AN ABSTRACT OF THE DISSERTATION OF

Charlotte Rasmussen for the degree of Doctor of Philosophy in Genetics presented on December 8, 1995.

Title: Cis– and Trans–acting Sequences involved in Baculovirus Transcription and Replication

Redacted for Privacy

Abstract approved: _____

George F. Rohrmann

To learn more about transcription in baculovirus–infected *Spodoptera frugiperda* cells (Sf9), the cDNA clone encoding the Sf9 TATA–binding protein (307 amino acids; 34 Kda) was isolated, sequenced and its expression in response to baculovirus infection monitored. Northern analysis showed that SfTBP transcription, like other host genes, was inhibited by baculovirus infection. Phylogenetic analysis indicated that TBP is a reasonable macromolecule to use to infer phylogenetic relationships though it may be more informative within kingdoms or phyla.

An *in vitro* transcription system, using nuclear extracts prepared from Sf9 cells at various times post–infection transcribed early, late and very late baculovirus gene promoters in a temporal manner similar to that seen during baculovirus infection. Nuclear extracts prepared at 16 hr p.i. were optimal for baculovirus late gene transcription and contained SfTBP and AcMNPV IE–1, a viral transactivator required for baculovirus DNA replication. Deletion analysis of the baculovirus late promoter, *vp39*, identified a minimal late promoter but failed to identify late promoter–specific DNA binding in gel retardation assays.

IE–1 was shown to specifically bind to the *ie–2* and *pe38* promoters and *hr*1a, one of eight regions in the AcMNPV genome that contains repeated imperfect palindromes that function as transcriptional enhancers and origins of baculovirus

replication. Comparison of the IE–1 binding sites revealed a putative IE–1 nucleotide binding sequence, 5′–ACBYGTAA–3′.

Since palindromic DNA sequences can potentially form cruciform structures, the relationship between cruciform formation, IE-1 binding and *hr* transcriptional activity was examined using two–dimensional gel electrophoresis of topoisomers, nuclease P1, β–glucuronidase reporter gene and gel retardation assays. An altered 42–bp *hr* palindrome containing no mismatches required 9.6 kcal/mole to form a cruciform whereas cruciform formation was not observed with the 42–bp AcMNPV consensus *hr* palindrome. Both the consensus *hr* and the altered *hr* bound IE–1 and functioned as enhancers of transcription equally well. The 42–bp sequence AcMNPV consensus *hr* palindrome is the smallest DNA sequence reported to date that binds IE–1 and functions as an enhancer indicating that the 42–bp encompassing the AcMNPV *hr* palindrome is all that is required for enhancer function.

Cis– and Trans–acting Sequences involved in Baculovirus Transcription and Replication

by

Charlotte Rasmussen

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Completed December 8, 1995

Commencement June 1996

Doctor of Philosophy dissertation of Charlotte Rasmussen

presented on December 8, 1995

APPROVED:

Redacted for Privacy

Major Professor, representing Genetics Program

Redacted for Privacy

Chair of Genetics Program

Redacted for Privacy

Dean of Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Redacted for Privacy

Charlotte Rasmussen, Author

ACKNOWLEDGEMENT

I would like to thank Dr. Ken van Holde, Dr. George Pearson, and Dr. Joann Leong for serving on my advisory committee. I would especially like to thank Dr. Lynda Ciuffetti for serving as my graduate school representative, above and beyond the call of duty.

I would like to thank the members, both past and present, of the Rohrmann lab for their advice during the course of this work. I would also like to thank Dr. P. Shing Ho, Dr. Gary Schroth and Dr. Meredith Howell for their invaluable advice concerning DNA structural analysis.

I want to thank my family, Kirsten, Carl and Chris for providing a place to go home to. Mostly, I want to thank Jim for being with me every step of the way.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Cis– and Trans–acting Sequences involved in Baculuovirus Transcription and Replication

Chapter 1
Introduction

Baculoviruses are a diverse group of large double-stranded DNA viruses that infect holometabolus insects primarily of the order *Lepidoptera* [87]. Since some members of this insect order are economically important pests, the virulence of certain baculoviruses for specific insects has created interest in their use as biological pesticides (for review see [168]). More recently, interest in baculovirus biology has increased due to the development of baculovirus expression systems that produce large quantities of foreign proteins in insect cells (for review see [188, 13, 131, 227]). The major advantage of baculovirus expression systems over bacterial expression systems is the abundant expression of recombinant proteins that in many cases are functionally similar to their authentic counterparts [188]. In addition, baculovirus expression systems produce large quantities of recombinant protein when compared to yeast and mammalian expression systems. Baculovirus expression systems utilize recombinant forms of the most–intensively studied baculovirus, *Autographa californica* multinucleocapsid nuclear polyhedrosis virus (AcMNPV), which grows to high titer in *Spodoptera frugiperda* cells [13].

1.1 Baculovirus Structure

Baculoviruses are characterized by enveloped, rod-shaped virions that contain covalently closed double-stranded viral DNA ranging in size from 88 to greater than 166 kbp [273, 7]. Within the genus Baculoviridae, morphologically distinct types of baculoviruses have been characterized. The occluded viruses consist

of the nuclear polyhedrosis viruses (NPVs), and the granulosis viruses (GVs). NPVs are composed of a viral nucleoprotein complex that is encased by capsid proteins forming the nucleocapsid. The nucleocapsid is surrounded by an envelope and embedded in a crystalline structure composed primarily of the protein, polyhedrin. The occlusion bodies or polyhedra, protect the virions from a number of environment agents, and as a result, occluded viruses are very stable in the soil. NPVs have many virions occluded within each single crystal of polyhedra. NPVs can be subdivided into SNPVs which contain a single nucleocapsid within an envelope, and MNPVs which have multiple nucleocapsids within a single envelope. Granulosis viruses are similar in morphology to NPVs except that a single virion is occluded within a crystalline matrix composed of the protein, granulin.

1.2 Baculovirus Infection Cycle

The archetype baculovirus, *Autographa californica* multinucleocapsid nuclear polyhedrosis virus, (AcMNPV) infects at least 32 species of insects including *Autographa californica* or the alfalfa looper [128]. AcMNPV, and all other occluded baculoviruses, begin their infection cycle [87, 17] upon ingestion of the occluded viruses by a susceptible insect. Upon ingestion, the high pH of the midgut dissolves the polyhedra and releases infectious virions. The virions then fuse with the midgut epithelial cells releasing nucleocapsids into the cells. Once inside the cell, nucleocapsids are transported to the nucleus where uncoating exposes the viral DNA to the host cell machinery. The viral DNA then undergoes a primary round of replication and the replicated DNA is packaged into progeny virions. The virions then bud through the nuclear membrane and travel through the cytoplasm to bud through the plasma membrane before being released into the tracheal system of the insect host [63, 143, 139]. The budded virus (BV) then infects many host cells producing a systemic infection resulting in a second

round of replication.

Approximately 24 hours post-infection, a second viral phenotype, polyhedra derived virus (PDV), is produced. The nucleocapsids produced at this time are enveloped de novo within the nucleus and occluded within polyhedra. The infection progresses producing large quantities of virus resulting in the death of the insect and subsequent release of occluded viruses (PDVs) into the environment. A typical baculovirus infection cycle is diagrammatically represented in Figure 1.1.

1.3 Structural Components of Baculoviruses

As described above, the infection of insect cells with a baculovirus results in the production of two viral phenotypes, BV and PDV, that have specific roles in baculovirus infection. The BV phenotype, produced early in infection, is responsible for the systemic spread of the infection within the insect. The PDVs, produced late in the infection cycle, are responsible for horizontal spread of the virus within an insect population. PDVs are capable of persisting in the environment for extended periods of time and therefore may be responsible for future disease outbreaks.

PDVs and BVs are structurally distinct [313] containing both common and phenotype–specific virion components (see Figure 1.2). Their nucleocapsids appear to be identical whereas their envelopes are distinctly different. Nucleocapsids consist a viral DNA core that is complexed with a 6.9 kDA arginine–rich DNA binding protein [324, 267, 189]. The viral capsid surrounding the nucleoprotein core is composed of at least two proteins of molecular weight 39 (vp39) [238, 266, 16, 302] and 87 kDA (p87) [212].

Nucleocapsids from both BVs and PDVs are enveloped. In BVs, the envelopes are derived from the viral–modified plasma membrane of the insect cells which contains the viral protein, gp64 (gp67) [19]. gp64 appears to function as

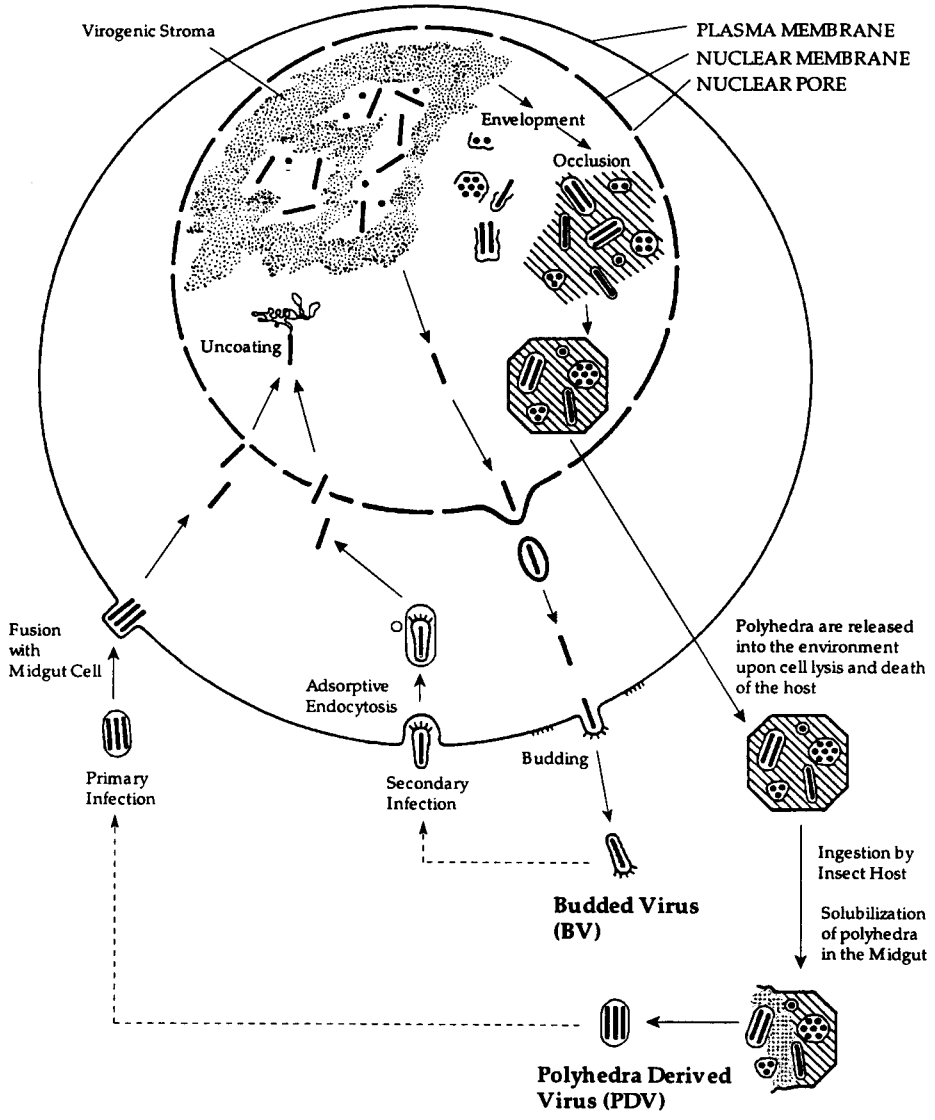


Figure 1.1. The Baculovirus Infection Cycle. This diagram is modified from [17].

a pH–dependent fusion protein [19, 204] facilitating BV entry into cells through endocytosis. PDVs are encased in polyhedrin, an alkali–soluble 29 kDa protein that forms the crystalline matrix surrounding the nucleocapsid (for review see [261]). Polyhedrin can represent up to 15% of the total coomassie stainable protein synthesized by baculovirus infected insects [247]. This high level of expression is the basis of the aforementioned baculovirus expression systems. Electron microscopy (EM) studies indicate that polyhedra are surrounded by electron dense envelope containing the polyhedron envelope protein [84, 321]. The *p10* gene product is thought to be involved in the formation of extensive fibrillar structures found in the cytoplasm and nucleus of infected cells [309]. Williams *et al.* [323] have shown that deletion of the *p10* gene resulted in the absence of the fibrillar structures and in the production of polyhedra that lacked or had fragmented polyhedron envelopes. Furthermore, in viral mutants that lack the *p10* gene, polyhedra are fragile and have irregular pitted surfaces [89] suggesting that p10 may have a role in the stability of polyhedra. The structural components of the two viral phenotypes of baculoviruses are shown in Figure 1.2.

1.4 Genome Organization

Physical maps showing the cleavage sites of a variety of different restriction endonucleases for a number of baculoviruses including AcMNPV [310, 150], *Bombyx mori* MNPV (BmMNPV) [190], and *Orygia pseudotsugata* MNPV (OpMNPV) [171, 35] have been constructed. Subsequent determination of the location and sequence of a number of baculovirus genes indicate that, although different baculoviruses have accumulated substantial sequence diversity, they are evolutionarily related [17].

Numerous genotypic variants of AcMNPV are used in many laboratories but differ only slightly from one another [227]. Recently, the complete genomic sequence of clone 6 of AcMNPV has been determined [7]. Analysis of the complete

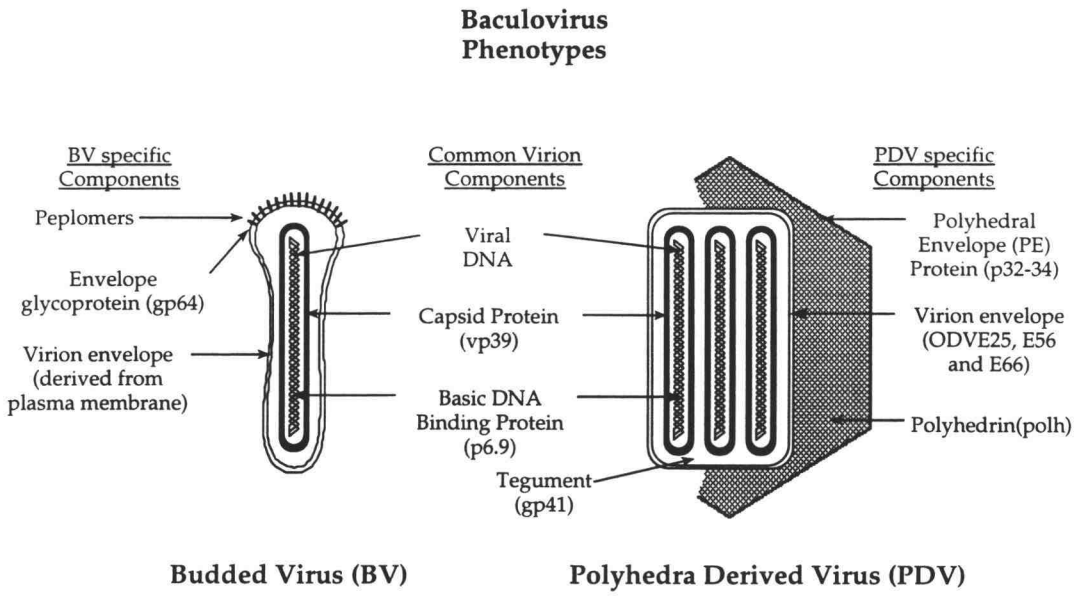


Figure 1.2. Baculovirus Structural Components. The two baculovirus virion phenotypes are shown; shared and virion–specific components are indicated. This diagram is modified from [17].

sequence of AcMNPV indicates that the genome has the potential to encode 154 methionine–initiated open reading frames (ORF)s of at least 150 nucleotides (50 amino acids). These ORFs are distributed evenly throughout genome on either DNA strand and arranged as adjacent, non–overlapping reading frames separated by short intragenic regions [7]. There appears to be little or no clustering of functionally–related genes. For example, viral structural genes [7, 150] and viral genes required for replication [7, 150] are dispersed throughout the genome. However, three genes involved in early gene expression, *ie–1*, *ie–2* and *pe38* are located within 5 kb of one another [7, 150].

The AcMNPV genome encodes a number of non–essential genes; the introduction of a null mutation into a viral gene that does not appear to affect the ability of the virus to replicate in tissue culture or insect hosts defines a non–essential gene. Non–essential genes are thought to confer growth advantages to the virus under specific conditions that are not detected under standard

laboratory conditions such as alternate insect hosts, harsh field conditions or a particular cell or tissue type.

A notable feature of the AcMNPV genome is the presence of eight related regions containing *Eco*R1 sites. These regions known as homologous regions or *hr*s contained two to eight copies of an imperfect palindrome each separated by less well conserved direct repeat elements [102, 7]. Originally identified by Cochran and Faulkner [47], who first suggested that *hr*s might function as origins of DNA replication, *hr*s appear to function as both enhancers for some early promoters [98, 102, 218] and as origins of replication (see section 1.7) [235, 149, 152, 170].

1.5 Expression of Baculovirus Genes

In baculovirus–infected insect cells, the expression of viral genes occurs in a temporally-controlled fashion resulting in a cascade of early, late and very late gene transcription. Expression of viral genes leading ultimately to the production and release of infectious baculoviruses occurs via an ordered cascade of events; each successive group of viral genes depends on the expression of a prior group for activation [64] (for review see [17, 262]). Most evidence suggests that the cascade is regulated at the level of transcription [17].

Transcription of baculovirus early genes begins before initiation of viral DNA replication and is mediated by the host RNA polymerase II transcription machinery [76, 122, 83]. Most early gene promoters contain a TATA–box; the common core–promoter element (consensus TATAAA) usually found between –25 and –30 relative to the start site of RNA polymerase II transcription in eukaryotic organisms. Transcription initiation usually occurs within the consensus sequence, CAGT [18, 15]. Analysis of a number of arthropod promoter elements indicate that the sequence element TCAGT is overrepresented [38] at the site of transcription initiation. The similarity of this sequence to those of vertebrate

initiators (Inrs) of the TdT and Adenovirus ML class [318] suggest that mechanisms similar to RNA polymerase II transcription of TATA–less promoters may be used during baculovirus early transcription.

1.5.1 The TATA–Binding Protein

Transcription initiation by eukaryotic RNA polymerase II has been extensively studied (for review see [88, 336, 81, 50]). Chromatographic fractionation and biochemical analysis of cell extracts have identified at least seven fractions required for transcription initiation. These fractions, termed TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH and TFIIJ, contain a number of general transcription factors required for RNA polymerase II transcription initiation [336]. The TFIID fraction contains a polypeptide, the TATA–binding protein (TBP), that binds to the TATA–box [32, 25] (for review see [272, 88]). TBP and a number of other proteins termed TATA–binding protein associated factors (TAFs) [59, 245, 85] comprise the basal transcription factor, TFIID. RNA polymerase II transcription initiation complexes at *Drosophila* promoters contain at least eight TAFs associated with TBP [36, 86, 59] and human transcription initiation complexes have at least ten TAFs associated with TBP [317, 146]. The binding of TBP and associated factors to the TATA–box is the first step in the assembly of an active RNA polymerase II transcription complex [25]. TFIID then interacts with a variety of other general transcription factors including TFIIA [163], TFIIB [43], the C–terminal domain of large subunit of RNA polymerase II [306], and the initiator binding factor (TFII–I) [137] (TFII–I may be the site–specific DNA binding component of the complex in TATA–less promoters) to form a functional pre–initiation complex (for review see [336]). Transcriptional stimulation via the initiator requires TFIID [282] and it appears that TBP is directed to bind to the –30 region regardless of the sequence [337].

Recent evidence indicates that TBP is also required for RNA polymerase

I [48, 249, 49] and RNA polymerase III [281, 183, 114] transcription, suggesting that TBP is an universal transcription factor (for review see [279, 320, 293]). RNA polymerase I transcription initiation complexes contain three polypeptides, TAF110, TAF63 and TAF48 that individually bind specifically to TBP to form the SL1 complex [49]. Efficient recognition and transcription of the ribosomal RNA promoter requires cooperative binding between SL1 and a second factor, UBF [8]. Another complex containing TBP and a different set of TAFs is responsible for RNA polymerase III transcription initiation [114]. There appears to be a diverse set of TBP–TAFs complexes that are responsible for promoter–specific RNA polymerase selection (for review see [85, 279]).

TFIID–promoter interactions may facilitate melting of the DNA helix resulting in the formation of an open complex. Mizutani *et al.* [203] have shown that DNA supercoiling promotes the formation of the preinitiation complex on the fibrion gene promoter and order of addition experiments suggest that DNA supercoiling facilitates the binding of the TFIID fraction, presumably TBP, to the promoter suggesting that binding of TFIID may accompany local melting resulting in the formation of an open transcription complex. Co–crystallization of the TBP/TATA–box complex showed that the binding of TBP to DNA introduces a 80 degree bend in the DNA molecule [140, 141, 34]. Mechanistically, DNA bending could allow the general transcription factors to form the necessary contacts to initiate transcription and/or may unwind the DNA duplex, thereby assisting in the formation of a an open transcription complex.

1.5.2 TBP–interactions

Interactions between TFIID and other proteins including viral regulators have been shown both genetically and biochemically. The TBP–containing TFIID fraction may be a major target for gene regulation. Recent evidence suggests that several of the TAFs may function as co–activators; $TAF_{II}110$ is required

for Sp1 activation [118] while $TAF_{II}150$ is required for NTF–1 mediated activation [36]. Therefore, components of the TBP–TAF or holo–TFIID complex serve as an important targets for some promoter specific activator proteins of both viral and cellular origin. Cellular transcription factors that appear to interact with holo–TFIID include ATF [124, 108], USF [271, 328, 202] and Gal 4 [123] and the Retinoic Acid Receptor [9]. Viral transactivators, including E1A of Adenovirus[164], IE2 and IE86 of Cytomegalovirus [106, 134], Zta of Epstein-Barr virus [180, 179], IE of pseudorabies virus [327, 1], VP16 and ICP4 of Herpes Simplex virus [290, 93, 85], have been shown to interact with proteins of the basic transcriptional machinery including TBP and associated TAFs. The acidic domain of the herpes simplex protein VP16 binds strongly and selectively to both the human and yeast TATA–box binding factor suggesting that the acidic domains in transcription factors may be required for interaction with TFIID [290]. Mutations in the TATA sequences in the adenovirus early control region dramatically reduces inducibility by E1A [330]. *In vitro* studies with the immediate early protein of pseudorabies virus (PvRIE) indicate that PvRIE stimulates TFIID binding to promoter sequences during nucleosome assembly [327, 1]. This result suggests that trans–activating viral factors may increase the transcription rate of viral genes by increasing the rate and/or stability of TFIID binding under conditions of direct competition between nucleosome assembly and transcriptional preinitiation complex formation *in vivo*. Alternatively, transactivators may interact with TFIID or other components of the basal transcriptional machinery [182, 43] to accelerate rate limiting steps in the formation of the pre–initiation complex (reviewed in [336]). Finally, Adenovirus E1A has been shown to transactivate RNA polymerase III transcription apparently by altering the binding properties of TFIIIC [55] indicating that viral regulatory proteins may interact with a variety of RNA polymerase complexes.

1.5.3 Baculovirus Transactivators

Baculovirus early genes are transcribed by uninfected insect cell nuclear extracts [122, 83] and require no viral gene products for expression when transfected into insect cells [102]. The binding of TATA–binding protein to early gene promoters is likely to be the first event in the baculovirus transcriptional cascade. At this point, baculovirus transactivating factors may interact with TBP in a manner similar to the transactivating factors of the Herpes viruses. The level of transcription from baculovirus early gene promoters is modulated by viral transactivating factors which include immediate early gene, *ie–1* [103], *ie–0* [156], *ie–2* (formerly called *ien*) [29, 30, 334], and possibly *pe38* [158]. The TATA sequence is sufficient to allow transactivation by IE–1 from a baculovirus early promoter suggesting interactions between TFIID and IE–1 [18]. In transient expression assays, the viral *hr* sequences function as enhancer elements to elevate the transcript levels of early genes [102, 98, 218, 28, 185]. *Hr*s have been shown to enhance expression of reporter genes under the control of baculovirus early promoters including those from the *39k* [28, 102], *p35* [218] and *p143* [185] genes, by more than 1000–fold when linked *in cis* with the promoter and co–transfected with the baculovirus regulatory gene *ie–1*.

Carson and coworkers [30] have shown that IE–1 may also function as a repressor of transcription. Comparison of the *ie–1* sequence to the sequences of immediate early genes of other viruses revealed no significant similarities [103]. However, the deduced amino acid sequence of IE–1 protein indicates that it is highly charged and contains an acidic N–terminal region. Gel retardation assays have shown that whole–cell extracts from Sf9 cells transfected with the *ie–1* gene contain protein(s) that bind to regions within *hr*5 [96, 97, 259]. Construction of N– and C–terminal deletion mutants indicated that IE–1 has at least two distinct regions: a N–terminal acidic domain necessary for transactivation and a C–terminal domain required for DNA binding [155]. Gel retardation analysis

with sequential N–terminal deletions of IE–1 resulted in the appearance of faster migrating complexes, suggesting that IE–1 is a component of the protein–DNA complex bound to *hr*5. Recently, it was shown that IE–1 is required for replication [187, 151, 148] (see section 1.7 below) and appears to have some sequence similarity with single–stranded DNA binding proteins [148].

1.6 Late and Very Late Gene Expression

By definition, late genes are those genes transcribed at the same time or shortly after the initiation of viral DNA synthesis [302]. Late gene expression is dependent on viral DNA replication and is inhibited when DNA replication is blocked by aphidicolin [73]. Concomitant with the initiation of viral DNA replication, is the appearance of an α–amanitin– and tagetoxin–resistant RNA polymerase [91, 82] responsible for the transcription of late, and the hyperexpressed late genes, *p10* and polyhedrin (*polh*). This late RNA polymerase initiates transcription within a late promoter element a/g/t/TAAG [261, 17] and has a unique subunit composition suggesting that it is a virally–encoded or a virus–modified host RNA polymerase [333]. During the late phase of viral transcription, steady-state levels of several host nuclear transcripts including actin, histone H2A, H3, and H4 and heat shock protein 70, decrease [225]. The mechanism of inhibition of host RNA synthesis is not understood. However, reduction in the steady–state levels of host mRNA appears to require a late viral protein [225]. The inhibition of host cell mRNA synthesis has been extensively studied in poliovirus–infected cells. Dasgupta and co-workers [265, 45, 54] have shown that the mechanism of inhibition involves cleavage of the TBP by the 3C protease encoded by poliovirus.

A number of baculovirus genes required for late and very late gene expression were identified by subtracting clones from an AcMNPV genomic library and assaying for the ability to transactivate late or very late promoter–containing reporter plasmids in transient expression assays [232]. This assay has lead to

the identification of eighteen genes involved in late and very late gene expression including *ie-1*, *ie-2*, *lef-2*, *lef-1*, *lef-3*, *lef-4*, *lef-5*, *p143*, lef-6, *lef-7*, *lef-8*, *lef-9*, *lef-10*, *lef-11*, *p47*, *p35*, *39K*, *lef-9*, *lef-10* and *dnapol* [186, 232, 230, 231, 177, 233, 234, 303, 208]. Some of these genes have been shown to be essential for replication while others appear to be required for late and very late gene expression (see section 1.7) [187, 148, 151].

Several of the genes identified using the subtraction transient–expression assay were previously also identified using temperature–sensitive (*ts*) viral mutants. Characterization of viral ts mutants in the *p47* [31], and *ie-1* [256] genes indicated that these genes were involved in regulating viral gene expression late in infection while *p143* [184] was defective in viral DNA replication.

Hyperexpressed late genes are those genes whose transcript levels increase and remain high even after the levels of expression of other late genes decreases [16]. These very late genes, *p10* and *polh* are unique to baculoviruses, having extremely abundant transcripts resulting in abundant production of the protein product. Cis–acting elements of late and very late promoters have been characterized by mutational analysis. Linker scanning analysis of the *vp39* late and the *polh* very late baculovirus promoters indicate that the sequence TAAG (where transcription initiates) is essential for promoter activity [226, 206, 251]. Additionally, the *vp39* promoter requires 8 bp upstream and 6 bp downstream of the TAAG sequence to maintain regulation [206]. However, *polh* expression requires the region encompassing 50 bp upstream from the TAAG transcriptional initiation site to the translation initiation codon indicating that the untranslated leader of *polh* transcripts is necessary for hyperexpression [226]. Characterization of the *p10* promoter region suggested that although the requirement for the TAAG transcription initiation site is similar, competition experiments and subtle differences in timing of expression indicate the regulation of *p10* may differ from *polh* [33, 260, 307].

In addition to characterizing the cis–acting promoter elements, several factors involved in very late gene expression have been identified. A very late expression factor, *vlf–1*, that regulates the expression of both the polyhedrin and to lesser extent the *p10* gene [199] was identified by characterizing a viral *ts* mutant. The deduced amino acid sequence of *vlf–1* had two regions that showed some sequence similarity to integrases and resolvases [199]. Gel retardation assays using nuclear extracts from AcMNPV–infected cells have identified a 30–kDa phoshorylated host–encoded protein that binds to the hexanucleotide sequence AATAAA immediately upstream of the TAAG transcription initiation site [26]. These proteins and others may be involved in the regulation of baculovirus very late gene expression.

1.7 Viral DNA Replication

Baculovirus origins of replication have been identified from examination of defective interfering particles (DIs) and by infection–dependent transient replication assays (for review of baculovirus replication see [147]). Undiluted serial passage of AcMNPV in insect cells results in the production defective interfering particles containing viral genomes with large deletions. Presumably, the cis–acting sequences necessary for DNA replication and packaging are retained in these particles. The presence of supermolar *Eco*RI fragments that hybridized to regions containing the *hr* sequences [149] suggested that *hr*s function as origins of replication. However, in other studies using DIs, Lee and Krell [166, 165] found multiple repeats of a non–*hr* containing sequence from the *Hind*III-K region; this sequence contained many unusual features including direct and inverted repeats and imperfect palindromic sequences.

A *Dpn*I assay [239] was used to test the ability of cloned baculovirus sequences to undergo replication when transfected into infected insect cells [235, 149, 152, 237, 5, 236]. Experiments using AcMNPV DNA indicated that plas-

mids containing *hr*s are capable of replication. The AcMNPV genome contains seven *hr*s (*hr*1, *hr*1a, *hr*2, *hr*3, *hr*4a, *hr*4b and *hr*5 that are distributed throughout the viral genome. Each of the *hr*s contains two to eight 30–bp imperfect palindromes with naturally occurring *Eco*R1 sites at their cores [7, 102, 98]. An eighth *hr* containing a single palindrome that lacks the *Eco*R1 site has recently been reported but has not been functionally characterized [7]. The relative levels of replication appear to be independent of the number of palindromes [170] but deletion of palindromes within an *hr* has been shown to affect replication efficiency [235].

Sequences similar to the AcMNPV *hr*s have been found in a number of other baculoviruses including BmNPV [191], *Choristoneura fumiferana* MNPV [161], OpMNPV [300] and *Lymantria dispar* MNPV (LdMNPV) [237]. One of the OpMNPV *hr*s has been shown to act as an enhancer [300] and function as a replication origin when linked to flanking sequences [5]. Two of the LdMNPV *hr*s functioned as origin of replication when linked to a second AT–rich domain that contains a 6–10 bp *Nru*I–containing palindrome [237]. Taken together, these experiments suggest that palindromic sequences may function as origins of replication in a number of baculoviruses.

In order to identify the baculovirus genes involved in replication, overlapping cosmid clones representing the entire genome were transfected together with an origin–containing reporter plasmid into uninfected insect cells [151, 148, 2, 187]. Kool *et al.* [148] showed that *lef–1*, *lef–2*, *p143*, *lef–3*, *ie–1*, and *dnapol* were essential while *p35*, *ie–2* and *pe38* were stimulatory. In another study, Lu and Miller [187] showed that *lef–1*, *lef–2*, *p143*, *lef–3*, *ie–1*, and *p–35* were essential while *lef–7*, *ie–2* and *dnapol* were stimulatory. The differences in these results may reflect the different assay conditions; under certain conditions host–encoded functions may act in combination with the supplied baculovirus genes to replicate the reporter plasmid to detectable levels.

Functions have been ascribed to several of the gene products identified as essential or stimulatory for baculovirus replication. A number of baculoviruses have been shown to encode a DNA polymerase gene, *dnapol*, that shares the common sequence motifs that are conserved within polymerases [14, 304, 52]. *p143* contains an NTP binding site and DNA/RNA unwinding motifs that are often associated with helicases [184]. The *ie–1* gene product has been shown to function as a transactivator [102, 98, 218, 28, 185], bind to *hrs* [96, 97, 169], and contain a single–stranded DNA binding motif [148], suggesting that IE–1 is an origin binding protein. *Lef–1*, *lef–2*, and *lef–3*, were originally characterized as activators of late gene expression [230, 232, 177]. However, recent investigations suggest that these gene are directly involved in replication although the functions of *lef–1* and *lef–2* are unknown [187, 148, 2]. *Lef–3* has been shown by functional assays to encode a single–stranded DNA binding protein [109]. The *p35* gene product likely functions to inhibit apoptosis [46] while *ie–2* and *pe38* may stimulate expression of baculovirus replication genes. Lastly, *lef–7*, contains two single–stranded DNA binding motifs and has 21% sequence similarity to the HSV–1 UL 29 gene that encodes a single–stranded DNA binding protein suggesting that *lef–7* is a single–stranded DNA binding protein.

1.8 Objectives

To learn more about the processes of transcription in baculovirus–infected insect cells, a host–encoded DNA binding protein, the TATA–binding protein, was isolated, sequenced and its expression in response to baculovirus infection characterized. Deletions of a baculovirus late gene promoter, the *vp39* promoter, were constructed to determine the minimal sequence necessary for late expression using an *in vitro* transcription system. The minimal late sequence was then used in gel–retardation assays in an attempt to isolate DNA–binding proteins specific for baculovirus late promoters. The DNA–binding activities of the baculovirus–

encoded protein, IE–I, was investigated. The specificity and structure of the DNA sequence that binds IE–1 was also characterized.

Chapter 2
Materials and Methods

2.1 Molecular Biology Products

Nucleotides were purchased from Pharmacia Biotech Inc. (Alameda, CA) or Boehringer Manneheim (Indianapolis, ID). DNA and RNA modifying enzymes and restriction endonucleases were purchased from GIBCO–BRL (Gaithersberg, MD), New England Biolabs (Beverly, MA) and Promega Corporation (Madison, WI). Radiolabeled nucleotides were purchased from DuPont NEN (Wilmington, DE).

2.2 Chemical Reagents

Chemicals were purchased from Sigma Chemical Company (St. Louis, MO), Fisher Scientific (Pittsburgh, PA), Bio–Rad Laboratories (Richmond, CA) and J.T. Baker Chemical Company (Phillipsburg, NJ). Bacto–agar, Bacto–yeast extract and Bacto–tryptone were purchased from Difco Laboratories (Detroit, MI). 4–MUB–β–D–Glucuronide was purchased from New Jersey Lab and Glove Supply (Livingston, NJ).

2.3 Virus and Cell Culture

Spodoptera frugiperda cells (Sf9, ATCC CRL 1711) were grown to density of 2 $\times 10^6$ cells/ml in 75 or 150 ml of serum-free Sf900II medium (GIBCO–BRL) in 250 or 500 ml sterile disposable Erlenmeyer flasks (Corning) on an orbital shaker (VWR Scientific model 2001) at 135 rpm. Sf9 cells were also grown in monolayers in TMN–FH medium [284] supplemented with 10% fetal bovine serum (FBS), penicillin G (50 U/ml), streptomycin (50 μg /ml) and fungizone (Amphotericin B 375 ng/ml) in T–75 or T–150 canted–neck flasks (Corning).

Lymantria dispar cells (Ld652–Y) [248] were grown in monolayers in TMN–FH media to mid–log densities of 2–3$\times 10^6$ cells/ml. Both cell types were incubated in a VWR Scientific model 2020 incubator at 27°C without CO_2.

Sf9 cells were infected with AcMNPV (E–2 strain) [284] obtained from Dr. Loy Volkman. Ld652–Y cells were infected with the OpMNPV isolate described by Leisy *et al.* [172].

2.4 Growth and Infection for Time–Course Experiments

Sf9 cells [295] grown in serum–free Sf900II medium (GIBCO-BRL) to density of 2 $\times 10^6$ cells/ml and harvested by centrifugation at 1500 rpm as described [83], or infected with AcMNPV at a multiplicity of infection (moi) of 10 and harvested as above at the appropriate times post-infection (p.i.). *Lymantria dispar* cells (Ld652–Y) were grown in monolayers in TMN–FH media to mid–log densities of 2–3$\times 10^6$ cells/ml or infected with OpMNPV at a moi of 10 and harvested at the appropriate times p.i. Both uninfected and infected Sf9 and Ld652–Y cells were then frozen in liquid nitrogen and stored at –80°C until the cells were subjected DNA (section 2.13), RNA (section 2.14) or protein (section 2.15) extraction procedures.

2.5 Sf9 cDNA Library

A Sf9 unidirectional prokaryotic–expression cDNA plasmid library, constructed in pCDNAII using size–selected Sf9 mRNA, was purchased from Invitrogen.

2.6 Monoclonal and Polyclonal Antibodies

Monoclonal antibodies, 58C9 and 16E8, made against the highly conserved C–terminal domain of TBP from *Drosophila melanogaster*, were a generous gift from Dr. Robert Weinzierl. IE–1 polyclonal antibodies were a generous gift from Dr.

Plasmid Name	Vector	Gene Promoter	Insert Size	Source
pCR1	pKS–	AcMNPV 39k	384 bp	CR
pCR2	pKS–	AcMNPV 39k	375 bp	CR
pCR3	pKS–	AcMNPV vp39	479 bp	CR
pCR4	pKS–	AcMNPV vp39	479 bp	CR
pCR5	pKS–	AcMNPV vp39	191 bp	CR
pCR5 deletions	pKS–	AcMNPV vp39	see Chapter 5	CR
PCR6	pKS–	AcMNPV vp39	396 bp	CR
PCR7	pKS–	AcMNPV vp39	24 bp	CR
PCR7–M	pKS–	AcMNPV vp39	24 bp	CR
PCR8	pKS–	AcMNPV p10	500 bp	CR
PCR9	pKS–	AcMNPV p10	380 bp	CR
pCG19A	pBS–	OpMNPV p26	bp	CHG
pAcIE–1	pBS–	AcMNPV ie–1	2100 bp	LAG
pBG1	pKS–	AcMNPV 39K	1000 bp	BG
pBG30	pKS–	AcMNPV vp39	500 bp	BG

Table 2.1. AcMNPV and OpMNPV Promoter Constructs. Sources of the plasmids were: CR, Charlotte Rasmussen; CHG, Christian H. Gross; BG; Barbara Glocker; LAG, Linda A. Guarino (Texas A and M University).

Claude Delsert [223]. OpMNPV polyhedrin rabbit polyclonal antibodies are described elsewhere [263].

2.7 Construction of Plasmids used in this Study

The AcMNPV fragments used for plasmid construction were derived from an AcMNPV–plasmid library obtained from the laboratory of Dr. M. D. Summers. The OpMNPV restriction fragments used for plasmid construction were derived from an OpMNPV–cosmid library [171]. The AcMNPV and OpMNPV promoter–containing constructs are listed in Table 2.1.

The *hr*1a derivatives and OpE–derived constructs are listed in Table 2.2. The construction of several of these vectors is described below. The AcMNPV and OpMNPV gene constructs used in this study are described in Table 2.3.

pBKS–(–Eco), a pBKS– vector which had the polylinker *Eco*RI site de-

stroyed (obtained from Dr. D. Leisy) was used to construct the *hr*1a derivative clones. pHr1a was constructed by isolating the 397 bp *Acc*I fragment from pHdN containing the *hr*1a region, treating it with T4 DNA polymerase to make it blunt–ended, and cloning it into *Sma*I–digested pBKS–(–Eco); pHr1ΔEco, containing one complete palindrome (the left half of the left palindrome ligated to the right half of the right palindrome), was constructed by digestion of pHr1a with *Eco*R1 and religation; pHr1aΔEcoFI (four base duplication) and pHr1aΔEcoCB (four base deletion) were constructed by digesting pHr1aΔEco with *Eco*R1 and filling in the recessed 3′–termini with the large fragment of DNA polymerase I (Klenow) or removal of the protuding 5′–termini with S1 nuclease, respectively, before religation. All clones containing mutations within the *hr*1a region were confirmed by DNA sequence analysis.

The β–glucuronidase (GUS) reporter vector, p39KGUSpl18, was used for measuring the enhancer effects of DNA fragments cloned into this plasmid. It was constructed by Dr. Douglas Leisy from p10–GUS+ [89] and pHR39k [83] as follows: p10–GUS+ was digested with *Bam*HI and the 5′–overhangs were end–filled using T4 DNA polymerase, followed by digestion with *Hind*III, and the plasmid pHR39k was digested with *Sac*I, the 3′–overhangs were removed with T4 DNA polymerase and then the DNA was then digested with *Hind*III. The 5.2–kb *Bam*HI–*Hind*III fragment from p10–GUS+, which contains the GUS gene, the 3′–end of the OpMNPV *p10* gene and flanking region, and the pBluescribe (Stratagene, Inc.) vector, was ligated with the 1–kb 39K promoter–containing *Sac*I–*Hind*III fragment from pHR39k to form p39KGUS. To obtain a set of restriction sites useful for subsequent subcloning of DNA fragments into the above described plasmid, the 1–kb *Sca*I–*Hind*III fragment of pUC18, which contains the complete pUC18 polylinker, was used to replace the 1–kb *Sca*I–*Hind*III fragment of p39KGUS, forming p39KGUSpl18.

Plasmid Name	Vector	Insert	Number of Palindromes	Insert Size	Source
pHr1a or pKSE–Acc3–1	pKS–	hr1a	2	397 bp	DJL
pCR–11 or pHrΔEcoFI	pKS–	hr1a	1	316 bp	CR
pCR–12 or pHr1aΔEco	pKS–	hr1a	1	313 bp	CR
pCR–12AT4	pKS–	hr1a	1/2	237 bp	JTE
pCR–12AS1 or pHr1aΔEco.5RB	pKS–	hr1a	1/2	233 bp	JTE
pCR–12BT4	pKS–	hr1a	1/2	76 bp	JTE
pCR–12BS1 or pHr1aΔEco.5LB	pKS–	hr1a	1/2	71 bp	JTE
pCR–13 or pHr1aΔEcoCB	pKS–	hr1a	1	305 bp	CR
pCR–14a	pks–	hr1a	1	321 bp	CR
pCR–14b	pks–	hr1a	1	321 bp	CR
pCR–14b30E or pHr1aΔEco21L	pKS–	hr1a	1	251 bp	CR
pCR–14a7510	pKS–	hr1a	1	118 bp	CR
pCR–14a906	pKS–	hr1a	1	109 bp	CR
pCR–14a9011 or pHr1aΔEco18R	pKS–	hr1a	1	99 bp	CR
pHr1aGUS	p39KGUSpI18	hr1a	2	397 bp	DJL
pCR–15 or pHr1aΔEcoGUS	p39KGUSpI18	hr1a	1	313 bp	CR
pCR–16	pKS–P+Sal	0	0	6 bp	CR
pCR–17b30E	pSK–	hr1a	1	251 bp	DJL
pCR–18–1	pKS–	hr1a	2 x 1/2	92 bp	CR
pCR–18–2	pKS–	hr1a	1 + 2 x 1/2	184 bp	CR
pCR–18–3	pKS–	hr1a	2 + 2 x1/2	276 bp	CR
pCR–18–4	pKS–	hr1a	3 + 2 x 1/2	368 bp	CR
pCR–20 or pHrIRGUS	p39KGUSI18	hr1a	2 x 1/2	92 bp	CR
pCR–SAB	pKS–	perfect	1	42 bp	CR
pCR–SCD	pKS–	consensus	1	42 bp	CR
pCR–HAB	pKS–P+Sal	perfect	1	42 bp	CR
pCR–HCD	pKS–P+Sal	consensus	1	42 bp	CR
pOPE–NB	pKS–	OpE	2 1/2	170 bp	DJL
pOPE–EN	pKS–	OpE	8	580 bp	DJL
pHr–imperfect	pKS–	consensus	1	42 bp	CR
pHr–perfect	pKS–	perfect	1	42 bp	CR
pKS–F	pKS–	none	none	-40 bp	CR
pHr–imperfectGUS	p39KGUSI18	consensus	1	42 bp	CR
pHr–perfectGUS	p30KGUSI18	perfect	1	42 bp	CR

Table 2.2. AcMNPV and OpMNPV Origin Constructs. Sources of the plasmids include: CR, Charlotte Rasmussen; JTE, Jay T. Evans and DJL, Douglas J. Leisy

Plasmid Name	Vector	Gene	Source
Rep1	pKS–	*lef-1*	DJL
Rep2	pKS–	*lef-2*	DJL
Rep3	pKS–	*lef-3*	DJL
Rep4	pKS–	*dnapol*	CHA
Rep5	pBS–	*p143*	CHA
Rep6 or pAcIE–1	pKS+	*ie-1*	DJL
Rep7	pKS–	*p35*	DJL
Rep8	pKS+	*ie-2*	DJL
Rep9	pKS–	*pe38*	DJL
pOpIE1	pBS–	*ie-1*	CHA
pOpIE2	pBS–	*ie-2*	CHA

Table 2.3. Replication Gene Constructs. Source of plasmid constructs include: DJL, Douglas J. Leisy and CHA, Christian H. Ahrens

2.8 Propagation of Plasmids in Bacteria

Plasmids were propagated in *E. coli* strain DH5α. Competent DH5α cells were prepared as described by the calcium chloride [269] or the rubidium chloride method [192] and transformed as described by Sambrook *et al.* [269].

2.9 Isolation of Plasmid DNA from Bacteria

Plasmids were isolated from DH5α using either the rapid–boiling method [121], alkaline–lysis procedure [12], cesium chloride (CsCl) centrifugation [269], polyethylene glycol (PEG) precipitation [216] or on Qiagen columns (Qiagen, Inc.) depending on the amount of supercoiled plasmid required and the level of contamination by chromosomal DNA, RNA and proteins that could be tolerated in subsequent manipulations. Generally, the rapid–boiling method was used without further purification to screen for plasmid constructs during cloning procedures. Alkaline–lysis preparations were used to prepare plasmid DNA for sequencing reactions. Both CsCl centrifugation and Qiagen columns were used

to isolate highly–purified supercoiled plasmid DNA for *in vitro* transcription reactions and transfections.

2.10 Polymerase Chain Reaction (PCR) Amplification

PCR was performed using 1 μg of DNA isolated using the alkaline–lysis procedure (see section 2.9) from the pCDNA II plasmid library that had been amplified in liquid culture. The pCDNA II plasmid library was amplified by addition of 200 μl of the pCDNA II library glycerol stock (1.0×10^10 colonies/ml) to 500 ml of LB Broth (10 g Bacto–tryptone, 5 g Bacto–yeast and 10 g NaCl per litre) containing 50μg /ml ampicillin followed by incubation on an orbital shaker at 37°C for 6 hours. A typical 100 μl PCR reaction contained 100μM of each deoxynucleotide triphosphate, 10μg of each primer, 10 mM Tris–HCl [pH 9.0 at 25°C], 50 mM KCl, 1.5 mM $MgCl_2$, .01% gelatin (w/v), 0.1% Triton X–100 and 3 units of Ampli–Taq DNA polymerase (Perkin–Elmer). DNA amplification was achieved by 35 cycles of denaturation at 95°C for 1 min, primer annealing at 48°C for 1 min, and extension at 70°C for 1 min.

The sequence of the degenerate primers that yielded specific amplified product were: primer 1, 5′–ATAGGATCCAAYGCNGARTAYAAYCCNAA–3′, and primer 2, 5′–ATAGGAACCNCCNACCATRTTYTGDATYTT–3′ (see Figure 3.3). Primers were chosen by alignment of the conserved carboxy–terminal region of previously determined TBP sequences, determination of small conserved regions containing amino acids with few synonyms and utilizing *Drosophila* codon biases. The 216 bp amplified fragment was cloned into the *Sma*I restriction endonuclease site of the vector pBS– (Stratagene).

2.11 Colony Hybridization

Approximately 2.5×10^5 recombinants from the Sf9 pcDNAII library were screened using a colony hybridization protocol [92] modified by Dr. I. Rajagopal (personal communication). Briefly, bacterial colonies were allowed to grow to the size of pinheads, usually 6–8 hr, at 37°C on YT (8 g Bacto–tryptone, 5 g Bacto–yeast, 5 g NaCl, 20 g agar per litre) medium. The plates were incubated at 4°C for 2–12 hr and then marked with asymmetric dots before a circular nylon membrane (MSI) was situated on the plate. The membranes were marked with a needle point corresponding to the asymmetric dots and lifted off the plate and positioned on a fresh plate (colony side up). The colonies were allowed to grow on the membrane to the size of small pinheads (up to 4 hr) before lysis. Four to six ml of 0.5 M NaOH/1.5 M NaCl was placed on plastic wrap and the membrane placed on top (colony side up) for 7–9 min. The membranes were neutralized by 2 successive transfers to 0.5M Tris-HCl [pH 7.4]/ 1.5 M NaCl for 5–7 min. Bacterial debris was removed from the membranes by immersion and gentle rocking in wash solution (0.1 M LiCl, 67% ethanol, 0.03 M Tris–HCl [pH 7.4]). Membranes were air dried and then either baked in a vacuum oven at 80°C for 1–2 hr or UV crosslinked using UV Stratalinker 1800 (Stratagene, Inc.).

Prehybridization and hybridization with a ^{32}P–labeled riboprobe [90] made from the clone containing the 216 bp PCR product (see section 2.10) were performed as described in section 2.14 below. Positive clones were re–screened a second and third time.

2.12 Unidirectional Digestion with Exonuclease III in DNA Sequence Analysis

The clone containing the TBP cDNA sequence, pSfTBP, was digested with the restriction endonucleases, *Kpn*I and *BamH*I and then incubated with exonuclease III to produce a unidirectional set of nested deletions [113]. A nested set of deletions in the opposite direction was created by digestion with *Apa*I and *Xho*I followed by treatment with exonuclease III.

Exonuclease III was used to create 5′ and 3′ deletions of one of the *vp39* late promoter region. pCR5 was digested with *Apa*I and *Xho*I to delete from the 5′ end of the proximal ATAAG and *Sst*I and *Xba*I were used to delete from the 3′ region.

Subsequent double–stranded dideoxy sequencing [305] of the exonuclease III generated deletion clones was carried out using Sequenase (United States Biochemical) or Taq polymerase (Promega) sequencing kits as per manufacturer's instructions. Typically, 200–300 nucleotides could be sequenced using Taq polymerase while 150–200 nucleotides could be determined using Sequenase.

2.13 Southern Analysis

Extraction of genomic Sf9 DNA was performed as described by Summers and Smith [295]. Southern analysis was performed as follows. Restriction endonuclease–digested genomic DNA was electrophoresed on a 1% agarose gel in 1 X TBE (89 mM Tris–borate, 89 mM boric acid and 2 mM EDTA) buffer and blotted to a Gene Screen Plus membrane (Dupont) in 10 X SSC (1.5M NaCl, 0.15M sodium citrate). A restriction fragment containing the full–length SfTBP cDNA sequence was radioactively labeled [65]. Prehybridization was carried out in 0.25 M $NaHPO_4$ [pH 7.2], 7% $NaDodSO_4$ and 100μg /ml of salmon sperm DNA at 65°C for 3 hrs to 6 hrs [44]. 1–2 $\times 10^6$ cpm per ml of radiolabeled probe was

added to the prehybridization solution and hybridization was carried out for 12–16 hrs at 65°C . Both high and low stringency blots were initially washed with 2 X SSC and 1% SDS at room temperature for 15 min. Blots were then washed at 65°C for 30 min in 0.2 X SSC, 0.1% SDS (high stringency) or at 42°C for 30 min (low stringency). Autoradiograms were prepared using Kodak XAR–5 film and Dupont intensifying screens.

2.14 Northern Analysis

RNA was isolated from uninfected and AcMNPV–infected Sf9 cells by the guanidine isothiocyanate–cesium chloride method [40]. Total RNA was resolved by electrophoresis through 1.2% agarose–formaldehyde gels and transferred to a Gene Screen Plus membrane (Dupont) as described [269]. The BRL 1 kb DNA ladder was end–labeled with α–^{32}P–dATP [269] and 100,000 cpm were electrophoresed adjacent to the RNA samples. Strand–specific cRNA (riboprobes) complementary to the open reading frame of the gene of interest (see Figure legends) were made as described [90]. Prehybridization and hybridization conditions were identical to those described for Southern analysis (see section 2.13). Following hybridization, membranes were then washed once with 2 X SSC, 1% SDS at room temperature for 15 min and once with 0.2 X SSC, 0.1% SDS at 65°C for 20 min. Autoradiograms were prepared as described above (see section 2.13).

2.15 Western Blot Analysis

Protein extraction and western blot analysis of total protein isolated from uninfected and infected insect cells was performed as described [248]. 4.0 $\times 10^4$ cell equivalents of protein was loaded per lane, electrophoresed, and blotted. An identical gel stained with Coomassie brilliant blue showed the typical pat-

tern of protein expression seen during both AcMNPV infection of Sf9 cells and OpMNPV infection of Ld652–Y cells.

2.16 Phylogenetic Analysis

Phylogenetic analysis was performed using the Genetic Data Environment (GDE). Alignments of the carboxy–terminal amino acids of TBP were generated using clustal protein alignment [115]. Phylogenetic relationships were inferred by a distance–matrix method [296] using a Sun 6/670 computer and the treetool program.

2.17 Nuclear Extract Preparation

Nuclear extracts were prepared from uninfected or AcMNPV–infected Sf9 cells at as described by Glocker *et al.* [82] with minor modifications. Sf9 cells were harvested by centrifugation for 7 min at 1500 rpm in a Beckman GP centrifuge at room temperature. All subsequent steps where carried out at 4°C . The harvested Sf9 cells were resuspended in 4 packed–cell volumes of Buffer A (10 mM Tris–HCl [pH 7.9], 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT), incubated on ice for 10 min and then lysed by 10 strokes of a Kontes homogenizer (B pestle). Nuclei were pelleted by low-speed centrifugation (1/2 speed in a clinical centrifuge) and resuspended in 1 nuclear–pellet volume of a 9:1 mixture of Buffer C (20 mM Tris–HCl [pH 7.9], 25% glycerol, 420 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM DTT) and a saturated solution of $(NH_4)_2SO_4$. The nuclei were lysed with 15 strokes of a Kontes homogenizer (B pestle), transferred to ultracentrifugation tubes and rocked slowly (1/2 speed on Hoefer Scientific Instruments Red Rocker) for 30 min. After ultracentrifugation in a Beckman 100.2 rotor for 1 hr at 100,000 X *g* at 4°C , the supernatant was removed. An equal volume of saturated $(NH_4)_2SO_4$ solution was added to the supernatant

in 100 μl aliquots with gentle mixing. The milky–white supernatant was then rocked on ice for 30 min and then centrifuged in a Beckman 100.2 rotor at 15,000 X *g* for 30 min. The pellet was dissolved in dialysis buffer (20 mM Tris–HCl, [pH 7.9], 100 mM KCl, 20% glycerol, 0.2 mM EDTA, 10 mM 2–mercapto–ethanol) and dialyzed (> 50 volumes of dialysis buffer per volume of nuclear extract; mw cut-off of dialysis membrane was 3500 Da) for 3 hr with two buffer changes. After a 3 min centrifugation at 10,000 rpm in a microfuge, the nuclear extract was aliquoted, frozen in liquid nitrogen and stored at –80°C . Protein concentrations of the nuclear extracts were determined using the Bradford protein assay [23].

