Role of upstream sequences on the *Ac*MNPV polyhedrin promoter transcription

Thesis submitted to

Department of Biochemistry
School of Life Sciences
University of Hyderabad
Hyderabad

For the degree of

Doctor of Philosophy

M Senthil Kumar

Centre for DNA Fingerprinting and Diagnostics
Hyderabad
2006

Registration Number: 2KLBPH18

CERTIFICATE

This is to certify that this thesis entitled "Role of upstream sequences on the *Ac*NPV polyhedrin promoter transcription" comprises the work done by *M Senthil Kumar* under my guidance at the Centre for DNA Fingerprinting and Diagnostics. The work is original and has not been submitted in part or full for any other degree or diploma of any other University.

Dean,
School of Life Sciences
University of Hyderabad

Dr. Seyed E Hasnain
Supervisor

Head,
Department of Biochemistry
University of Hyderabad

DECLARATION

I hereby declare that this thesis entitled "Role of upstream sequences on the *Ac*MNPV polyhedrin promoter transcription" has been carried out by me under the supervision of *Dr. Seyed E. Hasnain* at the Centre for DNA Fingerprinting and Diagnostics, Hyderabad. The work is original and has not been submitted in part or full for any other degree or diploma of any other University earlier.

M Senthil Kumar
Candidate

ACKNOWLEDGEMENTS

I thank, my supervisor *Dr. Seyed E Hasnain* for his guidance, enthusiasm & encouragement, *Dr. J Gowrishankar*, Director, Centre for DNA Fingerprinting & Diagnostics, for his support.

I also thank my co-supervisor, *Dr. Murali D Bashyam* for his guidance, useful discussions and his patience.

Thanks also goes to members of LMCB, both past[1] & present[2], LMO[3] and NGTF[4] for a friendly work atmosphere.

I thank *Dr. Nasreen Ehtesham* for allowing me to use the facilities at the National Institute of Nutrition. I also thank *Prof. KVA Ramaiah*, University of Hyderabad, for helpful suggestions.

Financial support from the Council of Scientific and Industrial Research is gratefully acknowledged. I thank *Dr. J Gowrishankar*, Director for funding me during the last stages of my work.

Senthil Kumar M
June, 2006

[1]Noman, Aruna, Priya, Sudip, Sharmistha, Sudhir, Prachee, Ravi, Ashok & Prashanth.

[2]Dr. Sangita, Dr. Niyaz, Mrs. Savithri, Niteen, Leena, Krishnaveni, Smanla, Sheeba, Nasreena, Abid, Aisha, Noor, Sreejit, Anil, Chandra Sekhar, Sandeep, Shiny, Yusuf, Kaiser, Khaled and Ghous. Also, thanks to lab assistants: N Sudhir, Surender, Zameer, Krishnamoorthy & Wasim.

[3]Ratheesh, Khursheed, Mahesh & Madhu.

[4]Gopi, Vartul, Chandrakanth & Ajay.

ABBREVIATIONS

°C	Degree celsius
AA	Amino acids
APS	Ammonium persulphate
ATP	Adenosine-5'-triphosphate
bp	Base pair
BSA	Bovine Serum Albumin
Ci	Curies
cm	Centimeter (10^{-2} meter)
dATP	2'-deoxyadenosine-5'-triphosphate
dCTP	2'-deoxycytidine-5'-triphosphate
dGTP	2'-deoxyguanosine-5'-triphosphate
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DTT	1,4-Dithiothreitol
dTTP	2'-deoxythymidine-5'-triphosphate
E. coli	*Escherechia coli*
EDTA	Ethylenediaminetetraacetic acid (disodium salt)
EtBr	Ethidium Bromide
gm	Gram
hpi	hours post infection
HSP	Heat shock protein
kb	Kilobase pair
KCl	Potassium Chloride
kDa	Kilodaltons
Km	Michaelis-Menten constant
KV	Kilovolts
L	Litre
LB	Luria-Bertani
M	Molar
MCS	Multiple cloning site
mg	Milligram (10^{-3} gram)

Mg^{++}	Magnesium ion
$MgCl_2$	Magnesium Chloride
$MgSO_4$	Magnesium sulphate
mins	Minutes
mL	Millilitres (10^{-3} litres)
mM	Millimolar
mmol	Millimoles (10^{-3} moles)
MOI	Multiplicity of infection
NaCl	Sodium Chloride
N. E.	Nuclear Extract
ng	Nano gram (10^{-9} gram)
nt	Nucleotide
OD	Optical density
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
pmoles	Picomoles (10^{-12} moles)
RNA	Riboxynucleic acid
rpm	Revolutions per minute
SDS	Sodium dodecyl sulphate
secs	Seconds
TAE	TrisCl Acetic acid EDTA buffer
Taq	*Thermus aquaticus*
TBE	TrisCl Boric acid EDTA buffer
TE	TrisCl EDTA
TEMED	N,N,N',N'-tetramethylethylenediamine
TGE	TrisCl Glycin EDTA buffer
TrisCl	Tris-(hydroxymethyl) aminomethane hydrochloride
U	Units
UTR	Untranslated region
UV	Ultra Violet
μG	Micro gram (10^{-6} grams)
μJ	Micro joules (10^{-6} joules)
μM	Micromolar (10^{-6} molar)

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Chapter 1

Introduction

1.1 Baculoviridae Family

The members of Baculoviridae family are a very large and assorted family of viruses and are grouped under two genus: Nucleopolyhedrosis and Granulovirus. The term *baculo* refers to the rod-shaped capsids of the virus particles. They infect arthropods and are not known to have any natural mammalian hosts. The hosts of baculoviruses include over 600 species of insects belonging mainly to the order of Lepidoptera as well as the orders of Hymenoptera, Diptera, Coleoptera and Trichoptera (9).

Nucleopolyhedrosis viruses (NPV) form occluded bodies either with singly (as in the case of BmNPV) or multiply (as in the case of *Ac*MNPV) embedded nucleocapsids, that are made in the nucleus of infected cells and are then enveloped by first the nuclear membrane and then the plasma membrane. The type species of NPV is *Autographa californica* MultiNucleopolyhedrosis virus (*Ac*MNPV). Table 1.1 lists the baculoviridae members as catalogued by the Universal Virus Database, version 3[1]. (Entry contributed by Blissard GW, Black B, Keddie BA, Possee R,

[1]ICTVdB: http://www.ncbi.nlm.nih.gov/ICTVdb/ICTVdB/

Rohrmann G, Theilmann DA, and Volkman LE.). This classification is based on the morphology of occlusion bodies (OB). Recently, Jehle and colleagues have analyzed the phylogenetic affiliations (Figure 1.1 in Page 17) of baculovirus members based on sequence comparison of 29 genes of baculoviral genomes sequenced so far (Table 1.2 in Page 13 (61)).

As an outcome of this work, they have proposed that the baculoviridae family should be divided into the following four genera, namely:

1. **Alphabaculovirus**: Lepidopteran-specific NPV.

   Type Species: Autographa californica x nucleopolyhedrosis virus

2. **Betabaculovirus**: Lepidopteran-specific GV.

   Type Species: Cydia pomonella granulovirus (CyGV)

3. **Gammabaculovirus**: Hymenopteran-specific NPV.

   Type Species: Neodiprion lecontei nucleopolyhedrosisvirus (NeleNPV).

4. **Deltabaculovirus**: (Possibly) Diptern-specific baculoviruses.

   Type Species: Culec nigripalpus nucleopolyhedrosis virus (CuniNPV).

1.1.1 Infection Cycle of Baculoviruses

Insect larvae get infected when they ingest baculovirus occluded bodies. In the midgut, the occlusion bodies are broken down by high pH and the released virions begin their infection life cycle by replicating in midgut cells. In contrast to NPVs pathogenic to Diptera and Hymenoptera where the infection is limited to midgut cells, in lepidopteran NPVs, the infection spreads from the midgut to other tissues through the formation of budded viruses (BV).

Table 1.1: Baculoviridae Family members

Virus Code	**Virus Species Names**
00.006.	**Baculoviridae Family**
00.006.0.01.	**Genus: Nucleopolyhedrovirus**
00.006.0.01.002.	*Anticarsia gemmatalis MNPV* (AgMNPV)
00.006.0.01.043.	*Agrotis ipsilon nucleopolyhedrovirus* (AgipMNPV)
00.006.0.01.003.	*Autographa californica MNPV* (**AcMNPV**) [1]
00.006.0.01.003.00.000.002.	Anagrapha falcifera NPV (AnfaNPV) [2]
00.006.0.01.003.00.000.003.	Galleria mellonella MNPV (GmMNPV)[2]
00.006.0.01.003.00.000.004.	Rachiplusia ou MNPV (RoMNPV)[2]
00.006.0.01.003.00.000.005.	Spodoptera exempta MNPV (SpexMNPV)[2]
00.006.0.01.003.00.000.006.	Trichoplusia ni MNPV (TnMNPV)[2]
00.006.0.01.004.	*Bombyx mori NPV* (BmNPV)
00.006.0.01.105.	*Buzura suppressaria nucleopolyhedrovirus* (BusuNPV)
00.006.0.01.005.	*Choristoneura fumiferana MNPV* (CfMNPV)
Continued on next page	

Table 1.1 – continued from previous page

Virus Code	Virus Species
00.006.0.01.006.	*Choristoneura fumiferana DEF nucleopolyhedrovirus* (CfDefNPV)
00.006.0.01.131.	*Choristoneura rosaceana nucleopolyhedrovirus* (ChroNPV)
00.006.0.01.012.	*Culex nigripalpus nucleopolyhedrovirus* (CuniNPV)
00.006.0.01.209.	*Epiphyas postvittana nucleopolyhedrovirus* (EppoNPV)
00.006.0.01.238.	*Helicoverpa armisgera nucleopolyhedrovirus* (HearNPV)
00.006.0.01.007.	*Helicoverpa zea single nucleopolyhedrovirus* (HzSNPV)
00.006.0.01.008.	*Lymantria dispar MNPV* (LdMNPV)
00.006.0.01.009.	*Mamestra brassicae MNPV* (MbMNPV)
00.006.0.01.309.	*Mamestra configurata nucleopolyhedrovirus* (MacoNPV)
00.006.0.01.323.	*Neodiprion lecontii nucleopolyhedrovirus* (NeleNPV)
00.006.0.01.017.	*Neodiprion sertifer NPV* (NeseNPV)
00.006.0.01.010.	*Orgyia pseudotsugata MNPV* (OpMNPV)
00.006.0.01.013.	*Spodoptera exigua MNPV* (SeMNPV)
00.006.0.01.014.	*Spodoptera frugiperda MNPV* (SfMNPV)
Continued on next page	

Table 1.1 – continued from previous page

Virus Code	**Virus Species**
00.006.0.01.457.	*Spodoptera littoralis nucleopolyhedrovirus* (SpliNPV)
00.006.0.01.015.	*Thysanoplusia orichalcea nucleopolyhedrovirus* (ThorNPV)
00.006.0.01.016.	*Trichoplusia ni single nucleopolyhedrovirus* (TnSNPV)
00.006.0.01.491.	*Wiseana signata nucleopolyhedrovirus* (WisiNPV)
Unassigned Members in the Genus	
00.006.0.81.035.	*Aedes sollicitans NPV* (AesoNPV)
00.006.0.81.267.	*Hyphantria cunea nucleopolyhedrovirus* (HycuNPV)
00.006.0.81.011.	*Orgyia pseudotsugata SNPV* (OpSNPV)
00.006.0.81.367.	*Panolis flammea NPV* (PaflNPV)
00.006.0.81.381.	*Penaeus monodon NPV* (PemoNPV)
00.006.0.81.457.	*Spodoptera littoralis NPV* (SpliNPV)
00.006.0.81.458.	*Spodoptera litura NPV* (SpltNPV)
00.006.0.81.476.	*Tipula paludosa NPV* (TipaNPV)
00.006.0.02.	**Genus: Granulovirus**
Continued on next page	

Table 1.1 – continued from previous page

Virus Code	Virus Species
00.006.0.02.002.	*Artogeia rapae granulovirus* (ArGV)
00.006.0.02.002.00.000.007.	*Pieris brassicae granulovirus* (PbGV)[2]
00.006.0.02.028.	*Choristoneura fumiferana granulavirus* (ChfuGV)
00.006.0.02.037.	*Cryptophlebia leucotreta granulavirus* (CrleGV)
00.006.0.02.003.	*Cydia pomonella granulovirus* (**CpGV**) [1]
00.006.0.02.064.	*Harrisina brillians granulavirus* (HabrGV)
00.006.0.02.065.	*Helicoverpa armisgera granulavirus* (HearGV)
00.006.0.02.077.	*Lathronympha phaseoli granulavirus* (LaphGV)
00.006.0.02.004.	*Plodia interpunctella granulovirus* (PiGV)
00.006.0.02.109.	*Plutella xylostella granulavirus* (PlxyGV)
00.006.0.02.114.	*Pseudaletia unipuncta granulavirus* (PsunGV)
00.006.0.02.005.	*Trichoplusia ni granulovirus* (TnGV)
00.006.0.02.006.	*Xestia c-nigrum granulovirus* (XecnGV)
Unassigned Members in the Genus	None reported.

Continued on next page

Table 1.1 – continued from previous page

Virus Code	**Virus Species**
List of Unassigned Viruses in the Family	None reported.
Similarity with Other Taxa	None reported.

[1]Type Species

[2]Isolates and Strains are not italicized

Table 1.2: List of baculovirus genomes sequenced till January 2006 and their characteristics. Reprinted with permission from (61). (Table 1 of (61))

Virus	Abbreviation	ORFs	Size (bp)	AT (%)	Accession
Autographa californica MNPV	AcMNPV	155	133,894	59.3	NC 001623
Choristoneura fumiferana MNPV	CfMNPV	145	129,609	49.9	NC 004778
Choristoneura fumiferana DEF NPV	CfDEFNPV	149	131,160	54.2	NC 005137
Rachiplusia ou MNPV	RoMNPV	146	131,526	60.9	NC 004323
Bombyx mori NPV	BmNPV	143	128,413	59.6	NC 001962
Epiphyas postvittana NPV	EppoNPV	136	118,584	59.3	NC 003083
Orgyia pseudotsugata MNPV	OpMNPV	152	131,990	44.9	NC 001875
Adoxophyes honmai NPV	AdhoNPV	125	113,220	64.4	NC 004690
Agrotis segetum MNPV	AgseNPV	153	147,544	54.3	NC 007921
Chrysodeixis chalcites	ChChNPV	151	149,622	60.9	NC 007151
Helicoverpa armigera NPV (G4)	HearNPV (G4)	135	131,403	61	NC 002654
Helicoverpa armigera NPV (C1)	HearNPV (C1)	134	130,760	61.1	NC 003094

Continued on next page

Table 1.2 – continued from previous page

Virus	Virus abbreviation	ORFs	Size(bp)	AT (%)	Accession
Helicoverpa zea SNPV	HzSNPV	139	130,869	60.9	NC 003349
Mamestra configurata NPV (A)	MacoNPV (A)	169	155,060	58.3	NC 003529
Mamestra configurata NPV (B)	MacoNPV (B)	168	158,482	60	NC 004117
Lymantria dispar MNPV	LdMNPV	166	161,046	42.5	NC 001973
Spodoptera exigua MNPV	SeMNPV	139	135,611	56.2	NC 002169
Spodoptera litura NPV	SpltNPV	141	139,342	57.2	NC 003102
Trichoplusia ni	TnSNPV	144	134,394	61	NC 007383
Cryptophlebia leucotreta GV	CrleGV	129	110,907	67.6	NC 005068
Cydia pomonella GV	CpGV	143	123,500	54.8	NC 002816
Phthorimea operculella GV	PhopGV	130	119,217	64.3	NC 004062
Plutella xylostella GV	PlxyGV	120	100,999	59.3	NC 002593
Adoxophyes orana GV	AdorGV	119	099,657	65.5	NC 005038
Xestia c-nigrum GV	XecnGV	181	178,733	59.3	NC 002331
Agrotis segetum GV	AgseGV	132	131,680	62.7	NC 005839

Continued on next page

Table 1.2 – continued from previous page

Virus	Virus abbreviation	ORFs	Size(bp)	AT (%)	Accession
Neodiprion lecontei NPV	NeleNPV	90	081,756	66.7	NC 005906
Neodiprion sertifer NPV	NeseNPV	90	086,462	66.2	NC 005905
Culex nigripalpus NPV	CuniNPV	109	108,252	49.1	NC 003084

Baculoviruses of both the genus (NPVs and GVs) produce two morphologically distinct viruses during their life cycle in insect hosts: budded viruses (BV) and occluded viruses (OV or ODV, occlusion derived viruses), also known as polyhedron derived viruses (PDV). The BV appears immediately following the infection and is necessary for the secondary infection of neighbouring cells through adsorptive endocytosis. The viruses formed later are occluded (ie ODV) and are thought to be more environmentally stable than BV since they are enclosed in a matrix composed of polyhedrin protein surrounded by a carbohydrate-rich layer called calyx.

BV are more infectious in cell culture and are only poorly infectious by oral infection route. The ODV are more infectious by the oral route and exhibit low infectivity for tissue culture cells. Both the forms contain virally encoded proteins (108, and references therein) in their envelope that differ significantly between BV and ODV and this might contribute to the difference in their infectivity profile.

The genome of baculoviruses is a double-stranded, closed covalent circular DNA molecule, the size ranging from 82kb in hymenopteran viruses (32; 76) to 180kb in *Xestia c-nigrum granulovirus* (52).

1.2 Baculoviruses: Applications in Molecular Biology and Biotechnology

The primary interest in baculoviruses stemmed from their ability to control insect pest populations and hence their use in bio-control. For the past three decades, their use as expression vectors for the production of recombinant proteins in insect

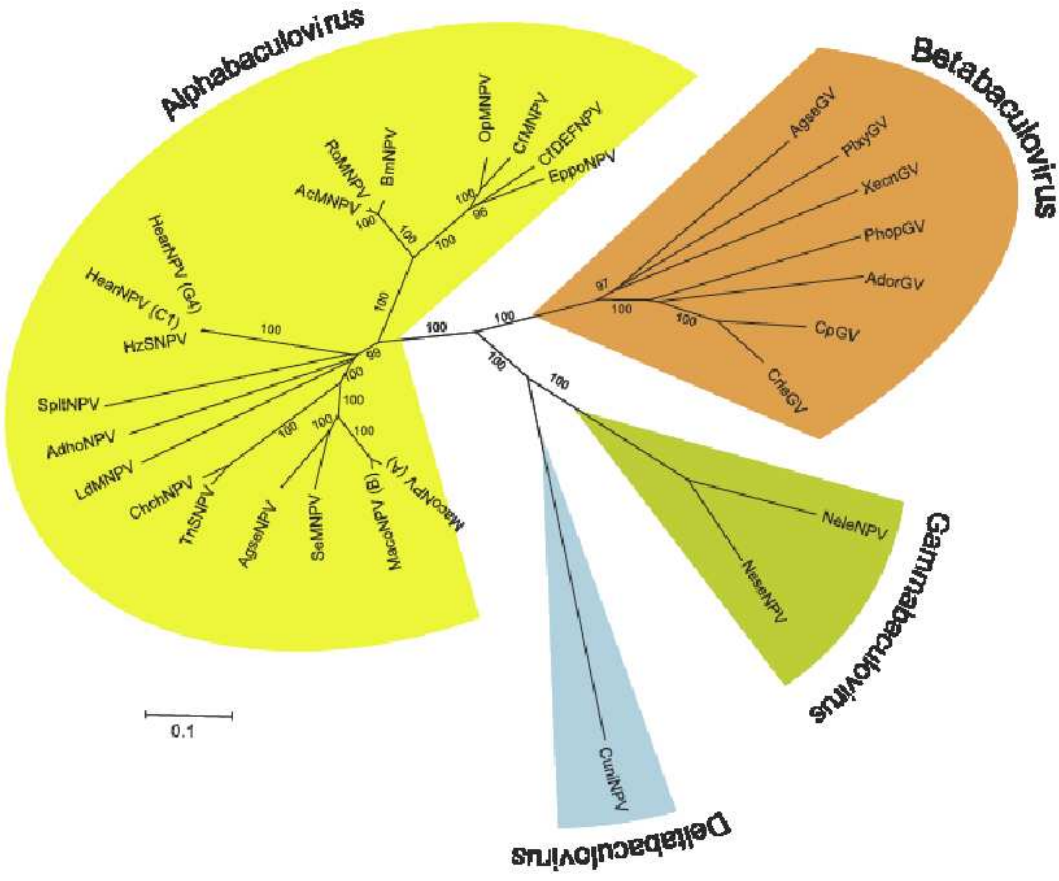


Figure 1.1: Unrooted Neighbor-Joining tree constructed by Jehle et al (61) based on sequence alignment of 29 core baculoviral genes belonging to five functional classes of genes namely, transcription, replication, genes encoding structural proteins, auxillary proteins or genes of unknown function. The sequences are from 29 genera of baculoviruses sequenced so far. Reprinted with permission (Fig 1. of (61)).

cells has been realised, which in turn has paved the way to their widespread use in many laboratories and an increasing interest in their biology.

1.2.1 Baculovirus Expression System

*Ac*MNPV and *Bm*NPV are the most well characterized and employed viruses for expressing foreign genes in the Baculovirus Expression Vector System (BEVS). The BEVS is the system of choice for the expression of proteins and has several advantages that are not present in bacterial expression. The key features are discussed below:

High Level of Expression

The baculovirus expression takes advantage of the fact that during the infection cycle, large amounts of viral proteins are synthesized during the very late stage of the life cycle, that are not essential for the replication of the virus and thus the genes coding for these highly expressed proteins can be replaced with the gene of interest. Generally, promoters of either p10 or the polyhedrin gene are used for the expression. The levels of expression from these promoters are very high ($\approx$ 20% of total mRNA) (90).

Eukaryotic Environment

The insect cell culture provides an eukaryotic environment for the proper folding of the expressed protein and also provides various post-translational modifications (N/O-glycosylation, amidation, signal cleavage, phosphorylation, prenylation, carboxymethylation, etc.) (50; 132) (reviewed in (4; 118)). Fall army worm

(*Spodoptera frugiperda* (*Sf*9) cell line) is the most commonly used host in BEVS, some of the other hosts are silkworm (*Bombyx mori*), cabbage looper (*Trichoplusia ni*), salt march caterpillar (*Estigmeneacrea*) and mosquito (*Aedes albopictus*).

The host insect cells are usually grown in suspension cultures and the protein can be purified either from lysed cells or from the growth media, if the protein carries a signal peptide. The additional advantage is the recombinant viruses can also be fed to insect larvae and in some cases, the larvae produced over 26 times more recombinant protein than High 5 cells (*Sf*9 cells) in culture (115).

Simultaneous Expression of Multi-Protein Complexes

Baculovirus vectors have been developed for simultaneous expression of more than one protein. This is useful particularly for studying as well as purifying multi-subunit protein complexes. Earlier the expression of two or more proteins was achieved by co-infecting two or more recombinant viruses. Genetically engineered transfer vectors were later developed to simultaneously express multiple genes under the control of different viral promoters both in cell culture (18; 50) and in *Spodoptera frugiperda* larvae (62). Baculovirus vectors now exist for the expression of two, three, four proteins (126, and references therein), and multiple protein subunits, using a single baculovirus that makes use of the Cre-mediated recombination at *loxP* sites to insert genes in transfer vector either sequentially or tandemly in a rapid, one-step procedure in *E. coli* (6).

Expression of Unspliced Genes

The genome of the baculovirus can accommodate large inserts and also allows the expression of unspliced genes. Davriche and collegues demonstrated that recombi-

nant baculovirus carrying the unspliced gene of the immediate early protein *(ie-1)* of the human cytomegalovirus that has two introns, can be properly spliced out in baculovirus infected cells (24).

1.2.2 Baculoviruses as Mammalian Cell Expression Vectors

The host range of *Ac*MNPV is limited not because of its inability to enter and uncoat, but because of its inability to replicate and transcribe genes in non-permissive cells. Manipulation of *Ac*MNPV to infect mammalian cells has resulted in gene transfer and expression mediated by recombinant *Ac*MNPV carrying mammalian gene regulatory elements. Modified vectors have been successfully developed using CMV promoter and high level reporter gene expression was observed in human and rabbit hepatocytes *in vitro* (56). Boyce and colleagues described a modified *Ac*MNPV virus containing *E. coli lacZ* under the control of Rous sarcoma virus and reported the expression from different mammalian cell lines. They concluded that since *Ac*MNPV can accommodate large insertions and can only replicate in insect cells, it is a highly effective gene delivery vehicle for hepatocytes. *Ac*MNPV has been employed as a useful vector for genetic manipulation of liver cells (13). For a review of the present status of baculovirus as gene delivery agents, please refer (34; 59).

1.2.3 Baculovirus Surface Display of Proteins

Boublik et al were the first to describe the use of *Ac*MNPV as a vector for the display of distinct proteins on the surface of budded virions in a manner analogous

to "phage display" (12). The baculovirus protein gp64 is abundantly expressed during the baculovirus infection cycle and transported to the cell surface to be incorporated into budded virions. When a small sequence encoding an antigenic epitope is fused with an extra copy of gp64 sequence and is expressed simultaneously along with the inherent gp64, both are displayed on the surfaces of the baculovirus infected cells. Baculoviruses have been employed to study protein–ligand selection by cloning antigenic epitopes along with an extra copy of gp64, expressed and thus displayed on the surface of baculovirus infected cells (27).

1.2.4 Role of Baculoviruses as Biopesticides

The ability of naturally occuring baculoviruses to control insect populations was realized much before the revolution of recombinant DNA technology. But the rapid progress made during the 1950s and 60s in the synthesis of new insecticides and pesticides that acted across a broad spectrum of insects and took less time to control insect pests when compared with baculoviruses that have a slow kill time and are host (pest)-specific, prevented baculoviruses in being employed commercially. But with the advent of genetic tools to manipulate baculoviruses, and the realization that excess use of chemicals to control pest populations poses environmental damage, many recombinant baculoviruses have been developed, that express various insecticidal toxins like *B. thuringiensis subsp. kurstuki* HD-73 endotoxin and *B. thuringiensis subsp. uizuwai* 7.21 CrylA(b) toxin, insect cell hormones that interfere with the life cycle of insect pests (eg. Juvenile hormone esterase (JHE)) etc. These recombinant viruses exhibit lesser killing time and show considerable efficacy as wild type viruses (reviewed in (10; 122; 148)). By far, baculoviruses

are the ideal candidates for use as biocontrol agents since they do not replicate in non-insect hosts, providing a safer and ideal way to control insect cell populations.

1.3 *Ac*MNPV: The Type Species of Baculoviridae Family

The *Autographa californica* Multiple Nucleopolyhedrosis Virus is the type member of the Nucleopolyhedrosis genus of baculoviridae family (Please refer Table 1.1 on Page 8 for a complete list of baculoviridae members). It is the most characterized and exploited among baculoviruses. The 133kb genome is a double-stranded, closed covalent circle DNA molecule and it has been sequenced completely (2).

The biology of the infection cycle of the *Ac*MNPV in the insect cell underlies the utility of baculoviruses as expression vectors. The *polh* gene is not essential for viral DNA replication and for the production of infectious particles in cell culture, and is expressed at very high levels during the final phase of infection. The polyhedrin protein is required for generation of occlusion bodies (occ^+ phenotype) which helps the virus to infect a new insect host, but is not required for formation of budded form of the virus that is capable of infecting insect cells in cell culture. The basic theme of the baculovirus expression system is to generate a $polh^-$ recombinant virus in which the *polh* gene is replaced by the desired foreign gene to be expressed and is placed directly under the control of *polh* promoter. Hence, the recombinant viruses are easily identified by their occ^- phenotype.

1.4 Genome organization of *Ac*MNPV

The *Ac*MNPV genome has been sequenced completely (2). The genome map of *Ac*MNPV is given in Figure 1.2. The genome is comprised of a circular DNA of 133,894bp and has an overall A + T content of 59%. The genome codes for atleast 154 genes that are distributed evenly on both the strands. Homologous regions (*hr*) act as origins of replication (*ori*) and are distributed periodically throughout the genome (Figure 1.2 on Page 24 and Figure 1.3 on Page 25). So far, nine homologous regions have been identified. Except for *hr4C*, all the other *hr*s contain two–eight 30bp palindromes with an *Eco*RI site as the palindrome core (72). *hr*s apart from acting as *ori* for DNA replication also function independently as enhancers (41; 143). Recently Willis et al, reported that the genome of *Trichoplusia ni* single nucleopolyhedrovirus, unlike all other nucleopolyhedrosis viruses, does not contain *hr* sequences (147).

The analysis of the nucleotide frequencies for start codons of the 154 ORFs suggested that all of them start with ATG as the start codon and the analysis of codon usage indicate that the predominant translation termination codon utlized is TAA (117 of 154 genes (76%)) (2).

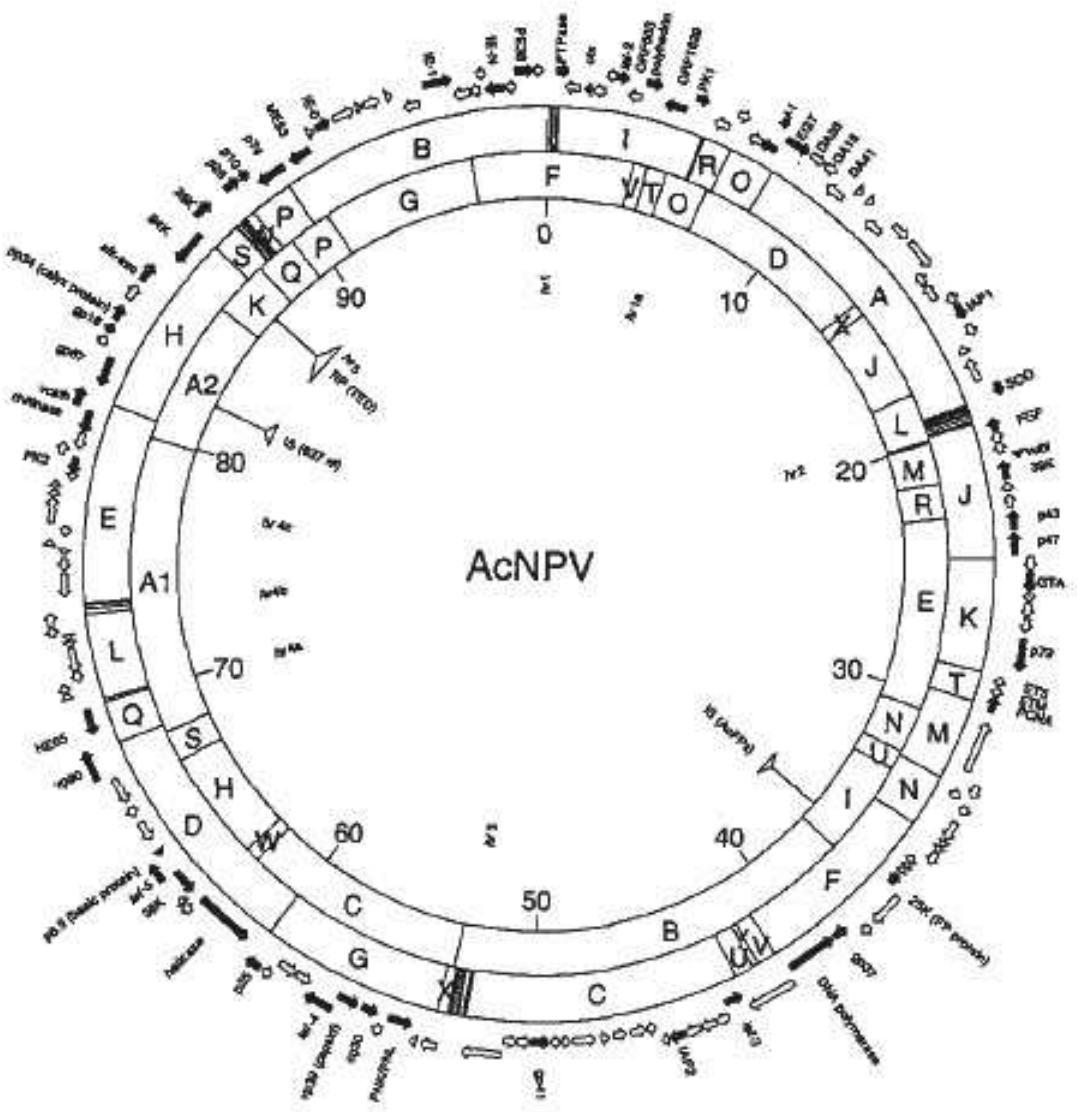


Figure 1.2: The genome map of *Autographa californica Multi Nucleocapsid Polyhedrosis Virus* (Taken from Fig 2. of (2)). The outer ring indicates the restriction fragments of EcoRI and the inner ring that of HindIII. The 154 ORFs are indicated as arrows representing the direction of the transcription of these genes. The homologous regions (*hr*s) are indicated in the inner circle.

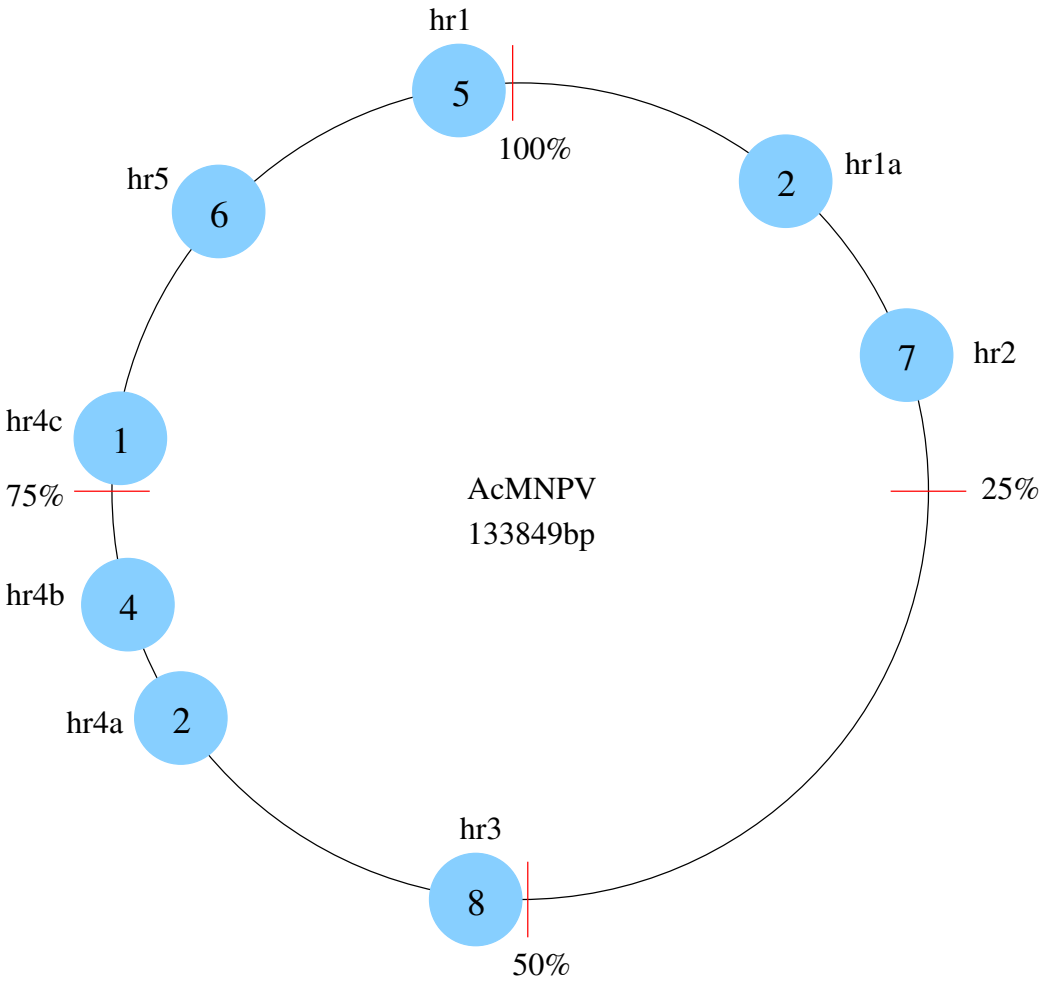


Figure 1.3: The location of homologous regions (*hr*s) in *Ac*MNPV genome (Modified with permission from Fig. 1 of (108)). The number inside the circles indicate the number of repeats present in a particular *hr* sequence.

1.5 *Ac*MNPV DNA Replication

DNA replication occurs between 6–20 hpi. The DNA replication of *Ac*MNPV requires both *cis*–acting sequence elements (*hr*) that initiate DNA replication and *trans*–acting gene products. The importance of *hr* sequences in DNA replication has been well documented. They were initially postulated to function as *ori* because of their symmetric location in the genome, palindromic structure, and high A + T content. A single palindrome with an intact core was shown to be sufficient for *hr* plasmid replication in *Ac*MNPV-infected cells (46; 80).

Deletion of the *Ac*MNPV *hr*2 and *hr*5 sequences abolished the ability of plasmids to replicate in the standard infection-dependent replication assay, while deletion of *hr1*, *hr3*, and *hr4a* sequences decreased but did not eliminate plasmid replication (150). Moreover, plasmids carrying 11 different early promoter regions or plasmids carrying complete *ie-2* and *pe38* genes or the *ie-1* upstream region were also found to be able to replicate in *Ac*MNPV-infected cells. Wu and colleagues proposed that DNA replication in *Ac*MNPV does not require *hr*s for initiation and virtually any sequence can be replicated when plasmids (including pUC18 that does not carry a *hr* sequence) and viral replication essential genes are introduced at the same time. However, when the plasmid was introduced inside the cells prior to viral infection, the replication machinery of *Ac*MNPV being sequence specific could only replicate those plasmids that carried certain viral sequences (eg. *hr*s) (150).

A Non-*hr ori* present in *Hin*dIII-K region has also been reported in *Ac*MNPV and is present as tandemly repeated sequences in defected genomes (77) and sequences within the *Hin*dIII-K region can also support plasmid replication in tran-

sient replication assays (72). Based on competitive PCR, comparison of the number of molecules representing the *Hin*dIII-K region and *ie-1* origins as well as the non-*ori polH* region in a size-selected nascent DNA preparation revealed that *Hin*dIII-K region is utilized 14 times more efficiently than the *ie-1* region during the late phase of infection and *Hin*dIII-K also remained the more active *ori* throughout the early and middle replication phases indicating that the replication of DNA in *Ac*MNPV involves multiple *ori* that are activated with vastly different efficiencies during the viral infection cycle (45).

A number of *Ac*MNPV genes involved in DNA replication have been identified and fall into either essential (*p143* (helicase), *lef-1*, *ie-1*, *lef-2* and *lef-3*) or stimulatory genes (*dnapol*, *p35*, *ie-2*, *lef-7* and *pe38*) (72; 81; 86). *p143* gene product binds to double-stranded DNA, but not to single-stranded DNA in a sequence independent manner, suggesting that its function in DNA replication is mediated by binding to DNA (75).

In transfection based plasmid replication assays, a subset of the *lef*s supported plasmid replication; IE-1, *LEF-1*, *LEF-2*, *LEF-3*, *p143*, and *p35* were essential for plasmid replication, while *ie-N*, *lef7*, and *dnapol* had stimulatory effects (86). Nine gene products have been implicated in *Ac*MNPV DNA replication (53), they are: LEF-1, LEF-2, LEF-3 (SSB-Single Stranded DNA Binding Protein), P143 (HEL), DNAPOL (DNA Polymerase), IE-1, IE2 and P35.

LEF-1 shows clear similarity to eukaryotic primases (3) and interacts with LEF-2 in yeast two-hybrid assays (29) and it has also been shown that mutation of an invariant aspartic acid in the putative primase domain abolished LEF-1 activity. LEF-2 co-purifies with LEF-1 in both ssDNA-cellulose and DEAE resin suggesting that LEF-2 might bind both to DNA as well as LEF-1 (95). LEF-3

is a single-stranded DNA binding protein and co-purifies with alkaline nuclease (AN) of *Ac*MNPV and this data suggested the possibility that this might function in homologous recombination in *Ac*MNPV genome (94). LEF7 is required for *Ac*MNPV DNA replication in *Sf*9 cells but not in *Tn*368 cell lines (19) and this phenomenon is similar to that of *p35* and *ie2*, but so far, the role of LEF7 in DNA replication is unclear (53). HEL protein has a leucine zipper motif, a putative nuclear localization signal, and seven amino acid motifs previously identified in a number of proteins involved in NTP binding and DNA/RNA unwinding (83).

Hefferon et al had reported that all LEFs invovled in replication with the exception of P143 and P35 resided in the nucleus of transfected cells and based on yeast two-hybrid data as well as subcellular localization in transfected cells by subcellular fractionation and immunofluorescent microscopy had constructed a model of *Ac*MNPV DNA replication (53) (Fig. 1.4 on Page 29). This model explains the role of LEF-3 which is a single-stranded DNA binding protein (SSB), primase-helicase complex (comprised of P143, LEF-1 and LEF-2) and the DNA polymerase on *Ac*MNPV DNA replication. Recent findings suggest that though the replication might be mediated through these proteins other factors also play important roles. The replication of baculovirus genome might be dependent on homologous recombination event to effectively replicate the genome of correct size (109). In many other viruses such as herpes simplex virus, T4 and lambda phages, homologous recombinaton is an essential event that precludes DNA replication (108).

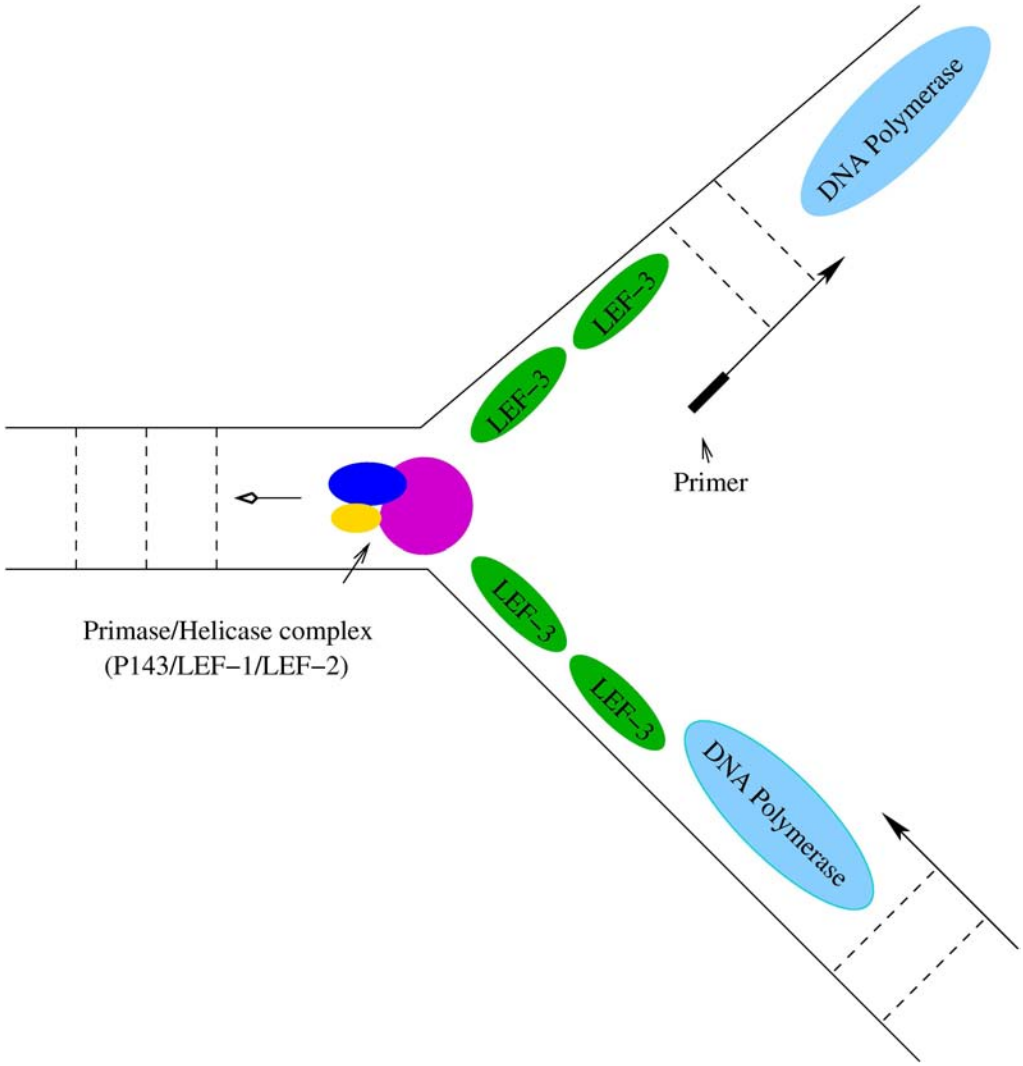


Figure 1.4: *Ac*MNPV DNA replication complex. DNA synthesis is indicated by long arrows. (Modified with permission from Fig 5. of (53))

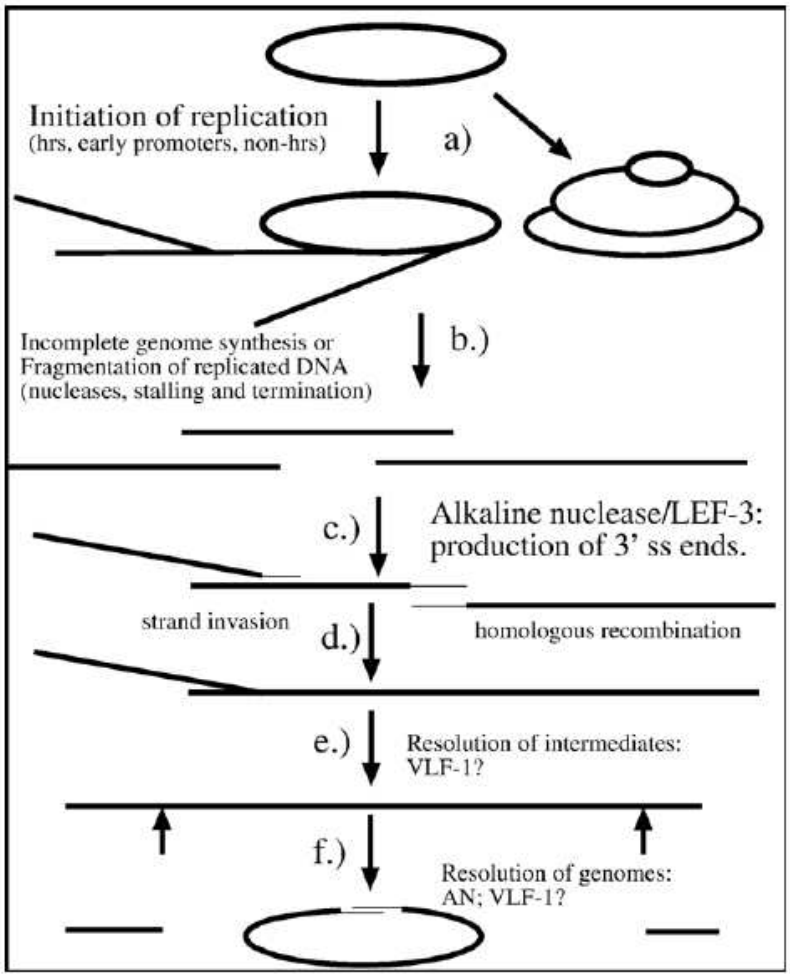


Figure 1.5: DNA replication in *Ac*MNPV involves homologous recombination. The DNA fragments generated by DNA replication followed by recombination are bigger than the viral genome. AN and LEF-3 are involved in homologous recombination. VL-F-1 has been shown to play a role in the resolution of these intermediates, but the mechanism of the action of VLF-1 is unclear. Reprinted with permission from Fig. 4 of (108).

According to this model (Fig. 1.5 Page 30), replication is initiated at various sequences (*hr*, non-*hr* and promoter sequences bound by RNA polymerase II enzyme) in the genome. The replication then proceeds through a rolling circle or theta mode. This results in incomplete synthesis because of stalling, termination or through the action of nucleases. The enzyme alkaline nuclease (AN) acts on linear DNA by digesting them in a 5'→3' fashion generating free 3' ends. AN is associated with LEF-3, a SSB protein. Through strand invasion, homologous recombination occurs and the resulting intermediates are much bigger than the actual genome in size. VLF-1 and AN have been implicated in the role of resolving intermediates. The deletion of the gene *an* results in a smaller genome size although DNA replication occurs unaffected (108). The exact mechanism of the action of these two proteins in processing the replication intermediates is unclear.

1.6 *Ac*MNPV Gene Transcription and Regulation

The infection cycle of the baculovirus takes place in three phases: early, late and very late and is regulated by a cascade of gene expression. The early gene expression takes place between 0-6h post infection (hpi). Transcription of early genes starts usually at an initiation sequence of CAGT, which is the first initiator element to be discovered in baculoviruses (9). The early promoters resemble typical eukaryotic class II promoters and are transcribed by the host RNA Polymerase II. The early gene products are required for viral DNA replication and late gene expression. *p35* gene, is an important early gene required to suppress the apop-

totic response of *Spodoptera frugiperda* cells to *Ac*MNPV infection by acting as a suicide inhibitor of caspases, as well as an antioxidant. Four early genes, *ie-0*, *ie-1*, *ie-2* (or *ie-N*) and *pe-38*, have been shown to be important for transactivating early baculovirus promoters in transient expression studies. The late gene expression spans about 6–20 hpi and the very late gene expression occurs during 20–60hpi. The late and very late genes are transcribed by an α-amanitin and tagetitoxin resistant, virally encoded RNA polymerase (35; 151). The four subunits of this enzyme have been identified and they are: LEF-4, LEF-8, LEF-9 and P47 (42). Eighteen AcMNPV genes called late expression factors(LEFs) have been implicated, by transient expression in co-transfections, to have regulatory roles in transcription from the *polyhedrin* promoter (137).

1.6.1 Early Gene Transcription

Early gene transcription takes place between 0–6 hpi and involves the transcription of genes responsible for very late gene transcription and DNA replication. The early gene transcription is mediated by host RNA polymerase II. Nuclear extracts from uninfected *Sf*9 could accurately initiate transcription from adenovirus promoter and nuclear extracts from a human cell line could initiate transcription from the baculovirus immediate early *gp64* promoter indicating the fact that the transcription of *Ac*MNPV early genes involves RNA polymerase II of the host cell (57).

Most of the early genes contain a conserved mRNA transcription initiation site that consists of the consensus sequence CAGT (8; 9) and possesses a TATA box located between 25–30bp upstream of the transcription start point (TSP) in addition

to DNA elements recognized by the host transcription machinery (68). Four early genes, *ie-0* , *ie-1*, *ie-2* (*ie-N*) and *pe38* are known to transactivate early baculovirus promoters in transient assays (121). IE-1 is the principle early transregulator of *Ac*MNPV and plays a central role in stimulating early viral transcription (79; 124). The *ie-1* gene has been shown to activate several early promoters including 39K, *ie-2* and p35. IE-1 can transactivate transcription from the *p143* promoter and the activation was found to be augmented by the *ie-n* gene product (84). The stimulatory action of IE-1 is further enhanced when the target promoters are linked to palindromic sequences from the *Ac*MNPV homologous repeat (*hr*) regions. IE-1 binds to the imperfect palindrome that constitutes repetitive sequences within the *hr*s that are dispersed throughout the *Ac*MNPV genome (39; 40; 41; 74; 123).

