection, and data analysis were performed using the Rotor-Gene 3000 systems (Corbett Life Science).

Promoter analysis and syntenic regions in regions of overlapping transcription (ROTs)

Promoters were searched in regions comprising 200 bp upstream of the putative start codon of each ORF by screening for known promoter elements. We chose 200 bp since this was determined by several independent studies to harbor all known late promoters (within the first 80 bp upstream from the first ATG) and most known functional early promoters of baculovirus genes. Early promoter (E) indicates a TATA box sequence (TATA or TATAWWW, W= A or T) followed by a CANT motif downstream, transcribed by the host RNA polymerase II [81]. To screen the late promoter motifs transcribed by the viral encoded RNA polymerase, the conserved pattern used was DTAAG [23,82]. Moreover, to investigate the genomic organization of clusters of genes with similar transcriptional profiles that we found in the AgMNPV, we did complete genome overlays including 13 group I alphabaculovirus with the Artemis Comparative Tool [83], while focusing on the conservation of syntenic regions of overlapping transcription (ROTs) that were first described in AcMNPV [28] and BmNPV [31].

GRN inference

A conceptual problem inherent to transcriptional data analyses is the assumption that gene expression dynamics is linear, deterministic and constant, while in fact it is mostly non-linear, infrequent and random [84,85]. Expression of genes most often happens as pulses of indeterminate duration [86], and depends on the interplay of different biological components (transcription factors, promoters motifs, RNA polymerases, splicing factors, etc.) at a given temporal (*i.e.*, developmental phase) and spatial (*i.e.*, tissular) context [85]. Moreover, levels of mRNA expression and its encoded protein translation are not necessarily correlated [87], while remarkable differences are observed in mRNA half-life [88]. Notwithstanding, clustering methods were used to analyze baculovirus gene expression transcriptional data [46,60,89]; but because transcript numbers experience iterative exponential fluctuations in time during successive infection cycles, gene expression is better understood as a non-linear

system [84,85], taking place on a log-scaled attractor [90-92]. We argue that these properties make simple least-square distance estimates among individual gene expression time-series uninformative and biased, being inadequate for process description [26,93]. Therefore, we used a prediction algorithm for the AgMNPV-2D gene regulatory network that is based on multiple regression and tree-based ensemble methods using our gene expression data (Additional file 2: Table S2) [94]. We chose the method implemented in GENIE3 instead of clustering because it: (i) was shown to perform well on reconstructing complex GRNs; (ii) does not make any assumption about the nature of gene regulation; (iii) infers directed GRNs; (iv) can deal with combinatorial and non-linear interactions among genes and; (v) provides a confidence ranking of regulatory links. Ranked lists of regulatory links were calculated with GENIE3 loaded in the free software environment for statistical computing and graphics R v.2.10.0 [95], then uploaded the ranked list output in Cytoscape 2.8.0 for network visualization and analysis [96]. To evaluate the importance of node connection in the modeled GRN, we used graph-theoretic methods implemented in the Hub Objects Analyzer (*Hubba*) plug-in for Cytoscape, which were shown for example to recover known yeast interactome data with greater than 70% accuracy [97]. With *Hubba*, we used Maximum Neighborhood Component (MNC) and Density of Maximum Neighborhood Component (DMNC) algorithms. This approach was helpful to obtain a minimal set of most important edges necessary to fully connect all viral nodes in the GRN.

Measuring complex network attributes

Modular networks have subsets of nodes that are densely connected within and weakly connected between subsets [61,98]. Therefore, betweenness centrality (BC), which estimates the number of shortest paths from all vertices to all others that pass through a node, is a more informative measure of topologic importance of a node than connectivity (k), which indicates the number of connections of a node irrespective of its neighbor's connectivity properties. Likewise, we also estimated other relevant topologic attributes that inform on network architecture such as: (i) the average clustering coefficient distribution (ACC) that gives the average of the clustering coefficients for all nodes with all quantities of neighbors, helping to find modularity in networks; (ii) the average neighborhood connectivity (ANC) that is the average connectivity of all neighbors of a node, which also informs on the sub-structuring of a network; (iii) stress centrality (SC) [99,100] that is the number of shortest paths passing through a node and; (iv) modularity (Q), which measures the number of edges falling within groups minus the expected number in an equivalent network with edges placed at random, quantifying the level of network compartmentalization in communities of nodes. While comparing genomic and GRN organizations we also investigated if three attributes (GRN communities, promoter motifs and functional categories) were spatially organized in the AgMNPV-2D. Hence, we estimated their probability of arbitrary co-occurrence by a Monte Carlo (MC) procedure, in which the attributes were randomized 100 million times without replacement, using a PERL script available from the authors upon request. During our randomization exercises, the co-occurrences were incremented when an attribute randomly sampled from the list was identical to the previously sampled. Thus, the chance to find co-occurrence events of an attribute depends on its frequency in the list. During MC simulations this approach should introduce compositional biases, since it does not reflect the real gene accretion process during baculovirus genome evolution [101] but critically, it does generate virtual attribute arrangements with the same total number of components as the real genome.