2.18 *In vitro* Transcription Reactions

In vitro transcription reactions were performed as described [122, 83, 82]. Final concentrations of reaction components for *in vitro* transcription of baculovirus early promoters were 3 mg/ml nuclear extract, 20 mM Hepes [pH 8.4 at 25°C], 6 mM $MgCl_2$, and 37.5 μg /ml of DNA template in a total volume of 20 μl . Reaction conditions were modified to support *in vitro* transcription from baculovirus late promoters as follows: the addition of 50 μg /ml of α–amanitin and 20 units of RNasin (Promega), and reduction of $MgCl_2$ concentration to 2mM. Transcription was initiated by the addition of nucleotides (600 μM each of ATP, CTP and GTP; 25 μM UTP; 5 μCi α–^{32}P–UTP) followed by incubation at 30°C for 30 min. DNA template containing baculovirus early promoters was preincubated with nuclear extract from uninfected Sf9 cells for 25 min at 30°C whereas the nuclear extracts from 16 hr p.i. (or later) and baculovirus late promoter templates were not subjected to preincubation. Reactions were stopped by the addition of 75 μl stop buffer (0.25% SDS, 5 mM EDTA, 50 mM sodium acetate [pH 5.2], 1 mg *E. coli* tRNA/ml), extracted once with an equal volume of 1:1 phenol/chloroform [pH 5.0] and precipitated by the addition of 2.5 volumes of 95% ethanol.

After incubation for 2–12 hr at -20°C , the RNA was pelleted by microcentrifugation for 30 min at 4°C . The pellets were air–dried for at least 20 min and resuspended in 15 μl of running dye (98% formamide, 2% xylene cyanole FF, 2% bromophenol blue). Samples were electrophoresed in 7 M urea–5% acrylamide gels for 35 min at 175 volts on a Bio–Rad minigel apparatus or for 1.5 hr at 550 volts on a Bio–Rad Protean II apparatus. *Hae*III restriction fragments of ϕX174 were radiolabeled using T4 DNA polymerase [192] and used as size standards. Gels were dried under vacuum and exposed using Kodak XAR film for 12–24 hr with a Dupont intensifying screen.

2.19 Primer Extension Analysis

Transcription start sites of the *in vitro* transcribed RNA were confirmed by primer extension analysis as described [122, 83, 82]. For this analysis, the *in vitro* transcribed RNA was prepared as described in section 2.18 except that radiolabeled UTP was omitted and the UTP concentration in the reaction mixture was increased to 200 μM . The forward (5′–GTAAAACGACGGCCAGT–3′), or reverse (5′–TCACACAGGAAACAGCTATGAC–3′), sequencing primers were used depending on the baculovirus promoter construct being assayed.

2.20 Gel Retardation Analysis of the *vp39* Late Promoter

Oligonucleotides were labeled using γ–^{32}P–dATP and T4 polynucleotide kinase [269] and then purified on G–50 Sephadex (Pharmacia) spin columns. Radiolabeled oligonucleotides were annealed by heating the oligonucleotides to 95°C for 5 min in 50mM Tris–HCl (pH 7.5), 50 mM NaCl, 1mM EDTA and placing the eppendorfs tubes in a beaker of water that had been heated to 75°C . The water in the beaker and consequently the oligonucleotides were allowed to cool to room temperature before use. The annealed radiolabeled probes (10,000

cpm) were incubated for 20 min at 4°C with the amount of extract indicated in a total volume of 20 μl containing. The radiolabeled probe and non–specific and specific competitor DNA were added simultaneously to the reaction mixtures. DNA–protein complexes were resolved by electrophoresis on 4.2% polyacrylamide 3% glycerol gels in 8.0 mM Tris–HCl [pH 7.9], Borate, 1 mM EDTA. Electrophoresis was carried out at 200 volts for 3–5 hr at 4°C .

2.21 Transfections

Log–phase Sf9 cells were seeded into six-well plates (1.25 x $\times 10^6$ cells/well) and allowed to attach for 4 to 12 hr. After the cells were attached, the medium was removed and replaced with 0.4 ml of Grace's Insect medium (GIBCO–BRL) with 10% FBS. An equal volume of transfection buffer (25 mM Hepes [pH 7.1], 140 mM NaCl, 125 mM $CaCl_2$–$2H_20$) containing plasmid DNA (see Figure legends for quantity of plasmid DNA transfected) was applied dropwise to the well and mixed thoroughly by gentle rotation of the titer plate. After 4 hr the transfection mixture was removed and replaced with 1 ml of fresh TNM–FH medium and incubated for 24 hr at 28°C .

2.22 Plasmid Replication Assay

Plasmid DNA was collected from infected and transfected Sf9 cells and assayed for replication by digestion with *Dpn*I [239] and the appropriate restriction endonucleases needed to separate the insert fragment from the vector. The subsequent Southern blotting and hybridization with ^{32}P–labeled pBKS– was as described [170]. Replication efficiencies were quantified with a PSI–486 Phosphorimager SI and Imagequant Workstation (Molecular Dynamics) using the Scanner Control SI–PDSI 1.0 and Imagequant 4.1 software packages.

2.23 β–glucuronidase Assay

1.25 $\times 10^6$ Sf9 cells, seeded in six-well plates, were transfected with varying amounts of pOpIe–1 or pAcIe–1 and β–glucuronidase (GUS) reporter constructs (see text and figure legends). After 24 hr, the cells were harvested by scraping, transferred to eppendorf tubes, and pelleted by centrifugation at 4000 RPM for 3 min in a microfuge. The cell pellets were resuspended in 200 ml PBS (120 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4 [pH 7.4]) and lysed by three freeze–thaw cycles. Insoluble debris was removed by centrifugation at 4000 RPM for three min and the supernatants assayed for GUS activity using the fluorogenic assay [132].

2.24 Preparation of Whole–Cell Extracts

Whole–cell extracts were prepared as described [97] with minor modifications. Briefly, 1 x $\times 10^6$ Sf9 or Ld652–Y cells were seeded in six-well plates and then transfected with 10 mg of the plasmids encoding the gene of interest as described in section 2.21. After 24 hr, the cells were transferred to eppendorf tubes and washed three times with PBS followed by centrifugation for 3 min at 2500 RPM in a microfuge. The cells were resuspended in 4 x the packed–cell volume of extraction buffer (10 mM Hepes [pH 7.2], 0.4 M NaCl, 0.1 mM EGTA, 0.5 mM DTT, 0.5 mM phenylmethylsulfonyl fluoride, 20% glycerol) and incubated on ice for 10 min. The cells were pelleted at 3000 RPM in a microfuge at 4°C and the supernatant was removed and immediately frozen in liquid nitrogen. The whole–cell extracts were stored at –80°C until needed.

2.25 Gel Retardation Assay using AcMNPV *hr* Sequences

Plasmids containing putative IE–1 binding sites were digested with the appropriate restriction endonucleases (usually sites within the polylinker region of the

vector) to liberate the DNA fragment of interest. The DNA fragments containing the region of interest were recovered from agarose gels using powdered glass [312] or from 5% non–denaturing polyacrylamide gels using the crush–soak method [192] and labeled by end–filling with either α–^{32}P–dATP or α–^{32}P–dCTP using Klenow [269]. After purification on G–50 Sephadex (Pharmacia) spin columns, the radiolabeled probes were used for gel retardation assays. Radiolabeled probes (10,000 cpm) were incubated for 20 min at 4°C with the amount of extract indicated in a total volume of 20 μl containing 10 mM Tris [pH 7.9], 100 mM NaCl, 1 mM dithiothreitol, 20% glycerol, 1 mg poly–dIdC (Sigma Chemical Co.) [96]. For the electrophoretic mobility supershift experiments, IE–1 polyclonal antibodies (a generous gift from Dr. Claude Delsert [223]), at the indicated dilutions, were added to the samples and further incubated for 10 min. DNA–protein complexes were resolved by electrophoresis on 4.2% polyacrylamide/3% glycerol gels in 8.0 mM Tris–HCl [pH 7.9], 6 mM Sodium Acetate, 1 mM EDTA. Electrophoresis was carried out at 200 volts for 3–5 hr at 4°C . The AcMNPV *p10* promoter fragment used in the competition experiments was isolated by *Xho*I digestion of a construct, pCR8, that contains the *Hinc*II–*Sst*I fragment from HindIII–Q cloned into *Hinc*II/*Sst*I–digested pBKS–. *Xho*I cleaves pCR8 once in the vector polylinker and once upstream of the *p10* promoter, producing a 297 bp DNA fragment encompassing the *p10* promoter region.

2.26 Nuclease P1 Assay

The single–stranded nuclease P1 (Pharmacia) was used to cleave supercoiled plasmids that contain hairpin or cruciform structures in a manner similar to Lilley [181]. P1 reactions were performed in 20 mM MES [pH 6.6], 0.1mM EDTA and 25 mM NaCl at 37°C for the times indicated. The units of P1 used varied between 0.3 to 1.4 and are indicated in the figure legends. After P1–digestion, the reactions were divided in half. One half of the sample was analyzed on a 1%

agarose gel while the remainder of the sample was digested with the restriction enzyme *ScaI* before electrophoresis on 1% agarose gels.

2.27 Two–Dimensional–Gel Electrophoresis of Circular DNA Topoisomers

Two–dimensional gel electrophoresis of circular DNA topoisomers was performed as described in Bowater *et al.* [22] with the following modifications. Topoisomerase I ladders of plasmid DNA were made by incubating 3.0 μg of Qiagen–prepared supercoiled plasmid DNA in 50 mM Tris–HCl [pH 7.5], 0.1 mM EDTA and 100 mM KCl in the presence of 1 μl Topoisomerase I (kindly provided by Dr. Gary Schroth) [220] and increasing amounts of a 50 μM ethidium bromide for 4 hr at 37°C . Topoisomerase I reactions were then extracted once with an equal volume of Tris–buffered phenol [pH 8.0] followed by extraction with 2 volumes of NaCl–saturated butanol. Plasmid DNA was precipitated by addition of 6 μl 5M NaCl and 250 μl of 100% ethanol, and subsequent incubation at –20°C for 30 min. The DNA was then pelleted by centrifugation at 4°C at 14,000 rpm for 30 min before being resuspended in 50 μl of TlowE (10 mM Tris-HCl [pH 7.5] and 0.1 mM EDTA). The range of plasmid DNA topoisomers obtained was evaluated by electrophorizing in TBlowE (89 mM Tris–borate, 89 mM boric acid and 0.25 mM EDTA) either on 0.9% agarose gels for 16 hours at 85 volts in the absence of intercalating agents. After electrophoresis, the gel was stained and the topoisomerase I reactions that contributed to a complete range of topoisomers, from slightly positively supercoiled to highly negatively supercoiled, were identified. These reactions were then combined and the DNA precipitated by addition of 0.06 volumes of 5 M NaCl and 2.5 volumes of 100% ethanol. The complete set of topoisomers was resuspended in 50 to 75 μl of TlowE.

Five μl of the topoisomer mix were combined with 1.25 μl of 5 X TBlowE buffer and incubated for at least 2 hr at 23 or 37°C . The sample was then

electrophoresed in the first dimension in absence of intercalating agents. Electrophoresis was carried out on 1.5% agarose gels for 30 hours [144] in TBlowE buffer at 85 volts. The gel was then turned 90°C and soaked in the indicated concentrations of the intercalating agent chloroquine for 6 hr to overnight. Electrophoresis in the second dimension was carried out for 20 hr at 85 volts with the same chloroquine concentration in the running buffer. The gel was then stained in ethidium bromide solution and photographed.

Chapter 3
The *Spodoptera frugiperda* TATA-Binding Protein

3.1 Introduction

Initiation of the baculovirus transcriptional cascade within *S. frugiperda* cells requires RNA polymerase II and its associated transcription factors [122]. One of the earliest events in the baculovirus transcriptional cascade is likely to be the binding of the TATA–binding protein (TBP) to early gene promoters. The association of the TBP–containing complex, TFIID, with the TATA–box is the first step in the assembly of an active RNA polymerase II transcription complex [25]. TBP then interacts with a variety of cellular co–activators, other general transcription factors, and the C–terminal domain of RNA polymerase II [306] to form a functional pre–initiation complex. Evidence suggests that the pre–initiation complex and specifically TBP and its associated factors (TAFs), serve as important targets for several promoter specific viral activator proteins including E1A of Adenovirus, IE2 of Cytomegalovirus, Zta of Epstein Barr virus, IE of Pseudorabies virus and VP16 of Herpes Simplex virus ([81] and references therein). Therefore, it is likely that SfTBP serves as a target for baculovirus transactivators during early RNA polymerase II–dependent transcription [83].[1]

To begin to understand the interaction of baculovirus early genes with the host cell transcriptional machinery, TBP from the *Spodoptera frugiperda* cell line, Sf9, was cloned and sequenced. To determine if the viral inhibition of host cell transcription that occurs in baculovirus–infected cells includes genes encoding transcription factors that are required for transcription of early genes, SfTBP mRNA levels were monitored during a time course of baculovirus infection using

---

[1] A modified version of this chapter was published as: C. Rasmussen and G.R. Rohrmann. (1994). Characterization of the *Spodoptera frugiperda* TATA–Binding Protein: Nucleotide Sequence and Response to Baculovirus Infection. Insect Biochem. Molec. Biol.24:699–708.

a strand–specific probe from the cloned SfTBP gene. Steady state levels of SfTBP mRNA were compared to steady state levels of selected viral and host messages. The effect of viral infection on the stability of SfTBP protein was determined by western blot analysis using anti–TBP antibody.

3.2 Isolation of a *S. frugiperda* TATA–Binding Protein cDNA Clone

The polymerase chain reaction, utilizing degenerate primers to the highly conserved carboxy–terminal region of the *Drosophila* TBP, was used to amplify a 216 bp carboxy–terminal fragment of *S. frugiperda* TBP (SfTBP) from a *S. frugiperda* cDNA library. This fragment was then cloned into the *Sma*I site of pKS– creating pSfTBP–216. A ^{32}P–labeled riboprobe [90] was made and used to screen the same Sf9 cDNA library. Ten positive clones were isolated (see Section 3.3 below). The most intense signal was from a clone containing a 1.21 kb insert. This clone, pSfTBP, was subsequently sequenced and found to be homologous to previously sequenced TATA–binding proteins. Nested sets of deletion clones were generated using pSfTBP, as described in section 2.12 to facilitate sequence analysis. Figure 3.1 illustrates the deletions used to determine the sequence of SfTBP. Sequencing of the appropriate deletions resulted in the determination of the complete sequence of SfTBP. Figure 3.2 diagrammatically depicts the overlapping sequences used to determine the complete SfTBP sequence.

The nucleotide sequence and deduced amino acid sequence are presented in Figure 3.3. SfTBP has an open reading frame of 921 nucleotides that encodes a protein with a predicted mass of 34 kDa and a predicted isoelectric point of 10.4. The sequence surrounding the initiator methionine of SfTBP conforms to the consensus sequence (CCA/GCCATGG) defined by Kozak [157], including a purine at the highly conserved –3 position relative to the start site of translation. This suggests that SfTBP transcripts are efficiently translated. A putative polyadenylation signal, AATAAA, is located near the 3′ terminus. Downstream

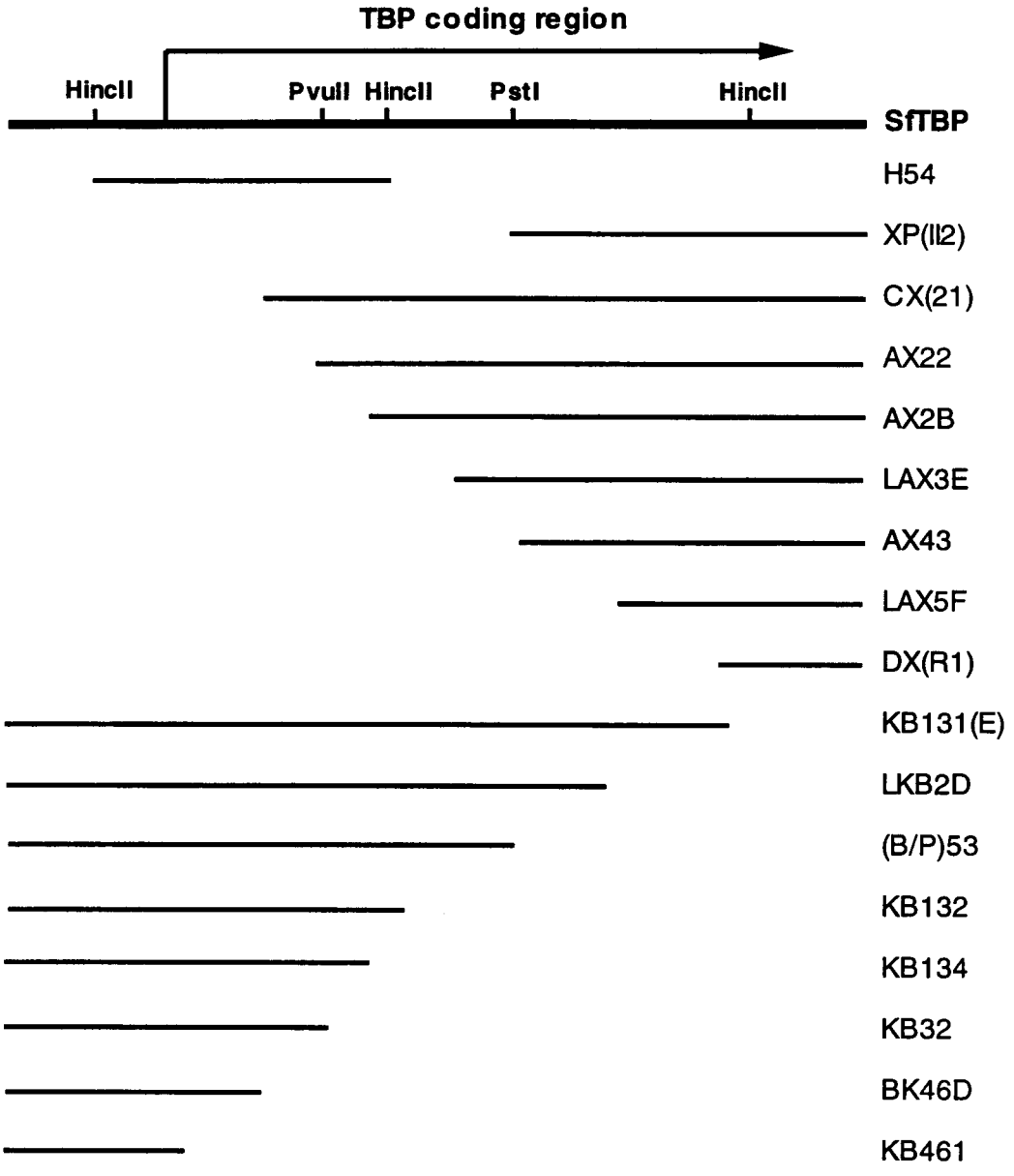


Figure 3.1. A Schematic Diagram of the SfTBP Deletion Clones used in Sequence Determination. The full length SfTBP cDNA clone is illustrated by the thick black line and the extent of the SfTBP coding region is indicated by the arrow. Representative restriction endonuclease sites are denoted.

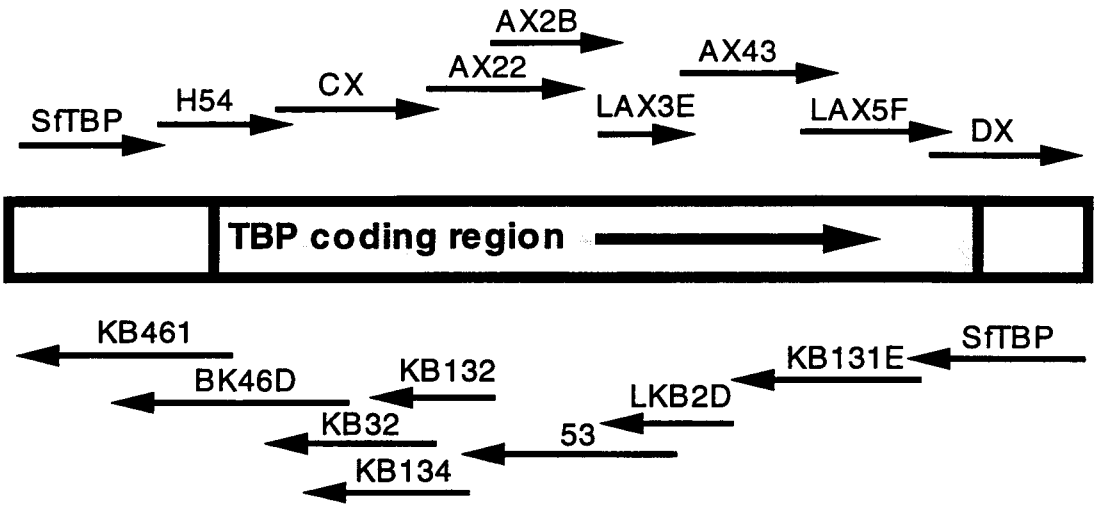


Figure 3.2. Strategy for Sequencing the SfTBP cDNA Clone. The shaded box represents the SfTBP coding region; the arrow indicates the direction of transcription. The deletion clones used to determine the complete sequence are shown. The regions of sequence overlap and the length the arrow for each deletion clone are drawn to scale and therefore are representative of the number of nucleotides sequenced.

from the polyadenylation signal are two regions that contain sequences, AT_{3-4}, thought to be involved in selective mRNA degradation suggesting that specific degradation may regulate SfTBP transcripts levels [280].

3.3 Isolation of TBP Cross–hybridizing Clones

During isolation of pSfTBP, a number of other clones produced a positive signal when probed with the pSfTBP–216 riboprobe during colony hybridization. However, none of the hybridization signals were as intense as the one generated for pSfTBP. This result is demonstrated by the accompanying Southern blot (see Figure 3.4). The intensity of cross–hybridization was determined using a PSI–486 phosphoimager. Hybridization signals from clones p227, p231, p236, p331, p334, p4152, p4153, and p5133 all were 3–5% of the control pSfTBP while clone p225 had a hybridization signal 14% of pSfTBP. The nine TBP–cross hybridizing

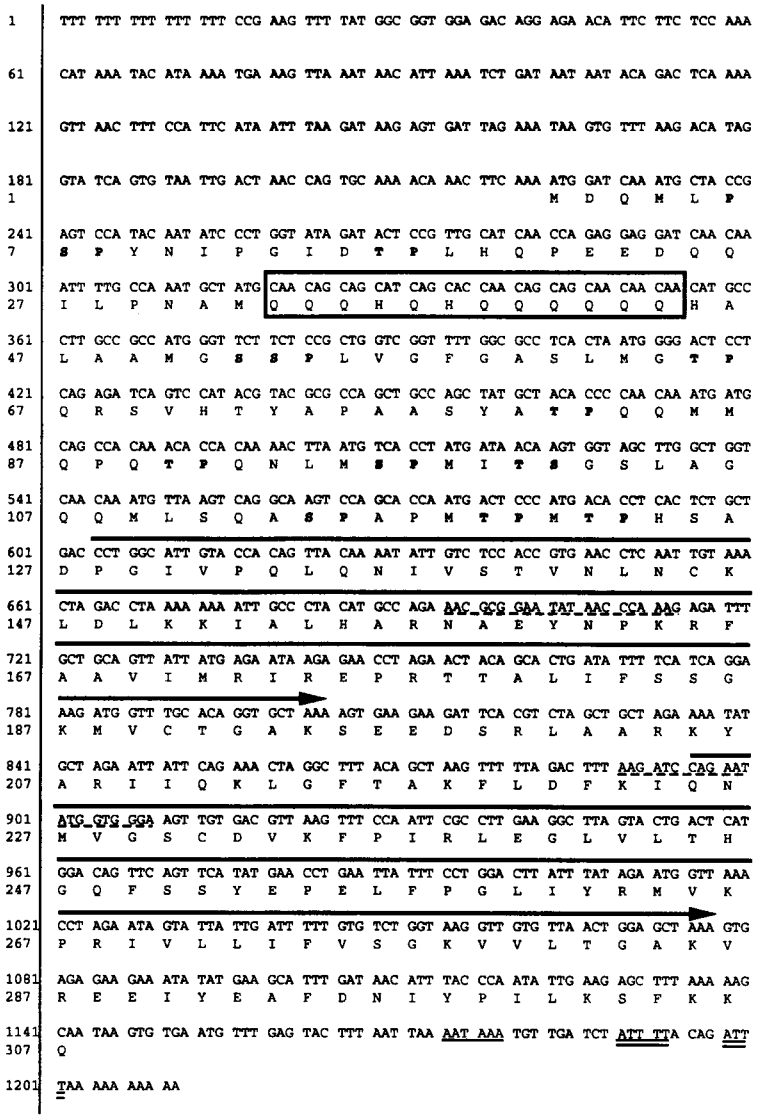


Figure 3.3. Nucleotide and Predicted Amino Acid Sequence of the *S. frugiperda* TATA–Binding Protein cDNA Clone. Shown is the cDNA sequence containing an open reading frame of 921 bp (307 amino acids) of the *S. frugiperda* TBP. Nucleotide position is shown on the left with the amino acid position shown immediately below. The carboxy–terminal direct repeats that flank a basic region are overlined by long arrows. The amino–terminal glutamine tract (Q–run) is boxed and STP sequences (see section 3.4) are indicated by bold type. The putative polyadenylation signal is underlined and the selective mRNA degradation signals are double underlined. The sequences encoded by the primers used for PCR are indicated with dashed lines. The Genbank accession number for this sequence is L22538.

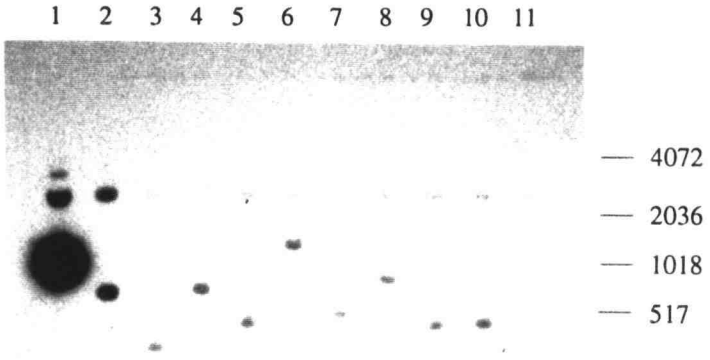


Figure 3.4. Detection of the *S. frugiperda* TBP Cross–hybridizing Clones by Southern Analysis. The plasmid DNA was digested with *Pvu*II, blotted to Gene Screen Plus (Dupont), and probed with a random-primer labeled full–length cDNA of the *S. frugiperda* TATA-binding protein using high stringency conditions. Lane 1, pSfTBP; lane 2, p225; lane 3, p227; lane 4, p231; lane 5, p236; lane 6, p331; lane 7, p334; lane 8, p4152; lane 9, p4153; lane 10, p5133; lane 11, pActin. The BRL 1 kb DNA ladder was used as a size standard. The numbers on the right indicate the sizes in bps.

clones were then characterized by digestion with 10 restriction enzymes (*Hind*III, *Eco*RI, *Xba*I, *Pst*I, *Bam*HI, *Pvu*II, *Hinc*II, *Cla*I and *Bgl*II) in order to determine whether these cross–hybridizing clones contained truncated portions of SfTBP cDNA. The restriction endonuclease data indicated none of these clones are truncated versions of SfTBP. The SfTBP cross–hybridizing clones may contain small regions of sequence similarity to SfTBP or may be cDNA clones of other *S. frugiperda* genes that are similar to TBP. Genes with considerable similarity to TBP have been characterized in *D. melanogaster* [53].

3.4 Analysis of the *S. frugiperda* TATA-Binding Protein Sequence

The sequence of cDNAs encoding TBP from a variety of organisms including *S. cerevisiae* [62, 107, 125], *S. pombe* [119], *P. carinii* [201], *D. melanogaster* [117, 209], *S. frugiperda* [252], *B. mori* [286], *H. sapiens* [241, 136], *M. musculus* [297], *M. auratus* [219], *A. castellanii* [325], *S. tuberosum* [120], *Z. mays* [104, 311], *A. thaliana* [79], *T. aestivum* [138, 6], *D. discoideum* [20], *X. laevis* [111], *C. elegans* [178], *O. vulvulus* [176], *P. falciparum* [197], *A. cliftonii* [71], *T. thermophila* [288], *E. histolytica* [228] and two archaebacteria species, *T. celer* [193] and *P. woesei* [264].

These sequences show that TBP is a bipartite molecule containing a highly conserved carboxy–terminal domain consisting of 180 amino acids and an amino–terminal domain that is highly variable in size and amino acid composition. The carboxy–terminal domain of TBP contains two direct repeats flanking a basic region. In addition, a region with similarity to prokaryotic sigma factors is located in the C–terminus of TBP. A number of studies have indicated that the carboxy–terminal domain of TBP contains the DNA binding activity and TBP binds as a monomer [126, 253, 292, 243, 51, 80].

The X–ray crystallographic structure of the *A. thaliana* and the yeast TBP was elucidated revealing a novel DNA binding motif [217, 34]. The C–terminal domain of TBP is a saddle–shaped protein; the concave side of the saddle is an antiparallel beta sheet that binds to DNA and the convex surface is thought to interact with other initiation factors and regulatory proteins. Co–crystallization of the TBP/TATA–box complex showed that the binding of TBP to DNA introduces a 80 degree bend in the DNA molecule [140, 141]. Mechanistically, DNA bending could allow the general transcription factors to form the necessary contacts to initiate transcription and/or may unwind the DNA duplex, thereby assisting in the formation of a single–stranded region needed to initiate transcription.

The carboxy terminus of the SfTBP amino acid sequence is at least 75% identical to the carboxy-terminal region of all other TBPs cloned to date (See Appendix 1); among insects it is 93% identical to the carboxy-terminus of a closely related species, *D. melanogaster* and 99% identical to the most closely related species, *Bombyx mori.* SfTBP contains all the conserved carboxy–terminal motifs with only a few conservative changes in amino acid composition compared to the consensus sequence derived from the other cloned TBPs, indicating that the DNA binding properties of SfTBP are similar to all other TBPs, including *A. thaliana.*

The amino-terminal domain of various TBPs differ greatly in length and primary structure. However, TBPs from higher eukaryotic organisms appear to contain some common elements that include glutamine tracts termed " Q-runs" and "STP" regions that are rich in serine, threonine and proline. Although the function of these regions are unknown, the TBP N–terminal domain in human and *Drosophila* appears to be required for interactions with the transcription factor, Sp1, that result in transactivation [241, 244]. Truncated proteins containing only the conserved C–terminal domain of human TBP can only restore basal level transcription to TFIID–depleted HeLa cell extracts suggesting that the N–terminal domain is also required for transactivation [241]. Experiments have shown that the N–terminus of TBP enhances TBP–induced bending and reduces the stability of the TBP–DNA complex [160]. However, *in vivo* replacement of the yeast TBP gene with the human or yeast–human hybrid TBP genes, indicated that crucial determinants of the functional differences between the yeast and human TBP were localized to the highly conserved C–terminal region [80, 51]. Therefore, both the variable N–terminal and the highly conserved C–terminal domain of TBP appear to be necessary for species specific protein–protein and protein–DNA interactions.

Comparing the N–terminal regions from closely related species may allow

one to identify regions that interact with common transcription factors and regions that may be responsible for species–specificity. As depicted in Figure 3.5, the *S. frugiperda* TBP and the *Bombyx mori* amino–terminal protein sequence are 82% identical. However, the *Drosophila*, amino–terminal sequence diverges significantly from both *B. mori* and *S. frugiperda.* Nevertheless, the TBP amino–terminal domains of these three insects contain the Q–runs and STP–like elements. Furthermore, the first 27 amino acids of the *Drosophila*, *S. frugiperda* and *B. mori* amino–terminal region are very similar (74%). The remainder of the amino–terminal domain of *Drosophila* does not align well with either *S. frugiperda* or *B. mori* except at three small regions rich in methionine, proline, threonine and glutamine. These alignments suggest that the amino terminal 27 amino acids and these three small regions may be required for interaction with conserved co–activators or other transcription factors.

3.5 TBP is Encoded by a Single Gene in *S. frugiperda*

S. frugiperda genomic DNA was digested with a number of restriction endonucleases, blotted, and hybridized with a labeled full–length probe to SfTBP (See Fig. 3.6). Both high and low stringency blots were identical; the probe hybridized to a single band per lane on each blot indicating that there is only one copy of TBP per haploid genome. This finding is consistent with data from other higher eukaryotic organisms. Furthermore, no cross–hybridizing bands were found on an low stringency AcMNPV genomic blot that was identically probed and washed under low stringency conditions indicating that the AcMNPV baculovirus does not encode a TBP homolog (data not shown).

Figure 3.5. Alignment of the Amino–terminal Region of *S. frugiperda*, *D. melanogaster* and *B. mori* TATA–binding Protein. The alignment of the amino–terminal amino acids of TBP from *D. melanogaster* (Dm), *S. frugiperda* (Sf), *Bombyx mori* (Bm) was generated using the clustal protein alignment program [115]. The arrowhead marks the beginning of the conserved carboxy–terminal domain of TBP and the vertical lines indicate amino acid residues conserved in both sequences. Upper case letters indicate amino acid identity with the SfTBP sequence. The numbers on the left indicate the position of the amino acid residue within the sequence.

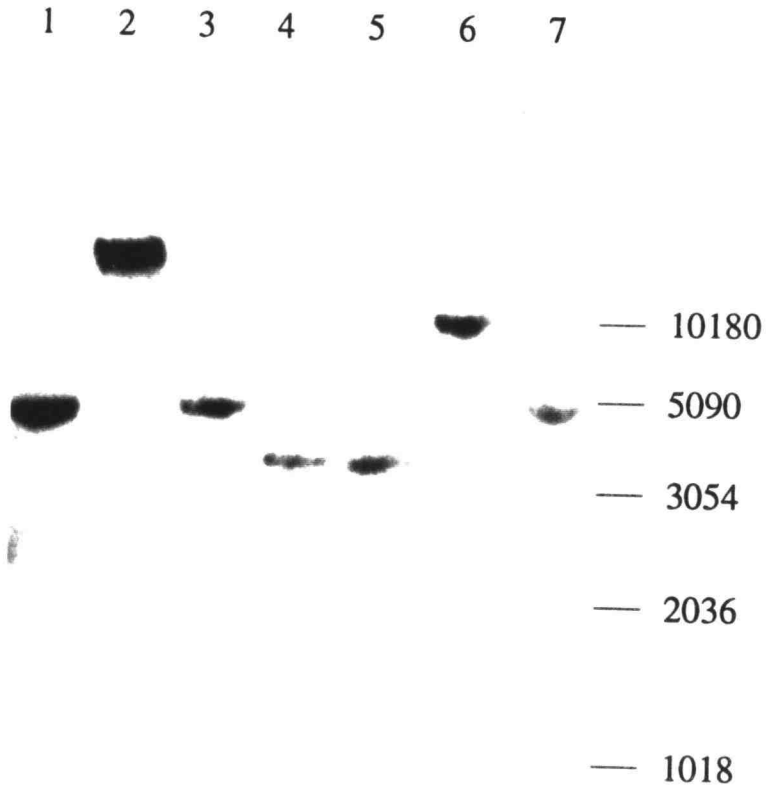


Figure 3.6. Detection of the *S. frugiperda* TBP Gene on a Genomic Southern Blot.*S. frugiperda* genomic DNA, digested with a variety of restriction endonucleases and blotted to Gene Screen Plus (Dupont), was probed with a random–primer labeled full–length cDNA of the *S. frugiperda* TATA-binding protein using low stringency conditions. Identical results were obtained using stringent conditions. *S. frugiperda* genomic DNA was digested as follows: lane 1, *Eco*RI; lane 2, *Bam*HI; lane 3, *Bam*HI and *Eco*RI; lane 4, *Bgl*II; lane 5, *Bgl*II and *Eco*RI; lane 6, *Hind*III; lane 7, *Hind*III and *Eco*RI. The BRL 1 kb DNA ladder was used as a size standard. The numbers on the right indicate the sizes in bps.

3.6 Northern Analysis of *S. frugiperda* and AcMNPV Transcripts during a Time Course of Baculovirus Infection

Sf9 cells were infected with AcMNPV at a moi of 10 and total RNA was isolated at various times p.i., blotted and hybridized with radiolabeled viral or host–specific probes. By 48 hours, the Sf9 cells had failed to divide and polyhedra were seen in 10–20% of the cells indicating that the infection was progressing. By 5 days, the infection was complete. Ethidium bromide staining of the RNA agarose–formaldelhyde gels after electrophoresis indicated that equal amounts of RNA were loaded in all lanes except 120 hr p.i. where the total RNA isolated showed extensive degradation.

The steady–state levels of the 1.3 kb SfTBP transcript were monitored during a time course of AcMNPV infection of Sf9 cells. Previous work by Ooi and Miller [225] has shown that the steady–state levels of a number of host–cell nuclear transcripts are reduced by 18 hr p.i. However, the mRNA levels for factors involved in host–cell transcription were not examined. The data (Fig. 3.7) show that the steady-state levels of the 1.3 kb SfTBP transcript decline by 24 hr p.i. indicating that the inhibition of host cell transcription includes factors required for the RNA polymerase II transcription process. Minor bands of 2.0 kb and 1.6 kb were also detected and may be unspliced TBP message, or another mRNA encoding a TBP–like protein [53]. The decline in steady–state levels of SfTBP transcripts after AcMNPV infection indicates that viral shutdown of host cell transcription includes messages for host–encoded transcription factors thought to be required for transcription of early baculovirus genes.

To determine if the decline in SfTBP mRNA levels was similar to the decrease in transcript levels of other host messages, a riboprobe complementary to the *D. melanogaster* actin 5C gene [75] was made and the same blot was reprobed. *S. frugiperda* transcripts of size 2.4, 1.9, and 1.4 kb cross–hybridized with *Drosophila* actin 5C (Fig. 3.8). All these heterologous actin transcripts had

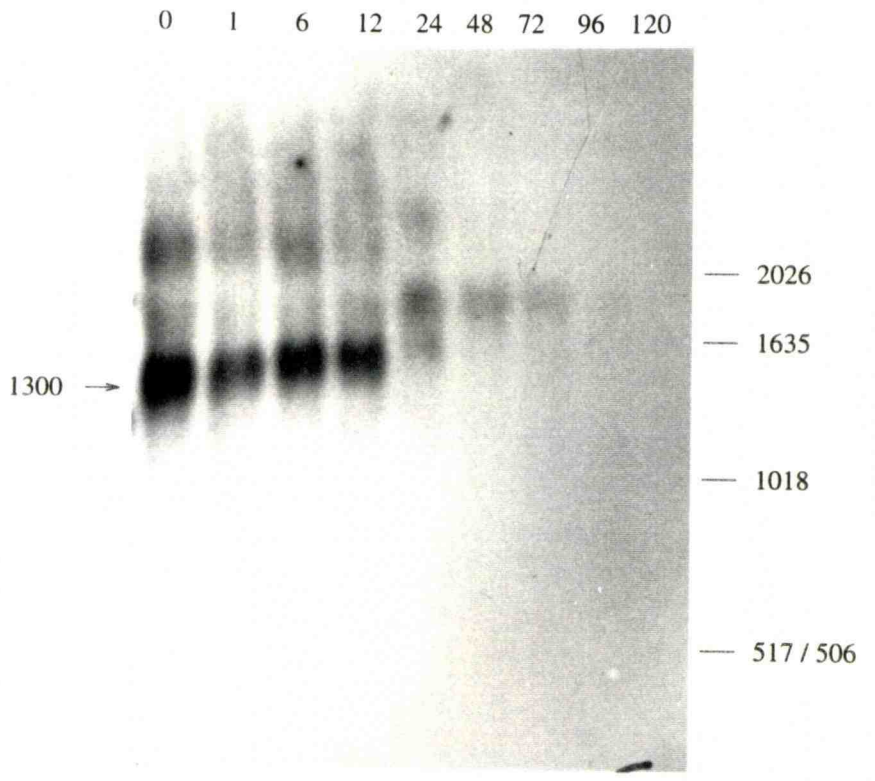


Figure 3.7. Northern Analysis of *S. frugiperda* TBP Transcripts During a Time Course of AcMNPV Infection. A strand-specific cRNA complementary to the SfTBP open reading frame was synthesized using SP6 polymerase and hybridized to a northern blot of total RNA (15 μg /lane) isolated from Sf9 cells at various times p.i. The numbers on the right indicate the sizes (bps) of selected markers of an end–labeled BRL 1 kb DNA ladder. The position and size of the SfTBP transcript is indicated on the left.

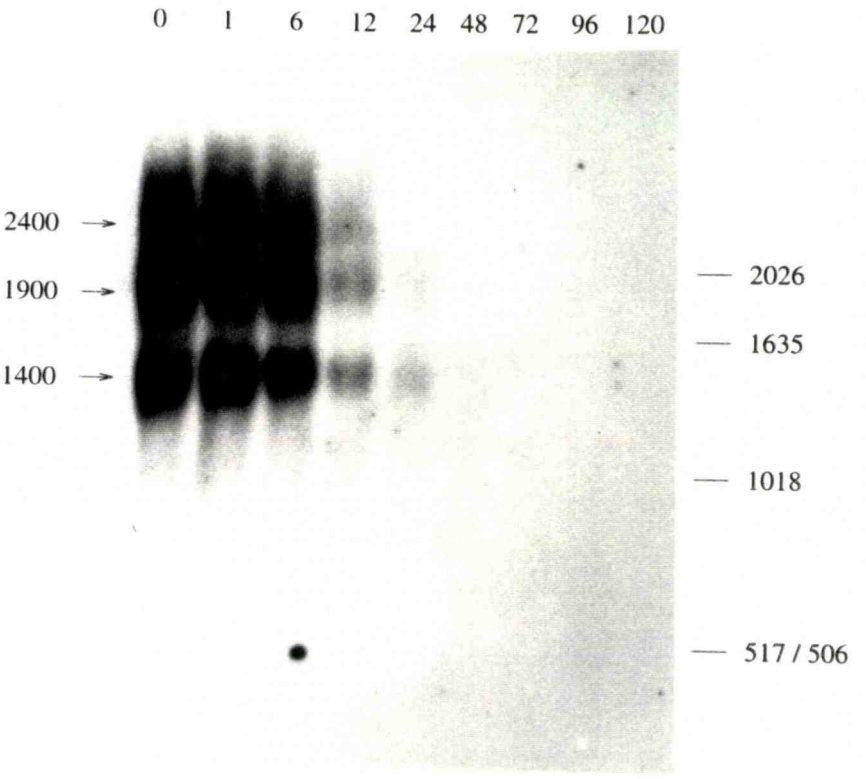


Figure 3.8. Northern Analysis of *S. frugiperda* Actin 5C Transcripts During a Time Course of Infection. The northern blot described in Fig 3.7 was stripped and rehybridized with a strand–specific cRNA complementary to the *Drosophila* actin 5C gene transcripts. The positions and sizes of the cross–hybridizing *S. frugiperda* actin 5C transcripts are indicated on the left. The markers are the same as in Fig 3.7.

a pattern of expression similar to that of SfTBP. However, actin 5C steady–state mRNA levels decreased somewhat earlier in the infection (by 12 hr) than SfTBP transcript levels. These results are consistent with those previously seen for actin 5C transcripts using dot blot analysis [225].

Concurrent with the decrease in steady–state levels of host messages, there was an increase in the level of expression of the AcMNPV viral messages, *p26* [250] and *ie–1* [103]. A riboprobe complementary to the 0.5 kb *Eco*RI fragment of AcMNPV*Hind*Q was constructed. This probe contains the *p26* gene without the *hr*5 enhancer region and was used to reprobe the same northern blot (Fig. 3.9).

Transcripts of length 1.2 kb, 1.4 kb and 2.1 kb were detected at 6 hr p.i.. The 1.2 kb band, that encodes the *p26* gene and a small portion of the 5' region of the *p10* gene [250], increases in abundance at 12 hr p.i. but is not detected at 24 hr p.i. indicating that the 1.2 kb band is an early baculovirus transcript whose transcription ceases by 24 hr p.i. The 1.4 kb transcripts, encoding the *p26* and *p10* genes, were detected between 6 and 72 hr p.i. A transcript of 2.1 kb was detected at 6 hr p.i., was maximally abundant at 12 hr p.i., and disappeared by 48 hr p.i. The 2.1 kb transcript has not been previously described in the literature [250, 74]. It is possible that this transcript is produced by a transcription initiation site within the *hr5* enhancer region or to non-termination of transcription at the *p10* termination site. Finally, it is possible that a small portion of the riboprobe contains vector sequences that cross-hybridize with another transcript that does not encode either the *p10* or *p26* gene.

A riboprobe complementary to the 5' terminus of the *ie-1* gene (a 752 bp *Cla*I–*Eco*RV fragment from Ac*Eco*RI–B) was constructed. This probe hybridized to a large number of transcripts by 24 hr p.i. Earlier in infection transcripts of approximately 1.9 kb and 2.1 kb could be discerned (See Fig. 3.10). The 1.9 kb transcript encoding the *ie-1* gene that was first detected 6 hr p.i, increased in intensity by 24 hr and was barely detectable by 72 hr p.i. [41, 103] when the level of all transcripts decreased significantly due to cell death caused by AcMNPV infection. Transcripts of approximately 2.1 kb were detectable 12 hr p.i. and degenerated to indistinct transcripts by 24 hr p.i. These bands may correspond to a number of spliced *ie-0* transcripts synthesized late in AcMNPV infection [154]. Later in infection, larger transcripts of 3.7, 3.5 and 3.1 kb that overlap the *ie-1* region have been reported [103]. Although these individual transcripts could not be distinguished in this blot, mRNA of these sizes hybridized to the *ie-1* probe. Similar results for *ie-1* and *ie-0* transcripts have been reported by others [103].

Northern analysis using AcMNPV viral probes demonstrated that the AcM-

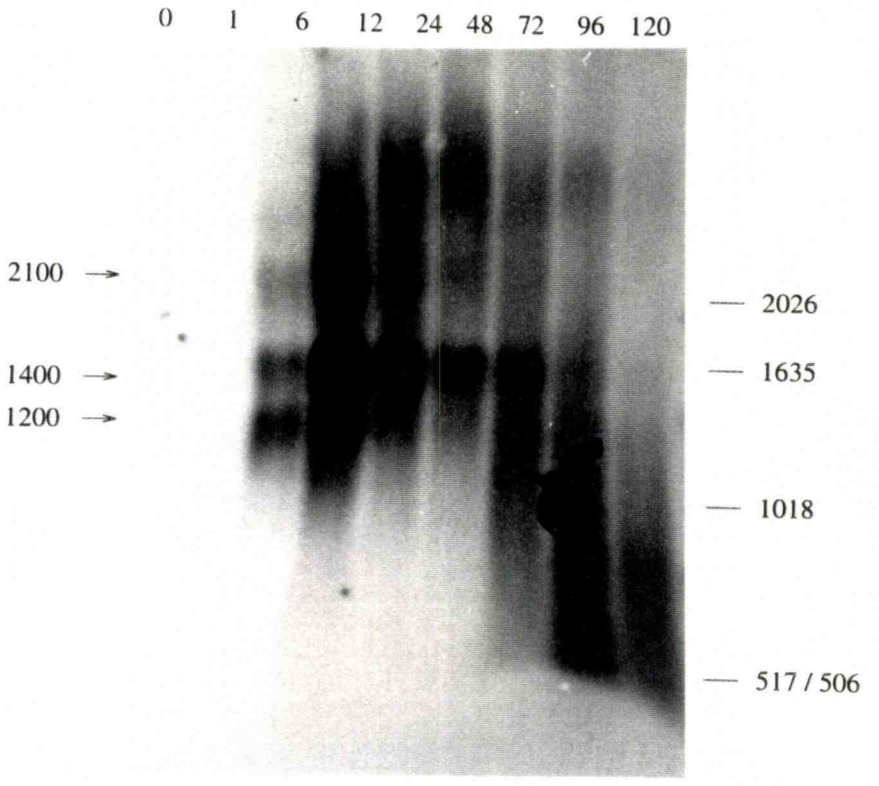


Figure 3.9. Northern Analysis of AcMNPV *p26* and *p10* Transcripts During a Time Course of Infection. The northern blot shown in Fig. 3.8 was stripped and rehybridized with a strand–specific cRNA complementary to the *p26* transcript. This riboprobe also detected *p10* messages. The positions and sizes of the *p26* and *p10* transcripts are indicated on the left.

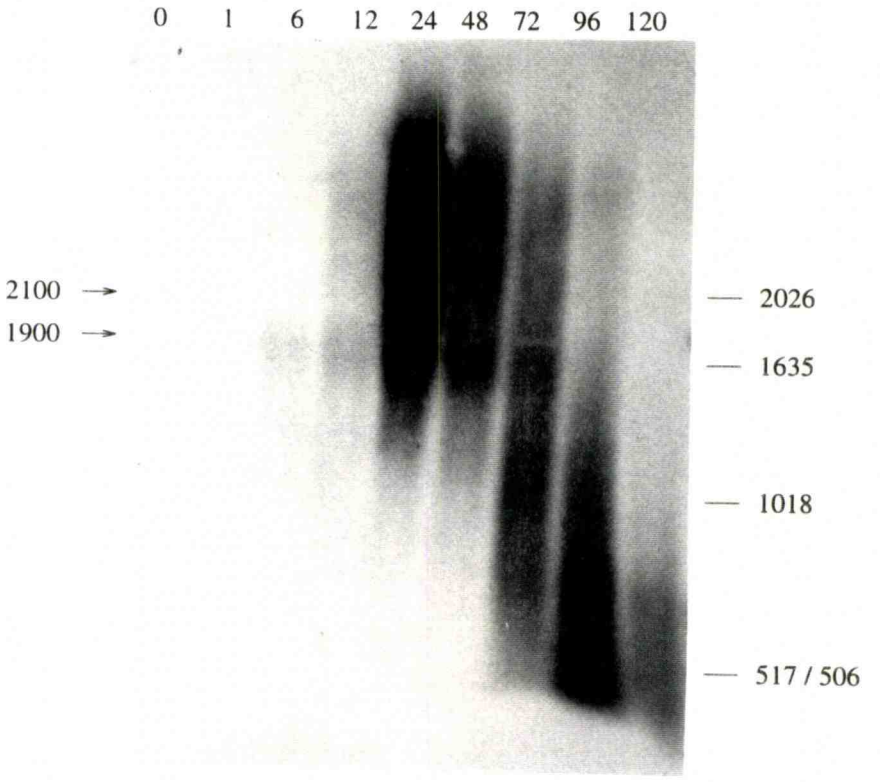


Figure 3.10. Northern Analysis of AcMNPV *ie-1* Transcripts During a Time Course of Infection. The northern blot described in Fig. 3.9 was stripped and rehybridized with a strand-specific cRNA complementary to the *ie-1* transcript. The positions and sizes of the *ie-1* containing transcripts are indicated on the left.

NPV infection progressed as expected indicating that the steady state transcript levels seen are indicative of the infection process. The decline in both actin and SfTBP transcript levels indicate that the inhibition of host cell transcription includes messages for host–encoded transcription factors required for the RNA polymerase II transcription process.

3.7 Detection of the TATA–Binding Protein during Baculovirus Infections of Sf9 and Ld Cells

In order to examine TBP protein levels during the course of baculovirus infection, western blot analysis of AcMNPV–infected Sf9 cells using an antibody made against the highly conserved C–terminal domain of TBP from *Drosophila melanogaster* was performed. Protein samples were isolated at the same times p.i. that the total RNA samples were isolated. As Figure 3.11 shows, SfTBP is present at relatively constant levels throughout the time course of infection. Although the open–reading frame of the SfTBP predicts a molecular weight of 34 kDa; the antibody detects proteins of molecular weight 36 kDa and 38 kDa indicating that SfTBP may be post–translationally modified.

The decline in steady–state SfTBP message levels does not result in a substantial decrease in SfTBP protein levels (See Fig. 3.7 and 3.11). SfTBP appears to be stable and not targeted for degradation. In poliovirus–infected HeLa cells, TBP is specifically cleaved by the poliovirus–encoded 3C protease [45, 54]. This cleavage plays a role in the inhibition of host cell RNA synthesis. Unlike poliovirus, the inhibition of host cell transcription during baculovirus infection does not appear to involve TBP cleavage or degradation.

Western blot analysis was also performed on protein samples isolated at the various times p.i. from another lepidopteran insect, *Lymantria dispar* infected with the baculovirus, OpMNPV. As Figure 3.12 shows, that LdTBP has a molecular weight of 33 KDa (3 kDa smaller than SfTBP) and was present at relatively

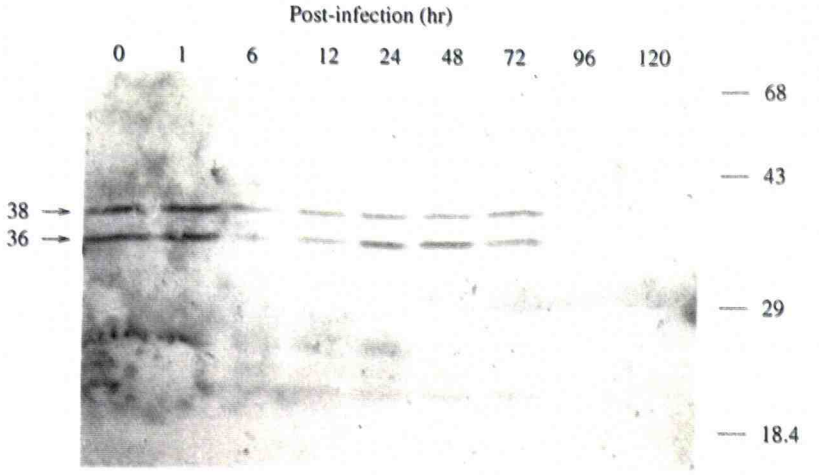


Figure 3.11. Western Blot Analysis of SfTBP During a Time Course of AcM-NPV Infection. Protein samples were isolated at the same times p.i. that the total RNA samples were isolated. The time in hr p.i. are indicated above the lanes. 4.0 $\times 10^4$ cell equivalents of protein was loaded per lane. The numbers on the right indicate the marker sizes in kDa (BioRad prestained low range protein ladder).

constant levels until 36 hr p.i. after which LdTBP was no longer detected. This result differs from the western analysis of SfTBP where protein levels persist until 72 hr p.i. The more rapid decrease in the levels of LdTBP may be due to more rapid degradation, decreased protein stability or a more rapid decline in TBP message levels upon OpMNPV infection.

In addition, only one form of LdTBP was detected and it was 3–5 kDa smaller (33 kDa) than SfTBP. This suggests that TBP may be variable in closely related insects possibly due to differences in N–terminal amino acid sequences and/or differences in post–translational modifications. There is a high degree of conservation between mouse and human N–terminal sequences [111]; there are five amino acids differences and a partial deletion of a Q–run (likely to be a single event) in the mouse sequence. The *B. mori* sequence has 21 amino acid changes when compared to the Sf9TBP sequence (See Fig. 3.5). Greater divergence among insect TBP sequences may be explained by the substantially higher rates of DNA evolution in insect genomes compared to those found in other groups, especially mammals and birds [27].

3.8 Summary

In order to begin characterizing early transcriptional events in the baculovirus infection cycle, the gene encoding the TATA–binding protein from the *S. frugiperda* cell line, Sf9 was cloned and sequenced. A cDNA clone containing an 921 bp open–reading frame (307 amino acids; 34 Kda) homologous to the TATA–binding protein (TBP) was isolated and sequenced from Sf9 cells which are commonly used in the baculovirus expression system. Southern blot analysis indicated that SfTBP was encoded by a single gene in the *S. frugiperda* genome. Northern blot analysis indicated that steady–state levels of the 1.3 kb SfTBP transcript declined by 24 hr p.i. corresponding to the time of virus–induced inhibition of other host–cell transcripts. However, corresponding western blot analysis showed that

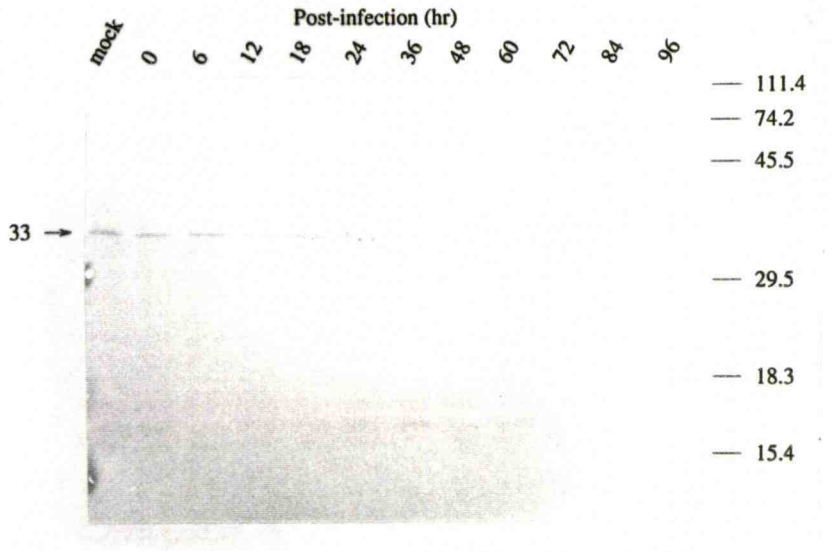


Figure 3.12. Western Blot Analysis of LdTBP During a Time Course of OpMNPV Infection. The time in hr p.i. are indicated above the lanes. 3.2 $\times 10^4$ cell equivalents of protein was loaded per lane. The numbers on the right indicate the marker sizes in kDa (BioRad prestained low range protein ladder).

the steady–state TBP protein levels remained unchanged until much later in infection. Therefore, if other components of the Sf9 RNA polymerase II transcriptional machinery behave in a similar manner, cessation of host–cell transcription would require active intervention by one or more baculovirus encoded factors.