Tandemly repeated *Bm*NPV *ie-1* promoter sequences were found to increase reporter gene activity by 4-fold (69). IE-2 can increase transcription from reporter genes linked to *ie*1 promoter and quantitative PCR assays established that co-transfection of plasmids carrying both *ie-1* and *ie-2* genes resulted in higher levels of *ie-1* mRNA than cells transfected with *ie-1* alone (153).

A domain rich in acidic residues and essential for transactivation (Acidic Activation Domain (AAD)) is located within the N-terminal 145 amino acids of the IE-1 polypeptide (73) and the C-terminal harbors a helix-turn-helix domain that is essential for interaction with DNA (124). Moreover, IE-1 AAD transactivation domain can enhance transcription when fused to heterlogous DNA-binding domains like that of *E. coli* Lac repressor (130) and the yeast GAL4 transactivator (124). IE-0 is an alternatively spliced product of IE-1 that has an additional 54 aminoacids (20) and the additional amino acids at the N terminus of IE-0 modify the transcriptional activation of this protein relative to IE-1 (134).

By using subtractive transient expression assays with a library of clones that span the entire *Ac*MNPV genome, Gong and colleagues reported that the gene product of p35 gene increased expression from 39K promoter (37). The baculovirus p35 gene acts as a universal suppresor of apoptosis that inhibits caspase activation in a cleavage-dependent mechanism (26) and also functions in an oxidant-dependent pathway (127). The p35 gene is stimulatory for DNA replication (70).

Hr regions, which are present at nine locations dispersed in the AcNPV genome, have been shown to act as enhancers of transcription of early genes (15; 21; 44; 78; 84; 106) and also in certain cases as origins of viral replication (44; 71; 72; 113).

1.6.2 Late and Very Late Gene Transcription

The late and very late gene expression is driven by a virally encoded, α-amanitin and tagetitoxin resistant RNA polymerase (5; 35; 42; 58). The fact that the late and very late promoters, unlike the early promoters that are transcribed by the host RNA polymerase, are transcribed by a virally induced RNA polymerase clearly suggests that this transition is the most distinct event that occurs in the baculovirus transcription cascade. The late phase of viral gene expression is defined as the viral transcription that occurs after or concurrently with viral DNA replication and does not occur when the cells have been previously treated with aphidicolin, an inhibitor of DNA polymerase (97; 144). Late gene products are present in the infected cells just after DNA replication and decrease later and very late mRNAs are hyperexpressed and appear after DNA replication and remain at a sustained level throughout infection.

The late promoters are transcribed from the consensus promoter element,

ATAAG or GTAAG and this core element functions both as the promoter and mRNA start site (9; 119; 125) and is similar to promoters recognized by RNA polymerases specific to yeast mitochondrial and T7, T3 bacteriophages (7; 9; 91; 129). Mutations extending from 8bp upstream to 6bp downstream from a TAAG sequence have a significant effect on expression from the *Ac*MNPV late gene *pcap99* promoter and a synthetic promoter comprising this 18bp is sufficient to drive the transcription (101).

Earlier, eighteen LEFs (late expression factors) were reported to be responsible for late gene expression (87) and later one more additional factor LEF-12 was also found to play a role (120). LEF-12 mutant *Ac*MNPV exhibited a two-fold reduction in late gene expression and was also shown to be not essential for DNA replication (43). These same eighteen genes were also reported to be involved in transient expression of the late basic *p6.9* gene (136), the very late *polh* (110) as well as the very late *p10* genes (136).

LEF-5 has been shown to contain a zinc ribbon domain with homology to RNA polymerase II elongation factor TFIIS from several taxa and is implicated in late gene transcription. It was also shown that it can interact with itself in yeast two-hybrid system as well as in glutathione-S transferase affinity assays (48).

The very late gene promoters of *polh* and *p10* require the 12bp initiator AATAAGTATTTT and TAAGTATT seems to be the major determinant for promoter activity. Mutations within this 8bp sequence blocked transcription initiation and resulted in a 2000-fold reduction in reporter activity (110). The transcription of very late promoters depends not only on the initiator promoter, but also by the presence of a downstream A+T rich region corresponding to the 5' untranslated leader sequence of their mRNAs (110; 146). This downstream sequence has been

given the name "burst" sequence since a burst of very late gene transcription is observed during the very late phase of *Ac*MNPV infection in *Sf*9 cells. The presence of the "burst" sequence is unique to very late promoters and has not been observed in any of the *Ac*MNPV late promoters. Mutations in the burst sequence resulted in a 20-fold reduction in reporter activity (110) in contrast to a 4-fold decrease in activity resulted by mutations in the polyhedrin upstream sequence (119). A sequence 69 nucleotides upstream to the normal position of the polyhedrin ATG translation initiation codon was sufficient for maximum promoter activity but this was reduced by 90% when only 56 nucleotides upstream remained (116).

McLachlin et al reported the identification of *VLF-1* (very late transcription factor1) that is responsible for the regulation of very late promoters like *polh* and *p10*, but does not play a major role in transcription of late gene promoters like *orf603* and *vp39* in a temperature sensitive mutant *ts*B837 which produced occlusion-defective virus (92). It was further shown that *VLF-1* can enhance transcription from very late, but not late promoters in transient expression assays (136).

Yang et al had shown by DNaseI footprinting that VLF-1 binds to burst sequences present in the very late promoters like *polh* and *p10* and linker scan mutations within these burst sequences impaired interactions between VLF-1 and the promoter (152). Purified VLF-1 could support transcription driven by the baculovirus induced RNA polymerase from very late but not late promoters *in vitro* (98). VLF-1 remains the only factor identified so far, that specifically regulates very late, but not late gene expression. VLF-1 belongs to the integrase (Int) family of proteins. Based on the binding ability of VLF-1 to various forms of DNA, it was found that VLF-1 exhibits DNA structure-dependent binding with

the highest affinity for cruciform DNA (96), indicating the possibility that VLF-1 may play a role in the processing of branched DNA molecules at a late stage of DNA replication as well as packaging of genomes (108). Significantly, recent evidences support this hypothesis. In VLF-1 knockout *Ac*MNPV, the replication of DNA is one-third of that of the wild type virus, but no DNA is detectable in the infected cell supernatant suggesting that the DNA synthesized by the mutant virus was not able to assemble into virions (139). Immunoelectron analysis of cells transfected with *vlf-1* knockout bacmid showed that the aberrant tubular structures containing the capsid protein vp39 were observed, suggesting that this virus construct was defective in producing mature capsids. Moreover, it was also evident from immunoelectron microscopy of *Sf*9 cells transfected with a recombinant virus containing a functional HA-tagged version of VLF-1, that VLF-1 localized to an end region of the nucleocapsid (140).

Guarino et al reported the purification of the virus-specific RNA polymerase (42). Earlier reports suggested that the virally-induced RNA polymerase was composed of either 8 (151) or 10 (5) putative subunits . Surprisingly, contrary to these reports the purifiedSiddharthan RNA polymerase was found to be composed of only four equimolar subunits, namely LEF-4, LEF-8, LEF-9 and P47, making it the simplest of the RNA polymerases reported till date from a eukaryotic source. Interestingly, these four subunits had not been shown to be functional in *in vitro* reconstitution experiments. Nevertheless, the late and very late promoters were transcribed with equal efficiency pointing to the possibility that additional factors are needed for correct temporal regulation (42).

LEF-8, the largest (101kDa) of the 4 subunits, had earlier been shown as harboring a conserved sequence motif, GXKX4HGQ/NKG found in DNA-dependent

RNA polymerases from a diversity of prokaryotic and eukaryotic organisms (112). Moreover, mutations within this 13AA motif, abrogated late gene expression (135).

LEF-4 contains signature motifs of GTP:RNA guanylyltransferases and exhibits RNA triphosphatase and NTPase activities similar to the mRNA capping enzymes of yeast and vaccinia virus (38), suggesting that baculoviruses may encode a full complement of capping enzymes (63). Thus, LEF-4 acts as a capping enzyme besides being a part of the RNA polymerase complex. P47 is localized to nucleus, supporting the hypothesis that it might be involved in late gene expression (16), but seems to be unrelated to other transcription factors (108). LEF-9 contains a motif, NTDCDGD which is similar in 5 out of 7 positions to a highly conserved Mg^{++} binding motif (NADFDGD) found in the largest subunits of DNA-directed RNA polymerases (85). The D residue is responsible for binding to Mg^{++} and is conserved in all these RNA polymerases (108).

1.7 Role of Host Factors in Very Late Gene Expression

Although, the components of the virally-induced RNA polymerase have been identified (42), the requirements of the host factors in driving the late and very late gene transcription cannot be ruled out completely. The earliest report that host proteins might regulate very late gene expression was by Etkin et al (28). They reported the presence of a host factor from uninfected *Sf* 9 cells binding to the 96bp minimal *polyhedrin* promoter (110). Based on DNaseI footprinting they showed that positions –86 to –72 of the 96bp fragment were protected by the binding

of the host factor. Co-transfection experiments involving a plasmid carrying 96bp promoter fragment (pUC19-96) and a reporter plasmid in which the β-gal reporter was driven by *polh* promoter, the expression of β-gal was higher when the amount of co-transfected pUC19-96 plasmid was higher indicating that the host protein might negatively regulate the *polh* promoter (28). There were no further reports regarding this host factor from *Sf*9 cells.

Our lab has been using the *Ac*MNPV *polh* promoter as a model to study the involvement of host factors in the very late gene expression in *Sf*9 cells and has identified the role of several host factors that play a crucial role in *polh* transcription (49).

The theme of our lab has been to identify the role of host factors in transcription from baculoviral very late promoters. The first protein to be identified from our lab was the polyhedrin promoter binding protein (PPBP). PPBP binds to the initiator sequence of the polyhedrin promoter and is required for the transcripion from the promoter. A hexanucleotide sequence (AATAAA) within the promoter is important for binding along with the neighboring TAAGTATT present at the transcription start point. The phosphorylation of PPBP is required for its function since dephosphorylation of PPBP abolished its binding to DNA (14). PPBP also exhibits a sequence-specific single-stranded DNA-binding activity. It binds to the coding strand of polyhedrin geSiddharthanne promoter with very high affinity but it does not bind to the non-coding strand (102). PPBP was also shown to bind to the baculovirus p10 promoter with similar binding affinity to that of polyhedrin promoter (60). It was also shown that PPBP is characteristically different from TBP present in *Spodoptera frugiperda* cells (33). The exact funtion of PPBP is unknown in insect cells, but it has been suggested that it can function as an

intiator binding protein. PPBP is required for *polh* transcription both *in vitro* and *in vivo*. A simplistic model has been proposed by Ghosh and colleagues on the role of PPBP on transcription from the *polh* promoter (33). According to this model, PPBP might function by binding to the hexa- and octa- motifs of the polyhedrin promoter and then can recruit the virally induced polymerase thus initiating transcription.

Sp-family of proteins were previously not known to occur in insect cells. A novel Sp family-like factor was reported for the first time from insect cells and was shown to bind to Sp1 binding sites present ≈700bp upstream to the polyhedrin promoter (117). The presence of Sp1 binding sites upstream to the promoter could rescue the transcription from a mutated polyhedrin promoter where the hexanucleotide motif had been mutated (from AATAAA→CCCCCC). Viruses carrying deletion in the 700bp region upstream to the promoter exhibited lower reporter activity. The activity could be however restored in a recombinant virus carrying the viral Sp1 binding element cloned in lieu of the deletion, confirming the role of these factors on the polyhedrin promoter transcription. Moreover, both the *Sf*9 and HeLa nuclear extracts showed a similar binding pattern to AcSp (the Sp1-binding site present upstream to the polyhedrin promoter) and consensus Sp1-binding motifs in EMSA suggesting that both the binding factors might be similar (117).

Host factors from *Sf*9 cells were also shown to bind *hr1* sequence present ≈5kb upstream to the promoter. A 38kDa novel hr1-BP (homologous region-binding protein) was reported from *Sf*9 nuclear extract and plasmids that carried *hr1* sequences, sequestered hr1-BP and abolished the *hr1*-mediated enhancement of reporter gene expression *in vivo* (44). The presence of hr1-BP is not restricted to insect cells alone and hr1-BP was also reported from the nuclear extracts of

mammalian cells. Mammalian hr1-BP exhibited differences with the insect cell hr1-BP in terms of ion requirements, DNA groove binding and molecular size (143). It was also shown that *hr1* can enhance transcription from heterologous promoters like CMV and *Drosophila* hsp70 (143).

1.8 Objectives of Present Research Work

The baculovirus expression vector system utilizes the *Autographa californica* multiple nuclear polyhedrosis virus (*Ac*MNPV) very late hyperactive polyhedrin gene (*polh*) promoter to drive the transcription of foreign genes. Despite the widespread use of the polh promoter, we are only beginning to understand the molecular basis for basal and hyperactivated transcription from the promoter. Although the role of several viral genes in transcription from the polyhedrin gene promoter has been elucidated, the role of host factors has only recently been shown. The current knowledge regarding the role of *cis*–acting elements upstream of the promoter and *trans*–acting proteins in hyperactivation from this promoter is limited. An understanding of regulation of *polh* expression would provide insights into host-pathogen interactions, which may also have implications in human pathogenic virus systems. It will also potentially lead to the establishment of a cell free gene expression system, which would be devoid of problems such as secretory load most often encountered in the baculovirus expression system. The aim of this study is to analyze the sequences upstream to the core promoter element for the presence of *cis*–acting elements and to identify and characterize host factors that may interact with these elements to augment transcription from the polyhedrin gene. This work describes the role of the 4kb upstream sequence on the activation of transcription

from the polyhedrin gene promoter. The objectives of this study are:

1. Functional assessment of baculovirus 4kb polyhedrin upstream sequence by transient expression studies using plasmid constructs.

2. Analysis of baculovirus 4kb upstream sequence for the presence of *cis*–acting elements by comparison with known eukaryotic *cis*–acting elements.

3. Biochemical characterization of the protein(s) interacting with these sequences.

Chapter 2

Materials and Methods

Chapter 2

Materials and Methods

2.1 Materials

Tissue culture plasticware, media and reagents, enzymes and kits, radioisotopes, membranes, X-ray films, Phosphorimagers and chemicals were obtained from the following sources:

Amersham Biosciences, UK (presently GE Healthcare Biosciences, UK) X-ray films (Hyperfilm MP), Hyperfilm cassettes, Rediprime DNA labelling kits, Hybond–N^+ nylon membrane.

Applied Biosystems : DNA sequencing reagents.

Boehringer Mannheim Gesellschaft mit beschränkter Haftung (GmbH), Germany (presently Roche Applied Science): Poly dI.dC, DNA-binding protein purification kits.

Gibco BRL Inc, USA (presently Invitrogen, USA): Grace's Insect cell

medium, Foetal Bovine Serum, Antibiotic-antimycotic solution, DNA Molecular Weight Markers, Lipofectin, Dulbecco's Minimal Essential Medium (DMEM).

Clontech Inc, USA : BacPAK™ Baculovirus Expression Vector System kits.

Hi-Media, India : Agar agar, tryptone, Yeast extract.

JONAKI, India : Radioisotopes.

NEN DuPont, USA : Radioisotopes- α ^{32}P dATP, α ^{32}P dCTP, γ ^{32}P ATP (3000 Ci/mmol).

New England Biolabs, USA : Restriction Enzymes, T4 DNA Ligase, T4 DNA polymerase, Klenow fragment of *E. coli* DNA polymerase I.

Novagen : Fastplax™ Titer Kit.

Perkin Elmer Inc, USA : dNTPs, Taq DNA polymerase.

Pierce, USA : Bicinchonic acid (BCA) kit for protein estimation.

Promega Inc, USA : Luciferase Assay kits, DNA molecular weight markers.

Qualigens, India : Acetic acid, acetone, ammonium acetate, boric acid, glucose, glycerol, HCl, isopropyl alcohol, KCl, KH_2PO_4, NaH_2PO_4, NaOH, Tris, etc.

Sigma, USA : Acrylamide, APS, β-mercaptoethanol, BSA, bromophenol blue, $CaCl_2$, DTT, EDTA, EGTA, Ethidium bromide, Ficoll, Formic acid, Lactalbumin, Low gelling temperature agarose, Magnesium acetate, $MgCl_2$, neutral red staining solution, NP-40, PMSF, piperidine, Sephadex G-50, TEMED, Tryphan blue, Urea, Xylene cyanol etc.

United State Biochemicals, USA : Agarose, bis-acrylamide, PVP, SDS, sonicated Salmon sperm DNA etc.

2.2 Methods

2.2.1 Cell culture

Maintenance of *Sf*9 Insect Cells

Spodoptera frugiperda cells (*Sf*9) were grown at 27°C in TNMFH medium containing 10% foetal bovine serum and 1% antibiotic-antimycotic solution (Invitrogen, USA). TNMFH medium is Grace's basal insect cell culture medium containing lactalbumin and yeast hydrolysate. 1L of TNMFH medium (incomplete medium-ICM) contains 46.3g of Grace's medium, 3.33g each of lactalbumin and yeast hydrolysate and 0.35g of Sodium bicarbonate ($NaHCO_3$). The pH was adjusted to 6.2 with 10N KOH and the medium was sterilized by filtration under aspetic conditions using 0.22μ filter (Millipore, USA) in a laminar flow hood. For monolayer cultures, 1×10^6 of *Sf*9 cells were added to a 25cm^2 tissue culture flask containing 5mL of complete medium (CM). Cells were checked for viability by staining with 10% tryphan blue and only cells with $> 95\%$ viability were used for the experiments.

Virus Infection and Amplification

The cells were infected with *Ac*MNPV C-6 strain at 10 MOI. Before infection, the CM was completely removed and the viral inoculum was added to achieve the 10^7 pfu/mL. The infection was allowed to proceed for 1 hour, with gentle intermittent

rocking to allow uniform cell infection. The viral inoculum was removed and CM was added and the cells were incubated for 60 hours at 27℃ for luciferase reporter assay (or for nuclear extract preparation) or for 5 days at 27℃ for amplifying the virus.

Determination of Viral Titer

The titration of wild type (ie C6) AcMNPV or recombinant viruses (vMAluc and vDSXluc) were determined according to the protocol provided by the manufacturer (Novagen, USA) using Fastplax titer kit.

Construction of recombinant viruses

1.5 million *Sf*9 cells were seeded in 35mm dishes and incubated at 27℃ for 2hrs. The recombinant virus, vDSXluc was constructed using the BacPak Baculovirus Expression System (Clontech, USA) with pDSXKNluc as the tranfer vector. pDSXKN-luc DNA was diluted to 100ng/μL with autoclaved MilliQ water and 5μL of Bac-Pak6 viral DNA(*Bsu*36I digest) was added. The volume was made upto 96μL with sterile water. 4μL of Bacfectin was added to the DNAs, mixed gently and incubated at room temperature for 15mins to allow the transfection reagent to form complexes with the DNA. Meanwhile, the complete medium was removed from cell monolayers and 1.5mL incomplete medium was added. The Bacfectin-DNA mixture was added dropwise to the medium while gently swirling the dish to mix. The dishes were incubated at 27℃ for 5hrs. 1.5mL of complete medium was added and the dishes were incubated at 27℃. $\sim$ 72hrs after addition of the Bacfectin-DNA mixture to the cells, the medium which contains virus produced by the transfected cells was transfered to a sterile container and stored at 4℃ for

Plaque Assay. To obtain more virus, a fresh aliquot (1.5mL) of complete medium was added to each dish and incubated at 27℃ for another 2 days and the medium was harvested as above.

2.2.2 Molecular Cloning

Standard procedures described by Maniatis et al were followed for the manipulation, transformation of plasmid DNA in *E. coli* DH5α cells (128). Positive clones were identified using restriction digestion analysis and the clones were confirmed by DNA sequencing (ABI Prism 377 system) .

Construction of Plasmids

pKNluc is a derivative of pVL1393 containing the Firefly luciferase gene cloned in the BamHI site downstream to the polyhedrin promoter (33; 117). pAJpolluc is a pUC-derived plasmid containing the polyhedrin promoter cloned in the HincII-BamHI sites and luciferase gene cloned downstream in the BamHI site (60). pKNluc carries both the 4kb upstream as well as the downstream sequences of the polyhedrin promoter. pAJpolluc does not carry either the upstream or downstream sequences.

Construction of pKNluc derivatives pKNluc vector was chosen to study the effect of deletions in the 4kb upstream region on the expression of reporter luciferase levels driven by the polyhedrin promoter. pKNluc was digested with the restriction enzymes shown in Figure 2.1, the sites were end-filled with Klenow fragment of *E. coli* DNA polymerase I and religated to generate the deletion constructs: pDMMKNluc (MluI–MluI deletion), pDMXKNluc (MluI–XhoI dele-

tion), pDSXKNluc (SacII–XhoI deletion) and pDSNpolluc (SacII–NdeI deletion). Oligonucleotides containing the binding sites of potential transcription factor binding sites (Listed in Table 2.1 on Page 51) were synthesized with SacII-XhoI restriction ends and were cloned in the SacII-XhoI sites of pKNluc in lieu of the SacII-XhoI region.

Construction of pAJpolluc derivatives pAJpolluc was used as the basal vector to clone upstream sequences from the 4kb upstream region. pSXpolluc, pSNpolluc were constructed by PCR amplification of the the SacII-XhoI and SacII-NdeI fragments respectively from the pKNluc plasmid and subsequently cloned in the PstI site of pAJpolluc. The primer sets respectively were: SX-F: 5'GATATCCTGCAGCCGCGGGGTATTGAACCGCGCGAT3', SX-R: 5' GATATCCTGCAGCTCGAGGTGCAGCGAGTCAACGCG and NS-F: 5' GATATCCTGCAGCATATGCGGTGTGAAATACCGCAC3', NS-R: 5'GATATCCTGCAGCCGCGGACAAGCGCAAACATCCGG3'.

The MluI-XhoI region of the polyhedrin upstream sequence was divided into three sub-regions (A:1-200bp, B:170-360bp, C:320-574bp) (Figure 3.8 on Page 74). The A-fragment was amplified using the primers A-f (5'GATATCCTGCAGCTCGCTGCACCTCGAGCAGTTCGT3') and A-r (5'GATATCCTGCAGTATGCGCAAACAACCCAACTGTAT3'), B-fragment using the primers B-f (5'GATATCCTGCAGAAAATATATACA GTTGGGTTGTTT3') and B-r (5'GATATCCTGCAGTTATTCCACACTTTGATCACTTGA3') and the C-fragment using the primers C-f (5'GATATCCTGCAGAATCGATGCAAGTGATCAAAGTGT3') and C-r (5'GATATCCTGCAGCAATCAAAGCTCGTGCCGGAACGC3'). All the three fragments were then cloned seperately in the PstI site of pAJpolluc.

Similarly, the SacII-XhoI region that contains the HSE-like elements, was also divided into three regions: full length (pSXpolluc contained the full length of the

SacII-XhoI region, described above), pSXBpolluc (containing the first 500bp of the SacII-XhoI region) and pSXApolluc (containing only the first 250bp of the SacII-XhoI region). pSXBpolluc was obtained by cloning the PCR product amplified with the primers b-sx-1-250-f: 5' GATATCCTGCAGGGTATTGAACCGCGCGATCCGACAAATCCA3' and b-sx-1-500-r: 5' GATATCCTGCAGCGACCCGCTGTATTTGCAGCCGCATACAGT3' in the PstI site of pAJpolluc. PSXApolluc was cloned by amplifying the the first 250bp of SacII-XhoI region using b-sx-1-250-f (described above) and b-sx-1-250-r 5'GATATCCTGCAGGCGCAAAAAACCGAGGAACTTGTTAAAAAA3' primers and the PCR amplified product was cloned in the PstI site of pAJpolluc.

Oligonucleotides containing the potential transcription factor binding sites (Listed in Table 2.2 on Page 52) were synthesized with PstI restriction ends and were cloned in the PstI site of pAJpolluc.

Colony Hybridization

In order to identify positive clones after transformation of *E. coli* with the ligation mixture, colony hybridization was carried out. Colonies grown overnight at 37℃ on LB plates were transferred to Hybond-N+ membrane (Amersham Biosciences, USA). The membrane was washed with 0.5N NaOH twice, to denature the DNA and then the alkali was neutralized by washing the membrane in 1M Tris-HCl (pH 7.7) for 5 minutes, followed by 0.5 Tris-HCl (pH 7.5). The membrane was then washed in 1.5M NaCl for 5 minutes, air dried and the DNA was cross-linked to the membrane by UV crosslinking (UVP CL-1000 crosslinker (1200 kJ)). Prehybridization, hybridization and autoradiography were carried out as described in Section 2.6.4 on Page 58.

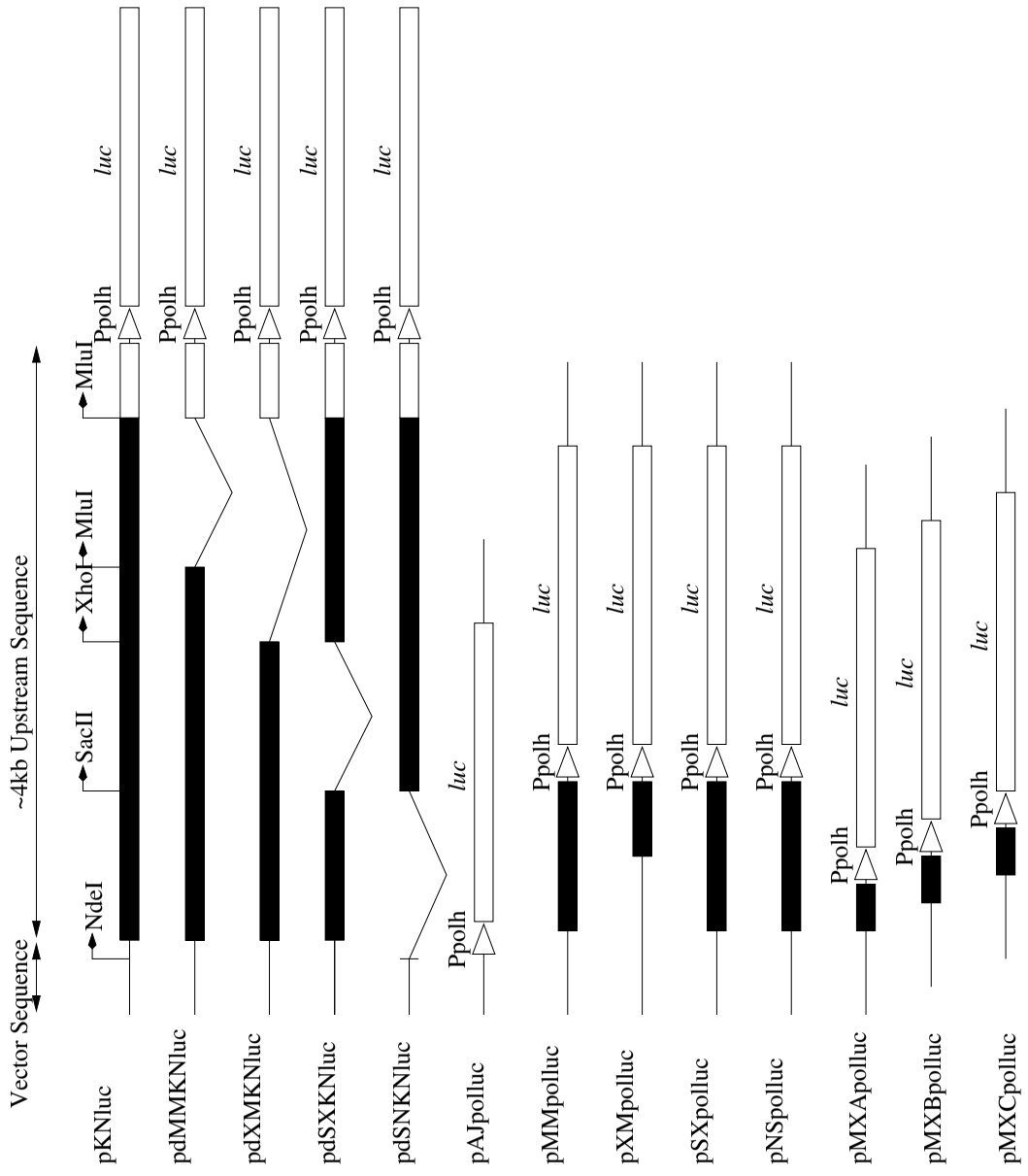


Figure 2.1: Derivatives of pKNluc and pAJpolluc

Table 2.1: Oligonucleotides cloned in pKNluc

Oligo Name	**Sites present**	**Top strand sequence (5'→3')**
HSE+	Combined HSE-like sequence and some other transcription factor binding sites (eg. GATA-1, AP-1, Cdx-A) from the SacII-XhoI region	GGTTTTCTGCTTTCTTCGCAATCAGCTTAGTCACCCTTCTTCTACATTCTTC
HSE	HSE-like elements present in the SacII-XhoI region	GGTTTTCTGCTTTCTTCTTCTTCTACATTCTTC
vHSE	Contains the exact HSE-like sequence present in the SacII-XhoI region	GGGTCTTTTTTCTGCATTATTTCGTCTTTCTTTTGC

Table 2.2: Oligonucleotides cloned in pAJpolluc

Oligo Name	**Sites present**	**Top strand sequence (5'→3')**
HSE	Combined HSE-like sequence from the SacII-XhoI region	AGCTTTTTTCTGCTTTCTTCGCAATCAGCTTAGTCACCCTTCTTCTACATTCTTCTGCA
HSE+O	Combined HSE-like sequence and some other transcription factor binding sites (eg. GATA-1, AP-1, Cdx-A) from the SacII-XhoI region	AGCTTTTTTCTGCTTTCTTCTTCTTCTACATTCTTCTGCA
dHSE	Consensus HSE present in Drosophila hsp70 promoter	AGCTTCGAGAAATTTCTCTCTCGTTGGTTCCAGAGACTCGAATGTTCGCGACTGCA
Ac-Oct1	Octamer binding site present in the MluI-XhoI region	GTGATTTGCATCTGCA

Computational Analyses

DNA sequences were analyzed using Sequence Analysis software (Version 3.3, PE Biosystems, USA) and DNAstar software (Version 4.4, DNAStar Inc, USA).

2.3 DNA-Protein Interaction Studies

2.3.1 Preparation of Nuclear Extract

All operations were carried out at 4℃. *Sf*9 cells from 4 T75 tissue culture flasks ($\approx$ 60 million cells) were pelleted at 2000rpm (Rotor no.12148, Sigma *2K15* centrifuge) for 10 mins, washed with PBS and lysed in a buffer containing 1% Nonidet-40, 30mM Tris-HCl, 10mM Magnesium acetate for 5mins. The nuclei were harvested at 1600rpm (Rotor no.12148, Sigma 2K15 centrifuge) for 3mins, suspended in protein extraction buffer containing 420mM NaCl, 10mM HEPES, 10mM Magnesium Chloride, 1mM EDTA pH 8.0, 1mM EGTA pH 8.0, 1mM DTT and 25% glycerol. After 30mins, the nuclei were pelleted at 11000rpm and the supernatant was dialyzed against 1X PBS, aliquoted and stored at -70 ℃.

2.4 Radiolabelling of Oligonucleotides

100ng of annealed double-stranded oligonucleotide harboring the putative transcription factor binding site was labelled using γ ^{32}PATP (30 μCi) and 10 units of T4 polynucleotide kinase enzyme (New England BioLabs, USA). The reaction was carried out at 37℃ for 1 hour. After the completion of the reaction, the labelled oligonucleotide was seperated from unincorporated $\gamma^{32}P$ ATP by passing

the reaction mixture through a G-50 Sephadex column equilibrated with 10mM Tris pH 8.0 and 1mM EDTA pH 8.0.

2.4.1 Electrophoretic Mobility Shift Assay

EMSA was carried out as described in (117). 1ng of the labelled oligonucleotide was incubated with 5–10μg of nuclear extract in the presence of binding buffer containing 10mM Tris-HCl pH 7.5, 120mM NaCl, 1mM EDTA pH 8.0, 10mM PMSF, 5% glycerol and 1μg of poly-dI.dC at 25℃ for 30mins. The DNA-protein complex was resolved at 4℃ in a 6% (29:1 acrylamide-bisacrylamide) polyacrylamide gel in 0.5X TGE buffer (Tris Glycine EDTA). After electrophoresis, the gel was covered with plastic wrap, dried and exposed overnight in either Hyperfilm cassette at -70℃ or PhosphorImager cassette (Amersham Biosciences, USA) at room temperature.

For HSE, the consensus *Drosophila* heat shock element (top strand: (5'→3') AGC TTC GAG AAA TTT CTC TCT CGT TGG TTC CAG AGA CTC GAA TGT TCG CGA CTG CA) was synthesized, annealed and used for EMSA reactions. The Oct1 sequence used for EMSA was (top strand: (5'→3' TGT CGA ATG CAA GCC ACT AGA A) (Santa Cruz, USA).

2.5 Determination of DNA Major/Minor Groove Interaction

To determine the interaction of Oct1-like factor from *Sf*9 cells with DNA, EMSA binding reactions were also carried out in the presence of Distamycin A (Sigma,

St.Louis, USA) and Methyl Green (PolySciences Inc, USA). 1ng of radiolabelled Oct-1 was incubated with difference concentrations of either Distamycin A (0.25,0.5,1,2mM), Methyl Green (0.25,0.5,1,2mM) or actinomycin D (0.01mM–0.3mM) for 30 mins at room temperature. 10μg of *Sf*9 nuclear extract was then added and the tubes were further incubated at room temperature for 15 mins. The results were then analyzed by resolving in a 5% polyacrylamide gel as described in Section 2.4.1.

2.5.1 UV Crosslinking Analysis of DNA-Protein Interaction

The binding reaction was carried out as described in Section 2.4.1 on page 54. After the binding reaction was completed, the reacting mixture containing the labelled oligonucleotide probe and *Sf*9 nuclear extract was transferred to a fresh sheet of Parafilm and exposed to 1200 kJ of UV in a UVP CL1000 UV cross linker. After UV exposure, equal volumes of SDS loading buffer (0.00625M Tris-Cl pH 6.8, 2% SDS, 5% β-mercaptoethanol, 10% glycerol, 0.025% bromophenol blue) was added to each tube. The tubes were boiled for 5 minutes at 100℃ in a heat block and the samples were electrophoresed on a 10% SDS-polyacrylamide (29:1 acrylamide:bisacrylamide) gel in a Tris-glycine buffer (25mM Tris, 192mM glycine pH 8.3, 0.1% SDS). After electrophoresis, the gel was dried in a gel dryer, wrapped in a plastic wrap and autoradiographed as described earlier.

2.5.2 Southwestern Analysis

100μg of crude nuclear extract was denatured at 37℃ in SDS sample loading buffer (10mM Tris.HCl pH 6.8, 2% SDS, 100mM DTT, 0.04% bromophenol blue) for 5

minutes before loading onto a 10% SDS-polyacryamide (29:1 acrylamide:bisacrylamide) gel. After the electrophoresis, the gel was equilibriated in transfer buffer (25mM Tris, 190mM glycine, 20% methanol) for 10 minutes. The proteins were transferred overnight at 30mA onto a nitrocellulose membrane at 4℃. The membrane was then washed for 30 minutes with three changes of binding buffer (10mM Tris.HCl pH 7.5, 100mM NaCl, 1mM $MgCl_2$, 0.5mM DTT and 0.1mM PMSF). The membrane bound protein was denatured in binding buffer containing 6M guanidine hydrochloride for 5 minutes. Half a volume of this buffer was then replaced by adding binding buffer to achieve 3M guanidine hydrochloride and the membrane was washed for 5 minutes. Subsequent washes were made similarly to get 1.5, 0.75, 0.375, 0.1875M of guanidine hydrochloride in binding buffer. The membrane was finally washed with the binding buffer without any guanidine chloride. The membrane was then incubated for 3 hours at 4℃ with gentle rocking in blocking buffer (binding buffer containing 0.5% of BSA, 100mg/mL salmon sperm DNA and 1% non-fat dried milk). The labelled Oct-1 oligonucleotide was then added (10^6 cpm/mL of binding buffer) and incubated overnight with gentle rocking at 4℃. After the hybridization was completed, the membrane was washed with ice cold binding buffer, wrapped in plastic wrap and autoradiographed as described earlier.

2.6 Transient Expression Assay

2.6.1 Transfection of Plasmid DNA Using Lipofectin in *Sf*9 Cells

Lipofectin (Invitrogen, USA) mediated transfection of *Sf*9 cells was done as described earlier (117; 143). For transient reporter assays, 1.5–2 million cells were seeded in a 6-well tissue culture plate (Corning, USA) and allowed to adhere for 1 hour. The complete medium was replaced with 2mL of incomplete medium (ie medium that lacks foetal bovine serum and antibiotics). After an incubation of 2 hours, the DNA-lipofectin mixture (5μg of plasmid DNA and 5μL of lipofectin suspension) was added in a total volume of 1mL of incomplete medium and the cells were further incubated for 6–8 hours. The incomplete medium containing plasmid DNA and lipofectin was removed and the cells were washed gently once with complete medium. *Ac*MNPV infection was then carried out by replacing the complete medium with *Ac*MNPV C6 virus (10 MOI). The tissue culture plate was given a gentle rock every 15 minutes for an hour to prevent the drying of the cells as well as to spread the virus evenly. After 1 hour of infection, the virus was removed and 2mL of complete medium was added and the cells were incubated in 27 ℃ for 60 hours.

2.6.2 Luciferase Assay

Luciferase assay was done at 60 hpi. The cells were pelleted at 1500rpm and washed once with 1X PBS. The cell pellet was resuspended in 500μL of 1X Reporter Lysis buffer (Promega, USA). After two cycles of freeze and thaw, 100μL of Luciferin

substrate (Promega, USA) was added to the lysate and the values were read in a Luminometer (Turner Designs, USA) with a delay time of 3 secs. After the readings were taken, the samples were stored in -20℃ for dot blot hybridization and protein estimation.

2.6.3 Random Primer Labelling of *luciferase* cDNA

The BamHI fragment of pAJpolluc containing the *luc* gene was labelled with ^{32}P dCTP using the ReadyPrimer random primer labelling kit (Amersham, USA). The 2kb *luc* fragment was diluted to a concentration of 2.5–25ng in 45μL of TE buffer (10mM Tris.HCl pH 8.0, 1mM EDTA pH 8.0). The DNA sample was denatured by heating to 95–100℃ in a boiling water bath, then snap cooled on ice for 5 minutes after denaturation. The denatured DNA was then added to the reaction tube containing random oligonucleotides, Klenow fragment of DNA Polymerase I, dATP, dGTP, dTTP and reaction buffer concentrate. 5μL of ^{32}P dCTP was added and the reaction tube was incubated at 37℃ for 60 minutes. The reaction was stopped by adding 5μL of 0.2M EDTA pH 8.0. Approximately 15μL of the probe was used per 5mL of hybridization buffer in dot blot hybridization.

2.6.4 Dot Blot Hybridization

To confirm that the results obtained in the transfection experiments were not due to differences in the amounts of plasmid DNA, dot blot hybridization was carried out. Either 20μL of undiluted (neat) or diluted (1/4 and 1/8) samples (ie cell lysates from luciferase assay) were blotted onto Hybond-N+ nylon membrane (Amersham, USA) using BioDot apparatus (BioRad, USA) by applying vaccum.

The membrane was air dried and UV-crosslinking was carried out in UVP CL-1000 crosslinker (1200 kJ). Prehybridization was carried out at 55℃ for 5–6 hours and then the denatured probe was added, the hybridization was allowed to proceed overnight. Prehybridization/Hybridization buffer contained 6X SSC pH 7.0, 5X Denhardt's solution, 0.05% SDS and 100mg/mL of salmon sperm DNA. After the hybridization, the membrane was washed twice in 2X SSC + 0.1% SDS for 10 minutes each at room temperature, followed by two washes for 10 minutes each in 1X SSC + 0.1% SDS at 65℃. The membrane was then air dried, wrapped in plastic wrap and kept for autoradiography.

2.6.5 Protein Estimation

Protein estimation was carried out either using Micro BCA Protein Assay kit (Pierce, USA) by following the manufacturer's instruction or by using Bradford's method. For Bradford's method, 2–5μL of the lysate was taken and 200μL of Bradford's reagent was added and the color intensity was read at 595nm in a BioRad ELISA reader. The luciferase assay values were normalized against the amount of protein present in the lysate.

Chapter 3

Results

Chapter 3

Results

3.1 Rationale

The theme of our lab's research had been to dissect the role of *cis*-acting sequence elements and *trans*-acting host factors that are involved in polyhedrin gene transcription. The first host factor characterized in our lab from *Sf*9 cells was the polyhedrin promoter binding protein (PPBP). PPBP from *Sf*9 cells binds to the core promoter element of the polyhedrin promoter (14). The core promoter (AATAAATAAGTATT) was earlier described by Morris and Miller as the minimal essential sequence required for transcription from the polyhedrin promoter (101). Apart from the core promoter sequence, other sequence elements had since been identified that play important roles in activating transcription from the polyhedrin promoter (117; 143). *hr1* which functions as an origin of replication also functions as an enhancer in an independent manner. *hr1* being a classical enhancer, functions in a position- and orientation- independent manner (44; 143). It can also enhance transcription from heterologous promoters in *Sf*9 cells as well as in mammalian

cells (143).

A host protein, hr1-BP (hr1- binding protein) had been reported from *Sf*9 cells that is essential for transcription from the polyhedrin promoter (44; 46). hr1-BP was also been reported from mammalian cells and it was shown that the mammalian and insect hr1-BP factors are distinct from each other in terms of salt tolerance, DNA groove binding, etc (143). Recombinant viruses carrying an additional copy of *hr1* can additively increase expression of foreign genes. Sequences resembling classical GC boxes that are bound by Sp-family of protein(s) were reported from *Sf*9 cells. These sequences were also been shown to rescue transcription from a mutated promoter where the core ATTAAA motif had been replaced with CCCCCC (117).

In spite of the identification of these sequence elements and the corresponding host factors,the molecular basis for high level of transcription from the polyhedrin promoter is still not well understood. The 4kb upstream sequence is bordered by hr1 at the 5' end and the polyhedrin promoter at the 3' end. This region contains the following ORFs namely, ORF504 (protein tyrosine phosphatase), ORF984 (baculovirus repeat ORF - bro), conotoxin-like peptide (ctx), ORF453 (ORF4), ORF327 (ORF5), ORF630 (LEF-2) and ORF603 (Figure 3.1).

In order to identify sequence elements in the 4kb upstream region that may play an important role in polyhedrin transcription, plasmids carrying deletions in the 4kb upstream sequences were constructed.

3.2 Deletion analysis of *polh* 4kb region

The plasmid pKNluc contains the complete 4kb upstream region of the polyhedrin gene (117). pKNluc was employed to study the possible role of this upstream sequence in driving the polyhedrin expression. Using convenient restrictions sites that were present, four separate regions were deleted from pKNluc. Plasmid constructs were derived from the parent plasmid pKNluc by restriction digestion, end filling by Klenow fragment of DNA polymerase, followed by self ligation of the resulting vector. The deletion constructs derived from pKNluc (Figure 2.1) namely, pdSXpolluc, pdMXKNluc, pdMMKNluc and pdSNKNluc were studied by transient expression in *Sf*9 cells.

3.2.1 Deletion analysis of the 4kb region showed two regions involved in the enhancement of transcription from the polyhedrin gene promoter

Transient expression assay of these deletion constructs was performed in *Sf*9 cells. pdSXKNluc that carries a 1kb deletion of the SacII-XhoI region showed a significant reduction (3–4 fold) in luciferase reporter expression, while the expression supported by pdMXKNluc that carries the MluI-XhoI deletion, was drastically reduced (10 fold) (Figure 3.4). The deletion of either the NdeI-SacII region (pdSNKNluc) or the MluI-MluI region (pdMMKNluc) did not affect expression from the polyhedrin promoter significantly suggesting that these two regions are not important for polyhedrin promoter driven transcription (Figure 3.5). In all transfections, equal amounts of plasmid constructs were used, which was confirmed

by dot blot Hybridization as described in Section 2.6.4 of Materials and Methods. A recombinant virus was constructed to confirm the role of the SacII-XhoI region in driving the polyhedrin gene expression. The plasmid pdSXKNluc was used as a transfer vector in order to construct the virus using the BacPAK baculovirus expression system. This virus carries deletion of the complete SacII-XhoI region(vdSXluc), and it showed a significant decrease in reporter gene expression when compared with the control recombinant virus (vMAluc) where the entire 4kb region is intact (Figure 3.6). In both these viruses, the polyhedrin promoter drives a luciferase gene in lieu of the polyhedrin gene. These results clearly demonstrate that the deletion of SacII-XhoI region results in significant reduction in polyhedrin transcription.

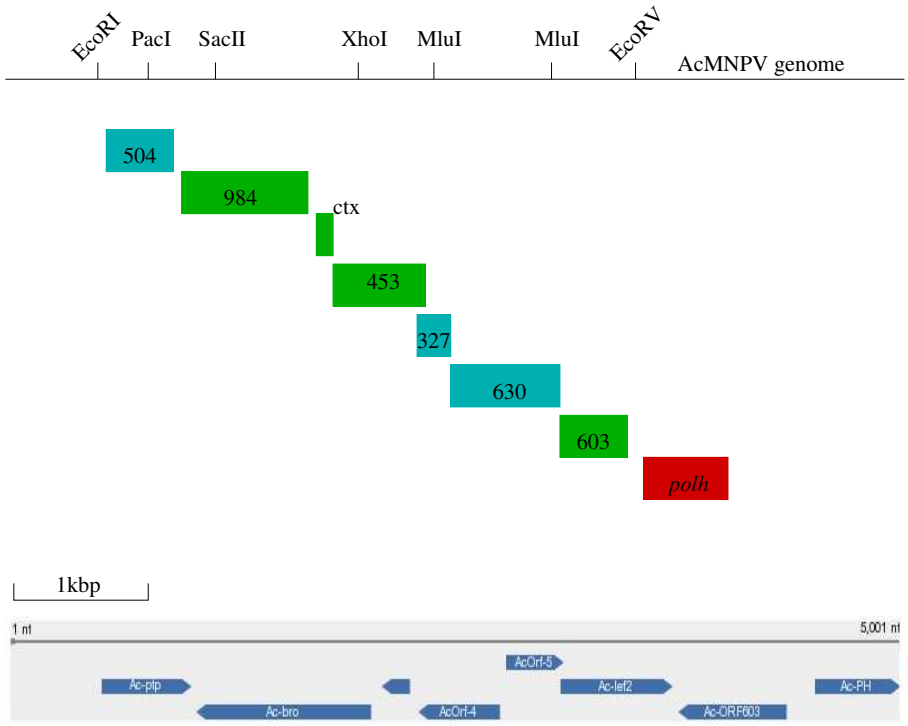


Figure 3.1: 4kb upstream region of the polyhedrin gene promoter. The ORFs that are present in the '+' strand are shown in Cyan and the ORFs that are present in the '-' strand are shown in Green. The polyhedrin gene is shown in Red ('+' strand). The Restriction Enzyme sites are shown above the ORFs and the scale is given below. Figure modified from (111). NCBI Sequence viewer map of the polyhedrin upstream region is given below that shows the corresponding new names of the ORFs.

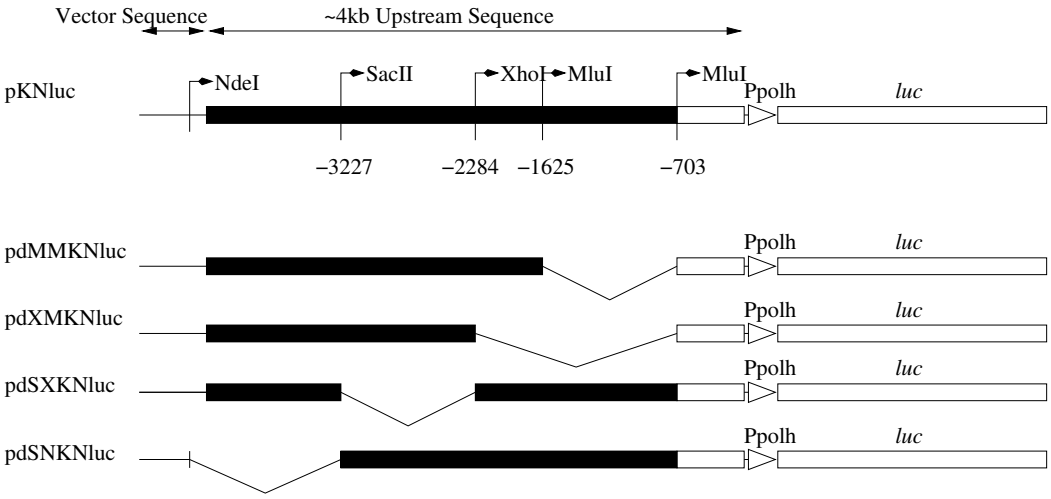


Figure 3.2: The deletion constructs were derived by digesting pKNluc with the restriction enzymes indicated above, followed by Klenow treatment and ligation of the vector backbone to generate the constructs (pdSXKNluc, pdMXKNluc, pdMMKnluc and pdSNKNluc respectively. The numbers below the restriction map indicates the position of the sites in relation to transcription start point (TS) of the polyhedrin gene.

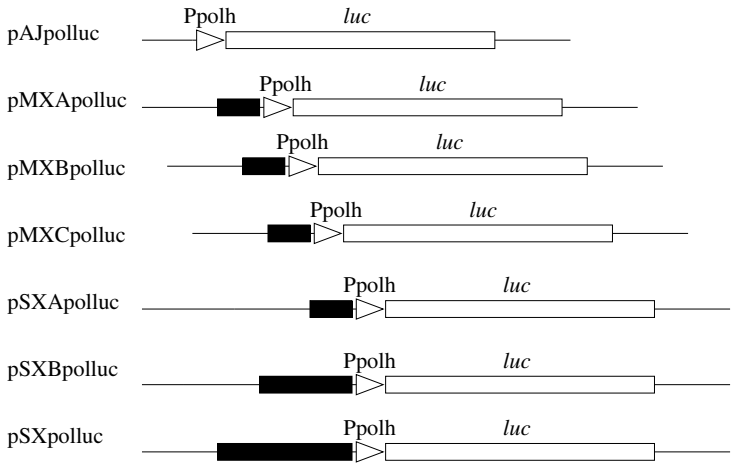


Figure 3.3: pAJpolluc is a pUC-derived plasmid containing the *polh-luc* cassette (117). The SacII-XhoI region and the MluI-XhoI region were divided into three small regions, amplified by PCR and were cloned in the PstI site of pAJpolluc upstream to the polyhedrin promoter. pSXApolluc, pSXBpolluc contain 250bp, 500bp respectively of the SacII-XhoI region. pSXpolluc contains the full-length SacII-XhoI region. MluI-XhoI region was similarly divided three regions, but these regions were divided in such a way that they overlap with one another. pMXApolluc, pMXBpolluc and pMXCpolluc contains 1–200bp, 170–336bp and 320–570bp of the MluI-XhoI region respectively.