Estimating the evolutionary change on each AgMNPV-2D ORF

We also sought to generate node attributes such as molecular evolution (*i.e.*, genetic diversity), to correlate with the estimated network parameters. To verify this property, we inferred the genetic diversity (π) in a comparative set of 121 coding sequences shared among 4 group I alphabaculoviruses: AgMNPV (*Anticarsia gemmatalis* MNPV, DQ813662), CfDEFNPV (*Choristoneura fumiferana* defective NPV, AY327402), EppoMNPV (*Epiphyas postvittana* MNPV, AY043265) and CfMNPV (*Choristoneura fumiferana* MNPV, AF512031). These baculoviruses were chosen due to their close phylogenetic relationship, and the homologues ORFs were selected according to Oliveira *et al.* [22]. Nucleotide sequences were extracted from GenBank database using the annotation software Artemis [102]. Multiple sequence alignments for each gene were generated using ClustalW [103] and manually adjusted using Bioedit [104]. Finally, we used a phylogeny-based Markov Chain Monte Carlo (MCMC) Bayesian method implemented in the Coalesce 1.5 beta program [105] to estimate the genetic diversity ($\pi = Ne \cdot \mu$) for each viral gene alignment and compared π to betweenness centrality (BC) estimates obtained for ‘core’ (replication and capsid) and ‘satellite’ (all the others) functions [101]. With Coalesce, an initial Watterson estimate of π was used and the MCMC consisted of 10 short chains and 2 long chains of 4000 steps each, during which the viral trees for each gene were optimized and π estimated.

Competing interests

The authors declare that they have no competing interests. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

JVCO and PMAZ designed the experiments, analyzed the data and wrote the manuscript. AFB, CTB and CCMF analyzed the data and wrote the manuscript, analyzed the data and wrote the manuscript. AI wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgments

JVCO, CTB and AI hold FAPESP scholarships (04/12456-0, 09/16740-8 and 12/04818-5), AFB and CCMF hold CAPES-MSc and PhD scholarships and PMAZ holds a CNPq-PQ scholarship. This work was supported financially by FAPESP (Fundação de Amparo a Pesquisa do Estado de São Paulo, process: 2007/55282-0).

References

1. Kitano H: **Biological robustness.** *Nat Rev Genet* 2004, **5**:826–837.
2. Schlosser G, Wagner GP: *Modularity in development and evolution.* London: University of Chicago Press; 2004.
3. West-Eberhard MJ: *Developmental Plasticity and Evolution.* New York: Oxford University Press; 2003.

4. Hartwell LH, Hopfield JJ, Leibler S, Murray AW: **From molecular to modular cell biology.** *Nature* 1999, **402**:C47–52.
5. Schilling CH, Edwards JS, Palsson BO: **Toward metabolic phenomics: analysis of genomic data using flux balances.** *Biotechnol Prog* 1999, **15**:288–295.
6. Varma A, Palsson BO: **Metabolic capabilities of Escherichia coli: I. synthesis of biosynthetic precursors and cofactors.** *J Theor Biol* 1993, **165**:477–502.
7. Wagner A: **Does selection mold molecular networks?** *Sci STKE* 2003, **2003**:PE41.
8. Vitkup D, Kharchenko P, Wagner A: **Influence of metabolic network structure and function on enzyme evolution.** *Genome Biol* 2006, **7**:R39.
9. Krakauer DC, Zanutto PMA: **Viral Individuality and Limitations of the Life Concept.** In *Protocells: Bridging Nonliving and Living Matter*. Edited by Rasmussen S, Bedau M, Packard N, Krakauer D, Stadler P, Deamer D. Cambridge: MIT Press; 2009:712.
10. O'Reilly DR, Miller LK, Luckow VA: *Baculovirus Expression Vectors*. New York: A Laboratory Manual; 1993.
11. Granados RR, Federici BA: *Biology Of Baculoviruses*. CRC Press; 1987.
12. Garavaglia MJ, Miele SA, Iserle JA, Belaich MN, Ghiringhelli PD: **The ac53, ac78, ac101, and ac103 genes are newly discovered core genes in the family Baculoviridae.** *J Virol* 2012, **86**:12069–12079.
13. Garcia-Maruniak A, Maruniak JE, Zanutto PM, Doumbouya AE, Liu JC, Merritt TM, Lanoie JS: **Sequence analysis of the genome of the Neodiprion sertifer nucleopolyhedrovirus.** *J Virol* 2004, **78**:7036–7051.
14. Herniou EA, Olszewski JA, Cory JS, O'Reilly DR: **The genome sequence and evolution of baculoviruses.** *Annu Rev Entomol* 2003, **48**:211–234.
15. Lauzon HA, Lucarotti CJ, Krell PJ, Feng Q, Retnakaran A, Arif BM: **Sequence and organization of the Neodiprion lecontei nucleopolyhedrovirus genome.** *J Virol* 2004, **78**:7023–7035.
16. Miele SA, Garavaglia MJ, Belaich MN, Ghiringhelli PD: **Baculovirus: molecular insights on their diversity and conservation.** *Int J Evol Biol* 2011, **2011**:379424.
17. Slack J, Arif BM: **The baculoviruses occlusion-derived virus: virion structure and function.** *Advances in virus research* 2007, **69**:99–165.
18. Szewczyk B, Hoyos-Carvajal L, Paluszek M, Skrzecz I, de Lobo Souza M: **Baculoviruses— re-emerging biopesticides.** *Biotechnol Adv* 2006, **24**:143–160.
19. Moscardi F: **Assessment of the application of baculoviruses for control of Lepidoptera.** *Annu Rev Entomol* 1999, **44**:257–289.