Analysis of the *S. frugiperda* TBP (SfTBP) sequence showed that the amino–terminal portion of SfTBP was 82% identical to the *B. mori* N–terminal TBP sequence. Both the *S. frugiperda* and *B. mori* N–terminal TBP sequences diverged significantly from the N–terminal region of all other TBP sequences, including *Drosophila melanogaster*. This analysis of TBP sequences of two closely related insect species suggests that the amino-terminal portion of TBP is variable in insect species.

Selectively targeting specific insect pests while leaving beneficial insects unaffected is a major goal of insect control programs. Fundamental life processes such as RNA polymerase II transcription that involve proteins with a high degree

of variability are potential targets for insect pest control. Utilizing information about proteins such as TBP that play critical roles in life processes may lead to the development of methods capable of selectively disrupting these processes in target insects.

Chapter 4
Inferred Phylogenetic Trees using Molecular Sequences encoding the TATA–Binding Protein

4.1 Molecular Evolution

The evolutionary history of organisms can be inferred using data derived from homologous genetic material. Genetic information may change through time due to events such as substitutions (transitions or transversions), deletions, and inversions. By studying the rates and patterns of change in homologous genetic material, the evolutionary history of organisms can be constructed.

The choice of the molecular sequence used to construct phylogenetic trees is important if we are to have confidence in the results. If a sequence is either nearly identical or shows very little identity, informative positions (a site is phylogenetically informative if it favours some trees over others) are rare and phylogenetic trees can not be inferred with any confidence. Homologous protein–coding genes, where synonymous substitutions (synonymous substitutions cause no change in the amino acid sequence but result in a change in the nucleotide sequence) can occur, are often highly conserved at the protein level but may contain informative sites at the nucleotide level. Therefore phylogenies using highly conserved proteins can often be reconstructed with confidence. The genetic data can be used to augment other data such as the morphology, embryology and the fossil record of the organisms to construct a phylogenetic tree.

4.2 Constructing Phylogenetic Trees

Methods of inferring phylogenetic trees can be classified according to the type of data, either character–based or distance–based, used to reconstruct the tree. Character–based methods use discrete characters in a DNA sequence to provide information about a particular species whereas distance–based methods calculate

the similarity between two sequences to determine the evolutionary distance between two species (for reviews of both character–based and distance–based methods see [10, 296, 116, 67]).

4.2.1 Distance Methods

Distance methods compute the evolutionary distances for all pairs of taxa by determining the number of nucleotide or amino acid substitutions separating the two taxa. The similarity score for the two sequences is corrected based on a model of evolution and an additive tree constructed. Distance methods assume that the lengths of the branches between taxa can be summed to yield a measure of the distance between the taxa and branching events. Several different distance methods were used in this analysis. They include: the Desoete tree with either the Jukes and Cantor distance [133] or the Olsen correction [142], the Fitch and Margoliash method [70] which uses a weighted least squares method to correct for the error in distance estimates to obtain branch length, and the Neighbor–Joining method [268] which sequentially identifies neighbor–pairs in a tree and uses the nodes generated to construct a tree.

4.2.2 Character Methods

Character–based methods use informative sites to determine which phylogenetic tree best explains the data. Two different character–based methods were used in this analysis, maximum parsimony [60, 69] and maximum likelihood [67]. Maximum parsimony finds the tree or trees that require the fewest number of evolutionary changes to explain the data sampled. Maximum parsimony can be bootstrapped; a subset of the data is used to create a new data set from which a number of most parsimonious trees are calculated. Bootstrapping allows you to determine if there is significant evidence for a particular branching order in your

data set. Maximum likelihood methods determine the probability of resultant phylogenetic tree given a certain evolutionary model. The inferred phylogenetic tree has the highest likelihood score.

4.3 Is the TATA–binding Protein an Informative Macromolecule?

The process of transcription, its regulation and fidelity are of central importance to a cell and therefore an organism. Since this process occurs in all extant life, DNA–dependent RNA polymerases and their associated protein factors may be homologous. Eukaryotic organisms utilize three different polymerases, RNA polymerase I, II, and III, that contain some common subunits [329] to transcribe DNA. Sequence data and immunochemical cross–reactivity indicate that the largest subunits of the RNA polymerases from eubacteria, eukaryotes and archaebacteria are homologous [246]. Since the RNA polymerases are homologous, the accessory transcription factors may are also homologous. One of the best characterized general transcription factors is the TATA–binding protein. TBP is a component of all three eukaryotic RNA polymerase complexes [320, 279, 293] and TBP–like sequences have recently been found in archaebacteria [264, 193]. The C–terminal region of the TBP consists of two conserved direct repeats that likely arose from an early duplication event. The amino–terminal regions of TBPs from different species vary greatly in size and sequence. The carboxy–terminal domain of TBP is highly conserved, having greater than 80% sequence identity in a wide variety of eukaryotic organisms, suggesting a universal, perhaps primordial, role for TBP in the transcription process. Therefore, TBP is a good candidate macromolecule for use in phylogenetic reconstruction. I reconstructed an evolutionary history of eukaryotes using the TBP sequence data available in Genbank as of May 31, 1995.

4.4 Phylogenetic Analysis using the TATA–Binding Protein

As of May 31, 1995, the TATA–binding protein had been cloned and sequenced in the following organisms: *S. cerevisiae* [62, 107, 125], *S. pombe* [119], *P. carinii* [201], *D. melanogaster* [117, 209], *S. frugiperda* [252], *B. mori* [286], *H. sapiens* [241, 136], *M. musculus* [297], *M. auratus* [219], *A. castellanii* [325], *S. tuberosum* [120], *Z. mays* [104, 311], *A. thaliana* [79], *T. aestivum* [138, 6], *D. discoideum* [20], *X. laevis* [111], *C. elegans* [178], *O. vulvulus* [176], *P. falciparum* [197], *A. cliftonii* [71], *T. thermophila* [288], *E. histolytica* [228] and two archaebacteria species, *T. celer* [193] and *P. woesei* [264]. Note that *A. thaliana*, and *Z. mays* have two copies of TBP that have likely arisen from a recent duplication event.

4.4.1 Clustal Alignment of TBP Sequences

The TATA–binding protein nucleotide sequences and their corresponding amino acid translations were aligned using the clustal alignment program in GDE [115]. Since the carboxy–terminus of all the TBP sequences is highly conserved, the computer–generated alignment of the both the nucleic acid sequence and the amino acid translations needed only minor modifications. However, the amino–terminal portion of all the TBP sequences did not align well, although certain sequence motifs appear in the amino–terminal regions of most of the TBPs from different species. Analysis of the genomic organization of yeast, mouse, and amoeba TBP suggest that the C–terminal and N–terminal regions may have evolved separately [294]. The alignment of the TBP protein sequences is shown in Appendix 1. Based on these alignments and the possibility that inclusion of the amino–terminal portion of TBP would bias the alignments, the amino–terminal portion of the TBP sequences were excluded from this analysis. Therefore, sequence masks were created for carboxy–terminal region of the TBP to examine

position 1, 2 and 3 within each codon.

4.4.2 Inferred Phylogenetic Trees using the TATA-Binding Protein

I constructed a Desoete unrooted tree (no correction) using the aligned carboxy-terminal region of the TBP protein sequence (See Figure 4.1). Figure 4.1 displays the phylogenetic relationships using a phenogram. This tree conformed to current beliefs about the evolutionary branching order of these organisms [145] with fungi, plants and animals clustered in groups of relatedness. The higher eukaryotic organisms, consisting of *M. musculus*, *X. laevis H. sapiens*, and *M. auratus* clustered in a manner suggestive of a radiation event occurring during the course of evolution [145]. The fungal cluster included *P. carinii*, an opportunistic pathogen, responsible for the majority of fatal pneumonia in AIDs patients [201]. This is consistent with the recent ribosomal RNA data that places *P. carinii* in the fungal kingdom [291]. The insects, *D. melanogaster*, *B. mori* and *S. frugiperda* and the nematodes, *O. vulvulus* and *C. elegans* have diverged from the other higher eukaryotic organisms. The plant TBP sequences *A. thaliana*, *S. tuberosum*, *Z. mays* and *T. aestivum* sequences cluster. However, further examination of the TBP–based phylogenetic tree reveals that the plants do not segregate into the monocot and dicot classes. However, selective pressures may have altered the rate of evolution in three of the four plant TBP sequences since these sequences are from domesticated crop species. However, phylogenetic trees based on the nucleic acid sequences does segregate the plants into monocot and dicot classes although fails, due to codon bias, to produce a biologically meaningful tree for several other organisms (see below).

Inspection of the clustering of the eukaryotic unicellular organisms including the ciliate, *T. thermophila*, the causative agent of malaria, *P. falciparum*, the amoeba, *A. castellanii*, the slime mold, *D. discoideum*, and an entamoeba that causes dysentery in humans, *E. histolytica*, indicated that these species have

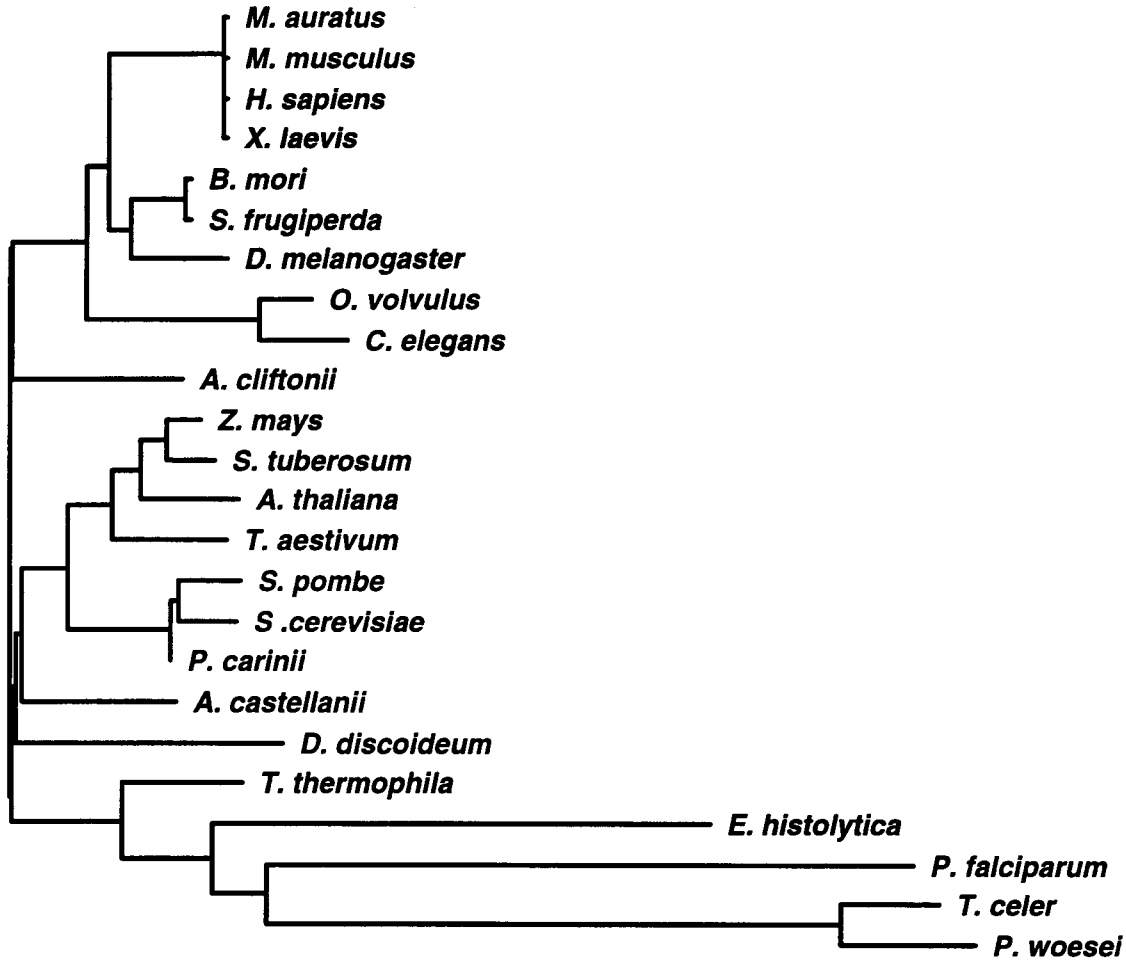


Figure 4.1. A Phylogenetic Tree Constructed Using the TATA–Binding Protein. Phylogenetic relationships were determined using an alignment of the entire conserved carboxy–terminal domain amino acids from the complete TBP sequences available from Genbank, EMBL and DDBJ on 31 May 1995.

diverged considerably supporting other evidence [198, 224, 145, 315] that protozoans are paraphyletic and diverged from the mainstream of eukaryotic descent early in eukaryotic evolutionary history. Interestingly, the disease causing agents, *P. falciparum* and *E. histolytica* appear to be evolving at a faster rate than other unicellular organisms. Studies have shown disparate rates of molecular evolution in hosts and their parasites [105] that correlated well with the difference in generation times. Furthermore, these agents have been subjected to selective pressures, such as insecticide treatment of insect vectors and drug treatments in recent times, that may have increased their mutation rates. Higher rates of mutation would effect the phylogenetic tree by moving the branching point for *P. falciparum* and *E. histolytica* deeper into the tree.

The branching order of the two archaebacteria sequences *T. celer* and *P. woesei* suggests that archaebacteria and eukaryotes diverged from each other after the divergence of eubacteria. Evidence that archaebacteria are phylogenetically more closely related to eukaryotes than eubacteria is supported by other data. Similar analysis with subset of TBP sequences [193], RNA polymerase data [246], analysis of translation elongation factor sequences [130], and the discovery of TFIIB–like proteins in archaea [229] indicates that eubacteria diverged first followed by the subsequent divergence of archaebacteria and eukaryotes. However, examination of the branch points for *T. celer* and *P. woesei* place the emergence of the archaea later than other evidence indicates [145]. This result can be explained if the rates of mutation are higher in *P. falciparum* and *E. histolytica*, and thus placed these species deeper in the tree than their actual date of emergence.[2]

---

[2] A phylogenetic tree containing the subset of TBP sequences available in May, 1993 was published as part of C. Rasmussen and G.R. Rohrmann (1994). Characterization of the *Spodoptera frugiperda* TATA–Binding Protein: Nucleotide Sequence and Response to Baculovirus Infection. Insect Biochem. Molec. Biol. 24:699–708.

The nucleotide sequence encoding the carboxy–terminal region of TBP was then used to construct a number of Desoete trees (Jukes and Cantor or Olsen correction) using different sequence masks. Both the Jukes and Cantor and Olsen corrections produced trees with the same branching order and similar branch lengths (data not shown). When all nucleotide positions in each codon and only the 3rd nucleotide position in each codon were considered, unexpected clustering was observed. The dipteran insect, *D. melanogaster*, and *A. castellanii*, an amoeba, formed a cluster. These results suggested that *Drosophila* is more closely related to amoeba than to the lepidopteran insect, *Spodoptera frugiperda.* This pattern always emerged when these positions are used to infer phylogenetic trees using other distance methods as well as maximum parsimony and maximum likelihood methods (data not shown). This result is unexpected since conventional wisdom based on morphology [159, 21] and a poor fossil record [326] indicate that Lepidoptera and Diptera share a recent common ancestor. However, this result can be explained by looking at the codon bias in the TBP sequence in these organisms. It appears that the codon bias for *A. castellanii* and *Drosophila*, are similar as both organisms prefer to have cytosine or guanine in the third position of each codon.

When the 1st and 2nd nucleotide position in each codon was considered, phylogenetic trees that more closely conforms to current beliefs of the evolutionary branching order of these organisms was obtained. Nevertheless, bootstrapped maximum parsimony analysis indicated that the branching order among the higher eukaryotes could not be determined with a high degree of confidence. Molecular phylogenies of eukaryotic organisms suggest patterns of episodic increases in biological diversity that correlates well with the fossil record [145]. If the higher eukaryotic organisms like *S. frugiperda*, *X. laevis* and *H. Sapiens*, all diverged from a common ancestor at approximately the same time, this could explain the uncertainty in the branching orders. However, clearly, Diptera and

Lepidoptera are morphologically more related to each other than to humans. Recent evidence [27] that suggests that insect genomes have a much higher rate of nucleotide substitutions than mammals could also account for the results found here. Higher rates of evolution would effect phylogenetic trees by moving the branch point deeper into the tree, similar to what we see in the phylogenetic trees inferred using sequence information from the TATA–binding protein. Nevertheless, the phylogenetic trees constructed using sequence data from the TATA–binding protein, in particular, the carboxy–terminal protein sequence data conformed closely to current beliefs on the evolutionary branching order of eukaryotic organisms.

4.5 Eubacterial Homologs of the TATA–Binding Protein

X–ray crystallography has shown that the TATA–binding protein binds to DNA in the minor groove of the DNA double helix created by the TATA–box element [289, 164, 140, 141, 34]. Unlike TBP, most eukaryotic sequence–specific DNA binding protein recognize and bind DNA in the major groove (for review see [110]). However, the histone–like proteins of eubacteria (for review see [58]) appear to bind to DNA in the minor groove of the DNA double helix [332]. A histone–like protein from *B. stearothermophilus*, HU, has been crystallized and its three–dimensional structure suggests that a pair of two–stranded β ribbons emerge from the body of the protein and encircle the DNA double helix in the minor groove [298]. The *E.coli* protein IHF (integration host factor) is thought to be a homolog of the histone–like proteins of eubacteria [58] and also contains the β ribbon DNA binding motif. However, unlike the histone–like proteins which bind DNA in a nonspecific manner, IHF binds to specific DNA sequences; IHF is required for the site–specific recombination of the bacteriophage lambda into the *E.coli* genome [58]. Nash and Granston [215] suggested that some sequence similarity exists at the protein level between the DNA binding domains of the

TATA–binding protein and IHF.

Nuclear magnetic resonance (NMR) studies using the MotA transcription factor from bacteriophage T4 also revealed a secondary structure consisting of a six–stranded antiparallel β pleated sheet with three α helices. This secondary structure is similar to the carboxy–terminal repeat in TBP including hydrophobic and aromatic residues on exposed sheet surface of protein [68]. Since both these proteins appear to bind to the minor groove of the DNA helix in sequence–specific manner and share some similarity at the protein level in the DNA binding domain, they may be homologous. I have tested the hypothesis that these eubacteria proteins may be homologs of TBP by aligning the TBP carboxy–terminal protein sequence with IHF and MotA protein sequence using the clustal protein alignment program. Alignments of both the amino acid and DNA sequence of *E. coli* IHF and MotA with the eukaryotic TBPs indicate that these genes are not homologous. Nevertheless, it is possible that this DNA binding motif (β–ribbon structure) arose only once in the course of evolution but the β–ribbon can tolerate extreme divergence and still maintain function. However, considering the degree of conservation in the eukaryotic and archaebacterial TBP β–ribbon structure, this is unlikely. Therefore, it is probable that the similarities between these molecules are a consequence of convergent evolution.

4.6 Summary

The carboxy–terminal region of the TATA–binding protein is highly conserved, suggesting that this gene is homologous. Using the TBP sequence data, I have reconstructed an evolutionary history of eukaryotic organisms. This phylogenetic analysis indicated that the TATA–binding proteins of both archaebacteria and eukaryotic species are indeed homologous. Furthermore, this analysis showed that TBP is a reasonable macromolecule to use to infer phylogenetic relationships though it may be more informative within kingdoms or phyla. Finally, the

molecular data suggest that the TATA–binding protein of eukaryotes and archaebacteria is not a homolog of the *E. coli* protein, IHF or MotA transcription factor from bacteriophage T4.

Chapter 5

In vitro Transcriptional Analysis of Baculovirus Early and Late Gene Promoters

5.1 Introduction

Much of the progress in defining components of the transcriptional machinery in eukaryotic cells has utilized biochemical approaches that involve the development of *in vitro* transcription systems followed by subsequent fractionation and reconstitution of active components (reviewed in [272]). The development of an *in vitro* transcription system for baculovirus genes using uninfected Sf9 cells [122], and infected insect cells [83, 82] should prove useful for the isolation and identification of host and viral genes involved in the regulation of baculovirus gene expression. Of particular interest are the interactions between the host–cell transcription machinery and viral regulatory proteins required for transactivation of early viral gene transcription. Later in the infection process, the mechanisms of transcription of the novel baculovirus late and hyperexpressed late gene promoters are also of interest. Detailed knowledge of these processes will lead to a better understanding of the infection process aiding in the development of more efficient expression systems and biological control agents.

This chapter summarizes the results of my work using the baculovirus *in vitro* transcription systems to study early and late gene promoters. I show that uninfected nuclear extracts are able to transcribe two baculovirus early gene promoters. Using extracts prepared at various times post–infection, early, late and very late baculovirus gene promoters were shown to be transcribed *in vitro* in a temporal manner similar to that seen during the course of baculovirus infection of insect cells. Nuclear extracts prepared at 16 hr p.i. were optimal for baculovirus late gene transcription. In an attempt to identify late–promoter specific DNA binding proteins, a minimal baculovirus late promoter was constructed and

used in gel–retardation assays. Deletion analysis of the *vp39* late promoter delineated a minimal baculovirus late promoter. However, subsequent gel–retardation assays failed to identify late–promoter specific DNA binding proteins.

5.2 *In Vitro* Transcription of Baculovirus Early Promoters

To show that uninfected nuclear extracts were able to transcribe several baculovirus early gene promoters using reaction conditions similar to those established for the OpMNPV *gp64* early promoter [122], the AcMNPV *ie–1* and the OpMNPV *p26* gene promoters were assayed.

The *ie–1* gene encodes a multifunctional protein that is transcribed early in infection [103, 218, 28, 185, 102, 98, 42] and is involved in both the transactivation of baculovirus early genes and replication of the plasmids containing the putative baculovirus origins of replication [187, 148, 2]. S1 nuclease mapping of the transcriptional start sites of the *ie–1* gene has shown that transcription initiates within a consensus CAGT sequence located 32 bp downstream from a TATA–box [103] (See Figure 5.1). The OpMNPV *p26* gene is transcribed early in infection and transcript analysis has shown that transcription initiation occurs about 22 nucleotides downstream from a TATA–like sequence element [11]. To assay for *in vitro* transcription activity in nuclear extracts, the pAcIE–1 and pCG19A (pCG19A contains the *p26* gene) plasmids were digested with *Hinf*I and *Hind*III respectively. If accurate *in vitro* transcription initiation with uninfected nuclear extracts occurs, run–off transcripts of 339 and 135 nucleotides would be synthesized from the AcMNPV *ie–1* promoter and the OpMNPV *p26* promoter, respectively (See Figure 5.1).

The results of the *in vitro* transcription assays are shown in Figure 5.2 and Figure 5.3. Major run–off transcription products of the expected size are present indicating that nuclear extracts from uninfected Sf9 cells are capable of specifi-

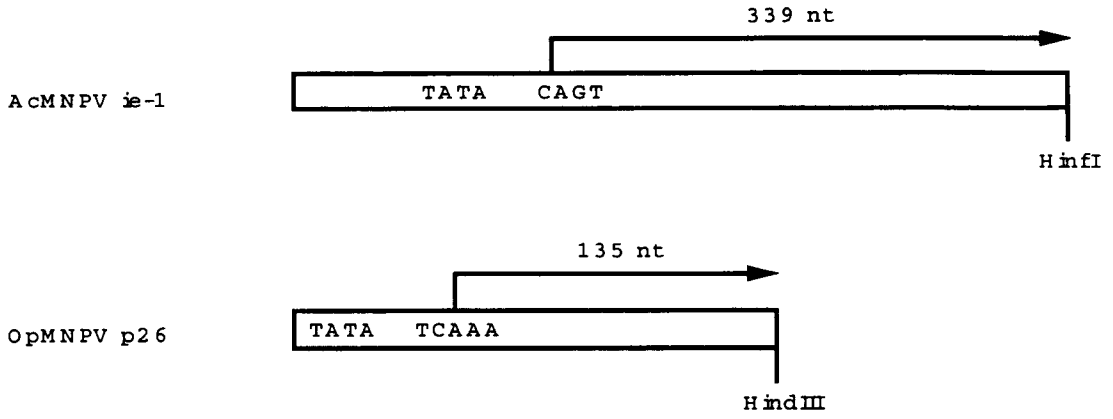


Figure 5.1. Schematic Representation of the AcMNPV *ie-1* and OpMNPV *p26* Run-off Transcripts. The arrows indicate the expected site of transcription initiation and size in nucleotides of the run-off transcripts.

cally initiating transcription at the AcMNPV *ie-1* and OpMNPV *p26* early gene promoters from the same sites used *in vivo*.

The influence of Mg^{2+} concentration on the efficiency of *in vitro* early transcript production using these two promoters was characterized. As shown in Figure 5.2, specific transcription initiation was observed from the AcMNPV *ie-1* promoter at Mg^{2+} concentrations from 1.2 mM to 11.2 mM with maximal activity between 5.2 and 7.2 mM Mg^{2+}. Transcription from the OpMNPV *p26* promoter was observed at similar concentrations of Mg^{2+} (See Fig. 5.3) with maximal production of run-off transcripts occurring between 3.2 and 5.2 mM Mg^{2+}. These results indicate that the *ie-1* and *p26* early baculovirus promoters have Mg^{2+} optima (4–6 mM) similar to that show for the OpMNPV *gp64* construct [122].[3]

---

[3] The AcMNPV *ie-1* transcription data was published in: R.R. Hoopes Jr., and G.F. Rohrmann. (1991). In vitro transcription of baculovirus immediate early genes: Accurate mRNA initiation by nuclear extracts from both insect and human cells. Proc. Natl. Acad. Sci. USA 88:4513–4517.

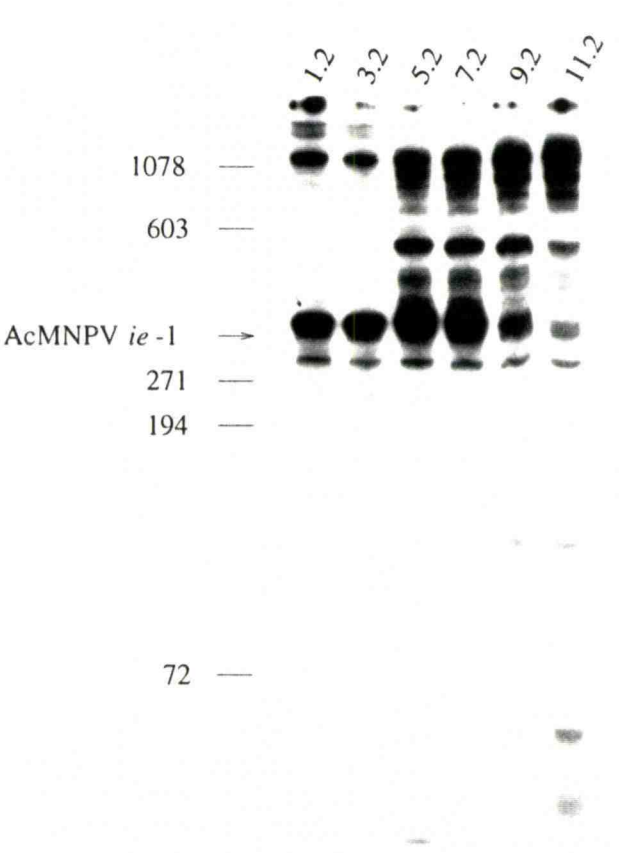


Figure 5.2. Optimization of Mg^{2+} Conditions for AcMNPV *ie-1 in vitro* Transcription. The numbers above the lanes indicate concentration of Mg^{2+} in mM.

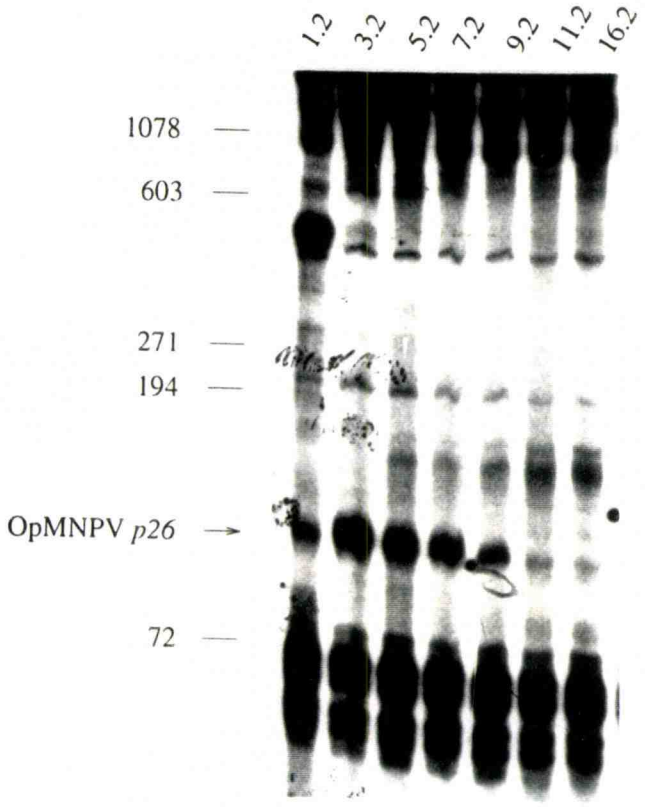


Figure 5.3. Optimization of Mg^{2+} Conditions for OpMNPV *p26 in vitro* Transcription. The numbers above the lanes indicate concentration of Mg^{2+} in mM.

5.3 *In vitro* Transcription of Baculovirus Late Promoters

Baculovirus late gene transcription is dependent on an α–amanitin– and tagetoxin–resistant RNA polymerase [76, 91, 82] that initiates transcription within an DTAAG promoter sequence after the onset of viral DNA replication [17]. This polymerase and any associated factors are probably responsible for the high levels of expression of the polyhedrin (*polh*) and *p10* genes. The development of an *in vitro* transcription system for baculovirus late promoters that can be used to identify and characterize factors involved in late gene expression is a major advancement for baculovirus transcription studies [82, 331].

5.4 Optimization of Late *In Vitro* Transcription Conditions

Glocker *et al.* [82] were able to prepare nuclear extracts from infected insect cells capable of transcribing the *39K*, *vp39*, *p10* and *polh* promoters. Conditions similar to those previously established for early *in vitro* transcription were found, except that late transcription signals were optimal at low concentrations (1 or 2 mM) of Mg^{2+}. This may be due to the presence of a Mg^{2+}–dependent DNase activity [82] that is activated at higher Mg^{2+} concentrations resulting in degradation of the DNA template.

To optimize late *in vitro* transcription signals, I tested a number of additional reaction conditions including substituting manganese (Mn^{2+}) for magnesium, template preparation methods, and the addition of EDTA to the *in vitro* transcription reactions. Primer extension analysis, performed on RNA transcribed from a *vp39* or *39K* promoter template using late transcription conditions, was used to evaluate the modifications. The M13 forward primer which is complementary to plasmid sequences, was used to prevent primer extension of any endogenous viral RNA present in infected–cell nuclear extracts.

Addition of EDTA and substitution of Mn^{2+} for Mg^{2+} were assayed in at-

tempts to reduce or inhibit the DNase activity. Concentrations of Mn^{2+} ranging from 0 to 8 mM were unable to support late *in vitro* transcription from the *vp39* promoter (data not shown). As shown in Figure 5.4a, addition of 2 mM EDTA to the *in vitro* transcription reactions resulted in the same late transcript signal intensity as was present at 0 mM EDTA but addition of 4 mM EDTA abolished late transcription from the *39K* promoter. Therefore, low concentrations of EDTA had no effect on the *in vitro* transcription reaction whereas higher concentrations were detrimental to late transcript production.

Three methods of template preparation, alkali lysis, CsCl gradients and purification on Qiagen columns were tested for ability to support late transcription. Data indicated that DNA template prepared either on CsCl gradients (Fig. 5.4b, lanes 1 and 2) or purified on Qiagen columns (data not shown) served equally well as DNA templates while templates prepared using the alkali lysis procedure generated less intense signals (Fig. 5.4b, lane 3) and on occasion, a number of extraneous bands (data not shown).

It has been reported that DNase activity can be eliminated from nuclear extracts by purification of intact nuclei through a 2 M sucrose cushion [195, 194]. Nuclear extracts were prepared from infected Sf9 cells at various times p.i. as described (see section 2.17) except nuclei were pelleted on sucrose cushions as described in Marzluff and Huang [194] before the high salt extraction. As shown in Figure 5.4c purification of nuclei on sucrose cushions did not improve the late transcription signal from the *39K* promoter.

Finally, nuclear extracts prepared 16 hr p.i. are rendered transcriptionally inactive by mild heat treatment. Nuclear extracts incubated at 47°C for 20 min, were not capable of producing transcripts while untreated extract produced transcripts from pCR5 (See Figure 5.4d). This result suggests that a factor(s) required for late *in vitro* transcription is heat–labile. Previously, Nakajima *et al.* found that the TBP–containing fraction, TFIID, of RNA polymerase

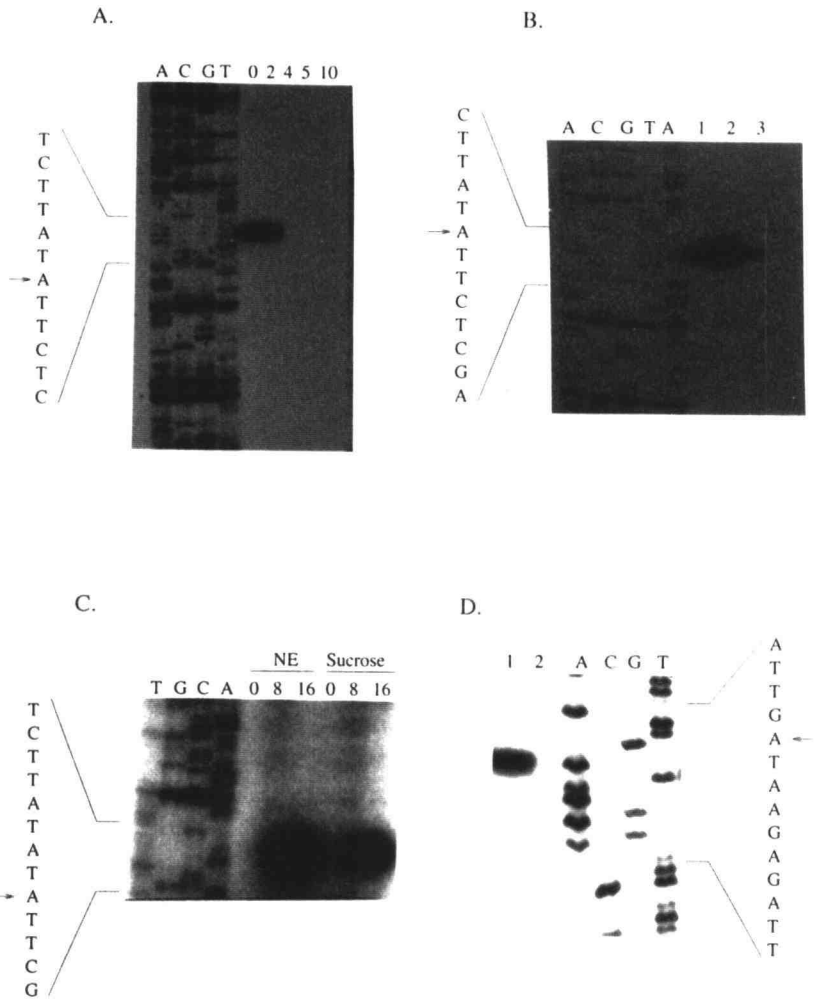


Figure 5.4. Optimization of Late *in vitro* Transcription Conditions. A. Effect of addition of EDTA. Numbers above the lane indicate concentration of EDTA in mM. B. Template preparation methods. Lane 1 and 2, pCR5 isolated from two separate CsCl gradients; lane 3, pCR5 isolated using alkali lysis procedure. C. Nuclear extracts prepared with (lanes 3 to 6) and without (lanes 1 to 3) a sucrose cushion. D. Heat inactivation of 16 hr p.i. nuclear extracts. Lane 1, 16 hr p.i. nuclear extract; Lane 2, heat–treated 16 hr p.i. nuclear extract.

II–containing nuclear extracts was preferentially inactivated under the same conditions. Using column fractions to supplement heat–treated extracts, they were able to purify TFIID approximately 300–fold [214]. Perhaps, TBP and its associated factors are the heat–labile components required for baculovirus late *in vitro* transcription. Further purification of the late transcription complex, possibly using supplementation of heat–inactivated extracts, would answer this question.

5.5 Nuclear Extract Time Course

The ability of the *in vitro* transcription system to mimic *in vivo* baculovirus gene expression using extracts prepared 0 to 40 hr p.i. was examined. Several baculovirus gene promoters that included the early and late promoters from the *39K* gene, the late promoters from the *vp39* gene and the hyperexpressed *p10* late promoter were assayed for *in vitro* transcriptional activity. The plasmids containing these promoters used in the *in vitro* assays are diagrammatically represented in Figure 5.5.

Expression of the *39K* gene is controlled by adjacent promoter elements; early expression is controlled by two TATA–box elements and late expression initiates within an ATAAG promoter sequence that overlaps the proximal TATA box element [99, 100]. Late transcription is completely dependent upon the presence of the TAAG motif [100]. Primer extension analysis of the *in vitro* transcripts generated using the *39K* promoter containing construct, pCR1, are shown in Figure 5.6a. Accurately initiating transcripts from the *39K* early promoters are seen in 0, 8 and 12 hr p.i. but early transcripts are not detected past 16 hr p.i. However, transcripts initiating within the late promoter element ATAAG are detected at 16 and 24 hr p.i.

Primer extension analysis of *in vitro* transcripts generated using the *vp39* promoter–containing construct, pCR5, are shown in Figure 5.6b. The *vp39* gene promoter contains three late promoter elements that all serve as sites of tran-

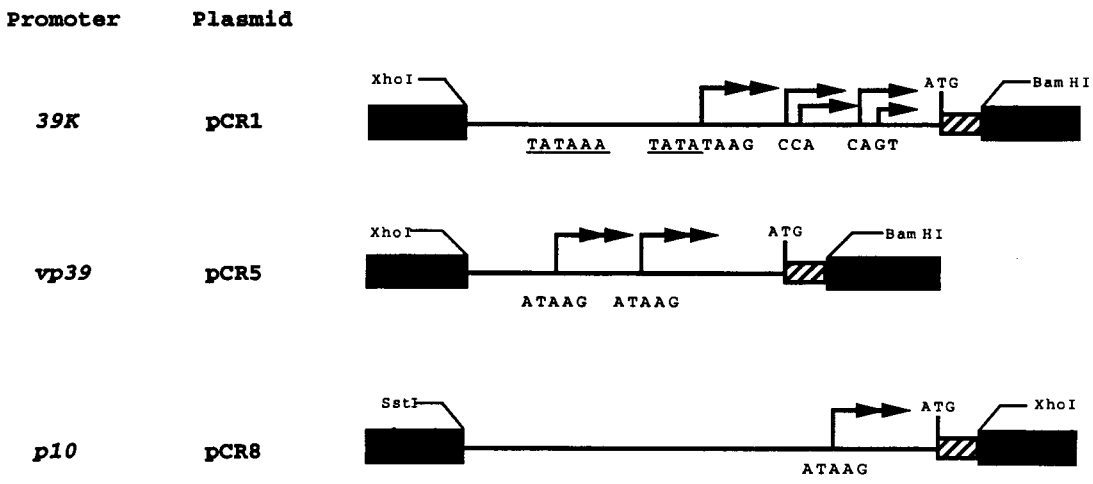


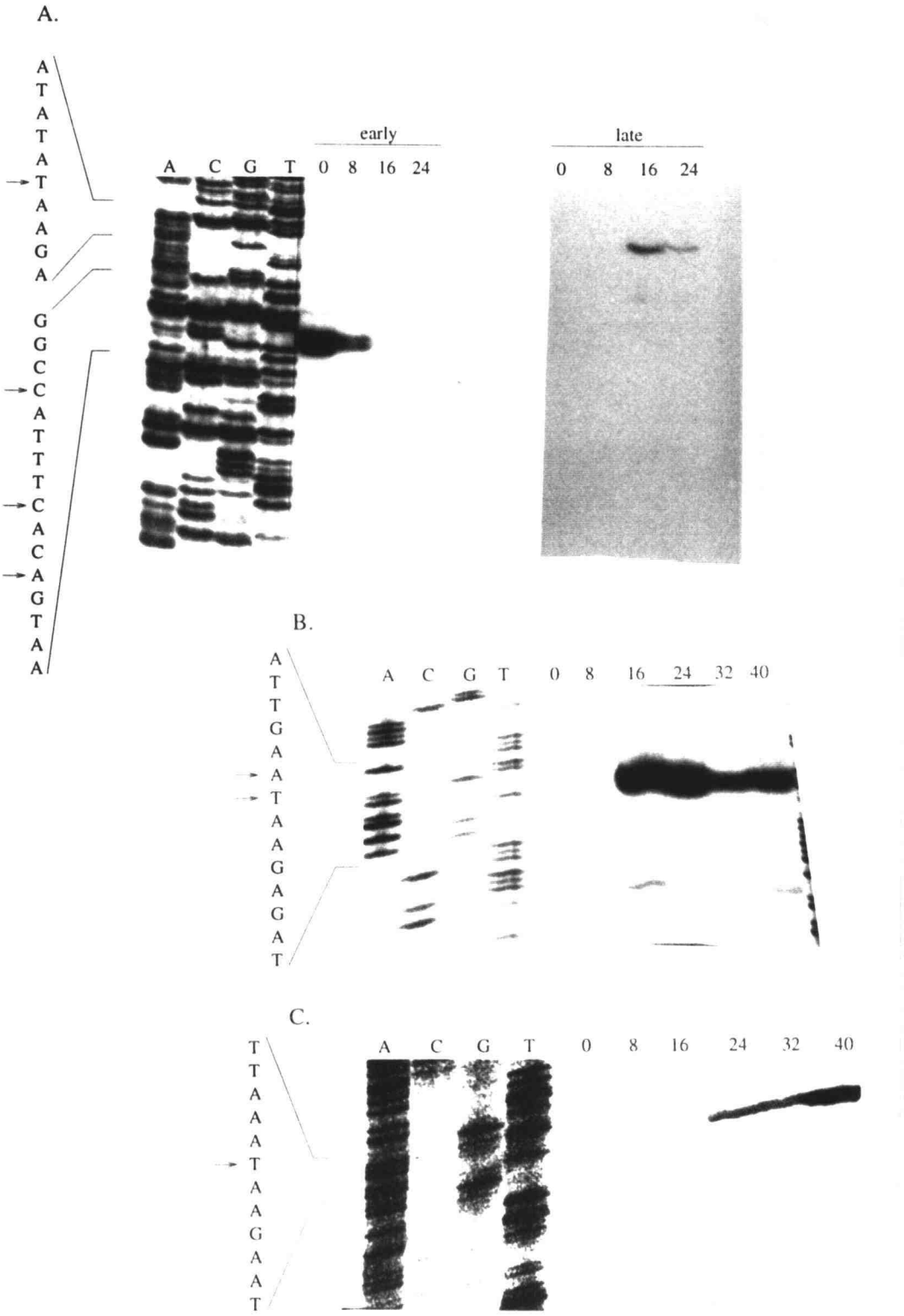
Figure 5.5. Schematic Diagrams of the *39K*, *vp39*, and, *p10* Promoter Region Constructs. pCR1 contains a 384 bp *BssH*II–*Sst*I fragment encompassing both early and late promoters from the *39K* gene. pCR5 contains two of the three late promoters of *vp39*. The pCR8 construct contains a 500 bp *Hinc*II–*Sst*I fragment encompassing the *p10* promoter. Single arrowheads indicate early transcription initiation sites while double arrowheads indicate late transcription initiation sites. Important consensus sequences are shown and the TATA elements are underlined. The crosshatched box represents coding sequence and thick black lines represents vector sequence.

scription initiation *in vivo* [301]. The pCR5 construct contains two of the three late promoter elements (See Figure 5.5). Transcription was detected from the *vp39* template between 16 and 40 hours post infection; maximal expression occurred between 16 and 24 hr p.i. (See Figure 5.6b).

The pCR8 construct contains the *p10* promoter region and was used as the DNA template in late *in vitro* transcription reactions. *p10* is a hyperexpressed late gene; *p10* transcript levels increase and remain high even after the levels of expression of other late genes decreases [16]. *p10* transcripts were detected by primer extension analysis 16 hr p.i. and were maximally expressed in 40 hr p.i. nuclear extracts. (See Figure 5.6c).

These results show that early, late and very late baculovirus gene promoters are transcribed *in vitro* in a temporal manner similar to the cascade of viral gene expression seen during the course of baculovirus infection. Furthermore, these data indicated that nuclear extracts prepared from Sf9 cells 16 hr p.i. were optimal for *in vitro* late gene transcription while extracts prepared 32–40 hr p.i. were optimal for very late gene transcription.

Figure 5.6. Primer Extension Analysis of *in vitro* Transcripts of Early, Late and Very Late Promoters. Transcripts were generated using nuclear extracts prepared from AcMNPV–infected Sf9 cells various times post infection. Numbers on the top of the lanes indicate the times post–infection that the nuclear extract was prepared. The sequencing ladders on the left were generated using the same template and primer used in the primer extension reactions. *In vitro* transcription of the A) *39K* promoter using early and late reaction conditions and B) *vp39* promoter using late conditions and C) *p10* promoter using late reaction conditions.


A.
A T A T A T A A G A G G C C A T T T C A C A G T A A A
early
A C G T 0 8 16 24
late
0 8 16 24
B.
A T T G A A T A A G A G A T
A C G T 0 8 16 24 32 40
C.
T T A A A T A A G A A T
A C G T 0 8 16 24 32 40

Clear differences between the *in vivo* expression of late and very late genes *in vivo* have been reported [301, 205, 226]. In particular, expression from *vp39* late promoters was detected between 12 hr p.i. and 24 hr p.i. with transcript levels declining by 48 hr p.i. [301, 205]. However, *polh* transcripts were not detected until 24 hr p.i. after which time transcript levels increased dramatically [301]. Understanding the mechanism of differential transcription of late and very late gene promoters during the course of baculovirus infection is of interest. My work shows that differences in the expression of late and very late genes seen *in vivo* can be reproduced using the *in vitro* transcription system. Subsequent fractionation and reconstitution of active components of these nuclear extracts at selected times post–infection may lead to the identification of viral and/or host factors responsible for temporal transcription of baculovirus genes.

5.6 Alignments of Late and Very Late Promoters

Consensus sequences for molecular binding sites such as promoter regions can often be determined by aligning promoter regions, and then choosing the most common bases at each site to create the consensus sequence. However, information is lost when the relative frequency of bases at each position is ignored. Schneider *et al.* [274] developed a function that can be used to assess the information content of the sites, that is the frequency of base pair occurrence at individual positions within a set of sites is included in the evaluation.

Information is measured in bits; one bit of information is capable of answering one yes/no question. Assuming that a sequence of nucleotides is composed of an equal number of A's, C's, G's and T's, one bit of information is needed to answer the question "Is the base a purine?". Two bits of information is required to ascertain the exact base at a given position (Is the base an A?). In order to transmit the sequence to another person using only yes/no questions , we need two bits of information for each position in the sequence. If another sequence is

composed of an equal number of A's and T's with no C's and G's, this sequence can be transmitted to another person using only one yes/no question or one bit of information for each position in the sequence (Is the base an A?).

The information in one column of a multiple sequence alignment can be computed by considering the bases in the column as a string and computing the information of that string. By computing the information content at each position of the multiple sequence alignment, the important positions within the alignment can be found even when the column is not composed of identical bases.

The uncertainty of a column S, in a multiple sequence alignment, is

$$E(S) = - \sum_{x \in \{A,C,G,T\}} \frac{N(x,S)}{L} \log_2 \left(\frac{N(x,S)}{L} \right)$$

where $N(x,S)$ is the number of the bases x in the column S and L is the length of the column S. The information of the column S is

$$I(S) = e(L,\mathcal{D}) - E(S)$$

where $\mathcal{D}$ is the set of all sequences in the multiple alignment and $e(L,D)$ is a function dependent on the maximum uncertainty of a column of length L and a distribution of bases equal to the distribution of bases in the sequences in the set $\mathcal{D}$ [274]. Computer programs developed by Dr. J. Holloway were used to compute and display the information content of multiple sequence alignments of baculovirus late and hyperexpressed late promoters.

Alignments of the *p10* promoter region from five different baculovirus sequences, available from Genbank as of June 6, 1995, were generated using the clustal V alignment program [115]. The computer–generated alignments were modified slightly and the information content of the aligned sequences was determined as described above. The *p10* promoter alignment is shown in Figure 5.7 and the information content is graphically displayed in Figure 5.8. The results of this analysis suggest that sequences immediately downstream of the ATAAG sequence may be important for gene expression. The sequence is very AT rich

and seems to be composed of alternating TA base pairs. Deletions of the AcMNPV *p10* 5' leader sequence that includes this region resulted in a substantial decreases in *p10* expression as measured by CAT activity [319]. These data suggest that this region is important in regulation of *p10* expression. Furthermore, this region contains a consensus palindromic sequence that could potentially result in a short cruciform structure that includes the ATAAG sequence. Interestingly, a GGvCCTyTdr sequence between 40 and 70 bp upstream of the ATAAG sequence also appears to be highly conserved. A limited study of the AcMNPV *p10* promoter region indicated that a deletion removing 13 bp past this sequence affected transient CAT expression three– to fourfold [319]. A similar motif, CCTwTyGT is found 10 bp downstream of the polyhedrin late promoter sequence rTAAG (see below).

Alignments of the *polh* promoter region from 13 sequences available in Genbank on June 6, 1995 were performed in a manner similar to that described for *p10*. The *polh* promoter alignment is shown in Figure 5.9 and the information content is graphically displayed in Figure 5.10. The alignment indicates that sequences downstream of the rTAAG sequence may be important for *polh* expression. The sequence is AT rich, having a high conservation of T residues immediately downstream while further downstream A residues appear to be conserved. Nested within the conserved sequence is a TnTNGTA sequence that has been previously suggested to be similar to the 5sRNA and tRNA promoters [335]. Linker scanning analysis of this region of the AcMNPV *polh* gene showed that sequence changes in these regions resulted in decreased expression from the *polh* promoter [226].

Forty–four exclusively late promoters regions from the complete sequence of the AcMNPV genome [7] could not be aligned using clustal V [115] and as a result were aligned by hand. Information was calculated from these alignments and the results are displayed in a logo format [275] in Figure 5.11. The rTAAG

```
          1          11          21          31          41
          |           |           |           |           |
   AcMNPV  GGACCTTTAA TTCAACCCAA CACAATATAT TATA------ ----------
    BmNPV  GGACCTTTAA TTCAGCCCAA CACAATATAT TACA------ ----------
    OpNPV  GGCCCTCTTG TAAGCCACGG T--------- ---------- ----------
    CfNPV  GGGCCTTTTG CTGGCCACAA C--------- ---------- ----------
    SeNPV  GGACCTCTGA TTACGAATCA GAACAAAGAC GACGACGACG ACGACAGTAT
consensus  GGvCCTyTdr ttvvvcmcva bammawakay kayr------ ----------

         51          61          71          81          91
          |           |           |           |           |
   AcMNPV  ---------- ----GTTAAA TAAGAATTAT TATC-AAATC ATTTGTATAT
    BmNPV  ---------- ----GCTAAA TAAGAATTAT TATT-AAATT ATTTGTATAT
    OpNPV  ---------- ----TTTAAA TAAGCACTAT TATA-AAATA TTAA-TATAT
    CfNPV  ---------- ----GTTAAA TAAGCATTAT TAAC-TAATT ATTATTATAT
    SeNPV  TGGCAATATC GTATAGAGAA TAAGT-TTAT TATTATAATT GTAATTATAT
consensus  ---------- ----dbtaAA TAAGhatTAT TAth-wAATh dTwwkTATAT

        101         111         121         131         141
          |           |           |           |           |
   AcMNPV  TAATTAAAAT ACTATACTGT AAATTACATT TTATTTACAA TCATG
    BmNPV  TAATTAAAAT CTTATACTGT AAATTACATT TTATTTACTA TCATG
    OpNPV  TTACCAAACT GACTGCAATA TGATCATGTC CAAGCCCAGC ATTTT
    CfNPV  TAAGATAACA ATTTCAATAT CATCATGTCC AAGCCCAGCA TTTTA
    SeNPV  TATACATTAT GAGTCAAAAT ATTTTACTTT TGATCCGAGC CGACA
consensus  Taanhaaamt vhbwbamwdt hdwthabdty hdabyyvvnm hbwtd
```

Figure 5.7. Alignment of *p10* Hyperexpressed Late Promoter Region. *p10* sequences avaibale in Genbank on June 6, 1995 were used in the analysis. Ac, *Autographa californica* MNPV; Bm, *Bombyx mori* MNPV; Op *Orygia pseudotsugata* MNPV and SNPV (sequences are identical so only one was included in the analysis); Cf, *Choristoneura fumiferana*; Se *Spodoptera exigua.* The consensus is listed below the alignment; a capital A, C, G, or T indicates that the nucleotide at that position is completely conserved; a lower case a, c, g or t indicates that the nucleotide 80% conserved and the remaining nucleotide combinations are indicated in the IUB/GCG nomeclature. See Appendix 2.