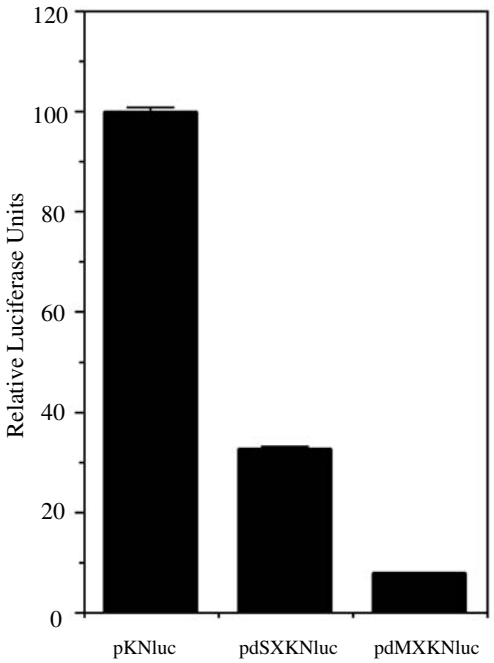


Figure 3.4: Transient expression of pdSXKNluc, pdMXKNluc, with pKNluc as control in *Sf*9 cells showed that the deletion of either SacII-XhoI or MluI-XhoI regions results in significant reduction in polyhedrin promoter driven trancription.

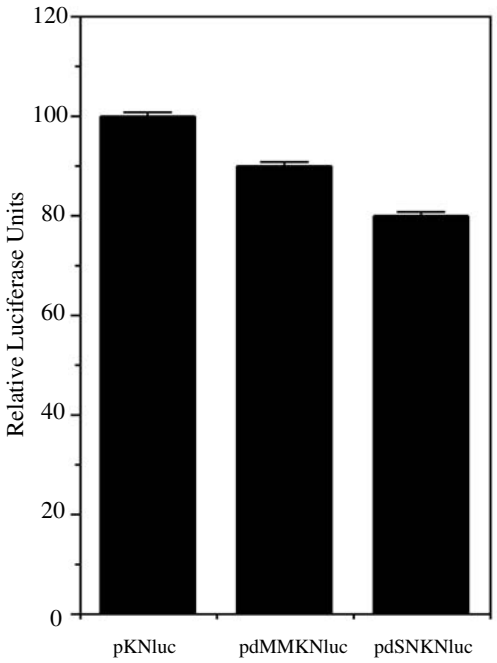


Figure 3.5: Transient expression of pdMMKNluc, pdSNKNluc, with pKNluc as control in *Sf*9 cells showed that the deletion of either MluI-MluI or SacII-NdeI regions does not result in significant reduction in polyhedrin promoter driven trancription.

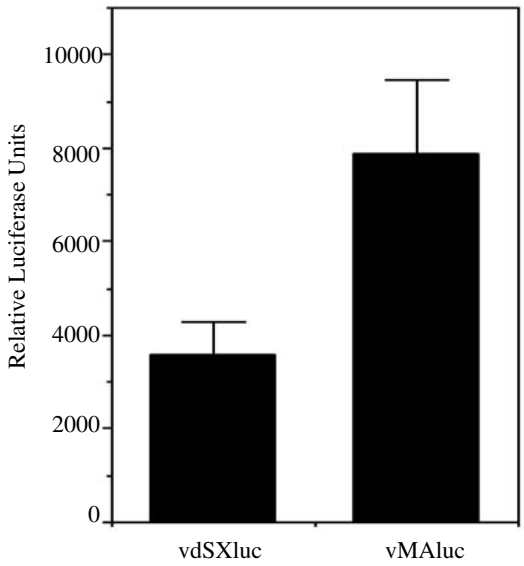


Figure 3.6: *Sf*9 cells were infected with equal MOI (10) of either recombinant viruses vdSXluc (carrying the SacII-XhoI region deletion) or vMAluc (carrying the intact 4kb region) and luciferase assay was done at 60hpi. vDSXluc showed significantly lower reporter gene expression when compared with vMAluc

3.2.2 Presence of either SacII-MluI or MluI-XhoI region upstream to *polh* promoter results in activation of transcription

Deletion of either the SacII-XhoI region or the MluI-MluI region in the 4kb upstream sequence of the polyhedrin promoter resulted in significant reduction in reporter gene expression. In order to determine, if these two regions harbor any *cis*-acting elements that regulate transcription from the polyhedrin promoter, plasmid derivatives from these two regions were constructed.

The SacII-XhoI region was divided into three regions and each region was separately cloned upstream of a *polh-luc* casette to generate three constructs. The plasmid pSXpolluc carries the entire SacII-XhoI region ($\approx$ 1kb) and the plasmids pSXApolluc and PSXBpolluc carry the first 250bp and 500bp relative to the 5' end of the 1kb SacII-XhoI region, respectively. The fragments pertaining to the corresponding SacII-XhoI region were amplified by PCR and subsequently cloned in pAJpolluc to generate the above three constructs, as described in Materials and Methods (Section 2.2.2).

All the three plasmids (pSXApolluc, pSXBpolluc and pSXpolluc) were transiently transfected into *Sf*9 cells. pAJpolluc which does not carry any upstream sequence was used as a control in these transfection experiments. When luciferase assay was carried out from *Sf*9 cells transfected with the above three plasmid constructs, all the three plasmids showed a consistent 2–3 fold increase in reporter gene expression when compared with the pAJpolluc control (Figure 3.7).

Similarly, the MluI-XhoI region (569bp) was divided into three overlapping fragments namely A (1–250bp), B (170–336bp) and C (320–574bp). Primers spe-

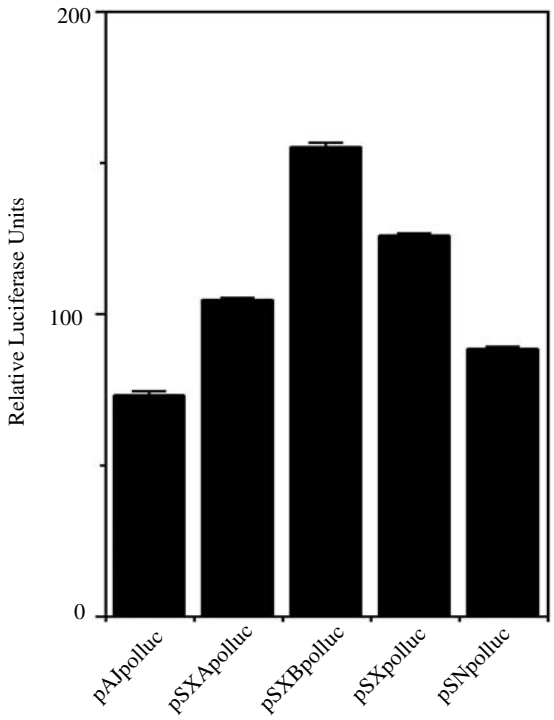


Figure 3.7: Transient expression of SacII-XhoI region in *Sf*9 cells. All the three derivatives (pSXApolluc, pSXBpolluc and pSXpolluc) showed higher luciferase reporter gene expression when compared with the control plasmid (pAJpolluc) that does not carry any upstream sequence. pSNpolluc that carries the SacII-NdeI region did not show any significant increase in reporter gene expression.

cific for these three fragments were synthesized and the corresponding region was amplified from the plasmid pKNluc. The PCR fragments were restriction digested with PstI and subsequently cloned in PstI restriction digested pAJpolluc to generate the following constructs: pMXApolluc, pMXBpolluc and pMXCpolluc respectively (Figure 3.8).

When transient expression of the MluI-XhoI region derived plasmids was carried out in *Sf*9 cells, only the plasmids pMXApolluc and pMXBpolluc showed upto two-fold increase in reporter gene expression from the polyhedrin promoter as compared to pAJpolluc. The plasmid pMXBpolluc that carries the 170–360bp fragment from the MluI-XhoI region showed the highest increment in luciferase expression. The plasmid pMXCpolluc, showed a drastic decrease in polyhedrin promoter driven luciferase reporter gene expression (Figure 3.9).

From these data, it can be concluded that the 170–360bp region corresponding to the middle portion of the MluI-XhoI region, is also important for polyhedrin promoter driven transcription. Whereas, the presence of the SacII-NdeI fragment upstream to the polyhedrin promoter had no significant increase in the reporter gene expression indicating that the SacII-NdeI region probably does not play a major role in expression from the polyhedrin gene promoter (Figure 3.7).

3.3 Identification of *cis*-acting elements in the 4kb upstream region of the polyhedrin promoter

The SacII-XhoI region and the Mlu-XhoI region were found to independently up-regulate polyhedrin transcription based on the results from transient expression assays in *Sf*9 cells. In order to identify transcription factor binding sites in these two regions, computational analysis was carried out. Sequences from both the MluI-XhoI and the SacII-XhoI regions were submitted online for the identification of transcription factor binding sites to TESS (Transcription Element Search System)(http://www.cbil.upenn.edu/cgi-bin/tess/tess). The default options for the basic parameters chosen for analysis, were as follows: Maximum allowable string mismatch % (t_{mm}) was 10, Minimum log-likelihood ratio score (t_{s_a}) was 12, Minimum string length(t_w) was 6, Minimum lg likelihood ratio (t_a) was 6.0, Maximum lg-likelihood deficit (t_d) was 8.0, Minimum core similarity (t_c) was 0.75, Minimum matrix similarity (t_m) was 0.85, Secondary Lg-Likelihood Deficit was 3.0, Count significance threshold was $1.0e^{-2}$. Sequences that showed significant similarity to already verified transcription factor binding sites are given a higher log-likelihood ratio score (t_{s_a}) in TESS. Table 3.1 and Table 3.2 show the results obtained for the SacII-XhoI region and the MluI-XhoI region, respectively.

In the Mlu-XhoI region, the Octamer binding protein (OBP or Oct-1) binding site was found to have a high log-likelihood score (20). Similarly, in the SacII-XhoI region, the heat shock elements (HSE) were found to have a high score (131).

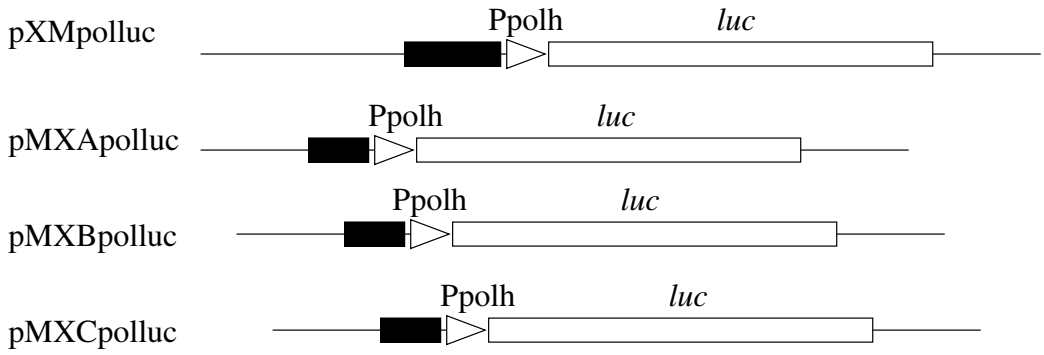


Figure 3.8: MluI-XhoI region derivatives cloned in pAJpolluc vector. The MluI-XhoI region was amplified by PCR from the parent plasmid, pKNluc and cloned in the PstI site of pAJpolluc.

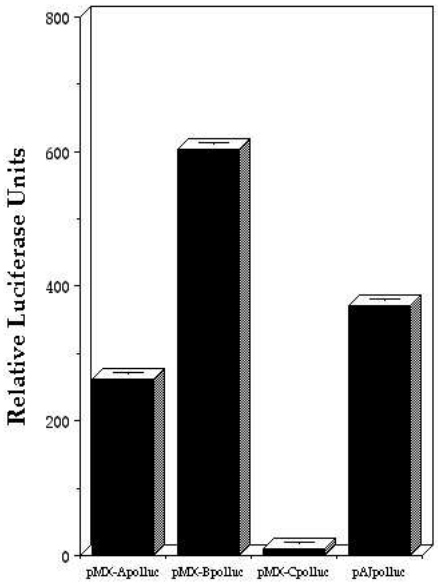


Figure 3.9: Transient expression of pMXApolluc, pMXBpolluc and pMXCpolluc in *Sf*9 cells

Table 3.1: Transcription factor binding sites present in the *Sac*II–*Xho*I region. (The position indicates the relative position of the binding site with respect to the 5‘ end of SacII-XhoI region)

Transcription Factor	Position	Length	Sequence(5'→3')	Log-Likelihood	Poisson model 'p' value (P_v)
c/EBP–α	245	10	TTGCGCCACt	18.00	-1.0e+00
HSTF	385	7	CTTCTAgAAG	18.00	-1.0e+00
AP-1	443	7	TCACCAATCTc	18.00	-1.0e+00
C/EBP–β	709	7	ATTgTCCAAT	18.00	-1.0e+00
E2F–1	907	6	CGCGGCAAA	18.00	-1.0e+00

Table 3.2: Transcription factor binding sites present in the *Mlu*I–*Xho*I region. (The position indicates the relative position of the binding site with respect to the 5‘ end of MluI-XhoI region)

Transcription Factor	Position	Length	Sequence(5'→3')	Log-Likelihood	Poisson model 'p' value (P_v)
c-myb	7	6	CAGTTC	12.00	-1.0e+00
PU.1	21	7	CTTCCTC	14.00	-1.0e+00
Sp1	50	6	GGGTGG	12.00	-1.0e+00
HES-1	86	6	CACAAG	12.00	-1.0e+00
CCAAT-binding factor	147	6	ATTGGC	12.00	-1.0e+00
C/EBP alpha	172	6	GTTGGG	12.00	-1.0e+00
GATA-3/GATA-4	188	6	ATATCT	12.00	-1.0e+00
GATA-1	191,294, 332, 434	6	TCTATC	12.00	-1.0e+00
C/EBP alpha	202	6	GTTGGG	12.00	-1.0e+00
Oct-1	227	10	TGATTTGCAT	20.00	-1.0e+00
AP-1	416	7	AGTTTCA	14.00	-1.0e+00

Hence, both these *cis*-elements were chosen for further study to identify their role in polyhedrin promoter transcription.

3.3.1 HSE-like elements are present in the SacII-XhoI region

Heat Shock Elements (HSEs) are present in the hsp70 promoter of *Drosophila* and other species and are responsible for driving the transcription of stress induced genes. HSE are bound by HSTF (Heat Shock Transcription Factor) (Reviewed in (131)). Although the SacII-XhoI region harbored numerous potential transcription factor binding sites, sequences resembling the Heat Shock Elements (HSE) with a high log-likelihood score of 18.00 were present (eg. position 385 relative to the 5' end of the SacII-XhoI region). Though sequences resembling HSE-like elements were found to be present throughout the 4kb region but with a lower score. Other potential transcription factor binding sites that were identified were for: E2F, AP-1, C/EBP-α etc. (Table 3.1).

3.3.2 HSF-like factors bind to HSE-elements present in the SacII-XhoI region

Many HSE sites with high similarity were found when the SacII-XhoI region was analyzed using TESS. Electrophoretic mobility shift assay was performed using labelled oligonucleotide containing HSE-like element from the SacII-XhoI region. A specific complex was observed when nuclear extract from *Sf*9 cells was used (Figure 3.10). Moreover, the complex could be competed out only by a 100-fold molar excess of unlabelled oligonucleotide but not by the presence of nonspecific DNA

(pUC18). This shows that factors probably resembling Heat Shock Transcription Factors (HSTF), from *Sf*9 cells bind with high specificity to the HSE-like elements present in the SacII-XhoI region.

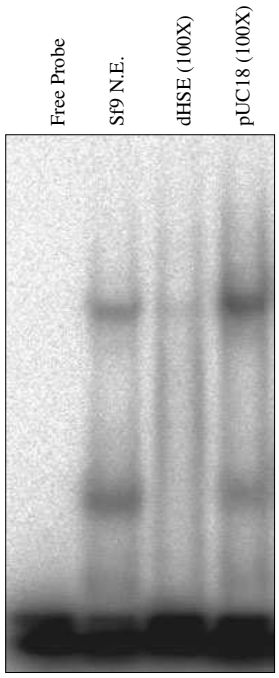


Figure 3.10: Electrophoretic mobility shift assay with dHSE-oligonucleotide. 5μg of *Sf*9 N.E. was incubated with radiolabelled dHSE oligonucleotide either alone (lane Sf9 N.E.) or in the presence of 100X excess of unlabelled dHSE oligonucleotide ((dHSE (100X)) or in the presence of 100X non-specific DNA ((pUC18(100X)). A specific complex of HSE-protein can be clearly seen.

3.3.3 Presence of HSE-elements upstream to the *polh* promoter results in increased transcription

In order to investigate if these sequences can enhance transcription from polyhedrin promoter, oligonucleotides were synthesized containing sequences corresponding to: 1) combined HSE+other transcription factor binding sites from SacII-XhoI region, 2) combined HSE-like elements from the SacII-XhoI region and 3) classical heat shock elements from the *Drosophila* hsp70 promoter, with PstI overhangs, annealed and cloned upstream to the polyhedrin promoter in the PstI site of the vector pAJpolluc (Table 3.3). The constructs thus derived were called as pHSE+polluc, pHSEpolluc and pdHSEpolluc respectively.

Transient expression of these three plasmids were carried out in *Sf*9 cells with pAJpolluc as a control. The results are shown in Figure 3.11. The presence of either HSE-like elements or the presence of *Drosophila* HSE sequences upstream to the polyhedrin promoter resulted in a consistent 2–3 fold increase in the reporter gene expression as compared to pAJpolluc and was comparable to the expression from plasmids carrying either the entire SacII-XhoI region (pSXpolluc) or the HSE + other TF binding sites. pHSE+polluc showed higher levels of reporter gene expression pointing out the possibility that, though the HSE-like elements are involved in the transcription, additional factor(s) might indeed play a possibly additive role in increasing the expression from the polyhedrin promoter.

To further determine if the HSE-like elements identified from the SacII-XhoI region can replace the enhancement that is due to the presence of the entire SacII-XhoI region, plasmid constructs were derived from the plasmid pKNluc. The experiment was to find out, if the abrogated expression from the plasmid pdSXKNluc

(carrying the deletion of the SacII-XhoI region) can be rescued by the presence of the HSE-like elements present in the SacII-XhoI region in lieu of the entire SacII-XhoI region.

Oligonucleotides corresponding to: 1) combined HSE and other elements from the SacII-XhoI region, 2) combined HSE-like elements from the SacII-XhoI region and 3) exact HSE-like sequence present in the SacII-XhoI rgion were synthesized with SacII-XhoI overhangs and cloned in lieu of the SacII-XhoI region in the plasmid pKNluc to generate, pHSE+KNluc, pHSEKNluc and pvHSEKNluc respectively. The list of constructs is given in Table 3.4. When transient expression of these plasmid constructs were carried out, all the upstream sequence elements (ie HSE, HSE+, vHSE) could enhance expression from the polyhedrin promoter and resulted in expression comparable to pKNluc which carries the entire 4kb upstream region (Figure 3.12).

These results clearly demonstrate that the presence of HSE-like elements upstream could enhance expression from the polyhedrin promoter.

Table 3.3: HSE constructs in pAJpolluc

S. No.	Construct	Binding Sites Present	Sequence(5'→3')
1.	pHSEpolluc	Combined HSE binding sites from SacII–XhoI region	AGCTTTTTTCTGCTTTCTTCGCAATCAGCTTAGTCACCCTTCTTCTACATTCTTCTGCA
2.	pHSE+polluc	HSE, C/EBP-α, E2F, AP-1, etc.	AGCTTTTTTCTGCTTTCTTCTTCTTCTACATTCTTCTGCA
3.	pdHSEpolluc	HSE from *Drosophila* hsp70 promoter	AGCTTCGAGAAATTTCTCTCTCGTTGGTTCCAGAGACTCGAATGTTCGCGACTGCA

Table 3.4: HSE constructs in pKNluc

S. No.	Construct	Binding Sites Present	Sequence(5'→3')
1.	pHSEKNluc	Combined HSE binding sites from SacII–XhoI region.	Similar to HSE oligonucleotide except instead of PstI overhangs have SacII-XhoI overhangs
2.	pHSE+KNluc	HSE, E2F, C/EBP-α, AP-1, etc	Similar to HSE+ oligonucleotide except instead of PstI overhangs have SacII-XhoI overhangs
3.	pdHSEKNluc	HSE from *Drosophila* hsp70 promoter	Similar to dHSE oligonucleotide except instead of PstI overhangs have SacII-XhoI overhangs
4.	pvHSEKNluc	exact HSE-like sequence presence in the SacII-XhoI region	GGGTCTTTTTTCTGCATTATTTCGTCTTTCTTTTGC

3.3.4 Octamer-like sequences present in MluI-XhoI region enhance *polh* gene transcription in an orientation independent manner

Based on the TESS results, it was found that the Octamer protein transcription factor binding site was present in the 170–360bp MluI-XhoI B fragment (ie pMXBpolluc).

In order to determine whether the presence of these Oct1 sites upstream to the promoter can influence the expression from the polyhedrin promoter, plasmid reporter constructs were constructs were made using synthesized oligonucleotides carrying the appropriate sequences and were cloned separately in pAJpolluc. Oligonucleotides harboring the Oct-1 binding site from the MluI-XhoI (which is 100% identical to the consensus Oct-1 sequence) were synthesized with PstI overhangs (Top strand sequence: (5‘→3‘) GTGATTTGCATCTGCA), annealed and cloned upstream to the polyhedrin promoter at the PstI site in pAJpolluc. Clones corresponding to both the forward and reverse orientations were obtained and were termed as pFoctpolluc and pRoctpolluc respectively.

Transient expression of these plasmids in *Sf*9 cells revealed that there was a consistent 4–5 fold increase in expression from plasmids carrying Oct1-binding sites either in the forward or the reverse orientation, as compared to the control plasmid, pAJpolluc. The reporter gene expression, however, was less when compared with the plasmid pMXBpolluc that carries the 170–360bp region of the Mlu-XhoI region (Figure 3.13), indicating that other sequence elements in this region probably also play a role in activation of transcription from the polyhedrin promoter.

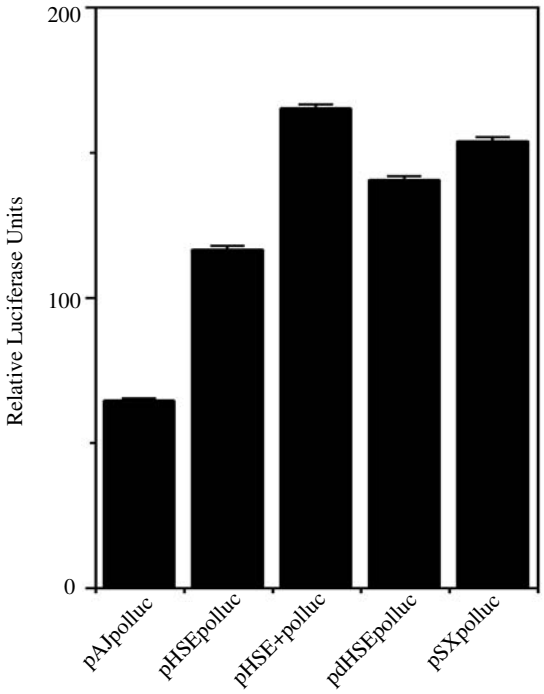


Figure 3.11: Transient expression of plasmids carrying either no upstream element (pAJpolluc), HSE-element (pHSEpolluc), HSE+other elements (pHSE+polluc) or the *Drosophila* consensus HSE (pdHSEpolluc), the full length SacII-XhoI region in *Sf*9 cells. The presence of either HSE-elements or the *Drosophila* HSE element results in a 2–3 fold increase in the reporter gene expression and is comparable to the enhancement due to the presence of the full length SacII-XhoI region.

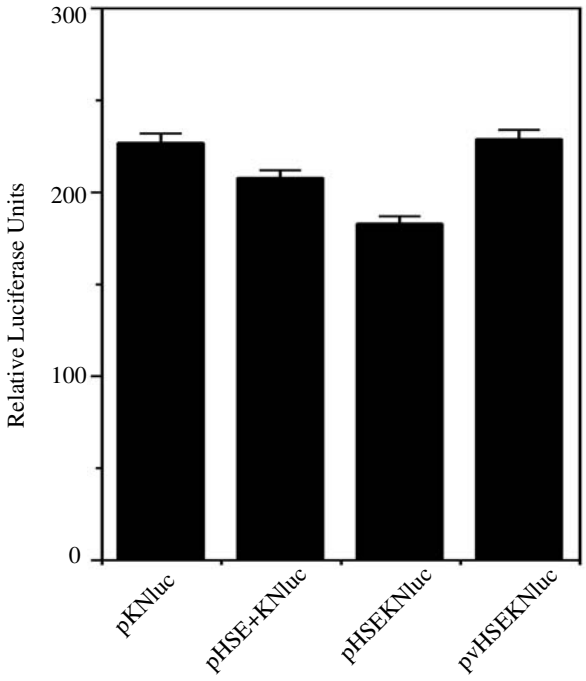


Figure 3.12: Transient expression of plasmids carrying various HSE elements in lieu of the SacII-XhoI region in the plasmid pKNluc. pHSE+KNluc that carries the HSE + other sites, pHSEKNluc that contains only the HSE-like elements and pvHSEKNluc that contains the exact HSE-like elements from the SacII-XhoI region showed comparable reporter gene expression with the plasmid pKNluc. This results show that the potential transcription factor binding sites present in the SacII-XhoI region may play a vital role in driving the expression from the polyhedrin gene promoter.

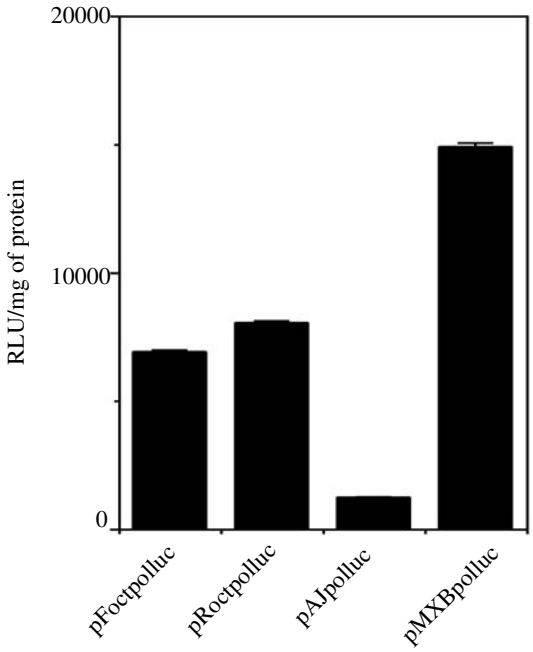


Figure 3.13: Transient expression of plasmids carrying either no upstream sequence (pAJpolluc) or octamer sequence present in the MluI-XhoI region in forward (pFoctpolluc) or reverse (pRoctpolluc) orientations. pMCBpolluc that carries the 170–360bp of the MluI-XhoI region harbors the octamer sequences. The reporter gene expression from octamer carrying plasmids either forward or reverse orientation was 4–5 fold in comparison with pAJpolluc, but was less when compared with the plasmid pMXBpolluc.

3.3.5 Oct1-like factor(s) from *Sf*9 nuclear extract bind to the octamer sequences present in the MluI-XhoI region

In order to determine if Oct-1 like protein(s) are indeed present in *Sf*9 cells, EMSA was carried out with labelled Oct-1 binding site oligonucleotide. The results show clearly that Oct-1 like POU homeobox proteins indeed are present in *Sf*9 cells and bind with high specificity to the Oct-1 site present in the MluI-XhoI region. The Oct1–*Sf*9 N.E. complex could not be competed out in the presence of a 100-fold molar excess of nonspecific competitor like pUC18 or a mutated oligonucleotide (mutOct1) in which two important bases in the consensus sequence TGATTTTGCAT had been replaced with "GC". The complex could however be competed out with a 100-fold molar excess of unlabelled Oct-1 oligonucleotide (Figure 3.14). Binding reactions were also carried out in presence of single-stranded DNA either from the top or the bottom strand of the consensus Oct1 oligonucleotide (Figure 3.15). The results showed that the bottom strand can compete effectively with the double stranded Oct1 oligonucleotide (lane 4) when compared with the top strand (lane 3).

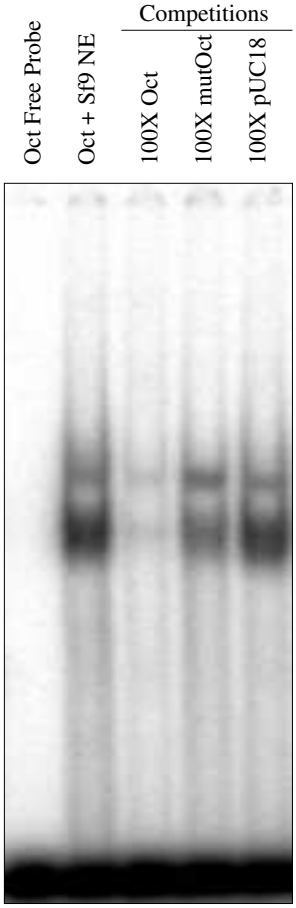


Figure 3.14: Electrophoretic mobility shift assay with radiolabelled Oct oligonucleotide shows that Oct1-like factor(s) from *Sf*9 N.E. can bind to the octamer sequence present in the MluI-XhoI region of the polyhedrin gene promoter. 10μg of N.E. from *Sf*9 cells was incubated with radiolabelled Oct1 oligonucleotide sequence. 100-fold molar excess of unlabelled Oct1 competes for binding, but mutOct1 oligonucleotide that carries a AT$\rightarrow$GC change in the sequence could not compete for binding. The binding could also could not be competed out with 100X excess of a non-specific DNA (pUC18).

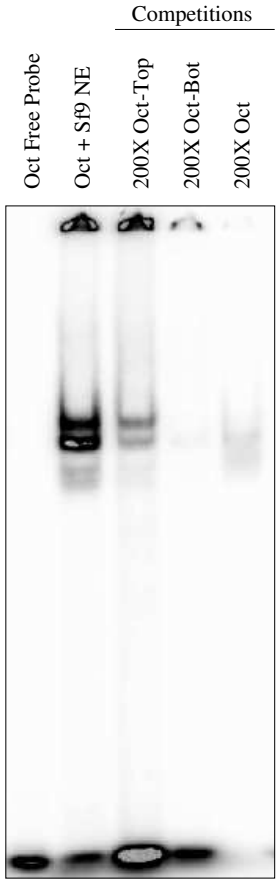


Figure 3.15: The factor(s) binding to the octamer sequence can be competed out by 200X excess of single-strand oligonucleotide of either the top or bottom strands. 10μg of *Sf*9 N.E. was incubated with radiolabelled consensus Oct1 oligonucleotide either alone (lane 2) or in the presence of unlabelled (200X) top (lane 3) or bottom strand (lane 4) or unlabelled ds-Oct1 oligonucleotide (lane 5). The complex could be competed out both by the top and the bottom strand of the consensus Oct1 binding sequence.

3.3.6 Biochemical characterization of factor(s) binding to the octamer sequence

Having shown that, Oct1-like protein from *Sf*9 nuclear extract can bind to Oct-1 binding site present in the polyhedrin upstream region, characterization of the nature of the DNA-protein interaction was carried out.

Divalent and monovalent cations inhibit the binding of the Oct-like protein to the octamer sequence

As shown in Figure 3.14, Oct1-like factors bind to the octamer sequences present in the MluI-XhoI region with high specificity. The binding could be seen even in the presence of 100X molar excess of a non-specific DNA (pUC18). However, the binding was inhibited in the presence of 500mM NaCl (Figure 3.16). Similar results were obtained when the binding reactions were carried out in presence of divalent cations like $MgCl_2$, $CaCl_2$ and $MnCl_2$ suggesting that the presence of either monovalent or divatlent cations inhibits binding (Figure 3.17, Figure 3.18, Figure 3.19).

To further support this hypothesis, binding reactions were carried out in the presence of EDTA, a divalent chelator. *Sf*9 nuclear extract was incubated with labelled Oct-1 oligonuleotide either alone or in the presence of increasing concentrations of EDTA (0,50,75,100mM). The binding however, did not decrease even at 100mM EDTA concentration (Figure 3.20 in Page 97). This experiment clearly shows that the presence of divalent cations are not important for the binding of the octamer binding protein and the effect of divalent or monovalent cations may infact be inhibitory to the binding of the Oct-like protein.

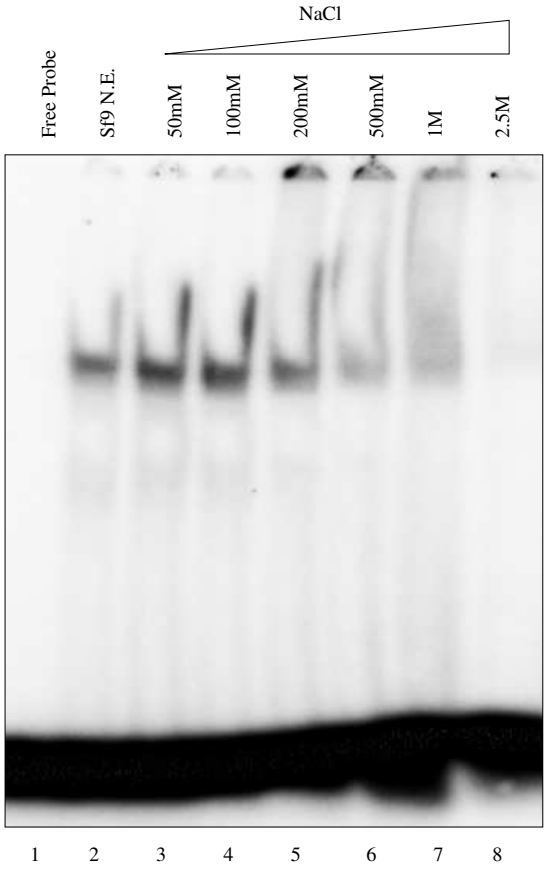


Figure 3.16: Labelled Oct-1 oligonucleotide was incubated either alone (1) or with 5μg of *Sf*9 N.E. (2). In order to study the effect of NaCl, the binding reactions were carried out in the presence of either 50mM (3), 100mM (4), 200mM (5), 500mM (6), 1M (7) or 2M (8) NaCl. The binding of the Oct-like factor diminished significantly only in the presence of 200mM (5) and was abolished in the presence of 500mM NaCl(6).

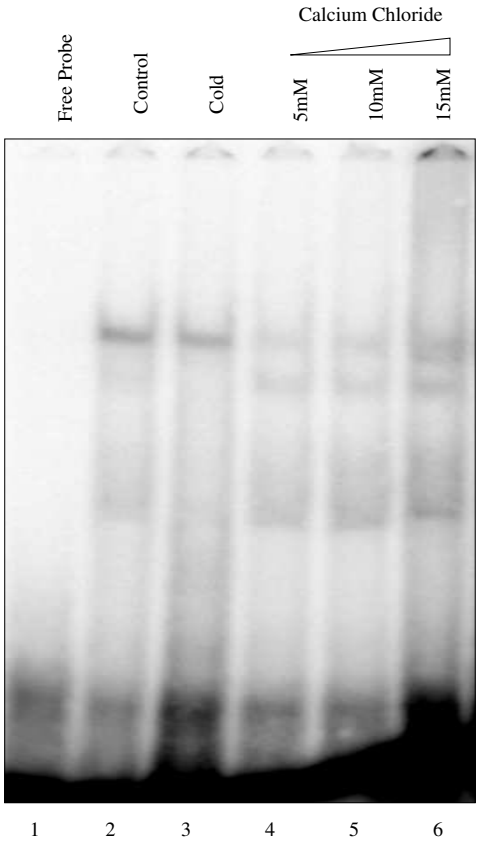


Figure 3.17: The presence of monovalent cations destabilize the octamer sequence binding. The binding reactions were carried out without $CaCl_2$ (lanes 2, 3) or with increasing amounts of $CaCl_2$(5, 10, 15mM $CaCl_2$, lanes 4,5,6 respectively). The binding was inhibited at the lowest concentration tested (5mM $CaCl_2$).

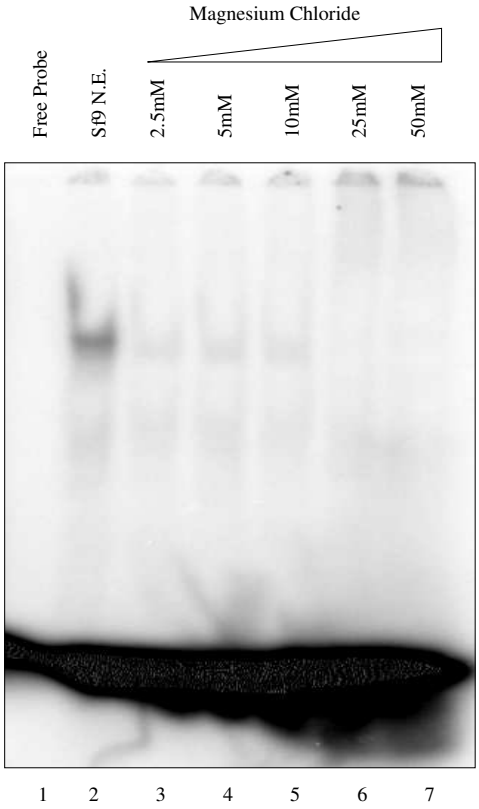


Figure 3.18: Binding reactions between radiolabelled Oct1 and *Sf*9 N.E. was carried out either alone (ie without any $MgCl_2$ (lane 2) or with increasing concentrations of $MgCl_2$ (2.5mM (lane 3), 5mM (lane 4), 10mM (5), 25mM (6) and 50mM (7). Binding of the Oct1-like factor was inhibited even in the presence of 2.5mM $MgCl_2$.

Figure 3.19: $MnCl_2$ inhibits Oct1-like factor binding. Binding reactions were carried out either in the absence of $MnCl_2$ (lane 2) or with varying concentrations of $MnCl_2$ (lane 4 5mM, lane 5 10mM, lane 6 25mM, lane 7 50mM). Lane 3 shows the cold (100X unlabelled Oct1 oligonucleotide) competition.

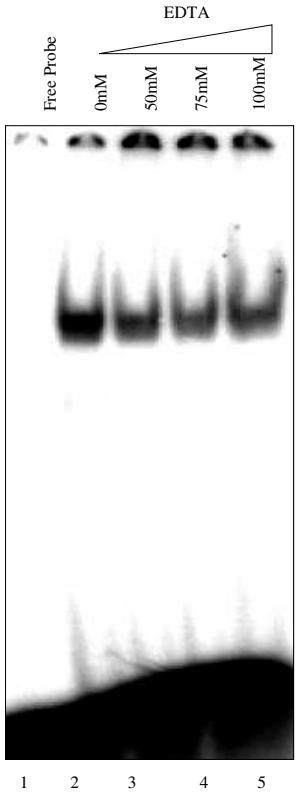


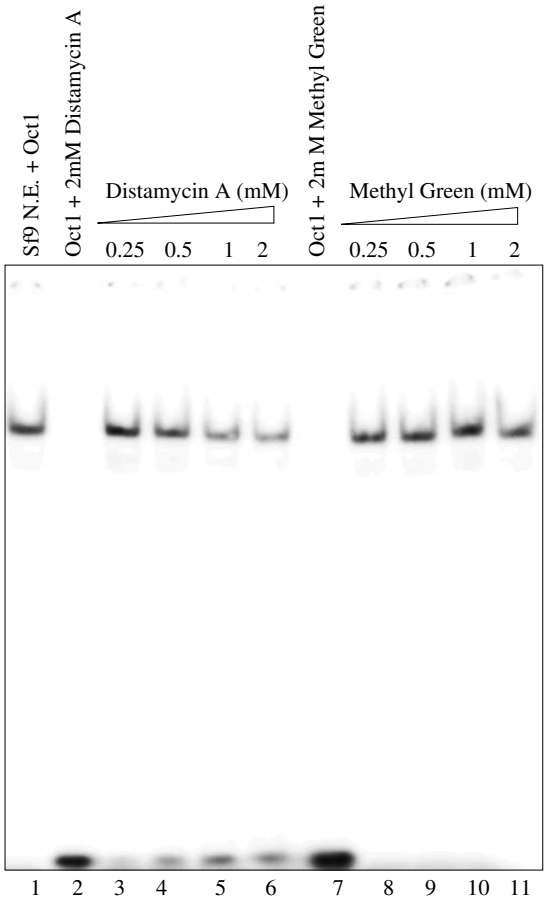
Figure 3.20: The presence of EDTA, a divalent chelator has no effect on the DNA binding activity of octamer binding protein. Binding reactions were carried out either in the absence (lane 2) or in the presence of 50mmM (lane 3), 75mM (lane 4) or 100mM EDTA (lane 5). The binding was not inhibited even in the presence of 100mM EDTA (lane 5).

Oct1-like factor binding is not inhibited by Methyl Green but marginally inhibited by Distamycin A

In order to determine the nature of interaction of the Oct1-like factor with DNA, EMSA reactions were carried out in the presence of both major groove binding dye, Methyl Green and minor groove binding drug, Distamycin A (Figure 3.21). The binding of Oct1-like factor was inhibited marginally in the presence of 1mM and 2mM Distamycin A (lane 5,6). But the presence of Methyl Green (lanes 8–11) even at a concentration of 2mM did not inhibit binding. These results suggest that the binding of the Oct1-like factor with DNA might involve minor groove.

UV-crosslinking and Southwestern analysis results indicate the size of the octamer binding protein to be $\approx$ 66 kDa

In order to determine the size of the protein factors that are binding to the octamer cognate sequence, UV-crosslinking and Southwestern analyses were carried out. *Sf*9 nuclear extract was incubated with either the radiolabelled consensus Oct1 oligonucleotide or the mut-Oct1 oligonucleotide. The reaction mixture was then cross-linked by exposing them to UV in a UV-crosslinker and then analyzed by SDS-PAGE.



b

Figure 3.21: The Oct1-like factor interaction with DNA involves minor groove. Labelled Oct oligonucleotide was incubated with 10μg of *Sf*9 N.E. (lane 1) or with 2mM of Distamycin A (lane 2) or with 2mM Methyl Green (lane 7). A specific band can be seen in lane 1 and the mobility of the oligonucleotide was not affected in the presence of either of Distamycin A or Methyl Green (lanes 2 and 7). The binding of Oct-like factor decreased in the presence of 0.5mM of Distamycin A (lane 4) and further decreased in the presence of 1mM or 2mM Distamycin A (lanes 5 and 6) but remained unchanged in the presence of 0.25–2mM Methyl Green (lanes 8–11).

A specific band corresponding to 66kDa was obtained in the reaction that contained the consensus oligonucleotide, but not when the nuclear extract was incubated with the mut-Oct1 oligonucleotide (Figure 3.22). This indicated that the size of the protein that is binding to the octamer sequence to be $\approx$ 66kDa.

Same results were obtained using Southwestern analysis which is a more accurate method to determine the molecular weight of the protein factors binding to a specific DNA element. About 20μg of nuclear extract were resolved on 10% SDS-PAGE, transferred to nitrocellulose membrane and then probed with either radio-labelled Oct1 or mutOct1 oligonucleotide. A specific band of $\approx$ 66kDa was obtained only with Oct1 oligonucleotide, confirming the size of the octamer sequence binding protein (Figure 3.23).

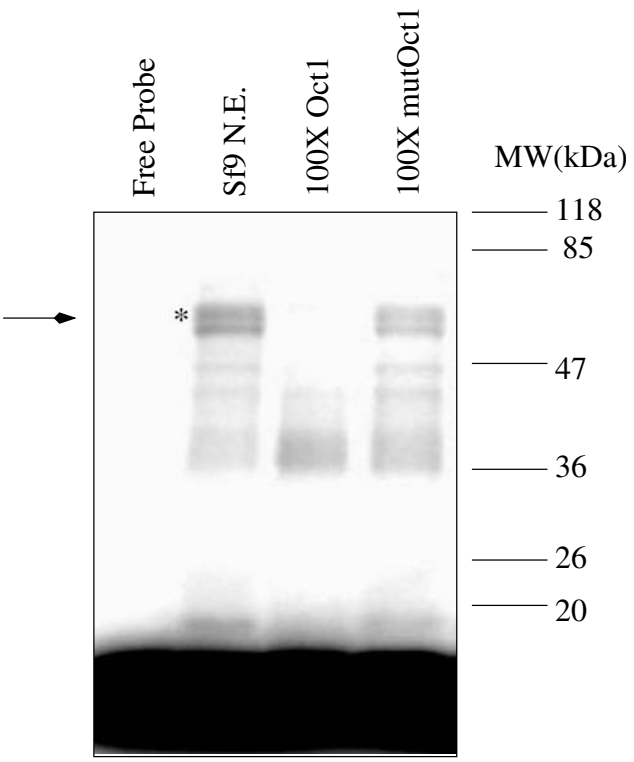


Figure 3.22: UV-crosslinking results show that the size of the octamer binding protein to be ≈ 66kDa. *Sf*9 N.E. was incubated with labelled Oct1 oligonucleotide either alone or in the presence of 100X excess of unlabelled Oct1 or unlabelled mutated Oct1 oligonucleotide. The reaction mixture was transferred to a sheet of parafilm, UV crosslinked for 2 mins in a UV crosslinker (12 KJoules/min) and then analyzed by SDS-PAGE analysis. A specific band corresponding to ≈ 66 kDa was obtained in the lane where no competitors were used. The migratory position of prestained protein molecular weight markers are shown on the right.

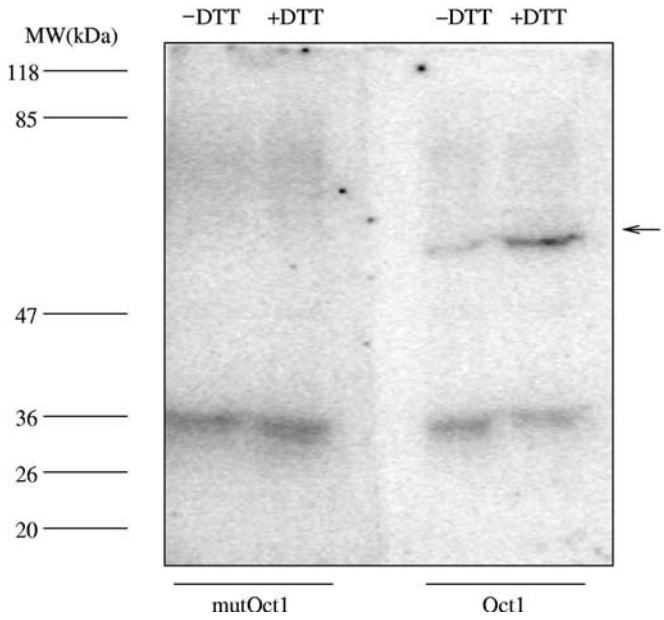


Figure 3.23: Southwestern analysis also shows the size of the octamer binding protein to be ≈ 66kDa. About 100μg of *Sf*9 N.E. was resolved in a 10% SDS gel, transferred to a nitrocellulose membrane. +DTT or -DTT indicates the presence or absence of DTT in the SDS loading dye respectively. The membrane was then probed with either radiolabelled consensus Oct1 (Oct1) or the mutated Oct1 (mutOct1) oligonucleotide, washed with excess of Southwestern buffer and then exposed by autoradiography. Both the mutOct1 and Oct1 probes revealed the presence of a non-specific DNA binding protein of ≈ 35 kDa. However, a specific band corresponding to ≈ 66 kDa could be seen in the membrane probed with the consensus Oct1 oligonucleotide but not in the mutOct1 membrane.

Chapter 4

Discussion & Summary

Chapter 4

Discussion and Summary

Regulation of gene expression is fundamental for all biological processes. The first mode of regulation in this important process, namely transcriptional regulation, is highly organized in *Ac*MNPV. The early gene transcription occurs immediately after infection and before DNA replication. Late and very late phases are characteristically different from the early phase of gene transcription. The late phase occurs concurrently with DNA replication whereas the very late phase does not commence if the DNA synthesis is incomplete. The late and very late genes are transcribed by a virally encoded RNA polymerase. The polyhedrin gene falls in the category of very late genes along with another important structural protein encoding gene, the *p10* gene. It is understandable that these structural genes, that form the polyhedra and the matrix respectively (for packaging of virus particles) are transcribed in the very late stages of infection i.e. after DNA replication. Although, the exact mechanism of how the polyhedrin gene promoter as well as other late and very late genes *as per se* are transcribed by a relatively simple, virally encoded, 4-subunit RNA polymerase is unclear, the role of other factors

that might function along with this virally encoded RNA polymerase has been the focus of this study.

The main question addressed in this study was, "What regulates the hyperactivation of transcription from the polyhedrin promoter?". For answering this, we have tried to elucidate the role of the 4kb upstream region in polyhedrin gene transcription. An Sp1 binding element had been identified (located ≈700 upstream to the polyhedrin promoter) that activates transcription from the polyhedrin promoter and characterized earlier from our lab (117). A comprehensive analysis of the 4kb upstream region had now been carried out, to identify additional regulatory elements that appear to play important roles in the activation of transcription from the polyhedrin promoter. Using a simple strategy, specific deletions in the 4kb upstream region were made to study the effect of the deleted region in reporter gene expression. Any deletion that results in consistent reduction in reporter gene expression would indicate that the deleted fragment may harbor *cis*-acting positive regulatory elements. Simultaneously, the deleted fragments were cloned upstream to the basal polyhedrin promoter. Again, if the sequence thus cloned harbors *cis*-acting postive regulatory elements, there will be consistent increase in reporter gene expression.

Deletion of the MluI-XhoI region resulted in 10 fold reduction in luciferase reporter gene expression when compared with the control plasmid, pKNluc that carries the entire 4kb upstream region (Figure 3.4). When this region was divided into three fragments and cloned upstream to the polyhedrin basal promoter in the vector pAJpolluc, only the construct, pMXBpolluc showed significant increase in the reporter gene expression (Figure 3.9). The presence of only the first 200bp of the MluI-XhoI region (in the plasmid construct, pMXApolluc) upstream to the

polyhedrin promoter did not result in a significant increase in expression when compared with the control plasmid, pAJpolluc (Figure 3.9).

The Oct1 binding site identified (Table 3.2) in the 'B' fragment of the MluI-XhoI region. The plasmid carrying the Oct1 containing region (pMXBpolluc) was able to enhance the expression from *Ac*MNPV polyhedrin promoter by 7–8 fold, as compared to pAJpolluc. However, the drastic reduction that is observed in the case of pMXCpolluc is unclear. Search for potential transcription factor in this MluI-XhoI region did not show any transcription factor binding site with high scores when submitted to TESS. Possibly some yet to be identified element with a negative regulation of polyhedrin promoter might be present in this region. However, the presence of other elements in the MluI-XhoI region might override the inhibition caused by elements present in this region.