20. Maruniak JE, Garcia-Maruniak A, Souza ML, Zanotto PM, Moscardi F: **Physical maps and virulence of *Anticarsia gemmatalis* nucleopolyhedrovirus genomic variants.** *Arch Virol* 1999, **144**:1991–2006.
21. Maruniak JE: **Molecular biology of *Anticarsia gemmatalis* baculovirus.** *Mem Inst Oswaldo Cruz* 1989:107–111.
22. Oliveira JV, Wolff JL, Garcia-Maruniak A, Ribeiro BM, de Castro ME, de Souza ML, Moscardi F, Maruniak JE, Zanotto PM: **Genome of the most widely used viral biopesticide: *Anticarsia gemmatalis* multiple nucleopolyhedrovirus.** *J Gen Virol* 2006, **87**:3233–3250.
23. Lu A, Miller LK: **Regulation of baculovirus late and very late gene expression.** In *The Baculoviruses*. Edited by Miller LK. New York: Plenum; 1997:193–216.
24. Happ B, Li J, Doerfler W: **Proteins encoded in the 81.2- to 85.0-map-unit fragment of *Autographa californica* nuclear polyhedrosis virus DNA can be translated in vitro and in *Spodoptera frugiperda* cells.** *J Virol* 1991, **65**:89–97.
25. Oellig C, Happ B, Muller T, Doerfler W: **Overlapping sets of viral RNAs reflect the array of polypeptides in the EcoRI J and N fragments (map positions 81.2 to 85.0) of the *Autographa californica* nuclear polyhedrosis virus genome.** *J Virol* 1987, **61**:3048–3057.
26. Smith I: **Misleading messengers? Interpreting baculovirus transcriptional array profiles.** *J Virol* 2007, **81**:7819–7820. author reply 7820–7811.
27. Friesen PD, Miller LK: **Temporal regulation of baculovirus RNA: overlapping early and late transcripts.** *J Virol* 1985, **54**:392–400.
28. Kool M, Vlak JM: **The structural and functional organization of the *Autographa californica* nuclear polyhedrosis virus genome.** *Arch Virol* 1993, **130**:1–16.
29. Lubbert H, Doerfler W: **Transcription of overlapping sets of RNAs from the genome of *Autographa californica* nuclear polyhedrosis virus: a novel method for mapping RNAs.** *J Virol* 1984, **52**:255–265.
30. Mainprize TH, Lee K, Miller LK: **Variation in the temporal expression of overlapping baculovirus transcripts.** *Virus Res* 1986, **6**:85–99.
31. Katsuma S, Kang W, Shin-i T, Ohishi K, Kadota K, Kohara Y, Shimada T: **Mass identification of transcriptional units expressed from the *Bombyx mori* nucleopolyhedrovirus genome.** *J Gen Virol* 2011, **92**:200–203.
32. Rodrigues JC, De Souza ML, O'Reilly D, Velloso LM, Pinedo FJ, Razuck FB, Ribeiro B, Ribeiro BM: **Characterization of the ecdysteroid UDP-glucosyltransferase (egt) gene of *Anticarsia gemmatalis* nucleopolyhedrovirus.** *Virus Genes* 2001, **22**:103–112.
33. Razuck FB, Ribeiro B, Vargas JH, Wolff JL, Ribeiro BM: **Characterization of the p10 gene region of *Anticarsia gemmatalis* nucleopolyhedrovirus.** *Virus Genes* 2002, **24**:243–247.