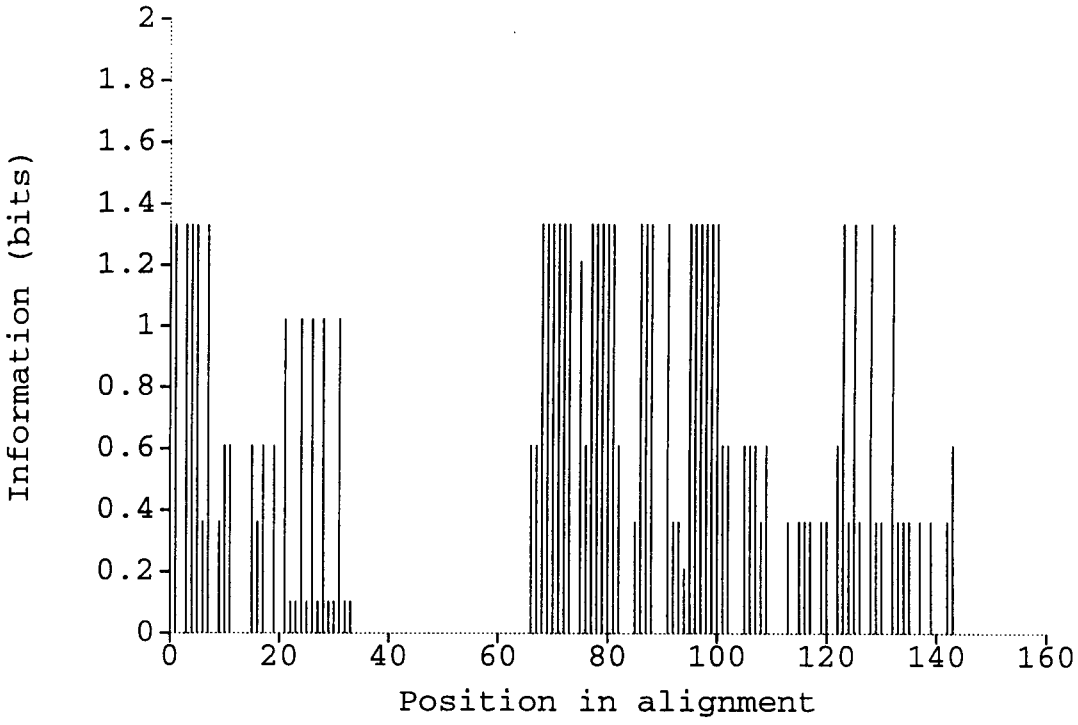


Figure 5.8. Information Content of the *p10* Promoter Region. The information content of each position in the alignment of the *p10* promoter region was determined. The y axis is the information content of that base position in the alignment measured in bits and the x axis is the position of the base in the *p10* promoter region corresponding to alignment generated in Figure 5.7. The late promoter sequence, ATAAG begins at position number 70.

```
          1          11          21          31          41          51
          |           |           |           |           |           |
   HcNPV  GGGTTGAATA AAATGCGTTT TTAAACATGT TTTTTTAATA ATTATAAGTA ATTTGCTGTT
   ApNPV  AACGTTTTTG ATTTACATCT -TTTATTTAC CTTTAAATAG ATAATAAGTA CATTGCTGTT
   BmNPV  TTATTAACGA TACAAATGGA AATAATAACC ATCTCGCAAA TAAATAAGTA TTTTACTGTT
  AcMNPV  GATATCATGG AGATAATTAA AATGATAACC ATCTCGCAAA TAAATAAGTA TTTTACTGTT
  SlMNPV  TATTGACATG CGATTTCTCA ATACCAATGA AGATCAAGTG ATGATAAGGA ATTTATTACT
   BsNPV  GATGTTTTAT AAATTATCAA AATTTGTTCA TTTACAATCT TCAATAAGTA TTTTTTTCCT
   LdNPV  CTAATCTTTT ATTTGTCGCG ATCGGGCGCG CGGGCGGTCT CCAATAAGTA TTTTATTCTT
  SfMNPV  TTTGCCGCAT TTTTATCTAA TCTTTTGCCG CACTGCGGGA ATTGTAAGTA ATTTTTTCCT
   SeNPV  TTTATCATCG ACGACGATAA GTAATGATAT GGTTGCGGGA ATTGTAAGTA ATTTTTTCCT
  PfMNPV  GTGTCGGTTT CGTCGACGCG TATTTATATG TTTTTTGGGA AATGTAAGTA ATTTTCTCCT
   MbNPV  GCTTCGGTAT AACCGACGCG TATTTATACG TTTTTTGGGA AATGTAAGTA ATTTTCTCCT
  OpMNPV  AAAATAAAAC ACTAGTTACT ATT-GGCGTT TCGTTTTTTA TTAATAAGTA ATTTCCTGTT
  PnMNPV  AAAATAAAAC ATTGGTTATT ATTTGGCGTT TTGTTTTTTA TTAATAAGTA ATTTGCTGTT
  OpSNPV  GAGTAATTCG ATTTTTGCGT GAGAATTTCA ACGACACACC TCAATAAGTA TTTTTGTCCT
consensus nnndnnnhnn hnnnnnnnnd dhnnnnnnnn nnntnnndnn hhdrTAAGtA htTTnbTvyT

         61          71          81          91         101
          |           |           |           |           |
   HcNPV  ATTGTAGCAA -TTTTGTAAT -AAAAATAT- CCTATAACT- --ATG
   ApNPV  ATTGTAGCAA -CTTTCTAGT --AAAATTT- CCTATAACT- --ATG
   BmNPV  TTCGTAACAG -TTTTGTAAT AAAAAAA--- CCTATAA--A T-ATG
  AcMNPV  TTCGTAACAG -TTTTGTAAT AAAAAAA--- CCTATAA--A TAATG
  SlMNPV  ATCGTTCTAG ATAGTGAAAA ATCAAATATC CC-------A TAATG
   BsNPV  ATTGTAAAAC ATTGTGAAAA ATCAAATACA AC-------A TAATG
   LdNPV  TTCGTAAAGA TTTTGGAAAA ATCAAATACA CCGT-----A AAATG
  SfMNPV  TTCGTAAAAC ATTGTGAAAA AATAAAT--- ---------A TAATG
   SeNPV  TTCGTAAAAC ATTGTGAAAA AATAAAT--- ---------A TAATG
  PfMNPV  TTCGTAGAAG ATTGTGAAAA ATAAAAT--- ---------A TAATG
   MbNPV  TTCGTAGAAG ATTGTGAAAA ATAAAAT--- ---------A TAATG
  OpMNPV  ATTGTAACAA -TTTTGTAAA A--AAATTT- CCTATAACC- --ATG
  PnMNPV  ATTGTAACAA -TTTTGTAAA A--AAATTT- CCTATAACC- --ATG
  OpSNPV  TTCGTAAAAC ATTGTGAAAT TTCAAATACA CC-------A TAATG
consensus wTyGTavhav wttktgwAaw awhAAAtwym mckwtaacya waATG
```

Figure 5.9. Alignment of the *polh* Hyperexpressed Late Promoter Region. *polh* sequences available in Genbank on June 6, 1995 were used in this analysis. Hc, *Hyphantria cunea* NPV; Ap, *Antheraea pernyi* NPV; Bm, *Bombyx mori* MNPV; Ac, *Autographa californica* MNPV; Sl, *Spodoptera littoralis* MNPV; Bs, *Buzura suppressaria* NPV; Ld, *Lymantria dispar* NPV; Sf, *Spodotera frugiperda* MNPV; Se, *Spodoptera exigua* NPV; Pf, *Panolis flammea* MNPV; Mb, *Mamestra brassicae* NPV; Op, *Orygia pseudotsugata* MNPV; Pn, *Perina nuda* MNPV and Op, *Orygia pseudotsugata* SNPV. The consensus is listed below; a capital A, C, G, or T indicates that the nucleotide at that position is completely; a lower case a, c, g or t indicates that the nucleotide 80% conserved and the remaining nucleotides are indicated using the IUB/GCG nomeclature. See Appendix 2.

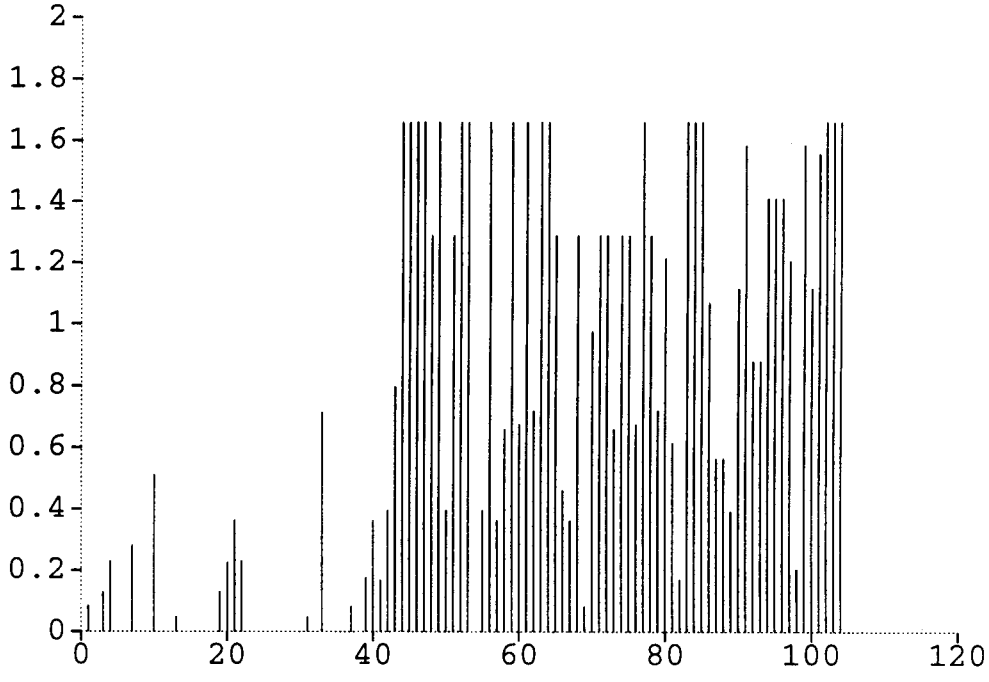


Figure 5.10. Information Content of the *polh* Promoter Region. The information content of each position in the alignment of the *polh* promoter region was determined. The y axis is the information content of that base position in the alignment measured in bits and the x axis is the position of the base in the *polh* promoter region corresponding to alignment generated in Figure 5.9. The late promoter sequence, rTAAG begins at position number 44.

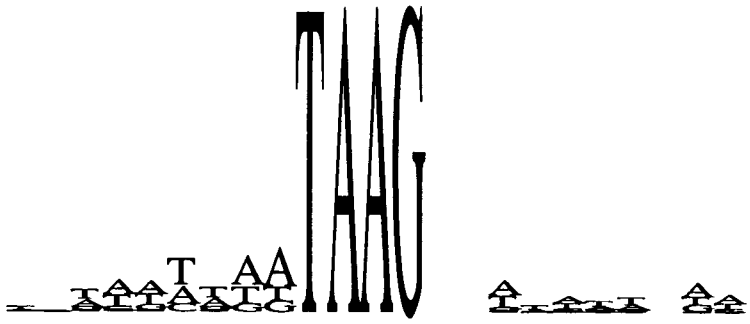


Figure 5.11. Information Content of 44 AcMNPV Exclusively Late Promoters. The information content determined by aligning 44 exclusively late promoters from AcMNPV is displayed in logo form.

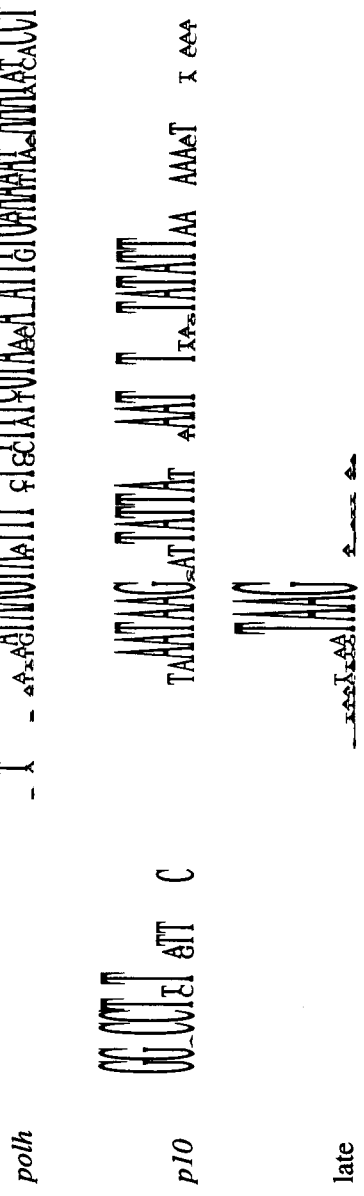


Figure 5.12. Information Content for the *p10*, *polh* and late promoters. The information content is displayed in logo form.

sequence is absolutely conserved and interestingly, the exclusively late promoters seem to have A/T rich sequences immediately preceding the late promoter motif. Alteration of this sequence by replacing the AT rich region with GC base pairs could be used to evaluate this hypothesis.

Figure 5.12 displays the conserved regions of the *p10* and *polh* promoter regions and illustrates the differences between the different late and hyperexpressed late promoters. The AT richness in front of the dTAAG promoter may be a requirement for all late promoters. Addition of more information may be required for hyperexpressed lates. The different motifs seen in the *p10* and *polh* promoters may be responsible for the dissimilar expression of the *polh* and *p10* genes [260]. Addition of these motifs to an exclusively late promoter and monitoring the level of expression is one approach to understanding the mechanism of late and very late gene expression.

5.7 *In vitro* transcription analysis of the AcMNPV *vp39* Late Promoter

To begin to dissect the differences between late and very late gene expression, deletion analysis was performed on the *vp39* promoter. The *vp39* promoter was chosen because it was readily expressed *in vitro* in nuclear extracts prepared 16 hr p.i. and did not contain any early promoters that could confound the results of the experiments. Previous studies on the *vp39* promoter region had shown that *vp39* contains three TAAG sequences. Each TAAG sequence serves as a site of transcription initiation *in vivo* [302]. Deletion analysis of the *vp39* promoter region showed that the ninety–nine nucleotides upstream of the initiator methionine, containing only a single late promoter, was sufficient for late expression [301].

Initially, the plasmid pCR5, containing two of the three late promoter elements, was constructed. Deletion of the upstream promoter did not appear to

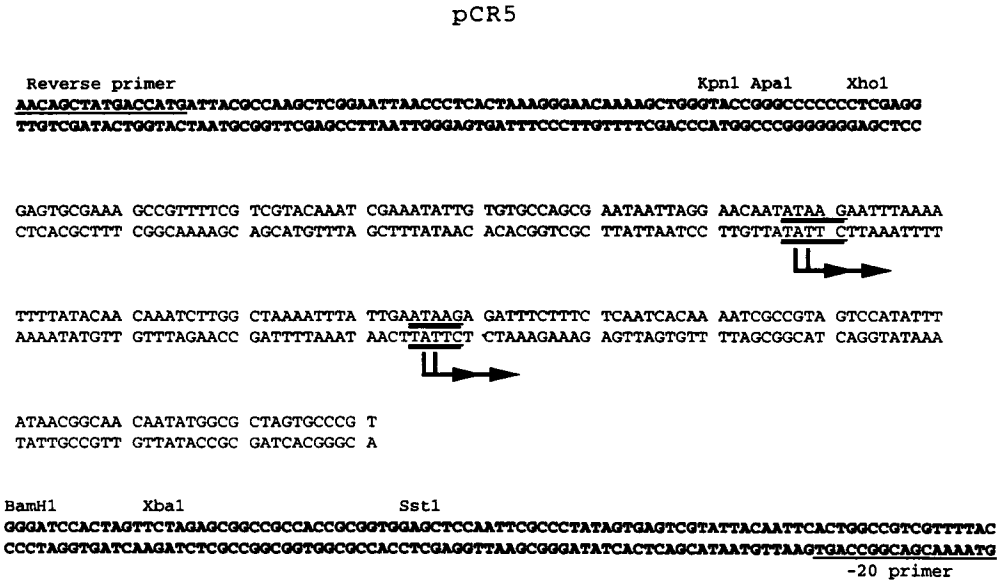


Figure 5.13. Nucleotide Sequence of the pCR5 *vp39* Late Promoter Construct. The two late promoter sites are underlined, and the transcription start sites and direction of transcription are indicated by the arrow. pKS– vector sequences are indicated in bold type. Restriction endonuclease sites denoted are in the multiple cloning region of pKS–. *Apa*I and *Xho*I were used to create exonuclease III deletions in one direction and *Sst*I and *Xba*I were used for the other direction.

have deleterious effects on *in vitro* transcript initiation from the remaining two late promoters (data not shown). Figure 5.13 displays the nucleotide sequence of the *vp39* promoter containing region of pCR5.

Exonuclease III was used to create 5' and 3' deletions of one of the *vp39* late promoter regions as described in section 2.12. pCR5 was digested with *Apa*I and *Xho*I to delete from the 5' end of the proximal ATAAG and *Sst*I and *Xba*I were used to delete from the 3' region (See Figure 5.13). Using this protocol, the 5' deletions contained a single late promoter element located nearest the initiator methionine. The 3' end deletions contained two late promoters; an upstream ATAAG promoter (see Figure 5.13) and the late promoter nearest the methionine that was subjected to deletion analysis. A number of clones

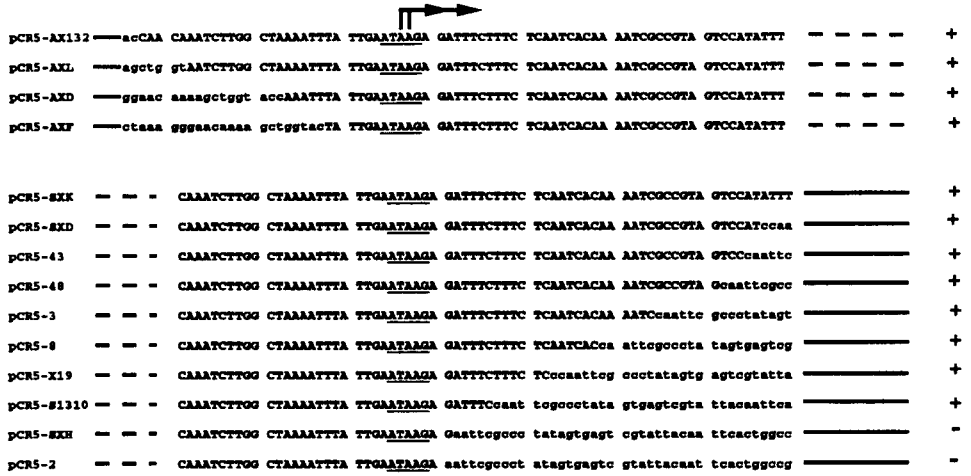


Figure 5.14. Nucleotide Sequence of the *vp39*–derived Constructs used in *in vitro* Transcription Analysis. The sequences of *vp39* deletion clones constructed using Exonuclease III are shown. The names of the clones are indicated on the left. The *vp39* nucleotide sequences are shown in capitol letters while the sequences derived from the pKS– vector are indicated by lower case letters. The solid line represents a continuation of the vector sequences and the dashed line represents the remainder of the *vp39* promoter region contained in the clone. A + indicates that an *in vitro* transcription signal was produced using this template whereas - indicates that no signal was detected.

were sequenced and *in vitro* transcription from the templates that contained the complete ATAAG sequence but had portions of the either the 5' or 3' flanking sequence deleted was assayed using primer extension. Figure 5.14 shows the sequence of the clones assayed for transcriptional activity using the late *in vitro* system.

Primer extension analysis of *in vitro* generated transcripts from the clones pCR5–AX132, pCR5–AXL, pCR5–AXD and pCR5–AXF is shown in Figure 5.15. When used as DNA templates in late *in vitro* transcription reactions, these four deletion constructs generated transcripts that initiated within the ATAAG promoter. A deletion as close as 6 bp upstream (pCR5–AXF) of the ATAAG sequence still resulted in the production of accurately initiating late transcripts

(Figure 5.15). Examination of the pCR5–AXF sequence showed that pCR5–AXF retained an AT rich sequence immediately upstream of the late promoter transcription initiation site. Information content analysis (See section 5.6) of AcMNPV late promoters indicated that sequences similar in nucleotide composition are found 6 bp preceding the ATAAG motif.

Deletions 3' to the *vp39* late promoter were also assayed(Figure 5.16). Due to the nature of these deletions constructs, primer–extension products became progressively shorter the closer the deletion was to the ATAAG promoter sequence. The lengths of the expected primer–extension products from the *in vitro* generated transcripts are indicated in Table 5.1. Accurately initiating transcripts were produced when plasmids pCR5–SXK, pCR5–SXD, pCR5–43, pCR5–48, pCR5–3, pCR5–8, pCR5–X19 and pCR5–S1310 were used as DNA templates. However, plasmids pCR5–SXH and pCR5–2 failed to produce any accurately initiating late transcripts when assayed using the *in vitro* transcription system. These results, shown in Figure 5.16, indicate that accurate initiation from the baculovirus *vp39* late promoter requires a maximum of 6 bp of sequence downstream from the ATAAG sequence; deletion of 4 additional base pairs resulted in cessation of late transcript production. Examination of the pCR5–SXH and pCR5–2 sequences indicated that a CGCCC sequence is found 5 bp downstream of the late transcription initiation site in both these clones. Information content analysis (See section 5.6) of AcMNPV late promoters indicated that AT rich sequences are common in this region. Alteration of these sequences may have contributed to the loss of the late transcription signal from these clones.

The primer extension analysis of the *vp39* late promoter deletion constructs delineated a minimal late promoter sequence necessary for accurate transcription initiation. The minimal late promoter element contained 6 bp upstream and 6 bp downstream of the consensus ATAAG. Oligonucleotides were synthesized (Center for Gene Research and Biotechnology, Oregon State University) containing the

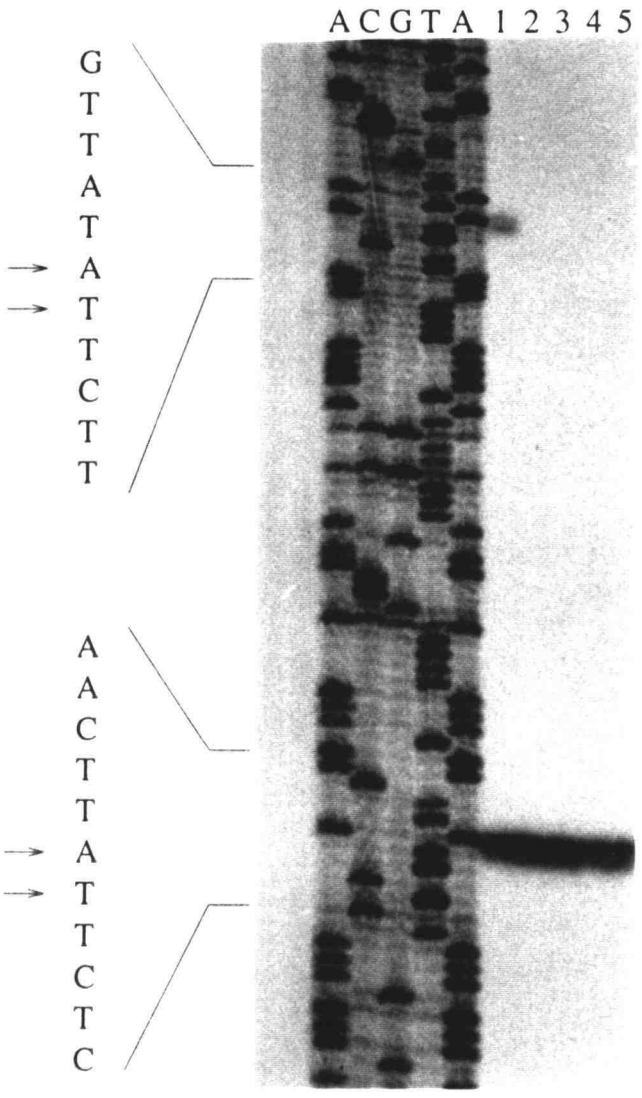


Figure 5.15. *In vitro* Transcription Analysis of the 5' *vp39* Deletion Constructs. The *vp39* deletion clones are as indicated. The arrows mark the position of the mRNA transcription start sites. A sequencing ladder generated from the pCR5 plasmid using the same primer, the M13 forward primer, as used in the primer extension reactions was used as size standards. A. Lane 1, pCR5; lane 2, pCR5–AX132; lane 3, pCR5–AXL; lane 4, pCR5–AXD; Lane 5, pCR5–AXF.

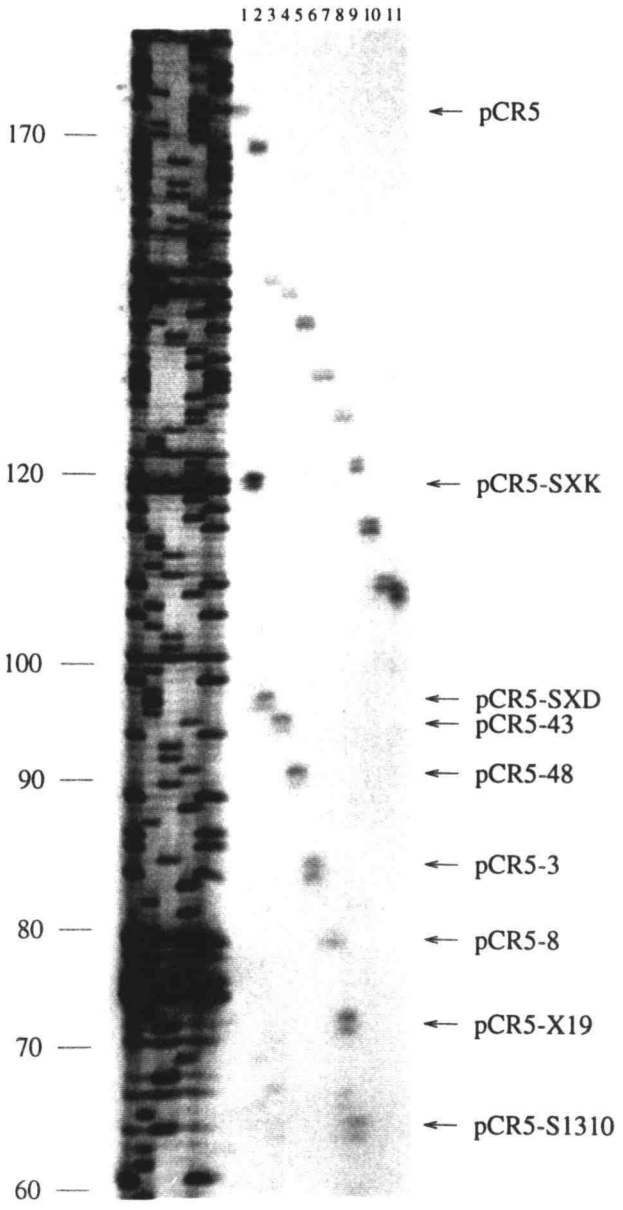


Figure 5.16. *In vitro* Transcription Analysis of the 3' *vp39* Deletion Constructs. The *vp39* deletion clones are as indicated. The arrows mark the position of the mRNA transcription start sites. A sequencing ladder generated from the pCR5 plasmid using the same primer, the M13 forward primer, as used in the primer extension reactions was used as size standards. Lane 1, pCR5; lane 2, pCR5–SXK; lane 3, pCR5–SXD; lane 4, pCR5–43; lane 5, pCR5–48; lane 6, pCR5–3; lane 7, pCR5–8; lane 8, pCR5–X19; lane 9, pCR5–S1310; lane 10, pCR5–SXH; lane 11, pCR5–2.

Plasmid Name	Expected Lengths	Actual Lengths
pCR5–SXK	116 and 117 bp	116 and 117 bp
pCR5–SXD	94 and 95 bp	94 and 95 bp
pCR5–43	91 and 92 bp	91 and 92 bp
pCR5–48	88 and 89 bp	88 and 89 bp
pCR5–3	81 and 82 bp	81 and 82 bp
pCR5–8	75 and 76 bp	75 and 76 bp
pCR5–X19	70 and 71 bp	70 and 71 bp
pCR5–S1310	63 and 64 bp	63 and 64 bp
pCR5–SXH	57 and 58 bp	0 bp
pCR5–2	55 and 56 bp	0 bp

Table 5.1. Expected and Actual Lengths of the 3' *vp39* Deletion Clone Primer–extension Products. Primer–extension products were generated using the 3' *vp39* deletion clones in the late *in vitro* transcription system.

minimal late promoter sequence (oligo*vp39*A and oligo*vp39*B) and a mutated version (oligo*vp39*AM and oligo*vp39*BM) in which the ATAAG was changed to AAAAA. Oligonucleotides *vp39*A and *vp39*B and oligonucleotides *vp39*AM and *vp39*BM were annealed and cloned into pKS–. The resultant clones, pCR7 and pCR7–M, were assayed for transcriptional activity. Figure 5.17 shows the sequence of the oligonucleotides used to construct pCR7 and pCR7–M. As shown in Figure 5.18, transcripts were detected from the minimal late promoter but not from the mutant late promoter indicating that a functional baculovirus late promoter consists of 17 bp including the consensus ATAAG sequence.

During the course of this work, Morris and Miller [207] reported an *in vivo* mutational study of the *vp39* promoter. Using linker scanning analysis, they showed that 8 bp upstream and 6 bp downstream of the consensus ATAAG were necessary for minimal expression from the *vp39* promoter when fused to the chloramphenicol acetyltransferase (CAT) reporter gene. My *in vitro* results indicate that the *in vitro* late transcription system accurately reflects *in vivo* baculovirus gene expression.

Figure 5.17. Nucleotide Sequence of the *vp39*AB and *vp39*ABM Oligonucleotides. This oligonucleotides were used to construct the minimal *vp39* late promoter clone, pCR7, and the mutated *vp39* promoter, clone pCR7–M. Bold type indicate the sequences derived from the *vp39* promoter region and plain type shows the additional nucleotides added to facilitate cloning into pKS–.

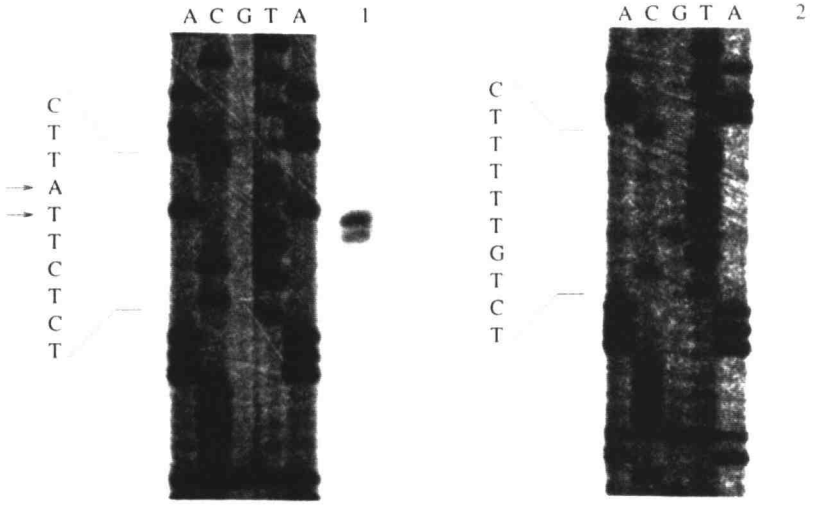


Figure 5.18. *In vitro* Transcription Analysis of a Minimal Late Promoter. The arrows mark the position of the mRNA transcription start sites. Sequencing ladders were generated using either the pCR7 or pCR7–M plasmid using the M13 reverse primer, which was the same primer used in the primer extension reactions. Lane 1, pCR7. Lane 2, pCR7–M.

5.8 Electrophoretic Mobility Shift Analysis of the *vp39* Late Promoter

Once the minimal late promoter sequence for the *vp39* gene was defined, oligonucleotides *vp39*A and *vp39*B and oligonucleotides *vp39*AM and *vp39*BM were annealed to create double–stranded oligonucleotides referred to as *vp39*AB and *vp39*ABM respectively. *vp39*AB and *vp39*ABM were then used in electrophoretic mobility shift analysis (EMSA) in an attempt to identify late promoter–specific DNA binding proteins. Nuclear extracts prepared from AcMNPV–infected Sf9 cells 0 and 16 hr p.i. were used in EMSA studies.

Reaction conditions similar to those that resulted in the production of accurately initiating late transcripts in the *in vitro* transcription system were used in EMSA assays. Therefore, binding reactions were performed in total volume of 20 μl containing 2mM Mg^{2+}, 20 mM Hepes [pH 8.4 at 25°C], 15% glycerol, and 25 mM KCl in the presence of 1 μg of non–specific competitor DNA (either sheared salmon sperm DNA or poly–dIdC) and 10,000 cpms of the labeled oligonucleotides. Figure 5.19 shows the EMSA results when increasing amounts of nuclear extract prepared 0 and 16 hr p.i. were used. Protein binding to the *vp39* minimal oligonucleotides in presence of 1 μg /μl salmon sperm DNA was observed in both 0 and 16 hr p.i. nuclear extracts. However, DNA–binding activity was much greater in 0 hr p.i. nuclear extracts than extracts prepared 16 hr p.i. suggesting that a host–encoded protein binds to baculovirus late promoters until late in infection.

To confirm that the protein complex binds specifically to the *vp39* minimal promoter region, complexes were competed by addition of increasing amounts of unlabeled *vp39*AB minimal oligonucleotides to the reaction in the presence of salmon sperm DNA (ssDNA) or poly–dIdC (See Figure 5.20). This resulted in competition of the shifted complexes, suggesting that the DNA–protein complex formation was specific. However, three complexes were seen when poly–dIdC

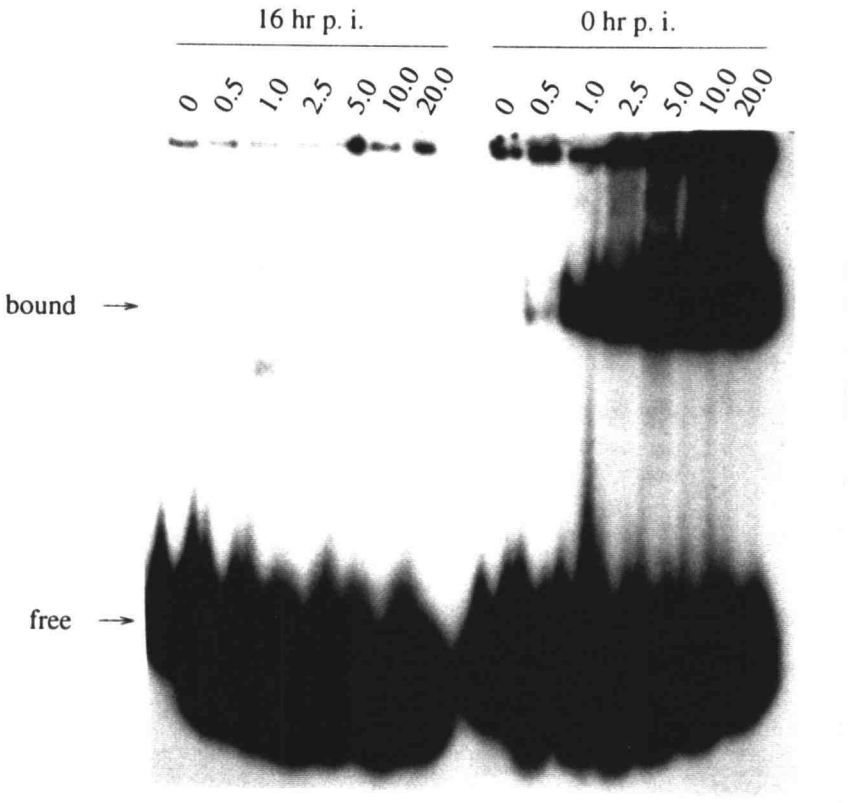


Figure 5.19. Gel Retardation Analysis of *vp39* Minimal Promoter Region. The radiolabeled oligonucleotides were incubated in the presence of increasing amounts of nuclear extract prepared 0 and 16 hr p.i. The protein concentration in μg added per 20 μl reaction is indicated above the lane.

was used as the non–specific competitor DNA whereas only one complex was observed with ssDNA. The single complex observed when ssDNA was used as the non–specific competitor appeared to have the same mobility as the fastest migrating band in poly–dIdC lanes. Identical complexes were observed using nuclear extracts prepared 0 and 16 hr p.i. although the equivalent amount of protein from 16 hr p.i. nuclear extracts contained significantly less DNA–binding activity. The difference in number of complexes is likely due to the decreased sequence complexity of poly–dIdC; ssDNA due to it's greater sequence complexity may have bind a greater number of specific and non–specific DNA binding proteins.

If a host–encoded protein is responsible for the DNA–binding activity seen in these nuclear extracts, the reduction in DNA–binding in late extracts may be due to viral–induced modification i.e. (de)phosphorylation or degradation of the host–encoded factor. To test for such activity, 0 and 16 hr p.i. extracts were combined and assayed for DNA–binding activity. As shown in Figure 5.20, mixing 2.5 μg of 0 hr p.i. and 2.5 μg 16 hr p.i. extracts resulted in the production of shifted complexes equivalent to those seen with 2.5 μg of O hr p.i. nuclear extracts indicating that the 16 hr p.i. nuclear extracts are unable to modify the DNA binding activity of 0 hr p.i. nuclear extracts.

Competition experiments using unlabeled *vp39*AB oligonucleotides suggested that the shifted complexes were specific for the *vp39* late promoter sequence. However, the shifted complex(es) were also competed by unlabeled *vp39*ABM mutant oligonucleotides and oligonucleotides of unrelated sequences (oligonucleotides, TCGAGGGTAGGGGTCAGAGGTCACTCG and its complement, were kindly provided by Dr. Mark Leid). Figure 5.21 shows that in 0 hr p.i. nuclear extracts, DNA–binding activity was competed with the *vp39*ABM mutant oligonucleotides at the same concentrations as the *vp39*AB oligonucleotides. These data indicate that the DNA–binding activity in these extracts is not spe-

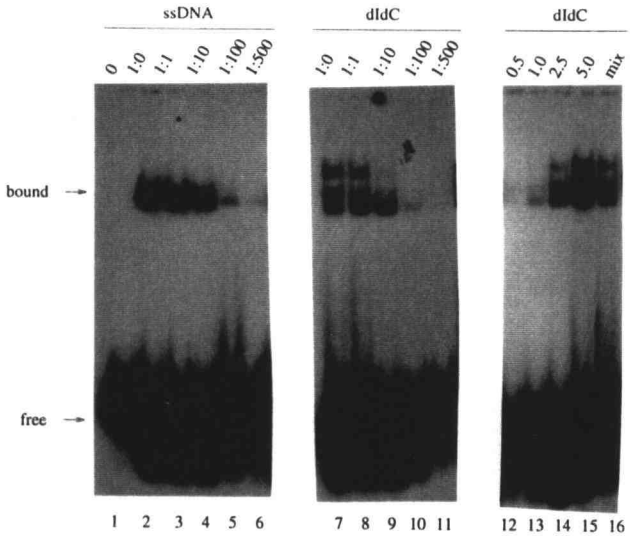


Figure 5.20. Gel retardation Competition Analysis of *vp39* Minimal Promoter Region. The radiolabeled oligonucleotides were incubated in the presence of 5 μg (lanes 1–11) or increasing concentrations of O hr p.i. nuclear extracts indicated in μg on top of lanes 12 to 15. Specific competitor DNA was added in excess molar amounts as indicated at the top of each lane (lanes 1–11) to reaction mixtures containing 1 μg of non–specific competitor DNA; either ssDNA (lanes 1–6) or poly–dIdC (lanes 7–16) as indicated. The results of mixing equal amounts of 0 and 16 hr p.i. (2.5 μg 0 hr p.i. and 2.5 μg 16 hr p.i.) nuclear extracts is shown in lane 16.

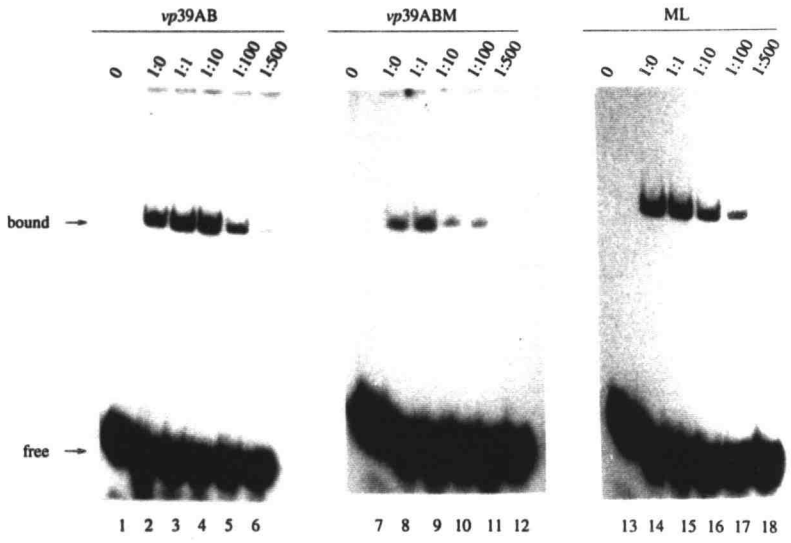


Figure 5.21. Gel Retardation Competition Analysis using Specific and Non–specific Oligonucleotides. The radiolabeled *vp39*AB oligonucleotides were incubated in the presence of 10 μg /μl nuclear extract prepared 0 hr p.i. The *vp39*AB, *vp39*ABM and ML oligonucleotides were added in excess molar amounts as indicated at the top of each lane. All binding reactions contained 1 μg of non–specific competitor DNA.

cific for the baculovirus late promoter. Furthermore, use of unrelated oligonucleotides of equivalent size also resulted in competition of the shifted complexes. Taken together, these results indicate that the DNA binding activity found using the *vp39*AB oligonucleotides is not specific for baculovirus late promoters.

Recently Burma *et al.* [26] reported that a host–encoded protein of 30 kDa bound to the AcMNPV polyhedrin gene promoter region. Using nuclear extracts prepared both 0 and 51 hr p.i., they showed that this factor bound to three overlapping fragments within the polyhedrin promoter region. Their results suggested that this 30 kDa protein was binding to an AATAAA sequence found in all three fragments. Interestingly, a shorter version of this sequence, AATAA, is found in *vp39* oligonucleotides used for gel–retardation analysis. It is possible that the 30 kDa protein identified by Burma *et al.* is binding to the

vp39 oligonucleotides. However, Burma *et al.* used a DNA fragment where the AATAAA sequence had been mutated to CCGCCC as a competitor in gel retardation assays and showed no specific competition of the shifted complex, whereas alteration of the AATAA sequence to AAAAA in the *vp39* oligonucleotides resulted in competition. Since the *vp39* mutant and wild–type promoter were assayed for *in vitro* transcriptional activity (see section 5.7) and the mutated version of the *vp39* promoter failed to produce accurately initiating transcripts, these results suggest that the DNA–binding identified in this work may not be associated with transcription of baculovirus late promoters. Further purification and fractionation of uninfected and infected nuclear extracts is necessary before the role (if any) of these host–encoded DNA–binding components can be elucidated.

5.9 Western Analysis of Nuclear Extracts

Nuclear extracts prepared from uninfected and infected (16 hr p.i.) Sf9 cells were examined using western blot analysis for the presence of SfTBP and AcMNPV IE–1. Monoclonal antibody made against the highly conserved C–terminal domain of TBP from *Drosophila melanogaster* and polyclonal antibodies to AcMNPV IE–1 were used. As Figure 5.22 shows, SfTBP is present in both uninfected and infected extracts whereas AcMNPV IE–1 is present in nuclear extracts prepared from infected but not uninfected cells. These results are as expected since SfTBP was detected in whole cell extracts until 72 hr p.i. and AcMNPV IE–1 was detected in AcMNPV–infected cells until 72 hr p.i. [223, 42]. These data are consistent with the involvement of both SfTBP and AcMNPV IE–1 in baculovirus late gene transcription but by no means establishes their involvement. Further fractionation and reconstitution of the components of the *in vitro* transcription system for baculovirus late promoters is required before both host and viral encoded late transcription factors can be identified positively.

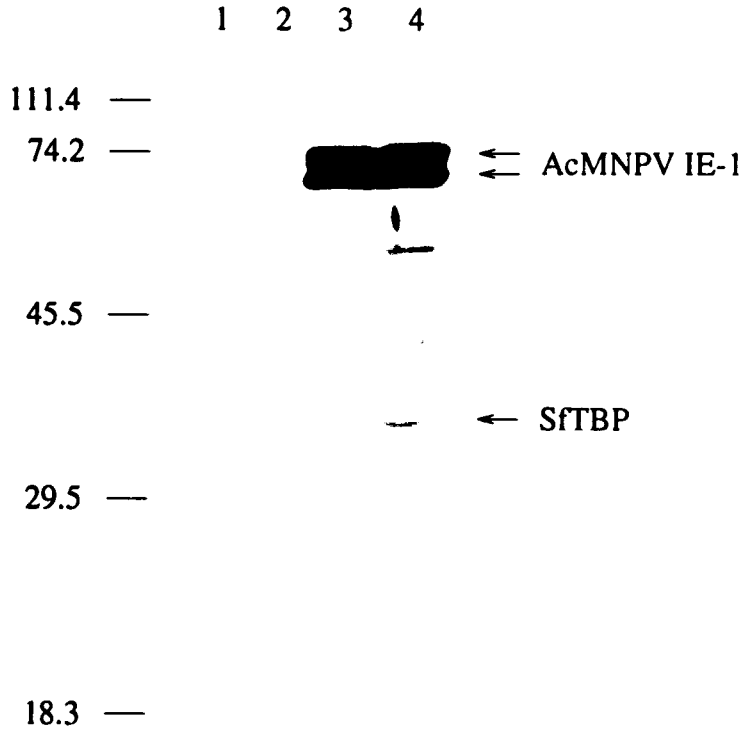


Figure 5.22. Detection of SfTBP and AcMNPV IE–1 in Uninfected and Infected Nuclear Extracts. Lanes 1 and 2 contain 30 and 35 μg , respectively, of nuclear extract prepared from uninfected Sf9 cells and lanes 3 and 4 contain 20 and 35 μg of nuclear extract, respectively, prepared 16 hr. p.i. from infected Sf9 cells. The numbers on the left indicate the marker sizes in kDa determined using the BioRad prestained low range protein ladder.

5.10 Summary

Both early and late *in vitro* transcription systems appear to mimic *in vivo* baculovirus gene expression. Using extracts prepared at various times post–infection, early, late and very late baculovirus gene promoters were shown to be transcribed *in vitro* in a temporal manner similar to that seen during the course of baculovirus infection of insect cells. Nuclear extracts prepared at 16 hr p.i. were optimal for baculovirus late gene transcription. Nuclear extracts competent for late transcription were shown by western blot analysis to contain SfTBP, a protein required by all three cellular RNA polymerases for transcription and AcMNPV IE–1, a transactivator of early gene expression and required for replication and/or late gene expression. Deletion analysis of the *vp39* late promoter identified a minimal baculovirus late promoter that can be used in further studies to identify factors required for late transcription. This minimal late promoter was used in gel–retardation assays in an attempt to identify late–promoter specific DNA binding proteins. Gel shift analysis failed to identify promoter–specific DNA binding at in nuclear extracts.

Chapter 6
Characterization of Protein–DNA Interactions at Enhancer Regions and Replication Gene Promoters

6.1 Introduction

*Hr1*a is the smallest of the *hr*s (with the exception of *hr*4c, which is not well conserved) containing only two 30–bp imperfect palindrome sequences separated by 58 bp. Characterization of plasmid subclones of the *hr*1a–containing AcMNPV *Hind*III–N fragment revealed that only plasmids containing *hr*1a underwent infection–dependent replication and were able to stimulate transcription. A schematic representation of the *Hind*III–N subcloned fragments used to study the replication, transcriptional enhancer and protein–binding activity is shown in Figure 6.1. Sequences mapping to the left of *hr*1a were required for maximal levels of replication. Plasmids containing only one half of a palindrome or disruptions of the central *Eco*RI core either did not replicate, or replicated very poorly, and did not exhibit enhanced transcriptional activity. Using gel–mobility shift analysis, whole–cell extracts made from *ie–1* transfected cells have been shown to contain a DNA–binding activity that is specific for *hr*s [97, 96]. Therefore, I attempted to correlate IE–1 binding to *hr*1a and *hr*1a derivatives with replication and transcription.

Transient expression assays have shown that the level of transcription from baculovirus promoters is modulated by the product of the *ie–1* gene alone [101, 103, 218] or in combination with *hr* elements [102, 98, 218, 28, 185, 30]. Interestingly, the *ie–1* gene product has also been shown to inhibit the expression of two other genes, *ie–2* [30] and a larger spliced form of *ie–1*, termed *ie–0* [156] while allowing continued expression of its own gene product. This is similar to the herpes simplex virus infected–cell polypeptide 4 (ICP4) which functions as both a repressor and an activator of RNA polymerase II transcription. Studies

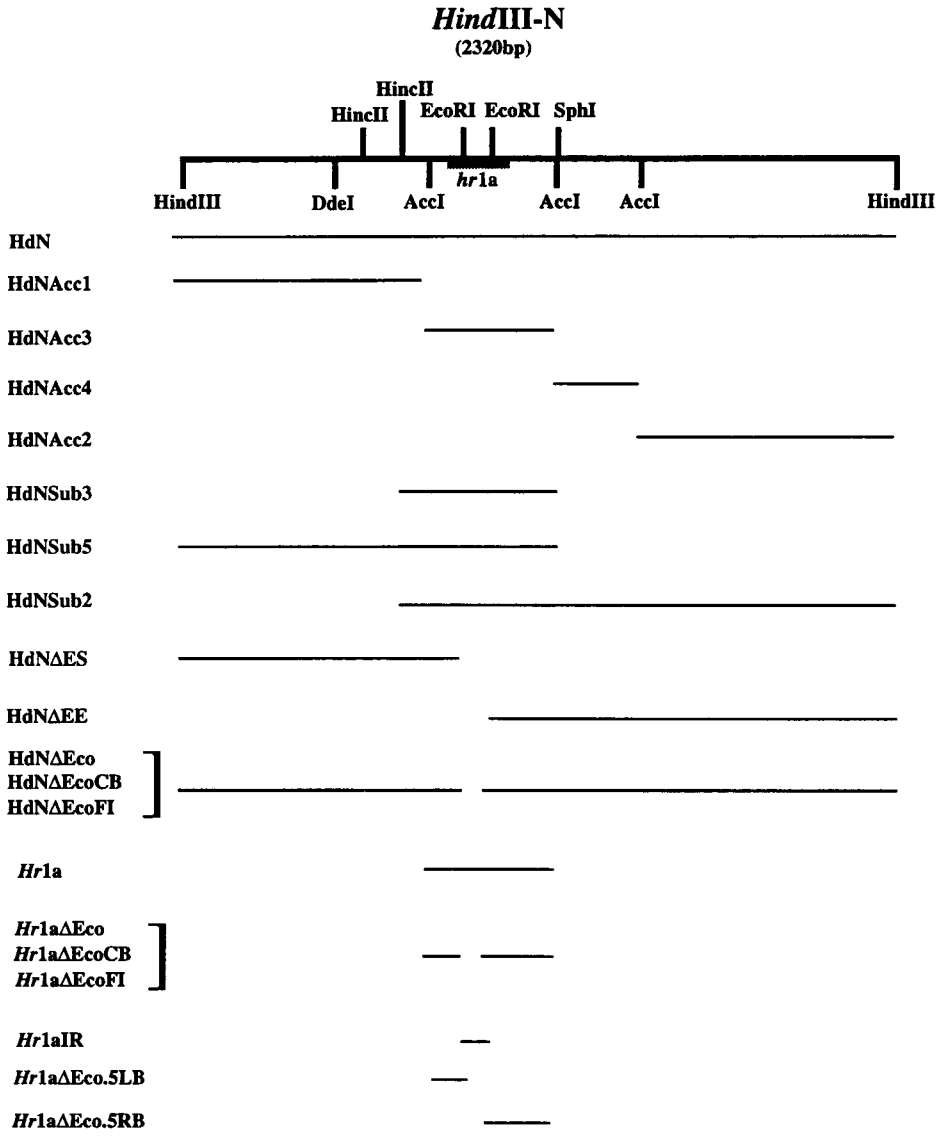


Figure 6.1. Schematic representation of the *hr*1a fragments used to study the replication, transcriptional enhancer and protein binding activity. A restriction map of the *Hind*III–N fragment is shown at the top. The *hr*1a region is depicted by a hatched bar. Regions contained in each subclone are indicated by solid lines below the map.

have indicated that ICP4 repression is mediated by ICP4 binding to a specific sequence near the site of transcription initiation [211, 95]. Therefore, to investigate the possibility that the *ie–1* gene product functions in a manner similar to ICP4, the binding activity of pAcIE–1 transfected whole–cell extracts for the promoter region of several genes involved in baculovirus replication (*ie–2*, *pe38*, *dnapol* and *helicase*) was studied.

Since *hr*s function as putative origins of replication, EMSA may be useful to probe for additional protein–DNA interactions between the proteins involved in baculovirus DNA replication. Therefore, I used whole–cell extracts made from Sf9 cells transfected with different combinations of the nine baculovirus genes involved in transient replication (including *ie–1*) [148] in gel–mobility shift analysis.

The transient replication assay was used in our laboratory to identify putative origins of replication in the OpMNPV genome [236, 5]. In addition, trans–acting sequences essential for transient replication of origin–containing plasmid DNA were also identified [2, 4]. Similar to the findings with the AcMNPV baculovirus, OpMNPV required six genes for transient replication of origin–containing plasmids [3]. Additional experiments were performed to test for the interchangeability of these genes. It was found that the OpMNPV *ie–1* gene could substitute for the AcMNPV *ie–1* gene in the transient replication assay when transfected with *hr*–containing plasmid DNA into Sf9 cells. However, the AcMNPV *ie–1* gene was unable to substitute for the OpMNPV *ie–1* gene when transfected with the OpMNPV origin containing plasmid DNA into Ld652–Y cells [3]. In order to better understand the non–reciprocity of these two homologous genes, I performed gel–mobility shift analysis, to determine whether OpMNPV IE–1 bound to AcMNPV *hr* sequences and OpMNPV putative origins of replication.

6.2 Binding of Proteins from pAcIE–1 Transfected Extracts to *hr*1a is Specific

Gel retardation assays were employed for examining the interactions of proteins from pAcIE–1 transfected cells with *hr*1a. Addition of increasing amounts of whole–cell extract from pAcIE–1 transfected cells to an intact *hr*1a–containing probe resulted in the formation of three DNA–protein complexes. Complex I, a faster migrating form was detected at low concentrations of extract (Figure 6.2a, lane 2), complex II, a slower migrating form was seen at higher extract concentrations (Figure 6.2a, lane 3), and complex III, the slowest migrating complex was only observed at very high extract concentration (Figure 6.2a, lane 6), and its presence was dependent on the quality of the whole–cell extract. When the *hr*1a–containing probe was incubated with whole–cell extracts from untransfected cells or pKS– transfected cells, we occasionally detected a shifted band, however, this band was only detected in trace amounts (Figure 6.2a, lanes 7–11). To confirm that the protein complex binds specifically to the intact *hr*1a region, complexes were competed by addition of increasing amounts of unlabeled *hr*1a–containing DNA fragments to the reaction (Figure 6.2b). This resulted in competition of the shifted complexes, indicating that the DNA–protein complex formation was specific. Competition experiments with an equivalently–sized DNA fragment (a 297 bp *Xho*I–*Xho*I fragment) from the AcMNPV *p10* promoter, did not result in a decrease in the amount of complex formed (Figure 6.2b, compare lane 7 to lane 2), indicating that the protein(s) were binding to *hr*1a in a sequence–specific manner.