Lo and colleagues reported an upstream sequence termed as *pu* that can enhance expression from a minimal CMV (CMVm) promoter from human cytomegalovirus as well as *Drosophila* hsp70 and the minimal *p35* promoter. The expression could be further enhanced by the presence of *hr*1 (82). The *pu* region comprises three ORFs namely, ORF4, ORF5 and *lef-2*. The deletion of *lef-2* sequence or the ORF4 sequence decreased the reporter gene expression, but the presence of *lef-2* gene alone could not activate the CMVm promoter. The truncation of ORF5 also did not decrease the reporter gene expression suggesting that this ORF might not be important for transcription from the CMVm promoter. However, any deletion in ORF4 resulted in total abolishment of activation capability of *pu* sequence (82). *pu* sequence could activate expression from all the three promoters tested. However, they did not study the role of *pu* sequence in context of the polyhedrin promoter which is its natural neighbor in the *Ac*MNPV genome. The MluI-MluI

region comprises of the entire *lef-2* ORF as well as partial sequences of ORFs, ORF5 and ORF603. Although, Lo et al observed a reduction in expression from CMVm promoter when the *lef-2* was deleted, the results presented here suggest that, atleast in the case of polyhedrin promoter, the deletion of *lef-2* does not cause any significant decrease in reporter gene expression (Figure 3.5).

Similarly, the deletion of the ORF PTP (ORF984), that comprises a major portion of the SacII-NdeI region, did not result in significant reduction in reporter gene expression (Figure 3.5). PTP (Protein Tyrosine Phosphatase) from *Bm*NPV had been demonstrated to help in the survival of the virus under natural conditions by increasing the ELA (enhanced locomotory action) in *Bombyx mori* larvae (65).

TESS searches yielded many potential transcription factor binding sites in both the SacII-XhoI and the MluI-XhoI region. However, two candidates namely, HSE and Oct-1 showed higher scores when compared to other transcription factor binding sites. The presence of either of these factors upstream to the polyhedrin promoter could enhance expression by 2-fold (ie without the aid of any other upstream sequence). In the case of HSE (heat shock element), the presence of either artificially created (ie combined HSE-like elements or combined HSE + other transcription factor binding sites) as well as the exact HSE-like element present in the SacII-XhoI region or the presence of classical HSE from *Drosophila* hsp70 promoter could independently enhance expression in a similar manner (Figure 3.11). In the case of Oct-1, the orientation of the binding element was found to be unimportant for the enhancement. Oct-1 binding site when present either in the "forward" or "reverse" orientation could enhance expression from the polyhedrin promoter with equal efficiency (Figure 3.13). EMSA with oligonucleotides carrying either the HSE or the Oct-1 sequences showed specific complexes with *Sf*9 nuclear ex-

tract, providing evidence that factors resembling HSTF (heat shock transcription factor, or HSF-heat shock factor) and Oct-1 are present in the nuclear extracts of *Sf*9 cells.

We have further carried out biochemical characterization of the binding of Oct1-like host factor(s) from *Sf*9 cells to the Oct-1 element present in the MluI-XhoI region. When the binding reactions of Oct-1 sequence from the MluI-XhoI region was carried out in the presence of either monovalent or divalent cations (NaCl, Figure 3.16, $CaCl_2$ Figure 3.17, $MgCl_2$ Figure 3.18 and $MnCl_2$ Figure 3.19, respectively) the binding of the Oct1-like factor was inhibited in the presence of 50mmM NaCl, 5mM $CaCl_2$, 2.5mM $MgCl_2$, 5mM $MnCl_2$. Therefore, the divalent cations were strong inhibitors of Oct1-like factor binding to DNA as compared to monovalent cations. When the binding reactions were carried out in the presence of EDTA, a divalent metal chelator, the binding was not inhibited even in the presence of 100mM EDTA (Figure 3.20). These results suggest that the binding of the Oct1-like factor to the octamer sequence does not require any divalent metal ions for binding. Earlier work from Westin et al suggested that the binding of the octamer binding factor (OTF/Oct-1) to the octamer sequence was not affected when the EMSA reactions were carried out with metal-depleted (EDTA treated) HeLa nuclear extract (145). Similar results were also obtained when the same reactions were carried out in the presence of 1,10-phenanthroline, an efficient metal chelator of Zn, Mn, Cu and Fe ions (145).

Oct1 EMSA reactions were carried out in the presence of Methyl Green (major groove binding dye) and distamycin A (minor groove binding drug). The results showed that the binding was inhibited marginally in the presence of 1mM or 2mM distamycin A, but not in the presence of Methyl Green (Figure 3.21). Crystal

structures of POU_s from Oct-1 protein suggests that it binds to the major groove of the DNA (25). The POU domain is a bipartite domain consisting of a POU-specific (POU_s) and POU-homeodomain (POU_{hd}) connected by a flexible linker. The left half of the consensus octamer sequence, ATGCAAAT, is recognised by the POU_s, whereas the right half by POU_{hd} (25). Comparison of the DNA binding properties of POU_s and POU_{hd} suggests that both POU_s and POU_{hd} independently regulate specificity at opposite ends of the DNA site (5'TATGCAAAT3') and both domains contact the central bases (5'TATGCAAAT3') where coordinate binding of the POU_s to the major groove overrides the intrinsic specificity of POU_{hd} in the minor groove (11). Moreover, in the model provided by Dekker et al, the N-terminal of the POU_{hd} is positioned at the minor groove (25). The POU_s domain is related to the DNA binding domains of λ and 434 repressors and 434 Cro proteins (25; 67).

Oct-1 is a ubiquitously expressed transcription factor that has been reported from nuclear extracts of many tissue types, while another factor, Oct-2 is lymphoid specific that binds to IgH promoter sequences. Octamer sequence binding proteins belong to POU (Pit-1 Oct-1 Unc-86– the members where the domain was first identified) family of proteins and contain a bipartite POU domain that has a POU specific domain and a POU homeodomain (25). Members of this family contain the highly conserved POU domain (54). Oct-1 belongs to this family of POU DNA-binding domain containing proteins (25). Other factors that belong to the POU family of proteins are: Oct-2 , Pit-1, Oct-6 etc. In *Drosophila* most of the POU domain proteins are involved in development control (22; 64; 89; 105; 138; 141; 154). Oct-1 factor had been implicated in the replication of adenovirus. The POU domain was found to interact with the terminal protein-DNA polymerase complex. The POU_{hd} was also found to interact but with this complex, but the

interaction was less than that in the case of the entire POU domain. The POU_s domain was not found to interact with the complex (23). Earlier, Hatfield and Hearing had shown that the presence of Oct-1 binding sites in the adenovirus 103bp inverted terminal repeat (ITR) were bound by cellular transcription factor Oct-1 and stimulates DNA replication *in vivo* (51). It was also shown that POU domain from different subclasses of transcription factors could stimulate adenovirus DNA replication *in vitro* (142). Taken together, these results suggest that Oct-1 or transcription factors belonging to POU domain containing proteins are essential for adenovirus replication.

Being an ubiquitously expressed protein, Oct-1 had been implicated in the regulation of transcription in a large number of genes. The implication of having a variable linker between the POU_s and POU_{hd} (24 residues in the case of Oct-1) allows for interactions of other factors that bind to DNA and Oct-1 itself (67).

Octamer sequences had been shown to be important for H2B gene expression in *Xenopus* oocytes along with CCAAT motifs during the development of frog embroys (55). VP16, a protein encoded by Herpes Simplex Virus interacts with Oct-1 along with a host factor HCT (host cell factor) from HeLa cells and this interaction is essential for the transcription of HSV *ie* gene (149). Using Oct-1 deficient mouse, Nogueiro et al showed that the HSV infection was arrested in Oct-1 deficient cells, further augmenting evidence that Oct-1 is required for the transcription of IE gene expression (107).

The promoters of RNA polymerase II and III human small nuclear RNA (snRNA)promoters have a common proximal sequence element (PSE). In addition to PSE, the promoters of all snRNA contain binding sites for octamer-binding transcription factors (Octs), and functional studies have indicated that the PSE

and octamer elements work cooperatively (103). Mittal et al had shown that POU domain of Oct-1, but not the related protein Pit-1 POU domain, can facilitate the binding of snRNAP$_c$ (a TBP protein-containing complex that binds to PSE) (99). The OCA-B (OBF-1, Bob1) coactivator protein functions in immunoglobulin transcription and B-cell differentiation by forming a complex with the Oct-1 or Oct-2 transcription factors and the octamer DNA sequence (17; 88; 133). Oct-1 binding sites were initially identified in the H2B promoters in chicken (47). They were found to be conserved in the 5' untranslated region of H2B genes *Xenopus laevis*, *Strongylocentrotus purpuratus* and *Psammechinus miliaris* and the position was between the CAAT box and the TATA box, and they were initially termed as "H2B specific Box" (47).

Recently, Mesplède and colleagues had shown that Oct-1 binds to octamer-like sequences present in the upstream region of Interferon-α gene. They showed that Oct-1 could bind to IFA11 (one of the IFN-α subtypes) promoter both *in vivo* and *in vitro* and repressed IFN-A overexpression. This might have a regulatory role in preventing the detrimental effect of IFN-A overexpression (93).

The important role of Oct-1 in replication and transcription in general, in various species has been well established. Results presented in this study show that Oct1-like factor from *Sf*9 cells may be important for regulating transcription from the *Ac*MNPV polyhedrin gene promoter.

In this work, HSE-like elements were identified in the SacII-XhoI region of the 4kb upstream of polyhedrin promoter. The presence of these elements either alone or with other transcription factor binding sites can enhance the expression from polyhedrin promoter (Figure 3.11). Also, factors resembling heat shock factors from *Sf*9 cell nuclear extract have been shown to bind to HSE (Figure 3.10).

The heat shock proteins were first detected as a new puffing pattern upon heat shock in *Drosophila buschii* polytene chromosomes in 1961 (30). The heat shock stress-induced transcription requires activation of heat shock factor (HSF) that binds to the heat shock promoter element (HSE) (131), characterized as multiple adjacent and inverse repeats of the motif 5'nGAAn3' (31). HSF also binds to long arrays of this motif and the footprint size increases with the addition of each 5bp unit of these arrays. Trimers of HSF bind to DNA and the number of HSF subunits in direct contact with DNA is determined by the number of correctly positioned 5bp recognition units (114). HSE elements had been reported in human hsp70 where they are bound by HSF and along with CCAAT box-binding factor (CTF) and regulate the activity of hsp70 gene expression (100). Of the total four HSTFs in humans, HSTFs1–3 play a positive role and HSTF4 might play a negative role in the regulation of heat shock induced gene expression (104).

The ORFs that lie in the 4kb upstream of the polyhedrin promoter encode the following proteins namely, ORF504 (Protein Tyrosine Phosphatase (PTP)), *bro*-baculovirus repeated ORF (ORF984), *ctx*-conototoxin-like peptide, ORF4 (ORF453), ORF5 (ORF327), LEF-2 (ORF630), ORF603 followed by the polyhedrin gene.

Of these ORFs, SacII-XhoI region contains the region that encodes most of the ORF984 protein (233 AA of the 328 AA) which encodes a protein belonging to the BRO family (Figure 3.1). Although not much is known about this protein from *Ac*MNPV, homologs from other members have been well characterized. *Bombyx mori* Multiplenucleopolyhedrosis Virus (*BmNPV*) has 5 members related to *Ac*MNPV *bro* namely ORFs 22, 80, 81, 131, and 132 (36; 66). The transcripts of these ORFs *bro a–e* respectively could be detected as early as 4hpi and decreased after 24hpi. Moreover, the expression of these transcripts required viral factor(s).

The *bro-c*, *bro-d*, and *bro-e* genes have a CAGT early-gene start motif, while *bro-a* and *bro-b* have a TAGT sequence instead of CAGT unlike ORF2 (*Ac*MNPV *bro* gene) which is classified as a late gene since it has late-gene promoter motif (ATAAG), but it is not shown to be expressed in the late stages in *Ac*MNPV (66).

The exact function of these *bro* genes is unknown. Zemskov and colleagues had reported that BRO-A, BR0-C and BRO-D could be isolated along with core histone proteins from *Bm*NPV infected *Bm*N cells using a histone extraction protocol and proposed that BRO-C and BRO-A could influence host DNA replication and transcription (155). Moreover, in *Bm*NPV, it was shown that *bro-d* is essential for viral multiplication in *Bm*N cells (66). Since the transcripts of *bro* had not been shown to occur in *Ac*MNPV infected *Sf*9 cells in late stages of infection, it appears very unlikely that the gene product of *bro* (233 AA of the 328 AA present in the SacII-XhoI region) can influence the transcription from the polyhedrin promoter. Hence, it can be safely concluded that the increase in the expression that is observed due to the presence of the SacII-XhoI region upstream to the polyhedrin promoter in the reporter transfection assays is not due to the partial protein coding region of *bro* (ORF984) but may be due to the presence of *cis*-acting elements present in the SacII-XhoI region.

The MluI-XhoI region is 569bp in length and consists of two ORFs namely ORF4 (ORF453) and ORF5 (ORF327) located in opposite orientations (Figure 3.1). ORF4 and ORF5 encode polypeptides of 151 and 109 amino acids respectively and of these, the MluI-XhoI region carries the majority of the N-terminus of ORF4 (140 of the 151 AA residues) and only about 50% of the N-terminus of ORF5 (40 of the 109 AA residues).

Acharya and Gopinathan had reported that a 293bp fragment that encompasses

these ORFs, when present in *cis* but not in *trans*, could enhance expression by 10-fold when compared with a minimal polyhedrin promoter in *Bm*NPV. Two AP-1 sites that had been identified in this region were found to be not important for the enhancer function. The same fragment, when present upstream to the *Bm*NPV *p10* promoter could enhance the luciferase reporter gene expression by two fold (1). Using RNase protection assays, they also showed that both these ORFs were expressed during the course of viral infection as delayed-late transcripts (between 36–72hpi). Moreover, extracts from uninfected *Bm*N cells did not yield a complex with a segment of the enhancer element in EMSA and hence they suggest that early or late viral activators, but not the insect cell host factors, might act through this enhancer element (1). However, they have not mentioned which segment of this enhancer-like element was used to test for the binding of host factor (1). From these results, it can be concluded that the ORFs present in the MluI-XhoI region may not play a role in the activation of transcription from the polyhedrin promoter.

From these results, it can be concluded that both the *cis*-acting elements and *trans*-acting host factors are important for driving the polyhedrin promoter. Moreover, the *Ac*MNPV is capable of employing the insect cell host factors in driving the expression of its own genes.

Summary

1. Based on deletion analysis, two regions namely SacII-XhoI region and MluI-XhoI region were identified to be important for polyhedrin gene transcription.

2. In addition, these two regions could enhance transcription when seperately cloned upstream of the *polh-luc* cassette.

3. A recombinant virus (vDSXluc) in which the entire SacII-XhoI region had been deleted showed significant reduction in luciferase reporter gene expression when compared with another virus (vMAluc) in which the entire 4kb upstream sequence is present, confirming that this region is indeed vital for transcription from the polyhedrin promoter.

4. Transcription element search system (TESS) revealed the presence of transcription factor binding sites with high sequence similarlity to heat shock element (HSE) and Oct1 in the SacII-XhoI region and MluI-XhoI region of the polyhedrin promoter, respectively.

5. Presence of either HSE or Oct1 enhances reporter gene expression when present upstream to the polyhedrin promoter. Oct1 element could function in either orientation.

6. Factor(s) resembling HSTF or Oct1 from *Sf*9 bind to HSE and Oct1 binding sites respectively, with high specificity, as evident from electrophoretic mobility shift assays.

7. Oct1-like factor from *Sf*9 cells was further characterized and it was determined that it interacts with DNA through the minor groove.

8. Single stranded Oct1 oligonucleotide could compete for complex formation with Oct1- like factor. Monovalent or divalent cations inhibited the binding of Oct1-like factor to DNA and the presence of EDTA did not hinder binding suggesting that monovalent or divalent cations are not important for binding.

9. The size of Oct1-like factor was determined to be $\approx$ 66kDa by UV-crosslinking and Southwestern analysis.

References

Bibliography

[1] A. Acharya and K. P. Gopinathan. Identification of an enhancer-like element in the polyhedrin gene upstream region of *Bombyx mori* nucleopolyhedrovirus. *J Gen Virol*, **82**(Pt 11):2811–9, Nov 2001. 4

[2] M. D. Ayres, S. C. Howard, J. Kuzio, M. Lopez-Ferber, and R. D. Possee. The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. *Virology*, **202**(2):586–605, Aug 1 1994. 1.3, 1.4, 1.2

[3] J. W. Barrett, H. A. Lauzon, P. S. Mercuri, P. J. Krell, S. S. Sohi, and B. M. Arif. The putative LEF-1 proteins from two distinct *Choristoneura fumiferana* multiple nucleopolyhedroviruses share domain homology to eukaryotic primases. *Virus Genes*, **13**(3):229–37, 1996. 1.5

[4] Murali D Bashyam, A Ramachandran, S Ghosh, P Viswanathan, and S E Hasnain. Multiple regulatory circuits for transcription from the polyhedrin gene promoter of the *Autogapha californica* Multiple Nucleocapsid Polyhedrosis Virus. *Proc. Indian. Natn. Sci. Acad. (PINSA)*, **B66**:83–96, 2000. 1.2.1

[5] H. Beniya, C. J. Funk, G. F. Rohrmann, and R. F. Weaver. Purification of a virus-induced RNA polymerase from *Autographa californica* nuclear polyhedrosis virus-infected spodoptera frugiperda cells that accurately initiates late and very late transcription in vitro. *Virology*, **216**(1):12–9, Feb 1 1996. 1.6.2

[6] Imre Berger, Daniel J. Fitzgerald, and Timothy J. Richmond. Baculovirus expression system for heterologous multiprotein complexes. *Nat Biotechnol*, **22**(12):1583–7, Dec 2004. 1.2.1

[7] T. K. Biswas, J. C. Edwards, M. Rabinowitz, and G. S. Getz. Characterization of a yeast mitochondrial promoter by deletion mutagenesis. *Proc Natl Acad Sci U S A*, **82**(7):1954–8, Apr 1985. 1.6.2

[8] G. W. Blissard and G. F. Rohrmann. Location, sequence, transcriptional mapping, and temporal expression of the gp64 envelope glycoprotein gene of the *Orgyia pseudotsugata* multicapsid nuclear polyhedrosis virus. *Virology*, **170**(2):537–55, Jun 1989. 1.6.1

[9] G. W. Blissard and G. F. Rohrmann. Baculovirus diversity and molecular biology. *Annu Rev Entomol*, **35**:127–55, 1990. 1.1, 1.6, 1.6.1, 1.6.2

[10] B. C. Bonning and B. D. Hammock. Development of recombinant baculoviruses for insect control. *Annu Rev Entomol*, **41**:191–210, 1996. 1.2.4

[11] M. C. Botfield, A. Jancso, and M. A. Weiss. Biochemical characterization of the Oct-2 POU domain with implications for bipartite DNA recognition. *Biochemistry*, **31**(25):5841–8, Jun 30 1992. 4

[12] Y. Boublik, Di P. Bonito, and I. M. Jones. Eukaryotic virus display: engineering the major surface glycoprotein of the *Autographa californica* nuclear polyhedrosis virus (*Ac*NPV) for the presentation of foreign proteins on the virus surface. *Biotechnology (N Y)*, **13**(10):1079–84, Oct 1995. 1.2.3

[13] F. M. Boyce and N. L. Bucher. Baculovirus-mediated gene transfer into mammalian cells. *Proc Natl Acad Sci U S A*, **93**(6):2348–52, Mar 19 1996. 1.2.2

[14] S. Burma, B. Mukherjee, A. Jain, S. Habib, and S. E. Hasnain. An unusual 30-kda protein binding to the polyhedrin gene promoter of *Autographa californica* nuclear polyhedrosis virus. *J Biol Chem*, **269**(4):2750–7, Jan 28 1994. 1.7, 3.1

[15] D. D. Carson, L. A. Guarino, and M. D. Summers. Functional mapping of an *Ac*NPV immediately early gene which augments expression of the IE-1 trans-activated 39K gene. *Virology*, **162**(2):444–51, Feb 1988. 1.6.1

[16] E. B. Carstens, A. L. Lu, and H. L. Chan. Sequence, transcriptional mapping, and overexpression of p47, a baculovirus gene regulating late gene expression. *J Virol*, **67**(5):2513–20, May 1993. 1.6.2

[17] D. Chasman, K. Cepek, P. A. Sharp, and C. O. Pabo. Crystal structure of an OCA-B peptide bound to an Oct-1 POU domain/octamer DNA complex: specific recognition of a protein-DNA interface. *Genes Dev*, **13**(20):2650–7, Oct 15 1999. 4

[18] U. Chatterji, R. Ahmad, B. Venkaiah, and S. E. Hasnain. A recombination-efficient baculovirus vector for simultaneous expression of multiple genes. *Gene*, **171**(2):209–13, Jun 1 1996. 1.2.1

[19] X. Chen, Z. Hu, J. A. Jehle, Y. Zhang, and J. M. Vlak. Analysis of the ecdysteroid UDP-glucosyltransferase gene of *Heliothis armigera* single-nucleocapsid baculovirus. *Virus Genes*, **15**(3):219–25, 1997. 1.5

[20] G. E. Chisholm and D. J. Henner. Multiple early transcripts and splicing of the *Autographa californica* nuclear polyhedrosis virus IE-1 gene. *J Virol*, **62**(9):3193–200, Sep 1988. 1.6.1

[21] J. Choi and L. A. Guarino. The baculovirus transactivator IE1 binds to viral enhancer elements in the absence of insect cell factors. *J Virol*, **69**(7):4548–51, Jul 1995. 1.6.1

[22] P. J. Clyne, S. J. Certel, de M. Bruyne, L. Zaslavsky, W. A. Johnson, and J. R. Carlson. The odor specificities of a subset of olfactory receptor neurons are governed by acj6, a POU-domain transcription factor. *Neuron*, **22**(2):339–47, Feb 1999. 4

[23] F. E. Coenjaerts, van J. A. Oosterhout, and van der P. C. Vliet. The Oct-1 POU domain stimulates adenovirus DNA replication by a direct interaction between the viral precursor terminal protein-DNA polymerase complex and the POU homeodomain. *EMBO J*, **13**(22):5401–9, Nov 15 1994. 4

[24] C. Davrinche, C. Pasquier, M. Cerutti, L. Serradell, D. Clement, G. Devauchelle, S. Michelson, and J. L. Davignon.

Expression of human cytomegalovirus immediate early protein IE1 in insect cells: splicing of RNA and recognition by CD4+ T-cell clones. *Biochem Biophys Res Commun*, **195**(1):469–77, Aug 31 1993. 1.2.1

[25] N. Dekker, M. Cox, R. Boelens, C. P. Verrijzer, van der P. C. Vliet, and R. Kaptein. Solution structure of the POU-specific DNA-binding domain of Oct-1. *Nature*, **362**(6423):852–5, Apr 29 1993. 4

[26] dela W. P. Cruz, P. D. Friesen, and A. J. Fisher. Crystal structure of baculovirus P35 reveals a novel conformational change in the reactive site loop after caspase cleavage. *J Biol Chem*, **276**(35):32933–9, Aug 31 2001. 1.6.1

[27] W. Ernst, R. Grabherr, D. Wegner, N. Borth, A. Grassauer, and H. Katinger. Baculovirus surface display: construction and screening of a eukaryotic epitope library. *Nucleic Acids Res*, **26**(7):1718–23, Apr 1 1998. 1.2.3

[28] E. Etkin, L. Carp-Weiss, and B. Z. Levi. *Spodoptera frugiperda Sf*-9 cells nuclear factor binds to a specific sequenc e on the baculovirus polyhedrin promoter. *Virus Res*, **31**(3):343–56, Mar 1994. 1.7

[29] J. T. Evans, D. J. Leisy, and G. F. Rohrmann. Characterization of the interaction between the baculovirus replication factors LEF-1 and LEF-2. *J Virol*, **71**(4):3114–9, Apr 1997. 1.5

[30] Ritossa F. A new puffing pattern induced by temperature shock and dnp in *Drosophila*. *Experientia*, **18**:571–573, 1962. 4

[31] M. Fernandes, H. Xiao, and J. T. Lis. Fine structure analyses of the drosophila and saccharomyces heat shock factor–heat shock element interactions. *Nucleic Acids Res*, **22**(2):167–73, Jan 25 1994. 4

[32] Alejandra Garcia-Maruniak, James E. Maruniak, Paolo A. M Zanotto, Aissa E. Doumbouya, Jaw-Ching Liu, Thomas M. Merritt, and Jennifer S. Lanoie. Sequence analysis of the genome of the *Neodiprion sertifer* nucleopolyhedrov irus. *J Virol*, **78**(13):7036–51, Jul 2004. 1.1.1

[33] S. Ghosh, A. Jain, B. Mukherjee, S. Habib, and S. E. Hasnain. The host factor polyhedrin promoter binding protein (PPBP) is involved in transcription from the baculovirus polyhedrin gene promoter. *J Virol*, **72**(9):7484–93, Sep 1998. 1.7, 2.2.2

[34] Sudip Ghosh, Md Khalid Parvez, Kakoli Banerjee, Shiv K. Sarin, and Seyed E. Hasnain. Baculovirus as mammalian cell expression vector for gene therapy: an emerging strategy. *Mol Ther*, **6**(1):5–11, Jul 2002. 1.2.2

[35] B. Glocker, R R Jr Hoopes, L. Hodges, and G. F. Rohrmann. In vitro transcription from baculovirus late gene promoters: accurate mRNA initiation by nuclear extracts prepared from infected *Spodoptera frugiperda* cells. *J Virol*, **67**(7):3771–6, Jul 1993. 1.6, 1.6.2

[36] S. Gomi, K. Majima, and S. Maeda. Sequence analysis of the genome of *Bombyx mori* nucleopolyhedrovirus. *J Gen Virol*, **80 (Pt 5)**:1323–37, May 1999. 4

[37] M. Gong and L. A. Guarino. Expression of the 39k promoter of *Autographa californica* nuclear polyhedrosis virus is increased by the apoptotic suppressor P35. *Virology*, **204**(1):38–44, Oct 1994. 1.6.1

[38] C. H. Gross and S. Shuman. RNA 5'-triphosphatase, nucleoside triphosphatase, and guanylyltransferase activities of baculovirus LEF-4 protein. *J Virol*, **72**(12):10020–8, Dec 1998. 1.6.2

[39] L. A. Guarino and W. Dong. Expression of an enhancer-binding protein in insect cells transfected with the *Autographa californica* nuclear polyhedrosis virus IE1 gene. *J Virol*, **65**(7):3676–80, Jul 1991. 1.6.1

[40] L. A. Guarino, M. A. Gonzalez, and M. D. Summers. Complete sequence and enhancer function of the homologous dna regions of *Autographa californica* nuclear polyhedrosis virus. *J Virol*, **60**, 1986. 1.6.1

[41] L. A. Guarino and M. D. Summers. Interspersed homologous DNA of *Autographa californica* nuclear polyhedrosis virus enhances delayed-early gene expression. *J Virol*, **60**:215–23, 1986. 1.4, 1.6.1

[42] L. A. Guarino, B. Xu, J. Jin, and W. Dong. A virus-encoded RNA polymerase purified from baculovirus-infected cells. *J Virol*, **72**(10):7985–91, Oct 1998. 1.6, 1.6.2, 1.7

[43] Linda A. Guarino, Toni-Ann Mistretta, and Wen Dong. Baculovirus LEF-12 is not required for viral replication. *J Virol*, **76**(23):12032–43, Dec 2002. 1.6.2

[44] S. Habib and S. E. Hasnain. A 38-kda host factor interacts with functionally important motifs within the *Autographa californica* multinucleocapsid nuclear polyhedrosis virus homologous region (*hr1*) DNA sequence. *J Biol Chem*, **271**(45):28250–8, Nov 8 1996. 1.6.1, 1.7, 3.1, 3.1

[45] S. Habib and S. E. Hasnain. Differential activity of two non-hr origins during replication of the baculovirus *Autographa californica* nuclear polyhedrosis virus genome. *J Virol*, **74**(11):5182–9, Jun 2000. 1.5

[46] S. Habib, S. Pandey, U. Chatterji, S. Burma, R. Ahmad, A. Jain, and S. E. Hasnain. Bifunctionality of the AcMNPV homologous region sequence (*hr1*): enhancer and ori functions have different sequence requirements. *DNA Cell Biol*, **15**(9):737–47, Sep 1996. 1.5, 3.1

[47] R. P. Harvey, A. J. Robins, and J. R. Wells. Independently evolving chicken histone H2b genes: identification of a ubiquitous H2b-specific 5' element. *Nucleic Acids Res*, **10**(23):7851–63, Dec 11 1982. 4

[48] S. H. Harwood, L. Li, P. S. Ho, A. K. Preston, and G. F. Rohrmann. AcMNPV late expression factor-5 interacts with itself and contains a zinc ribbon domain that is required for maximal late transcription activity and is homologous to elongation factor TFIIS. *Virology*, **250**(1):118–34, Oct 10 1998. 1.6.2

[49] S. E. Hasnain, A. Jain, S. Habib, S. Ghosh, U. Chatterji, A. Ramachandran, P. Das, B. Venkaiah, S. Pandey, B. Liang, A. Ranjan, K. Natarajan, and C. A. Azim. Involvement of host factors in tran-

scription from baculovirus very late promoters – a review. *Gene*, **190**(1):113–8, Apr 29 1997. 1.7

[50] S. E. Hasnain, B. Nakhai, N. Z. Ehtesham, P. Sridhar, A. Ranjan, G. P. Talwar, and P. K. Jha. Beta-subunit of human chorionic gonadotropin hormone and firefly luciferase simultaneously synthesized in insect cells using a recombinant baculovirus are differentially expressed and transported. *DNA Cell Biol*, **13**(3):275–82, Mar 1994. 1.2.1, 1.2.1

[51] L. Hatfield and P. Hearing. The NFIII/OCT-1 binding site stimulates adenovirus DNA replication in vivo and is functionally redundant with adjacent sequences. *J Virol*, **67**(7):3931–9, Jul 1993. 4

[52] T. Hayakawa, R. Ko, K. Okano, S. I. Seong, C. Goto, and S. Maeda. Sequence analysis of the *Xestia c-nigrum* granulovirus genome. *Virology*, **262**(2):277–97, Sep 30 1999. 1.1.1

[53] Kathleen L. Hefferon and Lois K. Miller. Reconstructing the replication complex of *Ac*MNPV. *Eur J Biochem*, **269**(24):6233–40, Dec 2002. 1.5, 1.4

[54] W. Herr, R. A. Sturm, R. G. Clerc, L. M. Corcoran, D. Baltimore, P. A. Sharp, H. A. Ingraham, M. G. Rosenfeld, M. Finney, and G. Ruvkun. The POU domain: a large conserved region in the mammalian *pit-1*, *oct-1*, *oct-2*, and *Caenorhabditis elegans unc-86* gene products. *Genes Dev*, **2**(12A):1513–6, Dec 1988. 4

[55] C. Hinkley and M. Perry. Histone H2b gene transcription during xenopus early development requires functional cooperation between proteins

bound to the CCAAT and octamer motifs. *Mol Cell Biol*, **12**(10):4400–11, Oct 1992. 4

[56] C. Hofmann, V. Sandig, G. Jennings, M. Rudolph, P. Schlag, and M. Strauss. Efficient gene transfer into human hepatocytes by baculovirus vectors. *Proc Natl Acad Sci U S A*, **92**(22):10099–103, Oct 24 1995. 1.2.2

[57] R R Jr Hoopes and G. F. Rohrmann. In vitro transcription of baculovirus immediate early genes: accurate mRNA initiation by nuclear extracts from both insect and human cells. *Proc Natl Acad Sci U S A*, **88**(10):4513–7, May 15 1991. 1.6.1

[58] N. E. Huh and R. F. Weaver. Identifying the RNA polymerases that synthesize specific transcripts of the *Autographa californica* nuclear polyhedrosis virus. *J Gen Virol*, **71 (Pt 1)**:195–201, Jan 1990. 1.6.2

[59] Andreas Huser and Christian Hofmann. Baculovirus vectors: novel mammalian cell gene-delivery vehicles and their applications. *Am J Pharmacogenomics*, **3**(1):53–63, 2003. 1.2.2

[60] A. Jain and S. E. Hasnain. A 30-kda host protein binds to two very-late baculovirus promoters. *Eur J Biochem*, **239**(2):384–90, Jul 15 1996. 1.7, 2.2.2

[61] J. A. Jehle, G. W. Blissard, B. C. Bonning, J. S. Cory, E. A. Herniou, G. F. Rohrmann, D. A. Theilmann, S. M. Thiem, and J. M. Vlak. On the classification and nomenclature of baculoviruses: A proposal for revision. *Arch Virol*, May 2 2006. 1.1, 1.2, 1.1

[62] P. K. Jha, R. Pal, B. Nakhai, P. Sridhar, and S. E. Hasnain. Simultaneous synthesis of enzymatically active luciferase and biologically active beta subunit of human chorionic gonadotropin in caterpillars infected with a recombinant baculovirus. *FEBS Lett*, **310**(2):148–52, Sep 28 1992. 1.2.1

[63] J. Jin, W. Dong, and L. A. Guarino. The LEF-4 subunit of baculovirus RNA polymerase has RNA 5'-triphosphatase and ATPase activities. *J Virol*, **72**(12):10011–9, Dec 1998. 1.6.2

[64] R. Kambadur, K. Koizumi, C. Stivers, J. Nagle, S. J. Poole, and W. F. Odenwald. Regulation of POU genes by *castor* and *hunchback* establishes layered compartments in the drosophila CNS. *Genes Dev*, **12**(2):246–60, Jan 15 1998. 4

[65] Shizuo G. Kamita, Koukichi Nagasaka, Josie W. Chua, Toru Shimada, Kazuei Mita, Masahiko Kobayashi, Susumu Maeda, and Bruce D. Hammock. A baculovirus-encoded protein tyrosine phosphatase gene induces enhanced locomotory activity in a lepidopteran host. *Proc Natl Acad Sci U S A*, **102**(7):2584–9, Feb 15 2005. 4

[66] W. Kang, M. Suzuki, E. Zemskov, K. Okano, and S. Maeda. Characterization of baculovirus repeated open reading frames (*bro*) in *Bombyx mori* nucleopolyhedrovirus. *J Virol*, **73**(12):10339–45, Dec 1999. 4

[67] J. D. Klemm, M. A. Rould, R. Aurora, W. Herr, and C. O. Pabo. Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA

recognition with tethered DNA-binding modules. *Cell*, **77**(1):21–32, Apr 8 1994. 4

[68] P. H. Kogan and G. W. Blissard. A baculovirus *gp64* early promoter is activated by host transcription factor binding to CACGTG and GATA elements. *J Virol*, **68**(2):813–22, Feb 1994. 1.6.1

[69] K. Kojima, T. Hayakawa, S. Asano, and H. Bando. Tandem repetition of baculovirus *ie1* promoter results in upregulation of transcription. *Arch Virol*, **146**(7):1407–14, Jul 2001. 1.6.1

[70] M. Kool, C. H. Ahrens, R. W. Goldbach, G. F. Rohrmann, and J. M. Vlak. Identification of genes involved in DNA replication of the *Autographa californica* baculovirus. *Proc Natl Acad Sci U S A*, **91**(23):11212–6, Nov 8 1994. 1.6.1

[71] M. Kool, C. H. Ahrens, J. M. Vlak, and G. F. Rohrmann. Replication of baculovirus DNA. *J Gen Virol*, **76 (Pt 9)**:2103–18, Sep 1995. 1.6.1

[72] M. Kool, R. W. Goldbach, and J. M. Vlak. A putative non-hr origin of DNA replication in the hindIII-K fragment of *Autographa californica* multiple nucleocapsid nuclear polyhedrosis virus. *J Gen Virol*, **75 (Pt 12)**:3345–52, Dec 1994. 1.4, 1.5, 1.6.1

[73] G. R. Kovacs, J. Choi, L. A. Guarino, and M. D. Summers. Functional dissection of the *Autographa californica* nuclear polyhedrosis virus immediate-early 1 transcriptional regulatory protein. *J Virol*, **66**(12):7429–37, Dec 1992. 1.6.1

[74] A. Kremer and D. Knebel-Morsdorf. The early baculovirus *he65* promoter: On the mechanism of transcriptional activation by IE1. *Virology*, **249**(2):336–51, Sep 30 1998. 1.6.1

[75] S. Laufs, A. Lu, K. Arrell, and E. B. Carstens. *Autographa californica* nuclear polyhedrosis virus *p143* gene product i s a DNA-binding protein. *Virology*, **228**(1):98–106, Feb 3 1997. 1.5

[76] Hilary M. A Lauzon, Christopher J. Lucarotti, Peter J. Krell, Qili Feng, Arthur Retnakaran, and Basil M. Arif. Sequence and organization of the *Neodiprion lecontei* nucleopolyhedrovirus genome. *J Virol*, **78**(13):7023–35, Jul 2004. 1.1.1

[77] H. Lee and P. J. Krell. Reiterated DNA fragments in defective genomes of *Autographa californica* nuclear polyhedrosis virus are competent for AcMNPV-dependent DNA replication. *Virology*, **202**(1):418–29, Jul 1994. 1.5

[78] D. J. Leisy, C. Rasmussen, H. T. Kim, and G. F. Rohrmann. The *Autographa californica* nuclear polyhedrosis virus homologous region 1 a: identical sequences are essential for DNA replication activity and transcriptional enhancer function. *Virology*, **208**(2):742–52, Apr 20 1995. 1.6.1

[79] D. J. Leisy, C. Rasmussen, E. O. Owusu, and G. F. Rohrmann. A mechanism for negative gene regulation in *Autographa californica* multinucleocapsid nuclear polyhedrosis virus. *J Virol*, **71**(7):5088–94, Jul 1997. 1.6.1

[80] D. J. Leisy and G. F. Rohrmann. Characterization of the replication of plasmids containing hr sequences in baculovirus-infected spodoptera frugiperda cells. *Virology*, **196**(2):722–30, Oct 1993. 1.5

[81] G. Liu and E. B. Carstens. Site-directed mutagenesis of the AcMNPV *p143* gene: effects on baculovirus DNA replication. *Virology*, **253**(1):125–36, Jan 5 1999. 1.5

[82] Huei-Ru Lo, Cheng-Chung Chou, Tzong-Yuan Wu, Joyce Pui-Yee Yuen, and Yu-Chan Chao. Novel baculovirus DNA elements strongly stimulate activities of exogenous and endogenous promoters. *J Biol Chem*, **277**(7):5256–64, Feb 15 2002. 4

[83] A. Lu and E. B. Carstens. Nucleotide sequence of a gene essential for viral DNA replication in the baculovirus *Autographa californica* nuclear polyhedrosis virus. *Virology*, **181**(1):336–47, Mar 1991. 1.5

[84] A. Lu and E. B. Carstens. Immediate-early baculovirus genes transactivate the *p143* gene promoter of *Autographa californica* nuclear polyhedrosis virus. *Virology*, **195**(2):710–8, Aug 1993. 1.6.1

[85] A. Lu and L. K. Miller. Identification of three late expression factor genes within the 33.8- to 43.4-map-unit region of *Autographa californica* nuclear polyhedrosis virus. *J Virol*, **68**(10):6710–8, Oct 1994. 1.6.2

[86] A. Lu and L. K. Miller. The roles of eighteen baculovirus late expression factor genes in transcription and DNA replication. *J Virol*, **69**(2):975–82, Feb 1995. 1.5

[87] Y. Lu, L. M. Tapay, P. C. Loh, J. A. Brock, and R. Gose. Development of a quantal assay in primary shrimp cell culture for yellow head baculovirus (YBV) of *Penaeid* shrimp. *J Virol Methods*, **52**(1-2):231–6, Mar 1995. 1.6.2

[88] Y. Luo, H. Fujii, T. Gerster, and R. G. Roeder. A novel B cell-derived coactivator potentiates the activation of immunoglobulin promoters by octamer-binding transcription factors. *Cell*, **71**(2):231–41, Oct 16 1992. 4

[89] L. Ma, N. Tamarina, Y. Wang, A. Kuznetsov, N. Patel, C. Kending, B. J. Hering, and L. H. Philipson. Baculovirus-mediated gene transfer into pancreatic islet cells. *Diabetes*, **49**(12):1986–91, Dec 2000. 4

[90] S. Maeda. Expression of foreign genes in insects using baculovirus vectors. *Annu Rev Entomol*, **34**:351–72, 1989. 1.2.1

[91] B. S. Masters, L. L. Stohl, and D. A. Clayton. Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7. *Cell*, **51**(1):89–99, Oct 9 1987. 1.6.2

[92] J. R. McLachlin and L. K. Miller. Identification and characterization of *vlf-1*, a baculovirus gene involved in very late gene expression. *J Virol*, **68**(12):7746–56, Dec 1994. 1.6.2

[93] Thibault Mesplède, Marie-Laure Island, Nicolas Christeff, Fahrettin Petek, Janine Doly, and Sebastien Navarro. The POU transcription factor Oct-1 represses virus-induced interferon A gene expression. *Mol Cell Biol*, **25**(19):8717–31, Oct 2005. 4

[94] V. S. MIkhailov. [replication of the baculovirus genome]. *Mol Biol (Mosk)*, **37**(2):288–99, Mar-Apr 2003. 1.5

[95] Victor S. Mikhailov and George F. Rohrmann. Baculovirus replication factor LEF-1 is a DNA primase. *J Virol*, **76**(5):2287–97, Mar 2002. 1.5

[96] Victor S. Mikhailov and George F. Rohrmann. Binding of the baculovirus very late expression factor 1 (VLF-1) to different DNA structures. *BMC Mol Biol*, **3**(1):14, Sep 26 2002. 1.6.2

[97] L. K. Miller, J. E. Jewell, and D. Browne. Baculovirus induction of a DNA polymerase. *J Virol*, **40**(1):305–8, Oct 1981. 1.6.2

[98] Toni-Ann Mistretta and Linda A. Guarino. Transcriptional activity of baculovirus very late factor 1. *J Virol*, **79**(3):1958–60, Feb 2005. 1.6.2

[99] V. Mittal, M. A. Cleary, W. Herr, and N. Hernandez. The Oct-1 POU-specific domain can stimulate small nuclear RNA gene transcription by stabilizing the basal transcription complex SNAPc. *Mol Cell Biol*, **16**(5):1955–65, May 1996. 4

[100] W. D. Morgan, G. T. Williams, R. I. Morimoto, J. Greene, R. E. Kingston, and R. Tjian. Two transcriptional activators, CCAAT-box-binding transcription factor and heat shock transcription factor, interact with a human hsp70 gene promoter. *Mol Cell Biol*, **7**(3):1129–38, Mar 1987. 4

[101] T. D. Morris and L. K. Miller. Mutational analysis of a baculovirus major late promoter. *Gene*, **140**(2):147–53, Mar 25 1994. 1.6.2, 3.1

[102] B. Mukherjee, S. Burma, and S. E. Hasnain. The 30-kda protein binding to the "initiator" of the baculovirus polyhedrin promoter also binds specifically to the coding strand. *J Biol Chem*, **270**(9):4405–11, Mar 3 1995. 1.7

[103] S. Murphy, J. B. Yoon, T. Gerster, and R. G. Roeder. Oct-1 and Oct-2 potentiate functional interactions of a transcription factor with the proximal sequence element of small nuclear RNA genes. *Mol Cell Biol*, **12**(7):3247–61, Jul 1992. 4

[104] A Nakai, M Tanabe, Y Kawazoe, J Inazawa, RI Morimoto, and K Nagata. HSF4, a new member of the human heat shock factor family which lacks properties of a transcriptional activator. *Mol. Cell. Biol.*, **17**(1):469–481, 1997. 4

[105] C. J. Neumann and S. M. Cohen. Boundary formation in *Drosophila* wing: Notch activity attenuated by the POU protein nubbin. *Science*, **281**(5375):409–13, Jul 17 1998. 4

[106] M. S. Nissen and P. D. Friesen. Molecular analysis of the transcriptional regulatory region of an early baculovirus gene. *J Virol*, **63**(2):493–503, Feb 1989. 1.6.1

[107] Mauricio L. Nogueira, Victoria E. H. Wang, Dean Tantin, Phillip A. Sharp, and Thomas M. Kristie. Herpes simplex virus

infections are arrested in Oct-1-deficient cells. *PNAS*, **101**(6):1473–1478, 2004. 4

[108] Kazuhiro Okano, Adam L. Vanarsdall, Victor S. Mikhailov, and George F. Rohrmann. Conserved molecular systems of the baculoviridae. *Virology*, **344**(1):77–87, Jan 5 2006. 1.1.1, 1.3, 1.5, 1.5, 1.6.2

[109] Kazuhiro Okano, Adam L. Vanarsdall, and George F. Rohrmann. Characterization of a baculovirus lacking the alkaline nuclease gene. *J Virol*, **78**(19):10650–6, Oct 2004. 1.5

[110] B. G. Ooi, C. Rankin, and L. K. Miller. Downstream sequences augment transcription from the essential initiation site of a baculovirus polyhedrin gene. *J Mol Biol*, **210**(4):721–36, Dec 20 1989. 1.6.2, 1.7

[111] David R. O'Reilly, Lois K. Miller, and Verne A. Luckow. *Baculovirus Expression Vectors: A Laboratory Manual.* W. H. Freeman and Company, New York, USA, 1992. 3.1

[112] A. L. Passarelli, J. W. Todd, and L. K. Miller. A baculovirus gene involved in late gene expression predicts a large polypeptide with a conserved motif of RNA polymerases. *J Virol*, **68**(7):4673–8, Jul 1994. 1.6.2

[113] M. Pearson, R. Bjornson, G. Pearson, and G. Rohrmann. The *Autographa californica* baculovirus genome: evidence for multiple replication origins. *Science*, **257**(5075):1382–4, Sep 4 1992. 1.6.1

[114] O. Perisic, H. Xiao, and J. T. Lis. Stable binding of drosophila heat shock factor to head-to-head and tail-to-tail repeats of a conserved 5 bp recognition unit. *Cell*, **59**(5):797–806, Dec 1 1989. 4

[115] P. L. Platteborze and C. A. Broomfield. Expression of biologically active human butyrylcholinesterase in the cabbage looper (*Trichoplusia ni*). *Biotechnol Appl Biochem*, **31 (Pt 3)**:225–9, Jun 2000. 1.2.1

[116] R. D. Possee and S. C. Howard. Analysis of the polyhedrin gene promoter of the *Autographa californica* nucl ear polyhedrosis virus. *Nucleic Acids Res*, **15**(24):10233–48, Dec 23 1987. 1.6.2

[117] A. Ramachandran, A. Jain, P. Arora, M. D. Bashyam, U. Chatterjee, S. Ghosh, V. K. Parnaik, and S. E. Hasnain. Novel sp family-like transcription factors are present in adult insect cells and are involved in transcription from the polyhedrin gene initiator promoter. *J Biol Chem*, **276**(26):23440–9, Jun 29 2001. 1.7, 2.2.2, 2.4.1, 2.6.1, 3.1, 3.1, 3.2, 3.3, 4

[118] Aruna Ramachandran, Murali D Bashyam, Priya Viswanathan, Sudip Ghosh, Senthil Kumar M, and Seyed E Hasnain. The bountiful and baffling baculovirus: The story of polyhedrin transcription. *Current Science*, **81**:998–1010, 2001. 1.2.1

[119] C. Rankin, B. G. Ooi, and L. K. Miller. Eight base pairs encompassing the transcriptional start point are the major determinant for baculovirus polyhedrin gene expression. *Gene*, **70**(1):39–49, Oct 15 1988. 1.6.2

[120] J. C. Rapp, J. A. Wilson, and L. K. Miller. Nineteen baculovirus open reading frames, including LEF-12, support late gene expression. *J Virol*, **72**(12):10197–206, Dec 1998. 1.6.2

[121] B. M. Ribeiro, K. Hutchinson, and L. K. Miller. A mutant baculovirus with a temperature-sensitive IE-1 transregulatory protein. *J Virol*, **68**(2):1075–84, Feb 1994. 1.6.1

[122] A. Richards, M. Matthews, and P. Christian. Ecological considerations for the environmental impact evaluation of recombinant baculovirus insecticides. *Annu Rev Entomol*, **43**:493–517, 1998. 1.2.4

[123] S. M. Rodems and P. D. Friesen. The hr5 transcriptional enhancer stimulates early expression from the *Autographa californica* nuclear polyhedrosis virus genome but is not required for virus replication. *J Virol*, **67**(10):5776–85, Oct 1993. 1.6.1

[124] S. M. Rodems, S. S. Pullen, and P. D. Friesen. DNA-dependent transregulation by IE1 of *Autographa californica* nuclear polyhedrosis virus: IE1 domains required for transactivation and DNA binding. *J Virol*, **71**(12):9270–7, Dec 1997. 1.6.1

[125] G. F. Rohrmann. Polyhedrin structure. *J Gen Virol*, **67 (Pt 8)**:1499–513, Aug 1986. 1.6.2

[126] Polly Roy. Baculovirus solves a complex problem. *Nat Biotechnol*, **22**(12):1527–8, Dec 2004. 1.2.1

[127] N. K. Sah, T. K. Taneja, N. Pathak, R. Begum, M. Athar, and S. E. Hasnain. The baculovirus antiapoptotic *p35* gene also functions via an oxidant-dependent pathway. *Proc Natl Acad Sci U S A*, **96**(9):4838–43, Apr 27 1999. 1.6.1

[128] Joseph Sambrook, David W. Russell, and Joe Sambrook. *Molecular Cloning: A Laboratory Manual (3-Volume Set)*. Cold Spring Harbor Laboratory Press, January 2001. 2.2.2

[129] A. H. Schinkel, Groot M. J. Koerkamp, Van der G. T. Horst, E. P. Touw, K. A. Osinga, Van der A. M. Bliek, G. H. Veeneman, Van J. H. Boom, and H. F. Tabak. Characterization of the promoter of the large ribosomal RNA gene in yeast mitochondria and separation of mitochondrial RNA polymerase into two different functional components. *EMBO J*, **5**(5):1041–7, May 1986. 1.6.2

[130] J. M. Slack and G. W. Blissard. Identification of two independent transcriptional activation domains in the *Autographa californica* multicapsid nuclear polyhedrosis virus IE1 protein. *J Virol*, **71**(12):9579–87, Dec 1997. 1.6.1

[131] P. K. Sorger. Heat shock factor and the heat shock response. *Cell*, **65**(3):363–6, May 3 1991. 3.3, 3.3.1, 4

[132] P. Sridhar and S. E. Hasnain. Differential secretion and glycosylation of recombinant human chorionic gonadotropin (beta hCG) synthesized using different promoters in the baculovirus expression vector system. *Gene*, **131**(2):261–4, Sep 15 1993. 1.2.1

[133] M. Strubin, J. W. Newell, and P. Matthias. OBF-1, a novel B cell-specific coactivator that stimulates immunoglobulin promoter activity through association with octamer-binding proteins. *Cell*, **80**(3):497–506, Feb 10 1995. 4

[134] D. A. Theilmann, L. G. Willis, B. J. Bosch, I. J. Forsythe, and Q. Li. The baculovirus transcriptional transactivator *ie0* produces multiple products by internal initiation of translation. *Virology*, **290**(2):211–23, Nov 25 2001. 1.6.1

[135] Jane S. Titterington, Tamara K. Nun, and A. Lorena Passarelli. Functional dissection of the baculovirus late expression factor-8 gene: sequence requirements for late gene promoter activation. *J Gen Virol*, **84**(Pt 7):1817–26, Jul 2003. 1.6.2

[136] J. W. Todd, A. L. Passarelli, A. Lu, and L. K. Miller. Factors regulating baculovirus late and very late gene expression in transient-expression assays. *J Virol*, **70**(4):2307–17, Apr 1996. 1.6.2

[137] J. W. Todd, A. L. Passarelli, and L. K. Miller. Eighteen baculovirus genes, including *lef-11*, *p35*, *39k*, and *p47*, support late gene expression. *J Virol*, **69**(2):968–74, Feb 1995. 1.6

[138] E. E. Turner. Similar DNA recognition properties of alternatively spliced drosophila POU factors. *Proc Natl Acad Sci U S A*, **93**(26):15097–101, Dec 24 1996. 4

[139] Adam L. Vanarsdall, Kazuhiro Okano, and G. F. Rohrmann. Characterization of a baculovirus with a deletion of *vlf-1*. *Virology*, **326**(1):191–201, Aug 15 2004. 1.6.2

[140] Adam L. Vanarsdall, Kazuhiro Okano, and George F. Rohrmann. Characterization of the role of very late expression factor 1 in baculovirus capsid structure and DNA processing. *J Virol*, **80**(4):1724–33, Feb 2006. 1.6.2

[141] G. J. Veenstra, van der P. C. Vliet, and O. H. Destree. POU domain transcription factors in embryonic development. *Mol Biol Rep*, **24**(3):139–55, Aug 1997. 4

[142] C. P. Verrijzer, van J. A. Oosterhout, and van der P. C. Vliet. The Oct-1 POU domain mediates interactions between Oct-1 and other POU proteins. *Mol Cell Biol*, **12**(2):542–51, Feb 1992. 4

[143] Priya Viswanathan, Betapudi Venkaiah, M. Senthil Kumar, Sheeba Rasheedi, Sudhanshu Vrati, Murali D. Bashyam, and Seyed E. Hasnain. The homologous region sequence (*hr1*) of *Autographa californica* multinucleocapsid polyhedrosis virus can enhance transcription from non-baculoviral promoters in mammalian cells. *J Biol Chem*, **278**(52):52564–71, Dec 26 2003. 1.4, 1.7, 2.6.1, 3.1, 3.1

[144] X. Wang, T. Lescott, De E. Clercq, and D. C. Kelly. Baculovirus replication: inhibition of *Trichoplusia ni* multiple nuclear polyhedrosis virus by [e]-5-(2-bromovinyl)-2'-deoxyuridine. *J Gen Virol*, **64 (Pt 6)**:1221–7, Jun 1983. 1.6.2

[145] G. Westin and W. Schaffner. Heavy metal ions in transcription factors from hela cells: Sp1, but not octamer transcription factor requires zinc for DNA binding and for activator function. *Nucleic Acids Res*, **16**(13):5771–81, Jul 11 1988. 4

[146] U. Weyer and R. D. Possee. Analysis of the promoter of the *Autographa californica* nuclear polyhedrosis virus *p10* gene. *J Gen Virol*, **70 (Pt 1)**:203–8, Jan 1989. 1.6.2

[147] Leslie G. Willis, Robyn Seipp, Taryn M. Stewart, Martin A. Erlandson, and David A. Theilmann. Sequence analysis of the complete genome of *Trichoplusia ni* single nucleopolyhedrovirus and the identification of a baculoviral photolyase gene. *Virology*, **338**(2):209–26, Aug 1 2005. 1.4

[148] H. A. Wood and R. R. Granados. Genetically engineered baculoviruses as agents for pest control. *Annu Rev Microbiol*, **45**:69–87, 1991. 1.2.4

[149] T. J. Wu, G. Monokian, D. F. Mark, and C. R. Wobbe. Transcriptional activation by herpes simplex virus type 1 VP16 in vitro and its inhibition by oligopeptides. *Mol Cell Biol*, **14**(5):3484–93, May 1994. 4

[150] Y. Wu and E. B. Carstens. Initiation of baculovirus DNA replication: early promoter regions can function as infection-dependent replicating sequences in a plasmid-based replication assay. *J Virol*, **70**(10):6967–72, Oct 1996. 1.5

[151] C. L. Yang, D. A. Stetler, and R. F. Weaver. Structural comparison of the *Autographa californica* nuclear polyhedrosis virus-induced RNA poly-

merase and the three nuclear RNA polymerases from the host, *Spodoptera frugiperda*. *Virus Res*, **20**(3):251–64, Aug 1991. 1.6, 1.6.2

[152] S. Yang and L. K. Miller. Activation of baculovirus very late promoters by interaction with very late factor 1. *J Virol*, **73**(4):3404–9, Apr 1999. 1.6.2

[153] S. Yoo and L. A. Guarino. The *Autographa californica* nuclear polyhedrosis virus *ie2* gene encodes a transcriptional regulator. *Virology*, **202**(2):746–53, Aug 1 1994. 1.6.1

[154] E. Zelzer and B. Z. Shilo. Interaction between the bHLH-PAS protein trachealess and the POU-domain protein drifter, specifies tracheal cell fates. *Mech Dev*, **91**(1-2):163–73, Mar 1 2000. 4

[155] E. A. Zemskov, W. Kang, and S. Maeda. Evidence for nucleic acid binding ability and nucleosome association of *Bombyx mori* nucleopolyhedrovirus BRO proteins. *J Virol*, **74**(15):6784–9, Aug 2000. 4

CURRICULUM VITÆ

Senthil Kumar M
Laboratory of Molecular & Cellular Biology
Centre for DNA Fingerprinting & Diagnostics,
ECIL Road, Hyderabad 500 076, INDIA
Phone No.: +91–40–27150008
Email.: senthil@cdfd.org.in

Educational Qualifications

Ph.D (Submitted to University of Hyderabad), 2006
Centre for DNA Fingerprinting & Diagnostics.
Title of thesis: "Role of upstream sequences on the *Ac*MNPV polyhedrin promoter transcription".