34. Slack JM, Shapiro M: **Anticarsia gemmatalis multicapsid nucleopolyhedrovirus v-trex gene encodes a functional 3' to 5' exonuclease.** *J Gen Virol* 2004, **85**:2863–2871.
35. de Lima L, Pinedo FJ, Ribeiro BM, Zanotto PM, Wolff JL: **Identification, expression and phylogenetic analysis of the Anticarsia gemmatalis multicapsid nucleopolyhedrovirus (AgMNPV) Helicase gene.** *Virus Genes* 2004, **29**:345–352.
36. Carpes MP, de Castro ME, Soares EF, Villela AG, Pinedo FJ, Ribeiro BM: **The inhibitor of apoptosis gene (iap-3) of Anticarsia gemmatalis multicapsid nucleopolyhedrovirus (AgMNPV) encodes a functional IAP.** *Arch Virol* 2005, **150**:1549–1562.
37. Belaich MN, Rodriguez VA, Bilen MF, Pilloff MG, Romanowski V, Sciocco-Cap A, Ghiringhelli PD: **Sequencing and characterisation of p74 gene in two isolates of Anticarsia gemmatalis MNPV.** *Virus Genes* 2006, **32**:59–70.
38. Castro ME, Souza ML, Araujo S, Bilimoria SL: **Replication of Anticarsia gemmatalis nuclear polyhedrosis virus in four lepidopteran cell lines.** *J Invertebr Pathol* 1997, **69**:40–45.
39. Sieburth P, Maruniak JE: **Growth characteristics of a continuous cell line from the velvetbean caterpillar, Anticarsia gemmatalis hübner (lepidoptera: Noctuidae).** *In Vitro Cel Dev Biol Plant* 1988, **24**:195–198.
40. Vaughn JL, Goodwin RH, Tompkins GJ, McCawley P: **The establishment of two cell lines from the insect Spodoptera frugiperda (Lepidoptera; Noctuidae).** *In vitro* 1977, **13**:213–217.
41. Davidson E, Levin M: **Gene regulatory networks.** *Proc Natl Acad Sci U S A* 2005, **102**:4935.
42. Cohen D, Marek M, Davies B, Vlak JM, van Oers M: **Encyclopedia of Autographa californica Nucleopolyhedrovirus Genes.** *Virol Sin* 2009, **24**:359–414.
43. Baculovirus Molecular Biology: *Baculovirus Molecular Biology*. <http://www.ncbi.nlm.nih.gov/books/NBK49500/>.
44. Wilson ME, Mainprize TH, Friesen PD, Miller LK: **Location, transcription, and sequence of a baculovirus gene encoding a small arginine-rich polypeptide.** *J Virol* 1987, **61**:661–666.
45. Kelly DC, Brown DA DA, Ayres MD, Allen CJ, Walker IO: **Properties of the Major Nucleocapsid Protein of Heliothis zea Singly Enveloped Nuclear Polyhedrosis Virus.** *J Gen Virol* 1983, **64**:399–408.
46. Iwanaga M, Takaya K, Katsuma S, Ote M, Tanaka S, Kamita SG, Kang W, Shimada T, Kobayashi M: **Expression profiling of baculovirus genes in permissive and nonpermissive cell lines.** *Biochemical and biophysical research communications* 2004, **323**:599–614.