6.3 IE–1 is a Component of the *hr*1a DNA–Protein Complex

Sf9 cells transfected with pAcIE–1 contain protein(s) that bind to *hr* sequences, suggesting that IE–1 is a component of the complex bound to *hr*s. Gel retar-

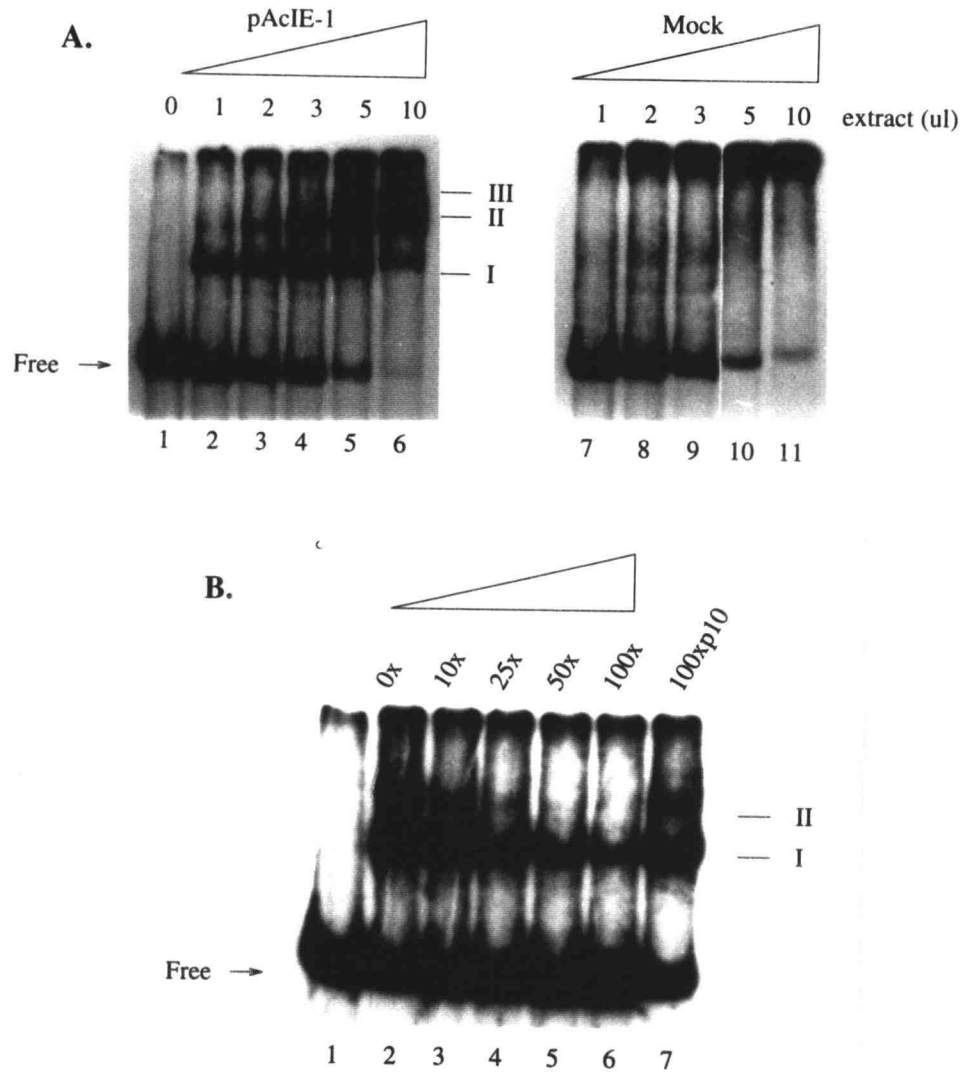


Figure 6.2. Gel Retardation Analysis of *hr*1a. A. An end-labeled 430 bp *BamH*I–*Hind*III fragment containing *hr*1a was incubated in the presence of increasing amounts of whole-cell extract prepared from either pAcIE-1 transfected (lanes 1–6) or mock-transfected (lanes 7–11) Sf9 cells as indicated. Whole-cell extract protein concentrations were typically 1 μg /μl . B. Binding reaction mixtures contained 5 μl of extract from pAcIE-1 transfected Sf9 cells. Both radiolabeled probe and competitor DNA were added simultaneously to the reaction mixtures. Competitor DNA was added in excess molar amounts as indicated at the top of each lane: lanes 3–6, unlabeled 430 bp *hr*1a fragment; lane 7, unlabeled 297 bp *Xho*1–*Xho*1 AcMNPV *p10* promoter fragment.

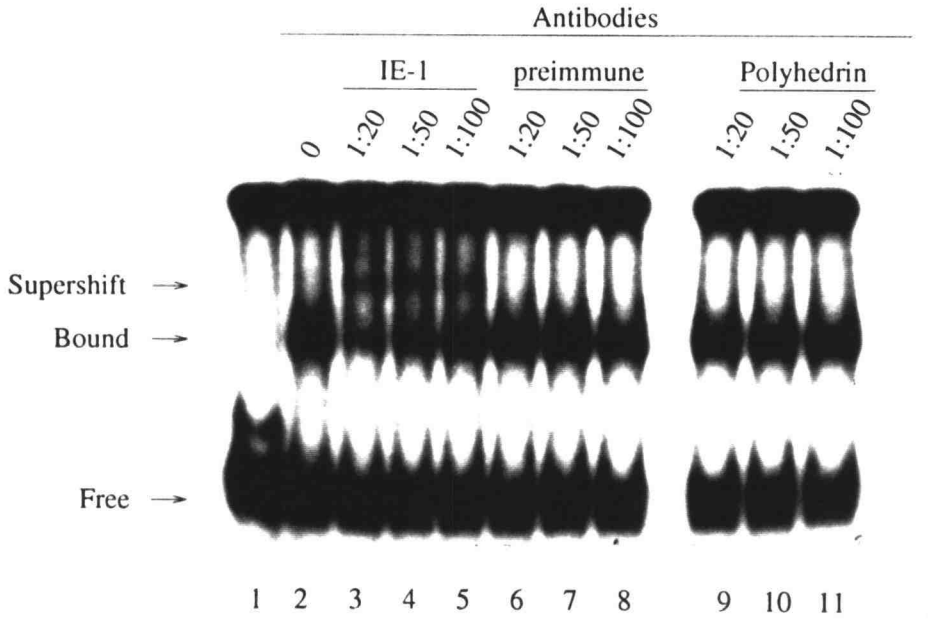


Figure 6.3. Gel Retardation Supershift Analysis of the *hr*1a DNA–protein Complex. The end–labeled 385 bp Hr1aΔEco fragment was incubated with 10 μl of whole–cell extract prepared from pAcIE–1 transfected Sf9 cells. After a pre–incubation period of 20 min, dilutions (indicated at the top of each lane) of rabbit polyclonal antibodies to AcMNPV IE–1 (lanes 3–5), preimmune serum (lanes 6–8) or rabbit polyclonal antibodies to OpMNPV polyhedrin (lanes 9–11) were added to the reaction mixture for a further 10 min before complexes were resolved by gel electrophoresis.

dation assays performed in the presence of polyclonal antibodies to IE–1 [223] were used to determine that IE–1 is a component of the DNA–protein complex. When the polyclonal IE–1 antibody was present in the reaction mixture we observed the appearance of a slower migrating complex or supershift (Figure 6.3, lanes 3–5), indicating that IE–1 is a component of the DNA–protein complex. Supershifts were not detected when preimmune serum (Figure 6.3, lanes 6–8) or polyclonal antibodies to the OpMNPV polyhedrin protein [263] (Figure 6.3, lanes 9–11) were added to the reaction mixture.

6.4 Host–encoded Protein Binding to *hr1*a Auxiliary Sequence

*Hr*1a is required for the replication of *Hind*III–N containing plasmids in infected Sf9 cells [169]. The efficiency of replication of pHdNAcc3 (Figure 6.1) was reduced compared to that of pHdN suggesting that sequences in the *Hind*III–N fragment lying outside of those contained in pHdNAcc3 contribute to the replication of *hr*1a–containing plasmids [169]. Subclones of pHdN were used to determine the location of the auxiliary sequences that augment the replication function of *hr*1a. pHdNSub5 (Figure 6.1) replicated with the same efficiency as pHdN indicating that auxiliary sequences mapping to the left of the pHdNAcc3 region can stimulate *hr*–dependent replication to near maximum levels. However, unlike the effect on replication, clones with an intact *hr*1a region but lacking sequences from the left portion of the HindIII–N fragment did not produce significantly reduced levels of GUS activity compared to clones containing this region [169] suggesting that this region enhances replication but not transcriptional activity. Deletion analysis of the pHdNSub5 region indicated that sequences within the *Hinc*II fragment were responsible for enhanced replication activity (H. Kim, personal communication).

To investigate possible DNA–protein interactions within this region, I used whole–cell extracts made from Sf9 cells transfected with each of the baculovirus genes involved in transient replication, including *ie–1*, in gel–mobility shift analysis. Gel–mobility shift analysis was performed as described in Chapter 2.25 using a 275 bp *Dde*I/*Acc*I from the pHdNΔES in attempts to identify replication proteins that interact with this auxiliary sequence. The results, shown in Figure 6.4, indicate that a host–encoded protein(s) binds to this auxiliary sequence. Mock–transfected extracts as well as extracts prepared using each of the replication genes contained a protein(s) that bound to this DNA fragment.

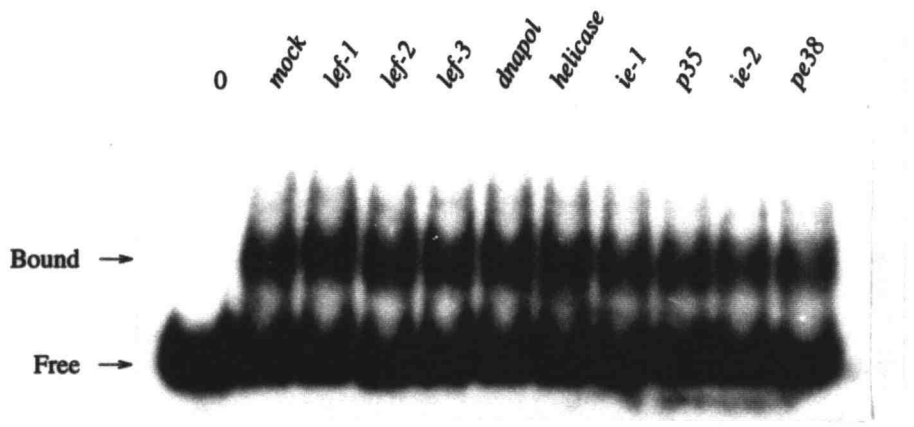


Figure 6.4. Gel Retardation Analysis of *hr*1a Auxiliary Sequence. An end–labeled 275 bp *Dde*I–*Acc*I fragment was incubated in the presence of 5.0 μl whole–cell extract (15 μg protein) prepared from Sf9 cells transfected with each of the replication genes. The replication gene used to transfect Sf9 cells is indicated above each well.

6.5 Binding of Proteins from pAcIE–1 Transfected Extracts to *hr*1a Mutants

The role of the 30–bp imperfect palindrome sequences and intervening sequences between the palindromes in the replication and transactivation ability of *hr*1a–containing plasmids was investigated by testing plasmids that had alterations within the imperfect palindrome [169]. A single palindrome was created by digesting with *Eco*RI and ligating the left half of the left palindrome with the right half of the right palindrome of *hr*1a, resulting in a construct containing a single palindrome that was identical to the *hr*1a right palindrome but lacks the 58 bp sequence between the two palindromes. The single palindrome construct, HdNΔEco was sufficient to allow infection–dependent plasmid replication. However, additional mutants within the single palindrome construct containing either a four base deletion (HdNΔCB) or a three base duplication (HdNΔFI) at the center of the single palindrome and derivatives containing only the left (HdNΔES)

or right half (HdNΔEE) of the single palindrome replicated at barely detectable levels. GUS reporter gene constructs containing these alterations were assayed to determine if the sequences required for replication also influenced enhancer function. These experiments showed that only mutants containing a single complete palindrome were able to transactivate; the central core mutants and the half palindrome constructs were unable to transactivate [169]. Taken together, these data suggest that a single palindrome is sufficient to allow *hr*1a–mediated replication and transactivation and that disruptions of the central *Eco*RI–core render the palindrome dysfunctional for replication and transactivation.

6.5.1 Binding of Proteins from pAcIE–1 Transfected Extracts to *hr*1a Central Core Mutants

I assayed *hr*1a mutants, similar to those tested for replication and enhancer functions, for the ability to form specific DNA–protein complexes. The sequences of these mutants are shown in Figure 6.5 and their construction is described in chapter 2.7.

Only one retarded complex was observed with a DNA fragment containing a single *hr*1a palindrome (construct pHR1aΔEco contains the left half of the first palindrome fused with the right half of the second palindrome) (Figure 6.6, lanes 2–6). Competition experiments identical to those described for the entire *hr*1a region (Figure 6.7, lanes 2–7) showed that the complex formed is sequence specific. The central *Eco*R1 core mutants showed a marked reduction in replication and transactivation efficiencies. However, gel retardation assays indicated that a single retarded DNA–protein complex was formed with the Hr1aΔEcoCB (Figure 6.6, lanes 7–12) and Hr1aΔEcoFI (Figure 6.6, lanes 13–18) DNA fragments containing the *Eco*RI core mutations. Competition experiments identical to those described above showed that the binding to the disrupted core mutants was specific (Figure 6.7, lanes 9–14 and 16–21). This data correlates well with

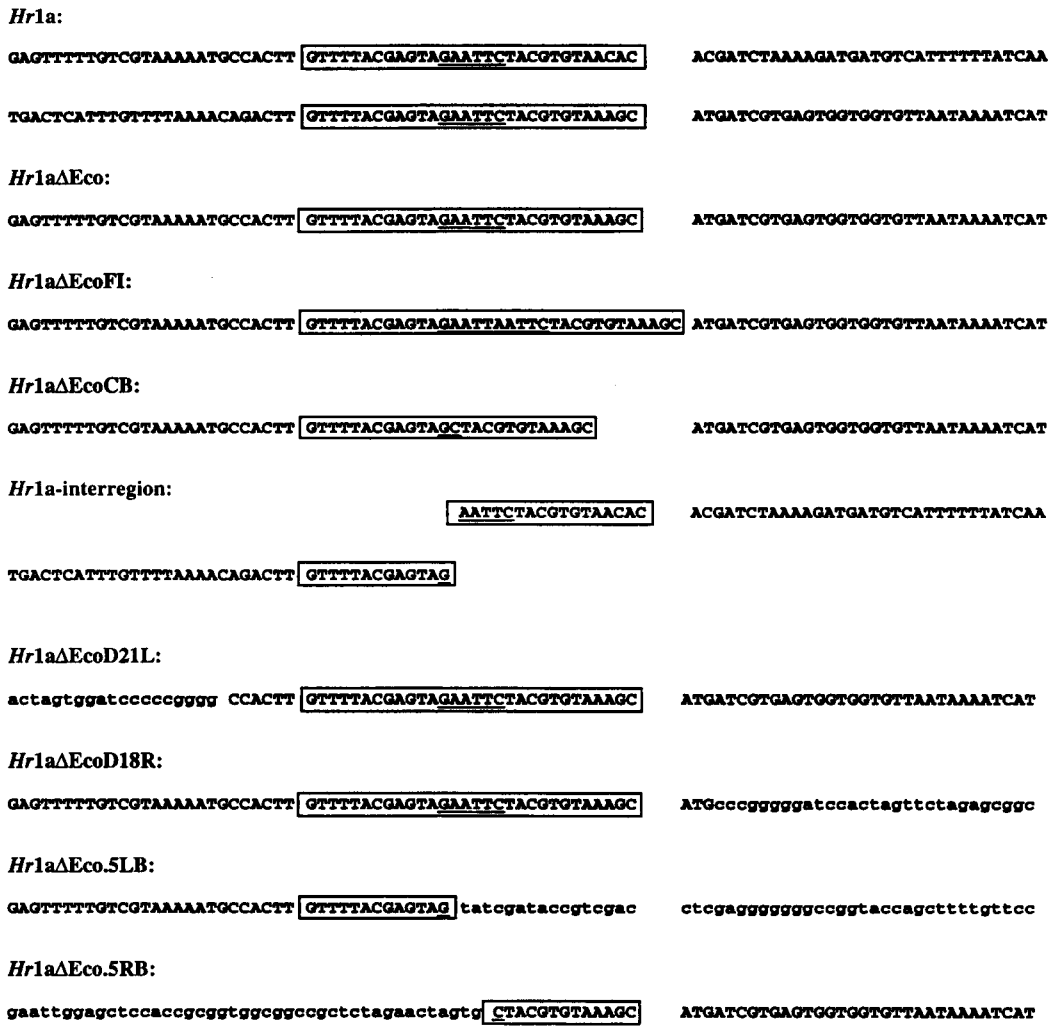


Figure 6.5. Sequences of the *hr*1a Mutants and Derivatives. Sequences from the 30 bp imperfect palindromes are enclosed in boxes. The *Eco*RI sites (or remaining portions of the *Eco*RI sites) are underlined. The nucleotide sequence derived from *hr*1a is shown in capitol letters while the sequences derived from the pKS– vector are indicated by lower case letters.

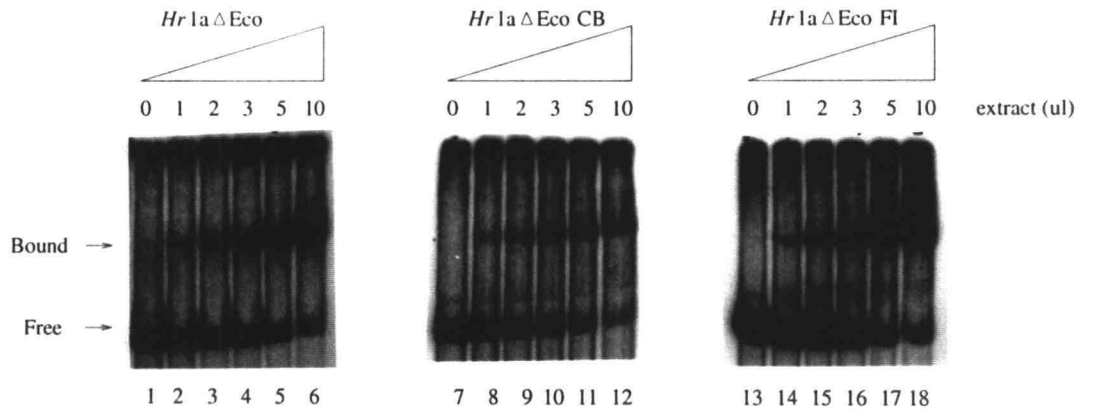


Figure 6.6. Gel Retardation Analysis of *hr*1a Central Core Mutants. End-labeled Hr1aΔEco (lanes 1–6), Hr1aΔEcoCB (lanes 7–12) and Hr1aΔEcoFI (lanes 13–18) fragments were incubated in the presence of 0 μl (lane 1, 7 and 13), 1 μl (lane 2, 8 and 14), 2 μl (lanes 3, 9 and 15), 3 μl (lanes 4, 10 and 16), 5 μl (lanes 5, 11 and 17) and 10 μl (lanes 6, 12 and 18) of whole-cell extract prepared from pAcIE-1 transfected Sf9 cells.

previous data [97, 259] indicating that protein(s) bind to regions flanking the *Eco*R1 core sequence. However, these results indicate that IE-1 binding by itself is not sufficient for activation of replication and enhancer functions. These central core mutations may alter sequences required for other factors to bind to *hr*1a, or disrupt essential DNA conformations such as hairpin structures and thus prevent both replication and trans-activation.[4]

---

[4] Data from sections 6.2, 6.3 and 6.5.1 were published in D.J. Leisy, C. Rasmussen, H. Kim and G.F. Rohrmann. (1994). The *Autographa californica* Nuclear Polyhedrosis Virus Homologous Region 1a: Identical sequences are essential for DNA Replication Activity and Transcriptional Enhancer Function. Virology. 208:742–752.

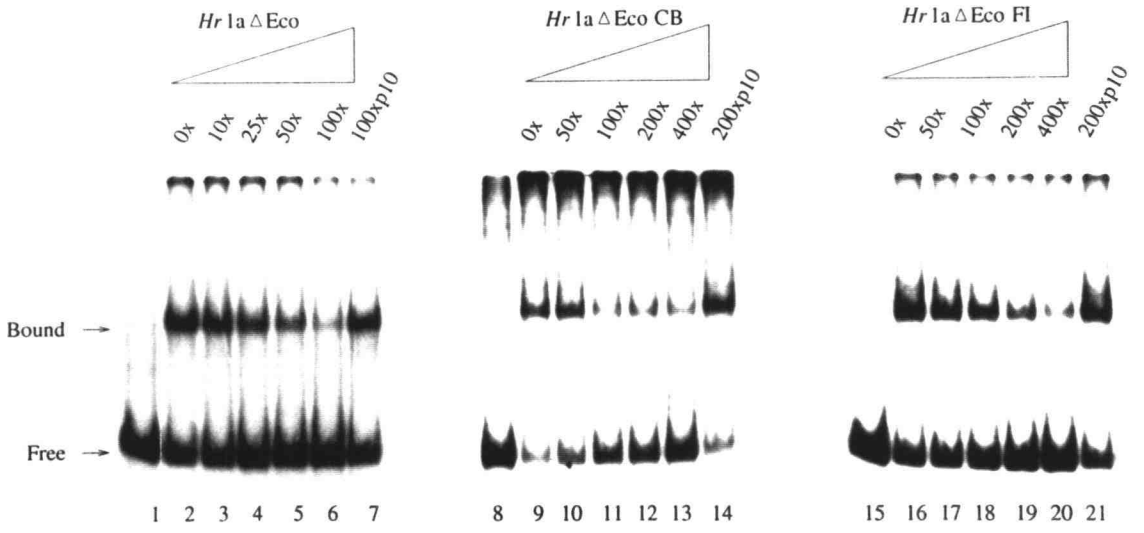


Figure 6.7. Gel Retardation Competition Analysis of *hr*1a Central Core Mutants. End-labeled Hr1aΔEco (lanes 1–7), Hr1aΔEcoCB (lanes 8–14) and Hr1aΔEcoFI (lanes 15–21) fragments were incubated in the presence of 5 μl of whole-cell extract prepared from pAcIE-1 transfected Sf9 cells. Both radiolabeled probe and competitor DNA were added simultaneously to the reaction mixtures. Competitor DNA was added in excess molar amounts as indicated at the top of each lane: lanes 3–6, unlabeled 385 bp Hr1aΔEco fragment; lanes 10–13, unlabeled 340 bp Hr1aΔEcoCB fragment; lanes 17–20, unlabeled 348 bp Hr1aΔEcoFI fragment; lanes 7, 14 and 21, unlabeled 297 bp *Xho*1–*Xho*1 AcMNPV *p10* promoter fragment.

6.5.2 Binding of Proteins from pAcIE–1 Transfected Extracts to Single Palindrome Derivatives

In order to delineate the specific sequences to which IE–1 was binding within the *hr*1a region, the Hr1aΔEco construct was subjected to deletion analysis. Two clones, pHR1aΔEco21L and pHR1aΔEco18R, that retained the central palindromic core but were significantly reduced in flanking sequence, were obtained. pHR1aΔEco21L and pHR1aΔEco18R contain a single *hr*1a palindrome (the left half of the first palindrome fused with the right half of the second palindrome) but portions of the 5' and 3' flanking sequence, respectively, have been deleted. The sequences of these derivatives are shown in Figure 6.5.

DNA fragments containing the HR1aΔEco21L and HR1aΔEco18R deletions were then used in gel mobility shift analysis to determine if IE–1 bound to the central region of the palindrome. Retarded complexes were observed when both HR1aΔEco21L and HR1aΔEco18R DNA fragments were used. Figure 6.8 shows that the addition of increasing amounts of pAcIE–1 transfected whole–cell extracts to reaction mixtures containing either pHR1aΔEco21L or pHR1aΔEco18R, resulted in the formation of a single retarded complex. Competition experiments similar to those previously described, showed that the complex formed is sequence specific (Figure 6.9). These data, together with the central core mutation data, suggest that IE–1 binds to regions immediately flanking the *Eco*R1 core sequence.

Two other derivatives of *hr*1a, that were truncated by digestion with *Eco*RI and therefore contained one half of a single palindrome (referred to as half sites) with 5' or 3' flanking regions were assayed for the ability to specifically bind IE–1. Figure 6.10 shows that that a single retarded DNA–protein complex was formed when labeled DNA fragments containing either hr1aΔEco.5RB and hr1aΔEco.5LB sequences were examined. To confirm that these complexes were specific, the complexes were competed by addition of unlabeled *hr*1a–containing

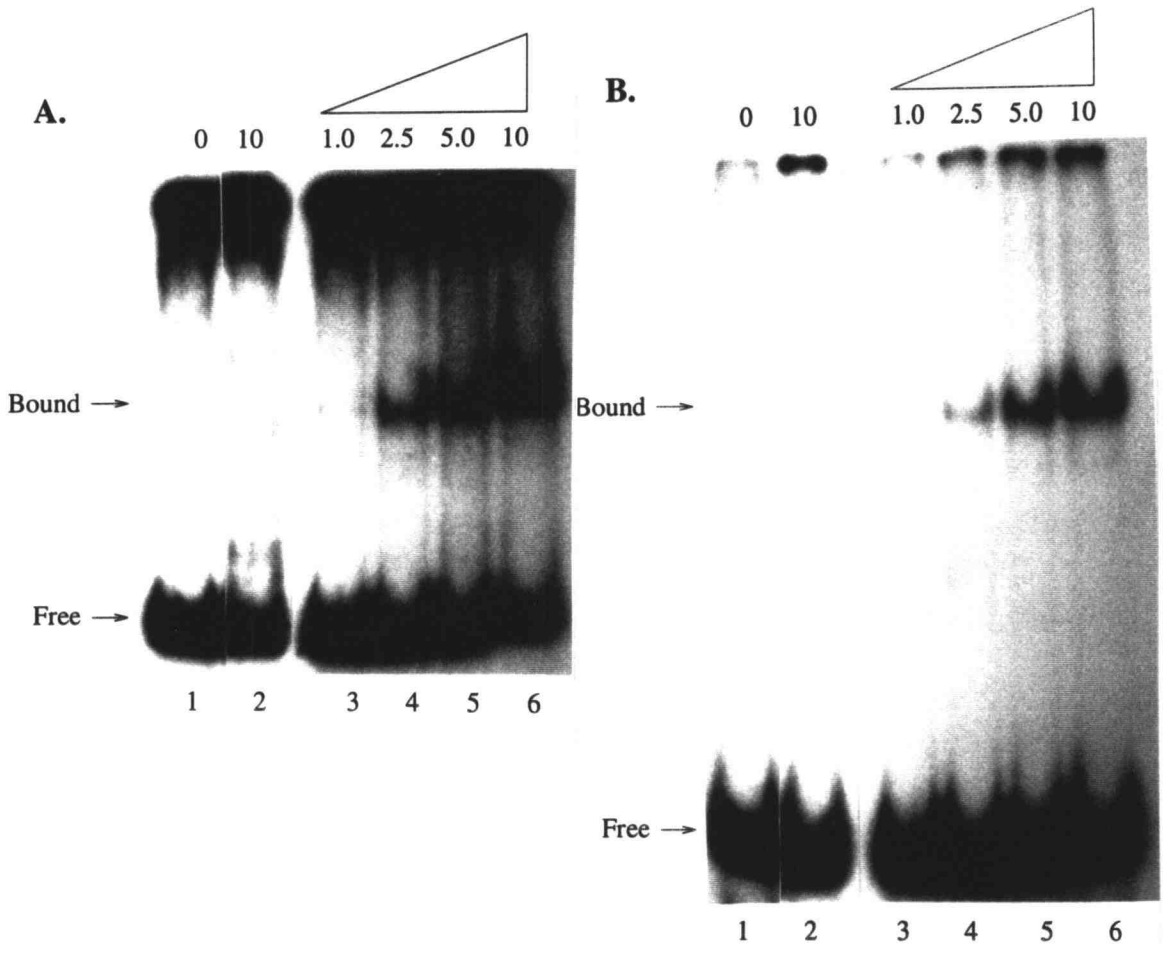


Figure 6.8. Gel Retardation Analysis of *hr*1a Single Palindrome Deletions. A. End–labeled Hr1aΔEco21L fragments were incubated in the presence of 0 μl (lane 1), 1 μl (lane 3), 2.5 μl (lane 4), 5 μl (lane 5), and 10 μl (lane 6) of whole–cell extract prepared from pAcIE–1 transfected Sf9 cells. 10 μl of whole–cell extract prepared from pKS– transfected Sf9 cells was used in lane 2. B. End–labeled Hr1aΔEco18R fragments were incubated in the presence of 0 μl (lane 1), 1 μl (lane 3), 2.5 μl (lane 4), 5 μl (lane 5), and 10 μl (lane 6) of whole–cell extract prepared from pAcIE–1 transfected Sf9 cells. 10 μl of whole–cell extract prepared from pKS– transfected Sf9 cells was used in lane 2.

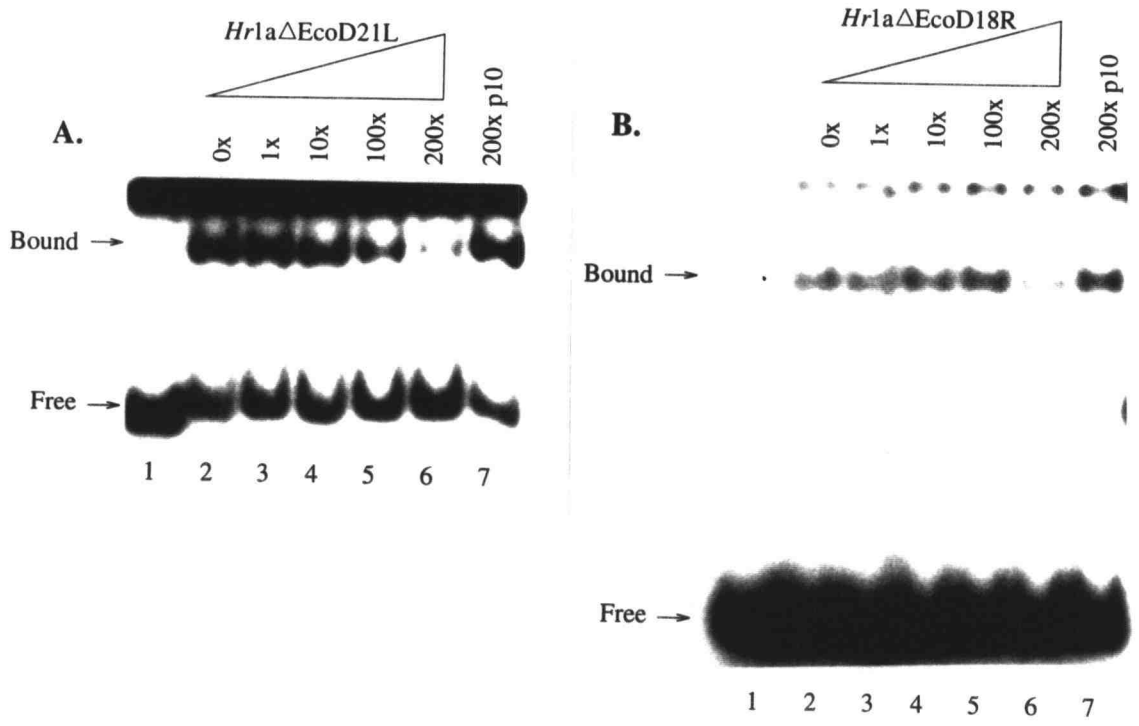


Figure 6.9. Gel Retardation Competition Analysis of *hr*1a Single Palindrome Deletions. A. End–labeled Hr1aΔEco21L fragments were incubated in the presence of 0 μl (lane 1) and 10μl (lane 2–7) of whole–cell extract prepared from pAcIE–1 transfected Sf9 cells. Both radiolabeled probe and competitor DNA were added simultaneously to the reaction mixtures. Competitor DNA was added in excess molar amounts as indicated at the top of each lane. An unlabeled 297 bp *Xho*1–*Xho*1 AcMNPV *p10* promoter was used a competitor in lane 7. B. End–labeled Hr1aΔEco18R fragments were incubated in the presence of 0 μl (lane 1) and 10μl (lane 2–7) of whole–cell extract prepared from pAcIE–1 transfected Sf9 cells. Competition was performed as described in A.

DNA fragments (Figure 6.10, lanes 3 and 4) but not by the addition of an equivalently sized DNA fragment containing the AcMNPV *p10* promoter region (Figure 6.10 lanes 5).

Lastly, the region between the two central palindromes, referred to as the interregion, was assayed for the ability to bind IE–1. Figure 6.11 shows that a host–encoded factor(s) binds to the interregion forming several retarded complexes. Complex I and II are detected at low concentrations of extract while complex III, a slower migrating form was seen at higher extract concentrations. These complexes were observed in pKS– and pAcIE–1 transfected whole–cell extracts. Competition experiments showed that the complexes observed, when either pKS– or pAcIE–1 transfected whole–cell extracts were used, are specific. One set of extracts prepared from pAcIE–1 transfected cells showed the formation of additional complexes that appeared to be specific for pAcIE–1 transfected extracts (data not shown). This result would suggest that both IE–1 and a host factor(s) are binding to this region. However, this pAcIE–1 transfected whole–cell extract contained very high IE–1 activity and these DNA binding results were not reproduced with other pAcIE–1 transfected whole–cell extracts.

The observation that host–factors bind to the IR is consistent with other data showing that *hr*–containing plasmids showed increased levels of expression compared to non–*hr* containing plasmids when transfected into insect cells without IE–1 [218, 185, 169] (see section 6.6). A host factor(s) may bind to DNA sequences within *hr*s and activate transcription.

6.6 Enhancer Activity of *hr*1a Derivatives in pAcIE–1 Transfected Sf9 Cells

Since *hr*s function as enhancers of early gene expression [28, 185, 218], I was interested in determining if several of the *hr*1a derivatives also functioned as enhancers. DNA fragments containing *Hr*1a, *Hr*ΔEco and *Hr*1aIR were subcloned

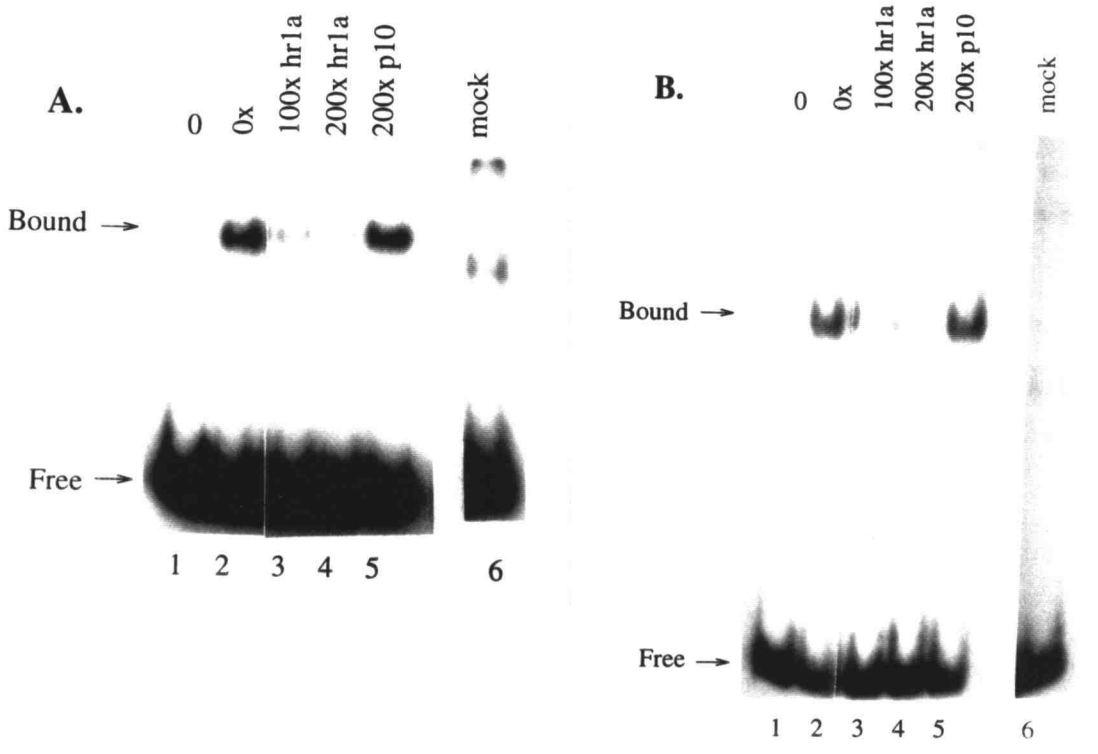


Figure 6.10. Gel Retardation Analysis of *hr*1a Half Sites. A. End-labeled Hr1aΔEco.5LB fragments were incubated in the presence of 0 μl (lane 1) and 10 μl of whole-cell extract prepared from pAcIE-1 transfected (lane 2–5) or mock-transfected (lane 6) Sf9 cells. Both radiolabeled probe and competitor DNA were added simultaneously to the reaction mixtures. Competitor DNA, a DNA fragment containing the entire *hr*1a region, was added in excess molar amounts as indicated at the top of each lane. An unlabeled 297 bp *Xho*1–*Xho*1 AcMNPV *p10* promoter fragment fragment was used as the competitor in lanes 5. B. Experiments were performed as described in A except end-labeled Hr1aΔEco.5RB DNA fragments were used.

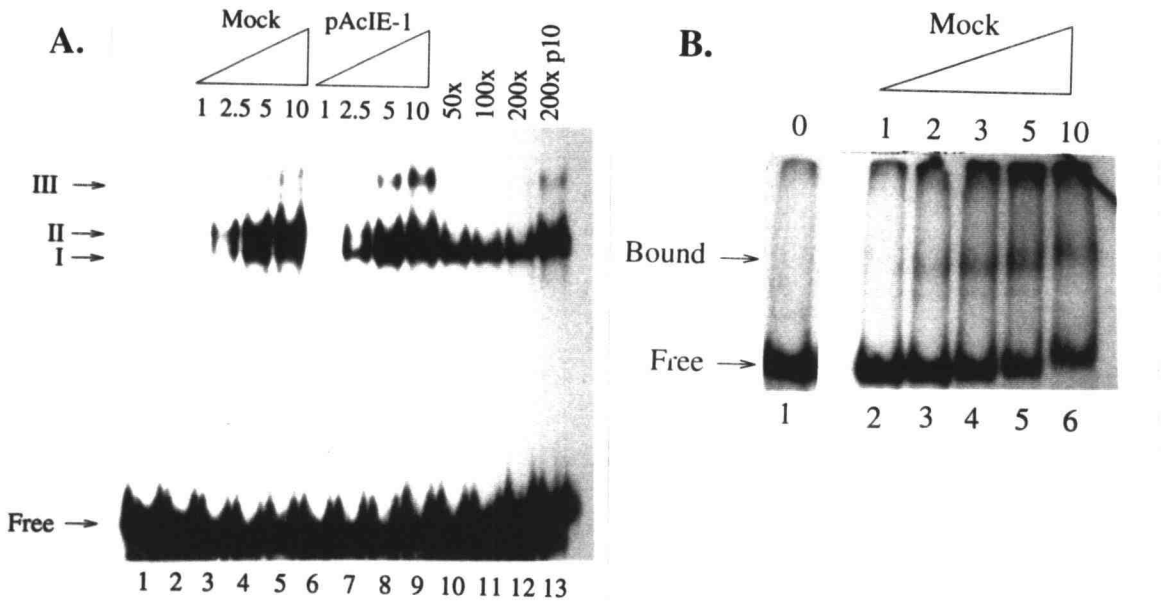


Figure 6.11. Gel Retardation Analysis of *hr*1a and the *hr*1a Interregion. A. End–labeled Hr1aIR was incubated in the presence of 0μl (lane 1), 2.5μl (lane 2 and 6), 5μl (lane 3 and 7) and 10μl (lane 5 and 9–13) of whole–cell extract prepared from pAcIE–1 transfected (lanes 6–13) and mock–transfected (lanes 2–6) Sf9 cells. Competitor DNA, a DNA fragment containing the *hr*1a interregion, was added in excess molar amounts as indicated at the top of lanes 10–12. Both radiolabeled probe and competitor DNA were added simultaneously to the reaction mixtures. An unlabeled 297 bp *Xho*1–*Xho*1 AcMNPV *p10* promoter fragment fragment was used as the competitor in lane 13. B. The end–labeled *hr*1a–containing DNA fragment was incubated with increasing amounts in μl of whole–cell extract prepared from mock–transfected Sf9 cells.

into the GUS reporter plasmid p39KGUSpI18. The enhancer activities of these clones were determined by measuring the GUS activities in extracts of Sf9 cells that had been co–transfected with these GUS reporter constructs and increasing concentrations of pAcIE–1.

The concentration of pAcIE–1 co–transfected with the GUS reporter constructs was varied between 0 and 5 μg per 1.25 $\times 10^6$ cells to determine which pAcIE–1 concentration yielded the highest levels of activation when compared to the same concentration of GUS reporter constructs. As previously shown, a concentration of 0.01 μg of pAcIE–1 per 1.25 $\times 10^6$ cells exhibited maximum enhancer effect [169]. As shown in Figure 6.12, the magnitude of the enhancer effect was diminished at both higher and lower concentrations of pAcIE–1. In the absence of pAcIE–1, an approximate 10–fold increase in GUS activity was observed with pHr1aGUS when compared to p39KGUSpI18 whereas the GUS levels observed using pHrΔEcoGUS and pHr1aIRGUS were not discernibly above p39KGUSpI18 levels.

Comparison of the transactivation ability of the three reporter gene constructs were performed using 0.01 μg of pAcIE–1 per 1.25 $\times 10^6$ cells. As shown in the Figure 6.13, the complete *hr*1a sequence enhances transcription 10–fold better than the construct containing just a single palindrome. Although, the GUS reporter construct containing the interregion (Hr1aIRGUS) does enhance transcription at least 10–fold when compared to the non–*hr*–containing plasmid, p39KGUSpI18, it is 200–fold lower than the complete *hr*1a sequence. These data suggest that synergistic interactions may be occurring between palindromes and/or the regions between the palindromes. The DNA sequences between the *hr* palindromes exhibit some sequence conservation [96, 98](See Appendix 3). Therefore, maximal enhancer activity appears to require at least two imperfect palindromes and may include the region between the palindromes that I have referred to as the interregion.

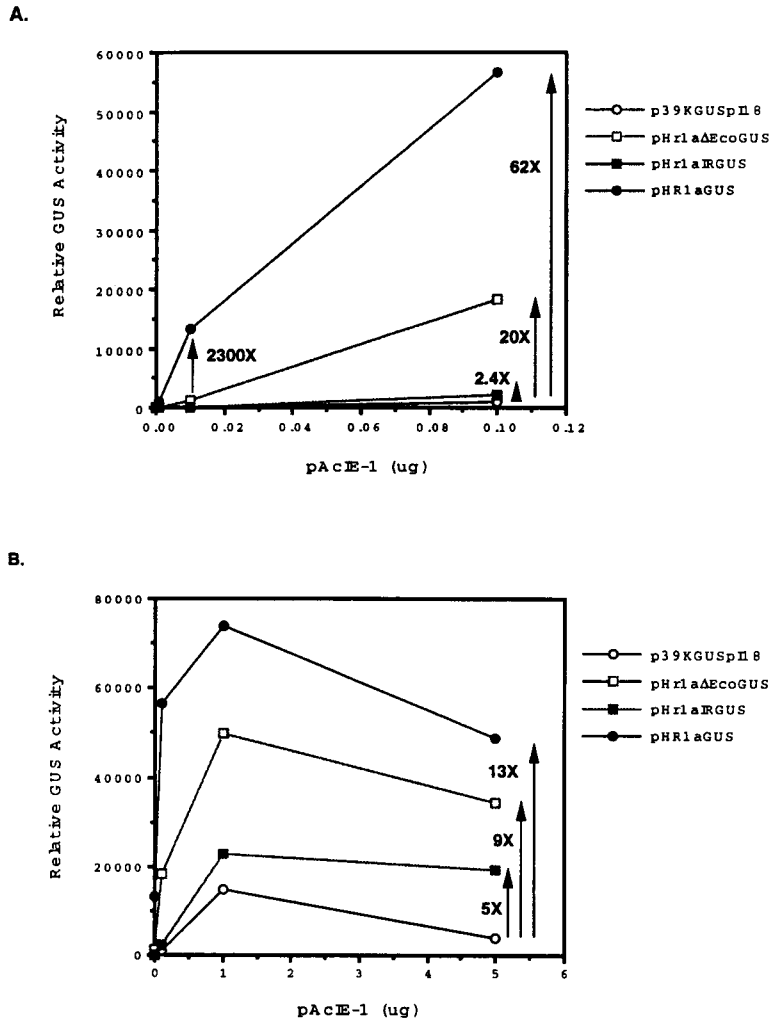


Figure 6.12. Enhancer Activity as a Function of pAcIE–1 Concentration. 0.5 μg of p39KGUSpI18, pHr1aGUS, pHR1aΔ and pHr1aIRGUS were co–transfected into 1.25 $\times 10^6$ Sf9 cells in the presence of varying concentrations of pAcIE–1. GUS activity was measured at 24 hr post–transfection. A. The influence of pAcIE–1 concentration on GUS activity are plotted relative to the activity obtained in p39KGUSpI18 transfected cells without the addition of pAcIE–1. The ratios of GUS activities of pHr1aGUS, pHR1aΔEcoGUS and pHR1aIRGUS relative to p39KGUSpI18 at 0.001 and 0.1 μg are shown. B. Same as A, only higher concentrations of pAcIE–1 were used. The ratios of GUS activities of pHr1aGUS, pHR1aΔEcoGUS and pHR1aIRGUS relative to p39KGUSpI18 at 5 μg is shown.

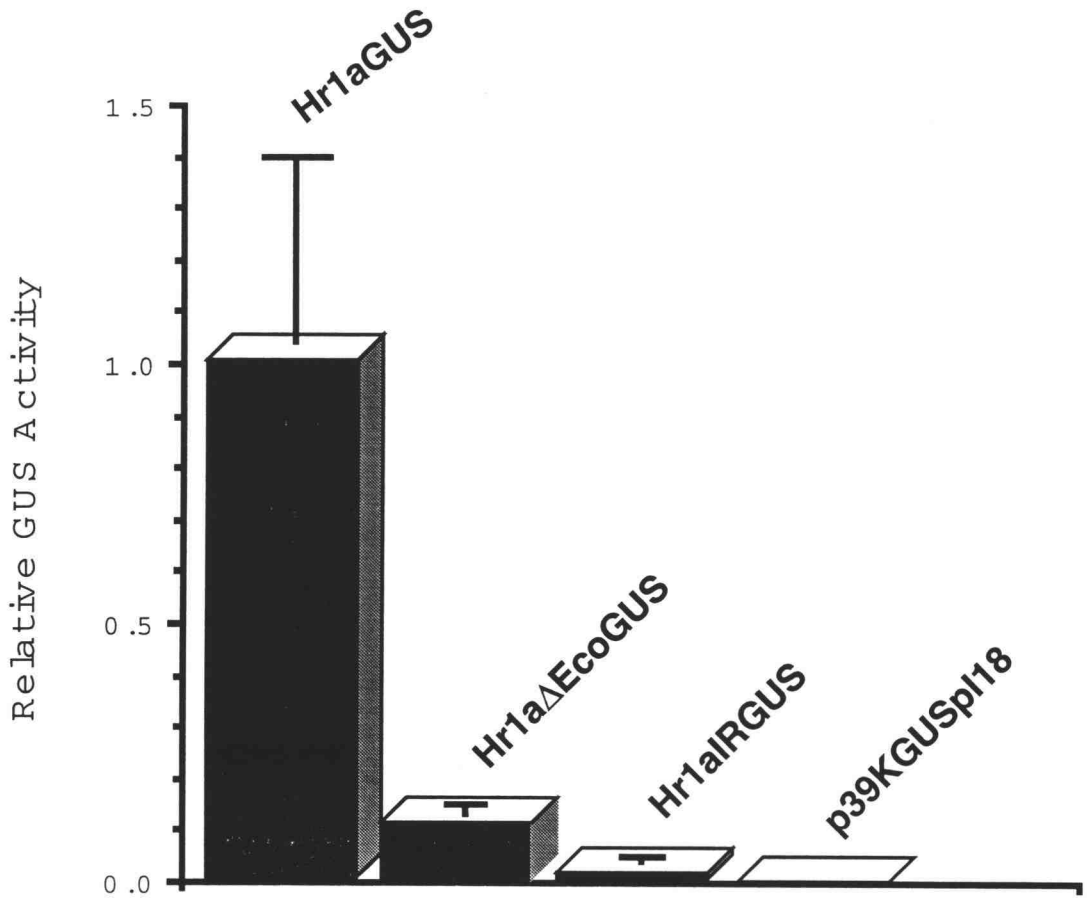


Figure 6.13. Enhancer Activity of *hr*1a derivatives. 1.25 $\times 10^6$ Sf9 cells were co–transfected with 0.5 μg of the indicated plasmids and 0.01 μg of pAcIE–1. The ratios of GUS activities relative to pHr1aGUS are shown. Each column represents the average of at least three independent transfections with the standard deviation indicated.

6.7 Binding of Proteins from pAcIE–1 Transfected Extracts to Replication Gene Promoters

As previously mentioned, the *ie–1* gene product has been shown to modulate the level of transcription from baculovirus promoters in the presence [102, 98, 218, 28, 185, 30] or absence [101, 103, 218] of *hr* elements. Furthermore, the *ie–1* gene product inhibits the expression of two other genes, *ie–2* [30] and a larger spliced form of *ie–1*, termed *ie–0* [156] while allowing continued expression of its own gene product. To investigate the possibility that IE–1 may bind to the promoter regions of several baculovirus genes, pAcIE–1 transfected whole–cell extracts were assayed for DNA binding activity specific for the promoter regions of several genes involved in baculovirus replication, including *ie–2*, *pe38*, *dnapol* and *helicase*.

As shown in Figure 6.14, pAcIE–1 transfected extracts contain a DNA binding activity specific for the *ie–2* gene promoter. Addition of pAcIE–1 transfected whole–cell extracts resulted in the formation of two DNA–protein complexes. A faster migrating complex was detected in both mock and pAcIE–1 transfected whole–cell extracts (Figure 6.14, compare lane 2 and 3) suggesting that a Sf9 encoded protein(s) may be interacting with the *ie–2* promoter region. However, it is interesting to note that the complex migration is not identical using mock and pAcIE–1 whole cell extracts, the addition of equivalent amounts of protein from pAcIE–1 transfected whole cell extracts resulted in increased binding (Figure 6.14, compare lane 2 and 3) and the complex was competed by addition of *hr*1a containing probe (Figure 6.14, lane 10). Complex II, a slower migrating form was specific for pAcIE–1 transfected extracts. Complex II was competed with increasing amounts of both the *ie–2* promoter and the *hr*1a fragment (Figure 6.14, lanes 6 and 10). Competition experiments with an equivalently sized DNA fragment (a 297 bp *Xho*I–*Xho*I fragment) from the AcMNPV *p10* promoter, did not result in a decrease in the amount of complex formed (Figure 6.14, com-

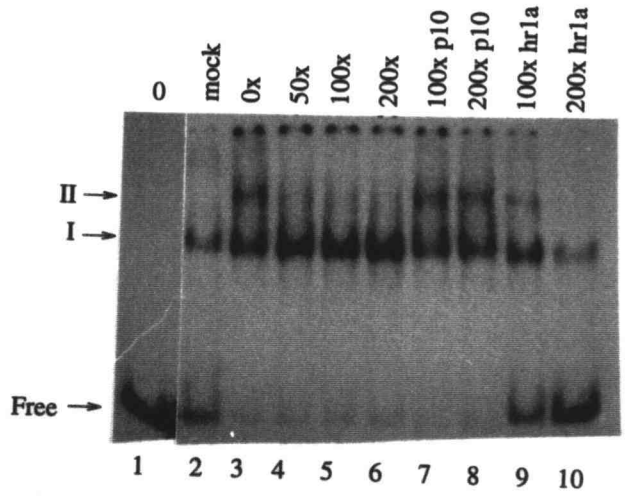


Figure 6.14. Gel Retardation Analysis of the *ie-2* Promoter Region. A 336 bp end-labeled *Nco*I-*Xba*I fragment containing the *ie-2* promoter region was incubated in the presence of 0μl (lane 1), 10μl (lane 2–10) of whole-cell extract prepared from pAcIE-1 (lanes 3–10) or mock (lane 2) transfected Sf9 cells. Extract protein concentrations were 2.5 μg /μl . Both radiolabeled probe and competitor DNA were added simultaneously to the reaction mixtures. Competitor DNA included the 336 bp *ie-2* promoter fragment (lanes 4–6), the 430 bp fragment containing the *hr*1a region (lane 9 and 10), and the 297 bp *Xho*1-*Xho*1 AcMNPV *p10* promoter fragment (lane 7 and 8) that were added in excess molar amounts as indicated at the top of each lane.

pare lane 8 to lane 3), indicating that the IE-1 is binding to the *ie-2* promoter in a sequence-specific manner.

When the *pe38* promoter region was used in gel mobility shift analysis, pAcIE-1 transfected extracts were shown to contain a DNA binding activity specific for the *pe38* gene promoter. Addition of pAcIE-1 transfected whole-cell extracts resulted in the formation of two DNA-protein complexes. Complex II, the slower migrating complex, was only formed when very high concentrations pAcIE-1 transfected extracts were used (Figure 6.15, lane 15). Complex I, a faster migrating complex, was formed by addition of pAcIE-1 transfected extracts. This complex was not detected when equal amounts of mock-transfected

extract was used. Competition by addition of increasing amounts of both the *pe38* promoter and the *hr*1a fragment (Figure 6.15) resulted in the disappearance of the shifted band. However, competition experiments with an equivalently sized DNA fragment (a 297 bp *Xho*I–*Xho*I fragment) from the AcMNPV *p10* promoter, did not result in a decrease in the amount of complex I or complex II (Figure 6.15, compare lane 8 to lane 3), indicating that the IE–1 is binding to the *pe38* promoter in a sequence–specific manner.

Electrophoretic mobility shift analysis using the *dnapol* and *p143* promoters failed to detect any sequence–specific binding in both mock and pAcIE–1 transfected extracts (data not shown) suggesting that the *ie–2* and *pe38* promoter regions contain sequences for IE–1 binding.

6.8 IE–1 is a Component of the *pe38* DNA–Protein Complex

Sf9 cells transfected with pAcIE–1 contain protein(s) that bind to replication promoters, suggesting that IE–1 is a component of the complex. Gel retardation assays performed in the presence of polyclonal antibodies to IE–1 [223] resulted in the appearance of a slower migrating complex or supershift (Figure 6.16, lanes 3–5), indicating that IE–1 is a component of the DNA–protein complex at the *pe38* promoter. Supershifts were not detected when pre–immune serum (Figure 6.16, lanes 6–8), polyclonal antibodies to the OpMNPV polyhedrin protein [263] (Figure 6.16, lanes 9–11) or mouse monoclonal antibodies to *Drosophila* TBP (data not shown) were added to the reaction mixture.

6.9 Western Analysis of Whole–Cell Extracts

Whole–cell extracts prepared from Sf9 cells transfected with pKS– or the AcMNPV *ie–1* gene were examined for the presence of IE–1 using western blot analysis. Polyclonal antibodies to AcMNPV IE–1 were used [223] to detect IE–1

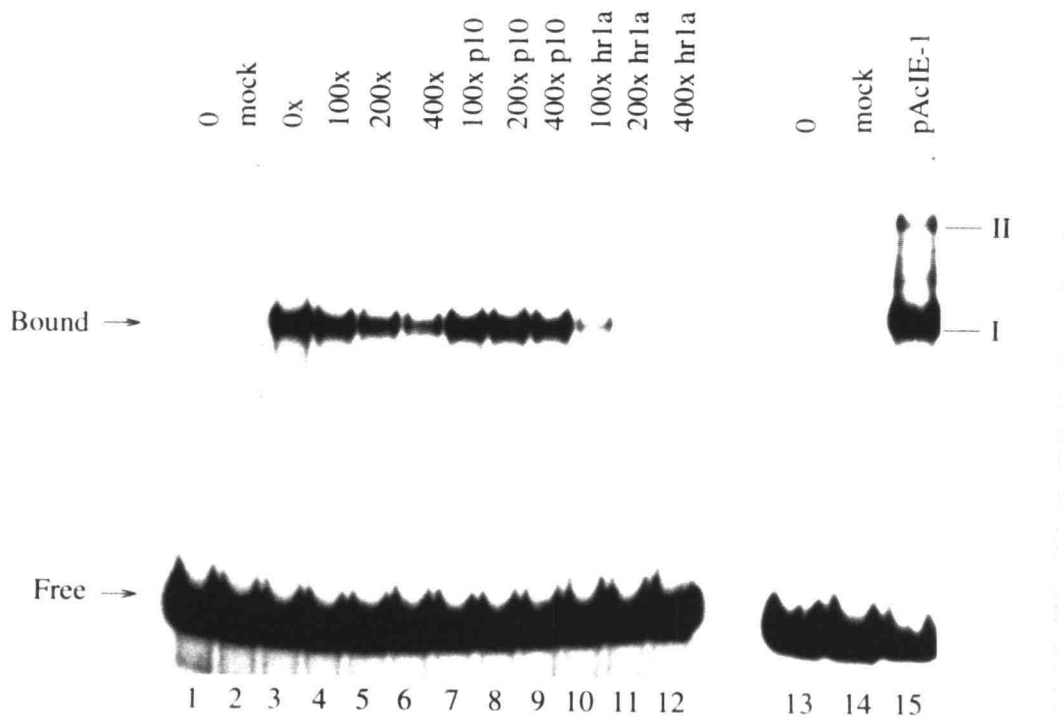


Figure 6.15. Gel Retardation Analysis of the *pe38* Promoter Region. A 120 bp end-labeled *Nco*I–*Kpn*I fragment containing the *pe38* promoter region was incubated in the presence of 0μg (lane 1 and 13), 2.5μg (lane 2–10) and 12.5μg (lane 14 and 15) of whole-cell extract prepared from pAcIE–1 transfected (lanes 3–12 and 15) and mock-transfected (lane 2 and 14) Sf9 cells. Both radiolabeled probe and competitor DNA were added simultaneously to the reaction mixtures. Competitor DNA, DNA fragment containing the *pe38* promoter (lanes 4–6) or the *hr*1a region (lanes 10–12), was added in excess molar amounts as indicated at the top of each lane. An unlabeled 297 bp *Xho*1–*Xho*1 AcMNPV *p10* promoter fragment fragment was used as the competitor in lanes 7–9.