Master of Science 1996 – 1998 M.Sc. in Plant Science at the Centre for Advanced Study in Botany, University of Madras. OPGA 4.2 out of 5.0

Bachelor of Science 1993 – 1996 B.Sc. in Botany at Ramakrishna Mission Vivekananda College, University of Madras (First Class).

Publications

- Viswanathan, P., Venkaiah, B., **Kumar, S.**, Rasheedi, S., Vrati, S., Bashyam, M.D. and Hasnain, S.E. (2003) The homologous region sequence (*hr1*) of *Autographa californica* multinucleocapsid polyhedrosis virus can enhance transcription from non-baculoviral promoters in mammalian cells. *Journal of Biological Chemistry* **278**: 52564 - 52571

- Aruna Ramachandran, Murali D Bashyam, Priya Viswanathan, Sudip Ghosh, **Senthil Kumar M** and Seyed E Hasnain (2001) The bountiful and baffling baculovirus: The story of polyhedrin transcription. *Current Science* **81**(8) : 998–1010

Other Publications

- Yellaboina S, Seshadri J, **Senthil Kumar M**, Ranjan A (2003) PredictRegulon: A web server for the prediction of the regulatory protein binding sites and operons in prokaryote genomes. *Nucleic Acids Research* **32**:W318-20.

- Murali D Bashyam, Gorinabele R Savithri, **Senthil Kumar M**, Calambur Narasimhan and Pratibha Nallari (2003). Molecular genetics of familial hypertrophic cardiomyopathy (FHC). *Journal of Human Genetics* **48**:55-64.

Poster/Conference/Oral presentations

- **Senthil Kumar M**, Seyed E Hasnain and Murali D Bashyam (2005). Influence of upstream sequence elements on hyperactivation of transcription from the *Ac*NPV polyhedrin gene promoter, In **The** *IX* **Asia-Pacific Conference on Transcription, National Health Research Institutes (NHRI), Taiwan.** (*Oral Presentation*).

- Sudip Ghosh, Priya Viswanathan, **M. Senthil Kumar**, And Seyed E. Hasnain (2002). Baculovirus Very Late Gene Transcription Is A Complex Interplay Between Viral And Host Factors. In **The** *VII* **Asia-Pacific Conference on Transcription (ACT-VII), Kuala Lumpur, Malaysia** (*Poster Presentation*).

The bountiful and baffling baculovirus: The story of polyhedrin transcription

Aruna Ramachandran, Murali Dharan Bashyam, Priya Viswanathan, Sudip Ghosh, M. Senthil Kumar and Seyed E. Hasnain*,†

Centre for DNA Fingerprinting and Diagnostics, ECIL Road, Nacharam, Hyderabad 500 076, India; †Also at National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi 110 067, India and Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur, Bangalore 560 064, India

Baculoviruses are a unique group of eukaryotic viruses that parasitize insects. The prototype member of the family *Baculoviridae* is *Autographa californica* multiple nuclear polyhedrosis virus (AcMNPV). Global interest in baculovirus biology stems from two important uses of baculoviruses – as biopesticides and as a highly favoured eukaryotic expression system for the large-scale production of recombinant proteins in the laboratory. Of late, baculoviruses have invited renewed interest by virtue of their potential use as a delivery system in gene therapy. Although the baculovirus expression vector system (BEVS) is extensively used worldwide, the transcriptional regulation of the hyperactive promoters used to drive foreign gene expression still remains shrouded in mystery. It is clear, however, that this regulation involves an intricate interplay of both host and viral factors. This review provides an overview of what we do know about the mechanisms of transcription of baculoviral genes, with special emphasis on the polyhedrin promoter, the workhorse promoter of the BEVS, and the insect cell host factors involved in enhancing transcription from it.

Baculoviruses: Versatile and effective biopesticides

The natural hosts of baculoviruses are insects belonging mainly to the classes Lepidoptera (butterflies and moths), Hymenoptera (sawflies) and Coleoptera (beetles). Many of these insects are plant pathogens, infecting agriculturally important crops and forest trees. Although chemical pesticides continue to be used to tackle this problem, farmers and agricultural scientists have of late recognized the importance of developing a safer and more eco-friendly alternative to such harsh chemical insecticides. Baculoviruses present the perfect biological solution to curb insect pest populations while simultaneously respecting the environmental balance. The viruses can be sprayed as a powder over the crops, whereby they are ingested by the feeding insect larvae, multiply in the host and ultimately kill the organism, releasing fresh virus particles into the environment to start the cycle all over again. There are several advantages of using baculoviruses as insecticides – they can be specifically targeted to certain pests, are self-propagating, safe for human handling and do not pollute the environment, thus preventing health hazards.

However, for a variety of reasons, it has proved more difficult than expected to develop effective baculovirus insecticide formulations. The reason for this is mainly the virus's low persistence in the environment, especially when recombinant baculoviruses are used. Another major drawback is the slow speed of killing in contrast to chemical insecticides which have a much more rapid knockdown effect. The widespread use of baculoviruses as pesticides is further hindered by their narrow host range, instability of insecticide formulations, and problems in registration and patentability. However, despite these limitations, the enormous ecological advantages of this approach makes the attempt to develop new-age and more efficient baculovirus biopesticides on a global scale well worth the effort[1].

The baculovirus expression vector system: A biofactory *par excellence* for the production of recombinant proteins

The second – and far more important – use of baculoviruses is as a vehicle for large-scale protein production. Baculoviruses have been very successfully used for the past couple of decades, for the expression of high levels of recombinant proteins[2,3]. Hundreds of proteins have been expressed to date by constructing recombinant baculoviruses. The heterologous gene is expressed usually under the control of the hypertranscribed *polyhedrin* (*polh*) or *p10* gene promoters that are turned on very late following viral infection – after 48 h or so.

The polyhedrin protein forms the crystalline matrix of viral polyhedral bodies (also called polyhedra), whereas the p10 protein forms large arrays of fibrous material,

*For correspondence. (e-mail: ehtesham@www.cdfd.org.in)

primarily in the nucleus but sometimes in the cytoplasm as well. Polyhedra formation is crucial for viral infection of insects in the wild, since they shield the delicate virions from harsh environmental stresses. Further, polyhedra act as a useful carrier of the virus particles; the crystalline polyhedrin matrix is solubilized in the mid-gut of the insect, thus releasing the virus particles which infect the mid-gut cells. Polyhedrin is thus an extremely important protein for virus survival and propagation in its native environment. However, under laboratory conditions, where growth parameters are tailored to suit the virus, the polyhedrin matrix is no longer a prerequisite for virus survival. Thus, the *polh* gene can be replaced with a heterologous gene of choice, which would then be efficiently expressed from the strong *polh* promoter. The same logic holds good for the *p10* gene. The hypertranscribed *polh* and *p10* promoters are thus the workhorse promoters of the baculovirus expression vector system (BEVS).

The BEVS owes its popularity to many more qualities other than just the unusual strength of the *polh* and *p10* promoters. Some of these are described briefly below:

Eukaryotic environment for protein production

The BEVS provides the necessary higher eukaryotic environment essential for the proper folding, post-translational modification, disulphide bond formation and other modifications required for the functional activity of many eukaryotic proteins. Post-translational modifications that have been reported to occur in the insect BEVS include signal cleavage, proteolytic cleavage, N-glycosylation, O-glycosylation, acylation, amidation, phosphorylation, prenylation and carboxymethylation[4]. All these modifications occur at sites identical to those in the wild type proteins, reinforcing the usefulness of the BEVS as one of the most favoured systems for expressing functionally active recombinant proteins.

Extremely high levels of expression

As mentioned earlier, the *p10* and *polh* promoters are the most commonly used promoters to drive the expression of foreign genes. Being unusually strong promoters, they hypertranscribe the gene(s) put under their control to the extent that the recombinant protein can account for about 25–50% of the total cellular protein. The *polh* promoter is the stronger of the two, and induces higher expression levels, but investigators have reported protein yields of up to a gram of recombinant protein her litre of insect cell culture, i.e. about 10^9 cells, using either promoter. Average protein yields lie in the range of 10–100 mg of protein per 10^9 cells. In terms of protein yield alone, the BEVS has been observed to consistently outperform other expression systems. The use of homologous enhancer-like sequences has been demonstrated to further enhance foreign gene expression levels several fold[5]. In addition, live caterpillars have also been used as a host for high level expression, to further improve on the economics of expression[6].

However, the expression kinetics differ from protein to protein and the promoter used[7], with the time of expression playing a critical role in proper post-translational modification and secretion[8]. Expression levels are also known to vary with the cell line used[9]. Another important parameter is the codon usage pattern of the recombinant gene: heterologous genes which use non-optimal codons of the insect host are observed to be poorly expressed[10,11]. Finally, the translation initiation context – as defined by the Ranjan–Hasnain consensus sequence[11] – was also found to play a major role in regulating protein expression levels. Recently, a recombinant baculovirus carrying different B and T cell epitopes from nine stage-specific antigens of *Plasmodium falciparum*, has been used to express, to very high levels, a recombinant multi-antigenic protein – a strong putative vaccine candidate for malaria[12,13].

Capacity for large inserts and simultaneous expression

The baculovirus nucleocapsid is predicted to be capable of accommodating inserts as large as 100 kb. Although this has not been practically tested (the largest insertion till date being only about 15 kb)[4], no investigator has been hampered by the size of the heterologous gene(s) used for insertion. Further, a number of transfer vectors have been genetically engineered to simultaneously express multiple genes under the control of different viral promoters both in cell culture[14–16] and in *Spodoptera* larvae[17].

Baculovirus-mediated gene transfer into mammalian cells

The BEVS has so far used only insect cells as a host for the expression of heterologous genes carried by recombinant baculoviruses. Interestingly, recent reports have demonstrated that they could also be used as gene delivery systems in mammalian cells. Although baculoviruses infect over 30 species of Lepidopteran insects, they are incapable of replicating in other insects or in any of over 35 mammalian cell lines studied[18,19]. However, the virus does enter mammalian cells and the viral DNA is capable of reaching the nucleus. Experimental studies have shown that when an exogenous promoter, such as that derived from Rous sarcoma virus or cytomegalovirus, is inserted into the baculovirus genome,

the modified virus becomes capable of gene expression in non-Lepidopteran cell lines, including various mammalian cells[20,21]. Boyce and coworkers[20] showed that reporter gene expression from a recombinant baculovirus was significantly higher in the human hepatocellular carcinoma cell line, HepG2, than in cell lines derived from other tissues like monkey kidney, human kidney, cervix and lung, B-cell, T-cell, rat adrenal, and mouse embryo fibroblast and muscle. Thus, new generation recombinant baculoviruses could well evolve into a gene delivery system of the future[22].

Safety and simplicity

Baculoviruses are relatively simple to use. Constructing recombinant viruses is much faster and easier than developing stable recombinant high-expressing cell lines, and the host insect cells can be grown at 27°C either as adherent or suspension cultures. Cells can be grown in a BOD incubator since CO_2 is not required for growth. Scale up has also been perfected with time, thus making it easy to produce large amounts of host insect cells in fermenters for subsequent viral infection and expression of recombinant proteins. Further, since baculoviruses have no non-arthropod hosts *in vivo*, they are harmless to humans and can be safely handled by investigators with no special precautions. However, a recent report by Gronowski *et al.*[23], have shown that *Ac*NPV is capable of provoking an anti-viral response in murine and human cell lines by inducing interferons. Although the possibility of baculoviruses infecting humans *in vivo* is remote, these findings nevertheless justify the use of greater precautionary measures in handling baculoviruses than in the past.

Gene expression in baculoviruses

During an NPV infection, more than a hundred viral genes are expressed in a cascade that can be broadly divided into three stages – early, late and very late. Each stage is characterized by the expression of a unique set of genes in a well-regulated cascade, with the products of one group of genes required for the expression of the next set[24]. By definition, early genes are expressed prior to viral DNA replication. Most, if not all, immediate-early genes encode transcriptional regulatory proteins[25–29]. DNA replication activates the viral template in a manner not yet defined and enables the late and very late classes of genes to be expressed, which encode proteins essential for virion assembly and viral occlusion formation. Although one of the most intriguing aspects of baculovirus biology concerns the control of the viral transcription cascade, we are still a long way from deciphering the precise mechanisms involved and the host and viral factors which play a role in this finely-orchestrated process.

Early gene expression

Early viral gene expression spans the time period from 0–6 h post-infection (hpi) and results in the transcription of genes encoding proteins required for viral DNA replication and late gene transcription. Most baculovirus early genes have a TATA box located 25 to 30 bp upstream of a conserved mRNA transcription initiation site that consists of the sequence CAGT[30]. The CAGT element, which is the first true initiator element to be discovered in baculoviruses, has been shown to be critical for the transcription of early genes. Substitution of the CAGT sequence resulted in a reduction of both reporter activity and *in vitro* transcripts, although transcripts initiated accurately[31]. However, not all early genes have a CAGT and/or TATA element. A notable example is the *dnapol* gene in both *Bm*NPV and *Ac*NPV, which is observed to initiate transcription from a GC rich region, with no canonical TATA box or CAGT motif present[31]. Early gene expression is dependent on a *a*-amanitin sensitive, tagetitoxin-insensitive host RNA polymerase II (ref. 32) and early gene promoters resemble typical eukaryotic class II promoters.

Four *Ac*NPV early genes, *ie-O, ie-1, ie-2* (or *ie-N*), and *pe-38*, have been shown to be important for transactivating early baculovirus promoters in transient expression assays. IE-0, IE-2, and PE-38 mRNAs are expressed only during the early phase of infection[25,33,34]. In contrast, IE-1 RNA is expressed during both the early and late phases of infection[28,33].

IE-1, a 582-amino acid long multifunctional transcriptional regulatory phosphoprotein, has been shown to transactivate a number of delayed-early genes including *39K*, *ie-2* and *p35* (refs 26–28) and at least one late gene, *39K* (ref. 29). The stimulatory effect of IE-1 is greatest when the target promoter is *cis*-linked to *Ac*NPV homologous repeat regions (*hrs*). Hr regions, which are present at nine dispersed locations in the *Ac*NPV genome, have been shown to act as enhancers of transcription[25,28,35–37] and also in certain cases as origins of viral replication[38–40]. However, competitive PCR methods used to map the activation profiles of *Ac*NPV *oris*, have demonstrated that the *ie-1* gene promoter also acts as an *ori*, and is activated in a temporal fashion[41]. IE1 has been shown to also negatively regulate certain promoter regions, e.g. the promoters of the *Ac*NPV *ie-0* and *ie-2* genes, which themselves are regulatory genes. Results indicate that IE1 brings about this downregulation of transcription by binding directly or as part of a complex to IE-1 binding motifs (5′-ACBYGTAA-3′) near the mRNA start site[42]. In addi-

tion, it has been shown that *ie-1* can activate the *he-65* gene promoter in insect *Trichoplusia ni* (*T. ni*) 368 and mammalian BHK21 cells[43], demonstrating that the activity of this multifaceted protein is not confined to a narrow host range, but probably involves a generalized mechanism conserved across various species.

Several other baculovirus early genes also have important regulatory functions. IE-0 is a protein identical to IE-1 except that it has an additional 54 amino acids at the N terminus as a result of a splicing event[33]. IE-0 mediated activation has been shown to require an *hr* enhancer[44] unlike IE-1, which can regulate transcription in the presence or absence of a *cis*-linked hr sequence. A further difference between IE-0 and IE-1 is that the former does not appear to downregulate the *ie-0* promoter[44].

ie-2, another early regulatory gene, both augments *ie-1* transactivation[25] and brings about transcription activation independent of *ie-1* (refs 26, 45). IE-2 activates the *ie-1* promoter approximately 2.5-fold in transient expression assays[45,46] and has been demonstrated to block cell cycle progression in a variety of cell lines, including *Sf*21 and *T. ni*[46].

The *pe-38* gene has been shown to transactivate the *p143* gene promoter[36]. Other *Ac*NPV genes such as *me-53* and *cg-30* are also thought to have some role in activating gene expression, based on their sequence similarity with transcriptional regulators from other systems[47,48].

The *p35* gene, an inhibitor of members of the ICE family of cysteine proteases and a major determinant of virus host range, is another essential early gene. p35 is required to suppress the apoptotic response of *S. frugiperda* cells to *Ac*NPV infection by acting as a suicide inhibitor of caspases[49,50], as well as an antioxidant[51,52], both in *Sf* 21 and *Sf* 9 cells and *S. frugiperda* larvae[49]. The important role played by the *p35* gene suggests that effective inhibition of apoptosis is required for both efficient viral DNA replication and gene expression[53,54].

The early *p143* gene is another important player in viral replication[53,55], transcription, shutdown of host protein synthesis and viral host range determination[54–56]. p143 has helicase-like motifs, a nuclear localization signal and a leucine zipper motif, is synthesized in virus-infected cells prior to the initiation of viral DNA replication and has been shown to bind nonspecifically to DNA[57]. *In vivo* complementation assays[58] revealed that some of the putative helicase motifs are not essential for p143 function; however, mutations within an ATP-binding motif, a potential helix-turn-helix region, and certain large amino acid deletions inactivated protein function. Recent reports have confirmed that p143 is indeed a DNA helicase with ATPase activity[59,60]. LEF-3, a single-stranded DNA-binding protein, has been shown to interact with p143 and help in localizing p143 to the nucleus[61]. p143 is also a crucial determinant of viral host range; interspecific replacement of a short sequence in the *Ac*NPV *p143* gene renders the virus capable of infecting *Bombyx mori* larvae[59]. It has been demonstrated that two key mutations in the *Ac*NPV p143 host specificity domain is enough to render the virus replication competent in *Bm5* cells, and kill *B. mori* larvae.

Finally, *orf121*, another baculovirus early gene has been observed to stimulate expression from the late *39K* gene promoter, an activity which is dependent on *ie-1* as well[62]. It was later discovered that *orf121* enhanced *ie-1* expression which in turn was responsible for upregulating expression of the *39K* gene[62].

Late and very late gene expression

The transition between early and late viral gene expression is the most distinctive regulatory event in the baculovirus transcription/infection cascade. Whereas early gene expression is dependent on an α-amanitin sensitive RNA polymerase present in uninfected cells, late and very late gene transcription involves a novel α-amanitin and tagetitoxin resistant, virally encoded RNA polymerase[32,63–66].

The baculovirus late and very late gene promoters resemble mitochondrial and bacteriophage T7 late promoters in that a short conserved sequence serves both as a promoter and an initiator element. The most conserved sequence element of *Ac*NPV late and very late promoters is the transcription initiator (A/T/G)TAAG[30]. The strength of expression from the promoters is critically dependent on the context of the TAAG sequence, with the 18 bp region encompassing the TAAG having been shown to be the minimal promoter determinant for basal transcription from the late *vp39* promoter[67]. However, essential promoter determinants for the very late *polh* and *p10* genes include not only the 12-bp initiator AATAAGTATTTT but also a downstream A + T rich region corresponding to the 5′ untranslated leader sequence of their mRNAs[68,69]. This untranslated sequence is responsible for the 'burst' in transcription observed during very late gene expression and is thus termed the 'burst' sequence. It is a defining element for very late gene promoters, being absent in late promoters. Mutations within the burst sequence reduce expression during the very late phase of infection by 10 to 20-fold and lower both the steady-state levels of *polh* mRNA and the rate of transcription initiation from the *polh* promoter[68]. In contrast, mutations in sequences upstream of the *polh* promoter TAAG motif have relatively milder effects on *polh* gene expression[69]. Progressive deletions of the *p10* promoter also suggest the presence of a burst sequence that is essential for strong expression during the very late phase[69].

The *Ac*NPV very late factor 1 (*vlf-1*) gene has been shown recently to play an important role in very late gene expression from the *polh* and *p10* gene promoters. *vlf-1*, a late gene, was originally identified by characterization of an occlusion-defective mutant virus, *ts*B837, which produced extremely low levels of *polh* and *p10* transcripts during the very late phase of infection. In transient expression assays, *vlf-1* was shown to stimulate expression from very late promoters but not from late promoters[70], making it probably the only well characterized viral factor known to date that regulates only very late but not late gene expression. Interestingly, recombinant baculoviruses with altered *vlf-1* expression revealed that the time of *vlf-1* expression and/or the concentration of VLF-1 in the cell was critical for switching on the *polh* and *p10* genes[71,72]. Recent studies have revealed that the VLF-1 protein binds directly to the untranslated regions of the polh and p10 promoters and is closely correlated with their transactivation[73]. Thus, VLF-1 seems to be a crucial player in the regulation of baculovirus very late gene expression.

Until recently, eighteen *Ac*NPV genes, called *lef*s (late expression factors), were identified[74–77] which are necessary to support transient expression of a reporter gene under the control of the late *vp39* promoter. The same set of *lef*s has also been shown to be involved in transient expression of the late basic 6.9-kDa protein gene[70] and the very late *polh*[68] and *p10* genes[70]. All these genes have been shown to affect steady state levels to reporter gene transcripts, implying that their effect is mainly at the transcriptional and not at the translational level[54].

Nine of the 18 genes (*ie-1*, *ie-2*, *lef-1*, *lef-2*, *lef-3*, *lef7*, *p143* (also called *dnahel*), *dnapol*, and *p35*) are necesary and sufficient for supporting replication of a plasmid containing a viral origin of replication[54]. Hence, they may act indirectly, by supporting viral DNA replication post-infection, which is also essential for late and very late gene expression.

The *ie-1* and *ie-2* genes, as explained above, are important immediate early genes which regulate the expression of their own genes and other immediate – and delayed early genes in the viral cascade. *p143* (*dnahel*) and *dnapol* encode polypeptides with sequence similarity to DNA helicases and polymerases, respectively, suggestive of their role in viral DNA replication. The viral late expression factor 3 (*lef-3*) gene product has been shown to have single-stranded DNA binding activity[78]. LEF-3 forms a homotrimer in solution[79] and helps to localize the P143 DNA helicase to the nucleus[61]. The *Bm*NPV LEF-3 protein has been demonstrated to have a helix-destabilizing activity, which may act in concert with P143 to facilitate strand separation during DNA replication. The involvement of *p35*, a known inhibitor of apoptosis, suggests that the apoptotic pathway needs to be blocked in order for DNA replication and subsequent transcription to occur.

The *lef-1*, *lef-2* and *lef-7* genes have been shown to be essential for DNA replication. The LEF-1 protein has primase-like motifs, which when mutated abrogates the ability of LEF-1 to support transient DNA replication[80]. Furthermore, two-hybrid screens have demonstrated, although the exact mechanism is unclear, that LEF-1 and LEF-2 interact with each other, and probably function synergistically in the replication process[80]. The *lef-2* gene also plays an important role in replication[56]. Merrington and coworkers[81] identified a mutant virus, VLD1 which was defective in late and very late gene expression which was subsequently found to be the result of a point mutation in the *lef-2* gene. Interestingly, the virus was not defective in DNA replication, suggesting that *lef-2* may play a dual role, both in DNA replication and very late gene expression, with different domains of the protein required for different functional roles[81]. The *lef-7* gene has been demonstrated not to be absolutely essential for DNA replication, but instead have a stimulatory effect. *lef-7*, like *lef-3*, is dependent on the multifunctional trans-regulatory gene, *ie-1*, for its activity in transient expression assays[75].

The remaining 9 *lef*s, i.e. *lef-4*, *lef-5*, *lef-6*, *lef8*, *lef-9*, *lef-10*, *lef-11*, *39K* and *p47* are directly involved in regulating late and very late gene expression, and not simply as a consequence of supporting DNA replication.

The *39K* gene encodes a phosphoprotein, pp31, which associates with the virogenic stroma, a virus-induced nuclear structure which appears to be the site for nucleocapsid assembly[82]. *39K*, like several other genes, is also regulated by the transactivator IE-1. ORF121 and IE-2 also upregulate expression of 39K, although this was found to be by virtue of their enhancement of IE-1 expression[62]. P35 also enhances expression of 39K by a mechanism which is thought to be different from that adopted by ORF 121, IE2 and IE-1 (ref. 62).

Recently, the virus-specific RNA polymerase was purified[63] and was apparently found to be composed of equimolar subunits of 4 lefs: LEF-4, LEF-8, LEF-9, and p47. LEF-8 and LEF-9 were earlier described as having some, though not extensive, sequence similarity with other DNA-dependent RNA polymerases[74,77]. This polymerase has been suggested to be the simplest DNA-directed RNA polymerase reported till date from any eukaryotic source. The polymerase supported transcription from late and very late promoters but was not active on early promoters. Interestingly, both late and very late promoters were transcribed with equal efficienty, highlighting the fact that the polymerase lacked the factors to bring about temporal expression of the late and very late genes in the sequential order required during the viral cascade of infection. Significantly, this complex of 4 equimolar subunits has not been shown to

function in *in vitro* reconstitution experiments but has been demonstrated to act only within the insect cell environment. The role of other host factor(s) in this process, *in vivo*, has thus not been excluded.

The LEF-4 protein was recently shown to have guanylyltransferase activity. It could hydrolyse the gamma-terminal phosphate of the 3′ end of RNA and also ATP and GTP to their respective dinucleotide forms[83]. These activities and the fact that LEF-4 has a KXDG motif, and homology with motifs common to viral and cellular guanylyltransferases, suggest that it may be part of a baculovirus RNA-capping complex. The triphosphatase, guanylyltransferse and methyltransferase components of the capping apparatus are organized differently in metazoan, viral and fungal systems. However, vaccinia virus capping enzyme has been shown to combine all three properties in a single multifunctional protein. LEF-4 combines the first two functions of the capping apparatus and is thus thought to be the major player in baculoviral mRNA capping[83] besides being part of the core RNA polymerase complex.

Of the remaining constituents of the viral RNA polymerase, the LEF-8 and LEF-9 proteins are hypothesized to constitute the catalytic core of the RNA polymerase since they possess amino acid sequence motifs with homology to other polymerases[63]. However this has not been proved experimentally, and the role of the p47 protein is still unknown.

No information is currently available on the functions of *lef-6*, *lef-10* and *lef-11* and the mechanism by which they are involved in late and very late gene expression. Although the exact role of *lef-5* too has not been delineated, it has been recently reported that the C-terminal end of the protein contains a novel domain which is homologous to the zinc ribbon domain of RNA polymerase elongation factor IIS (TFIIS) from a variety of taxa[84]. The same report also documents the interaction of the LEF-5 protein with itself and suggests that LEF-5 may be involved in transcript elongation.

Apart from these 18 *lefs*, a new factor, *lef* 12, has been identified of late, which is also supposed to be essential for late and very late gene transcription[85]. A set of plasmids was constructed in which each of the 18 *lef* open reading frames (ORFs) was controlled by the *Drosophila melanogaster* heat shock protein 70 (hsp70) promoter and epitope tagged. However, this set of plasmids failed to support transient late gene expression. The inability of the *p47* ORF to replace the *p47*-containing plasmid supplied in the *lef* plasmid library led to the identification of a 19th late expression factor gene (*lef-12*) located adjacent to the *p47* gene. The sequence of *lef-12* is predicted to encode a 21 kDa protein with no homology to any previously identified protein. The function of *lef-12* is yet to be elucidated.

Role of host factors in very late gene expression

All of the factors with an established role so far in viral late and very late gene expression, including the *polh* and *p10* genes, are viral proteins. However, many insect cell host factors too are critically important for this process. Our laboratory has been working on this aspect for the past several years and has identified several cellular proteins which play a crucial role in very late gene expression, using the *polh* promoter as a model[86].

Etkin and coworkers[87] had earlier identified a 200 kDa protein present in *Sf9* cells which binds to the polh promoter and was implicated in negative regulation of the promoter. However, the functional importance of this putative factor was not elucidated and there were no further reports about it thereafter.

The first host protein to be clearly identified as having an authentic role in *polh* transformation is the unusual 30-kDa transcription factor, the polh promoter binding protein (PPBP) identified in our laboratory[88,89]. This phosphoprotein binds with very high affinity and specificity to a hexanucleotide sequence motif, AATAAA, present within the minimal promoter immediately 5′ to the octanucleotide motif TAAGTATT which encompasses the transcription start point[90]. PPBP probably acts as an initiator binding protein (IBP) involved in the recruitment of the transcription machinery. PPBP specifically binds to the coding strand of the promoter[91] with increased affinity, compared to the duplex promoter, thus maintaining the promoter at the initiation point in a 'melted' state and allowing for increased rounds of transcription. Sequestering PPBP using its cognate binding motif – the polh promoter B domain – resulted in a drastic reduction in transcription

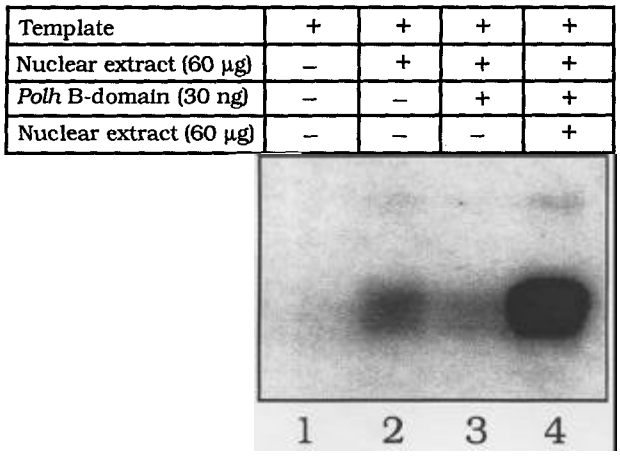


Figure 1. PPBP is required for transcription from the polh promoter. *In vitro* transcription from Ppolh was carried out using a C-free template plasmid. Transcription was reduced drastically when PPBP was titrated out using its cognate binding motif (compare lanes 3 and 2) and restored when the reaction was replenished with *Sf*9 nuclear extract containing PPBP (lane 4). Lane 1 is a control reaction carried out in the absence of template DNA.

Probe	p10 promoter B domain						
NE-*Sf*9	-	+	+	+	+	+	+
Competitor	-	-	10B	29B	mO	mH	pUC

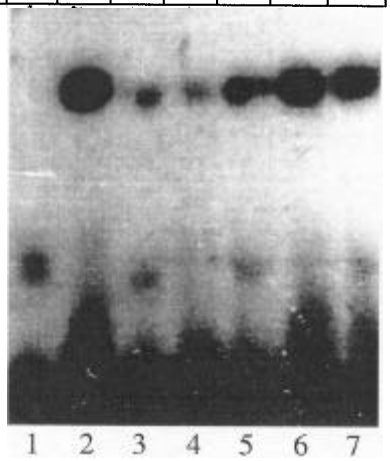


Figure 2. PPBP also binds to the *p10* gene promoter. EMSA showing a PPBP–DNA complex using the p10 B domain promoter sequence as probe. The complex (lane 2) could be specifically competed out using either the cold p10 or p29 (polh) promoter sequences (lanes 3 and 4, respectively) but not by polh promoter oligonucleotides carrying mutations in the TAAGTATT (mO) or AATAAA (mH) motifs (lanes 5 and 6, respectively). pUC18 DNA, used as a non-specific competitor, did not compete for binding either (lane 7). Lane 1 represents the free probe.

in vitro and *in vivo*[90] (Figure 1). Interestingly, PPBP has been shown to interact with the transcriptionally important AATAAA and TAAGTATT motifs of the *p10* promoter also[92] (Figure 2), suggesting that it may have a role in the regulation of gene expression from very late promoters in general (Jain, A., Ph D thesis).

The second host factor identified in our laboratory is the 38 kDa homologous region-1 (*hr1*) binding protein (hrBP), which binds at multiple sites within the *Ac*NPV *hr1* enhancer element, with high specificity and affinity[37]. There are nine homologous region sequences (*hrs*) dispersed throughout the *Ac*NPV genome, viz. *hr1*, *hr1a*, *hr2*, *hr2a*, *hr3*, *hr4a*, *hr4b*, *hr4c* and *hr5* (refs 93, 94), which act as origins of replication (*ori*s) as well as enhancers of transcription for some baculovirus early and late genes. Our laboratory has shown that the *hr1* enhancer enhances transcription from the *polh* promoter and also act as a putative *ori*, with both functions having distinct sequence requirements[40]. The 750 bp hr1 sequence element contains 5 imperfect palindromes with an *Eco*RI site at the centre of each palindrome. An intact palindrome along with the flanking sequence is the minimal requirement for the enhancer function of hr1 (Figure 3, ref. 40). This is in contrast to its replication function, where a palindrome alone was found to be both necessary and sufficient for the *ori* function of hr1 in transfected cells[40]. Hr1BP requires phosphorylation

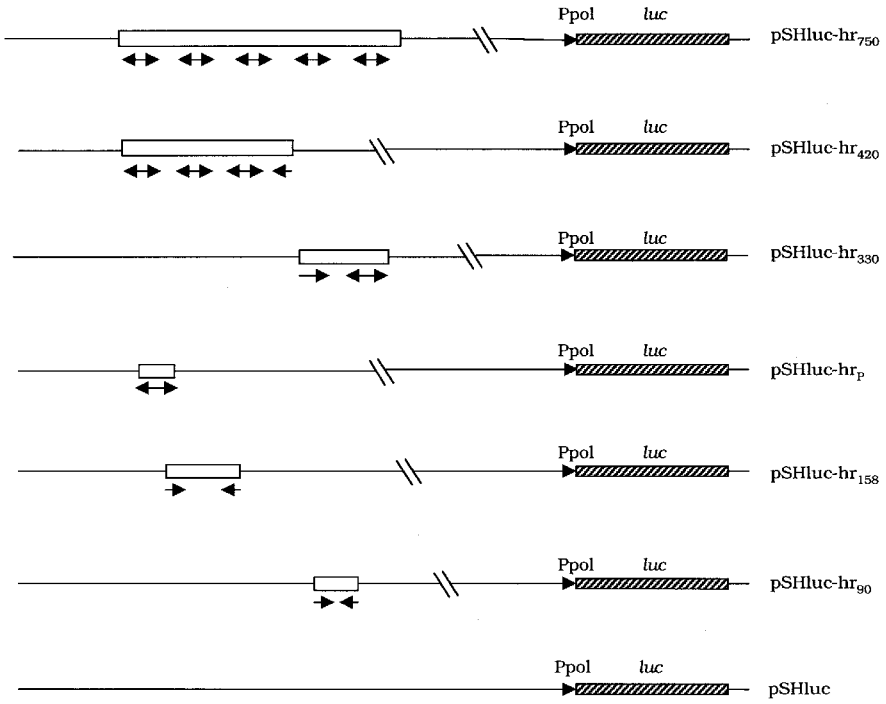


Figure 3 *a*. Enhancement of luciferase expression is a function of hr1 modules. Schematic representation of plasmid constructs carrying different components of hr1 used for analysis in transient expression assays.

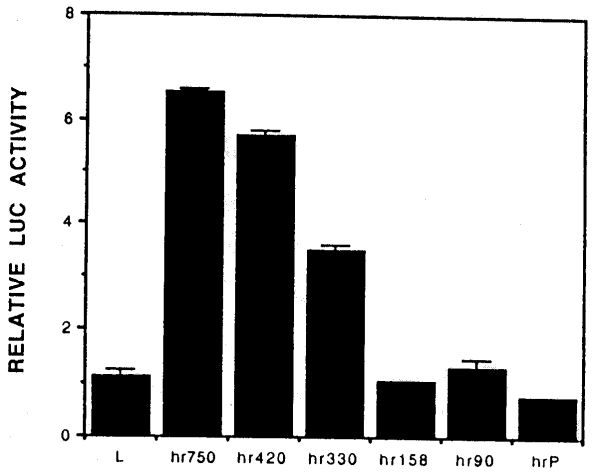


Figure 3 *b*. Enhancement of luciferase expression is a function of hr1 modules. Luciferase activity from these constructs represented as fold-enhancement over pSHluc (L). The different hr reporter constructs are indicated.

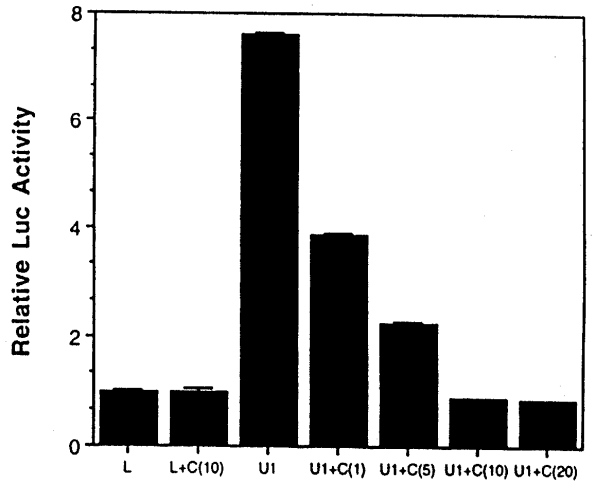


Figure 4. Hr1BP is required for the enhancer function of hr1. Luciferase expression of hr1-containing constructs (pSHluc-hrU1, U1) co-transfected with varying amounts of competitor plasmid (C, pSH-hrU1) carrying the hr1 sequence (bars 3–7). Bars, 4, 5, 6, and 7, corresponding to 1, 5, 10, and 20 µg of competitor plasmid respectively, show a drop in luciferase expression relative to bar 3, which has no competitor. Bars 1 and 2 are controls showing that there is no effect on luc expression with (lane 2) or without (lane 1) competitor in the case of a reporter plasmid which carries no hr1 sequence (L, pSHluc).

for binding and is essential for the enhancer function of *hr1*, as demonstrated by *in vivo* competition experiments[37] (Figure 4).

The host factor PPBP, which binds to the polh promoter and probably helps recruit the RNA polymerase[90], specifically recognizes the transcriptionally important AATAAA and TAAGTATT motifs within the polh promoter. Promoter vector constructs (Figure 5, ref. 95) were made, where the AATAAA was mutated to CCCCCC and the mutant promoter used to drive transcription of a downstream luciferase reporter gene. It was expected that reporter gene expression would be drastically affected, if not reduced to zero. Surprisingly, however, the drop in luciferase reporter expression (Figure 5, ref. 95) from the mutant promoter construct *vis-à-vis* the wild type promoter was not as sharp as expected. On closer analysis, it was apparent that an ~ 766 bp stretch present upstream of Ppolh could compensate for mutations within the promoter. When a control plasmid containing the mutant promoter with no upstream regions was analysed, it showed almost no luciferase expression, underscoring the importance of upstream sequences in the regulation of Ppolh. In a separate experiment, *Bal*31 deletion analyses of a 4 kb region upstream of the *polh* promoter identified two transcriptionally important regions, region I and region II (containing the 766 bp upstream sequence), spanning map units 0 to 1.5 and 2.5 to 3.12 respectively on the *Eco*RI 'I' fragment of the viral genome (Ch. Anser Azim, unpublished data). The deletion of these regions resulted in a significant reduction in *polh* promoter-driven reporter gene expression. These findings, coupled with the 'promoter-knockout' analysis results described above, promoted a more detailed dissection of region II.

The Sp family of proteins

A careful analysis of region II revealed a sequence motif, which we termed AcSp (for *Autographa californica* nuclear polyhedrosis virus Sp-like sequence), which carried GC and GT box-like motifs which are known to be bound by the Sp family of proteins (Figure 6). Keeping in mind the functional significance of the upstream sequences and the fact that Sp-family proteins have so far not been demonstrated in insect cells, it was pertinent to explore this region further to determine the importance, if any, of the AcSp sequence notif and any trans-acting factors that may bind to it. It was observed that AcSp and the consensus Sp1 sequence (cSp) specifically bound factor(s) in HeLa and *Spodoptera frugiperda* (*Sf*9) insect cell nuclear extracts to generate identical binding patterns, indicating the similar nature of the factor(s) interacting with these sequences. Recombinant plasmid constructs carrying the AcSp and cSp oligonucleotides enhanced *in vivo* expression of a polh promoter-driven luciferase gene (Figure 7). *In vivo* mopping of these factor(s) significantly reduced transcription from the polh promoter (Figure 8, ref. 95), and recombinant viruses carrying deletions in the upstream sequences containing AcSp confirmed the requirement of these factor(s) in polh promoter-driven transcription in the viral context (Figure 9, ref. 95). Our results thus document, for the first time, DNA–protein interactions involving novel members of the Sp-family of proteins in adult insect cells and their involvement in transcription from the polh promoter[95].

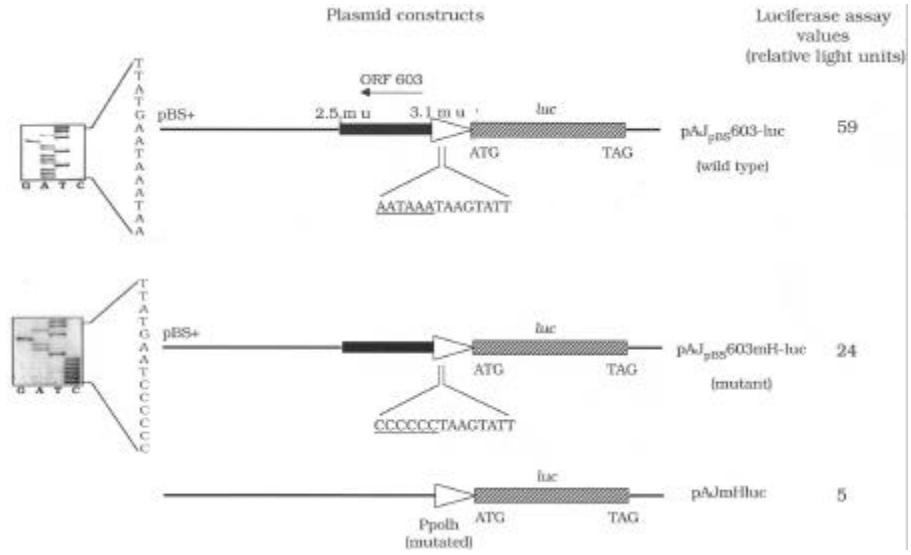


Figure 5. A ~ 766 bp region (containing the ORF603) can compensate for mutations within Ppolh. pAJpBS603-luc contains the wild type Ppolh driving expression of a luciferase reporter, with a ~ 766 bp upstream region. PAJpBS603mH-luc and pAJmHluc are the corresponding plasmids carrying a mutated promoter, with and without the 766 bp region, respectively. The corresponding luciferase values are shown alongside.

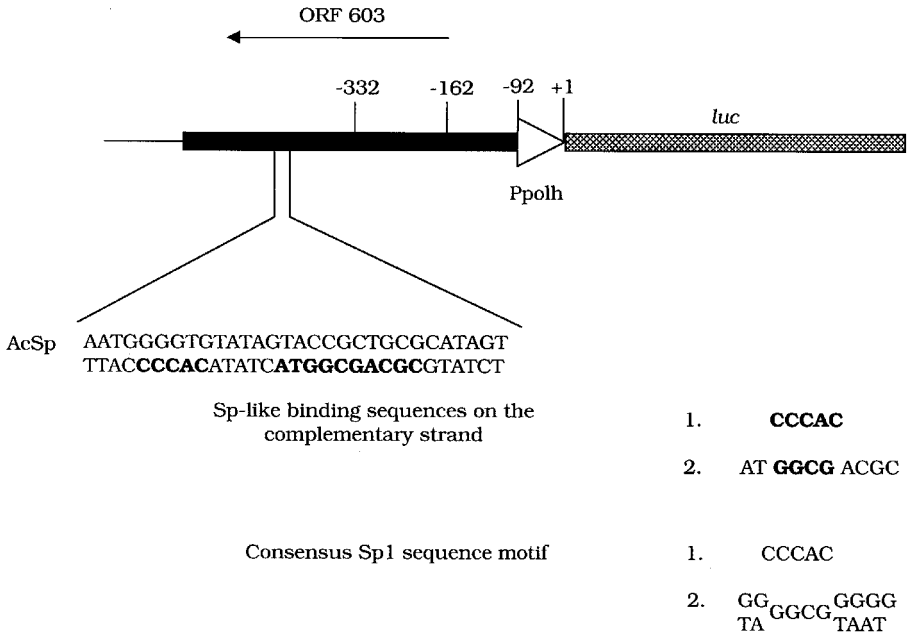


Figure 6. An Sp-like binding sequence (AcSp) is present upstream of Ppolh. The AcSp sequence bears a consensus CACCC motif and a loose GC box, depicted in boldface, on the non-coding strand.