47. D'Costa SM, Yao H, Bilimoria SL: **Transcription and temporal cascade in Chilo iridescent virus infected cells.** *Arch Virol* 2001, **146**:2165–2178.
48. Chinchar VG, Han J, Mao J, Brooks I, Srivastava K: **Instability of frog virus 3 mRNA in productively infected cells.** *Virology* 1994, **203**:187–192.
49. Chinchar VG, Yu W: **Metabolism of host and viral mRNAs in frog virus 3-infected cells.** *Virology* 1992, **186**:435–443.
50. Tjia ST, Carstens EB, Doerfler W: **Infection of Spodoptera frugiperda cells with Autographa californica nuclear polyhedrosis virus II. The viral DNA and the kinetics of its replication.** *Virology* 1979, **99**:399–409.
51. Liu HS, Bilimoria SL: **Infected cell specific protein and viral DNA synthesis in productive and abortive infections of Spodoptera frugiperda nuclear polyhedrosis virus.** *Arch Virol* 1990, **115**:101–113.
52. Riegel CI, Slavicek JM: **Characterization of the replication cycle of the Lymantria dispar nuclear polyhedrosis virus.** *Virus Res* 1997, **51**:9–17.
53. Moore RC, Purugganan MD: **The early stages of duplicate gene evolution.** *Proc Natl Acad Sci U S A* 2003, **100**:15682–15687.
54. Ohta T: **Time for spreading of compensatory mutations under gene duplication.** *Genetics* 1989, **123**:579–584.
55. Blissard GW, Wenz JR: **Baculovirus gp64 envelope glycoprotein is sufficient to mediate pH-dependent membrane fusion.** *J Virol* 1992, **66**:6829–6835.
56. Hoover K, Grove M, Gardner M, Hughes DP, McNeil J, Slavicek J: **A gene for an extended phenotype.** *Science* 2011, **333**:1401.
57. Foster KR: **The sociobiology of molecular systems.** *Nat Rev Genet* 2011, **12**:193–203.
58. Fan TJ, Han LH, Cong RS, Liang J: **Caspase family proteases and apoptosis.** *Acta Biochim Biophys Sin (Shanghai)* 2005, **37**:719–727.
59. Viana MP, Tanck E, Beletti ME, Costa Lda F: **Modularity and robustness of bone networks.** *Mol Biosyst* 2009, **5**:255–261.
60. Jiang SS, Chang IS, Huang LW, Chen PC, Wen CC, Liu SC, Chien LC, Lin CY, Hsiung CA, Juang JL: **Temporal transcription program of recombinant Autographa californica multiple nucleopolyhedrosis virus.** *J Virol* 2006, **80**:8989–8999.
61. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL: **Hierarchical organization of modularity in metabolic networks.** *Science* 2002, **297**:1551–1555.
62. Hurst LD, Pal C, Lercher MJ: **The evolutionary dynamics of eukaryotic gene order.** *Nat Rev Genet* 2004, **5**:299–310.

63. Lubbert H, Doerfler W: **Mapping of Early and Late Transcripts Encoded by the *Autographa californica* Nuclear Polyhedrosis Virus Genome: Is Viral RNA Spliced?** *J Virol* 1984, **50**:497–506.
64. Jin J, Guarino LA: **3'-end formation of baculovirus late RNAs.** *J Virol* 2000, **74**:8930–8937.
65. Adhya S, Gottesman M: **Promoter occlusion: transcription through a promoter may inhibit its activity.** *Cell* 1982, **29**:939–944.
66. Dinger ME, Gascoigne DK, Mattick JS: **The evolution of RNAs with multiple functions.** *Biochimie* 2011, **93**:2013–2018.
67. Hughes TA: **Regulation of gene expression by alternative untranslated regions.** *Trends Genet* 2006, **22**:119–122.
68. Friesen PD, Miller LK: **Divergent transcription of early 35- and 94-kilodalton protein genes encoded by the HindIII K genome fragment of the baculovirus *Autographa californica* nuclear polyhedrosis virus.** *J Virol* 1987, **61**:2264–2272.
69. Ooi BG, Miller LK: **Transcription of the baculovirus polyhedrin gene reduces the levels of an antisense transcript initiated downstream.** *J Virol* 1990, **64**:3126–3129.
70. Kang ST, Leu JH, Wang HC, Chen LL, Kou GH, Lo CF: **Polycistronic mRNAs and internal ribosome entry site elements (IRES) are widely used by white spot syndrome virus (WSSV) structural protein genes.** *Virology* 2009, **387**:353–363.
71. Olszewski J, Miller LK: **Identification and characterization of a baculovirus structural protein, VP1054, required for nucleocapsid formation.** *J Virol* 1997, **71**:5040–5050.
72. Summers MD, Smith GE: **A manual of methods for baculovirus vectors and insect cell culture procedures.** *Texas Agric Exper Sta Bull* 1987:1–57.
73. Abot AR, Moscardi F, Fuxa JR, Sosa-Gomez DR, Richter AR: **Susceptibility of populations of *Anticarsia gemmatilis* from Brazil and the United States to a nuclear polyhedrosis virus.** *JEntomolSci* 1995, **30**:62–69.
74. Abot AR, Moscardi F, Fuxa JR, Sosa-Gomez DR, Richter AR: **Development of resistance by *Anticarsia gemmatilis* from Brazil and the United States to a nuclear polyhedrosis virus under laboratory selection pressure.** *Biol Control* 1996, **7**:126–130.
75. Allen GE, Knell JD: **A nuclear polyhedrosis virus of *Anticarsia gemmatilis*: I, ultrastructure, replication, and pathogenicity.** *Fl Entomol* 1977:233–240.
76. Fuxa JR, Richter AR: **Distance and rate of spread of *Anticarsia gemmatilis* (Lepidoptera: Noctuidae) nuclear polyhedrosis virus released into soybean.** *Environ Entomol* 1994, **23**:1308–1316.