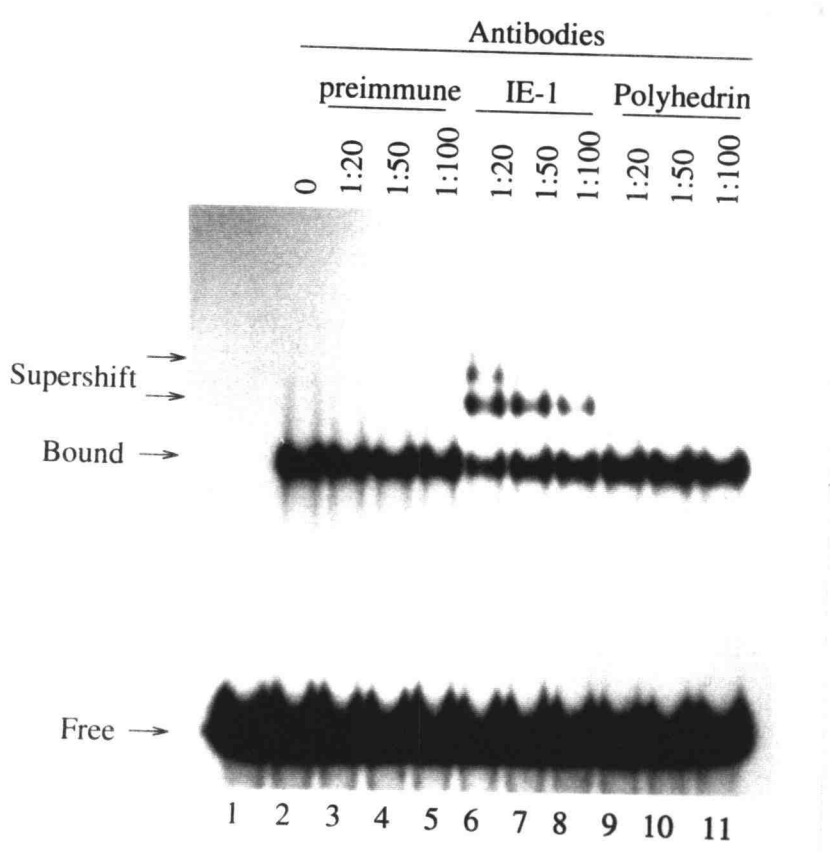


Figure 6.16. Gel Retardation Supershift Analysis of the *pe38* DNA–protein Complex. The end–labeled 120 bp *pe38* promoter containing fragment was incubated with 2.5 μg of whole–cell extract prepared from pAcIE–1 transfected Sf9 cells. After a pre–incubation period of 20 min, dilutions (indicated at the top of each lane) of rabbit polyclonal antibodies to AcMNPV IE–1 (lanes 6–8), preimmune serum (lanes 3–5) or rabbit polyclonal antibodies to OpMNPV polyhedrin (lanes 9–11) were added to the reaction mixture for a further 10 min before complexes were resolved by gel electrophoresis.

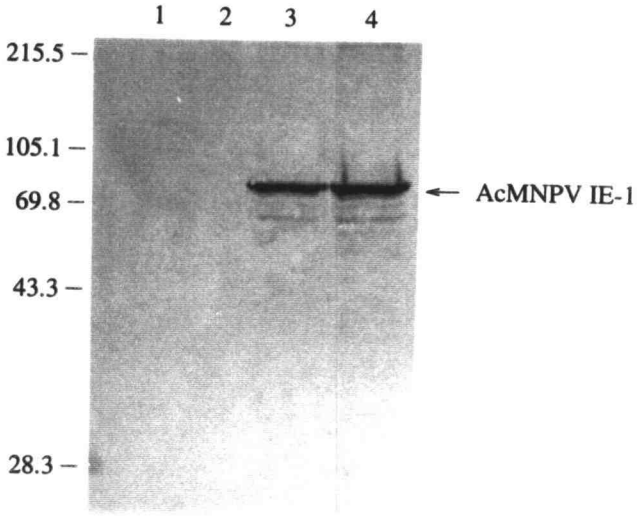


Figure 6.17. Detection of AcMNPV IE–1 in Whole–Cell Extracts. Lanes 1 and 2 contain 20 and 40 μg , respectively, of whole–cell prepared from pKS– transfected Sf9 cells and lanes 3 and 4 contain 20 and 40 μg of whole–cell extract, respectively, prepared from *ie–1* transfected Sf9 cells. The numbers on the left indicate the marker sizes in kDa determined using the BioRad prestained low range protein ladder.

protein production. As Figure 6.17 shows, IE–1 is present in whole–cell extracts prepared from *ie–1*–transfected but not pKS– transfected cells. The AcMNPV IE–1 polyclonal antibodies fail to cross–react with OpMNPV IE–1 (data not shown).

6.10 Analysis of DNA Sequences that Exhibit IE–1 Binding Activity

6.10.1 Alignments of AcMNPV *hr* Regions

Alignments of all the *hr* regions in the complete genome of AcMNPV [7] were generated using the clustal V alignment program [115]. The computer–generated alignments were modified and the information content of the aligned *hr* sequences was determined as described in Chapter 5.6. The *hr* alignment is shown in Ap-

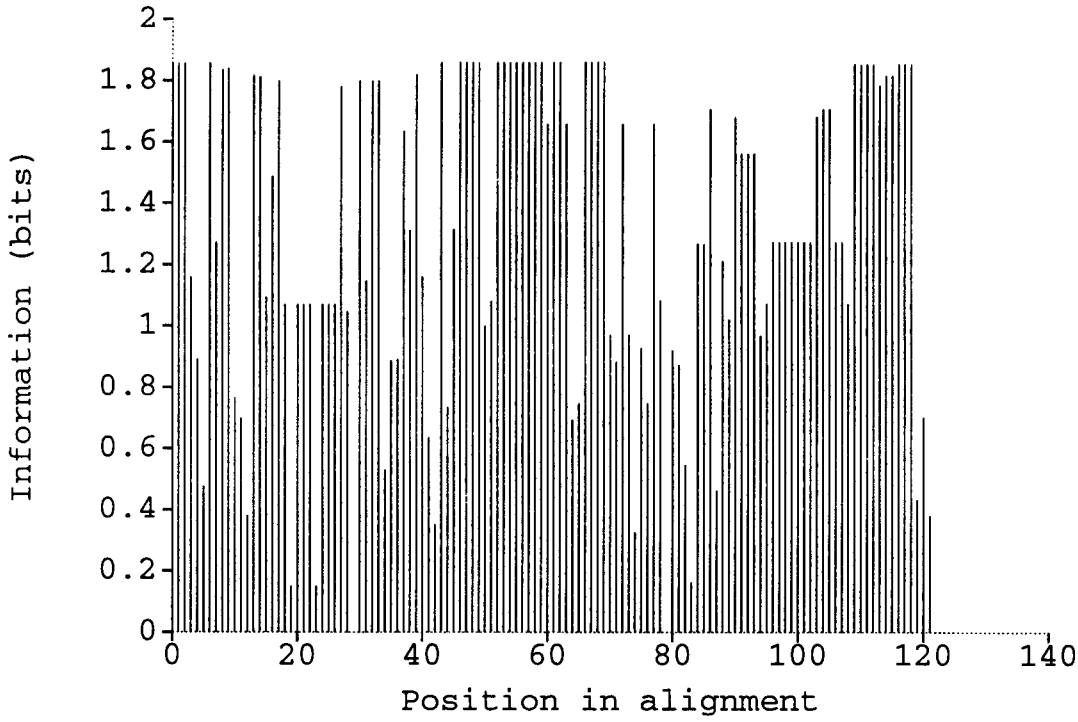


Figure 6.18. Information Content of the AcMNPV *hr*s. The information content of each position in the alignment of the *hr* region was determined. The y axis is the information content of that base position in the alignment measured in bits and the x axis is the position of the base in the *hr* region corresponding to alignment generated in Appendix 3. Position 56 corresponds to the G in the *Eco*RI site (GAATTC) found in the center of the imperfect palindromes.

pendix 3 and the information content is graphically displayed in Figure 6.18. In addition, the results obtained for the region including, and immediately flanking, the central palindromic core are displayed in a logo format [275] in Figure 6.19.

Using the aligned AcMNPV *hr* regions (See Appendix 3), a distance matrix was computed between all pairs of *hr*s using the Jukes and Cantor model of evolution [133]. Cluster analysis using PHYLIP was used to construct the tree (See Figure 6.20). Analysis of the tree indicated that the *hr*s segregated into three clusters and that the repeated palindromic elements within each *hr* region largely comprise these clusters. *Hr*3, *hr*1 and *hr*4a cluster, the *hr*2 region formed

Figure 6.19. Information Content for the Central Palindromic Region of AcMNPV *hr*s displayed in Logo Form.

a cluster and the *hr*5, *hr*1a and *hr*4b regions were split between two groups. The cluster containing the *hr*3, *hr*1 and *hr*4a repeats has accumulated a greater number of nucleotide differences than the other two clusters. *Hr*3, *hr*1 and *hr*4a repeat elements may have been present in the AcMNPV genome for a longer period of time and therefore had greater opportunity to accumulate nucleotide changes or alternatively selective pressures on sequence conservation may be less at these locations in genome. Recent duplication events may have given rise to *hr*2 as the repeated palindromic elements are very homogenous. The sequence differences between the *hr*s may be a result of random mutations in non–essential regions of the *hr*s. Alternatively, the sequence differences may also be important for the individual function of the *hr*s for instance in the timing of origin replication.

6.10.2 Determination of IE–1 Binding Site

Mutational analysis of both *hr*1a and *hr*5 [97, 259] suggest that IE–1 binds to a specific sequence immediately flanking the central palindromic core. Using a Gibbs sampling strategy for multiple sequence alignments [162], I compared the nucleotide sequences that have been shown to specifically bind IE–1 from my work using *hr*1a and that of others [97, 259] using *hr*5. By employing this algorithm, I was able compare the different DNA fragments and the number of complexes formed per DNA fragment. For instance, this algorithm can be used to find two identical sequence motifs within a DNA fragment that formed two (or more) complexes when used in EMSA. By comparing a number of the DNA fragments that have been tested for IE–1 binding activity, possible IE–1 binding motifs were determined. These nucleotide motifs were centered around a 5′–ACTCGTAA–3′ core sequence contained within the conserved imperfect palindrome.

Analysis of the promoter regions of the baculovirus genes involved in DNA

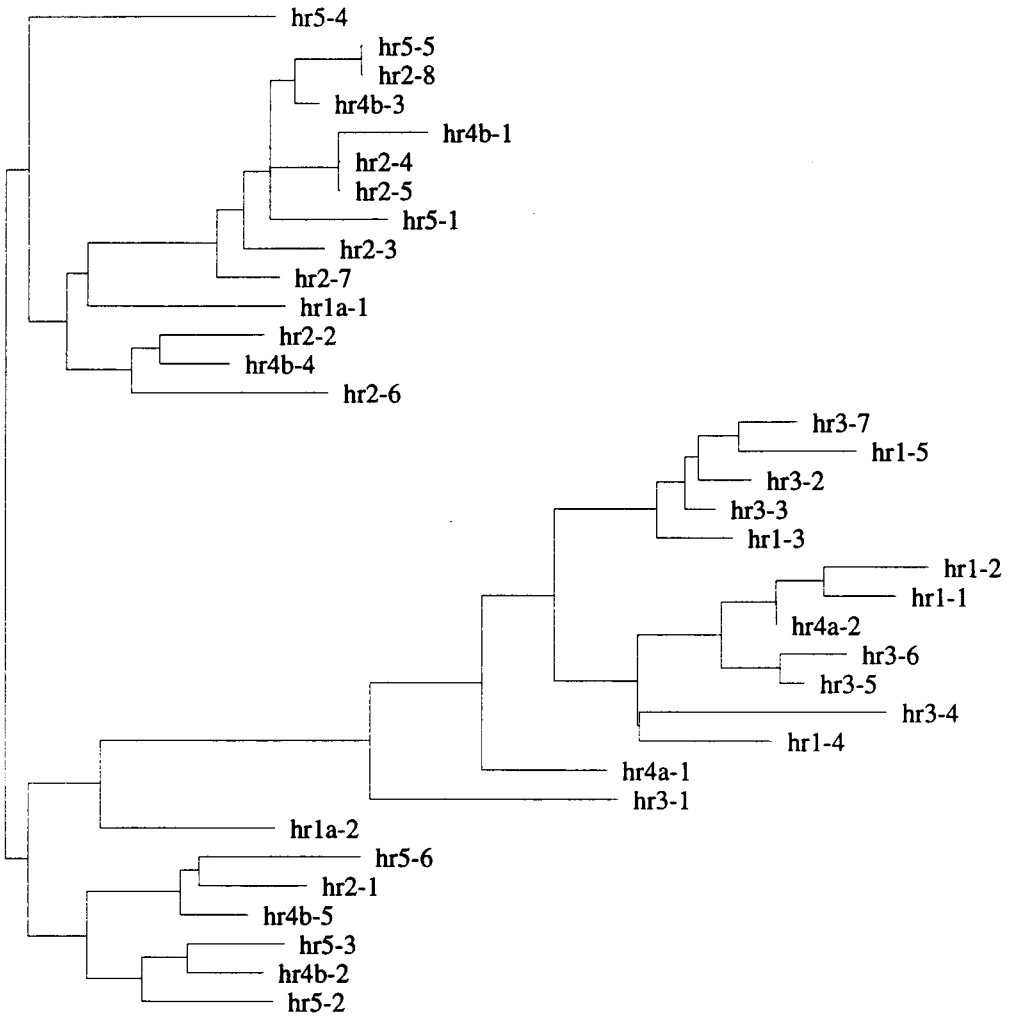


Figure 6.20. Cluster Analysis of the AcMNPV *Hr* Sequences

Figure 6.21. The Putative IE–1 Nucleotide Binding Sequence. Alignment of the *hr* consensus and the replication gene promoters that contain similar sequences. The putative IE–1 nucleotide binding motif is show in bold and underlined. The arrowheads mark the location of transcription initiation.

replication indicated that the *lef–1*, *ie–2* and *pe38* promoters contained the putative IE–1 binding motif identified above. Since the IE–1 binding site is palindromic, these data suggest that IE–1 binds to two sites containing the 5′–ACTCGTAA–3′ sequence motif within the palindrome (See Figure 6.21). A search of the ACMNPV genome revealed that this sequence is found in all the *hr*s with the exception of *hr*4c (which is not well conserved) and because the *hr*s are palindromic, the same sequence is present on the complementary strand. In addition, this motif was found in the promoter region of the *ie–0* gene. However, the sequence is only found once within the promoter regions of *lef–1*, *ie–2*, *pe38* and *ie–0*; it is not palindromic. Figure 6.21 shows the location of the motif within each of the promoter regions. The putative IE–1 binding motif is located 3′ to the TATA Box and either comprises or is a few bases upstream of the site of transcription initiation. The position of the putative IE–1 binding motif may be important for IE–1 regulation (see discussion). This sequence was also found in several other locations in AcMNPV genome but all these appear to be within coding regions and not located near promoter regions.

6.11 Protein–*hr*1a Interactions with Baculovirus Genes Required for Transient Replication

Whole–cell extracts made from cells transfected with each of the nine baculovirus genes involved in DNA replication were tested for ability to bind to *hr* sequences using EMSA. As shown in Figure 6.22, only whole–cell extracts made from *ie–1* transfected cells contained a DNA–binding activity. When both *ie–1* and each one of the other genes involved in replication were co–transfected no change in the mobility of the complex bound to *hr*1aΔEco was observed. Similarly, no change in mobility of the bound complex was observed when whole–cell extracts where prepared from Sf9 cells transfected with all the replication genes or when the complete *hr*1a probe was used (data not shown). Results from these experiments showed that under the conditions used only IE–1 bound to *hr*1a sequences indicating that other methods may be more useful in determining any additional protein–protein–DNA interactions between the replication proteins and *hr* sequences.

6.12 Does the OpMNPV IE–1 bind DNA in a manner similar to AcMNPV ?

The OpMNPV IE–1 predicted protein sequence is 21% identical at the amino terminus (maintaining an acidic profile) and 55% identical at the C–terminal to the AcMNPV IE–1 predicted protein sequence [299]. Transient assays showed that OpIE–1 was able to transactivate an AcMNPV *39K* CAT reporter construct in both Ld652–Y cells and Sf9 cells. Expression was enhanced by the presence of AcMNPV *hr*s, although OpMNPV IE–1 enhancement was less than that seen when AcMNPV IE–1 was transfected [299]. Constructs containing the OpMNPV enhancer region, referred to as OpE [300], showed that OpMNPV IE–1 also functioned as a transactivator in both Ld652–Y and Sf9 cells [300]. Therefore,

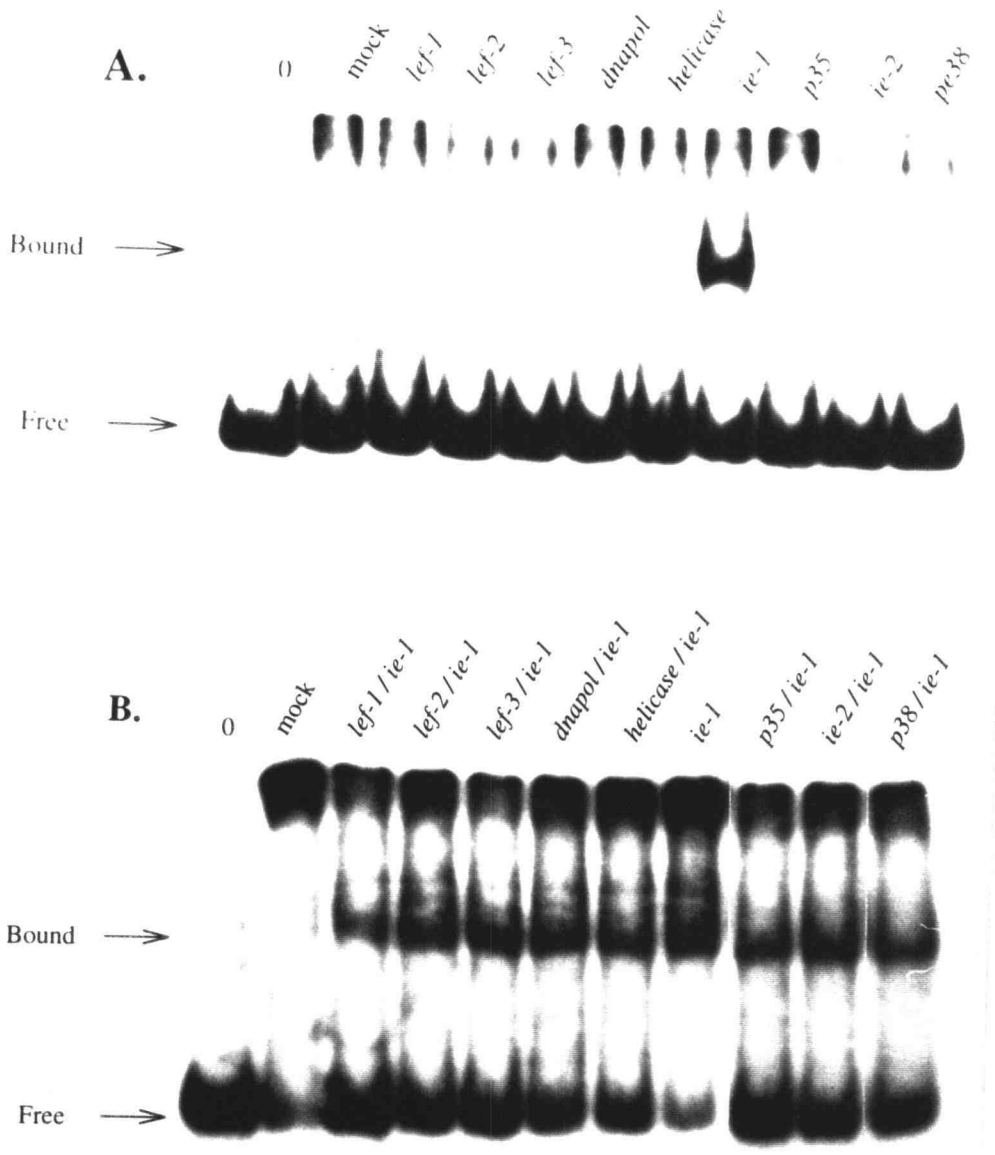


Figure 6.22. Gel Retardation Analysis of the Hr1aΔEco DNA–Replication Protein Complex. End–labeled Hr1aΔEco fragments were incubated in the presence of 5 μl of whole–cell extract prepared from Sf9 cells transfected with each replication gene or in combination with *ie–1*. A. Whole–cell extracts made from Sf9 cells transfected with a single replication gene. The name of the gene is indicated above the lane. B. Whole–cell extracts made from Sf9 cells transfected with *ie–1* and another gene involved in replication as indicated above the lane.

I performed gel–mobility shift analysis, to determine whether OpMNPV IE–1 bound to AcMNPV *hr* sequences and the OpMNPV enhancer, OpE.

6.13 Binding of Ld652–Y host–encoded protein to the AcMNPV *hr*1a is specific.

Gel retardation assays using the same experimental conditions as those established for AcMNPV pAcIE–1 transfected extracts were employed for examining the interactions of proteins from pOpIE–1 transfected Ld652–Y cells with *hr*1a. As shown in Figure 6.23, addition of increasing amounts of whole–cell extract from pOpIE–1 or pKS– transfected cells to *hr*1a resulted in the formation of a DNA–protein complex. However, no retarded bands were observed when the *hr*1aΔEco probe was used suggesting that a host–encoded protein(s) bound to the *hr*1a interregion. Gel retardation assays using *hr*1aIR indicated that host protein(s) binding to the interregion formed several complexes that were competed by addition of increasing amounts of unlabeled *hr*1aIR–containing DNA fragments (Figure 6.24). Competition experiments with an equivalently sized DNA fragment (a 297 bp *Xho*I–*Xho*I fragment) from the AcMNPV *p10* promoter, did not result in a decrease in the amount of complex formed (Figure 6.24), indicating that the host–encoded protein(s) were binding to *hr*1a interregion in a sequence–specific manner. These results indicate that both Sf9 and Ld652–Y cells produce proteins that bind to the *hr*1a interregion.

6.14 Host–encoded Protein Binding to OpMNPV Enhancer Sequences

Gel–mobility shift analysis was performed to determine whether OpMNPV IE–1 bound to the OpMNPV enhancer, OpE. OpE consists of a 66–bp element that is tandemly repeated partially or completely 12 times and has some sequence

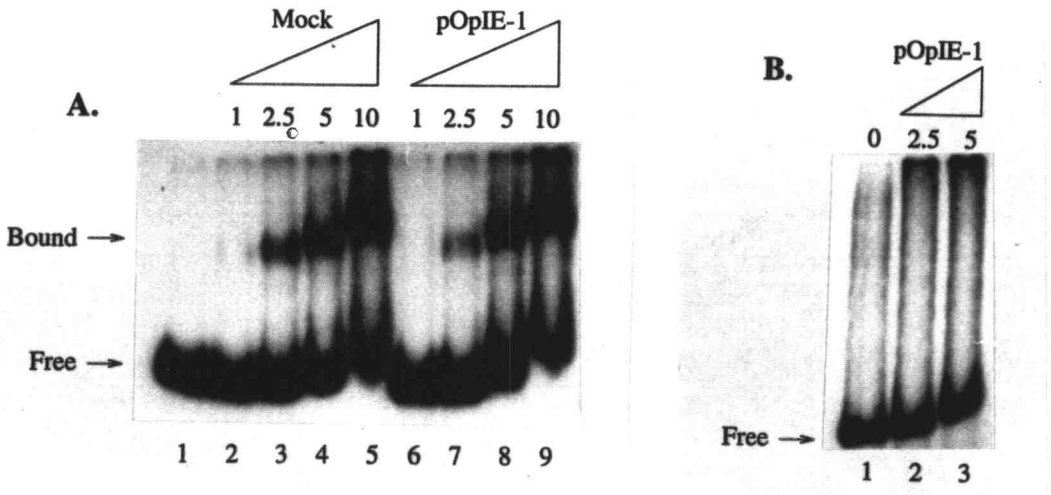


Figure 6.23. Gel Retardation Analysis of the *Hr*1a–Ld652–Y Protein Complex. A. End–labeled *hr*1a was incubated in the presence of 0 μl (lane 1), 1 μl (lane 2 and 6) 2.5 μl (lane 3 and 7), 5.0 μl (lanes 4 and 8) and 10 μl of whole–cell extract prepared from pOpIE–1 transfected Ld652–Y cells (lanes 6–9) or pKS–transfected Ld652–Y cells (lane 2–5). B. End–labeled Hr1aΔEco was incubated in the presence of 0 μl (lane 1), 2.5 μl (lane 2) and 5.0 μl (lane 3) of whole–cell extract prepared from pOpIE–1 transfected Ld652–Y cells

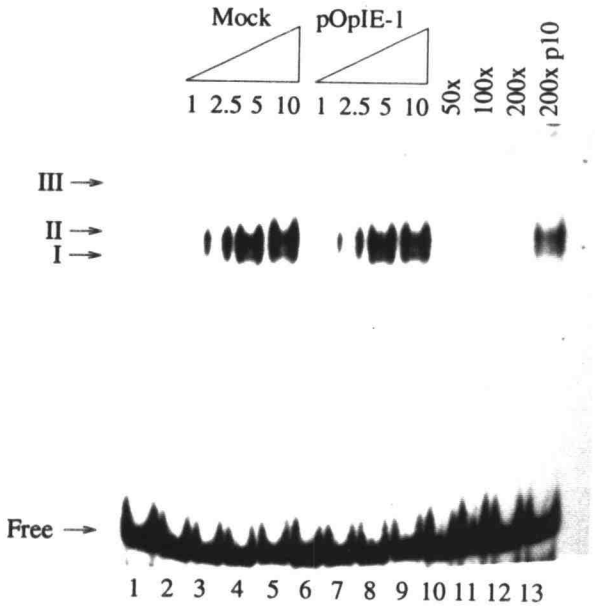


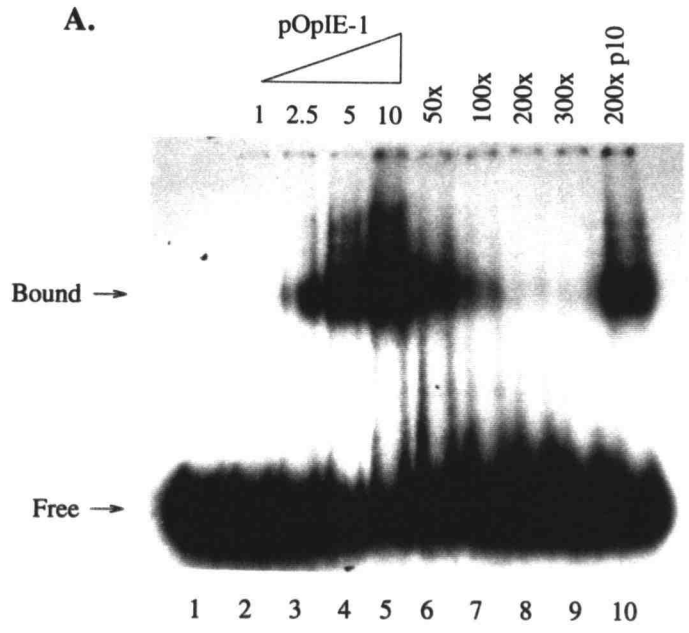
Figure 6.24. Gel Retardation Analysis of the *hr*1aIR–Ld652–Y Protein Complex. End–labeled *hr*1aIR was incubated in the presence of 0 μl (lane 1), 1.0 μl (lanes 2 and 6), 2.5 μl (lanes 3 and 7) 5.0 μl (lanes 4 and 8) and 10 μl (lanes 5 and 9–13) of whole–cell extract prepared from pOpIE–1 (lanes 6–13) or pKS– (lanes 2–5) transfected Ld652–Y cells. Both radiolabeled probe and competitor DNA were added simultaneously to the reaction mixtures. Competitor DNA, the unlabeled *hr*1aIR fragment, was added in excess molar amounts as indicated at the top of each lane. An unlabeled 297 bp *Xho*1–*Xho*1 AcMNPV *p10* promoter fragment was used as the competitor in lane 13.

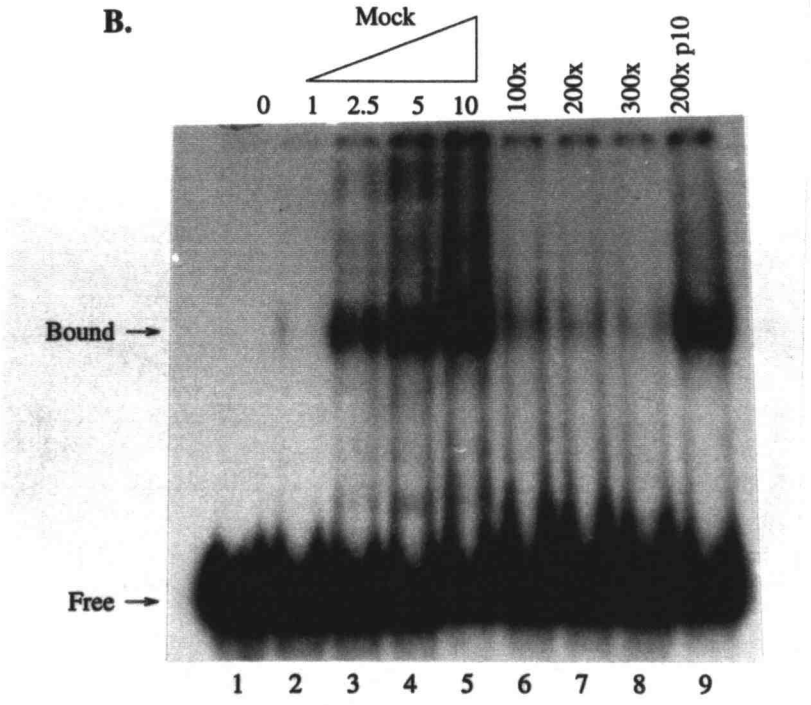
```
                                        CGCCGCTGACGCACCGCT------AGCACGCG
CCG TTCGAGAGCGTTCGCACCCGAA AAGCAGGGTCGCCGCTGACGCACCGCTAAAAATAGCACGCG
CCT TTCGAGAGCGTTCGCACCCGAA AAGCAGGGTCGCCGATGACGCACCGCTAAAGTAA
```

Figure 6.25. Sequence of the OpE–NB Enhancer Construct. The repeated elements are aligned and the palindromic region is boxed.

similarity to AcMNPV *hr* elements [300]. Southern Blot hybridization indicated that similar regions were found in four other regions in the OpMNPV genome. When the OpE region was cis–linked to an OpMNPV *vp39*–CAT reporter gene construct and co–transfected with *ie–1* into both Sf9 or Ld652–Y cells, expression of *vp39* was increased at least two–fold indicating that OpE does function as an enhancer element [300]. More recently, a 2.3 kb region containing 7 complete repeats and 5' flanking sequence of the OpE region was found to a function as a putative origin of replication in transient DNA replication assays [5]. Therefore, to determine if OpIE–1 transfected extracts contain a DNA binding activity specific for OpE repeats, I performed gel mobility shift analysis using a DNA fragment containing 2.5 repeat elements. This construct was referred to as OpE–NB and is shown in Figure 6.25. I was unable to detect specific OpMNPV IE–1 complex formation using the experimental conditions established for the homologous AcMNPV protein. However, I did detect host factor(s) from the Ld652–Y cells that bound specifically to the OpE region (See Figure 6.26). Similar results were seen if whole–cell extracts were made using pOpIE–1 transfected Sf9 cells (data not shown). Finally, Sf9 mock–transfected whole–cell extracts also contained a DNA binding activity specific for OpMNPV enhancer elements (data not shown).

Figure 6.26. Gel Retardation Analysis of the OpMNPV Enhancer Element. A. End–labeled OpENB was incubated in the presence of 0μl (lane 1), 1.0μl (lane 2), 2.5μl (lane 3) 5.0μl (lane 4) and 10μl (lanes 5 and 6–10) of whole–cell extract prepared prepared from pOpIe–1 transfected Ld652–Y cells. Both radiolabeled probe and competitor DNA were added simultaneously to the reaction mixtures. Unlabeled OpENB DNA fragments were added in excess molar amounts as indicated at the top of each lane. An unlabeled 297 bp *Xho*1–*Xho*1 AcMNPV *p10* promoter fragment fragment was used as the competitor in lanes 10 and 19. B. End–labeled OpENB was incubated in the presence of 0μl (lane 1), 1.0μl (lane 2), 2.5μl (lane 3) 5.0μl (lane 4) and 10μl (lanes 5–9) of whole–cell extract prepared prepared from pKS– transfected Sf9 cells. Competition analysis was performed as described in A.


A.
pOpIE-1
1 2.5 5 10
50x
100x
200x
300x
200x p10
Bound
Free
1 2 3 4 5 6 7 8 9 10


B.
Mock
0 1 2.5 5 10
100x
200x
300x
200x p10
Bound
Free
1 2 3 4 5 6 7 8 9

6.15 Summary

The results of the gel retardation assays indicate that pAcIE–1 transfected cells contain factors that bind specifically to *hr*1a. Supershift assays using polyclonal antibodies directed against IE–1 show that IE–1 either binds *hr*1a directly, or as part of a complex that includes host factors. Gel retardation assays using a number of different *hr*1a mutants indicated that IE–1 bound specifically to regions flanking the central palindromic *Eco*RI site. However, mutant *hr*1a sequences in which small deletions or insertions were introduced into the central *Eco*RI core region showed either greatly diminished or no replication or enhancer functions suggesting that IE–1 binding by itself is not sufficient for activation of replication and enhancer functions.

Gel retardation assays indicated that pAcIE–1 transfected cells contain factors that bind specifically to the *ie–2* and *pe38* promoters. Comparison of the IE–1 binding data for *hr* elements and analysis of the promoter regions of the *ie–2* and *pe38* genes revealed a putative IE–1 nucleotide binding sequence, 5′–ACTCGTAA–3′.

The gel retardation assay was also used to investigate additional protein–DNA interactions between the nine baculovirus genes required for transient replication. Results from these experiments indicated that only IE–1 bound to *hr*1a sequences under the conditions used.

Lastly, I performed gel mobility shift analysis, to determine whether the homologous protein in the OpMNPV baculovirus bound to AcMNPV *hr* sequences and OpMNPV putative origins of replication. Results from these experiments showed that a Ld652–Y host encoded protein(s) bound to the interregion of AcMNPV *hr*1a and that host encoded protein(s) from both Sf9 and Ld652–Y cells bound specifically to the OpMNPV enhancer element.

Chapter 7
Effect of Palindromic Mismatches on AcMNPV *Hr* Function

7.1 Introduction

Numerous studies have shown that B–form double–stranded DNA can be altered depending on the local nucleotide sequence and environmental conditions such as ionic strength, temperature, pH and topology (for review see [213]). DNA structural polymorphisms include Z–DNA [316], bent DNA [277, 276], tetraplex DNA [278] and cruciform formation [181, 200]. It has long been recognized that palindromic sequences found in double–stranded DNA might convert to hairpin structures known as cruciforms [242]. Thermodynamic calculations indicate that depending on length and number and type of mismatches within the palindromic sequence, cruciforms may form at moderate negative superhelicities [213]. Furthermore, statistical mechanical calculations suggest that a number of cruciforms found in plasmids and bacteriophages may form under conditions typical of supercoiled DNAs from natural sources [314]. Nevertheless, the biological significance of DNA structural perturbations is only now being explored.

The *hr* regions of the AcMNPV have been shown to function as enhancers of RNA polymerase II–mediated transcription [98, 102, 218, 259] and as putative origins of DNA replication in transient replication assays [235, 149, 152, 170, 169]. The AcMNPV genome contains eight *hr*s composed of varying numbers of highly conserved repeat elements. Each *hr* region has two to eight copies of a 30–bp imperfect palindrome flanked on both sides by approximately 20–bp of a direct repeat element [102, 7]. Conversion of the consensus 30–bp imperfect palindrome into a cruciform structure would result in a 13–bp stem containing two sites of mismatches.

DNA transitions such as local base pair melting are necessary for the processes of transcription [61] and replication [57]. Both of these processes re-

quire single–stranded DNA to serve as templates either for RNA transcript or DNA daughter–strand production. Transitions such as cruciform formation at *hr* regions in AcMNPV may facilitate local base pair melting thereby assisting the processes of baculovirus transcription and replication. Therefore, investigations were undertaken to study the possible role of cruciform formation in baculovirus replication and transcription. Oligonucleotides containing the AcMNPV *hr* consensus imperfect palindrome sequence and a perfect palindrome sequence were synthesized. The oligonucleotide sequences and their putative cruciform structures are shown in Figure 7.1. These oligonucleotides where then cloned into pKS– and used in assays to study cruciform formation including nuclease P1 assays and two–dimensional (2–D) electrophoresis of topoisomer ladders. The *ie–1* gene product is involved in transcriptional activation [101, 103, 218, 102, 98, 28, 185, 30] and is essential for baculovirus DNA replication [187, 151, 148]. Furthermore, IE–1 has been shown to bind to *hr*s [97, 96, 259, 169]. To examine the role of palindrome formation may play in IE–1 binding to *hr*s, gel retardation analysis was employed to characterize the ability of IE–1 to bind to the imperfect and perfect oligonucleotides. The ability of the perfect and imperfect palindrome to act as an enhancer was assayed by cloning the perfect and imperfect oligonucleotides into the GUS reporter plasmid, p39KGUSpI18, and monitoring GUS activities in extracts of Sf9 cells that had been co–transfected with these reporter constructs and increasing concentrations of pAcIE–1.

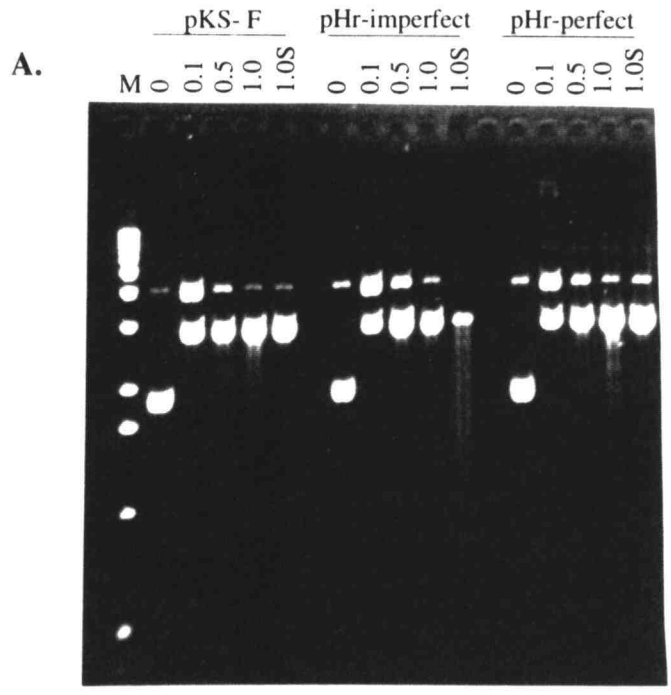
7.2 Nuclease P1 Assays

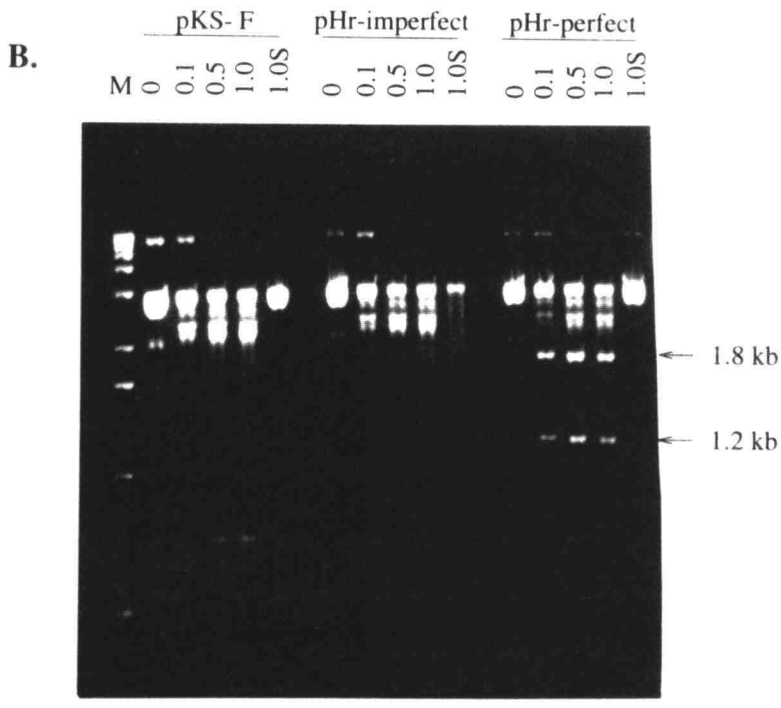
P1 nuclease, a single–strand specific nuclease, can be used to determine the presence and location of single–stranded DNA regions, (such as those found at the apex of cruciforms) in circular supercoiled plasmids. If P1 cleavage occurs at specific single–stranded sites within the plasmid DNA, it would result in a

Figure 7.1. Potential Cruciform Structure of the *hr* Perfect and Imperfect Palindromes. The imperfect and perfect oligonucleotide sequences are shown in capitol letters while the sequences derived from the pKS– vector are indicated by lower case letters. The stars indicate the sites of mismatched base pairs.

distinct banding pattern following digestion with restriction endonucleases and gel electrophoresis. Figure 7.2a shows the results of P1 cleavage of supercoiled plasmids pKS–F (the vector control), pHr–imperfect and pHr–perfect. While the starting material contains a minor amount of relaxed or nicked plasmid, cleavage of each sample by increasing amounts of nuclease P1 resulted in the conversion of the supercoiled plasmids to nicked and linear molecules. Figure 7.2b shows the results of gel electrophoresis of the P1–treated plasmids following digestion with the restriction enzyme *Sca*I. *Sca*I cleaves at a single site in the plasmids. Nuclease P1/*Sca*I digestion of the control plasmid, pKS–F, resulted in the production of several distinct bands. This result is not unexpected since the parent vector of pKS–F is a derivative of pBR322. pBR322 has been shown to contain a major and two minor sites of cleavage that correlate with sequences containing inverted repeats [181]. The digestion pattern of the pHr–imperfect construct was indistinguishable from pKS–F, the control vector, indicating that the consensus *hr* palindrome containing two mismatches does not form a cruciform under these experimental conditions. However, as Figure 7.2b clearly shows, the pHr–perfect digestion pattern is different from that of the vector control and pHr–imperfect indicating that an additional nuclease P1–sensitive site occurs within this construct. The sizes of the specific fragments produced by *Sca*I digestion (1.2 kb and 1.8 kb, see Figure 7.2b) are consistent with the formation of a cruciform structure centered at the *Eco*R1 of the perfect palindrome indicating that the perfect palindrome does form a cruciform structure under these experimental conditions.

Figure 7.2. Nuclease P1 Assay of the *Hr* Perfect and Imperfect Constructs. A. Nuclease P1 digestion of supercoiled plasmid DNA. B. *Sca*I digestion of the nuclease P1–treated reactions. The numbers above the wells indicate the units of nuclease P1 used in the reaction. The lane marked M contains the BRL 1 Kb DNA ladder. Samples in the lanes marked 1.0S were digested with *Sca*I before digestion with 1.0 unit of nuclease P1.


A.
pKS- F
pHr-imperfect
pHr-perfect
M 0 0.1 0.5 1.0 1.0S 0 0.1 0.5 1.0 1.0S 0 0.1 0.5 1.0 1.0S


B.
pKS- F
pHr-imperfect
pHr-perfect
M 0 0.1 0.5 1.0 1.0S 0 0.1 0.5 1.0 1.0S 0 0.1 0.5 1.0 1.0S
1.8 kb
1.2 kb

The pHr1aΔEco construct was also examined for nuclease P1 sensitive sites. This construct, derived from *hr*1a, contains a single palindrome created by fusing the left half of the first *hr*1a palindrome with the right half of the second palindrome and 55 bp of 5′ and 210 bp of 3′ flanking sequences (See Figure 6.1). Figure 7.3 shows that the pHr1aΔEco digestion pattern is different from that of the vector control, pKS– indicating that an additional nuclease P1–sensitive site occurs within this construct. The sizes of the specific fragments produced by *Sca*I digestion, 2.1 and 1.2 kb, are consistent with the formation of a cruciform structure centered at the *Eco*R1 of the single *hr* contained in this construct. This result indicates that the *hr*1a palindrome does form a cruciform structure under these experimental conditions. Furthermore, since the *hr*1a palindrome contains 2 mismatched base pairs similar to pHr–imperfect, these results suggest that the *hr*1a flanking sequences may play a critical role in cruciform formation.

7.3 Two–dimensional Gel Electrophoresis of Topoisomers

Circular double–stranded DNA can exist as topoisomers that differ in number of times one DNA strand wraps around the other strand and is measured as the linking number, L_K. L_K is composed of two components, the twist (T_W) which is the number of times the DNA helix crosses itself and the writhe (W_r) which is the twisting of the DNA duplex axis in three–dimensional space. The relationship between these terms is defined by the equation:

$$L_k = T_w + W_r \tag{7.1}$$

Treatment of supercoiled DNA with topoisomerase I in the presence of intercalating agents such as ethidium bromide (See Chapter 2.27) results in the production of a topoisomer ladder with each member of the ladder differing from the others by a linking number of one. The difference between the topoisomers results in different mobilities during electrophoresis on agarose gels; the greater the degree

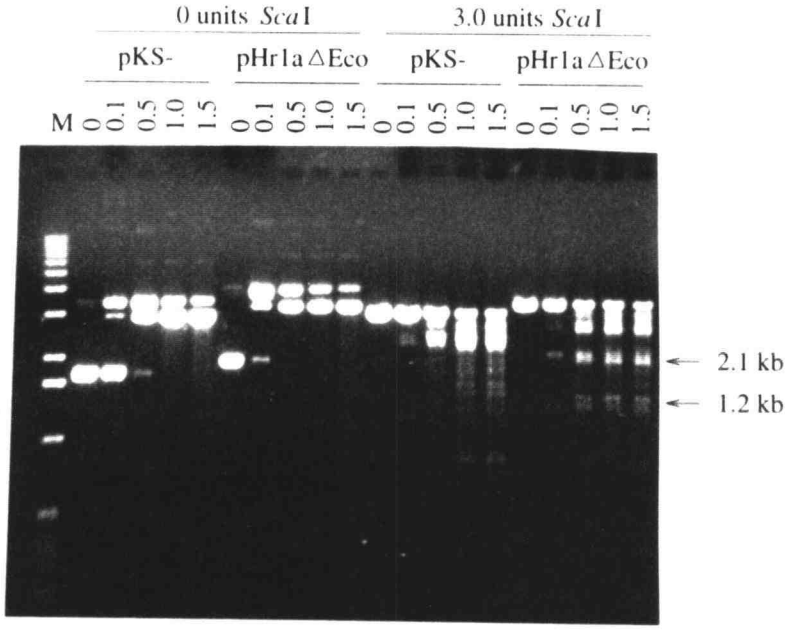


Figure 7.3. Nuclease P1 Assay of the pHr1aΔEco Construct. The numbers above the wells indicate the units of nuclease P1 used in the reaction. The plasmid construct and the units of the restriction endonuclease, *Sca*I, used in the reactions are also indicated above the wells. The lane marked M contains the BRL 1 Kb DNA ladder.

of supercoiling the faster the migration of the topoisomer. 2–D electrophoresis is performed to improve the resolution of the different topoisomers and permits the identification of DNA structural transitions.

The energy associated with DNA supercoiling can stabilize the formation of an altered DNA structure such as a cruciform. Formation of a cruciform would result in a change in the twist of the molecule. This results in a change in writhe because, according to aforementioned mathematical relationship, the linking number remains constant for a closed–circular molecule. The change in writhe alters the shape of the molecule and hence its electrophoretic mobility. Therefore, a plasmid containing sequences with the ability to form a cruciform would migrate more slowly than a plasmid with the same L_K. The amount of retardation seen in plasmids containing altered DNA structures such as cruciforms can be used to estimate the energy required for the DNA structural transition.

2–D gel analysis was performed on topoisomers of the pHr–imperfect, pHr–perfect and the vector control, pKS–F. The results are shown in Figure 7.4. DNA structural transitions were not observed when both the pKS–F and pHr–imperfect constructs were analyzed. However, a DNA structural transition was observed when pHr–perfect topoisomers where subjected to 2–D gel analysis (See Figure 7.4c). The free energy required for cruciform formation can be calculated using the following equation (for discussion of the derivation of this equation see [22]):

$$\Delta G = \frac{1100\mathrm{RT}}{\mathrm{N}} (\mathrm{L}_k - \mathrm{L}_w) \tag{7.2}$$

where R is the gas constant with a value of 1.9865 calories/mole/degrees Kelvin, T is the temperature in degrees Kelvin, N is the size of the plasmid in bp and L_k is the linking number of the topoisomer at the midpoint of the DNA structural transition and L_w is the linking number of the unshifted topoisomer of the same mobility.

The length of the extruded DNA structure can be determined by the equa-

tion:

$$\Delta T_w = \frac{N}{10.5} + 1 \qquad (7.3)$$

where N is the length of the extrusion. Calculations from equation 7.2 and 7.3 indicated that 9.6 kcal/mole was required to form the *hr* perfect cruciform and that 30–bp were extruded from the plasmid. This result correlates well with expected cruciform structure which consists of 30–bp. These data are consistent with the nuclease P1 analysis which indicated that the pHr–perfect construct contained a cruciform centered at the *Eco*R1 site of the *hr* palindrome.

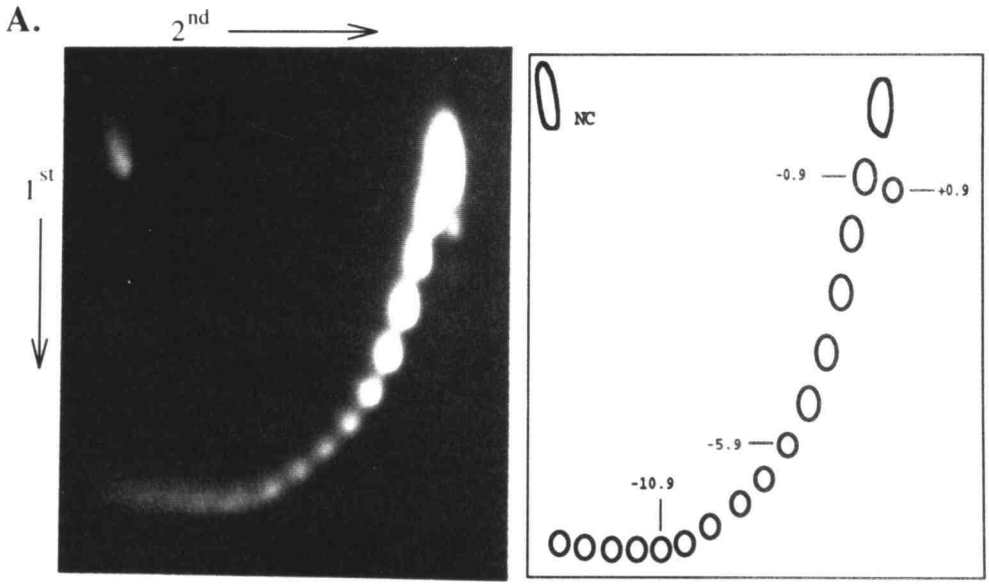


Figure 7.4. Two-Dimensional Electrophoresis of pHr-perfect and pHr-imperfect Topoisomers. For each construct, the ethidium bromide-stained agarose gel and a schematic representation of the gel is shown. Topoisomers were electrophoresed on 1.2% agarose gels for 20 hr in the first dimension and for 16 hr in the presence of 6.0μg per ml chloroquine in the second dimension. In the schematic representation, the linking numbers of selected topoisomers are indicated, the arrows show the DNA structural transition and the numbers in parentheses indicate the change in twist associated with the transition. NC denotes nicked circular DNA. A. pKS-F, the vector control B. pHr-imperfect C. pHr-perfect.

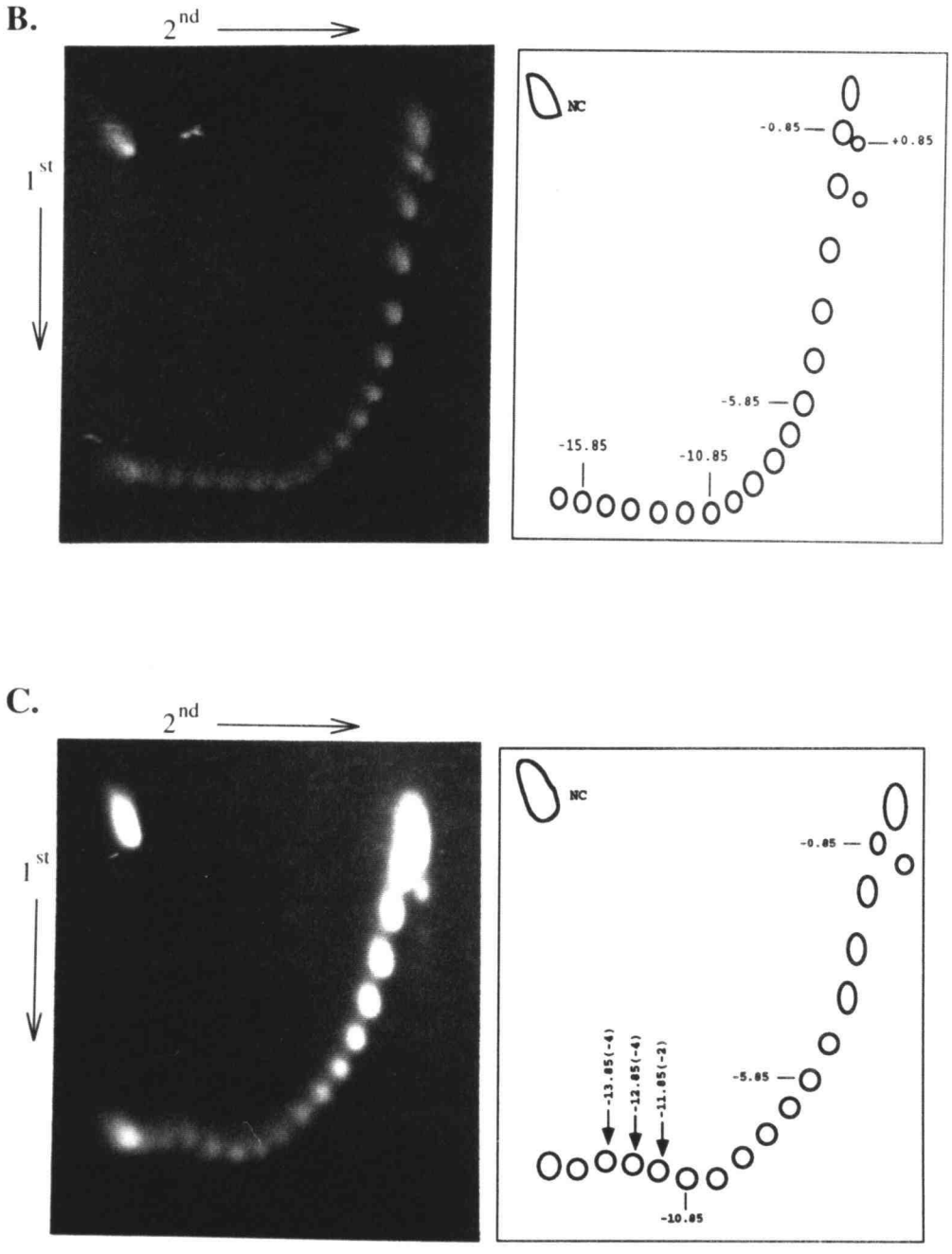


Figure 7.4 (continued)

2–D gel analysis using different electrophoresis conditions was performed on topoisomers of the pHr–imperfect, pHr1aΔEco and pKS–F . The results of these experiments are shown in Figure 7.5. Again, both the pKS–F and pHr–imperfect constructs did not exhibit any DNA structural transitions. However, calculations indicate that pHr1aΔEco extrudes 63–bp with a free energy requirement of 19.6 kcal/mole.