One of the most common regulatory elements present in eukaryotic promoters, enhancers and locus control regions is the GC box (GGGCGG), or the related GT(GGGTGG)/CACCC boxes. The first major advance in our understanding of how these sequences contribute to the control of gene expression was the isolation and identification of the GC-box binding protein, Sp1. As it turned out, Sp1 is simply the first cloned and identified member of a large and still growing family of proteins which bind to similar GC/GT box sequences and share a highly conserved DNA-binding zinc finger domain[96,97]. The superfamily is referred to as the Sp or XKLF (Krüppel-like factor) family, since the zinc finger DNA binding domains of all the members share homology with those found in the *Drosophila melanogaster* regulator protein Krüppel[97]. Currently, the Sp/XKLF family comprises at least 16 different mammalian family members, and is rapidly expanding. Although some Sp-like

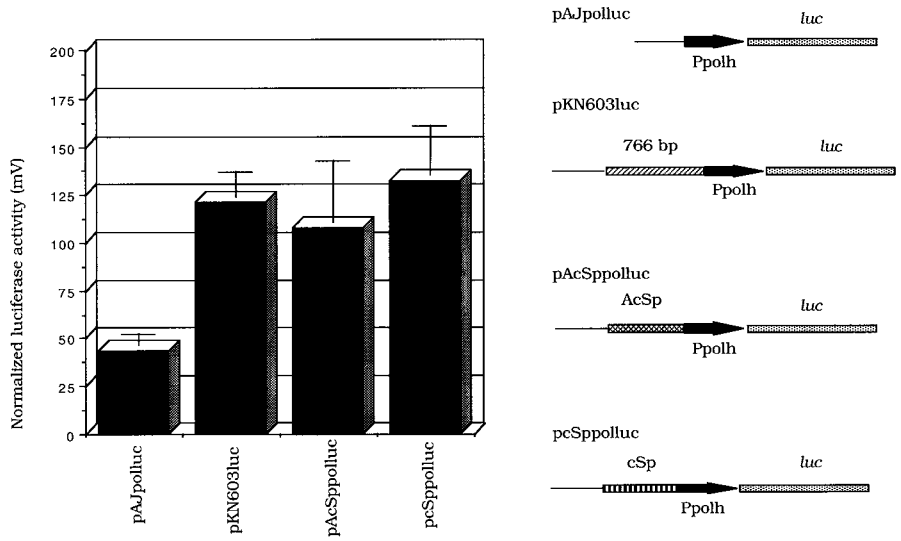


Figure 7. The AcSp and cSp sequence motifs enhance reporter gene expression when placed upstream to the polyhedrin promoter. Luciferase activity in transient expression analyses using pAcSp.pol.luc and pcSp.pol.luc, carrying the AcSp and cSp oligonucleotides respectively, upstream of the polh promoter. The relative luciferase levels of pcSp.pol.luc and pAcSp.pol.luc were compared with those of pAJpolluc (with no upstream sequences) and pKN603luc (carrying ~ 4 kb upstream sequences).

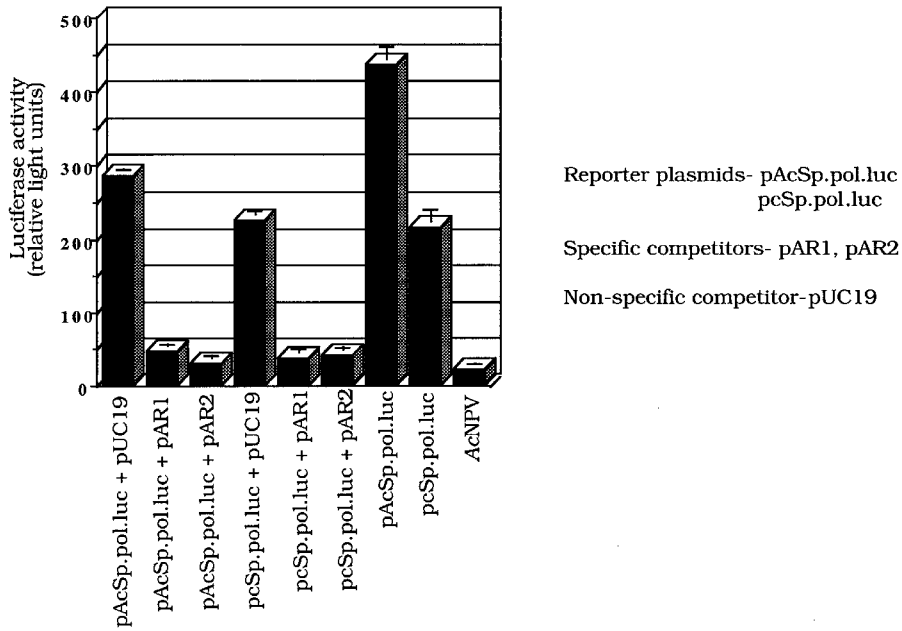


Figure 8. Mopping of the insect Sp-family protein(s) *in vivo* causes a reduction in polh promoter-driven reporter gene expression. Luciferase expression levels using 20 µg of pAcSp.pol.luc (lane 7) and pcSp.pol.luc (lane 8) plasmids after transfection into *Sf*9 cells were compared in the presence of specific or non-specific co-transfected competitor plasmids. Lanes 1 and 4 show luciferase expression using the reporter plasmids pAcSp.pol.luc and pcSp.pol.luc respectively, with pUC19 used as a non-specific competitor. Lanes 2 and 3 depict luciferase expression using pAcSp.pol.luc in the presence of competitor plasmids pAR1 or pAR2 respectively. Likewise, the competition with pAR1 or pAR2 using pcSp.pol.luc as reporter is shown in lanes 5 and 6 respectively. Lane 9 depicts *Ac*NPV infection carried out in the absence of any transfected plasmid. pAR1 and pAR2 plasmids carry the AcSp and cSp oligonucleotides respectively, cloned into pUC19.

proteins have been identified in *Drosophila* embryos only during the blastoderm stage, there are no reports of such factors being present in adult insect tissue.

Our findings show both an enhancement (in the case of an intact promoter) and a rescue of transcription (in the presence of a mutant promoter) with Ppolh upstream

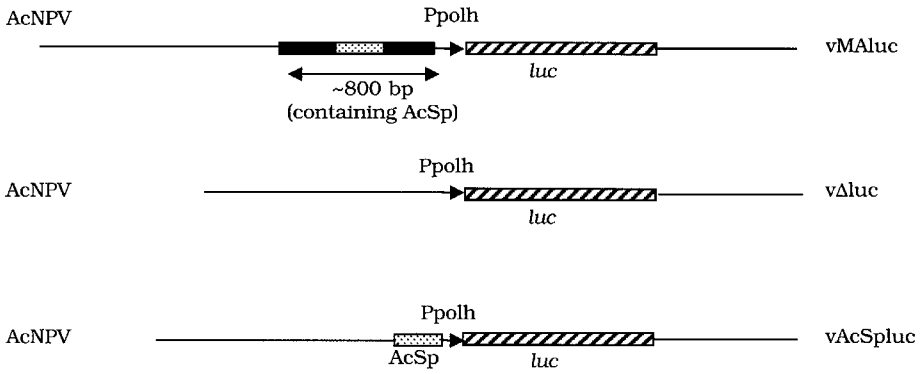


Figure 9 *a*. AcSp is required for enhancement of transcription from the polh promoter in the viral context. Schematic representation of the recombinant baculoviruses vMAluc, vΔluc and vAcSpluc carrying the polyhedrin promoter-driven luciferase gene, with varying sizes of upstream sequences.

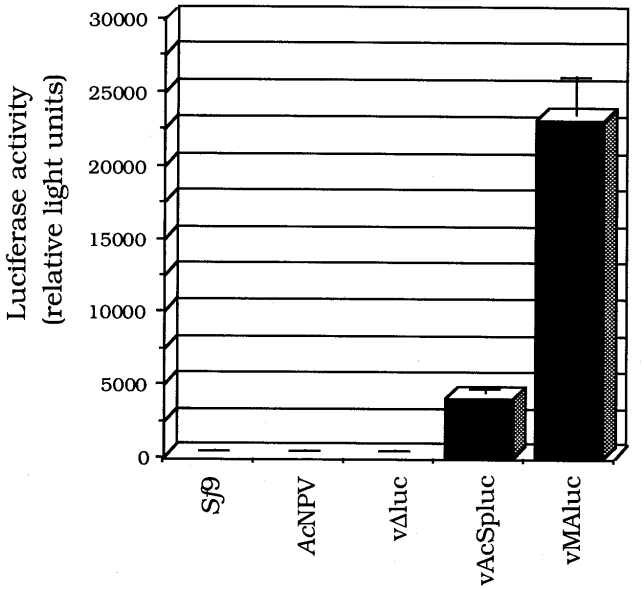


Figure 9 *b*. AcSp is required for enhancement of transcription from the polh promoter in the viral context. Luciferase levels recorded in *Sf* 9 cells, or after infection with *Ac*NPV or recombinant viruses, assayed 65 hpi.

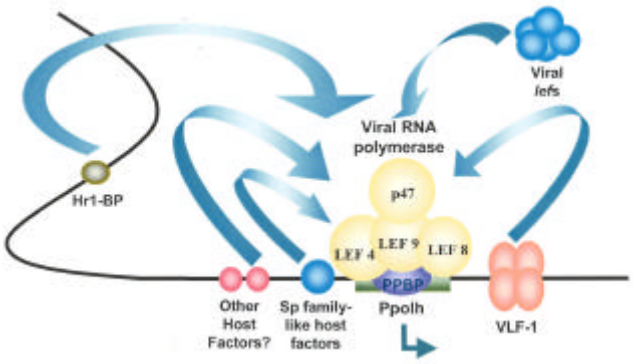


Figure 10. Regulation of transcription from the *Ac*NPV polyhedrin promoter. The hypertranscribed polh promoter recruits viral as well as insect cell host factors to enhance transcription. PPBP, which recruits the viral RNA polymerase to the transcription initiation site, Sp family-like proteins which bind to a promoter-proximal *cis*-sequence, and hr1BP, which interacts with the powerful upstream hr1 enhancer element play critical roles in this process.

regions carrying the AcSp or cSp motifs. Sp1 has been found to activate transcription from both TATA – and Inr – (initiator element) containing promoters. Further, in the bovine papillomavirus E2-responsive promoters, the TATA box or the initiator can be functionally replaced by Sp1 binding sites[98]. These data provided the first suggestion of an interaction between Sp1 and the general transcription machinery, particularly TFIID. Consistent with this observation, it was subsequently demonstrated that the human TBP-associated factor, $hTAF_{II}130$ (ref. 99), and its *Drosophila* homolog, $dTAF_{II}110$ (ref. 100), interact with the glutamine-rich activation domains of human Sp1. Thus, Sp1 is thought to function by recruiting the RNA polymerase complex to promoters via its interaction with TFIID. The most obvious explanation of the enhancement of reporter gene expression by the insect Sp family-like protein(s) is that they interact with the basal transcription machinery directly or indirectly to bring about these effects. Electromobility shift assays using the consensus TFIID oligonucleotide point to a possible interaction between the insect SP-like proteins and TFIID, hinting at a similar mode of action in *Sf* 9 cells too (Ramachandran, A., unpublished observations).

Conclusion

Intensive research has gone into the elucidation of the mechanisms underlying baculovirus transcription regulation. However, the ways of the powerful BEVS remain as enigmatic as ever. An understanding of all the players involved in this process would allow us to recreate *in vitro* the conditions and factors governing polh or p10 promoters transcription, thus permitting the synthesis of foreign proteins to the desired extent while by-passing laborious tissue culture or *in vivo* systems.

In this context, host factors have emerged as a crucial component involved in regulating transcription from the baculovirus very late promoters. In addition to PPBP,

and hrBP, our observations on the presence and involvement of Sp family-like host factor(s) in insect cells is novel. Their involvement in the regulation of a gene so critical for baculovirus survival in the environment adds an important dimension to the complexity of polh promoter-driven transcription. Given that the polh promoter is a TATA-less initiator promoter, coupled with the known involvement of Sp1 in initiator-mediated transcription, these results reveal another facet of the regulation of polyhedrin-initiator transcription.

Studies are in progress to further characterize the Sp-like factor(s) which bind to AcSp and elucidate in greater detail the transcription mechanisms by which they operate. A model of the major trans-acting factor(s) influencing transcription from the polh promoter (Figure 10) thus involves an interplay of host and viral factors. Coupled with structural information and knowledge of the Sp protein(s) and other host factors' cross-talk with various cellular or viral partners, we can expect more pieces of the complex and fascinating jigsaw puzzle of baculovirus gene regulation to fall into place in the near future.

1. Mishra, S., *Curr. Sci*., 1998, **75**, 1015–1022.
2. Kidd, M. and Emery, V. C., *Appl. Biochem. Biotechnol*., 1993, **42**, 137–159.
3. Sridhar, P., Awasthi, A. K., Azim, C. A., Burma, S., Habib, S., Jain, A., Mukherjee, B., Ranjan, A. and Hasnain, S. E., *J. Biosci*., 1994, **19**, 603–614.
4. O'Reilly, D. R., Miller, L. K. and Lucknow, V. A., in *Baculovirus Expression Vector: A Laboratory Manual*, W. H. Freeman and Co, New York, 1992.
5. Hasnain, S. E., Venkaiah, B. and Habib, S., US Patent #08/886,595.
6. Jha, P. K., Nakhai, B., Sridhar, P., Talwar, G. P. and Hasnain, S. E., *FEBS Lett*., 1990, **274**, 23–26.
7. Sridhar, P. and Hasnain, S. E., *Gene*, 1993, **131**, 261–264.
8. Sridhar, P., Panda, A. K., Pal, R., Talwar, G. P. and Hasnain, S. E., *FEBS Lett*., 1993, **315**, 282–286.
9. Mukherjee, B., Burma, S., Talwar, G. P. and Hasnain, S. E., *DNA Cell Biol*., 1995, **14**, 7–14.
10. Ranjan, A. and Hasnain, S. E., *Indian J. Biochem. Biophys*., 1995, **32**, 424–428.
11. Ranjan, A. and Hasnain, S. E., *Virus Genes*, 1995, **9**, 149–153.
12. Shi, Y. P., Hasnain, S. E., Sacci, J. B., Holloway, B. P., Fujioka, H., Kumar, N., Wohlhueter, R., Hoffman, S. L., Collins, W. E. and Lal, A. A., *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 1615–1620.
13. Shi, Y. P., Das, P., Holloway, B., Udhayakumar, V., Tongren, J. E., Candal, F., Biswas, S., Ahmad, R., Hasnain, S. E. and Lal, A. A., *Vaccine*, 2000, **18**, 2902–2914.
14. Hasnain, S. E., Nakhai, B., Ehtesham, N. Z., Sridhar, P., Ranjan, A., Talwar, G. P. and Jha, P. K., *DNA Cell Biol*., 1994, **13**, 275–282.
15. Roy, P., *Ann. NY Acad. Sci*., 1996, **791**, 318–332.
16. Chatterji, U., Ahmad, R., Venkaiah, B. and Hasnain, S. E., *Gene*, 1996, **171**, 209–213.
17. Jha, P. K., Pal, R., Nakhai, B., Sridhar, P. and Hasnain, S. E., *FEBS Lett*., 1992, **310**, 148–152.
18. Tjia, S. T., Zu Altenschildesche, G. M. and Doerfler, W., *Virology*, 1983, **125**, 107–117.
19. Volkman, L. E. and Goldsmith, P. A., *Appl. Environ. Microbiol*., 1983, **45**, 1085–1093.
20. Boyce, F. M. and Bucher, N. L., *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 2348–2352.
21. Hofmann, C., Sandig, V., Jennings, G., Rudolph, M., Schlag, P. and Strauss, M., *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 10099–10103.
22. Ghosh, S., Parvez, M. K., Banerjee, K., Sarin, S. K. and Hasnain, S. E., 2001, manuscript submitted.
23. Gronowski, A. M., Hilbert, D. M., Sheehan, K. C., Garotta, G. and Schreiber, R. D., *J. Virol*., 1999, **73**, 9944–9951.
24. Bashyam, M. D., Ramachandran, A., Ghosh, S., Viswanathan, P. and Hasnain, S. E., *Proc. Indian Natl. Sci. Acad*., 2000, **B66**, 83–96.
25. Carson, D. D., Guarino, L. A. and Summers, M. D., *Virology*, 1988, **162**, 144–151.
26. Carson, D. D., Summers, M. D. and Guarino, L. A., *Virology*, 1991, **182**, 279–286.
27. Guarino, L. A. and Summers, M. D., *J. Virol*., 1986, **57**, 563–571.
28. Guarino, L. A. and Summers, M. D., *J. Virol*., 1987, **61**, 2091–2099.
29. Guarino, L. A. and Summers, M. D., *J. Virol*., 1988, **62**, 463–471.
30. Blissard, G. W. and Rohrmann, G. G., *Annu. Rev. Entomol*., 1990, **35**, 127–155.
31. Chaeychomsri, S., Ikeda, M. and Kobayashi, M., *Virology*, 1995, **206**, 435–447.
32. Huh, N. E. and Weaver, R. F., *J. Gen. Virol*., 1990, **71**, 195–202.
33. Chisholm, G. E. and Henner, D. J., *J. Virol*., 1988, **62**, 3193–3200.
34. Krappa, R. and Knebel-Morsdorf, D., *J. Virol*., 1991, **65**, 805–812.
35. Leisy, D. J., Rasmussen, C., Kim, H.-T. and Rohrmann, G. F., *Virology*, 1995, **208**, 742–752.
36. Lu, A. and Carstens, E. B., *Virology*, 1993, **195**, 710–718.
37. Habib, S. and Hasnain, S. E., *J. Biol. Chem*., 1996, **271**, 28250–28258.
38. Kool, M., Ahrens, C. H., Vlak, J. M. and Rohrmann, G. R., *J. Gen. Virol*., 1995, **76**, 2103–2118.
39. Pearson, M., Bjornson, R., Pearson, G. and Rohrmann, G., *Science*, 1992, **257**, 1382–1384.
40. Habib, S., Pandey, S., Chatterji, U., Burma, S., Ahmad, R., Jain, A. and Hasnain, S. E., *DNA Cell Biol*., 1996, **15**, 737–747.
41. Habib, S. and Hasnain, S. E., *J. Virol*., 2000, **74**, 5182–5189.
42. Leisy, D. J., Rasmussen, C., Owusu, E. O. and Rohrmann, G. F., *J. Virol*., 1997, **71**, 5088–5094.
43. Murges, D., Kremer, A. and Knebel-Morsdorf, D., *J. Gen. Virol*., 1997, **78**, 1507–1510.
44. Kovacs, G. R., Guarino, L. A. and Summers, M. D., *J. Virol*., 1991, **65**, 5281–5288.
45. Yoo, S. and Guarino, L. A., *Virology*, 1994, **202**, 746–753.
46. Yoo, S. and Guarino, L. A., *Virology*, 1994, **202**, 164–172.
47. Knebel-Morsdorf, D., Kremer, A. and Jahnel, F., *J. Virol*., 1993, **67**, 753–758.
48. Thiem, S. M. and Miller, L. K., *J. Virol*., 1989, **63**, 4489–4497.
49. Clem, R. J. and Miller, L. K., *J. Virol*., 1993, **67**, 3730–3788.
50. Hershberger, P. A., Dickson, J. A. and Friesen, P. D., *J. Virol*., 1992, **66**, 5525–5533.
51. Hasnain, S. E., Taneja, T. K., Sah, N. K., Mohan, M., Pathak, N., Sahdev, S., Athar, M. and Begum, R., *J. Biosci*., 1999, **24**, 101–107.

52. Sah, N. K., Taneja, T. K., Pathak, N., Begum, R., Athar, M. and Hasnain, S. E., *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 4838–4843.
53. Kool, M., Ahrens, C., Goldbach, R. W., Rohrmann, G. F. and Vlak, J. M., *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 11212–11216.
54. Lu, A. and Miller, L. K., *J. Virol.*, 1995, **69**, 975–982.
55. Lu, A. and Carstens, E. B., *Virology*, 1991, **181**, 336–347.
56. Croizier, G., Croizier, L., Argaud, O. and Poudevigne, D., *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 48–52.
57. Laufs, S., Lu, A., Arrell, L. K. and Carstens, E. B., *Virology*, 1997, **228**, 98–106.
58. Liu, G. and Carstens, E. B., *Virology*, 1999, **253**, 125–136.
59. McDougal, V. V. and Guarino, L. A., *J. Virol*., 2000, **74**, 5273–5279.
60. McDougal, V. V. and Guarino, L. A., *J. Virol*., 2001, **75**, 7206–7209.
61. Wu, Y. and Carstens, E. B., *Virology*, 1998, **247**, 32–40.
62. Gong, M., Jin, J. and Guarino, L. A., *Virology*, 1998, **244**, 495–503.
63. Guarino, L. A., Xu, B., Jin, J. and Dong, W., *J. Virol*., 1998, **72**, 7985–7991.
64. Beniya, H., Joel, Funk, C., Rohrmann, G. F. and Weaver, R. F., *Virology*, 1996, **216**, 12–19.
65. Glocker, B., Hoopes, R. R. and Rohrmann, G. F., *J. Virol.*, 1993, **67**, 3771–3776.
66. Yang, C. L., Stetler, D. A. and Weaver, R. F., *Virus Res*., 1991, **20**, 251–264.
67. Morris, T. D. and Miller, L. K., *Gene*, 1994, **140**, 147–153.
68. Ooi, B. G., Rankin, C. B. and Miller, L. K., *J. Mol. Biol*., 1989, **210**, 721–736.
69. Weyer, U. and Possee, R. D., *J. Gen. Virol*., 1989, **70**, 203–208.
70. Todd, J. W., Passarelli, A. L., Lu, A. and Miller, L. K., *J. Virol*., 1996, **70**, 2307–2317.
71. Yang, S. and Miller, L. K., *Virology*, 1998, **248**, 131–138.
72. Yang, S. and Miller, L. K., *Virology*, 1998, **245**, 99–109.
73. Yang, S. and Miller, L. K., *J. Virol*., 1999, **73**, 3404–3409.
74. Lu, A. and Miller, L. K., *J. Virol*., 1994, **68**, 6710–6718.
75. Morris, T. D., Todd, J. W., Fisher, B. and Miller, L. K., *Virology*, 1994, **200**, 360–369.
76. Passarelli, A. L. and Miller, L. K., *J. Virol.*, 1994, **68**, 1186–1190.
77. Passarelli, A. L., Todd, J. W. and Miller, L. K., *J. Virol*., 1994, **68**, 4673–4678.
78. Hang, S., Dong, W. and Guarino, L. A., *J. Virol.*, 1995, **69**, 3924–3928.
79. Evans, J. T. and Rohrmann, G. F., *J. Virol*., 1997, **71**, 3574–3579.
80. Evans, J. T., Leisy, D. J. and Rohrmann, G. F., *J. Virol*., 1997, **71**, 3114–3119.
81. Merrington, C. L., Kitts, P. A., King, L. A. and Possee, R. D., *Virology*, 1996, **217**, 338–348.
82. Guarino, L. A., Dong, W., Xu, B., Broussard, D. R., Davis, R. W. and Jarvis, D. L., *J. Virol*., 1992, **66**, 7113–7120.
83. Jin, J., Ding, W. and Guarino, L. A., *J. Virol*., 1998, **72**, 10011–10019.
84. Harwood, S. H., Li, L., Ho, P. S., Preston, A. K. and Rohrmann, G. F., *Virology*, 1998, **250**, 118–134.
85. Rapp, J. C., Wilson, J. A. and Miller, L. K., *Virology*, 1998, **72**, 10197–10206.
86. Hasnain, S. E., Jain, A., Habib, S., Ghosh, S., Chatterji, U., Ramachandran, A., Das, P., Venkaiah, B., Pandey, S., Liang, B., Ranjan, A., Natarajan, K. and Azim, C. A., *Gene*, 1997, **190**, 113–118.
87. Etkin, E., Carp-Weiss, L. and Levi, B. Z., *Virus Res*., 1994, **31**, 343–349.
88. Burma, S., Mukherjee, B., Jain, A., Habib, S. and Hasnain, S. E., *J. Biol. Chem*., 1994, **269**, 2750–2757.
89. Hasnain, S. E., Habib, S., Jain, A., Burma, S. and Mukherjee, B., *Methods Enzymol*., 1996, **274**, 20–32.
90. Ghosh, S., Jain, A., Mukherjee, B., Habib, S. and Hasnain, S. E., *J. Virol*., 1998, **72**, 7484–7493.
91. Mukherjee, B., Burma, S. and Hasnain, S. E., *J. Biol. Chem*., 1995, **270**, 4405–4411.
92. Jain, A. and Hasnain, S. E., *Eur. J. Biochem.*, 1996, **239**, 384–390.
93. Ahrens, C. H., Leisy, D. J. and Rohrmann, G. F., in *Baculovirus DNA Replication* (ed. De Pamphilis, M.), CSHL Press, Cold Spring Harbor, New York, 1996.
94. Possee, R. D. and Rohrmann, G. F., Baculovirus genome organization and evolution, in *The Baculoviruses* (ed. Miller, L. K.), Plenum, New York, 1997, pp. 109–138.
95. Ramachandran, A., Jain, A., Arora, P., Bashyam, M. D., Chatterjee, U., Ghosh, S., Parnaik, V. K. and Hasnain, S. E., *J. Biol. Chem*., 2001, **276**, 23440–23449.
96. Turner, J. and Crossley, M., *Trends Biochem. Sci*., 1999, **24**, 236–241.
97. Philipsen, S. and Suske, G., *Neucleic Acids Res*., 1999, **27**, 2991–3000.
98. Ham, J., Steger, G. and Yaniv, M., *EMBO J*., 1994, **13**, 147–157.
99. Tanese, N., Saluja, D., Vassallo, M. F., Chen, J.-L. and Admon, A., *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 13611–13616.
100. Hoey, T., Weinzierl, R. O., Gill, G., Chen, J.-L., Dynlacht, B. D. and Tjian, R., *Cell*, 1993, **72**, 247–260.

ACKNOWLEDGEMENTS. We thank all members of the Hasnain lab at NII and CDFD for helpful discussions and suggestions. We also acknowledge the Department of Biotechnology, and the Department of Science and Technology, Govt. of India, for generous financial support over the years. We thank the ASBMB, USA for permission to reproduce Figures 4, 5, 8, 9a and 9b from *J. Biol. Chem*.[37,95]; Mary Ann Liebert Inc., USA for reproducing Figure 3a and 3b from *DNA Cell Biol*.[40]; the American Society for Microbiology, USA for reproducing Figure 1 from *J. Virol*.[90]; and Blackwell Science Ltd., UK for reproducing Figure 2 from *Eur. J. Biochem*.[92].

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 278, No. 52, Issue of December 26, pp. 52564–52571, 2003
Printed in U.S.A.

The Homologous Region Sequence (hr1) of *Autographa californica* Multinucleocapsid Polyhedrosis Virus Can Enhance Transcription from Non-baculoviral Promoters in Mammalian Cells*

Received for publication, August 22, 2003, and in revised form, October 14, 2003
Published, JBC Papers in Press, October 21, 2003, DOI 10.1074/jbc.M309351200

Priya Viswanathan‡§, Betapudi Venkaiah¶, M. Senthil Kumar‡§, Sheeba Rasheedi‡§, Sudhanshu Vrati¶, Murali D. Bashyam‡, and Seyed E. Hasnain‡‖**

From the ‡Laboratory of Molecular and Cellular Biology, Centre for DNA Fingerprinting and Diagnostics, Hyderabad 500076, India, the ¶National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi 110067, India, and the ‖Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore 560064, India

The *Autographa californica* multinucleocapsid polyhedrosis virus homologous region sequence hr1 enhances transcription from the viral polyhedrin promoter in *Spodoptera frugiperda* insect cells and independently functions as an origin of replication (*ori*) sequence. The binding of the host nuclear protein, hr1-binding protein (hr1-BP), is crucial for the enhancer activity (Habib, S., Pandey, S., Chatterji, U., Burma, S., Ahmad, R., Jain, A., and Hasnain, S. E. (1996) *DNA Cell Biol.* 15, 737–747 and Habib, S., and Hasnain, S. E. (1996) *J. Biol. Chem.* 271, 28250–28258). We demonstrate that hr1 can also enhance transcription from non-baculoviral promoters like cytomegalovirus and *hsp70* in mammalian cells but does not support *ori* activity in these cells. Unlike insect cells, hr1 can also function in mammalian cells as an enhancer when present in *trans*. hr1 DNA sequence binds with high affinity and specificity to nuclear factors in the mammalian cells. The insect hr1-BP- and the hr1-BP-like proteins from mammalian cells (mhr1-BP) have different properties with respect to ion requirements, DNA groove binding, and molecular size. When mammalian cells are infected with a recombinant baculovirus containing two promoters, the baculovirus polyhedrin and *Drosophila hsp70* gene promoter, the *hsp70* gene promoter alone is active in these cells, and this activity is further enhanced by the presence of an additional hr1 in the recombinant virus. hr1 may thus also have a role in baculovirus-mediated gene delivery in mammalian cells.

Enhancers are genetic elements that up-regulate gene expression by enhancing transcription of genes, even when placed several thousand nucleotides away in a position or orientation-independent fashion. The SV40 enhancer, the first enhancer to be discovered (1, 2) and the most extensively studied, represents the best example of a prototypical enhancer that functions in a wide variety of cell types from different species. It is even active in non-mammalian systems like *Xenopus laevis* kidney cells or even in the green alga *Acetabularia*, and, in some cases, it can confer cell specificity or host-range response to the promoter (3, 4).

The baculovirus *Autographa californica* multinucleocapsid nuclear polyhedrosis virus (AcMNPV)[1] genome consists largely of unique sequences. Interspersed, however, are nine homologous region (hr) sequences called hr1, hr1a, hr2, hr2a, hr3, hr4a, hr4b, hr4c, and hr5 (5, 6). The hrs vary in length from 150 to 800 bp and have a modular structure containing two to eight imperfect, 30-bp palindromes with an EcoRI site at the center of each palindrome. Homologous region 1 (hr1) has been shown to be an enhancer for the immediate early gene *ie-N* and delayed early gene *39k* (7). We earlier reported that hr1 enhances transcription from the polyhedrin promoter in a classic enhancer-like manner (8–10). hr1-mediated enhancement is position- and orientation-independent and can activate transcription to about 10-fold in a plasmid-based transient expression system and to more than 90-fold when inserted as an additional copy within the viral genome (10). Enhancement mediated by hr1 follows the temporal profile of very late gene expression (20–60 h.p.i.) and was a direct function of the number of modules (8). hr1 is bifunctional, *i.e.* it is both a transcriptional enhancer and an *ori* of replication (8), and these functions of hr1 are independent of each other. Although the palindrome alone is sufficient for the *ori* function, the palindrome plus the flanking sequence is essential for the enhancer function. A host cell protein, named hr1-BP (9), which binds very specifically to the palindrome plus the flanking sequence, is required for the enhancer function. We also showed that hr1 can stimulate expression from other homologous promoters like the *Ac*MNPV late promoter *cor* (8) and very late promoter *p10* and from heterologous promoters such as Drosophila *hsp70* in insect cells.[2]

In this study, we describe the ability of the baculovirus hr1 to function as an enhancer, both when placed in *cis* as well as *trans*, but not as an *ori* sequence in heterologous mammalian cells. These cells also contain protein factors that bind to hr1, but these binding proteins are distinct from hr1-BP present in insect cells. This enhancement of gene expression can be exploited for the development of improved vectors for baculovirus-mediated gene delivery in mammalian cells.

* This work was supported in part by core support from the Department of Biotechnology, Government of India to the National Institute of Immunology and Centre for DNA Fingerprinting and Diagnostics. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Recipients of Research Fellowships from the Council of Scientific and Industrial Research, Government of India.

** To whom correspondence should be addressed. Tel.: 91-40-2715-5604; Fax: 91-40-2715-5610; E-mail: ehtesham@cdfd.org.in.

[1] The abbreviations used are: AcMNPV, *A. californica* multinucleocapsid polyhedrosis virus; hr1, homologous region 1; hr1-BP, hr1-binding protein; mhr1-BP, hr1-binding protein from mammalian cells; *polh*, polyhedrin; EMSA, electrophoretic mobility shift assay; h.p.i., hours post infection; CMV, cytomegalovirus.

[2] P. Viswanathan and S. E. Hasnain, unpublished observation.

EXPERIMENTAL PROCEDURES

Cells and Virus—Spodoptera frugiperda cells (*Sf*9) were grown in TNMFH medium (Invitrogen) containing 10% fetal bovine serum as described previously (11). All the mammalian cell lines used namely, Vero (African monkey kidney epithelial cell line), COS-1 (African green monkey kidney fibroblast-like, SV40-transformed cell line), HepG2 (human hepatocyte cell line), HeLa (human adenocarcinoma cell line), and PS-1 (porcine kidney cell line) cell lines were cultured in Dulbecco's modified Eagle's medium containing high glucose. The C6 strain of wild type *Ac*MNPV was used in infection experiments. Recombinant viruses vAclacluc and vAclacluc-hr1 were constructed (10, 11) using pBacPAK8 (Clontech, Palo Alto, CA) baculovirus transfer vectors as per standard procedures. The recombinant viruses were purified and titrated, and viral infection was carried out at a multiplicity of infection of 10 for each virus. To ascertain that equal amounts of viral DNA from the different recombinants had entered the insect cells, equal amounts of the reaction mixtures were dot-blotted onto a nylon membrane after the luciferase assay, followed by probing with the *luc* cDNA and densitometric scanning (8).

Plasmid Constructs—Plasmid pBVluc was constructed by cloning a 3.695-kb fragment from pCEP-X2-luc (containing the CMV promoter, luciferase reporter, and SV40 polyadenylation signal) into the SalI site within the multiple cloning site of pUC18 to generate a 6.38-kb plasmid. pBVluc was linearized by partially digesting with SalI and was ligated to a 750-bp hr1 fragment, released as SalI fragment from pSHhr1 (9), to construct various plasmids with different positions, orientation and number of hr1 (Fig. 1*A*). pBVluc-hr1-U_1 (7.13 kb) and pBVluc-hr1-U_2 (7.13 kb) have hr1 cloned upstream to the CMV promoter in the same or opposite orientation, respectively, as present in the viral genome relative to the polyhedrin promoter. pBVluc-hr1-U_1-U_1 (7.88 kb) and pBVluc-hr1-D_1-D_1 have two copies of hr1 in the correct orientation upstream or downstream to the promoter, respectively. pBVluc-hr1-D_1 and pBVluc-hr1-D_2 have a single hr1 placed downstream to the *CMV-luc* construct in the right or wrong orientation, respectively.

Transient Transfection Assays—Transfection of reporter plasmids into *Sf*9 insect cells was carried out using Lipofectin as described previously (8). Mammalian cell transfections were performed using the reagent LipofectAMINE Plus (Invitrogen). Equal numbers of cells were plated onto 6-well plates. Before transfection, the cells were repeatedly washed with serum-free media to remove all traces of sera. 3 μg of plasmid was diluted in 100 μl of serum-free medium, 6 μl of Plus reagent was added to this, and the DNA-Plus mix was incubated for 15 min. This was then added to another tube containing 4 μl of LipofectAMINE diluted in 100 μl of serum-free medium and incubated together for another 15 min, and the DNA LipofectAMINE complex was added to cells and the volume was made up to 1 ml. After 6 h, 1 ml of 20% fetal bovine serum-containing media was added to the cells followed by infection with either wild type or recombinant viruses or transfected with transfer vector plasmid DNAs. Aliquots of infected and/or transfected cells were harvested at different time points, and luciferase or β-galactosidase enzyme expression levels were monitored as described (8, 11). All the experiments, both transfections and viral infections, were repeated in duplicate, at least three times. To completely rule out artifacts caused by unequal amounts of transfected DNA entering the cells, it was ascertained that equal amounts of plasmid DNA from the different constructs had entered the insect cells. For this, equal amounts of the reaction mixture, after the respective reporter enzyme assays, were dot-blotted on a nylon membrane and probed with the *luc* DNA or *lacZ* DNA as described, followed by densitometric scanning of the dot blots.

Electrophoretic Mobility Shift Assay—Nuclear protein extracts were prepared from *Sf*9 cells or the mammalian cells using a modified method (9, 12). Binding reactions with nuclear extracts from both insect and mammalian cells were carried out by incubating 2 μg of nuclear extract with end-labeled fragments of hr1. The reaction was carried out in a buffer containing 250 mM NaCl, 10 mM Tris-Cl, pH 7.5, 1 mM dithiothreitol, 20% glycerol, and 1 μg (~1000-fold excess) of poly(dI-dC) at 4 °C for 15 min. The reaction was loaded on a 4.5% polyacrylamide gel and fractionated at 150 V in TAE buffer (40 mM Tris-acetate and 1 mM EDTA). In competition experiments, a 20- or 40-fold molar excess of the unlabeled competitor was added to the reaction prior to the addition of the nuclear extract. For studying the effects of various ions on the DNA-protein binding, the salt containing the ion was added to the reaction buffer at the required concentration. To study the inhibition with minor and major groove-binding drugs, the hr1 probe was incubated for 30 min at room temperature with varying concentrations of distamycin A or methyl green and then analyzed for binding by incubation with the nuclear extract.

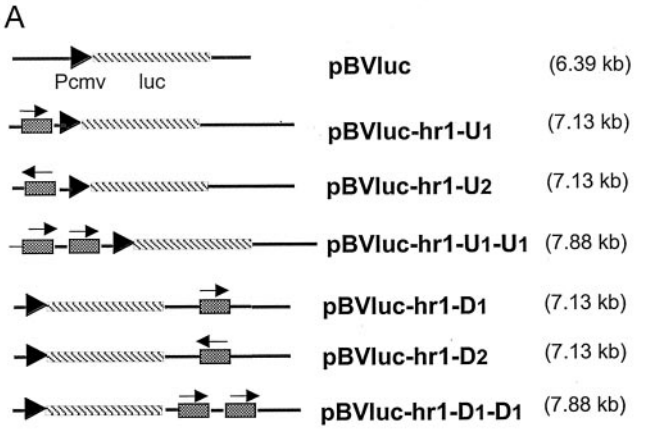
DpnI Sensitivity Assay—Plasmid DNA (2 μg) was transfected into insect and mammalian cells, followed by infection with wild type AcMNPV virus after 24 h or no infection at all as mentioned in the respective figure legends. The total cell DNA was isolated at different time points post transfection or infection as described (8) and resuspended in 25 μl of water. To assay for replication, 5 μl of DNA was digested with HindIII to linearize the plasmid and with both HindIII and DpnI to score for DpnI sensitivity (9). The digested DNA was Southern-transferred and hybridized to ^{32}P-labeled pUC18 probe.

RESULTS

hr1 Enhances Transcription from the CMV Promoter in a Position- and Orientation-independent Manner in Mammalian Cells—hr1 is a classic enhancer of transcription from the AcMNPV promoters in *Sf*9 insect cells. To check if hr1 can enhance CMV promoter-driven expression in mammalian cells, transient transfections were performed. The luciferase expression from Vero cells, transiently transfected with reporter plasmids (Fig. 1*A*) carrying hr1 sequence, was compared with cells transfected with pBVluc lacking the hr1 sequence. The presence of hr1 upstream and in the same orientation relative to the CMV promoter as in the viral genome (pBVluc-hr1-U_1), or upstream but in opposite orientation or downstream in either orientation results in ~3-fold increase in luciferase expression over the control plasmid pBVluc (Fig. 1*B*). Interestingly, constructs carrying two copies of the hr1 sequence in an upstream or downstream position did not generate any additional effect. Transient transfection experiments with the same constructs were performed in other mammalian cells. About similar enhancement of hr-1-mediated luciferase expression was observed in HepG2 cells and PS-1 cells irrespective of the position and orientation of hr1 relative to the CMV promoter (results not shown). When the same CMV promoter-containing plasmids were transfected into *Sf*9 cells, a 2.6-fold enhancement is observed (data not shown). These results suggest that hr1 can function as an enhancer of CMV promoter in both mammalian cells and *Sf*9 cells in a classic enhancer-like fashion.

hr1 Can Enhance Expression from Other Promoters and Reporters in Mammalian Cells—To check if hr1-mediated transcriptional enhancement works with other promoters and reporters, the plasmid pBVlacluc-hr1 was used. This plasmid contains the *hsp70* promoter-driven β-galactosidase, and the baculovirus polyhedrin promoter-driven luciferase along with a copy of the hr1 sequence element. The expression of the *lacZ* reporter from cells transfected with this plasmid was compared with a plasmid pUlacluc that did not carry the hr1 sequence. In Vero cells, about 3-fold increase in β-galactosidase expression in the hr1-containing plasmid over the control plasmid is observed (Fig. 2). The enhancement effected by hr1 is more or less similar irrespective of the constructs used (data not shown). COS-1 cells and HepG2 cells also show hr1-mediated increase in transcription, although to a slightly reduced level compared with Vero cells (data not shown). hr1 can, therefore, enhance expression of β-galactosidase from *hsp70* promoter in different heterologous host cells.

hr1 Can Also Function When Present in trans in Mammalian Cells—Enhancers are basically *cis*-acting DNA elements. In the recent years, there have been various reports of enhancers capable of working also in *trans* (13, 14). Experiments were accordingly designed to check if hr1, unlike in insect cells, can also function in *trans* in mammalian cells. pBVluc (a vector lacking hr1 sequence but with the CMV promoter driving luciferase reporter gene transcription) was co-transfected into Vero cells with different amounts of pSHhr1 (a plasmid containing only the hr1 element cloned in pUC18). pUC18 in appropriate concentration was added to the transfection mix to



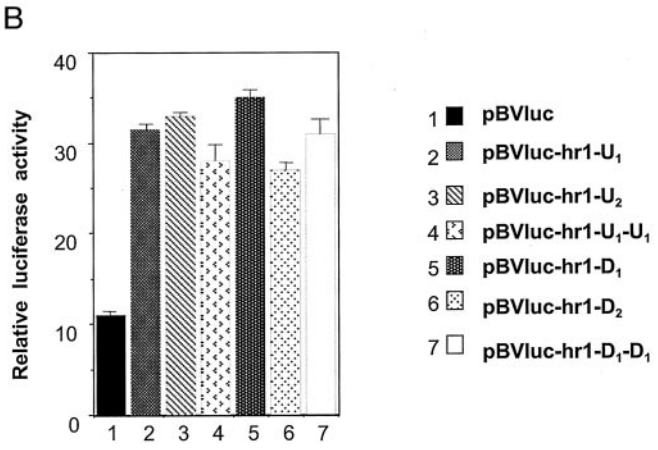


FIG. 1. *A*, a schematic representation of the plasmids used for transfections. pBVluc carries the CMV promoter driving the luciferase reporter. pBVluc-hr1-U_1 has hr1 cloned upstream of the CMV promoter in the same orientation as that in the virus. In pBVluc-hr1-U_2, hr1 is upstream but in the opposite orientation. pBVluc-hr1-U_1-U_1 has two hr sequences upstream, pBVluc-hr1-D_1 has hr1 downstream in the correct orientation, and pBVluc-hr1-D_2 has hr1 downstream but in the opposite orientation. pBVluc-hr1-D_1-D_1 has two hr sequences downstream to the promoter-reporter cassette. The *numbers* in the *parentheses* denote the sizes (in kb) of the respective plasmids. *B*, expression from the CMV promoter in mammalian cells is enhanced by hr1 in transient expression assays. Vero cells were transfected with control plasmid pBVluc (without hr1) and plasmids carrying hr1 placed in different positions and orientation as shown in Fig. 1*A*. The *first bar* represents the luciferase activity of the cells transfected with pBVluc, assayed 36 h post transfection. The *second*, *third*, *fourth*, *fifth*, *sixth*, and *seventh bars* show the luciferase activity after transfection with plasmid pBVluc-hr1-U_1, pBVluc-hr1-U_2, pBVluc-hr1-U_1-U_1, pBVluc-hr1-D_1, pBVluc-hr1-D_2, and pBVluc-hr1-D_1-D_1, respectively.

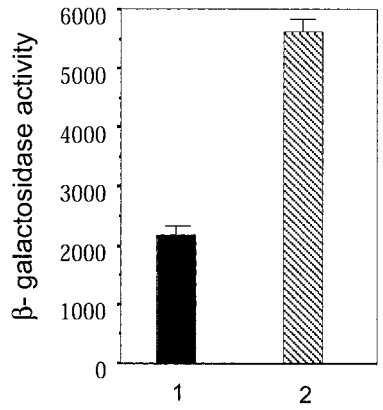
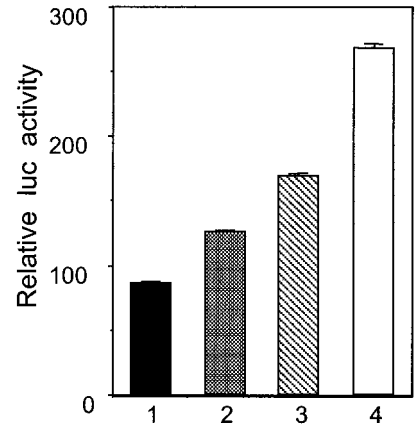


FIG. 2. **hr1 can enhance transcription from *hsp70* promoter in mammalian cells.** Vero cells were transfected with a control plasmid pUClacluc, containing *hsp70* promoter and β-galactosidase reporter (*filled bar*) and pBVlacluc-hr1, which has hr1 cloned downstream of the promoter-reporter (*shaded bar*). The levels of β-galactosidase activity in Vero cells assayed 36 h post transfection are shown.



pBVluc	2	2	2	2
pSHhr1	-	4	8	16

FIG. 3. **hr1 can as well function in *trans* in mammalian cells.** The luciferase activity in Vero cells were assayed 36 h after transfection with 2 μg of pBVluc alone (*1*) or along with 4 μg (*2*) or 8 μg (*3*) or 16 μg of pSHhr1 (*4*).

ensure identical amount of total transfected DNA. When Vero cells were transfected with the same amount of pBVluc and increasing amounts (4, 8, and 16 μg) of pSHhr1, an increase in luciferase expression is observed as a direct function of pSHhr1 concentration (Fig. 3). These results demonstrate that hr1 sequence can act as an enhancer when present in *trans*. Furthermore, this enhancement is a direct function of the concentration of the enhancer-carrying plasmid, pSHhr1, in transiently transfected cells.

hr1-BP-like Factors in the Mammalian Nuclear Extract Interact Specifically with hr1—The complete hr1 is a 750-bp fragment containing five imperfect palindromes (shown as *double-outward arrows* in Fig. 4*A*) with EcoRI site at the center of each palindrome. A complete enhancer module carries an EcoRI palindrome and flanking sequences. Using suitable restriction enzyme sites hr1 was digested into smaller fragments to generate constructs with varying number of palindromes and flanking sequences (Fig. 4*A*). The 170-bp single-enhancer module was used in an electrophoretic mobility shift assay (EMSA) to detect for the presence of hr1-binding proteins in mammalian cells. A distinct retardation in the migration of the labeled 170-bp fragment can be seen when the fragment is incubated with mammalian Vero cell nuclear extract (Fig. 4*B*). The binding of one or more of the Vero cell nuclear factors to hr1 is specific, and this is evident from the observation that the complex can be competed with cold 170-bp hr1 but not with heterologous competitor like pUC18, even when present in high molar excesses. 20- and 40-fold molar excess of cold 170-bp hr1 competitor abolishes the binding (*lanes 3* and *4*), whereas similar molar excesses of pUC18 (a heterologous competitor) has no effect (*lane 5* and *6*), demonstrating the specificity of the binding. 90-, 330-, and 420-bp fragments could also bind with high specificity to the mammalian hr1-binding protein (mhr1-BP) present within the mammalian cell nuclear extract (data not shown). These results demonstrate that factors present in mammalian cells interact with high specificity with the hr1 element.

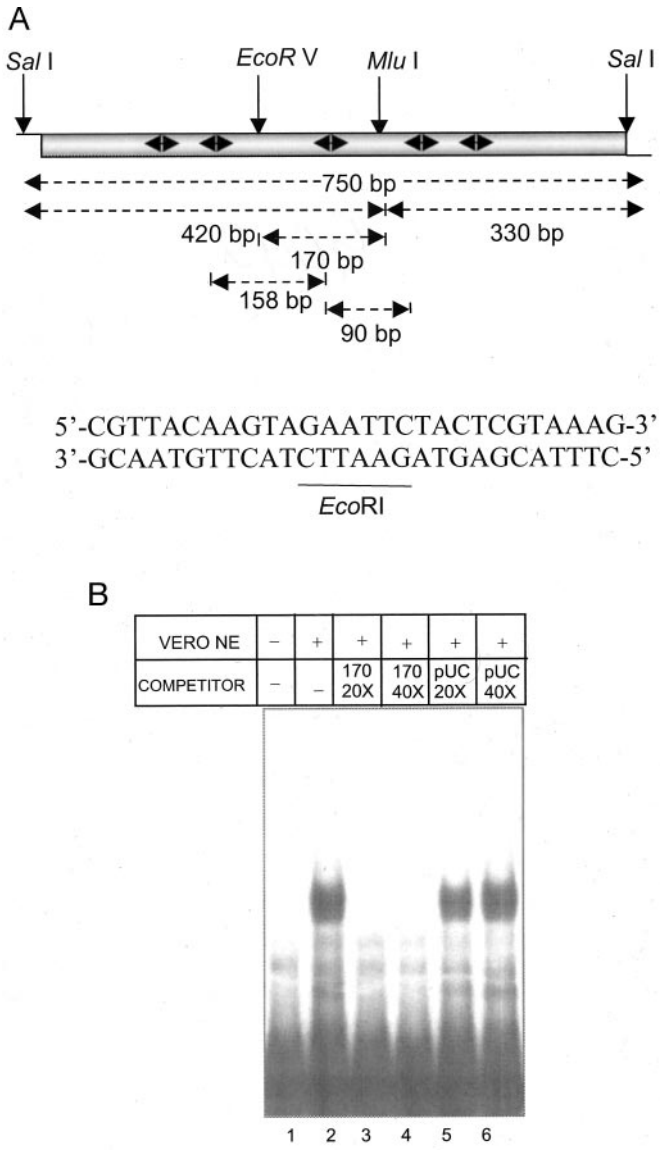


FIG. 4. *A*, a schematic map of the 750-bp hr1. The five 28-bp core palindromes are shown as *filled double arrows*. There is an EcoRI site at the center of each palindrome the sequence of which is shown. Using different restriction enzymes, hr1 can be fragmented to generate different combinations of palindrome and inter-palindrome sequences. *B*, hr1 binds with high affinity to a nuclear factor in mammalian cells. Nuclear extract (2 μg) from Vero cells was incubated with 1 ng of the 170-bp radiolabeled hr1 probe in binding reactions for electrophoretic mobility shift assays. The mobility of free probe is shown in *lane 1*. A distinct retardation in the mobility of the probe is seen when it is incubated with Vero nuclear extract (*lane 2*). 20- and 40-fold molar excess of cold 170-bp hr1 competitor abolishes the binding (*lane 3* and *4*). Similar molar excesses of pUC18 (a heterologous competitor) did not affect the retardation, showing the specificity of the binding (*lane 5* and *6*).

Effect of DNA Binding Drugs on the Enhancer-Nuclear Factor Interaction—Actinomycin D is a drug that binds to the minor groove of the DNA, whereas methyl green binds to the major groove. The 170-bp hr1 probe was incubated alone (Fig. 5, *lane 1*), or with 0.5 mM actinomycin D (*lane 2*), or 0.5 mM methyl green (*lane 8*), or with the addition of Vero cell nuclear extract along with different concentrations of actinomycin D (*lanes 4–7*) or methyl green (*lanes 10–13*). Complex formation was analyzed by gel retardation assay. It could be seen that increasing concentrations of actinomycin D or methyl green causes a reduction of the DNA-protein complex (Fig. 5, shown by an *arrow*), although this effect was more evident at higher concentrations of the drugs. As both the major and minor groove-binding drugs at higher concentrations inhibit the hr1-host factor interaction, it appears likely that the interaction of mhr1-BP with the hr1 DNA overlaps both the grooves.