77. Johnson DW, Maruniak JE: **Physical map of Anticarsia-gemmatalis nuclear polyhedrosis-virus (AgMNPV-2) DNA.** *J Gen Virol* 1989;1877–1883.
78. Klein D: **Quantification using real-time PCR technology: applications and limitations.** *Trends Mol Med* 2002, **8**:257–260.
79. Mackay IM, Arden KE, Nitsche A: **Real-time PCR in virology.** *Nucleic Acids Res* 2002, **30**:1292–1305.
80. Yang D-H, Barari M, Arif BM, Krell PJ: **Development of an oligonucleotide-based DNA microarray for transcriptional analysis of Choristoneura fumiferana nucleopolyhedrovirus (CfMNPV) genes.** *J Virol Methods* 2007, **143**:175–185.
81. Friesen PD: **Regulation of baculovirus early gene expression.** In *The Baculoviruses*. Edited by Miller LK. New York: Plenum; 1997:141–170.
82. Guarino LA, Xu B, Jin J, Dong W: **A virus-encoded RNA polymerase purified from baculovirus-infected cells.** *J Virol* 1998, **72**:7985–7991.
83. Carver T, Berriman M, Tivey A, Patel C, Bohme U, Barrell BG, Parkhill J, Rajandream MA: **Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database.** *Bioinformatics* 2008, **24**:2672–2676.
84. Kurakin A: **Self-organization vs Watchmaker: stochastic gene expression and cell differentiation.** *Dev Genes Evol* 2005, **215**:46–52.
85. Ross IL, Browne CM, Hume DA: **Transcription of individual genes in eukaryotic cells occurs randomly and infrequently.** *Immunol Cell Biol* 1994, **72**:177–185.
86. Larson DR: **What do expression dynamics tell us about the mechanism of transcription?** *Curr Opin Genet Dev* 2011, **21**:591–599.
87. Maier T, Guell M, Serrano L: **Correlation of mRNA and protein in complex biological samples.** *FEBS Lett* 2009, **583**:3966–3973.
88. Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M: **Global quantification of mammalian gene expression control.** *Nature* 2011, **473**:337–342.
89. Yamagishi J, Isobe R, Takebuchi T, Bando H: **DNA microarrays of baculovirus genomes: differential expression of viral genes in two susceptible insect cell lines.** *Arch Virol* 2003, **148**:587–597.
90. Huang S: **Cell lineage determination in state space: a systems view brings flexibility to dogmatic canonical rules.** *PLoS Biol* 2010, **8**:e1000380.
91. Huang S, Eichler G, Bar-Yam Y, Ingber DE: **Cell fates as high-dimensional attractor states of a complex gene regulatory network.** *Phys Rev Lett* 2005, **94**:128701.
92. Kauffman SA: *The origins of order: self-organization and selection in evolution.* Oxford University Press; 1993.

93. D’Haeseleer P, Liang S, Somogyi R: **Genetic network inference: from co-expression clustering to reverse engineering.** *Bioinformatics* 2000, **16**:707–726.
94. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P: **Inferring regulatory networks from expression data using tree-based methods.** *PLoS ONE* 2010, **5**.
95. *The R Project for Statistical Computing.*
96. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T: **Cytoscape 2.8: new features for data integration and network visualization.** *Bioinformatics* 2010, **27**:431–432.
97. Lin YY, Qi Y, Lu JY, Pan X, Yuan DS, Zhao Y, Bader JS, Boeke JD: **A comprehensive synthetic genetic interaction network governing yeast histone acetylation and deacetylation.** *Genes Dev* 2008, **22**:2062–2074.
98. Holme P: **Metabolic robustness and network modularity: a model study.** *PLoS ONE* 2011, **6**:e16605.
99. Brandes U: **A faster algorithm for betweenness centrality.** *Journal Math Sociol* 2001, **25**:163–177.
100. Shimbel A: **Structural parameters of communication networks.** *Bull Math Biophys* 1953, **15**:501–507.
101. Zannotto PMA, Krakauer DC: **Complete genome viral phylogenies suggests the concerted evolution of regulatory cores and accessory satellites.** *PLoS ONE* 2008, **3**:e3500.
102. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B: **Artemis: sequence visualization and annotation.** *Bioinformatics* 2000, **16**:944–945.
103. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**:4673–4680.
104. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucl Acids Symp Ser* 1999:95–98.
105. *Fetching the COALESCE program.*
<http://evolution.genetics.washington.edu/lamarc/coalesce.html>.

Additional files

Additional_file_10 as DOCX

Additional files 1 Table S1, showing the temporal expression of AgMNPV-2D genes in UFL-AG-286 cell line.