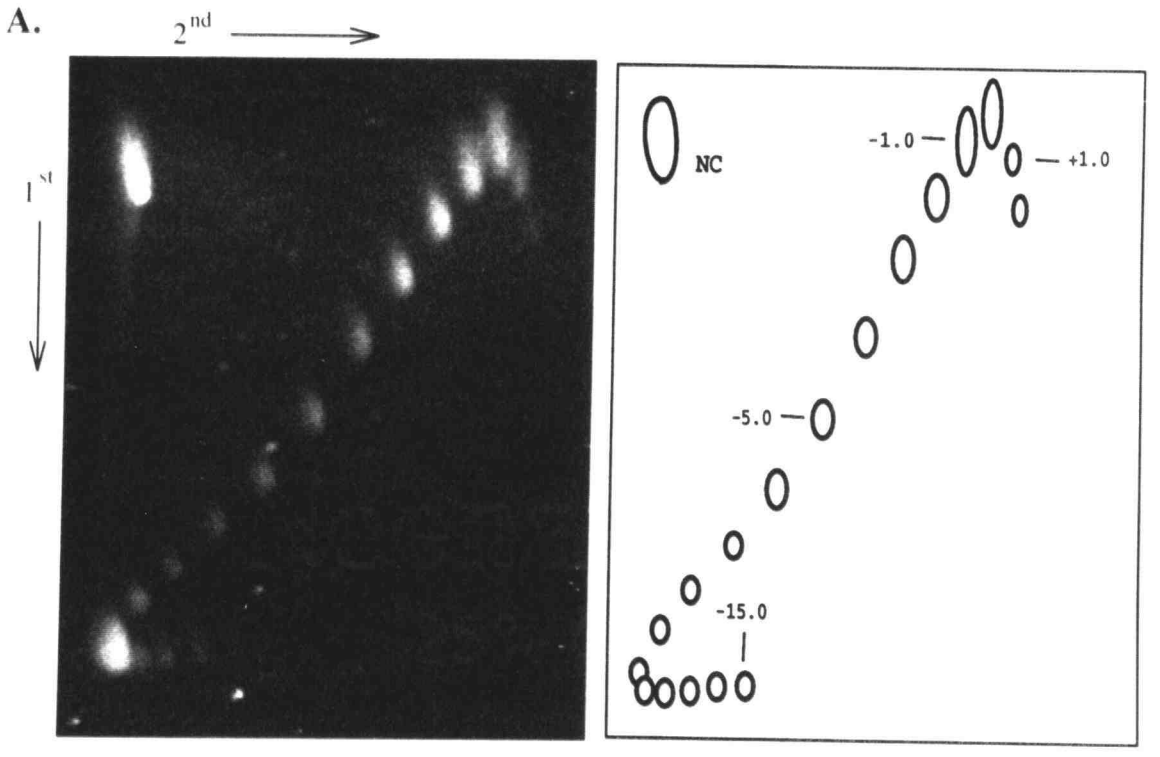
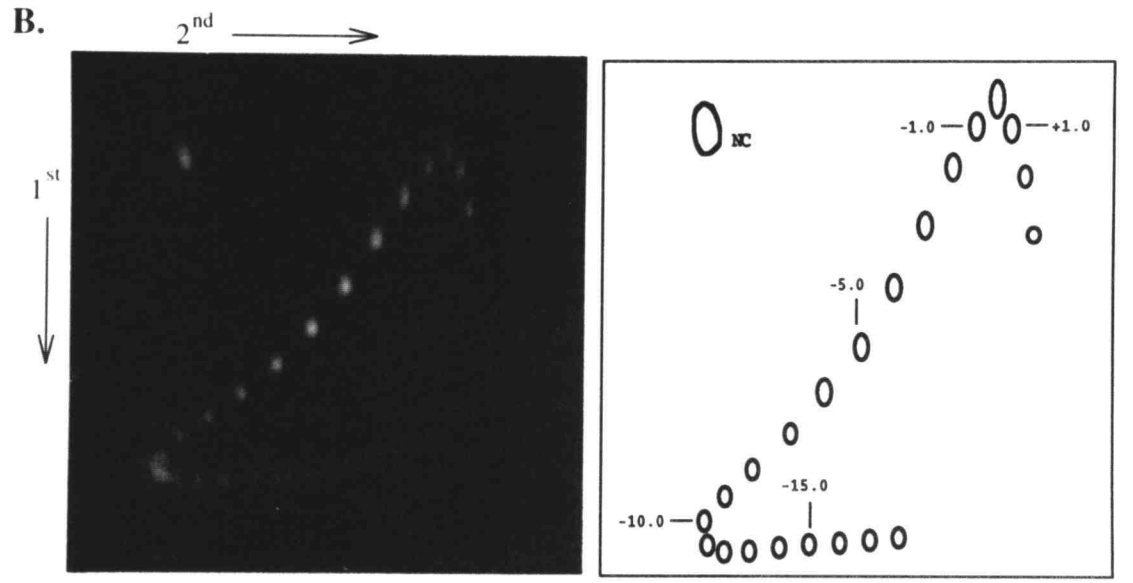


Figure 7.5. Two–Dimensional Electrophoresis of pHr–imperfect and pHr1aΔEco Topoisomers. For each construct, the ethidium bromide–stained agarose gel and a schematic representation of the gel is shown. Topoisomers were electrophoresed on 1.5% agarose gels for 30 hr in the first dimension and for 24 hr in the presence of 1.8μg per ml chloroquine in the second dimension. In the schematic representation, the linking numbers of selected topoisomers are indicated, the arrows show the DNA structural transition and the numbers in parentheses indicate the change in twist associated with the transition. NC denotes nicked circular DNA. A. pKS–F, vector control B. pHR–imperfect C. pHr1aΔEco.



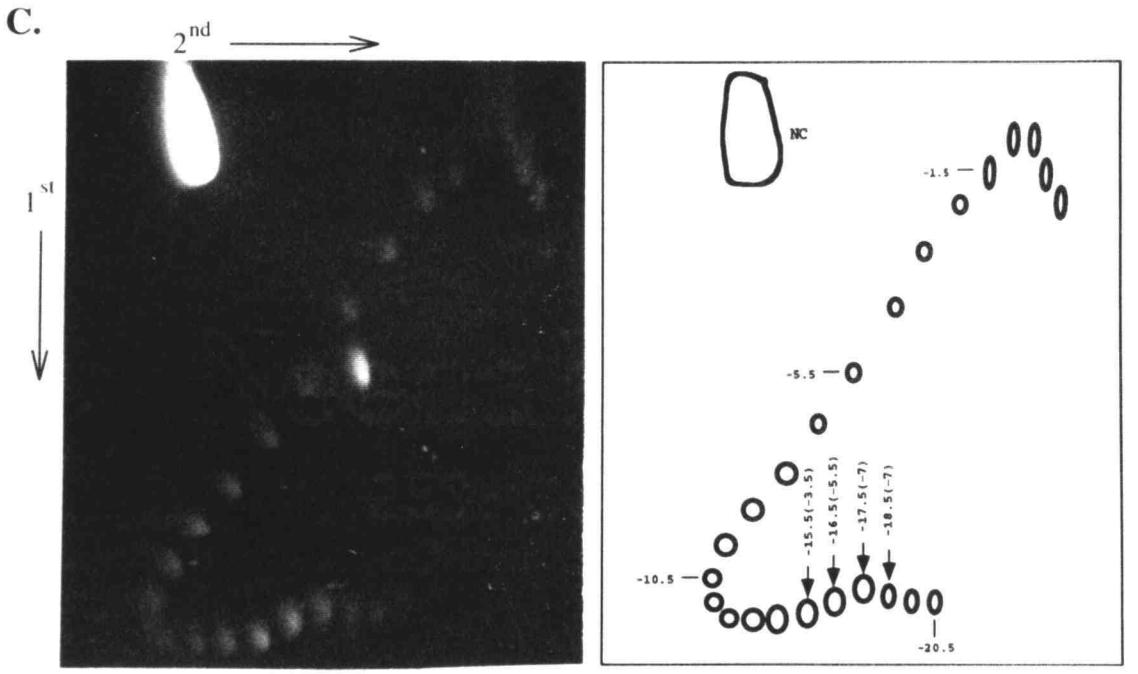


Figure 7.5 (continued)

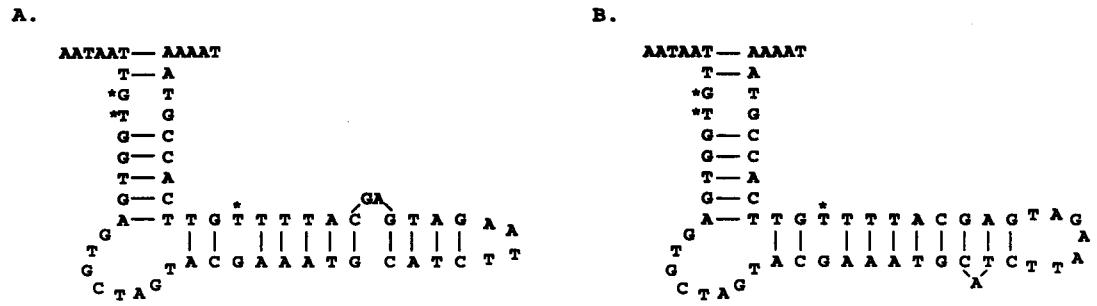


Figure 7.6. Potential DNA Secondary Structures of the *hr*1a Single Palindrome.

Analysis of this region using the M–FOLD program in GDE [339] revealed that the pHr1aΔEco insert has the potential to form two 56 bp cruciform–like structures; these stem–loop structures were centered around the central *hr* palindrome but contained an additional stem and loop (Figure 7.6). Alternatively, the AT–rich flanking sequences could allow additional melting of the DNA in this region. These sequences, flanking the conserved region of the AcMNPV *hr* consensus are not well–conserved suggesting that cruciform formation by pHr1aΔEco may be fortuitous. Furthermore, the free energy requirement for pHr1aΔEco cruciform formation suggests that the cruciform may not exist *in vivo*. The free energy of formation of cruciforms thought to form *in vivo* is typically between 13 and 18 kcal/mole [213]. Since a DNA structural transition for pHr1aΔEco but not pHr–imperfect was observed, a lower limit of 19.6 kcal/mole on the energy required for AcMNPV *hr* cruciform formation can be assumed. Therefore, it is unlikely to occur *in vivo* without thermodynamic assistance from DNA–binding proteins.

7.4 Binding of Proteins from pAcIE–1 Transfected Extracts to the *Hr* Perfect and Imperfect Oligonucleotides

To determine if IE–1 was capable of interacting with the imperfect and perfect palindrome constructs, gel retardation assays were employed using extracts from pAcIE–1 transfected cells. As shown in Figure 7.7, addition of increasing amounts of whole–cell extract from pAcIE–1 transfected cells to annealed radiolabeled oligonucleotides containing either the imperfect palindrome consensus or the altered perfect palindrome resulted in the formation of a DNA–protein complex. Quantitation performed using a PSI–486 Phosphoimager SI and Imagequant workstation (Molecular Dynamics) showed that doubling the amount of extract used doubled the portion of shifted complex when both Hr–imperfect and Hr–perfect probes were used. Furthermore, proteins from the pAcIE–1 extracts seemed to bind either probe with similar affinities. These results were unexpected because Guarino and Dong [97] had failed to detect DNA–protein interactions using a 40–bp oligonucleotide containing the central 24–bp palindrome. However, using 42–bp oligonucleotides, I was able to detect IE–1 binding to *hr*s.

To confirm that the protein complex binds specifically to the perfect and imperfect palindromes, the complex was competed by addition of increasing amounts of unlabeled perfect and imperfect oligonucleotides and the *hr*1a–containing DNA fragment to the reaction. These results, shown in Figure 7.8 and Figure 7.9 indicate that the DNA–protein complex formation was specific. Quantitation indicated that the complex was competed equally well with either the Hr–imperfect or Hr–perfect oligonucleotides. The ratio of bound to free probe decreases at the same rate whether the Hr–perfect or Hr–imperfect oligonucleotides are used as competitors. This is shown graphically in Figure 7.10. Addition of 60– and 120–fold molar excess of the *hr*1a–containing fragment resulted in competition of the shifted complexes, while 60– and 120–fold molar

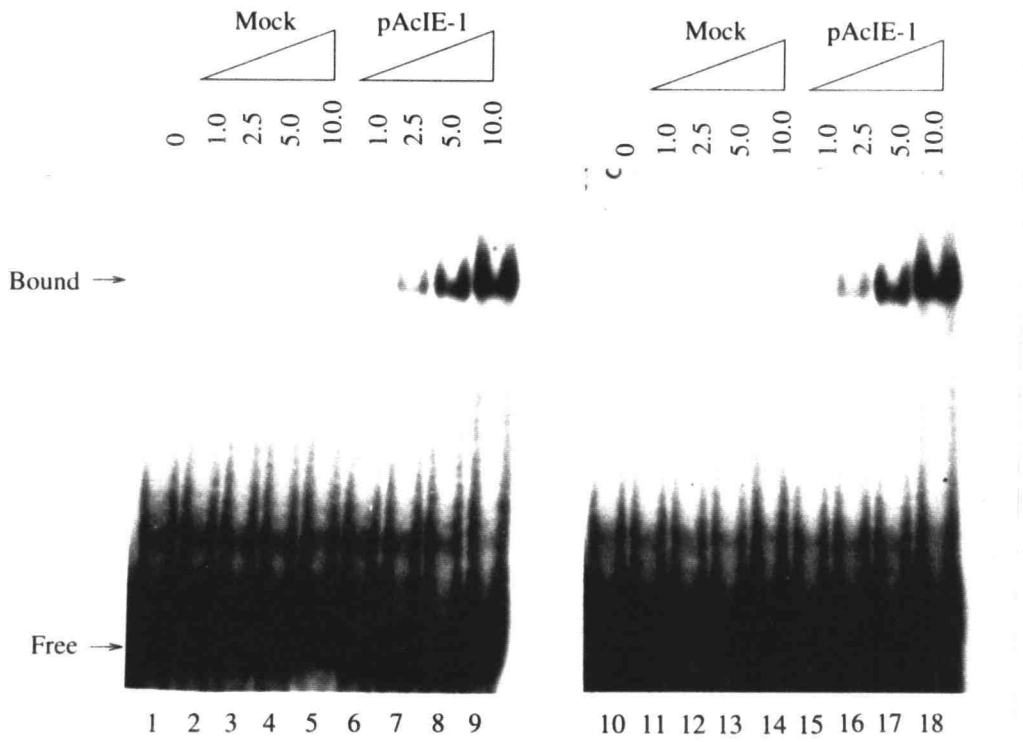


Figure 7.7. Gel Retardation Analysis of *Hr* Perfect and Imperfect palindromes. The radiolabeled oligonucleotides containing the perfect (lanes 1–9) and the imperfect (lanes 10–18) palindromes were incubated in the presence of increasing amounts in μl of whole-cell extract prepared from either pAcIE-1 transfected (lanes 6–9 and 15–18) or mock-transfected (lanes 2–5 and 11–14) Sf9 cells as indicated. Extract protein concentration was 2.5 μg /μl .

excess of the DNA fragment containing the AcMNPV *p10* promoter did not result in a decrease in the amount of complex formed (See Figures 7.8 and 7.9, compare lanes 5 and 9 to lanes 11 and 12). These results indicated that the protein(s) were binding to both the Hr–perfect and Hr–imperfect oligonucleotides in a sequence–specific manner. Finally, the annealed *vp39* and ML oligonucleotides (see Chapter 5.7 and 5.8) failed to compete IE–1 binding to the Hr–perfect and Hr–imperfect oligonucleotides (data not shown).

7.5 IE–1 is a Component of the DNA–Protein Complex

Sf9 cells transfected with pAcIE–1 contain protein(s) that bind to the perfect and imperfect oligonucleotides, suggesting that IE–1 is a component of the complex. Gel retardation assays performed in the presence of polyclonal antibodies to IE–1 [223] resulted in the appearance of a slower migrating complexes or supershifts (Figure 7.11 and Figure 7.12, lanes 3–5), indicating that IE–1 is a component of the DNA–protein complex bound to both the consensus *hr* palindrome and the perfect palindrome. Supershifts were not detected when pre–immune serum (Figure 7.11 and Figure 7.12, lanes 9–11), polyclonal antibodies to the OpMNPV polyhedrin protein [263] (Figure 7.11 and Figure 7.12, lanes 6–8) or mouse monoclonal antibodies to *Drosophila* TBP (data not shown) were added to the reaction mixture.

It has been reported that a cruciform structure containing mismatched base pairs is required for high–affinity binding of the cAMP response element–binding protein [285]. The AcMNPV *hr* central core palindrome is highly conserved and includes two regions of mismatched base pairs. Therefore, I was interested in testing whether the predicted *hr* cruciform structure containing these mismatches was a high–affinity binding site for AcMNPV IE–1. Using gel retardation analysis, I failed to detect specific binding to the radiolabeled single oligonucleotides that had been boiled in 50 mM NaCl, 50 mM Tris–HCl pH 7.5, 1 mM EDTA and

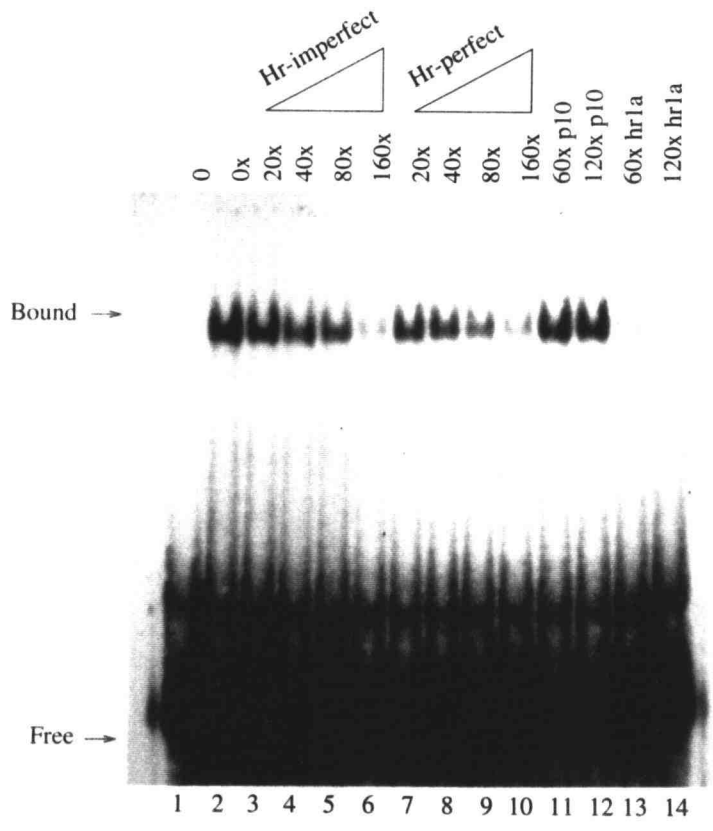


Figure 7.8. Gel Retardation Analysis of *Hr* Imperfect palindrome. The radiolabeled oligonucleotides containing the imperfect palindrome were incubated in the presence of 12.5 μg whole-cell extract prepared from pAcIE-1 transfected Sf9 cells. Both radiolabeled probe and competitor DNA were added simultaneously to the reaction mixtures. Competitor DNA was added in excess molar amounts as indicated at the top of each lane: lanes 3–6, unlabeled imperfect palindrome oligonucleotides; lanes 7–10, unlabeled perfect palindrome oligonucleotides; lanes 11 and 12, unlabeled 297 bp *Xho*1–*Xho*1 AcMNPV *p10* promoter fragment; lanes 13 and 14, unlabeled 430 bp *Hr*1a-containing fragment.

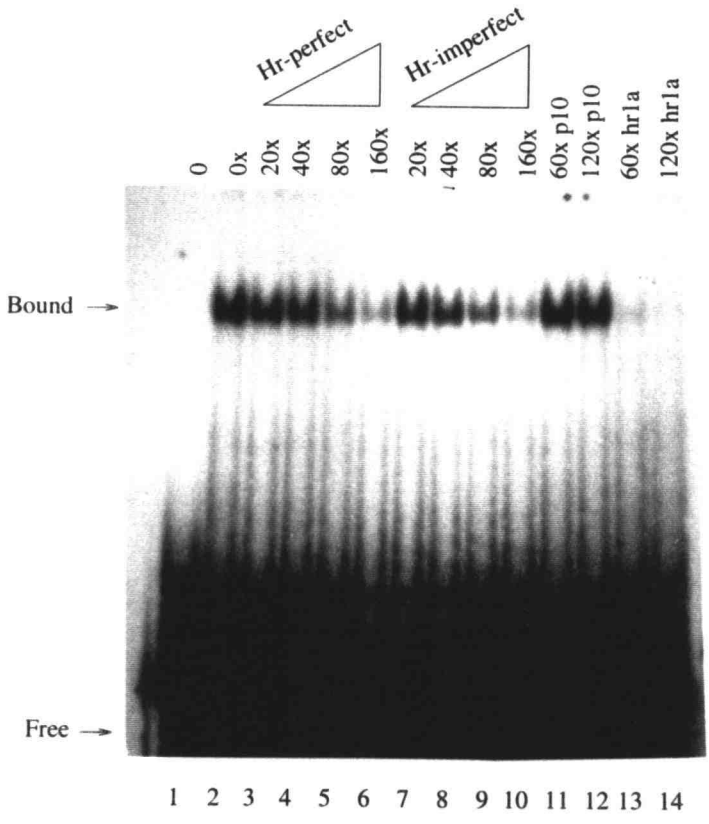


Figure 7.9. Gel Retardation Analysis of *Hr* Perfect Palindrome. The radiolabeled oligonucleotides containing the perfect palindrome were incubated in the presence of 12.5 μg whole-cell extract prepared from pAcIE-1 transfected Sf9 cells. Both radiolabeled probe and competitor DNA were added simultaneously to the reaction mixtures. Competitor DNA was added in excess molar amounts as indicated at the top of each lane: lanes 3–6, unlabeled imperfect palindrome oligonucleotides; lanes 7–10, unlabeled perfect palindrome oligonucleotides; lanes 11 and 12, unlabeled 297 bp *Xho*1–*Xho*1 AcMNPV *p10* promoter fragment; lanes 13 and 14, unlabeled 430 bp *Hr*1a–containing fragment.

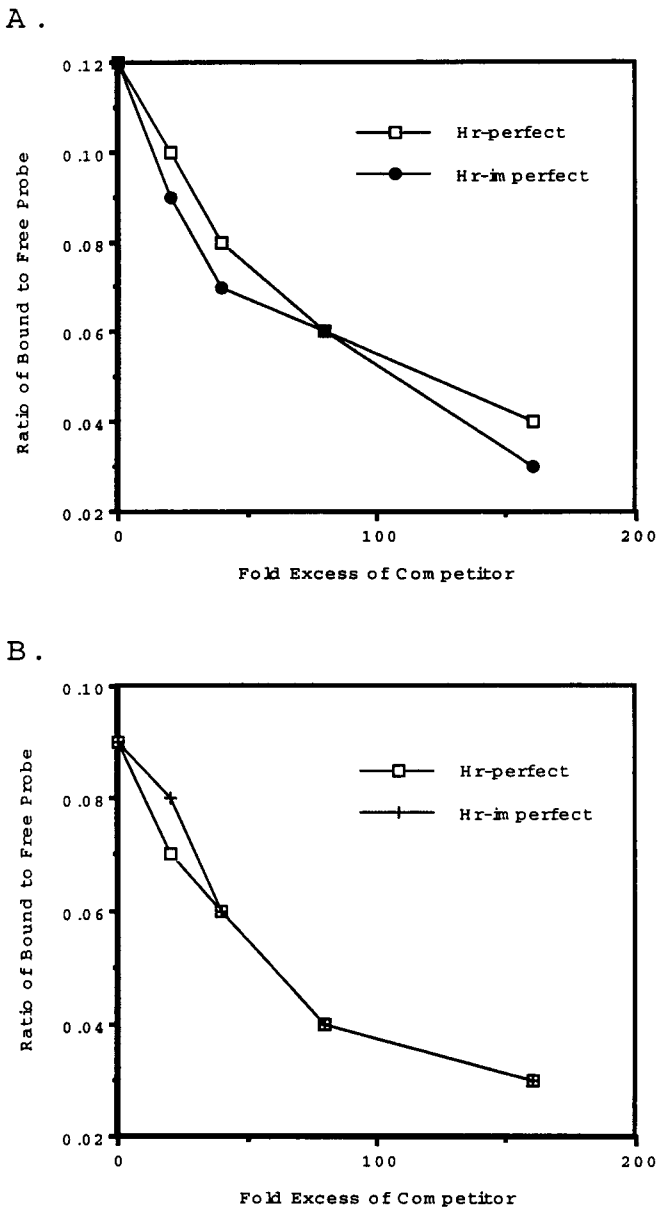


Figure 7.10. Comparison of Competitor Efficiencies. DNA binding assays shown in Figure 7.8 and Figure 7.9 were analyzed using PSI–486 Phosphoimager SI and Imagequant workstation (Molecular Dynamics). A. Competition of IE–1 binding to the radiolabeled Hr–imperfect oligonucleotides. B. Competition of IE–1 binding to the radiolabeled Hr–perfect oligonucleotides.

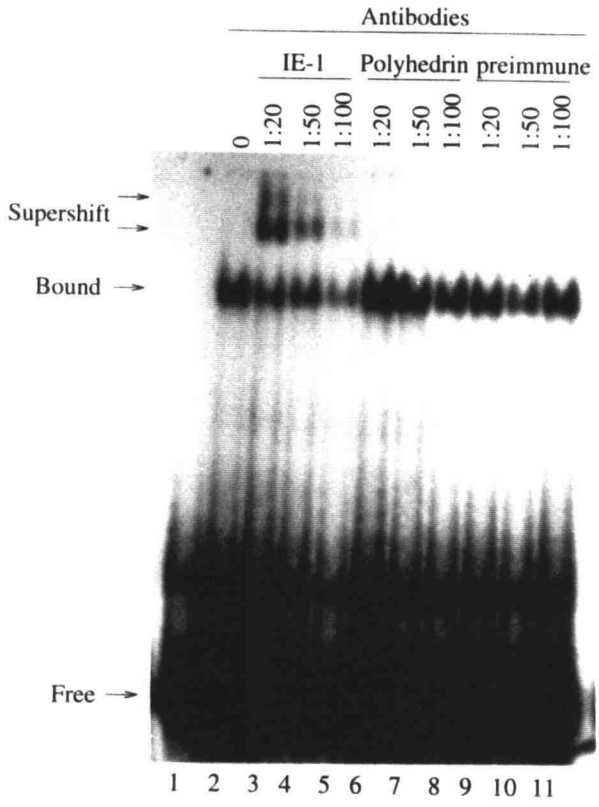


Figure 7.11. Gel Retardation Supershift Analysis of the *Hr* Imperfect Palindrome. The radiolabeled oligonucleotides containing the imperfect palindrome were incubated in the presence of 12.5 μg whole–cell extract prepared from pA-cIE–1 transfected Sf9 cells. After a pre–incubation period of 15 min, dilutions (indicated at the top of each lane) of rabbit polyclonal antibodies to AcMNPV IE–1 (lanes 3–5), preimmune serum (lanes 9–11) or rabbit polyclonal antibodies to OpMNPV polyhedrin (lanes 6–8) were added to the reaction mixture for a further 10 min before complexes were resolved by gel electrophoresis.

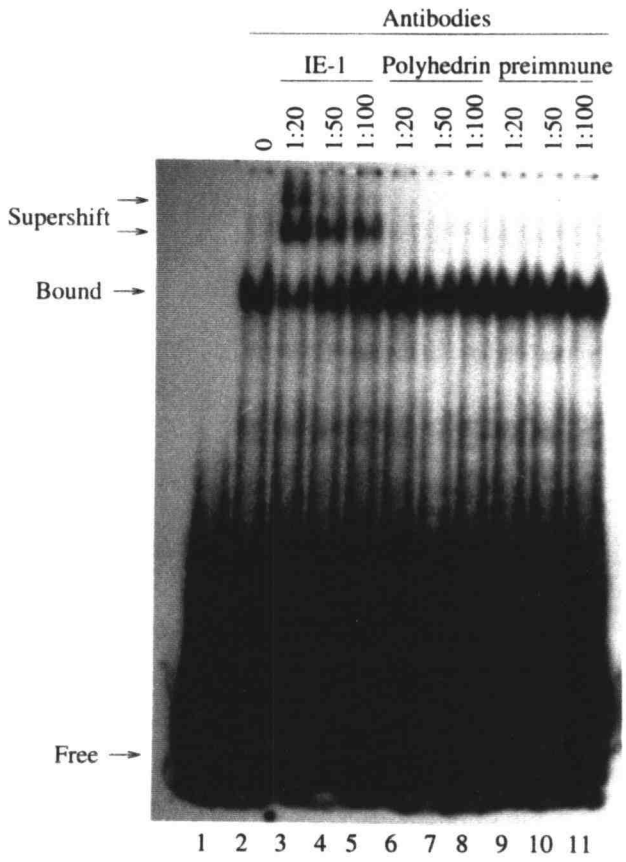


Figure 7.12. Gel Retardation Supershift Analysis of *Hr* Perfect Palindrome. The radiolabeled oligonucleotides containing the perfect palindrome were incubated in the presence of 12.5 μg whole–cell extract prepared from pAcIE–1 transfected Sf9 cells. After a pre–incubation period of 15 min, dilutions (indicated at the top of each lane) of rabbit polyclonal antibodies to AcMNPV IE–1 (lanes 3–5), preimmune serum (lanes 9–11) or rabbit polyclonal antibodies to OpMNPV polyhedrin (lanes 6–8) were added to the reaction mixture for a further 10 min before complexes were resolved by gel electrophoresis.

slowly annealed to promote formation of the cruciform structure (See Figure 7.13 and Figure 7.14, lanes 15–20). Experiments shown in Figures 7.13 and 7.14 using unlabeled slowly annealed single oligonucleotides as competitors indicated that the single oligonucleotides failed to act as competitors whereas the double–stranded Hr–imperfect and Hr–perfect oligonucleotides did. The sequences of the oligonucleotides are as follows:

1. imperfect oligonucleotide A:

   5′–GAACTCGCTTTACGAGTAGAATTCTACTTGTAAAACACAATC–3′,

2. imperfect oligonucleotide B:

   5′–GATTGTGTTTTACAAGTAGAATTCTACTCGTAAAGCGAGTTC–3′,

3. perfect oligonucleotide C:

   5′–GAACTCGTTTTACAAGTAGAATTCTACTTGTAAAACACAATC–3′,

4. perfect oligonucleotide D:

   5′–GATTGTGTTTTACAAGTAGAATTCTACTTGTAAAACGAGTTC–3′,

These results indicate that IE–1 does not bind to a cruciform structure containing mismatched bases.

7.6 Enhancer Activity of Perfect and Imperfect Palindromes in pAcIE–1 Transfected Sf9 Cells

Since *hr*s function as enhancers of early gene expression [28, 185, 218], I was interested in determining if altering the AcMNPV *hr* sequence from an imperfect to a perfect palindrome had an effect on enhancer activity. The oligonucleotides were subcloned into the GUS reporter plasmid p39KGUSpI18, creating pHr–imperfectGUS and pHr–perfectGUS. The enhancer activities of these clones were determined by measuring the GUS activities in extracts of Sf9 cells that had been

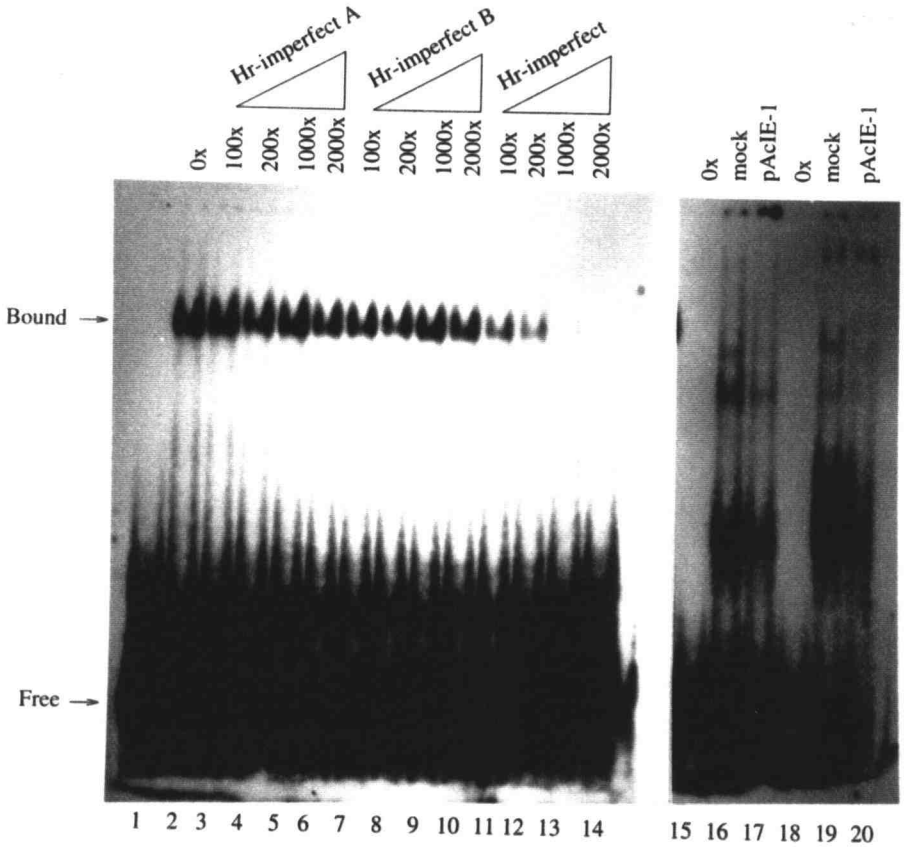


Figure 7.13. Electrophoretic Mobility Shift Analysis using the Hr-imperfect Oligonucleotides Singly as Competitor. The radiolabeled oligonucleotides containing the imperfect palindrome were incubated in the presence of 12.5 μg whole-cell extract prepared from pAcIE-1 transfected Sf9 cells. Both radiolabeled probe and competitor DNA were added simultaneously to the reaction mixtures. Competitor DNA was added in excess molar amounts as indicated at the top of each lane: lanes 3–6, unlabeled imperfect palindrome oligonucleotide A; lanes 7–10, unlabeled imperfect palindrome oligonucleotide B; lanes 11–14, unlabeled annealed imperfect palindrome oligonucleotides. Gel retardation analysis using oligonucleotide A (lanes 15–17) and B (lanes 18–20) was also performed. 12.5 μg of mock (lane 16 and 19) or pAcIE-1 (lane 17 and 20) transfected extracts was used.

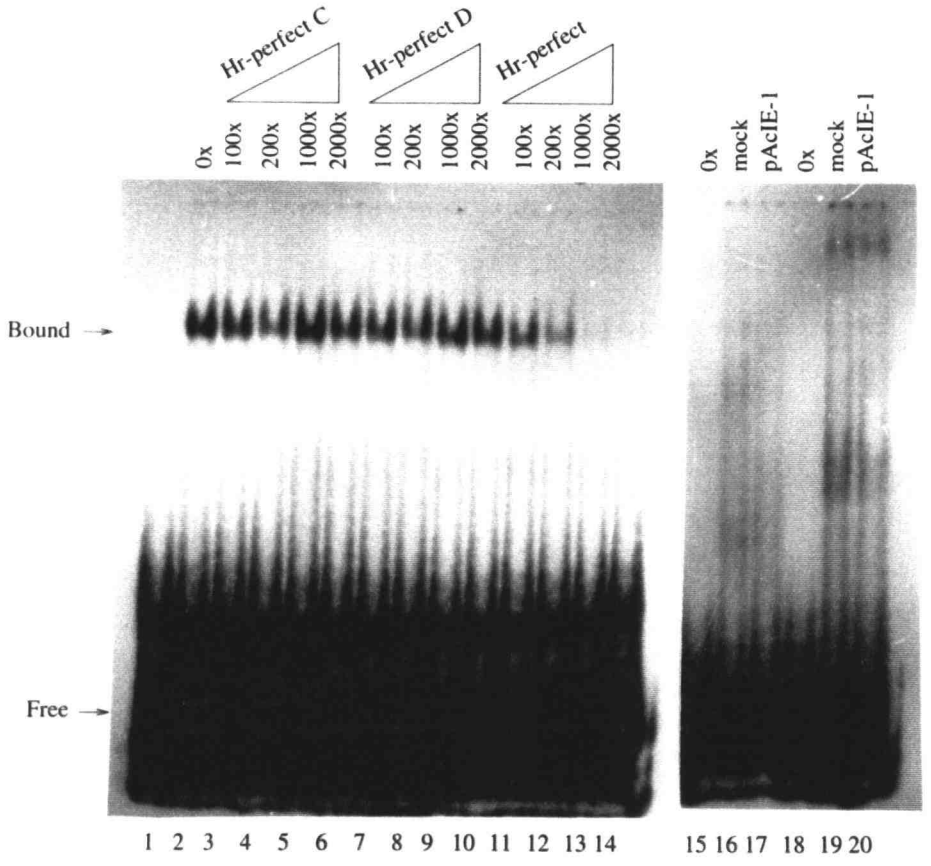


Figure 7.14. Electrophoretic Mobility Shift Analysis using the Hr–perfect Oligonucleotides Singly as Competitors. The radiolabeled oligonucleotides containing the perfect palindrome were incubated in the presence of 12.5 μg whole–cell extract prepared from pAcIE–1 transfected Sf9 cells. Both radiolabeled probe and competitor DNA were added simultaneously to the reaction mixtures. Competitor DNA was added in excess molar amounts as indicated at the top of each lane: lanes 3–6, unlabeled perfect palindrome oligonucleotide C; lanes 7–10, unlabeled perfect palindrome oligonucleotide D; lanes 11–14, unlabeled annealed imperfect palindrome oligonucleotides. Gel retardation analysis using oligonucleotide C (lanes 15–17) and D (lanes 18–20) was also performed. 12.5 μg of mock (lane 16 and 19) or pAcIE–1 (lane 17 and 20) transfected extracts was used.

co–transfected with these GUS reporter constructs and increasing concentrations of pAcIE–1.

The concentration of pAcIE–1 co–transfected with the GUS reporter constructs was varied between 0 and 5 μg per 1.25 $\times 10^6$ cells to determine which pAcIE–1 concentration yielded the highest levels of activation when compared to the same concentration of GUS reporter constructs. As previously shown, a concentration of 0.01 μg of pAcIE–1 per 1.25 $\times 10^6$ cells exhibited maximum enhancer effect [169]. As shown in Figure 7.15, the magnitude of the enhancer effect was diminished at both higher and lower concentrations of pAcIE–1. In the absence of pAcIE–1, no discernible difference in GUS activity was observed between p39KGUSpI18, pHr–perfectGUS, pHr–imperfectGUS and pHr1aΔEcoGUS.

Comparison of the transactivation ability of the three reporter gene constructs were performed using 0.1 μg of pAcIE–1 per 1.25 $\times 10^6$. This concentration of pAcIE–1 was chosen because assays performed at this concentration were highly reproducible. As shown in the Figure 7.16, the complete *hr*1a sequence enhances transcription 3–fold when compared to the single palindrome construct, pHr1aΔEco. The GUS reporter constructs containing the perfect and imperfect palindromes enhance transcription as well as the pHr1aΔEco construct. GUS activity for these constructs was 30–fold greater than p39KGUSpI18. In at least 3 separate assays, no differences in the levels of GUS activity between pHr1aΔEco, pHr–imperfectGUS and pHr–perfectGUS were observed. These data indicate that changes in the base composition at these positions within the *hr* sequence do not effect the ability of the *hr* to enhance transcription.

7.7 Summary

The results of the nuclease P1 assay and two–dimensional electrophoresis of the pHr–imperfect topoisomers showed that the AcMNPV consensus *hr* palindrome does not form a cruciform *in vitro*. However, the same assays performed using the

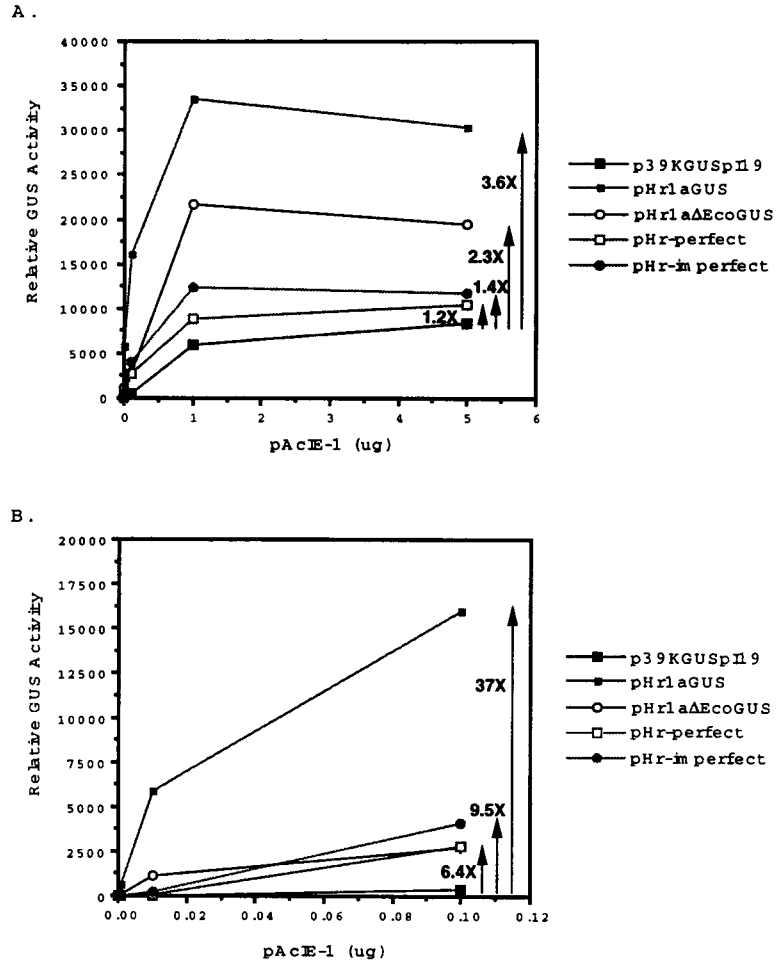


Figure 7.15. Enhancer Activity as a Function of pAcIE–1 Concentration. 0.5 μg of p39KGUSpI18, pHr1aGUS, pHR1aΔEcoGUS, pHr–perfectGUS and pHr–imperfectGUS were co–transfected into 1.25 $\times 10^6$ Sf9 cells in the presence of varying concentrations of pAcIE–1. GUS activity was measured at 24 hr post–transfection. A. The influence of pAcIE–1 concentration on GUS activity is plotted. The ratios of GUS activities of pHr1aGUS, pHr1aΔEcoGUS, pHr–perfectGUS and pHr–imperfectGUS relative to p39KGUSpI18 at 5 μg is shown. B. Same as A, except that only the lower concentrations of pAcIE–1 were plotted. The ratios of GUS activities of pHr1aGUS, pHr1aΔEcoGUS, pHr–perfectGUS and pHr–imperfectGUS relative to p39KGUSpI18 at 0.1 μg is shown.

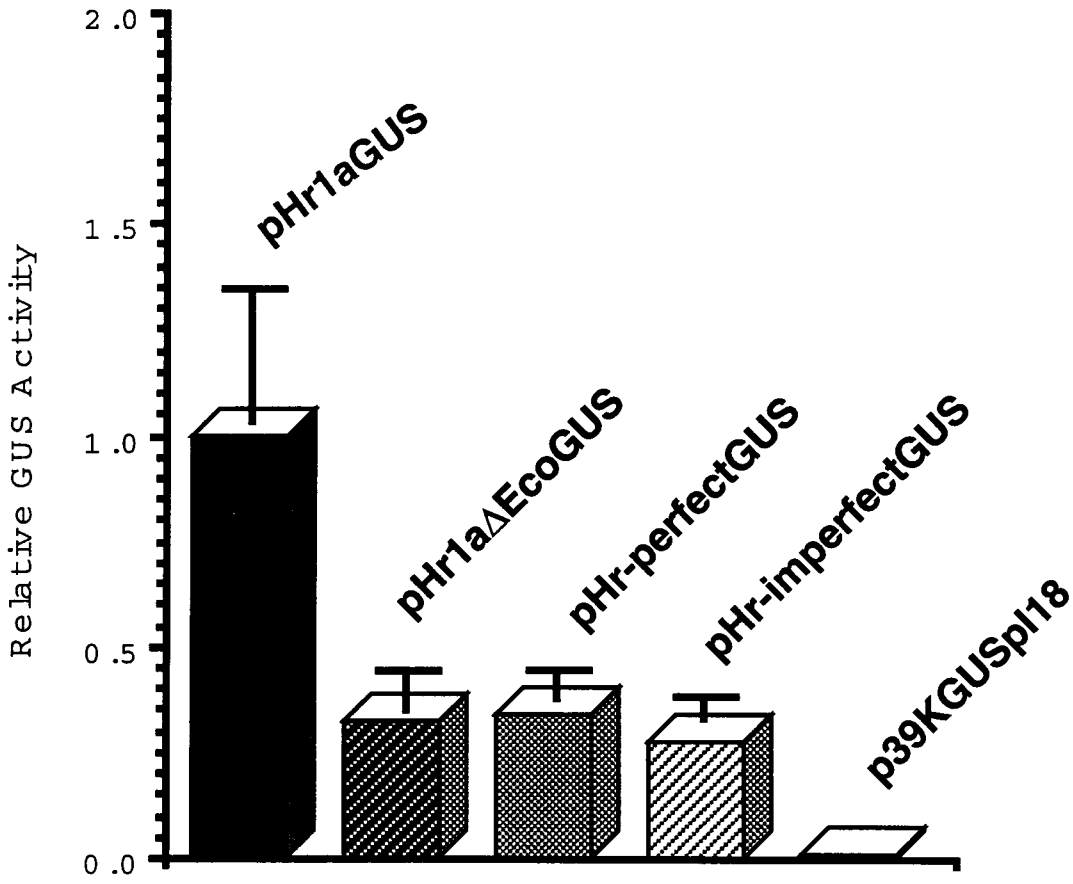


Figure 7.16. Enhancer Activity of the Hr–perfect and Hr–imperfect derivatives. 1.25 $\times 10^6$ Sf9 cells were co–transfected with 0.5 μg of the indicated plasmids and 0.1 μg of pAcIE–1. The ratios of GUS activities relative to pHr1aGUS are shown. Each column represents the average of three independent transfections with the standard deviation indicated.

pHr–perfect and pHr1aΔEco plasmids indicated that these constructs contained an altered DNA structure, centered at the *Eco*R1 site within the palindromic sequence. Characterization of the pHr–perfect cruciform using 2–D gel electrophoresis indicated that cruciform required 9.6 kcal/mole to form and extruded 30 bp. These findings are consistent with a the formation of a cruciform structure containing the altered AcMNPV *hr* sequence. The construct, pHr1aΔEco, required 19.6 kcal/mole to extrude 63 bp indicating that either a larger DNA secondary element is extruded from the plasmid or that the AT–rich flanking sequences allow additional melting of the DNA.

The results of the gel retardation assays indicate that pAcIE–1 transfected cells contain factors that bind specifically to the 42 bp perfect and imperfect oligonucleotides. Supershift analysis using polyclonal antibodies directed against IE–1 show that IE–1 binds the imperfect and perfect oligonucleotides. This is the smallest DNA sequence reported to date to which IE–1 has been shown to bind. The nucleotide differences between the perfect and imperfect palindromes occur in the variable region of the putative IE–1 binding sequence. The putative IE–1 binding site consensus sequence on one side of the palindrome changed from 5′–ACTCGTAA–3′ to 5′–ACTTGTAA–3′ in the perfect oligonucleotides. The putative IE–1 binding site consensus sequence on the other side of the palindrome remained unchanged. Competition experiments performed with annealed single–stranded perfect and imperfect oligonucleotides indicated that cruciform formation does not appear to have a direct role in IE–1 binding.

Comparison of the transactivation ability of the perfect and imperfect palindromic sequences using GUS reporter gene constructs showed that the perfect and imperfect palindromes functioned as enhancers of transcription equally well and as well as the pHr1aΔEcoGUS construct. These results, taken together with the gel retardation data, indicate the 42 bp encompassing the AcMNPV *hr* central palindrome is all that is required for enhancer function.

Chapter 8
Conclusions

Recent studies have shown that a number of viruses encode multifunctional cis– and trans–acting elements (for review see [37, 153, 57]) involved in both transcription and replication. Previous studies and this work indicate that baculoviruses also contain cis– and trans–acting elements that function in transcription and replication. The *ie–1* gene product transactivates RNA polymerase II–mediated baculovirus gene expression [83] and this expression is enhanced by the presence of cis–linked *hr* sequences [169, 259, 218]. This work and that of others has shown that IE–1 binds specifically to *hr*s suggesting that *hr* enhancer function is mediated through IE–1 binding activity. IE–1 is also involved in baculovirus DNA replication. In transient replication assays, the *ie–1* gene along with *dnapol*, *helicase*, *lef–1*, *lef–2*, and *lef–3* are essential for replication of *hr*–containing plasmids [148]. It is likely that IE–1 functions directly and indirectly in baculovirus replication. IE–1 may regulate the levels of expression of genes involved in DNA replication process and since *hr*s appear to function as origins of replication, IE–1 may be an origin–binding protein and serve to form the nucleus for assembly of the replication complex.

8.1 Cis– and Trans–acting Sequences Involved in Baculovirus Transcription

8.1.1 IE–1 Binding Studies

The results of the gel retardation assays indicate that pAcIE–1 transfected Sf9 cells contain factors that bind specifically to *hr*1a. Supershift assays using polyclonal antibodies directed against IE–1 show that IE–1 either binds directly, or as part of a complex that includes host factors. When a DNA fragment contain-

ing the complete *hr*1a sequence was used as a probe, three distinct complexes were detected at high cell extract concentrations, whereas a mutant, pHr1aΔEco, containing a single palindrome, showed only a single retarded complex. These results are similar to those reported for *hr*5 fragments containing either one or two complete palindromes [97, 259]. Guarino and Dong [97] suggested that the formation of complex I was due to the interaction of proteins with one of the two palindromes, and the formation of complex II was due to the interaction of proteins with both palindromes simultaneously. Complex III was interpreted as being formed from the interaction of proteins with an additional half palindrome that was present on their probe. Since the *hr*1a probe contains only two complete palindromes without an additional half palindrome, we suggest that complex III may be due to cooperative interactions between the factors bound at both palindromes that promote the binding of additional IE–1 and/or other factors. Alternatively, there may be additional binding sites for IE–1 (or the IE–1 containing complex) in the intervening DNA between the two palindromes. Analysis of this sequence, however, did not reveal any obvious sequence similarities between the intervening region and the palindromic sequences.

Construction of a number of derivatives of *hr*1a that included a single palindrome, mutations within the central *Eco*RI site, 5′ and 3′ deletions of the single palindrome, and palindrome half sites indicated that IE–1 bound to sequences within a palindrome half site. Using a Gibbs sampling strategy for multiple sequence alignments [162], a putative IE–1 binding motif, centered around a 5′–ACTCGTAA–3′ core sequence contained within the conserved imperfect *hr* palindrome, was determined. Information analysis of the *hr*s indicated that the TC dinucleotide (nucleotides 3 and 4) has some variability (See Figure 6.18). Therefore, the putative IE–1 binding motif is 5′–ACBYGTAA–3′ .

As stated earlier, IE–1 may function to regulate the levels of expression of genes involved in DNA replication. The *ie–1* gene product has been shown to

modulate the level of transcription from other baculovirus gene promoters in the absence [101, 103, 218, 156, 30] of *hr* elements. Analysis of the promoter regions of the baculovirus genes involved in DNA replication indicated that the *lef–1*, *ie–2* and *pe38* promoters contained the putative IE–1 binding motif, 5′–ACTCGTAA–3′. This sequence was also found upstream of the *ie–0* gene, the spliced form of *ie–1* that has been shown to be down–regulated by the *ie–1* gene product [156]. Furthermore, the putative IE–1 binding motif is located 3′ to the TATA box and either comprises or is a few bases upstream of the site of transcription initiation. This positioning of the putative IE–1 binding motif may be important for IE–1 transcriptional repression (see below). The *ie–2*, *pe38*, *dnapol* and *helicase* promoters were assayed for IE–1 DNA binding activity. IE–1 specific binding was observed when the *ie–2* and *pe38* promoters but not when the *dnapol* and *helicase* promoters were assayed indicating that the IE–1 is binding to the 5′–ACTCGTAA–3′ motif.

To determine if the putative IE–1 binding site does in fact bind IE–1 and that IE–1 binding is important for transcriptional regulation of the replication gene promoters, mutation of the putative IE–1 binding site within one of these promoters should be performed. This altered promoter should then be assayed for IE–1 binding activity and gene expression using GUS reporters constructs. Failure of IE–1 to bind to the mutated replication gene promoter and changes in measurable GUS activity, would indicate that IE–1 binds to this sequence and IE–1 binding is important for gene expression. Further studies including mutational analysis of the nucleotides within the IE–1 binding site should be conducted, together with footprinting experiments would to determine the nucleotides important for IE–1 binding. Footprinting experiments could be attempted using pAcIE–1 transfected whole–cell extracts, but may require purification of IE–1.

When performing gel retardation assays with mock–transfected Sf9 whole–cell extracts, retarded complexes were frequently detected suggesting that host

factors may also bind to *hr*s. Gel retardation studies showed that Sf9 cell factors bound specifically to the region between the *hr* central palindrome, referred to as the interregion. Consistent with this observation, is the finding that reporter gene constructs containing *hr* sequences transfected into Sf9 cells are transactivated even in the absence of *ie–1* [30, 218, 185, 169]. The question of what these host factors are and if they are important for *hr* function could be approached experimentally as outlined in section 8.1.3.

8.1.2 IE–1 Binding and Transcriptional Regulation

A number of the *hr*1a constructs examined for IE–1 binding activity were also examined for the ability to transactivate baculovirus gene expression. DNA fragments containing the entire *hr*1a region, the single palindrome, central core mutations, half sites and the interregion were subcloned into a GUS reporter plasmid. The enhancer activities of these clones were determined by measuring the GUS activities in extracts of Sf9 cells that had been co–transfected with these GUS reporter constructs and increasing concentrations of pAcIE–1. The entire *hr*1a region containing two palindromes greatly enhanced transcriptional activity; GUS activity is at least 10–fold higher than constructs containing just a single palindrome. The single palindrome and the 42–mer imperfect and perfect single palindrome constructs all enhance expression transcription when compared to the non–*hr* containing GUS reporter plasmid. However, mutant *hr*1a sequences in which small deletions or insertions were introduced into the central *Eco*RI core region and constructs containing half–sites showed either greatly diminished or no enhancer functions, although DNA probes containing core mutations were capable of binding IE–1 [169]. This suggests that IE–1 binding by itself is not sufficient for enhancer function. These mutations may alter sequences required for other factors to bind to *hr*1a, disrupt essential DNA conformations such as hairpin structures, or alter the stereospecific conformation of the IE–1

dimer that may be critical for interaction with other proteins.