Divalent Cations Destabilize the Binding of hr1 to hr1-BP, Whereas Monovalent Cations Stabilize the Binding—In an EMSA where the Vero cell nuclear extract and the 170-bp hr1 probe is incubated with increasing concentration (5–125 mM) of $MgCl_2$, a gradual decrease in the binding is apparent (Fig. 6 shown by an *arrow*). At concentrations higher than 50 mM a total loss of binding is seen (Fig. 6). Incubation with monovalent cation KCl generates an opposite pattern. At lower than 125 mM concentration of KCl, binding is very weak, however, at a concentration of 1000 mM, a very tight binding is observed (data not shown). These results demonstrate that, although divalent cations destabilize the hr1-mhr1-BP interactions, monovalent cations stabilize the same.

The One or More Nuclear Factors That Bind to hr1 in Sf9 and Mammalian Cells Are Not the Same—The AcMNPV hr1 binds to the *Sf*9 insect cell host protein, hr1-BP, and this binding is crucial for the enhancer function. Having shown that mammalian cells, even though they are not natural hosts to *Ac*MNPV infection, do contain one or more factors that binds to hr1 we investigated whether these factor(s) is(are) similar to those present in *Sf*9 insect cells. In an EMSA, the 170-bp probe was incubated with 2 μg of *Sf*9 nuclear extract and 2 μg of Vero nuclear extract. The DNA-protein complex in the two cases has different mobilities (Fig. 7, compare *lane 1* with *lane 2*). When 2 μg of the Vero and *Sf*9 extracts are incubated together, independent complexes corresponding to complexes generated by the individual extracts are seen (Fig. 7, *lane 3*). Moreover, when binding reactions were carried out under limiting concentrations of probe and 2 μg each of *Sf*9 and Vero extracts, the results were similar. This implies that the hr1-binding proteins in the two nuclear extracts are of different molecular sizes and perhaps involve different protein-protein interactions, whereas complexing with hr1 and/or bind to different regions within the hr1, hence they do not compete out each other. These results suggest that, although the mammalian cells indeed contain one or more factors that bind to hr1, these factors are different from the *Sf*9 hr1-BP in terms of their molecular sizes and possible cognate sequence motifs and/or protein partners involved in such an interaction.

hr1 Cannot Function as a Replication Origin in Mammalian Cells—In *Sf*9 cells, hr1 also functions as an origin of replication. The *ori* and enhancer activities are independent of each other and are detected using different transfection regimes (8). The *ori* activity is scored by a standard method, which is based on the sensitivity of the isolated plasmid to DpnI restriction digestion, which recognizes the target site to be methylated with Dam methylase (15). *Escherichia coli*, which is Dam^+ will methylate adenine residue within the GATC recognition site for DpnI, and consequently all its DNA will be digested by the enzyme. However, plasmid DNA that is replicated in mammalian cells (Dam^-) will not be methylated at GATC sequence and will therefore be resistant to DpnI digestion. Thus, DpnI can differentiate input plasmid DNA from that which has indeed replicated in the eukaryotic cell. DNA was isolated from mammalian cells transfected with pUC18 and pSHhr1, followed by AcMNPV viral infection after a 24-h gap to ensure the presence of viral factors required for *ori* function. A control where cells were transfected with pSHhr1 but not infected with virus was also included to check if the hr1 *ori* activity can occur even in the absence of viral infection. Total DNA was isolated after different time points (6, 12, 24, and 36 h) post-infection or post-transfection (in case of uninfected cells) and assayed for DpnI digestibility. In the control *Sf*9 cells, a high molecular

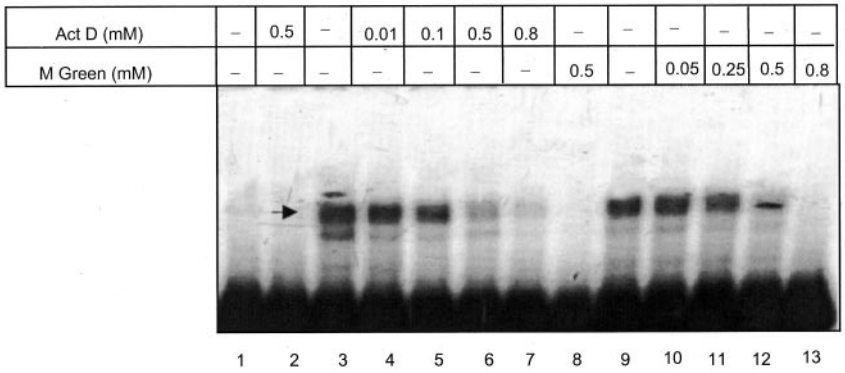


FIG. 5. **Both the major and minor groove binding drugs have an effect on the interaction of hr1 to mhr1-BP.** The 170-bp hr1 probe was incubated alone (*lane 1*), or with 0.5 mM actinomycin D (*lane 2*), or with 0.5 mM methyl green (*lane 8*) or with just the nuclear extract (*lanes 3* and *9*). The probe was incubated with different concentrations of actinomycin D as shown (*lanes 4–7*) or methyl green (*lanes 10–13*), and Vero cell nuclear extract was added subsequently. Complex formation (shown by an *arrow*) was analyzed by gel retardation assay.

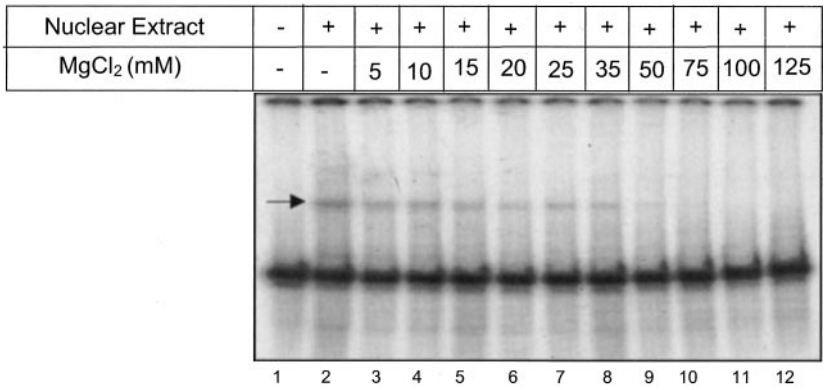


FIG. 6. **Divalent cations destabilize the binding of hr1 to mhr1-BP.** Binding reactions were carried out with Vero nuclear extract and the 170-bp probe at $MgCl_2$ concentrations ranging from 5 to 125 mM (*lanes 3–12*). *Lane 1*, free probe. *Lane 2*, probe with nuclear extract with no additional $MgCl_2$. With increasing amounts of $MgCl_2$ there is a decrease in binding (shown by an *arrow*).

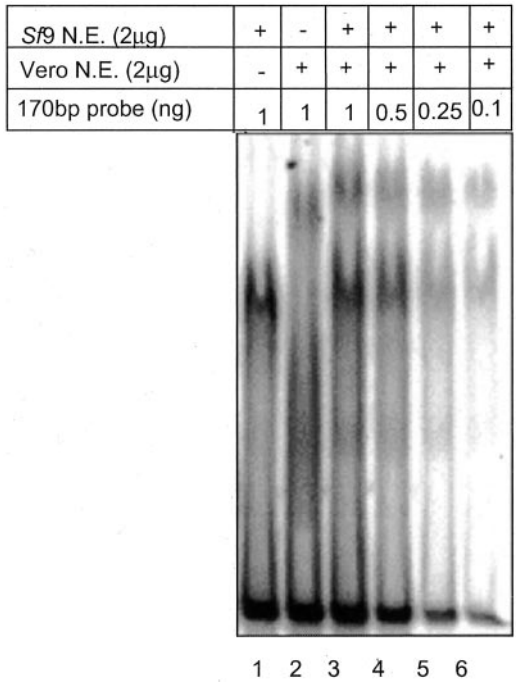


FIG. 7. **The *Sf*9 hr1-BP and mhr1-BP are distinct from each other.** 1 ng of 170-bp hr1 probe was incubated with 2 μg each of *Sf*9 (*lane 1*), Vero (*lane 2*), or both Vero and *Sf*9 extracts (*lane 3*). In competition experiments, 2 μg each of *Sf*9 and Vero extracts were incubated with either 0.5 ng (*lane 4*) or 0.25 ng (*lane 5*) or 0.1 ng (*lane 6*) of 170-bp hr1 probe.

weight DpnI-resistant form is seen in cells transfected with pSHhr1 followed by viral infection (Fig. 8*A*, *lanes 6*, *12*, *18*, and *24*), whereas cells transfected with pSHhr1 but not followed by viral infection do not show any resistant form but were easily digested with DpnI (Fig. 8*A*, *lanes 4*, *10*, *16*, and *22*) pointing to their unreplicated status. Cells transfected with pUC18 followed by viral infection are also sensitive to DpnI (Fig. 8*A*, *lanes 2*, *8*, *14*, and *20*) reflecting their unreplicated status. In contrast, a DpnI-resistant band is not detected at any point in mammalian HepG2 cells, even in cells transfected with pSHhr1 followed by viral infection (Fig. 8*B*, *lanes 6*, *12*, *18*, and *24*). Thus, although in the control *Sf*9 insect cells, the hr1-containing plasmid can support replication in the presence of viral infection hr1 fails to function as an origin of replication in mammalian cells.

A Recombinant Virus Containing Two Promoters Can Infect Mammalian Cells, However, Only the hsp70 Promoter Is Functional Not the Polyhedrin Promoter—Baculoviruses do not replicate in mammalian cells but can enter the cells through a passive endosomal pathway. A recombinant virus containing two promoters and two reporters, luciferase under the control of the baculoviral *polh* promoter and β-galactosidase under the *hsp*70 promoter, was used to infect different mammalian cells and control *Sf*9 cells. After different time points post infection luciferase and β-galactosidase assays were performed. In the control *Sf*9 cells (Fig. 9*A*), luciferase activity follows the typical time profile of *polh* promoter. Luciferase is detectable at 12 h.p.i., increases steadily from 24 to 36 h.p.i., and is very high at 48 h.p.i. In mammalian cells, there is no luciferase activity (data not shown), which is expected given the fact that *polh* promoter requires baculovirus encoded factors. The β-galactosidase activity follows a completely different profile. In *Sf*9 cells this activity, detectable as early as 6 h.p.i., increases only marginally from 12 h.p.i. up to 48 h.p.i. (Fig. 9*B*, *filled bars*). The β-galactosidase activity in mammalian cells is very much lower than in insect cells. In COS-1 cells (Fig. 9*B*, *white bars*) and HepG2 cells (Fig. 9*B*, *shaded bars*), the β-galactosidase activity is detectable at 6 h.p.i., increasing only marginally from 6 to 48 h.p.i. It is interesting to highlight that the observed overwhelming quantitative differences between *Sf*9 insect cells *vis-à-vis* mammalian cells is a combined reflection of the replication of baculovirus in the insect cells as well as increased efficiency of the *hsp*70 promoter. The Drosophila

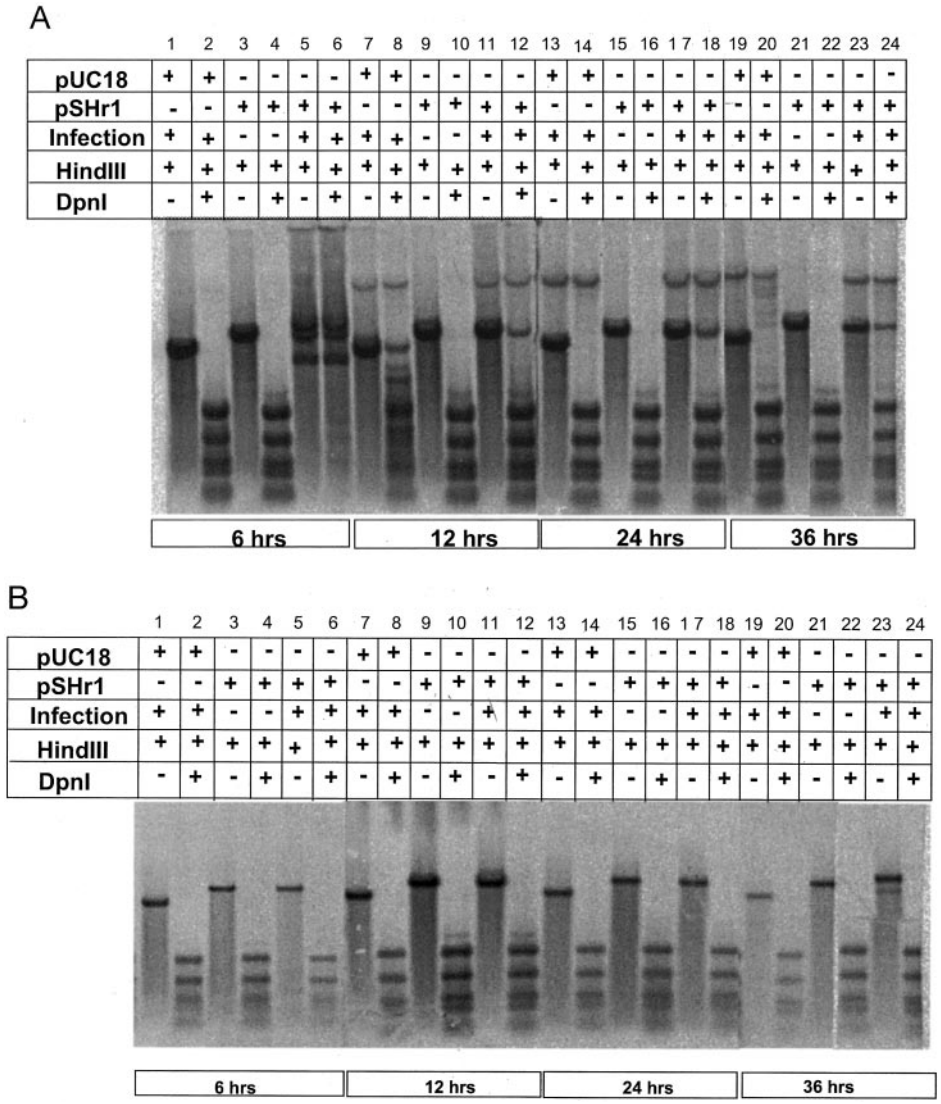


FIG. 8. *A*, hr1 does not function as an *ori* of replication in mammalian cells. The control *Sf*9 cells were transfected with pUC18 followed by infection with AcMNPV (*lanes 1*, *2*, *7*, *8*, *13*, *14*, *19*, and *20*), and total DNA was isolated at different time points (6, 12, 24, and 36 h) post infection. In another experiment, cells were transfected with an hr1-containing plasmid (*pSHhr1*) without any viral infection (*lanes 3*, *4*, *9*, *10*, *15*, *16*, *21*, and *22*) or followed by infection (*lanes 5*, *6*, *11*, *12*, *17*, *18*, *23*, and *24*). The total DNA isolated was further digested with HindIII alone to linearize the plasmid (all *odd lanes*) or digested with both HindIII and DpnI (all *even lanes*) and analyzed for DpnI sensitivity. Lanes where cells were transfected with pSHhr1 followed by infection with virus show the appearance of a DpnI-insensitive band indicating that hr1 works as an *ori* of replication in these cells. *B*, DpnI sensitivity assay to score for hr1 *ori* activity in HepG2 mammalian cells. HepG2 cells were transfected with pUC18 (*lanes 1*, *2*, *7*, *8*, *13*, *14*, *19*, and *20*) followed by infection with *Ac*MNPV. The total DNA was isolated at different time points (6, 12, 24, and 36 h) post infection. In parallel experiment, HepG2 cells were transfected with a hr1 containing plasmid (*pSHhr1*) without any viral infection (*lanes 3*, *4*, *9*, *10*, *15*, *16*, *21*, and *22*) or followed by infection (*lanes 5*, *6*, *11*, *12*, *17*, *18*, *23*, and *24*). The total DNA isolated was further digested with HindIII alone to linearize the plasmid (all *odd lanes*) or digested with both HindIII and DpnI (all *even lanes*) and analyzed for DpnI sensitivity. Absence of any DpnI-insensitive band implies that hr1 cannot function as an *ori* in mammalian cells.

*hsp*70 promoter is expectedly more active in a homologous insect cell environment than in a mammalian cell.

Introduction of an Additional hr1 in the Recombinant Virus Increases the Expression from the hsp70 Promoter in Mammalian Cells—Having shown that the hr1 sequence element can enhance transcription from *hsp*70 promoter in transient transfections in mammalian cells, we checked for hr1-mediated enhancement when it is present as an additional copy in the virus. A recombinant virus (vAclacluc-hr1) was constructed by introducing an additional copy of the hr1 enhancer stably integrated into the vAclacluc (10, 11). More than 2-fold enhancement of expression of β-galactosidase in mammalian cells from *hsp70* promoter was seen when vAclacluc-hr1 was used as compared with vAclacluc (Fig. 10). It is important to note that, unlike *Sf*9 insect cells, mammalian cells do not support baculovirus replication and even then hr1 could enhance expression of a reporter gene driven by a relatively inefficient heterologous insect promoter.

DISCUSSION

The baculovirus AcMNPV hr1, the enhancer under study, is bifunctional, in that it can function both as an enhancer of transcription and an origin of replication. *Spodoptera frugiperda* insect cells (*Sf*9) are the natural hosts for *Ac*MNPV, which follows a temporal profile of gene expression and has been used to express heterologous genes. hr1 can act to enhance the expression of genes during all the different phases of infection; the immediate early gene *ie-N* (or *ie-2*), the delayed early *39k* gene (7, 15, 16), the late gene *cor* and very late *polh* (8). A number of reasons such as the dual role of hr1, its ability to enhance transcription from a wide range of promoters within the host system by recruiting a host factor hr1-BP (9), the

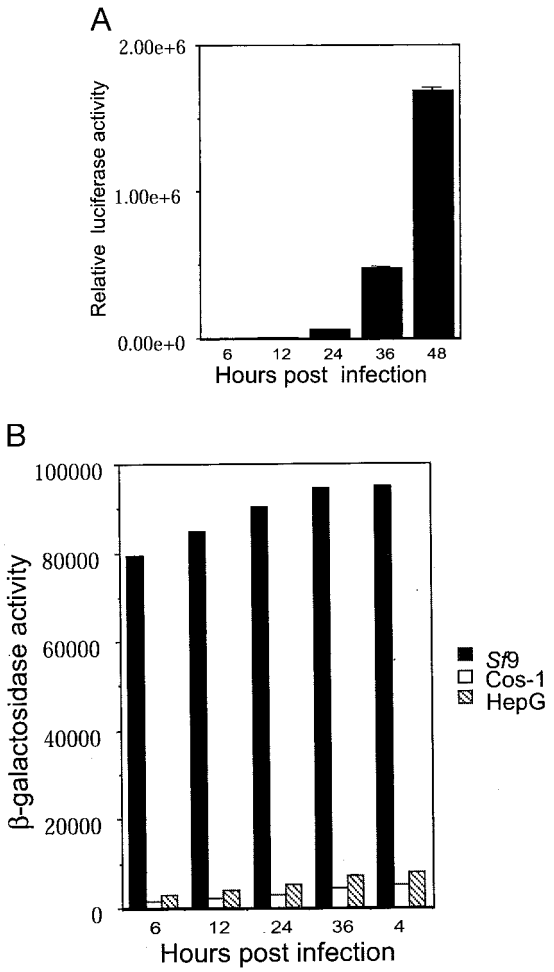


FIG. 9. **Differential reporter gene expression in *Sf*9 and mammalian cells.** Mammalian cells and control *Sf*9 cells were infected with the virus vAclacluc at 10 multiplicity of infection. *A*, luciferase activity of *Sf*9 cells was assayed at different time points post infection 6 h (*bar 1*), 12 h (*bar 2*), 24 h (*bar 3*), 36 h (*bar 4*), and 48 h (*bar 5*). *B*, the β-galactosidase of activity of mammalian cells (COS-1 cells, *white bars*; HepG2 cells, *shaded bars*) and control *Sf*9 cells (*filled bars*) at different h.p.i. is shown.

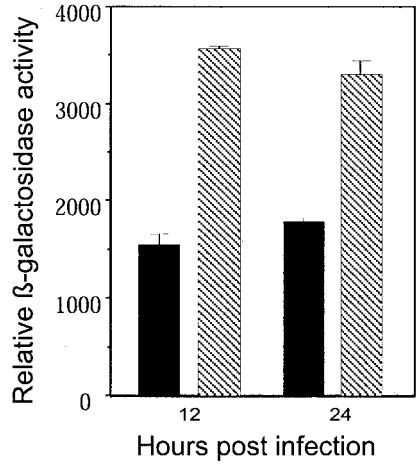


FIG. 10. **Presence of an additional copy of hr1 in the virus enhances the expression of hsp70 promoter-driven β-galactosidase expression in mammalian cells.** Mammalian cells were infected with two recombinant viruses, vAclacluc (*filled bars*) and vAclacluc-hr1 (*shaded bars*) and assayed for β-galactosidase activity at different time points post infection. The virus containing additional hr1 expresses to higher levels than the virus with only one copy of hr1.

utility of baculovirus-mediated gene transfer into mammalian cells, etc., tempted us to analyze the role of hr1 in mammalian cells and the identification of host factors that could interact with hr1. A search for sequence motifs that interact with known enhancer binding proteins and transcription factors in other viral and animal systems revealed that indeed quite a few of these motifs were present in hr1. The most interesting were the motifs with high homology to the consensus enhancer element motif recognized by the C/EBP family of transcription factors in the long terminal repeat enhancer of the avian leukemia virus (17). This motif was repeated (with 1- or 2-bp changes) 3′ to each palindrome in hr1. This motif is incidentally within the region in the flanking sequences, which exhibit high levels of homology.

We have demonstrated that hr1 behaves as a classic enhancer of transcription from the CMV promoter in mammalian cells such as kidney cells (Vero and COS-1), liver cells (HepG2), HeLa and porcine cells (PS-1). Enhancement was also seen when the Drosophila *hsp70* promoter is used. Furthermore, hr1-mediated action was independent of the reporter gene used. The maximum enhancement was observed in Vero cells, where a consistent 3-fold enhancement was observed in hr1-containing plasmids over control non-hr1 plasmids. In *Sf*9 cells, the binding of host factor hr1-BP to hr1 is critical for its enhancer function. Identification of specific nuclear factors present in mammalian cells (mhr1-BP) that bind with high affinity and specificity to different regions of hr1 was consistent with the ability of hr1 to function in non-insect host environment. mhr1-BP and hr1-BP are distinct from each other in terms of their molecular sizes, their ion requirements for binding, and the manner in which they approach the DNA.

Enhancers are usually *cis*-acting DNA sequences that upregulate gene expression, the exact mechanism of which is not clear though several models have been put forward. We have demonstrated for the first time that hr1 is capable of enhancing transcription even in *trans*, *i.e.* when it is present on another plasmid during co-transfections. There are reports that an enhancer can have effect on another DNA molecule when present in *trans* (13, 14). The long range action of enhancers include their ability to stimulate transcription across homologues. In *Drosophila*, these effects are called transvection. In certain combinations the cuticle pigmentation gene, called *yellow*, display intra-allelic complementation, as a result of the action of tissue-specific *yellow* enhancers present on one chromosome stimulating transcription from the *yellow* promoter located on the paired homologue (18). Although hr1 can exert its influence when present both in "*cis*" and "*trans*" in mammalian cells, it can act only in "*cis*" in insect cells. This could be due to the difference in the properties of the enhancer binding proteins in the two cases (*Sf*9 hr1-BP and mhr1-BP). mhr1-BP by virtue of its protein-protein interactions may be able to bring about enhancement of transcription from promoters present on another plasmid molecule.

hr1-BP has little or no role in the *ori* function of hr1 (8, 9), therefore the presence of hr1-BP like factors in mammalian cells should not be of any consequence as far as the *ori* function is concerned. This was indeed the case as was evident from the inability of the hr1 to support replication in mammalian cells. This also explains why baculoviruses can though enter and transiently express foreign genes in mammalian cells, they cannot replicate in these cells. One could speculate that replication is a more host-specific phenomenon than transcription involving hr and non-hr origins of replication (19). Several recent reports have demonstrated that primary human hepatocytes, cell lines of hepatic origin, and other tumor cell lines were able to efficiently take up and transiently express foreign genes under the control of constitutive mammalian promoters

using baculoviruses as delivery vehicles (20–24). High level reporter gene expression from heterologous promoters (8, 21) was observed in human and rabbit hepatocytes *in vitro*, whereas mouse hepatocytes and some other epithelial cell types are targeted at lower rates. The efficiency of baculovirus-mediated entry of foreign genes via an endosomal pathway considerably exceeds that by lipid transfection (20). Another report shows that baculovirus promoters in mammalian cells can be activated by adenovirus functions (25). Baculoviruses, therefore, offer an interesting alternative for a safe, self-limiting gene delivery system for gene therapy (24, 26). These viruses provide an added advantage by their ability to package inserts greater than 10 kb. The dual promoter reporter recombinant virus used here provides an ideal tool to study baculovirus-mediated gene delivery into mammalian cells. The baculovirus polyhedrin gene promoter requires insect cell host factors (27–33) for its activation and therefore cannot function in a non-insect environment. The transcription ability of the constitutive *hsp70* promoter, however, is relatively uniform in different cell lines, except that HepG2 cells show higher levels of reporter activity, consistent with earlier reports that these cells are able to take up baculoviruses more efficiently than other cell lines. The presence of an additional copy of hr1 increases the β-galactosidase expression thereby providing a system for improving the expression of genes delivered via baculovirus in mammalian cells. hr1 therefore can play a major role in enhancing expression of foreign gene in these cells.

Baculoviruses have long been used as biopesticides, owing to their small host range. It is possible now to manipulate the virus so as to increase its host range. Although the ability of the hr1 enhancer (10, 34) to activate transcription and support DNA replication in insect cells can be further exploited (35), the fact that hr1 can positively regulate non-baculoviral promoters both in transient transfections and viral infections without affecting replication makes it an ideal candidate for value addition in baculovirus-mediated gene delivery in mammalian cells (22, 36).

Acknowledgment—We thank all members of the Hasnain laboratory for helpful discussions.

REFERENCES

1. Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M. P., and Chambon, P. (1981) *Nucleic Acids Res.* **9,** 6047–6068
2. Banerji, J., Rusconi, S., and Schaffner, W. (1981) *Cell* **27,** 299–308
3. Laimins, L. A., Khoury, G., Gorman, C., Howard, B., and Gruss, P. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79,** 6453–6457
4. de Villers, J., Olson, L., Tyndall, C., and Schaffner, W. (1982) *Nucleic Acids Res.* **10,** 7965–7976
5. Ayers, M. D., Howard, S. C., Kuzio, J., Lopez-Ferber, M., and Possee, R. D. (1994) *Virology* **202,** 586–605
6. Rasmussen, C., Leisy, D. J., Ho, P. Y., and Rohrmann, G. F. (1996) *Virology* **224,** 235–245
7. Carson, D. D., Summers, M. D., and Guarino, L. A. (1991) *J. Virol.* **65,** 945–951
8. Habib, S., Pandey, S., Chatterji, U., Burma, S., Ahmad, R., Jain, A., and Hasnain, S. E. (1996)) *DNA Cell Biol.* **15,** 737–747
9. Habib, S., and Hasnain, S. E. (1996) *J. Biol. Chem.* **271,** 28250–28258
10. Hasnain, S. E., Venkaiah, B., and Habib, S. (October 12, 1999) U. S. Patent 5,965,393
11. Chatterji, U., Ahmad, R., Venkaiah, B., and Hasnain, S. E. (1996) *Gene* (*Amst.*) **171,** 209–213
12. Hasnain, S. E., Habib, S., Jain, A., Burma, S., and Mukherjee, B. (1996) *Methods Enzymol.* **274,** 20–32
13. Dunaway, M., and Droge, P. (1989) *Nature* **341,** 657–659
14. Chen, J.-L., Huisinga, K. L., Viering, M. M., Ou, S. A., Wu, C.-T., and Geyer, P. K. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 3723–3728
15. Guarino, L. A., Gonzalez, M. A., and Summers, M. D. (1986) *J. Virol.* **60,** 224–229
16. Guarino, L. A., and Summers, M. D. (1986) *J. Virol.* **60,** 215–223
17. Bowers, W. J., and Ruddell, A. (1992) *J. Virol.* **66,** 6578–6586
18. Morris, J. R., Chen, J. L., Geyer, P. K., and Wu, C. T. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 10740–10745
19. Habib, S., and Hasnain, S. E. (2000) *J. Virol.* **74,** 5182–5189
20. Hofmann, C., Sandig, V., Jennings, G., Rudolph, M., Schlag, P., and Strauss, M. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 10099–10103
21. Boyce, F. M., and Bucher, N. L. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 2348–2352
22. Condreay, J. P., Witherspoon, S. M., Clay, W. C., and Kost, T. A. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 127–132
23. Kost, T. A., and Condreay, J. P. (1999) *Curr. Opin. Biotechnol.* **5,** 428–433
24. Ghosh, S., Parvez, M. K., Banerjee, K., Sarin, S. K., and Hasnain, S. E. (2002) *Mol. Ther.* **6,** 5–11
25. Knebel, D., and Doerfler, W. (1987) *Virus Res.* **4,** 317–326
26. Kost, T. A., and Condreay, J. P. (2002) *Trends Biotechnol.* **20,** 173–180
27. Burma, S., Mukherjee, B., Jain, A., Habib, S., and Hasnain, S. E. (1994) *J. Biol. Chem.* **269,** 2750–2757
28. Hasnain, S. E., Jain, A., Habib, S., Ghosh, S., Chatterji, U., Ramachandran, A., Das, P., Venkaiah, B., Pandey, S., Liang, B., Ranjan, A., Natarajan, K., and Azim, C. A. (1997) *Gene* (*Amst.*) **190,** 113–118
29. Jain, A., and Hasnain, S. E. (1996) *Eur. J. Biochem.* **239,** 384–390
30. Mukherjee, B., Burma, S., and Hasnain, S. E. (1995) *J. Biol. Chem.* **270,** 4405–4411
31. Ramachandran, A., Jain, A., Arora, P., Bashyam, M. D., Chatterji, U., Ghosh, S., Parnaik, V. K., and Hasnain, S. E. (2001) *J. Biol. Chem.* **276,** 23440–23449
32. Ghosh, S., Jain, A., Mukherjee, B., Habib, S., and Hasnain, S. E. (1998) *J. Virol.* **72,** 7484–7493
33. Ramachandran, A., Bashyam, M. D., Viswanathan, P., Ghosh, S., Senthil Kumar, M., and Hasnain, S. E. (2001) *Curr. Sci.* **81,** 998–1010
34. Lo, H. R., Chou, C. C., Wu, T. Y., Yuen, J. P., and Chao, Y. C. (2002) *J. Biol. Chem.* **277,** 5256–5264
35. Jarvis, D. L. (2003) *Virology* **310,** 1–7
36. Huser, A., and Hofmann C. (2003) *Am. J. Pharmacogenomics* **3,** 53–63

W318–W320 *Nucleic Acids Research, 2004, Vol. 32, Web Server issue*
DOI: 10.1093/nar/gkh364

PredictRegulon: a web server for the prediction of the regulatory protein binding sites and operons in prokaryote genomes

Sailu Yellaboina[1], Jayashree Seshadri[1], M. Senthil Kumar[2] and Akash Ranjan[1,*]

[1]Computational & Functional Genomics Group and [2]Molecular Oncology Laboratory, Centre for DNA Fingerprinting and Diagnostics, EMBnet India Node, Hyderabad 500076, India

Received February 15, 2004; Revised and Accepted March 4, 2004

ABSTRACT

An interactive web server is developed for predicting the potential binding sites and its target operons for a given regulatory protein in prokaryotic genomes. The program allows users to submit known or experimentally determined binding sites of a regulatory protein as ungapped multiple sequence alignments. It analyses the upstream regions of all genes in a user-selected prokaryote genome and returns the potential binding sites along with the downstream co-regulated genes (operons). The known binding sites of a regulatory protein can also be used to identify its orthologue binding sites in phylogeneticaly related genomes where the *trans*-acting regulator protein and cognate *cis*-acting DNA sequences could be conserved. PredictRegulon can be freely accessed from a link on our world wide web server: http://www.cdfd.org.in/predictregulon/.

INTRODUCTION

With over 100 bacterial genomes sequenced, a key challenge of post-genomic research is to dissect the complex transcription regulatory network which controls the metabolic and physiological process of a cell. A first step towards this goal is to identify the genes within a genome that are controlled by a specific transcription regulatory protein. This paper describes a web server tool—PredictRegulon—for genome-wide prediction of potential binding sites and target operons of a regulatory protein for which few experimentally identified binding sites are known. This technique could utilize the available experimental data on binding sites of transcription regulatory proteins from various bacterial species (1–3) for identification of regulons in phylogenetically related species.

PREDICTREGULON METHOD

The program, PredictRegulon, first constructs the binding site recognition profile based on ungapped multiple sequence alignment of known binding sites. This profile is calculated using Shannon's positional relative entropy approach (4). The positional relative entropy Q_i at position i in a binding site is defined as

$$Q_i = \sum_{b=\mathrm{A,T,G,C}} f_{b,i} \log_{10} \frac{f_{b,i}}{q_b},$$

where b refers to each of the possible bases (A, T, G, C), $f_{b,i}$ is observed frequency of each base at position i and q_b is the frequency of base b in the genome sequence. The contribution of each base to the positional Shannon relative entropy is calculated by multiplying each base frequency by positional relative entropy as follows:

$$W_{b,i} = f_{b,i} \cdot Q_i,$$

where $W_{b,i}$ refers to the weighted Shannon relative entropy of the base b (A, T, G, C) at position i. Finally, a $4 \times \mathrm{L}$ entropy matrix (L is the length of the binding site) is constructed representing the binding site recognition profile, where each matrix element is the weighted positional Shannon relative entropy of a base.

The profile, encoded as the matrix, is used to scan the upstream sequences of all the genes of the user-selected genome. The entropy score of each site is calculated as the sum of the respective positional nucleotide entropy ($W_{b,i}$). A maximally scoring site is selected from the upstream sequence of each gene. The score may represent the strength of interaction between regulatory protein and binding site (5). The lowest score among the input sites is considered as the cut-off score. The sites scoring higher than the the cut-off value are reported as potential binding sites conforming to the consensus profile.

*To whom correspondence should be addressed. Tel: +9140 27171454; Fax: +9140 27155610; Email: akash@cdfd.org.in

Co-directionally transcribed genes downstream of the predicted binding site were selected as potential co-regulated genes (operons) according to one of the following criteria: (i) co-directionally transcribed orthologous gene pairs conserved in at least three genomes (6); (ii) genes belong to the same cluster of orthologous gene function category and the intergenic distance is <200 bp (7); (iii) the first three letters in gene names are identical (the gene names for all the bacterial species were assigned using the COG annotation); (iv) intergenic distance is <90 bp (8).

This method has two specific requirements: a few experimentally determined regulatory protein binding sites should be available for developing the binding site recognition profile, and the profile should be applicable to the genome where the regulator or its homologue is present. In the absence of any experimental information on the regulatory sites in a given genome one may look up the known regulatory motifs from other related species from one of the four online databases which host the information about known transcription regulatory protein binding sites in prokaryote genomes (1–3).

A limitation of this approach is that it may predict a few false positive sites as candidates. However, this limitation can be overcome by experimental validations, by either *in vitro* binding studies with double strand oligonucleotides containing the binding sites (designed based on prediction) and regulatory proteins or real-time PCR analysis of candidate co-regulated genes.

Table 1. Known LexA binding sites of *Bacillus subtilis* from the PRODORIC database

Binding site	Gene
AGAACAAGTGTTCG	*din*C
AGAACTCATGTTCG	*din*B
CGAACTTTAGTTCG	*din*A
CGAATATGCGTTCG	*rec*A
CGAACGTATGTTTG	*din*C
CGAACCTATGTTTG	*din*R
CGAACAAACGTTTC	*din*R
GGAATGTTTGTTCG	*din*R

Table 2. Output of PredictRegulon web server (predicted LexA binding sites)

Score	Position	Site	Gene	Synonym	COG	Product
5.37	−8	CGAACGTATGTTCG	—	Rv3776[a]	—	Hypothetical protein Rv3776
5.32	−100	CGAACATGTGTTCG	—	Rv3073c[a]	COG3189	Uncharacterized conserved protein
5.32	−144	CGAACATGTGTTCG	pyrR	Rv1379[a]	COG2065	Pyrimidine operon attenuation protein
5.22	−8	CGAACACATGTTCG	—	Rv3074[a]	—	Hypothetical protein Rv3074
5.2	−142	CGAACAATTGTTCG	—	Rv3371[a]	—	Hypothetical protein Rv3371
5.2	−64	CGAACAATTGTTCG	dnaE2	Rv3370c[a]	COG0587	DNA polymerase III
5.19	−36	CGAACGATTGTTCG	ruvC	Rv2594c[a]	COG0817	ruvC
5.14	−32	CGAAAGTATGTTCG	—	Rv0336[a]	—	Hypothetical protein Rv0336
5.14	−32	CGAAAGTATGTTCG	—	Rv0515[a]	—	Hypothetical protein Rv0515
5.14	−105	CGAACACATGTTTG	lexA	Rv2720[a]	COG1974	SOS-response transcriptional repressors
5.11	−122	CGAACAGGTGTTCG	recA	Rv2737c[a]	COG1372	recA
5.08	−87	CGAACAATCGTTCG	—	Rv2595[a]	COG2002	Hypothetical protein Rv2595
5.06	−44	CGAATATGCGTTCG	dnaB	Rv0058[a]	COG0305	Replicative DNA helicase
5.04	−263	GGAACTTGTGTTGG	ubiE	Rv3832c	COG2226	Methylase involved in ubiquinone biosynthesis
5.04	−23	AGAACGGTTGTTCG	splB	Rv2578c[a]	COG1533	DNA repair photolyase
5.02	−6	CGAATATGAGTTCG	—	Rv0071[a]	COG3344	Retron-type reverse transcriptase
5.01	−255	CGAACAAGTGTTGG	—	Rv1414	COG3616	Predicted amino acid aldolase or racemase
4.99	−181	GGAACGCGTGTTTG	—	Rv0750	—	Hypothetical protein Rv0750
4.98	−105	CGAACAACAGTTCG	baeS	Rv0600c	COG0642	Signal transduction histidine kinase
4.98	−186	CGAAGATGCGTTCG	rpsT	Rv2412	COG0268	Ribosomal protein S20
4.95	−242	TGAACGCAAGTTCG	fbpB	Rv1886c	COG0627	fbpB
4.95	−192	CGAACGGGAGTTCG	—	Rv1455	—	Hypothetical protein Rv1455
4.94	−270	AGAACCACCGTTCG	phd	Rv3181c	COG4118	Antitoxin of toxin–antitoxin stability system
4.94	−213	CGAACGACGGTTCG	pe	Rv2099c[a]	—	PE
4.92	−118	CGAACAGGTGTTGG	—	Rv0004	COG5512	Zn-ribbon-containing
4.92	−163	CGAACTTGCGTTCA	—	Rv1887	—	Hypothetical protein Rv1887
4.91	−239	GGAACGCGAGTTCG	fadB2	Rv0468	COG1250	3-hydroxyacyl-CoA dehydrogenase
4.91	−7	TGAACGAATGTTCC	—	Rv0039c	—	Hypothetical protein Rv0039c
4.9	−237	CGAAGCCTTGTTCG	dltE	Rv3174	COG0300	Short-chain dehydrogenase
4.89	−225	GGAAGGTGCGTTCG	frnE	Rv2466c	COG2761	Predicted dithiol-disulfide isomerase
4.88	−8	GGAAGCCATGTTCG	—	Rv0769	COG1028	Hypothetical protein Rv0769
4.88	−186	CGAAGAGGTGTTCG	coxS	Rv0374c	COG2080	Aerobic-type carbon monoxide dehydrogenase
4.88	−186	CGAACCGCAGTTCG	leuA	Rv3534c	COG0119	Isopropyl malate/citramalate synthases
4.85	−195	CGAACGGCTGTTGG	—	Rv2061c	COG3576	Hypothetical protein Rv2061c
4.85	−85	AGAACGGTTGTTGG	accA1	Rv2501c	COG4770	COG4770
4.84	−151	CGAAATTGTGTTCC	nuoB	Rv3146	COG0377	NADH:ubiquinone oxidoreductase
4.84	−217	CAAACATGTGTTCG	—	Rv2719c[a]	—	Hypothetical protein Rv2719c
4.84	−5	CGAACATGTATTCG	—	Rv1702c[a]	—	Hypothetical protein Rv1702c
4.84	−199	CGAAATCTTGTTTG	—	Rv1375	COG1944	Hypothetical protein Rv1375

Score: score of the binding sites, Position: position of the binding site relative to the translation start site, Site: binding site of a regulatory protein, Gene: gene downstream to the binding site, Synonym: synonym of the gene, COG: Cluster of Orthologous Gene code, Product: Gene product. [a] represents the ORFs known to be regulated by the regulator. 'a' symbols are not part of the orginal output of the web server. Source of Genome: NCBI ftp site (ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Mycobacterium_tuberculosis_H37Rv/), Accession no. NC_000962.

EXAMPLE: PREDICTION OF LEXA REGULON IN *MYCOBACTERIUM TUBERCULOSIS*

To demonstrate a typical usage of PredictRegulon, we predicted the LexA binding sites and LexA regulon of *M.tuberculosis* using the LexA binding sites of *Bacillus subtilis*. LexA regulators from *B.subtilis* and *M.tuberculosis* share a high sequence identity (45%) at protein level (data not shown). Table 1 lists the known LexA binding sites from *B. subtilis* given as input to the program (2) and Table 2 shows the output of predicted LexA binding sites in *M.tuberculosis*. The site column in Table 2 represents the predicted binding sites of LexA in *M.tuberculosis*. In a typical output the perfect match to the known binding sites and the downstream genes are highlighted with a yellow background, and the rest with score greater than cut-off is shown with a blue background (colours not shown in the table). Eighteen of these genes (indicated by 'a') belonging to the LexA regulon were also observed in data obtained by experimental means by others (9–12). The rest of the matches are potential novel regulatory sites which could be confirmed experimentaly.

The web output of PredictRegulon also contains the hyperlinked gene-synonym and COG number. A click on the former shows the predicted operon context of the regulatory motif while a click on the latter opens a new page showing a description of this gene in the NCBI Conserved Domain Database, which is in turn linked to Pubmed for published information on this gene. These additional links provides users a simple way to browse and understand the functional/physiological implication of the genes that are part of predicted regulon.

ACKNOWLEDGEMENTS

This work is partially supported by the Council of Scientific and Industrial Research (CSIR) NMITLI Grant to A.R. Y.S. and S.K. were recipients of Senior Research Fellowships from CSIR, Govt. of India.

REFERENCES

1. Salgado,H., Santos-Zavaleta,A., Gama-Castro,S., Millan-Zarate,D., Diaz-Peredo,E., Sanchez-Solano,F., Perez-Rueda,E., Bonavides-Martinez,C. and Collado-Vides,J. (2001) RegulonDB (Version 3.2): transcriptional regulation and operon organization in *Escherichia coli* K-12. *Nucleic Acids Res.*, **29**, 72–74.
2. Munch,R., Hiller,K., Barg,H., Heldt,D., Linz,S., Wingender,E. and Jahn,D. (2003) PRODORIC: prokaryotic database of gene regulation. *Nucleic Acids Res.*, **31**, 266–269.
3. Ishii,T., Yoshida,K., Terai,G., Fujita,Y. and Nakai,K. (2001) DBTBS: a database of *Bacillus subtilis* promoters and transcription factors. *Nucleic Acids Res.*, **29**, 278–280.
4. Shannon,C.E. (1948) A mathematical theory of communication. *Bell Sys. Tech. J.*, 379–423 and 623–656.
5. Benos,P.V., Bulyk,M.L. and Stormo,G.D. (2002) Additivity in protein-DNA interactions: how good an approximation is it? *Nucleic Acids Res.*, **30**, 4442–4451.
6. Ermolaeva,M.D., White,O. and Salzberg,S.L. (2001) Prediction of operons in microbial genomes. *Nucleic Acids Res.*, **295**, 1216–1221.
7. Salgado,H., Moreno-Hagelsieb,G., Smith,T.F. and Collado-Vides,J. (2000) Operons in *Escherichia coli*: genomic analyses and predictions *Proc. Natl Acad. Sci., USA*, **97**, 6652–6657.
8. Strong,M., Mallick P., Pellegrini,M., Thompson,M.J. and Eisenberg,D. (2003) Inference of protein function and protein linkages in *Mycobacterium tuberculosis* based on prokaryotic genome organization: a combined computational approach. *Genome Biol.*, **4**, R59.
9. Durbach,S.I., Andersen,S.J. and Mizrahi,V. (1997) SOS induction in mycobacteria: analysis of the DNA-binding activity of a LexA-like repressor and its role in DNA damage induction of the recA gene from *Mycobacterium smegmatis*. *Mol. Microbiol.*, **26**, 643–653.
10. Brooks,P.C., Movahedzadeh,F. and Davis,E.O. (2001) Identification of some DNA damage-inducible genes of *Mycobacterium tuberculosis*: apparent lack of correlation with LexA binding. *J. Bacteriol.*, **183**, 4459–4467.
11. Dullaghan,E.M., Brooks,P.C. and Davis,E.O. (2002) The role of multiple SOS boxes upstream of the *Mycobacterium tuberculosis* lexA gene—identification of a novel DNA-damage-inducible gene. *Microbiology*, **148**, 3609–3615.
12. Boshoff,H.I., Reed,M.B., Barry,C.E. and Mizrahi,V. (2003) DNAE2 polymerase contributes to *in vivo* survival and the emergence of drug resistance in *Mycobacterium tuberculosis*. *Cell*, **113**, 183–193.

J Hum Genet (2003) 48:55–64

MINIREVIEW

Murali D. Bashyam · Gorinabele R. Savithri
Murugapiran S. Kumar · Calambur Narasimhan
Pratibha Nallari

Molecular genetics of familial hypertrophic cardiomyopathy (FHC)

Received: October 15, 2002 / Accepted: November 18, 2002

Abstract Familial hypertrophic cardiomyopathy is an autosomal dominant disease with a wide range of clinical features from benign to severe, and is the most common cause of sudden death in otherwise healthy individuals. The two prominent clinical features are left ventricular hypertrophy and myocyte/myofibrillar disarray. The former is responsible for clinical symptoms such as breathlessness and angina, whereas the latter may lead to sudden cardiac death. The last decade has seen an enormous improvement in our understanding of the molecular genetics of this disorder. The clinical heterogeneity has been linked to genetic heterogeneity; mutations in nine genes encoding sarcomere proteins have been shown to be the molecular basis for the disorder. However, attempts to establish a genotype–phenotype correlation for each of the more than 100 mutations that have been identified have not been highly successful. Additional genetic loci, as well as nongenetic factors such as lifestyle, sex, and age, have also been shown to play a role in modulating the clinical presentation of the disease. How each mutation results in hypertrophy and/or myofibrillar disarray is unclear. The present review discusses the current status of the molecular genetic characterization of this important disorder.

Key words Hypertrophy · Cardiomyopathy · Myosin · Troponin · Tropomyosin · Actin · Myocyte disarray

M.D. Bashyam (✉) · M.S. Kumar
Molecular Oncology, Centre for DNA Fingerprinting and Diagnostics (CDFD), Nacharam, Hyderabad 500076, India
Tel. +91-40-7150008; Fax +91-40-7155610
e-mail: bashyam@www.cdfd.org.in

G.R. Savithri
Laboratory of Molecular and Cellular Biology, Centre for DNA Fingerprinting and Diagnostics (CDFD), Hyderabad, India

C. Narasimhan
Care Hospital, Hyderabad, India

P. Nallari
Department of Genetics, Osmania University, Hyderabad, India

Introduction

Familial hypertrophic cardiomyopathy (FHC) is an autosomal dominant disease characterized mainly by left ventricular hypertrophy. The left ventricular architecture in FHC is disorganized, with hypertrophic cardiac myocytes having abnormal shapes and multiple intercellular connections and being arranged in a chaotic pattern (myocyte disarray). Myocyte loss and replacement fibrosis are prominent features (Davies 1994; Maron et al. 1987), which can result in arrhythmia and altered myocardial hemodynamics. The term "hypertrophy" denotes a thickening of tissue due to an increase in the size of the constituent cells. The term "cardiomyopathy" means a disease of heart muscle (*cardio*, heart; *myo*, muscle; *pathy*, disease). Hypertrophy, in general, is recognized as a compensatory thickening of the left ventricle wall that occurs as a physiologic response to either excessive hemodynamic burden (such as pressure or volume overload due to hypertension and aortic valve disease) or a pathologic state that compromises function of the heart muscle (Grossman et al. 1975). FHC however, is a heritable disorder that occurs in the absence of hemodynamic burden. FHC is the most common cause of sudden death in otherwise healthy young individuals (Maron et al. 1978, 1986). The frequency of hypertrophic cardiomyopathy is approximately 0.1%–0.2% in Europe and the United States. The genetic basis for this disorder has been shown to be mutations in the cardiac sarcomere protein-coding genes. However, it is not clear how these mutations actually induce the hypertrophic response in the heart. Efforts are underway also to establish a genotype–phenotype correlation for this disorder.

Clinical symptoms

The disease is characterized by extreme clinical and morphological heterogeneity, ranging from benign to severe, and is further complicated by a variable age of presentation.

Clinically, the patients may present with some or all of the following symptoms.

Dyspnea (breathlessness)

Dyspnea is usually linked to physical activity but occasionally may manifest even at rest and may be more pronounced following meals. Shortness of breath results from an elevation of left ventricular diastolic pressure due to the thickening of the ventricle wall, and this increased pressure is relayed to the lungs via the pulmonary capillaries, leading to pulmonary congestion characterized clinically by dyspnea.

Chest pain

Also called angina, chest pain is a common symptom. It is usually brought on by exertion and relieved by rest, but may also occur at rest or during sleep and may persist. The greatly thickened muscle demands an increased oxygen supply and chest pain results when this demand cannot be met.

Syncope (fainting)

Patients may experience light-headedness, dizziness, and, more seriously, blackouts. Presyncope denotes dizziness, but the patient does not pass out completely. Episodes may occur in association with exercise, with palpitations, or without any apparent association. Fainting episodes may be due to an irregularity of the heartbeat, or a fall in blood pressure leading to insufficient blood supply to the brain.

Palpitations

Patients may occasionally feel an extra beat or a skipped beat, suggesting an irregular heart rhythm. Palpitations may start suddenly, may appear to be very fast, and may be associated with sweating or light-headedness.