Apêndice B - Triplicatas Géis Bidimensionais (2DE)

Figura B1- Análise das triplicatas de eletroforese bidimensional do fenótipo ODV purificado de células UFLAG em fitas de pH 3-10

Análise no software Platinum 7.0 (Ge Healthcare), demonstrando as proteínas estruturais. Os *spots* que não apareceram em todos os géis não foram cortados.

Figura B2- Análise das triplicatas dos géis bidimensionais do fenótipo ODV purificado de células UFLAG em fitas de pH 4-7.

As eletroforeses analisadas no software Platinum 7.0 (Ge Healthcare) demonstraram 16 spots para o fenótipo ODV. Os *spots* que não apareceram em todos os géis não foram cortados.

Os espectros de massas mostraram baixa cobertura ou baixa pontuação contra o banco teórico da AgMNPV-2D utilizando fitas de pH 4_7 por MALDI-Q-TOF.

Tabela B1- Proteínas estruturais do AgMNPV-2D no ODV isolado de UFL-AG-286 nas triplicatas 2DE em fitas de pH 4_7 por MALDI-Q-TOF .

Spot	AgMNPV-ORF	MALDI-Q-TOF	kDa	pI	(%)*	Score	NCBI BLAST
1	20	AG20	38	8.6	7	40	gi 116326089
1	140	ODV-E27	34	7.7	12	73	gi 116326208
2	97	P40	41	7.0	8	37	gi 116326165
3	86	VP39	39	6.5	12	114	gi 116326154
4	103	PNK/PNL	80	8.7	6	60	gi 116326171
6	151	PK1	32	8.9	10	22	gi 116326070
8	103	PNK/PNL	80	8.7	4	43	gi 116326153
9	103	PNK/PNL	80	8.7	4	12	gi 116326171
11	92	Helicase	142	7.1	3	14	gi 116326160
15	107	ODV-EC43	46	8.2	2	23	gi 116326175

* Porcentagem de cobertura da sequência gerado pelo banco de dados Mascot (www.matrixscience.com).

Apêndice C - Reação de PCR e diversidade da GP64

Tabela C1 - Reação para PCR para fragmentos de *gp64* (ORF-124)

Componente	Volume	[final]
10x Tampão com Mg ²⁺	5 µl	1x (2.5 mM Mg ²⁺)
10 mM dNTP Mix	1 µl	0.25 mM
Oligo A(+)	1 µl	0.5 µM
Oligo B (-)	1 µl	0.5 µM
DNA		100-200 ng
High Fidelity Taq DNA	0.4 µl	2 U
H ₂ O	completando 50 µl	-

Tabela C2 - Sequência dos oligonucleotídeos utilizados para amplificação da ORF 124 e localizações destes no genoma do AgMNPV-2D

Identificação	Posição	Sequência (5'→3')
ORF124 F	104496-104516	GACAAGGTTGCGAGTTCAAG
ORF124 R	106116-106140	TAAATCAGTTACAATTGAGCTGTG

Tabela C3 - Sequência dos oligonucleotídeos utilizados o sequenciamento dos clones gerados em vetor TOPO TA contendo a ORF 124

Identificação	Sequência oligo (5' - 3')
Ag124.901U17	ATGCACATTCAGGAGGA
Ag124.1180L15	CGCAGAGTCCGTGTT
Ag.gp64.seqF	GCGGCTTGAGTGCGATGACG
Ag.gp64.seqR	CCGAATTTGGGTCTGATTTG
M13 F	TGTAACACGACGGCCAGT
M13 R	CAGGAAACAGCTATGAC

Tabela C4- Análise das frequências de mutações: sinônimas e não sinônima na *gp64* entre os isolados geográficos sequenciados de AGMNPV por sequenciamento de alta cobertura