Arrhythmias

Arrhythmias are irregularities of the heartbeat, which occur because of disruption of the electrical conduction system of the heart. Myocyte disarray may be the primary cause for this disruption. Two types of arrhythmias, namely, ventricular tachycardia (arising from the ventricles) and atrial fibrillation (occurring in the atria) are particularly important and may require treatment. In atrial fibrillation, the normal regular rhythm of the heartbeat is lost and is replaced by an irregular rhythm that may be episodic (paroxysmal atrial fibrillation) or persistent. The loss of normal atrial contraction produces a risk of clot formation in the auricle. Sometimes the heart may need to be shocked back into normal rhythm. Heart block may result if the normal electrical signal travels down to the ventricles slowly or is completely blocked. This is uncommon, but if this occurs, a pacemaker may be required. The importance of FHC stems from the fact that a large number of patients may have an increased risk of premature death (due to arrhythmia), which may occur with little or no warning. Ventricular fibrillations usually precede sudden death.

Hypertrophy

The characteristic feature of FHC is the hypertrophied heart. However, there is a high degree of variability in the location and kind of hypertrophy. It could be concentric (spread throughout the left ventricle wall), apical (only the base of the left ventricle), or septal (affecting only the septum). There is an additional complication associated with septal hypertrophy. If hypertrophy occurs in the proximal region of the septum, the inward movement of the hypertrophied septum during systole pulls the anterior leaflet of the mitral valve toward the septum (SAM, systolic anterior motion), causing obstruction in the left ventricular outflow tract. Obstruction may also be a result of the septum thickening itself but is accentuated by SAM. SAM can sometimes be diagnosed by listening to a "murmur" sound in the heartbeat. SAM may also result in flow of blood back into the left atrium from the left ventricle, which is termed "mitral regurgitation."

Diagnosis

Clinically, FHC is diagnosed based on the medical history, physical examination, and electrocardiographic and echocardiographic identification of the septal and left ventricular hypertrophy. The electrocardiogram (ECG) usually shows an abnormal electrical signal due to muscle thickening and disorganization of the muscle structure. In a minority of patients (5%–10%), however, the ECG may be normal or show only minor changes. ECG abnormalities are not specific to FHC and may be encountered in other heart conditions as well. An accurate diagnosis of FHC is therefore based on an ultrasound scan of the heart called an echocardiogram (Echo). The Echo produces a picture of the heart in which excessive thickness of the muscle can be easily measured. Additional equipment called Doppler ultrasound can produce a color image of blood flow within the heart and measure the heart's contraction and filling. Turbulent flow inside the heart can also be detected by this method.

Genetic basis of FHC

Myofibrils, which consist of repeating units known as sarcomeres, make up the contractile elements in muscle. A sarcomere consists of seven major proteins and several minor proteins organized into thick and thin filaments. Thick filaments comprise myosin heavy and light chains and the

Table 1. FHC disease genes and mutation frequencies

Gene name	Gene symbol	Chromosomal location	No. of mutations
Beta-cardiac myosin heavy chain	MYH7	14q12	76
Myosin-binding protein C	MyBPC-3	11p11.2	33
Troponin T	TnT2	1q32	15
Troponin I	TnI3	19q13.4	8
Alpha-tropomyosin	Tpm	15q22.1	6
Myosin essential light chain	MYl3	3p21.2–3p21.3	8
Myosin regulatory light chain	MYl2	12q23–12q24.3	2
Alpha-cardiac actin	ACTC	15q11–q14	5
Titin	TTN	2q24.3	1

Compiled from the FHC mutation database (http://www.angis.org.au/Databases/Heart/heartbreak.html)
FHC, Familial hypertrophic cardiomyopathy

Table 2. Location of major mutations in the *MYH7* gene

Exon	No. of mutations	Functional domain
9	5	ATPase pocket
16	7	Actin binding
19	6	SH3 helix
20	8	Converter
22	6	Tail helix/rod
23	7	Tail helix/rod

myosin-binding protein C. Thin filaments consist predominantly of actin and of lesser amounts of regulatory proteins, namely, tropomyosin and troponin. In the resting muscle, binding of tropomyosin to the active site of actin prevents cross-bridge formation between actin and myosin, thus preventing muscle contraction. The process of contraction is initiated by the release of calcium ions from the sarcoplasmic reticulum of the heart muscle cells. The calcium ions bind to troponin, causing the latter to move the tropomyosin molecule away from the cross-bridge binding sites of actin. Subsequently, adenosine triphosphate (ATP) molecules bound to myosin are hydrolyzed by adenosine triphosphatase (ATPase) to produce an energized myosin molecule and the cross-bridge of the energized myosin molecule binds to actin to trigger the release of the energy stored in the myosin head. Energy release corresponds to the movement of the bound cross-bridge in a rowing motion, bringing the actin molecules toward the center of the sarcomere. An ATP molecule binds to the myosin molecule and releases the cross-bridge attachment; the myosin is now ready to bind to another actin site, as long as calcium is present.

Recent advances in biochemistry and molecular genetics have allowed a better understanding of the molecular basis for FHC. By using linkage analysis and molecular genetic studies, nine disease genes for FHC have so far been identified. These include β-cardiac myosin heavy chain (*MYH7*) (Geisterfer-Lowrance et al. 1990; Watkins et al. 1992), cardiac myosin-binding protein C (*MyBPC-3*) (Watkins et al. 1995a), cardiac troponin T (*TnT2*) (Thierfelder et al. 1994), cardiac troponin I (*TnI3*) (Kimura et al. 1997), α-tropomyosin (*Tpm*) (Watkins et al. 1995b), titin (*TTN*) (Satoh et al. 1999), α-cardiac actin (*ACTC*) (Mogensen et al. 1999), and the essential and regulatory myosin light chains (*Myl3*/*Myl2*) (Poetter et al. 1996) (Table 1). Because all the mutations in FHC patients so far identified are present in the structural proteins of the muscle, the disease has been termed "a disease of the sarcomere" (Thierfelder et al. 1994).

FHC is inherited as an autosomal dominant trait. Thus, individuals with FHC must have one affected parent and frequently have multiple affected relatives. Sporadic occurrence is relatively rare. The variable clinical phenotypes of the disease can in part be explained by the underlying genetic heterogeneity. The genetic heterogeneity is accentuated by the fact that mutations in several exons of the sarcomere genes can cause FHC. To some extent, it has been possible to correlate different mutations with different prognoses and variable echocardiographic findings.

β-Cardiac myosin heavy chain gene

Myosin is the major contractile protein in the heart muscle and is responsible for force generation as a result of its pulling action on actin at the expense of ATP. The first mutation to be identified in FHC was a missense mutation in the β-cardiac myosin heavy chain gene (*MYH7*) (Tanigawa et al. 1990; Geisterfer-Lowrance et al. 1990). Several distinct missense mutations have since been identified in this gene that result in the substitution of conserved amino acids (Watkins et al. 1992; Vikstrom and Leinwand 1996). Because each of the affected amino acids is evolutionarily highly conserved, it has been suggested that these mutations may compromise normal functioning of the protein (Watkins et al. 1992). Several mutations are clustered in the globular head region, and may occur in regions that form functionally important domains (Table 2). A significant fraction of mutations detected so far fall in the tail-helix that lies in the rod region of the protein (Table 2).

Recently, two families were shown to harbor mutations in the light meromyosin domain in the rod region of the protein (Blair et al. 2002). The *MYH7* mutations account for about 30% of the reported FHC cases (Table 1) (Watkins et al. 1995c). To date more than 70 different mutations have been identified in the *MYH7* gene (Table 1). Some are associated with sudden death (p.R403Q), but several also have a benign phenotype (p.V606M) (Epstein et al. 1992; Fananapazir and Epstein 1994).

Characterization of specific MYH7 *mutations*

Although a lot of research has gone into the characterization of several *MYH7* gene mutations, including in vitro studies and animal models, we still know very little about the link between specific mutations, and hypertrophy and myocyte disarray. The most well characterized *MYH7* mutation is p.R403Q. This is primarily because it is associated with sudden death in most cases. Sweeny and coworkers (1994) demonstrated that the p.R403Q mutation resulted in an 80% reduction in the filament sliding speed in vitro, whereas Lankford and coworkers (1996) showed reduced power output in the muscle containing the mutated protein.

Rayment and colleagues (1995) analyzed the effect of 29 mutations in *MYH7* that are implicated in FHC by using the crystal structure of the chicken skeletal myosin S1 peptide. They showed that most of the mutations were clustered around specific sites, viz, the actin-binding interface, the nucleotide-binding pocket, or the interface of the heavy chain with the essential light chain (ELC). Some of the mutations in the actin-binding interface could either affect the velocity of filament sliding, as seen in an in vitro motility assay, or lead to a reduced actin-activated Mg-ATPase activity (Cuda et al. 1993). The mutations in the nucleotide-binding pocket can compromise the catalytic function of the protein. Other mutations that lie in the binding site for the ELC show the importance of the interface between the ELC and the myosin heavy chain because this is required for coupling ATP hydrolysis with movement. However, as more and more mutations are being identified and characterized, it is becoming obvious that the mutations are actually spread throughout the gene.

Sata and Ikebe (1996) expressed and purified mutated forms of the cardiac myosin heavy chain and analyzed their motor and enzymatic properties in vitro. They specifically checked four mutant proteins, all of which showed decreased motor function in terms of ATPase activity and actin translocation. Interestingly, the degree of dysfunction in vitro correlated with the prognosis of the disease resulting from the mutation. The survival values for p.V606M (actin–myosin interface), p.R249Q (ATP pocket), p.R403Q (actin–myosin interface), and p.R453C (ATP pocket) were 94%, 79%, 36%, and 34%, respectively. This correlated with the extent of decrease in myosin motor activity. However, despite differences in prognosis, all mutations resulted in comparable levels of hypertrophy, indicating that other factors may influence cardiac hypertrophy resulting from this mutation. Roopnarine and Leinwand (1998) analyzed the ATPase activities of three mutations (p.R249Q, p.R403Q, and p.V606M). The mutant proteins were expressed and purified and then assayed for their ATPase activities. The p.R249Q mutant (which falls in the base of the ATPase pocket) exhibited a 1.7-fold reduction in maximum Velocity (Vmax). A more severe affect was observed with the p.R403Q mutation (located at the base of a loop that appears to play an important role during the stereospecific actin–myosin interaction). The Vmax decreased 3.5-fold, suggesting that the actin–myosin interaction is weakened substantially. The Vmax of the p.V606M (present in the actin-binding domain) was affected to a lesser degree compared with the wild-type protein. Therefore, these results correlate with the moderate prognosis of the p.R249Q mutation, the poor prognosis of the p.R403Q mutation, and the good prognosis of the p.V606M mutation. A study on a Korean family revealed that the p.G716R mutation may be associated with sudden and early death (Tae-Hong et al. 1998). These results indicate that there may be a somewhat specific genotype–phenotype correlation for the different *MYH7* mutations.

Geisterfer-Lowrance and coworkers (1996) developed a mouse model for the p.R403Q mutation and showed that the clinical features in the mice were similar to those observed in human patients. A transgenic rabbit model (harboring the p.R403Q mutation in the *MYH7* gene) for human hypertrophic cardiomyopathy has also been developed (Marian et al. 1999). The symptoms, including premature death, were similar to those observed in human, thus showing that the rabbit could be a desirable model for studying FHC. The mouse model was also used to understand the basis for the highly malignant phenotype of the p.R403Q mutation (Tyska et al. 2000). The work of Tyska et al. (2000) revealed that the mutant protein actually exhibits a 2.3-fold higher actin-activated ATPase activity, a 2.2-fold greater average force generation, and a 1.6-fold faster actin filament sliding in the motility assay. Therefore, an abnormal power output rather than a compensatory response could lead to hypertrophic response in this mouse model.

Richard and coworkers (1999) showed the presence of double heterozygosity in a French Caribbean family suffering from FHC. They analyzed 15 subjects and reported the presence of a mutation in the *MYH7* gene in exon 15 (p.E483K) in four subjects, a mutation in the *MyBPC3* gene in exon 30 (p.E1096 termination codon) in two subjects, and double heterozygosity in two subjects. As expected, the latter two subjects exhibited a significantly greater left ventricular hypertrophy than did the other affected subjects, although the double mutation was not lethal. Compound heterozygotes for the *MYH7* gene resulting in FHC have also been described (Nishi et al. 1995; Jeschke et al. 1998). Most studies on FHC resulting from mutations in the *MYH7* gene indicate that the mutated protein gets incorporated into the sarcomere and exerts a dominant negative effect, leading to a compensatory hypertrophic response.

Cardiac myosin-binding protein C (*MyBPC3*) gene

Cardiac myosin-binding protein C (*MyBPC3*) is the predominant myosin-binding protein in the heart muscle and, unlike myosin, is not expressed in other tissues (Gautel et al. 1995). It is thought to participate in thick filament assembly by binding myosin and titin, thus probably contributing to sarcomere stability (Schultheiss et al. 1990). The protein also has regulatory activities (Moos and Feng 1990; Schlender and Bean 1991). Unlike *MYH7*, mutations in the *MyBPC3* appear to be in general associated with significant hypertrophy but with good prognosis. The phenotypic expression of mutations occurs late in the life of affected individuals, the reasons for which are not yet clear.

The mutations identified in *MyBPC3* are mostly splice site mutations and, to a lesser extent, exon duplications (Bonne et al. 1995; Watkins et al. 1995a). In a majority of the cases, the mutation results in the formation of a truncated protein, which may lose the ability to bind to the myosin heavy chain (Bonne et al. 1995; Freiburg and Gautel 1996; Okagi et al. 1993). A few missense mutations have also been identified in this gene, including a study on a South-African family (Moolman-Smook et al. 1998). Yu and his colleagues (1998) analyzed two different mutations in two Australian families. One mutation (generation of a premature stop codon instead of glutamine at position 969) resulted in a truncated protein (devoid of the C-terminal 75 amino acids), which was unable to bind to cardiac myosin, and led to an accumulation of disorganized myofibrils resulting in a compensatory hypertrophic response. The second mutation (p.N755K) was a missense mutation located in a highly conserved region of the protein (linker region between the phosphorylation site and the myosin/titin binding region), which resulted in alteration of the function of the protein, leading to cardiac hypertrophy.

Animal models have been used to determine the development and progression of disease resulting from mutations in the *MyBPC3* gene. Transgenic mice expressing mutant forms of the protein exhibited a decreased shortening velocity and power output and significant changes in the ultrastructure of the heart (Yang et al. 1998, 1999). The results also showed that the mutant protein behaved as a poison polypeptide (Yang et al. 1999). The mouse model of a homozygous mutant *MyBPC3* gene was shown to result in dilated cardiomyopathy (McConnell et al. 1999). A knock-in mouse model of the N-terminal deleted gene was created recently (Witt et al. 2001). The left ventricular muscle fibers in the transgenic mice exhibited an increased sensitivity to Ca^{2+}, whereas overall force production was not altered (Witt et al. 2001). This provides a clue as to why the N-terminal mutant proteins lead to a hypercontractile state of the affected muscle fibers.

Troponin T gene

The function of troponin is to move the tropomyosin away, resulting in exposure of the myosin-binding sites on actin. Missense mutations as well as splice site mutations have been identified in the troponin T gene (*TnT2*; Thierfelder et al. 1994; Watkins et al. 1995b). Unlike *MYH7* and *MyBPC3*, mutations in *TnT2* appear to be largely associated with normal ventricular thickness or mild hypertrophy but with a significantly greater frequency of sudden death (Watkins et al. 1995b). A missense mutation (p.R92Q) was tested by Oberst and colleagues (1998) in a transgenic mouse model, and the results once again showed that hypertrophic cardiomyopathy was caused by a dominant-negative effect exerted by the mutant peptide. Watkins and coworkers (1996) studied the IVS15G>A splice site mutation that inactivates a 5′ splice donor site, leading to either skipping of exon 15 or activation of a cryptic splice site. Both aberrant cDNAs encode a truncated TnT2 peptide lacking the conserved C terminus. The analysis of Watkins et al. (1996) revealed that the truncated protein was incorporated into the sarcomere. However, the force generated was less than the sarcomere containing the wild-type full-length TnT2 peptide, indicating that the truncated peptide exerted a dominant-negative effect.

Surprisingly, the study of the p.I91N mutation in rat TnT2 (Lin et al. 1996) showed that the mutated protein exhibited a 50% faster thin filament movement rate. This result is similar to the one obtained by Tyska and coworkers (2000) for the *MYH7* p.R403Q mutation. Another study involving incorporation of mutant proteins in rabbit cardiac myofibrils indicated increased contractility of the cardiac muscle owing either to Ca^{2+} sensitization or potentiation of the maximum level of ATPase activity (Yanaga et al. 1999). Sweeney and coworkers (1998) suggested an alternative mechanism for the disease (distinct from the dominant-negative mechanism). Their work on two mutations, viz., p.I91N and p.R92Q, revealed that these mutations might lead to an increase in the cost of force production (less force generated per ATP molecule). However, the total reduction in the force generation may be minimal, thus explaining the moderate to nil hypertrophy seen in patients having mutations in *TnT2*. Thus, even in the absence of hypertrophy, increased cardiac output could result in energy demands that cannot be met, leading to arrhythmias and sudden death. Montgomery and coworkers (2001) have shown that the extent of mutation-induced dysfunction depends not only on the nature of the *TnT2* mutation, but also on the concentration of the mutant protein in the sarcomere. A study of the p.S179F mutation in exon 11 revealed that it results in sudden death in affected individuals when present in a homozygous state (Ho et al. 2000). Sehnert and colleagues (2002) made an important observation regarding the coexpression of different thin-filament constituents. By using the zebra fish model, they showed that TnT2 null mutants resulted in a significant reduction in levels of alpha tropomyosin and troponin I (Sehnert et al. 2002). The p.R141W *TnT2* mutant has been shown to cause dilated cardiomyopathy (Li et al. 2001). Interestingly, it has been shown that the altered thin-filament regulation caused by mutant TnT2, which causes dilated cardiomyopathy, is different from that caused by mutants that result in hypertrophic cardiomyopathy (Robinson et al. 2002). Recently, an interesting observation of coexistence of Friedreich's ataxia triplet repeat expansion and cardiac troponin T

mutation has been reported in a 5-year-old boy (Cuda et al. 2002).

Troponin I (*TnI3*) gene

The TnI3 is the inhibitory subunit of the troponin complex. The inhibition is released when calcium binds to the troponin C subunit. Several mutations leading to FHC have been identified in this gene, the majority of which are missense mutations. Three mutations lie in the inhibitory region, whereas the rest are located in the C-terminal region. Some of these mutations may result in apical hypertrophy (Kimura et al. 1997). One of them, viz., p.R145G, was studied in a mouse model. The results showed cardiomyocyte disarray, interstitial fibrosis, and premature death (James et al. 2000). The functional alterations that seemed to be responsible for the development of cardiac disease included increased skinned fiber sensitivity to calcium and, at the whole organ level, hypercontractility with diastolic dysfunction (James et al. 2000). It has been shown that the p.R145G and p.R162W mutations of the *TnI* gene, which have been associated with FHC, may cause the disease via impaired relaxation rather than impaired contraction, as seen with some other classes of mutants (Elliott et al. 2000).

Takahashi-Yanaga and colleagues (2001) studied the effect of several *TnI3* mutations. Their results demonstrated that most of the FHC-linked cTnI mutations did affect the regulatory processes involving the TnI3 molecule, and that at least five mutations (p.R145G, p.R145Q, p.R162W, K183del, and p.K206Q) increased the Ca^{2+} sensitivity of cardiac muscle contraction (Takahashi-Yanaga et al. 2001). The R145G mutation has been shown to result in severe diastolic dysfunction and somewhat decreased contractility leading to hypertrophy as a compensatory mechanism (Lang et al. 2002). The K183del is the most common mutation found in this gene. Kokado and colleagues (2000) showed in a study of 25 individuals from seven families that this mutation resulted in variable clinical presentation (Kokado et al. 2000), once again pointing to the role of other factors in the prognosis of the disease. Recently, the lysine 183 deletion was shown to lead to septal wall thinning and systolic dysfunction (Shimizu et al. 2002). Studies on this mutation have also shown generation of abnormal Q waves in patients, which could be an indication for sudden cardiac death (Shimizu et al. 2002). Burton and coworkers (2002) have shown that two mutations, viz., p.R145G and p.G203S, exhibited different effects on motility and ATPase activity. They set up in vitro assays using recombinant mutant proteins expressed in *Escherichia coli* that revealed that the p.G203S mutant exhibited reduced inhibition in the motility assay. However, the action of p.R145G was indistinguishable from that of the wild-type protein, indicating that this mutant protein led to hypertrophic cardiomyopathy without affecting its function in force production.

α-Tropomyosin (*Tpm*) gene

The function of tropomyosin is to cover the myosin-binding sites on actin. Exon 5 and exon 2 of the gene are involved in a majority of the missense mutations. A study of about 60 Japanese families revealed that the clinical phenotypes of different mutants are similar, reflecting a common mechanism by which dysfunctional thin filaments trigger cellular hypertrophy (Yamauchi-Takihara et al. 1996). Investigators showed that four different α-tropomyosin mutants resulted in an inappropriate increased force output at submaximal levels of calcium. Michele and coworkers (1999) have shown that the severity of a direct calcium-sensitizing effect of mutations in α-tropomyosin correlated with the extent of severity of FHC (Michele et al. 1999). Recently, Prabhakar and colleagues (2001) studied a mouse model for the Q180G mutation that exhibited a very severe phenotype of the disease. A novel deletion mutation in the α-tropomyosin gene has been identified in a Japanese patient with hypertrophic cardiomyopathy (Nakajima-Taniguchi et al. 1995). Some mutant forms of α-tropomyosin may also result in dilated cardiomyopathy (Olson et al. 2001).

Essential and regulatory myosin light chain genes (*Myl3/Myl2*)

Poetter and colleagues (1996) hypothesized that, because mutations in the ELC-binding region of the myosin heavy chain have been implicated in FHC, it was possible that mutations in the ELC itself could also trigger FHC. Therefore, they amplified the seven exons of the human cardiac/skeletal ELC gene separately and showed the presence of a mutation (p.M149V) in 1 of the 383 families screened. They showed that the rate of actin translocation in myosin isolated from these patients was higher compared with normal. Similar results were obtained when a mutated myosin heavy chain (p.R719Q), which lies adjacent to the ELC mutation, was used in the assay. The cardiac phenotype in these patients was a rare subtype of FHC, which involved massive hypertrophy of the cardiac papillary muscles, causing a midcavity obstruction. They also screened 399 unrelated FHC cases for mutations in the adjacent myosin regulatory light chain (RLC) and found three independent mutations (p.A3T, p.E22K, p.P94R). The cardiac morphology in these patients was strikingly similar to the patients with mutations in the ELC gene. These studies show that mutation in those regions of ELC and RLC that may be involved in binding to the myosin heavy chain may lead to FHC.

α-Cardiac actin (*ACTC*) gene

Actin is the major constituent of thin filaments and is directly involved in force generation. Mogensen and coworkers (1999) have identified α-cardiac actin as a novel disease gene for FHC. The gene has a p.A295S mutation in exon 5, which results in impaired myosin binding by the protein, leading to impaired force generation and, hence, compensatory hypertrophy. Mutations in this gene can also cause inherited idiopathic dilated cardiomyopathy (IDC) (Takai et al. 1999). Mogensen and coworkers (1999) have shown in their study that mutations affecting sarcomere contraction

led to FHC, whereas mutations affecting force transmission from the sarcomere to the surrounding syncytium led to IDC.

Titin

The titin gene is perhaps the second-largest human gene known to date and is an important component of the cardiac sarcomere. Satoh and coworkers (1999) identified *titin* as a candidate FHC gene based on analysis of 82 FHC patients who did not harbor mutation in any of the other known genes. The p.R740L mutation resulted in an increase in the binding affinity of titin for alpha-actin in the yeast two-hybrid system. These observations suggest that the titin mutation may cause hypertrophic cardiomyopathy in this patient via altered affinity to alpha-actin.

Evidence for involvement of multiple factors

Although FHC is predominantly recognized as a unigene disorder, the overall clinical symptoms may be modulated by other factors distinct from the sarcomeric mutations. In other words, patients with identical mutations may sometimes have different clinical features, indicating factors such as modifier genes, sex, physical activity, nutrition, ethnic background, and other environmental factors may affect the phenotypic expression of FHC. The most obvious factor is the extent of physical activity in the patient (Maron et al. 1998). In a few cases, the disease has been observed to have a somewhat sex-specific phenotype, resulting from a variation in the development of hypertrophy and progression to heart failure. However, there does not appear to be a contrasting difference between the two sexes. Our studies on one Indian family, which harbors a mutation in the 22nd exon of the *MYH7* gene, revealed clinical heterogeneity among various affected members of the same family (M.D. Bashyam et al. 2002, unpublished observations). Several studies have revealed the involvement of other genetic loci in the prognosis of FHC. The most important among these appear to be the genes that constitute the renin-angiotensin system. The angiotensin-converting enzyme gene (*ACE*) has been shown specifically to affect the prognosis in some cases (Lechin et al. 1995; Marian et al. 1993). This issue has also been addressed by Ortlepp and coworkers (2002), who showed that penetrance of a myosin-binding protein C mutation was dependent on a genetic polymorphism associated with the renin-angiotensin system. The DD allele of the ACE I gene has been shown to be associated with a severe form of hypertrophy and sudden death in patients with FHC (Marian et al. 1993; Iwai et al. 1994). It is possible that the extent of the contribution of this allele may be different in different mutations. Tesson and colleagues (1997) established an association of the *Ace* I D allele with the *MYH7* p.R403Q mutation, but not with the *MyBPC3* mutations, raising the possibility of a role for additional genetic modifiers. Recently, Semsarian and coworkers (2001) used a mouse model to address the question of other genetic markers that may influence phenotypic expression of the disease.

Interesting posers

The wide spectrum of genetic heterogeneity makes it essential to use a high-throughput screening technology for mutation detection. Recently, microarray technology was adapted for this purpose with good results (Waldmuller et al. 2002). As mentioned earlier, the clinical diversity of the disease can in some cases be explained by the underlying genetic heterogeneity. However, it is difficult at this time to state with any degree of certainty that a particular mutation will result in a specific phenotype. However, some definite patterns have emerged over the past few years pertaining to mutations in the *MYH7*, *TnT2*, and *MyBPC3* genes, probably because mutations in these genes are more common and consequently more research has been carried out on them. For *MYH7*, the prognosis appears to be different for different mutations. For example, the p.R403Q mutation is usually associated with a severe phenotype, leading to sudden death in most cases, whereas the p.V606M mutation is usually more benign. Most mutations in the *TnT2* gene are associated with low penetrance and subclinical hypertrophy but a very high incidence of sudden death. Mutations in the *MyBPC3* gene are usually associated with a better prognosis and a delayed onset of symptoms. Recently, Niimura and colleagues (2002) showed that hypertrophic cardiomyopathy in the elderly was caused mainly by mutations in the cardiac myosin-binding protein C, troponin I, or the alpha-cardiac myosin heavy chain genes, and was not due to genes such as beta-cardiac myosin heavy chain, troponin T, and alpha tropomyosin.

There are several interesting aspects of the disease that are worth investigating. For example, the molecular basis for the hypertrophic response is not clear. The link between the mutations and different kinds of hypertrophic response also remains to be elucidated. Numerous studies indicate that Ca^{2+} could be a primary signal for cardiac hypertrophy, not only because the release of calcium is the primary step in muscle contraction, but also because hypertrophic agonists are known to activate calcium-dependent signaling. It has been shown that a calcium-regulated transcription factor, nuclear factor of activated T cells (NFAT) interacts with the cardiac-specific transcription factor involved in the hypertrophic response, viz., GATA4. This suggests a possible link between hypertrophy and regulation of calcium levels (Reddy 1997). Experimental induction of cardiac hypertrophy (such as by using mechanical stress) has been shown to induce intracellular signaling pathways involving the Ras-dependent mitogen-activated protein kinase cascade (Sadoshima and Izumo 1993). The role of c-H-ras and c-myc in the hypertrophic response has also been studied. c-H-ras expression is normal throughout cardiac development and after birth (Komuro et al. 1988), whereas c-myc is expressed only during embryonic stages and is undetectable after birth (Simpson 1988). Recent studies

have indicated that c-H-ras expression is up-regulated during hypertrophy, and c-myc-positive patients show greater myocyte hypertrophy than do c-myc-negative patients (Kai et al. 1998). Therefore, the induction of these two genes may be an important step leading to hypertrophy. Calcium-calmodulin activated phosphatase-Calcineurin has been implicated in playing a major role in cardiac hypertrophy (Wilkins and Molkentin 2002). Several other growth factors, such as c-fos, c-jun, atrial and brain natriuretic peptides, and endothelin I, have been shown to be possibly involved in not only pressure overload hypertrophy but also in hypertrophy resulting from sarcomere gene mutations (Kai et al. 1998; Derchi et al. 1992; Hasegawa et al. 1993, 1996). Adult cardiomyocytes in FHC have been shown to reactivate fetal isoforms and to down-regulate adult forms of several cardiac genes, including the cardiac myosin heavy chain gene (Komuro et al. 1988; Chien et al. 1993). However, the cascade of events leading to this switch is yet to be determined.

Given the underlying genetic heterogeneity, it is possible that other disease genes for FHC may be identified in the future. This may actually further complicate attempts to establish a genotype–phenotype correlation. Therefore, it is important to include clinical data for each mutation identified so far in existing FHC mutation databases to facilitate such a correlation. A study of FHC in relation to its epidemiology and nongenetic factors may give a clue as to the etiology and help in better prognosis and preclinical and presymptomatic diagnosis.

References

Blair E, Redwood C, de Jesus Oliveira M, Moolman-Smook JC, Brink P, Corfield VA, Ostman-Smith I, Watkins H (2002) Mutations of the light meromyosin domain of the beta-myosin heavy chain rod in hypertrophic cardiomyopathy. Circ Res 90:263–269

Bonne G, Carrier L, Bercovici J, Cruaud C, Richard P, Hainque B, Gautel M, Labeit S, James M, Beckmann J (1995) Cardiac myosin binding protein-C gene splice acceptor site mutation is associated with familial hypertrophic cardiomyopathy. Nat Genet 11:438–440

Burton D, Abdulrazzak H, Knott A, Elliott K, Redwood C, Watkins H, Marston S, Ashley C (2002) Two mutations in troponin I that cause hypertrophic cardiomyopathy have contrasting effects on cardiac muscle contractility. Biochem J 362:443–451

Chien KR, Zhu H, Knowlton KU, Miller-Hance W, van-Bilsen M, O'Brien TX, Evans SM (1993) Transcriptional regulation during cardiac growth and development. Annu Rev Physiol 55:77–95

Cuda G, Fananapazir L, Zhu WS, Sellers JR, Epstein ND (1993) Skeletal muscle expression and abnormal function of beta-myosin in hypertrophic cardiomyopathy. J Clin Invest 91:2861–2865

Cuda G, Mussari A, Concolino D, Costanzo FS, Strisciuglio P (2002) Co-existence of frataxin and cardiac troponin T gene mutations in a child with Friedreich Ataxia and familial hypertrophic cardiomyopathy. Hum Mutat 19:309–310

Davies MJ (1984) The current status of myocardial disarray in hyptertrophic cardiomyopathy. Br Heart J 51:361–366

Derchi G, Bellone P, Chiarella F, Randazzo M, Zino V, Vecchio C (1992) Plasma levels of atrial natriuretic peptide in hypertrophic cardiomyopathy. Am J Cardiol 70:1502–1504

Elliott K, Watkins H, Redwood CS (2000) Altered regulatory properties of human cardiac troponin I mutants that cause hypertrophic cardiomyopathy. J Biol Chem 275:22069–22074

Epstein ND, Cohn GM, Cyran F, Fananapazir L (1992) Differences in clinical expression of hypertrophic cardiomyopathy associated with two distinct mutations in the beta-myosin heavy chain gene. A 908LeuVal mutation and a 403ArgGln mutation. Circulation 86:345–352

Fananapazir L, Epstein ND (1994) Genotype-phenotype correlation in hypertrophic cardiomyopathy. Insights provided by comparisons of kindreds with distinct and identical beta-myosin heavy chain gene mutations. Circulation 89:22–32

Freiburg A, Gautel M (1996) A molecular map of the interactions between titin and myosin-binding protein-C. Implications for sarcomeric assembly in familial hypertrophic cardiomyopathy. Eur J Biochem 235:317–323

Gautel M, Zuffardi O, Freiburg A, Labeit S (1995) Phosphorylation switches specific for the cardiac isoform of myosin binding protein-C: a modulator of cardiac contraction? EMBO J 14:1952–1960

Geisterfer-Lowrance A, Kass S, Tanigawa G (1990) A molecular basis for familial hypertrophic cardiomyopathy: a beta cardiac myosin heavy chain gene missense mutation. Cell 62:999–1006

Geisterfer-Lowrance AA, Christe M, Conner DA, Ingwall JS, Schoen FJ, Seidman CE, Seidman JG (1996) A mouse model of familial hypertrophic cardiomyopathy. Science 272:731–734

Grossman W, Jones D, McLaurin LP (1975) Wall stress and patterns of hypertrophy in the human left ventricle. J Clin Invest 56:56–64

Hasegawa K, Fujiwara H, Doyama K, Mukoyama M, Nakao K, Fujiwara T, Imura H, Kawai C (1993) Ventriclar expression of brain natriuretic peptide in hypertrophic cardiomyopathy. Circulation 88:372–380

Hasegawa K, Fujiwara H, Koshiji M, Inada T, Ohtani S, Doyama K, Tanaka M, Matsumori A, Fujiwara T, Shirakami G, Hosoda K, Nakao K, Sasayama S (1996) Endothelin-I and its receptor in hypertrophic cardiomyopathy. Hypertension 27:259–264

Ho CY, Lever HM, DeSanctis R, Farver CF, Seidman JG, Seidman CE (2000) Homozygous mutation in cardiac troponin T: implications for hypertrophic cardiomyopathy. Circulation 102:1950–1955

Iwai N, Ohmichi N, Nakamura Y, Kinoshita M (1994) DD genotype of the angiotensin-converting enzyme gene is a risk factor for left ventricular hypertrophy. Circulation 90:2622–2628

James J, Zhang Y, Osinska H, Sanbe A, Klevitsky R, Hewett TE, Robbins J (2000) Transgenic modeling of a cardiac troponin I mutation linked to familial hypertrophic cardiomyopathy. Circ Res 87:805–811

Jeschke B, Uhl K, Weist B, Schroder D, Meitinger T, Dohlemann C, Vosberg HP (1998) A high risk phenotype of hypertrophic cardiomyopathy associated with a compound genotype of two mutated β-myosin heavy chain genes. Hum Genet 102:299–304

Kai H, Muraishi A, Sugiu Y, Nishi H, Seki Y, Kuwahara F, Kimura A, Kato H, Imaizumi T (1998) Expression of proto-oncogenes and gene mutation of sarcomeric proteins in patients with hypertrophic cardiomyopathy. Circ Res 83:594–601

Kimura A, Harada H, Park JE, Nishi H, Satoh M, Takahashi M, Hiroi S, Sasaoka T, Ohbuchi N, Nakamura T, Koyanagi T, Hwang TH, Choo JA, Chung KS, Hasegawa A, Nagai R, Okazaki O, Nakamura H, Matsuzaki M, Sakamoto T, Toshima H, Koga Y, Imaizumi T, Sasazuki T (1997) Mutations in the cardiac troponin I gene associated with hypertrophic cardiomyopathy. Nat Genet 16:379–382

Kokado H, Shimizu M, Yoshio H, Ino H, Okeie K, Emoto Y, Matsuyama T, Yamaguchi M, Yasuda T, Fujino N, Ito H, Mabuchi H (2000) Clinical features of hypertrophic cardiomyopathy caused by a Lys183 deletion mutation in the cardiac troponin I gene. Circulation 102:6636–6639

Komuro I, Kurabayashi M, Takaku F, Yazaki Y (1988) Expression of cellular oncogenes in the myocardium during the developmental stage and pressure-overloaded hypertrophy of the rat heart. Circ Res 62:1075–1079

Lang R, Gomes AV, Zhao J, Housmans PR, Miller T, Potter JD (2002) Functional analysis of a troponin I (R145G) mutation associated with familial hypertrophic cardiomyopathy. J Biol Chem 277:11670–11678

Lankford EB, Epstein ND, Fananapazir L, Sweeney HL (1996) Abnormal contractile properties of muscle fibers expressing beta-myosin heavy chain gene mutations in patients with hypertrophic cardiomyopathy. J Clin Invest 95:1409–1414

Lechin M, Quinones MA, Omran A, Hill R, Yu QT, Rakowski H, Wigle D, Liew CC, Sole M, Roberts R, Marian AJ (1995) Angiotensin-I converting enzyme genotypes and left ventricular hypertrophy in patients with hypertrophic cardiomyopathy. Circulation 92:1808–1812

Li D, Czernuszewicz GZ, Gonzalez O, Tapscott T, Karibe A, Durand JB, Brugada R, Hill R, Gregoritch JM, Anderson JL, Quinones M, Bachinski LL, Roberts R (2001) Novel cardiac troponin T mutation as a cause of familial dilated cardiomyopathy. Circulation 104:2188–2193

Lin D, Bobkova A, Homsher E, Tobachman L (1996) Altered cardiac troponin T in vitro function in the presence of a mutation implication in familial hypertrophic cardiomyopathy. J Clin Invest 97:2842–2848

Marian AJ, Yu QT, Workman R, Greve G, Roberts R (1993) Angiotensin-converting enzyme polymorphism in hypertrophic cardiomyopathy and sudden cardiac death. Lancet 342:1085–1086

Marian AJ, Wu YU, Lim DS, McCluggage M, Youker K, Yu QT, Brugada R, DeMayo F, Quinones M, Roberts R (1999) A transgenic rabbit model for human hypertrophic cardiomyopathy. J Clin Invest 104:1683–1692

Maron BJ, Roberts WC, Edwards JE, McAllister HA, Foley DD, Epstein SE (1978) Sudden death in patients with hypertrophic cardiomyopathy. Characterization of 26 patients without functional limitation. Am J Cardiol 41:803–810

Maron BJ, Roberts WC, Epstein S (1986) Causes of sudden death in competitive athletes. J Am Coll Cardiol 7:204–214

Maron BJ, Bonow RO, Cannon RO, Leon MB, Epstein SE (1987) Hypertrophic cardiomyopathy: interrelations of clinical manifestation, pathophysiology and therapy. N Engl J Med 316:780–789

Maron BJ, Mitten MJ, Quandt EF, Zipes DP (1998) Competitive athletes with cardiovascular disease: the case of Nicholas Knapp. N Engl J Med 339:1632–1635

McConnell BK, Jones KA, Fatkin D, Arroyo LH, Lee RT, Aristizabal O, Turnbull DH, Gergakopoulos D, Kass D, Bond M, Niimura H, Schoen FJ, Conner D, Fischman DA, Seidman CE, Seidman JG, Fischman DH (1999) Dilated cardiomyopathy in homozygous myosin-binding protein-C mutant mice. J Clin Invest 104:1235–1244

Michele DE, Albayya FP, Metzger JM (1999) Direct, convergent hypersensitivity of calcium-activated force generation produced by hypertrophic cardiomyopathy mutant α-tropomyosins in adult cardiac myocytes. Nat Med 5:1413–1417

Mogensen J, Klausen IC, Pedersen AK, Egeblad H, Bross P, Kruse TA, Gregersen N, Hansen PS, Baandrup U, Borglum AD (1999) α-cardiac actin is a novel disease gene in familial hypertrophic cardiomyopathy. J Clin Invest 103:R39–R43

Montgomery DE, Tardiff JC, Chandra M (2001) Cardiac troponin T mutations: correlation between the type of mutation and the nature of myofilament dysfunction in transgenic mice. J Physiol 15:583–592

Moolman-Smook JC, Mayosi B, Brink P, Corfield VA (1998) Identification of a new missense mutation in MyBP-C associated with hypertrophic cardiomyopathy. J Med Genet 35:253–254

Moos C, Feng IM (1990) Effect of C-protein on actomyosin ATPase. Biochim Biophys Acta 632:141–149

Nakajima-Taniguchi C, Matsui H, Nagata S, Kishimoto T, Yamauchi-Takihara K (1995) Novel missense mutation in alpha-tropomyosin gene found in Japanese patients with hypertrophic cardiomyopathy. J Mol Cell Cardiol 27:2053–2058

Niimura H, Patton KK, McKenna WJ, Soults J, Maron BJ, Seidman JG, Seidman CE (2002) Sarcomere protein gene mutations in hypertrophic cardiomyopathy of the elderly. Circulation 105:446–451

Nishi H, Kimura A, Harada H, Harada H, Imaizumi T, Toshima H, Sasazuki T, Kimura A (1995) A myosin missense mutation, not a null allele, causes familial hypertrophic cardiomyopathy. Circulation 91:2911–2915

Oberst L, Ahao G, Park JT, Brugada R, Michael LH, Entman ML, Roberts R, Marian AJ (1998) Dominant-negative effect of a mutant cardiac troponin T on cardiac structure and function in transgenic mice. J Clin Invest 102:1498–1505

Okagi T, Weber E, Fischman DA, Vaughan KT, Mikawas T, Reinach FC (1993) The major myosin-binding domain of skeletal muscle MyBP-C (C protein) resides in the COOH-terminal, immunoglobulin C2 motif. J Cell Biol 123:619–626

Olson TM, Kishimoto NY, Whitby FG, Michels VV (2001) Mutations that alter the surface charge of alpha-tropomyosin are associated with dilated cardiomyopathy. J Mol Cell Cardiol 33:723–732

Ortlepp JR, Vosberg HP, Reith S, Ohme F, Mahon NG, Schroder D, Klues HG, Hanrath P, McKenna WJ (2002) Genetic polymorphisms in the renin-angiotensin-aldosterone system associated with expression of left ventricular hypertrophy in hypertrophic cardiomyopathy: a study of five polymorphic genes in a family with a disease causing mutation in the myosin binding protein C gene. Heart 87:270–275

Poetter K, Jiang H, Hassanzadeh S, Master SR, Chang A, Dalakas MC, Rayment I, Sellers JR, Fananapazir L, Epstein ND (1996) Mutations in either the essential or regulatory light chains of myosin are associated with a rare myopathy in human heart and skeletal muscle. Nat Genet 13:63–69

Prabhakar R, Boivin GP, Grupp IL, Hoit B, Arteaga G, Solaro JR, Wieczorek DF (2001) A familial hypertrophic cardiomyopathy alpha-tropomyosin mutation causes severe cardiac hypertrophy and death in mice. J Mol Cell Cardiol 33:1815–1828

Rayment I, Holden HM, Sellers JR, Fananapazir L, Epstein ND (1995) Structural interpretation of the mutations in β-cardiac myosin that have been implicated in familial hypertrophic cardiomyopathy. Proc Natl Acad Sci USA 92:3864–3868

Reddy DS (1997) Cellular and molecular biology of cardiac hypertrophy. Curr Science 72:13–30

Richard P, Isnard R, Carrier L, Dubourg O, Donatien Y, Mathieu B, Bonne G, Gary F, Charron P, Hagege M, Komajda M, Schwartz K, Hainque B (1999) Double heterozygosity for mutations in the β-myosin heavy chain and in the cardiac myosin binding protein C genes in a family with hypertrophic cardiomyopathy. J Med Genet 36:542–545

Robinson P, Mirza M, Knott A, Abdulrassak H, Willott R, Marston S, Watkins H, Redwood C (2002) Alterations in thin filament regulation induced by a human cardiac troponin T mutant that causes dilated cardiomyopathy are distinct from those induced by troponin T mutants that cause hypertrophic cardiomyopathy. J Biol Chem 277:40710–40716

Roopnarine O, Leinwand LA (1998) Functional analysis of myosin mutations that cause familial hypertrophic cardiomyopathy. Biophys J 75:3023–3030

Sadoshima J, Izumo S (1993) Mechanical stretch rapidly activates multiple signal transduction pathways in cardiac myocytes: potential involvement of an autocrine/paracrine mechanism. EMBO J 12:1681–1692

Sata M, Ikebe M (1996) Functional analysis of the mutations in the human cardiac beta-myosin that are responsible for familial hypertrophic cardiomyopathy. Implication for the clinical outcome. J Clin Invest 98:2866–2873

Satoh M, Takahashi M, Sakamoto T, Hirle M, Marumo F, Kimura A (1999) Structural analysis of the titin gene in HCM: identification of a novel disease gene. Biochem Biophys Res Commun 262:411–417

Schlender KK, Bean LJ (1991) Phosphorylation of chick cardiac C-protein by calcium/calmodulin-dependent protein kinase II. J Biol Chem 266:2811–2817

Schultheiss T, Lin ZX, Lu MH, Murray J, Fischman DA, Weber K, Masaki T, Imamura M, Holtzer H (1990) Differential distribution of subsets of myofibrillar proteins in cardiac nonstriated and striated myofibril. J Cell Biol 110:1159–1172

Sehnert AJ, Huq A, Weinstein BM, Walker C, Fishman M, Stainier DY (2002) Cardiac troponin T is essential in sarcomere assembly and cardiac contractility. Nat Genet 31:106–110

Semsarian C, Healey MJ, Fatkin D, Giewat M, Duffy C, Seidman CE, Seidman JG (2001) A polymorphic modifier gene alters the hypertrophic response in a murine model of familial hypertrophic cardiomyopathy. J Mol Cell Cardiol 33:2055–2060

Shimizu M, Ino H, Okeie K, Yamaguchi M, Nagata M, Hayashi K, Itoh H, Iwaki T, Oe K, Konno T, Mabuchi H (2002) Septal wall thinning and systolic dysfunction in patients with hypertrophic cardiomyopathy caused by a cardiac troponin I gene mutation. Am Heart J 143:690–695

Simpson PC (1988) Role of proto-oncogenes in myocardial hypertrophy. Am J Cardiol 62:13G–19G

Sweeney HL, Straceski AJ, Leinwand LA, Tikunov BA, Faust L (1994) Heterologous expression of a cardiomyopathic myosin that is defective in its actin interaction. J Biol Chem 269:1603–1605

Sweeney HL, Feng HS, Yang Z, Watkins H (1998) Functional analyses of troponin T mutations that cause hypertrophic cardiomyopathy: insights into disease pathogenesis and troponin function. Proc Natl Acad Sci USA 95:14406–14410

Tae-Hong H, Lee W-H, Kimura A, Satoh M, Nakamura T, Kim MK, Choi SK, Park JE (1998) Early expression of a malignant phenotype of familial hypertrophic cardiomyopathy associated with a Gly716Arg myosin heavy chain mutation in a Korean family. Am J Cardiol 82:1509–1513

Takahashi-Yanaga F, Morimoto S, Harada K, Minakami R, Shiraishi F, Ohta M, Lu QW, Sasaguri T, Ohtsuki I (2001) Functional consequences of the mutations in human cardiac troponin I gene found in familial hypertrophic cardiomyopathy. J Mol Cell Cardiol 33:2095–2107

Takai E, Akita H, Shiga N, Kanazawa K, Yamada S, Terachima M, Matsuda Y, Iwai C, Kawai K, Tokota Y, Yomoyama M (1999) Mutational analysis of the cardiac actin gene in familial and sporadic dilated cardiomyopathy. Am J Med Genet 86:325–327

Tanigawa G, Jarcho JA, Kass S, Solomon SD, Vosberg HP, Seidman JG, Seidman CE (1990) A molecular basis for familial hypertrophic cardiomyopathy: an alpha/beta cardiac myosin heavy chain hybrid gene. Cell 62:991–998

Tesson F, Dufour C, Moolman JC (1997) The influence of the angiotensin I converting enzyme genotype in familial hypertrophic cardiomyopathy varies with the disease gene mutation. J Mol Cell Cardiol 29:831–838

Thierfelder L, Watkins H, MacRae C, Lamas R, McKenna W, Vosberg HP, Seidman JG, Seidman CE (1994) Alpha-tropomyosin and cardiac troponin T mutations cause familial hypertrophic cardiomyopathy: a disease of the sarcomere. Cell 77:701–712

Tyska MJ, Hayes E, Giewat M, Seidman CE, Seidman JG, Warshaw DM (2000) Single-molecule mechanics of R403Q cardiac myosin isolated from the mouse model of familial hypertrophic cardiomyopathy. Circ Res 86:737–744

Vikstrom KL, Leinwand LA (1996) Contractile protein mutations and heart disease. Curr Opinion Cell Biol 8:97–105

Waldmuller S, Freund P, Mauch S, Toder R, Vosberg HP (2002) Low-density DNA microarrays are versatile tools to screen for known mutations in hypertrophic cardiomyopathy. Hum Mutat 19:560–569

Watkins H, Rosenzweig A, Hwang DS (1992) Characteristics and prognostic implications of myosin missense mutations in familial hypertrophic cardiomyopathy. N Engl J Med 326:1108–1114

Watkins H, Conner D, Thierfelder J, Jarcho JA, MacRae C, McKenna WJ, Maron BJ, Seidman JG, Seidman CE (1995a) Mutations in the cardiac myosin binding protein-C gene on chromosome 11 cause familial hypertrophic cardiomyopathy. Nat Genet 11:434–437

Watkins H, McKenna WJ, Thierfelder L, Suk HJ, Anan R, O'Donoghue A, Spirito P, Matsumori A, Moravec CS, Seidman JG (1995b) Mutations in the genes for cardiac troponin T and α-tropomyosin in hypertrophic cardiomyopathy. N Engl J Med 332:1058–1064

Watkins H, Seidman JG, Seidman CE (1995c) Familial hypertrophic cardiomyopathy: a genetic model of cardiac hypertrophy. Hum Mol Genet 4:1721–1727

Watkins H, Seidman CE, Seidman JG, Feng HS, Sweeney HL (1996) Expression and functional assessment of a truncated cardiac troponin T that causes hypertrophic cardiomyopathy. J Clin Invest 98:2456–2461

Wilkins BJ, Molkentin JD (2002) Calcineurin and cardiac hypertrophy: Where have we been? Where are we going? J Physiol 541:1–8

Witt CC, Gerull B, Davies MJ, Centneri T, Linke WA, Thierfelder L (2001) Hypercontractile properties of cardiac muscle fibers in a knock-in mouse model of cardiac myosin-binding protein-C. J Biol Chem 276:5353–5359

Yamauchi-Takihara K, Nakajima-Taniguchi C, Matsui H (1996) Clinical implications of hypertrophic cardiomyopathy associated with mutations in the α-tropomyosin gene. Eur J Heart 76:63–65

Yanaga F, Morimoto S, Ohtsuki I (1999) Ca^{2+} sensitization and potentiation of the maximum level of myofibrillar ATPase activity caused by mutations of troponin T found in familial hypertrophic cardiomyopathy. J Biol Chem 274:8806–8812

Yang Q, Sanbe A, Osinska H, Hewett TE, Klevitsky R, Robbins J (1998) A mouse model of myosin binding protein C human familial hypertrophic cardiomyopathy. J Clin Invest 102:1292–1300

Yang Q, Sanbe A, Osinska H, Hewett TE, Klevitsky R, Robbins J (1999) In vivo modeling of myosin binding protein C familial hypertrophic cardiomyopathy. Circ Res 85:841–847

Yu B, French JA, Carrier L, Jeremy RW, McTaggart DR, Nicholson MR, Hambly B, Semsarian C, Richmond DR, Schwartz K, Trent RJ (1998) Molecular pathology of familial hypertrophic cardiomyopathy caused by mutations in the cardiac myosin binding protein C gene. J Med Genet 35:205–210