Isolados	N-Terminal										Central							C-Terminal			
	4	96	99	120	149	186	237	287	381	420	669	679	687	701	757-759	757-759	795	1020	1050	1078	1479
	A>G	G>A	T>C	C>T	G>T	A>G	G>A	A>G	C>T	A>G	G>A	T>C	T>C	C>T	TTA>CTG	T>C	T>C	C>T	T>C	C>T	T>A
26	47.7%	29.9%	31.1%	73%	79%	50.5%	50%	*	*	51%	19.5%	19.2%	18.4%	19.3%	22%	20.9%	*	69.1%	14.3%	27.4%	
27	21.1%	93.3%	94.8%	95%	95%	87.3%	87.3%	*	25.9%	12.4%	*	*	*	*	*	*	*	11.5%	*	*	
28	65.3%	87%	87%	90%	90%	42.3%	37.1%	*	*	62.6%	23.7%	24.1%	24.3%	23.2%	22.5%	28.9%	*	41.1%	33.3%	27.6%	
29	35.7%	68%	68.5%	65%	65%	33%	70.7%	*	*	72.2%	20.6%	20.9%	21.5%	23.5%	27.6%	29.8%	*	79.1%	*	*	
30	99.3%	100%	100%	98%	100%	100%	99.2%	*	*	*	*	*	*	*	*	*	*	*	*	*	
31	72.2%	75.5%	76%	57%	62%	*	*	*	*	69%	19.6%	79.8%	17.7%	22.0%	23.6%	19.2%	*	52.1%	17.9%	*	
32	92.7%	78.7%	78.7%	79%	82%	70.9%	71.1%	*	55.1%	32.5%	17.8%	17.9%	17.7%	17.8%	17.6%	14.8%	*	19.7%	11.4%	16.2%	
33	99.1%	100%	100%	100%	100%	*	*	*	*	99.6%	*	*	*	*	*	*	*	*	*	*	
34	59.4%	10.9%	10.8%	48%	51%	13.1%	14.2%	*	*	92.8%	95.1%	95.2%	95.7%	96.2%	96.2%	96.1%	*	89.1%	85.1%	*	
35	98%	26.6%	26.9%	25%	25%	*	*	*	*	96.2%	68.2%	69.5%	66.9%	70.7%	71.0%	69.1%	*	60.9%	47.6%	*	
36	100%	29%	27%	32%	33%	*	*	*	*	94.5%	46.2%	46.4%	46.4%	50.5%	47.5%	44.7%	*	51.3%	46.5%	*	
37	12%	96%	96%	97%	55%	84%	85.5%	*	*	10.4%	*	*	*	*	86.8%	*	69.5%	*	*	*	
38	100%	98%	98%	98%	100%	100%	100%	*	*	*	*	*	*	*	*	*	*	*	*	*	
39	100%	*	*	*	*	*	*	17.5%	*	100%	98.3%	100%	100%	100%	62.7%	63.8%	*	56.1%	47.8%	*	
40	47.8%	67%	63%	68%	74%	14.5%	21.3%	*	*	72.9%	34.5%	33.3%	33.3%	32.6%	40.4%	34.7%	*	50.6%	20.5%	47.4%	
43	91%	70%	71%	75%	74%	12.1%	15.7%	*	*	72.9%	32.9%	33.3%	30.6%	31%	35%	37.7%	*	39.7%	33.9%	35.8%	

ANEXO - Lista de publicações

Ramsden C., Melo F. L., Figueiredo L. M., Holmes C. E., Zanutto, P. M., VGDN Consortium (BRACONI C. T.) High Rates of Molecular Evolution in Hantaviruses. *Molecular Biology and Evolution*, v. 25, p. 1488-1492, 2008.

Figueiredo L. M., Moreli. M., SOUSA, R. L., Borges A., Figueiredo G. G., Machado, A. M., Bisordi I., Nagasse-Sugara T.K., Suzuki A., Pereira L. E., Souza, R. P. Souza, L. T., BRACONI, C.T., Hari C.M., Zanutto, P. M.. Hantavirus Pulmonary Syndrome, Central Plateau, Southeastern, and Southern Brazil. *Emerging Infectious Diseases*, v. 15, p. 561-567, 2009.

Tomasini-Grotto, R-M, Montes, B; Triglia, D; BRACONI, C. T.; Aliano-Block, J.; Zanutto, P.M.; Pardini, M-I M. C.; Segondy, M. Variability of the conserved V3 loop tip motif in HIV-1 subtype B isolates collected from Brazilian and French patients. *Brazilian Journal of Medical and Biological Research (Impresso)*, v. 41, p. 720-728, 2010.

Iamarino A, de Melo, F.L., BRACONI, C. T, Zanutto P.M. BF Integrase Genes of HIV-1 Circulating in Sao Paulo, Brazil, with a Recurrent Recombination Region. *Plos One*. 2012;7(4):e34324.

Artigo aceito para publicação:

Oliveira J.V.C., Brito A. F., BRACONI, C. T, Freire, C. C. M., Iamarino A, Zanutto PM. Modularity and evolutionary constraints in a baculovirus gene regulatory network”. *BMC Systems Biology*;

Artigo sob revisão:

BRACONI, C. T, Ardisson-Araújo D. M. P., Leme A. F. P., Oliveira J. V. C., Pauletti B. A., Garcia-Maruniak A., Ribeiro B. M., Maruniak J. E., P. M. A. Zanutto. “Cell modulation factors are carried along structural proteins in the baculovirus AgMNPV-2D budded and occluded virions”. *Proteomics*;