The maximum enhancer effect of cis–linked *hr*–regions occurred when relatively low levels of co–transfected pAcIE–1 plasmid DNA were used. This suggests that the *hr* elements specifically bind IE–1, facilitating the interaction of IE–1 with the plasmid DNA, where it functions to stimulate transcription. At high concentrations, IE–1 may interact non–specifically with the DNA allowing near maximal levels of transcriptional stimulation in the absence of a specific interaction with an *hr* sequence. The relatively high levels of stimulation of early genes linked in cis to *hr* enhancer sequences by limited concentrations of IE–1 may reflect the conditions present at the beginning of a baculovirus infection. The successful initiation of AcMNPV infection may be dependent upon the ability of low levels of IE–1 to highly activate transcription of single copies of early genes linked in cis to *hr* enhancer sequences in the baculovirus genome.

8.1.3 Model for IE–1 Regulation of Baculovirus Gene Expression

Studies on multifunctional regulatory proteins from other DNA viruses may provide clues as to the mechanisms of baculovirus transcriptional regulation by IE–1. Infected–Cell polypeptide 4 (ICP4) from herpes simplex virus type 1 activates the transcription of some HSV genes [56, 221] and represses the transcription of others [56, 222, 258]. Studies, using reconstituted *in vitro* transcription systems, show that ICP4 activates transcription in a complex manner that involved DNA binding, interactions with general transcription factors including TBP, TFIIB and TBP–associated factors (TAFs) and possibly proteins that bind to the start site of transcription [93]. Similar studies using ICP4 repressable promoters indicated that ICP4 inhibited the stimulatory effect of the transactivators Sp1, VP16, USF, and ICP4 itself without affecting basal transcription [95, 257, 94]. Moreover, ICP4 inhibited transcription by binding to a specific sequence near the site of transcription initiation [211, 95] and thereby interfering with forma-

tion of transcription initiation complexes [94]. Smith *et al.* [283] suggested a model of how ICP4 regulates transcription that involves both protein–protein and protein–DNA interactions. ICP4 forms a tripartite complex with TBP and TFIIB and the resulting interactions stabilize the formation of the preinitiation complex. Formation of the complex is a rate–limiting step in the formation of the transcription preinitiation complex, and ICP4 stimulates formation at least five–fold [283]. Thus, one mechanism by which ICP4 enhances viral gene expression may be by recruiting and stabilizing the formation of the preinitiation complex. However, mutational analysis of ICP4 suggests that ICP4 may have additional interactions with other proteins that also result in an increase in gene expression [283]. At certain HSV promoters, ICP4 represses activated but not basal levels of transcription [95, 94, 257]. Gu *et al.* [94] showed that repression is mediated by ICP4 binding to DNA at or near the site of transcription initiation suggesting that ICP4 interacts with the preinitiation complex in a manner that inhibits activated transcription.

A number of studies have reported similar interactions between other viral transactivators and the transcription preinitiation complex. For instance, the human cytomegalovirus IE protein, the HSV VP16 early transactivator protein, the E1a protein of Adenovirus and Zta from Epstein–Barr virus interact directly with TBP [106, 290, 127, 180, 167]. Recent studies indicate that the synergistic action of the viral transactivators, Zta and VP16, correlates with the assembly of the TFIID:TFIIA (DA) complex and the ability of the DA–complex to bind TFIIB. The binding of TFIIB appears to be activator–dependent and once bound, TFIIB enhances the transcription by increasing complex stability [39]. These results imply a similar regulatory pathway for a number of viral systems. Moreover, viral transactivators may function in a similar manner in heterologous systems. In fact, Adenovirus E1A mutants defective in early gene expression can be complemented by the pseudorabies IE gene [66]. Substitution

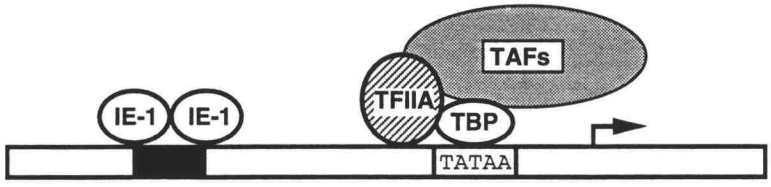
of the baculovirus *ie–1* gene with viral transactivators from other systems may be of interest in answering the following questions. Can heterologous viral transactivators function in the baculovirus system? Is baculovirus gene expression enhanced or reduced using heterologous transactivators? Could these experiments reveal functional domains of viral transactivators and be used to engineer a more potent AcMNPV IE–1 ?

Using the AcMNPV IE–1 DNA binding and transcriptional data obtained during the course of this work and the published data concerning AcMNPV IE–1 and other viral transactivators, a model for IE–1 transcriptional regulation of baculovirus promoters can be proposed. This model is diagrammatically shown in Figure 8.1. Binding of the AcMNPV IE–1 dimer to a single *hr* element results in a stimulation of transcription as measured using GUS reporter constructs. However, two copies of the *hr* palindrome elevate the levels of GUS expression 10–fold when compared to the single palindrome suggesting synergistic interactions. Binding of AcMNPV IE–1 dimers to several *hr*s may stimulate formation of the DA complex by specific interactions with multiple members of the preinitiation complex, including TBP and TAFs resulting in increased gene expression. A single IE–1 dimer may not be able to promote as efficient DA complex formation or TFIIB recruitment. Mutations in the *hr* palindromes that alter the spacing between the IE–1 binding sites may alter the conformation of the IE–1 DNA complex preventing productive interactions with the preinitiation complex. Finally, IE–1 mediated repression may be a function of IE–1 binding at or near the site of transcription initiation resulting in altered interactions with the preinitiation complex that leads to the repression of activated transcription.

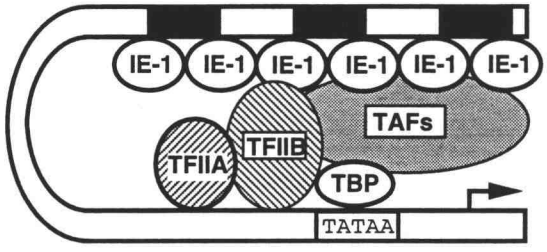
Future work to examine the mechanisms of IE–1 transcriptional regulation should prove interesting. Initially the experiments described above (section 8.1.1) to confirm the IE–1 binding motif, should be performed. In addition, the GUS activity assays can be used to study the effect of IE–1 on the expression

Activation:

-single *hr*



-multiple *hrs*



Repression:

-*hr* half site

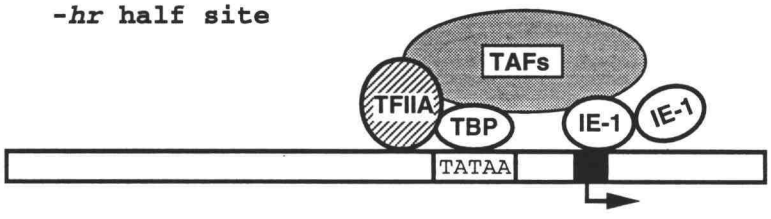


Figure 8.1. Schematic Representation of a Model for IE–1 Regulation of Gene Expression. General transcription factors involved in the formation of the preinitiation complex are indicated. The filled–in boxes represent the *hr*s or a portion of the *hr*. The arrowheads indicate the site of transcription initiation.

of replication gene promoters with and without the IE–1 binding motif located near the start site of transcription. These reporter constructs could also be assayed with and without *hr* enhancer sequences to determine if the position of putative IE–1 binding motif near the start site of transcription represses activated but not basal transcription levels.

The *in vitro* transcription system can be used to observe the effects of addition of AcMNPV IE–1 on baculovirus gene expression. Using different DNA templates, the role of IE–1 in the regulation of different genes can be examined. To determine if IE–1 addition promotes the formation of transcription initiation complexes, DNA templates and nuclear extracts from uninfected Sf9 cells can be incubated for different lengths of time in the presence or absence of IE–1 before the addition of the detergent, sarkosyl. Low concentrations of sarkosyl allows elongation of a committed transcription complex but inhibits formation of initiation complexes [112]. The quantity of primer extension products is a measurement of the number of transcription complexes formed before addition of sarkosyl and therefore to can be used to compare the rate of initiation complex formation in the presence and absence of IE–1.

These experiments would require purified AcMNPV IE–1. AcMNPV IE–1 could be purified as a fusion protein provided the fusion protein is functional in replication assays, gel retardation assays and GUS activity assays. However, although histidine–tagged IE–1 functioned in transient replication assays (Jay Evans, personal communication), and as a transactivator in GUS activity assays (data not shown), whole–cell extracts transfected with the histidine–tagged IE–1 failed to show reproducible DNA binding activity (data not shown). Therefore, IE–1 may have to be purified from transfected or infected Sf9 cells. Purification of IE–1 from infected or transfected Sf9 cells may not be too difficult as IE–1 polyclonal antibodies [77] and DNA oligonucleotides [135] containing the IE–1 binding motif could be used in affinity chromatography to purify IE–1. However,

any mutational analysis of IE–1 would require purification of the mutant IE–1 protein for subsequent functional assays, limiting the purification protocol to procedures that do not utilize IE–1 functions that may be altered by introduced mutations. Hence, overexpression and purification of the IE–1 protein from *E. coli* may provide a viable option. Alternatively, small quantities of IE–1 can be made using an *in vitro* transcription/translation system.

To examine the effects of IE–1 on the assembly of the transcription preinitiation complex, experiments similar to those performed with ICP4 and Zta viral transactivators can be performed [283, 179]. The general transcription initiation factors can be fractionated from nuclear extracts following well–established protocols [254, 270] and since SfTBP was cloned during the course of this work, recombinant SfTBP can be made. Gel retardation assays and DNAse I footprinting can be used to examine the protein–protein–DNA interactions between various general transcription factors and IE–1 during reconstitution experiments using various baculovirus promoter templates including those containing *hr* elements. Moreover, once this system is established, it can also be used to examine the effect the addition of IE–1 has on transcription initiation complex formation on pre–assembled nucleosome templates [327, 328].

The *ie–1* gene was identified as a late expression factor using an assay that did not differentiate between genes required for DNA replication and transcription of baculovirus late genes. It is possible that IE–1 regulates the levels of expression of genes involved in DNA replication and also functions as the origin–binding protein. Immunoprecipitation of IE–1 from nuclear extracts prepared from AcMNPV–infected Sf9 may allow for the isolation of IE–1 associated proteins. This experiment may result in the isolation of a late transcription complex or components of the baculovirus origin recognition complex that may include host factors.

Several of the molecular tools necessary for this work have been developed

during the course of my work. The gene encoding the *S. frugiperda* TATA–binding protein was cloned and sequenced and its the expression monitored during the course of AcMNPV infection. This clone can be used to make recombinant SfTBP to study IE–1 and TBP interactions. The *in vitro* transcription systems developed using uninfected and AcMNPV–infected nuclear extracts from Sf9 cells can be used in studies to look at IE–1 function during baculovirus gene expression. In addition, a gel–mobility shift assay for IE–1 DNA binding was developed and can be used in further investigations.

8.2 Transcription Factors and DNA Replication

The same cis– and trans–acting sequences that stimulate baculovirus early gene expression are also required for baculovirus DNA replication. *Hr*s act both as enhancers of transcription and as origins of replication. The *ie–1* gene is a trans–acting factor that is essential *hr*–dependent replication [147, 148] and also functions as a transactivator of early gene expression [101, 103, 218]. The ability of IE–1 to transactivate baculovirus early gene expression is augmented by the presence of *hr*s cis–linked to the early gene promoters [102, 98, 218, 28, 185, 30]. The involvement of transcription factors in DNA replication has been observed in a number of systems [240, 175, 57]. Since the processes of replication and transcription share many similarities including the formation of a multicomponent nucleoprotein complex in a nucleosomal environment, DNA unwinding, RNA synthesis, and DNA–template dependent polymerization of similar substrates in a 5′ to 3′ direction, it is not surprising that replication and transcription share common factors. Furthermore, coordinate regulation of these two processes may be achieved by the use of common factors.

Studies on transcription factor involvement in the replication in other DNA virus systems may provide clues to possible roles for AcMNPV IE–1 in baculovirus replication. Adenovirus utilizes several cellular transcription factors

that stimulate replication via two different mechanisms. The transcription factor, NFI, stimulates replication by positioning and stabilizing the binding of adenovirus–encoded DNA polymerase to the origin of replication [210]. Oct–1 binding to the adenovirus origin results in bending of DNA template which is thought to stimulate replication by facilitating protein–protein interactions and/or inducing structural distortions in the DNA template that facilitate initiation [210, 308]. The papovavirus SV40 T antigen is a multifunctional protein that binds to SV40 origins of replication [196], has DNA helicase activity [287] and stimulates the expression of viral genes [338]. Auxiliary cellular transcription factors appear to stimulate SV40 replication by alleviating repression caused by nucleosomes at the site of replication initiation [37]. The E2 protein of bovine papilloma virus also appears to function by alleviating nucleosomal repression possibly by recruiting cellular single–stranded DNA binding proteins to the origin [173, 175, 174]. E2 also interacts with E1, the origin recognition factor, to stabilize E1 binding to the origin DNA [72] and the E1/E2 complex may direct binding of the DNA polymerase–primase complex to the origin [240].

Unlike the viral origins of replication described above where transcriptional enhancer sequences are located adjacent to the origin, the Epstein–Barr virus (EBV) origin of replication, *oriP*, contains two–cis acting elements that are separated by nearly a kb of DNA [78, 255]. A 30–bp family of repeats that functions as a transcriptional enhancer and is transactivated by the the viral EBNA–1 gene product [255] is an essential component of the EBV DNA replication origin. A dyad symmetry element that contains 4 EBNA–1 binding sites is located approximately 900 bp away from the family of repeats is thought to function as a DNA unwinding element (DUE) and site of replication initiation [78, 322].

From studies utilizing other viral systems, it is clear that the AcMNPV *ie–1* gene product may have multiple functions in both transactivation and DNA replication. In addition to functioning as a viral transactivator, IE–1 may be

the origin recognition protein and thus facilitate interactions with other replication proteins. The purification of factors involved baculovirus replication and subsequent reconstitution of an *in vitro* replication system, should contribute to understanding the role of IE–1 in baculovirus replication.

8.3 Role of *Hr* Palindromes in Baculovirus Transcription and Replication

Regions of DNA to which proteins bind can contribute specific sequences as well as structural elements for the assembly of proteins required to carry out essential functions in replication and transcription. DNA structure can play a role in regulation. For instance, high–affinity binding of the cAMP response element–binding protein requires a cruciform structure containing mismatched base pairs suggesting that cruciform formation is important for transcriptional activation [285]. 2-D gel electrophoresis of constructs containing the perfect and imperfect *hr* palindromes indicated that the perfect palindrome is likely to form cruciforms *in vivo*, whereas the imperfect palindrome is unlikely to form this structure without thermodynamic assistance from DNA–binding proteins. Gel retardation assays indicated that the IE–1 did not require mismatched base pairs, nor a cruciform structure for binding to the *hr* palindrome. Furthermore, GUS reporter gene constructs failed to detect any difference in the levels of transactivation between the mismatched and perfect palindromes indicating that sequence–specificity, not DNA structure, was critical for IE–1 function. These data, together with the *hr*1a [169] and *hr*5 mutational analysis [259] suggest that IE–1 binding requires critically spaced half sites within a palindrome, rather than cruciform formation, for enhancer function.

Initiation of replication requires unwinding of the DNA to allow access of the replication complex to DNA template. A general model for replication is that the origin recognition protein recognizes specific sites at a replication origin and

induces localized unwinding by itself (SV40 T antigen [287]) or in combination with other replication factors [240]. DNA unwinding elements, regions of DNA that are intrinsically easier to unwind, can be found near origins of replication in *E. coli* [24], yeast [129] and in viral systems [322, 57, 240].

The Epstein–Barr virus, *oriP*, contains a family of 20 imperfect copies of 30–bp repeats that was predicted to form cruciform structures with a 22–bp stem containing 2 mismatches and a 12–bp loop. The dyad symmetry region of *oriP* can form a single 15–bp stem cruciform containing 1 mismatch with a 15–bp loop. P1 nuclease experiments indicated the region of dyad symmetry was P1 sensitive whereas the repeated family was P1 sensitive when 9 or more repeats were present [322].

I performed similar studies on the AcMNPV *hr*s. A construct containing 42–bp encompassing a single imperfect AcMNPV palindrome was not P1 sensitive. However, the pHr1aΔEco construct which contains a single palindrome (created by fusing the left half of the first *hr*1a palindrome with the right half of the second palindrome) with 55 bp of 5′ and 210 bp of 3′ *hr*1a–derived flanking sequences and the complete *hr*1a construct (data not shown) were P1 sensitive. The pHr1aΔEco flanking sequences are not conserved in the AcMNPV palindromic regions and result in a larger extrusion of DNA than predicted by the AcMNPV *hr* consensus sequence. However, *hr*1a may function in a slightly different manner than the other *hr* regions since *hr*1a contains only two palindromes and has been shown to have an auxiliary sequence (not present in these constructs) to the left of the palindrome that enhances the levels of replication [169]. The auxiliary sequence may have evolved in *hr*1a to enhance replication whereas multiple copies of palindrome in other origins may alleviate the need for auxiliary sequences.

2–D gel electrophoresis of topoisomers and nuclease P1 assays showed that the 42–bp single AcMNPV imperfect palindrome is unlikely to form a cruciform without thermodynamic assistance. Nevertheless, multiple copies of the AcM-

NPV *hrs* may, in a manner similar to the EBV *orip*, function as a DUE. However, AcMNPV *hr* palindromes would form cruciforms with a shorter stem (13–bp) and would contain at least two mismatches. This suggests that AcMNPV *hrs* may be less likely to unwind than EBV *orip* regions. Furthermore, the EBV *orip* family of repeats required at least 9 repeated elements before P1 sensitivity was observed. Since, AcMNPV *hrs* contain only one to eight palindromic elements, the energetics for DNA unwinding may not be favorable without aid from DNA binding proteins. To determine if *hrs* function as specific DNA binding sites and/or DNA unwinding elements, the AcMNPV *hr* regions, particularly those containing the greatest number of palindromes, should be assayed for P1 sensitivity. Furthermore, these experiments can be conducted in the presence and absence of purified AcMNPV IE–1 to observe the effect of IE–1 binding on DNA structure. Finally, electron microscopy could be used to observe the structure of AcMNPV *hr* regions in the presence and absence of purified AcMNPV IE–1.

BIBLIOGRAPHY

[1] S. M. Abmayr, J. L. Workman, and R. G. Roeder. The pseudorabies immediate early protein stimulates in vitro transcription by facilitating TFIID: promoter interactions. *Genes and Development*, 2:542–53, 1988.

[2] C. H. Ahrens and G.F. Rohrmann. Identification of essential trans–acting regions required for DNA replication of the *Orygia psuedostguata* multinucleocapsid nuclear polyhedrosis virus: *lef–1* is an essential replication gene. *Virology*, 207:417–428, 1995.

[3] C.H. Ahrens. *Identifcation of Essential Cis– and Trans–acting Sequences Involved in Baculovirus Replication.* PhD thesis, Oregon State University, 1995.

[4] C.H. Ahrens, C. Carlson, and G.F. Rohrmann. Identification, sequence and transcriptional analysis of *lef–3*, a gene essential for *Orygia pseudotsugata* baculovirus DNA replication. *Virology*, 210:372–382, 1995.

[5] C.H. Ahrens, M.N. Pearson, and G.F. Rohrmann. Identification and charaterization of a second putative origin of DNA replication in a baculovirus of *Orygia pseudotsugata.* *Virology*, 207:572–576, 1995.

[6] V. Apsit, J. A. Freeburg, M. R. Chase, E. A. Davis, and S. Ackerman. Wheat TFIID TATA binding protein. *Nucleic Acids Research*, 21:1494, 1993.

[7] 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*, 202:586–605, 1994.

[8] D.P. Bazett-Jones, B. Leblanc, M. Herfort, and T. Moss. Short–range DNA looping by the *Xenopus* HMG–box trancription factor, xUBF. *Science*, 264, 1994.

[9] A. Berkenstam, M.M. Vivanco Ruiz, D. Barettino, M. Horikoshi, and H.G. Stunnenberg. Cooperativity in transactivation between retinoic acid receptor and TFIID requires an activity analogous to E1A. *Cell*, 69:401–412, 1992.

[10] S.H. Berlocher. Insect molecular systematics. *Annual Review of Entomology*, 29:403–433, 1984.

[11] J.N. Bicknell, D.J. Leisy, G.F. Rohrmann, and G.S. Beaudreau. Comparison of the p26 gene region of two baculoviruses. *Virology*, 161:589–592, 1987.

[12] H.C. Birnboim and J. Doly. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*, 7(1513–1523), 1979.

[13] D.H.L. Bishop. Baculovirus expression vectors. *Seminars in Virology*, 3:253–264, 1992.

[14] R.M. Bjornson, B. Glocker, and G.F. Rohrmann. Characterization of the nucleotide sequence of the *Lymantria dispar* nuclear polyhedrosis virus DNA polymerase gene region. *Journal of General Virology*, 73:3177–3183, 1992.

[15] G.W. Blissard, P.L. Kogan, R.Wei, and G.F. Rohrmann. A synthetic early promoter from a baculovirus: roles of the TATA box and conserved start site CAGT sequence in basal levels of transcription. *Virology*, 190:783–793, 1992.

[16] G.W. Blissard and G.F. Rohrmann. Location, sequence, transcriptional mapping and temporal expression of the gp64 envelope glycoprotein of the *orgyia psuedotsugata* mutlicapsid nuclear polyhedrosis virus. *Virology*, 170:537–555, 1989.

[17] G.W. Blissard and G.F. Rohrmann. Baculovirus diversity and molecular biology. *Annual Review of Entomology*, 35:127–55, 1990.

[18] G.W. Blissard and G.F. Rohrmann. Baculovirus gp64 gene expression: Analysis of sequences modulating early transcription and transactivation by IE1. *Journal of Virology*, 65:5820–5827, 1991.

[19] G.W. Blissard and J.R. Wenz. Baculovirus gp64 envelope protein is sufficient to mediate pH–dependent membrane fusion. *Journal of Virology*, 66:6829–6835, 1992.

[20] J.E. Blume. Genbank accession number m64861. 1994.

[21] H.B. Boudreaux. *Arthropod pylogeny with special reference to insects.* J. Wiley and Sons, Inc., 1979.

[22] R. Bowater, F. Aboul-ela, and D.M.J. Lilley. Two–Dimensional gel electrophoresis of circular DNA topoisomers. *Methods in Enzymology*, 212:105–120, 1992.

[23] M.M. Bradford. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Analytical Biochemistry*, 72:248–258, 1976.

[24] D. Bramhill and A. Kornberg. Duplex opening by dnaA protein at novel sequences of initiation of replication at the origin of the *E. coli* chromosome. *Cell*, 52:743–755, 1988.

[25] S. Buratowski, S. Hahn, L. Guarente, and P.A. Sharp. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell*, 56:549–561, 1989.

[26] S. Burma, B. Mukherjee, A. Jain, S. Habib, and S. E. Hasain. An unusual 30-kda protein binding to the polyhedrin gene promoter of *Autographa californica* nuclear polyhedrosis virus. *Journal of Biological Chemistry*, 269:2750–2757, 1994.

[27] A. Caccone and J.R. Powell. Extreme rates of heterogeneity in insect DNA evolution. *Journal of Molecular Evolution*, 30:273–280, 1990.

[28] D.D. Carson, L.A. Guarino, and M.D. Summers. Functional mapping of an AcNPV immediate-early gene which augments expression of the IE-1 *trans*-activated 39K gene. *Virology*, 162:444–451, 1988.

[29] D.D. Carson, M.D. Summers, and L.A. Guarino. Molecular analysis of a baculovirus regulatory gene. *Virology*, 182:279–286, 1991.

[30] D.D. Carson, M.D. Summers, and L.A. Guarino. Transient expression of *Autographa californica* nuclear polyhedrosis virus immediate-early gene, IE-N is regulated by three viral elements. *Journal of Virology*, 65:945–951, 1991.

[31] E.B. Carstens. Sequence, transcriptional mapping, and overexpression of p47, a baculovirus gene regulating late gene expression. *Journal of Virology*, 67:2513–2520, 1993.

[32] B. Cavallini, J. Huet, J. Plassat, A. Sentenac, J. Egly, and P. Chambon. A yeast activity can substitute for the HeLa cell TATA box factor. *Nature*, 334:77–80, 1988.

[33] H. Chaabihi, M. Ogliastro, M. Martin, C. Giraud, G. Devauchelle, and M. Cerutti. Competition between baculovirus polyhedrin and p10 gene expression during infection of insect cells. *Journal of Virology*, 67:2664–2671, 1993.

[34] D.I. Chasman, K.M. Flaherty, P.A. Sharp, and R.D. Kornberg. Crystal structure of yeast TATA–binding protein and model for interaction with DNA. *Proceedings of the National Academy of Sciences USA*, 90:8174–8178, 1993.

[35] D.D. Chen, M.H. Nesson, G.F. Rohrmann, and G.S. Beaudreau. The genome of the multinucleocapsid baculovirus of *Orygia psuedotsugata*: restriction map and analysis of two sets of GC-rich repeated sequences. *Journal of General Virology*, 69:1375–1381, 1988.

[36] J. Chen, L.D. Attardi, C.P. Verrijzer, K. Yokomori, and R. Tjian. Assembly of recombinant TFIID reveals differential coactivator requirements for distinct transcriptional activators. *Cell*, 79:93–105, 1994.

[37] L. Cheng, J.L. Workman, R. Kingston, and T.J. Kelly. Regulation of DNA replication *in vitro* by the transcriptional activation domain of GAL4–VP16. *Proceedings of the National Academy of Sciences USA*, 83:589–593, 1992.

[38] L. Cherbas and P. Cherbas. The arthropod initiator: the capsite consensus plays an important role in transcription. *Insect Biochemistry and Molecular Biology*, 23:81–90, 1993.

[39] T. Chi, P. Lieberman, K. Ellwood, and M. Carey. A general mechanism for transcriptional synergy by eukaryotic activators. *Nature*, 377:254–257, 1995.

[40] J.M. Chirgwin, A.E. Przybyla, R.J. MacDonald, and W.J. Ritter. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry*, 18:5294–5299, 1979.

[41] G.E. Chisholm and D.J. Henner. Multiple early transcripts and splicing of the *Autographa californica* nuclear polyhedrosis virus IE-1 gene. *Journal of Virology*, 62:3193–3200, 1988.

[42] J. Choi and L.A. Guarino. Expression of the IE1 transactivator of *Autographa californica* nuclear polyhedrosis virus. *Virology*, 209:99–107, 1995.

[43] B. Choy and M.R. Green. Eukaryotic activators function during multiple steps of the preinitiation complex assembly. *Nature*, 366:531–536, 1993.

[44] G.M. Church and W. Gilbert. Genomic sequencing. *Proceedings of the National Academy of Sciences USA*, 81:1991–1995, 1984.

[45] M.E. Clark, P.L. Liebermann, A. J. Berk, and A. Dasgupta. Direct cleavage of human TATA–binding protein by poliovirus protease 3C in vivo and in vitro. *Molecular and Cellular Biology*, 13:1232–1237, 1993.

[46] R.J. Clem, M. Fechheimer, and L.K. Miller. Prevention of apoptosis by a baculovirus gene during infection of insect cells. *Science*, 254:1388–1390, 1991.

[47] M.A. Cochran and P. Faulkner. Location of homologous DNA sequences interspersed at five regions in the baculovirus AcMNPV genome. *Journal of Virology*, 45:961–970, 1982.

[48] L. Comai, N. Tanese, and R. Tijan. The TATA–binding protein and associated factors are integral components of the RNA polymerase I transcription factor, SL1. *Cell*, 68:965–976, 1992.

[49] L. Comai, J. C. B. M. Zomerdijk, Holger Beckmann, S. Zhou, A. Admon, and R. Tjian. Reconstitution of transcription factor SL1: Exclusive binding of TBP by SL1 or TFIID subunits. *Science*, 266:1966–1972, 1994.

[50] R.C. Conaway and J.W. Conaway. *Annual Review of Biochemistry*, volume 62, chapter General intiation factors for RNA polymerase II, pages 161–190. Annual Reviews Inc., 1993.

[51] B.P. Cormack, M. Strubin, A.S. Ponticelli, and K. Struhl. Functional differences between yeast and human TFIID are localized to the highly conserved region. *Cell*, 65:341–348, 1991.

[52] P. Cowan, D. Bulach, K. Goodge, A. Roberston, and D.E. Tribe. Nucleotide sequence of the polyhedrin gene region of the *Helicoverpa zea* single nucleocapsid nuclear polyhedrosis virus: placement of the virus in lepidopteran nuclear polyhedrosis virus group II. *Journal of General Virology*, 75:3211–3218, 1994.

[53] T.E. Crowley, T. Hoey, J. Liu, Y.N. Jan, L. Y. Jan, and R. Tjian. A new factor related to TATA-binding protein has highly restricted expression patterns in *Drosophila*. *Nature*, 361:557–561, 1993.

[54] S. Das and A. Dasgupta. Identification of the cleavage site and determinants required for Poliovirus 3C–catalyzed cleavage of human TATA–binding transcription factor TBP. *Journal of Virology*, 67:3326–3331, 1993.

[55] S. Datta, C. Soong, D.M. Wang, and M.L. Harter. A purified 289-amino-acid E1A protein activates RNA polymerase III transcription in vitro and alters transcription factor TFIIIC. *Journal of Virology*, 65:5297–5304, 1991.

[56] N.A. DeLuca and P.A. Schaffer. Activation of immediate–early, early and late promoters by temperature–sensitive and wild–type forms of herpes simplex virus type 1 protein ICP4. *Molecular and Cellular Biology*, 5(1997–2008), 1985.

[57] M.L. DePamphilis. *Annual Review of Biochemistry*, volume 62, chapter Eukaryotic DNA replication: Anatomy of an origin, pages 29–63. Annual Reviews Inc., 1993.

[58] K. Drlica and J. Rouviere-Yaniv. Histone–like proteins of bacteria. *Microbiological Reviews*, 51:301–319, 1987.

[59] B.D. Dynlacht, T. Hoey, and R. Tjian. Isolation of coactivators associated with the TATA–binding protein that mediate transcriptional activation. *Cell*, 66:563–576, 1991.

[60] R.V. Eck and M.O. Dayhoff. Atlas of protein sequence and structure. Technical report, National Biomedical Research Foundation, Silver Spring, MD, 1966.

[61] D. Eick, A. Wedel, and H. Heumann. From initiation to elongation: comparison of transcription by prokaryotic and eukaryotic RNA polymerases. *Trends in Genetics*, 10:292–296, 1994.

[62] D.M. Eisenmann, C. Dollard, and F. Winston. SPT15, the gene encoding the yeast TATA binding factor TFIID, is required for normal transcription initiation in vivo. *Cell*, 58:1183–91, 1989.

[63] E.K. Engelhard, L.N.W. Kam-morgan, J.O. Washburn, and L.E. Volkman. The insect tracheal system: a conduit for the systemic spread of *Autographa californica* M nuclear polyhedrosis virus. *Proceedings of the National Academy of Sciences USA*, 91:3224–3227, 1994.

[64] M.A. Erlandson, J. Gordon, and E.B. Carstens. Size and map locations of early transcription products on the *Autographa californica* nuclear polyhedrosis virus genome. *Virology*, 142:12–23, 1985.

[65] A.P. Feinberg and B. Vogelstein. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochemistry*, 132:6–13, 1983.

[66] L.T. Feldman, M.J. Imperiale, and J.R. Nevins. Activation of early adenovirus transcription by the herpesvirus immediate early gene: evidence for a common cellular control factor. *Proceedings of the National Academy of Sciences USA*, 79:4952–4956, 1982.

[67] J. Felsenstein. Numerical methods for inferring evolutionary trees. *The Quarterly Review of Biology*, 57:379–404, 1982.

[68] M.S. Finnin, D.W. Hoffman, and S.W. White. The DNA–binding domain of the MotA transcription factor from bacteriophage T4 shows struc-

tural similarity to the TATA–binding protein. *Proceedings of the National Academy of Sciences USA*, 91:10972–10976, 1994.

[69] W.M. Fitch. On the problem of discovering the most parsimonious tree. *American Naturalist*, 111:223–257, 1977.

[70] W.M. Fitch and E. Margoliash. Construction of phylogenetic trees. A method based on mutation distances estimated from cytochrome c sequences is of general applicability. *Science*, 155:279–284, 1967.

[71] S. Franck. Genbank accession number z28644. 1993.

[72] M. G. Frattini and L. A. Laimins. The role of E1 and E2 proteins in the replication of human papillomavirus type 31b. *Virology*, 204:799–804, 1994.

[73] P.D. Friesen and L.K. Miller. The regulation of baculovirus gene expression. *Current Topics in Microbiology and Immunology*, 131:31–49, 1986.

[74] P.D. Friesen and L.K. Miller. Divergent transcription of the early 35- and 94-kilodalton protein genes endcoded by the *Hind*III K genome fragment of the baculovirus *Autographa californica* nuclear polyhedrosis virus. *Journal of Virology*, 61:2264–2272, 1987.

[75] E.A. Fryberg, K. L. Kindle, and N. Davidson. The actin genes of Drosophila: a dispersed multigene family. *Cell*, 19:365–378, 1980.

[76] L.Y. Fuchs, M.S. Woods, and R.F. Weaver. Viral transcription during *Autographa californica* nuclear polyhedrosis infection: a novel RNA polymerase induced in infected *Spodoptera frugiperda* cells. *Journal of Virology*, 48:641–646, 1983.

[77] M.A. Hutchinson G. Walter, T. Hunter, and W. Eckhart. Purification of polyoma virus medium–size tumor antigen by immunoaffinity chromatography. *Proceedings of the National Academy of Sciences USA*, 79:4025–4029, 1982.

[78] T. A. Gahn and C.L. Schildkraut. The epstein–barr virus origin of plasmid replication, *orip*, contains both the initiation and termination sites of DNA replication. *Cell*, 58:527–535, 1989.

[79] A. Gasch, A. Hoffmann, M. Horikoshi, R.G. Roeder, and N.H. Chua. Arabidopsis thaliana contains two genes for TFIID. *Nature*, 346:390–4, 1990.

[80] G. Gill and R. Tijan. A highly conserved domain of TFIID displays species specificity in vivo. *Cell*, 65:333–340, 1991.

[81] G. Gill and R. Tjian. Eukaryotic co-activators associated with the TATA box binding protein. *Current Opinion in Genetics and Development*, 2:236–242, 1992.

[82] B. Glocker, R.R. Hoopes Jr., 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. *Journal of Virology*, 67:3771–3776, 1993.

[83] B. Glocker, R.R. Hoopes Jr., and G.F. Rohrmann. In vitro transactivation of baculovirus early genes by nuclear extracts from *Autographa californica* nuclear polyhedrosis virus-infected *Spodoptera frugiperda* cells. *Journal of Virology*, 66:3476–3484, 1992.

[84] A.F. Gombart, M.N. Pearson, G.F. Rohrmann, and G.S. Beaudreau. A baculovirus polyhedral envelope–associated protein: genetic location, nucleotide sequence, and immunoctyochemical characterization. *Virology*, 169:182–193, 1989.

[85] J.A. Goodrich, T. Hoey, C.J. Thut, A. Admon, and R. Tjian. Drosophila TAFII40 interacts with both a VP16 activation domain and the basal transcription factor TFIIB. *Cell*, 75:519–530, 1993.

[86] J.A. Goodrich and R. Tjian. TBP–TAF complexes: selectivity factors for eukaryotic transcription. *Current Opinion in Cell Biology*, 6:403–409, 1994.

[87] R.R. Granados and K. A. Lawler. In vivo pathway of *Autographa californica* baculovirus invasion and infection. *Virology*, 108:297–308, 1981.

[88] J. Greenblatt. Roles of TFIID in transcriptional initiation by RNA polymerase II. *Cell*, 66:1067–1070, 1991.

[89] C.H. Gross, R.L.Q. Russell, and G.F. Rohrmann. *Orgyia psuedotsugata* baculovirus p10 and polyhedron envelope protein genes: analysis of their relative expression levels and role in polyhedron structure. *Journal of General Virology*, 75:1115–1123, 1994.

[90] M.K. Gross, M.S. Kainz, and G.F. Merrill. The chicken thymidine kinase gene is transcriptionally repressed during terminal differentiaition: The associated decline in TK mRNA cannot account fully for the disappearance of TK enzyme activity. *Developmental Biology*, 81:439–451, 1987.

[91] M.A. Grula, P.L. Buller, and R.F. Weaver. α–amanitin–resistant viral RNA synthesis in nuclei isolated from nuclear polyhedrosis virus-infected *Heliothis zea* larvae and *Spodoptera frugiperda* cells. *Journal of Virology*, 38:916–921, 1981.

[92] M. Grunstein and D. Hogness. Colony hybridization: A method for the isolation of cloned DNAs that contain a specific gene. *Proceedings of the National Academy of Sciences USA*, 72:3961–3965, 1975.

[93] B. Gu and N. DeLuca. Requirements for the activation of the Herpes Simplex Virus Glycoprotein C promoter in vitro by the viral regulatory protein ICP4. *Journal of Virology*, 68:7953–7965, 1994.

[94] B. Gu, R. Kuddus, and N.A. Deluca. Repression of activator–mediated transcription by herpes simplex virus ICP4 via a mechanism involving interactions with the basal transcription factors TATA–binding protein and TFIIB. *Molecular and Cellular Biology*, 15:3618–3626, 1995.

[95] B. Gu, R. Rivera-Gonzalez, C.A. Smith, and N.A. DeLuca. Herpes simplex virus infected cell polypeptide 4 preferentially represses Sp1–activated over basal transcription from its own promoter. *Proceedings of the National Academy of Sciences USA*, 90:9528–9532, 1993.

[96] L.A. Guarino and W. Dong. Expression of an enhancer-binding protein in insect cells transfected with the *Autographa californica* nuclear polyhedrosis virus. *Journal of Virology*, 65:3676–3680, 1991.

[97] L.A. Guarino and W. Dong. Functional dissection of the *Autographa californica* nuclear polyhedrosis virus enhancer element *hr*5. *Virology*, 200:328–335, 1994.

[98] 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. *Journal of Virology*, 60:224–229, 1986.

[99] L.A. Guarino and M. Smith. Nucleotide sequence and charaterization of the 39K gene region of *Autographa californica* nuclear polyhedrosis virus. *Virology*, 179:1–8, 1990.

[100] L.A. Guarino and M. Smith. Regulation of delayed–early gene transcription by dual TATA boxes. *Journal of Virology*, 66:3733–3739, 1992.

[101] L.A. Guarino and M.D. Summers. Functional mapping of a *trans*-activating gene required for expression of a baculovirus delayed-early gene. *Journal of Virology*, 57:563–571, 1986.

[102] L.A. Guarino and M.D. Summers. Interspersed homologous DNA of *Autographa californica* nuclear polyhedrosis virus enhances delayed-early gene expression. *Journal of Virology*, 60:215–223, 1986.

[103] L.A. Guarino and M.D. Summers. Nucleotide sequence and temporal expression of a baculovirus regulatory gene. *Journal of Virology*, 61:2091–2099, 1987.

[104] M.M. Haab and G. Feix. Two different cDNAs encoding TFIID proteins of maize. *FEBS Letters*, 301:294–298, 1992.

[105] M.S. Hafner, P.D. Sudman, F. X. Villblanca, T.A. Spradling, J.W. Demastes, and S. A. Nadler. Disparate rates of molecular evolution in cospeciating hosts and parasites. *Science*, 265:1087–1090, 1994.

[106] C. Hagemeir, S. Walkerand R. Caswell, T. Kouzarides, and J. Sinclair. The human cytomegalovirus 80–kilodalton but not the 72–kilodalton immediate–early protein transactivates heterologous promoters in a TATA box–dependent mechanism and interacts directly with TFIID. *Journal of Virology*, 66:4452–4456, 1992.

[107] S. Hahn, S. Buratowski, P.A. Sharp, and L. Guarente. Isolation of the gene encoding the yeast TATA binding protein TFIID: a gene identical to the SPT15 suppressor of Ty element insertions. *Cell*, 58:1173–81, 1989.

[108] T. Hai, M. Horikoshi, R.G. Roeder, and M.R. Green. Analysis of the role of the transcription factor ATF in the assembly of a functional preinitiation complex. *Cell*, 54:1043–1051, 1988.

[109] X. Hang, W. Dong, and L.A. Guarino. The *lef–3* gene of *Autographa californica* nuclear polyhedrosis virus encodes a single–stranded DNA–binding protein. *Journal of Virology*, 69:3924–3928, 1995.

[110] S.C. Harrison. A structural taxomony of DNA–binding domains. *Nature*, 353:715–717, 1991.

[111] S. Hashimoto, H. Hisakazu, S. Hasegawa, R.G. Roeder, and M. Horikoshi. Conserved structural motifs within the N–terminal domain of TFIID from *Xenopus*, mouse and human. *Nucleic Acids Research*, 20:3788, 1992.

[112] D.K. Hawley and R.G. Roeder. Separation and partial characterization of three functional steps in transcription intiation by human RNA polymerase II. *Journal of Biological Chemistry*, 260:8163–8172, 1985.

[113] S. Henikoff. Unidirectional digestion with Exonuclease III in DNA sequence analysis. *Methods in Enzymology*, 155:156–166, 1987.

[114] N. Hernandez. TBP, a universal eukaryotic transcription factor. *Genes and Development*, 7:1291–1308, 1993.

[115] D.G. Higgins, A.J. Beasbly, and R. Fuchs. Clustal V: improved software for multiple sequence alignment. *Computer applications in the Biosciences*, 8:189–191, 1992.

[116] D.M. Hillis, J.P. Hueslsenbeck, and C.W. Cunningham. Application and accuracy of molecular phylogenies. *Science*, 264:671–677, 1994.

[117] T. Hoey, B.D. Dynlacht, M.G. Peterson, B.F. Pugh, and R. Tjian. Isolation and characterization of the Drosophila gene encoding the TATA box binding protein, TFIID. *Cell*, 61:1179–86, 1990.

[118] T. Hoey, R.O.J. Weinzierl, G. Gill, J. Chen, B.D. Dynlacht, and R. Tjian. Molecular cloning and functional analysis of Drosophila TAF110 reveal properties expected of coactivators. *Cell*, 72:247–260, 1993.

[119] A. Hoffmann, M. Horikoshi, C.K. Wang, S. Schroeder, P.A. Weil, and R.G. Roeder. Cloning of the *Schizosaccharomyces pombe* TFIID gene reveals a strong conservation of functional domains present in *Saccharomyces cerevisiae* TFIID. *Genes and Development*, 4:1141–8, 1990.

[120] M.J. Holdsworth, C. Grierson, W. Schuch, and M. Bevan. DNA–binding properties of cloned TATA–binding protein from potato tubers. *Plant Molecular Biology*, 19:455–464, 1992.

[121] D.S. Holmes and M. Quigley. A rapid boiling method for the preparation of bacterial plasmids. *Analytical Biochemstry*, 114:193–197, 1981.

[122] R.R. Hoopes Jr. and G.F. Rohrmann. *In vitro* transcription of baculovirus immediate early genes: Accurate mRNA initiation by nuclear extracts from both insect and human cells. *Proceedings of the National Academy of Sciences USA*, 88:4513–4517, 1991.

[123] M. Horikoshi, M.F. Carey, H. Kakidani, and R.G. Roeder. Mechanism of action of a yeast activator: direct effect of GAL4 derivatives on mammalian TFIID–promoter interactions. *Cell*, 54:665–9, 1988.

[124] M. Horikoshi, T. Hai, Y. Lin, M.R. Green, and R.G. Roeder. Transcription factor ATF interacts with the TATA factor to facilitate establishment of a preinitiation complex. *Cell*, 54:1033–1042, 1988.

[125] M. Horikoshi, C.K. Wang, H. Fujii, J.A. Cromlish, P.A. Weil, and R.G. Roeder. Cloning and structure of a yeast gene encoding a general transcription initiation factor TFIID that binds to the TATA box. *Nature*, 341:299–303, 1989.

[126] M. Horikoshi, T. Yamamoto, Y. Ohkuma, P.A. Weil, and R.G. Roeder. Analysis of structure–function relationships of yeast TATA box binding factor TFIID. *Cell*, 61:1171–8, 1990.

[127] N. Horikoshi, K. MacGuire, A. Kralli, E. Maldonado, D. Reinberg, and R. Weinmann. Direct interaction bewteen adenovirus E1A protein and the TATA–box binding transcription factor IID. *Proceedings of the National Academy of Sciences USA*, 88:5124–5128, 1991.

[128] D.L. Hostetter and B. Puttler. A new broad host spectrum nuclear polyhedrosis virus isolated from celery looper, *anagrapha falcifera* (kirby), (lepidoptera: Nocturnidae). *Environmental Entomology*, 20:1480–1488, 1991.

[129] R.Y. Huang and D. Kowalski. A DNA unwinding element and an *ars* consensus comprise a replication origin within a yeast chromosome. *EMBO Journal*, 12:4521–4531, 1993.

[130] N. Iwabe, K. Kuma, M. Hisegawa, S. Osawa, and T. Miyata. Evolutionary relationships of archaebacteria, eubacteria and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proceedings of the National Academy of Sciences USA*, 86:9355–9359, 1989.

[131] B.R. Jasny. Insect viruses invade biotechnology. *Science*, 238:1653, 1988.

[132] R.A. Jefferson. Assaying chimeric genes in plants: The GUS gene fusion system. *Plant Molecular Biology Reporter*, 5:387–405, 1987.

[133] T.H. Jukes and C.R. Cantor. *Mammalian Protein Metabolism*, chapter Evolution of protein molecules, pages 21–132. Academic Press, New York, 1969.

[134] R. Jupp, S. Hoffmann, R.M. Steinberg, J.A. Nelson, and P. Ghazal. Human cytomegalovirus IE86 protein interacts with promoter–bound TATA–binding protein via a specific region distinct from the autorepression domain. *Journal of Virology*, 67:7539–7546, 1993.

[135] J. T. Kadonga. Purification of sequence–specific binding proteins by DNA affinity chromatography. *Methods in Enzymology*, 208:10–24, 1991.

[136] C.C. Kao, P.M. Lieberman, M.C. Schmidt, Q. Zhou, R. Pei, and A.J. Berk. Cloning of a transcriptionally active human TATA binding factor. *Science*, 248:1646–50, 1990.

[137] J. Kaufmann and S.T. Smale. Direct recognition of initiator elements by a component of the transcription factor IID complex. *Genes and Development*, 8:821–829, 1994.

[138] T. Kawata, M. Minami, T. Tamura, K. Sumita, and M. Iwabuchi. Isolation and characterization of a cDNA clone encoding the TATA box–binding protein (TFIID) from wheat. *Plant Molecular Biology*, 19:867–872, 1992.

[139] B.A. Keddie, G.W. Aponte, and L.E.Volkman. The pathway of infection of *Autographa californica* nuclear polyhedrosis virus in an insect host. *Science*, 243:1728–1730, 1989.

[140] J.L. Kim, D.B. Nikolov, and S.K. Burley. Co–crystal structure of TBP recognizing the minor grove of a TATA element. *Nature*, 365:520–527, 1993.

[141] Y. Kim, J.H. Geiger, S. Hahn, and P. B. Sigler. Crystal structure of a yeast TBP/TATA–box complex. *Nature*, 365:512–520, 1993.

[142] M. Kimura. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, 16:111–120, 1980.

[143] B.A. Kirkpatrick, J.O. Washburn, E.K. Engelhard, and L.E. Volkman. Primary infection of insect tracheae by *Autographa californica* M nuclear polyhedrosis virus. *Virology*, 203:184–186, 1994.

[144] M.P. Kladde, Y. Kohwi, T. Kowhi-Shigematsu, and J. Gorski. The non–B DNA structure of d(CA/TG) differs from that of Z–DNA. *Proceedings of the National Academy of Sciences USA*, 91:1898–1902, 1994.

[145] A.H. Knoll. The early evolution of eukaryotes: a geological perspective. *Science*, 256:622–627, 1992.

[146] T. Kokubo, D.W. Gong, S. Yamashita, S. Horikoshi, R.G. Roeder, and Y. Nakatani. *Drosophila* 230 kDa TFIID subunit, a functional homolog of the human cell cycle gene product negatively regulates DNA binding of the TATA box–binding subunit of TFIID. *Genes and Development*, 7:1033–1046, 1993.

[147] M. Kool, C. H. Ahrens, J.M. Vlak, and G.F. Rohrmann. Replication of baculovirus DNA. *Journal of General Virology*, 76:2103–2118, 1995.

[148] 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. *Proceedings of the National Academy of Sciences USA*, 91:11212–11216, 1994.

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

[150] M. Kool and J.M. Vlak. The structural and functional organization of the *Autographa californica* nuclear polyhedrosis virus genome: an overview. *Archives of Virology*, 130:1–16, 1993.

[151] M. Kool, J.T.M. Voeten, R.W. Golbach, and J.M. Vlak. Functional mapping of regions of the *Autographa californica* nuclear polyhedrosis viral genome required for DNA replication. *Virology*, 168:680–689, 1994.

[152] M. Kool, J.T.M. Voeten, R.W. Goldbach, J. Tramper, and J.M. Vlak. Identification of seven putative origins of *Autographa californica* multiple nuclear polyhedrosis virus DNA replication. *Journal of General Virology*, 74:2661–2668, 1993.

[153] A. Kornberg and T.A. Baker. *DNA Replication.* W.H. Freeman and Company, New York, 2nd edition, 1992.

[154] G. R. Kovacs, L.A. Guarino, B.L. Graham, and M.D. Summers. Identification of spliced baculovirus RNAs expressed late in infection. *Virology*, 185:633–643, 1991.

[155] 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. *Journal of Virology*, 66:7429–7437, 1992.

[156] G.R. Kovacs, L.A. Guarino, and M.D. Summers. Novel regulatory properties of the IE1 and IE0 transactivators encoded by the baculovirus *Autographa californica* multicapsid nuclear polyhedrosis virus. *Journal of Virology*, 65:5281–5288, 1991.

[157] M. Kozak. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Research*, 15:8125–8132, 1987.

[158] R. Krappa and D. Knebel-Morsdorf. Identification of the very early transcribed baculovirus gene PE-38. *Journal of Virology*, 65:805–812, 1991.

[159] N. P. Kristensen. Phylogeny of insect orders. *Annual Review of Entomology*, 26:135–157, 1981.

[160] R. Kuddos and M.C. Schmidt. Effect of the non–conserved N terminus on the DNA binding activity of the yeast TATA binding protein. *Nucleic Acids Research*, 21:1789–1796, 1993.

[161] J. Kuzio and P. Faulkner. Region of repeated DNA in the genome of *Choristoneura fumiferana* nuclear polyhedrosis virus. *Virology*, 139:185–188, 1984.

[162] C.E. Lawrence, S.F. Altschul, M.S. Boguski, J.S. Liu, A.F. Neuwald, and J.C. Wootton. Detecting subtle sequence signals: A gibbs sampling strategy for multiple alignment. *Science*, 262:208–213, 1993.

[163] D.K. Lee, J. Dejong, S. Hashimoto, M. Horikoshi, and R.G. Roeder. TFIIA induces confromational changes in TFIID via interactions with the basic repeat. *Molecular and Cellular Biology*, 12:5189–5196, 1992.

[164] D.K. Lee, M. Horikoshi, and R.G. Roeder. Interaction of TFIID in the minor groove of the TATA element. *Cell*, 67:1241–1250, 1991.

[165] H. Lee and P.J. Krell. Generation and analysis of defective genomes of *Autographa californica* nuclear polyhedrosis virus. *Journal of Virology*, 66:4339–4347, 1992.

[166] H. Lee and P.J. Krell. Reitereated DNA fragments in defective genomes of *Autographa californica* nuclear polyhedrosis virus are competent for AcMNPV–dependent replication. *Virology*, 202:418–429, 1994.

[167] W.S. Lee, C.C. Kao, G.O. Bryant, X. Liu, and A.J. Berk. Adenovirus E1A activation domain binds the basic repeat in the TATA box transcription factor. *Cell*, 67:365–376, 1991.

[168] D.J. Leisy and N. Van Beek. Baculoviruses: Possible alternatives to chemical insecticides. *Chemistry and Industry*, 7:250–254, 1992.

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

[170] D.J. Leisy and G.F. Rohrmann. Characterization of the replication of plasmids containing *hr* sequences in baculovirus–infected *Spodoptera frugiperda* cells. *Virology*, 196:722–730, 1993.

[171] D.J. Leisy, G.F. Rohrmann, and G.S. Beaudreau. Conservation of genome organization in two multicapsid nuclear polyhedrosis viruses. *Journal of Virology*, 52:699–702, 1984.

[172] D.J. Leisy, G.F. Rohrmann, M. Nesson, and G.S. Beaudreau. Nucleotide sequencing and transcriptional mapping of the *Orygia pseudotsugata* multicapsid nuclear polyhedrosis virus p10 gene. *Virology*, 153:157–167, 1986.

[173] R. Li and M. R. Botchan. The acidic transcriptional activation domains of VP16 and p53 bind the cellular replication protein A and stimulate *in vitro* BPV–1 DNA replication. *Cell*, 73:1207–1221, 1993.

[174] R. Li and M.R. Botchan. Acidic transcription factors alleviate nucleosome–mediated repression of DNA replication of bovine papillomavirus type I. *Proceedings of the National Academy of Sciences USA*, 91:7051—7055, 1993.

[175] R. Li, L. Yang, E. Fouts, and M.R. Botchan. Site–specific DNA–binding proteins important for replication and transcription have multiple activities. *Cold Spring Harbor Symposia on Quantitative Biology*, LVIII:403–413, 1993.

[176] S. Li and J.E. Donelson. The gene for the TATA box–binding protein of *Onchocera volvulus*. *Molecular and Biochemical Parasitology*, 61:321–324, 1993.

[177] Y. Li, A.L. Passarelli, and L.K. Miller. Identification, sequence and transcriptional mapping of *lef*–3, a baculovirus gene involved in late and very

late gene expression. *Journal of Virology*, 67:5260–5268, 1993.

[178] S. Lichtsteiner and R. Tjian. Cloning and properties of the *Caenorhabditis elegans* TATA–box–binding protein. *Proceedings of the National Academy of Sciences USA*, 90:9673–9677, 1993.

[179] P. Lieberman. Identification of functional targets of the Zta transcriptional activator by formation of stable pre-initiation complex intermediates. *Molecular and Cellular Biology*, 14:8365–8375, 1994.

[180] P.M. Lieberman and A. J. Berk. The Zta trans–activator protein stabilizes TFIID association with promoter DNA by direct protein–protein interaction. *Genes and Development*, 5:2441–2454, 1991.

[181] D.M.J. Lilley. The inverted repeat as a recognizable structural feature in supercoiled DNA molecules. *Proceedings of the National Academy of Sciences USA*, 77:6468–6472, 1980.

[182] Y. Lin and M.R. Green. Mechanism of action of an acidic transcriptional activator in vitro. *Cell*, 64:971–981, 1991.

[183] S.M. Lobo, J. Lister, M.L. Sullivan, and N. Hernandez. The cloned RNA polymerase transcritpion factor IID selects RNA polymerase III to transcribe the human U6 gene in vitro. *Genes and Development*, 5:1477–1489, 1991.

[184] 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*, 181:336–346, 1991.

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

[186] 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. *Journal of Virology*, 68:6710–6718, 1994.

[187] A. Lu and L.K. Miller. The roles of eighteen late expression factor genes in transcription and DNA replication. *Journal of Virology*, 69:975–982, 1995.

[188] V.A. Luckow and M.D. Summers. Trends in the development of baculovirus expression vectors. *Biotechnology*, 6:47–55, 1988.

[189] S. Maeda, S.G. Kamita, and H. Kataoka. The basic DNA–binding protein of *Bombyx mori* nuclear polyhedrosis virus: the existence of an additional arginine repeat. *Virology*, 180:807–810, 1991.

[190] S. Maeda and K. Majima. Molecular cloning and physical mapping of the genome of *Bombyx mori* nuclear polyhedrosis virus. *Journal of General Virology*, 71:1851–1855, 1990.

[191] K. Majima, R. Kobara, and S. Maeda. Divergence and evolution of homologous regions of *Bombyx mori* nuclear polyhedrois virus. *Journal of Virology*, 67:7513–7521, 1993.

[192] T. Maniatis, E.F. Fritsch, and J. Sambrook. *Molecular cloning: a laboratory manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982.

[193] T.L. Marsh, C. I. Reich, R.B. Whitlock, and G.J. Olsen. Transcription factor IID in Archea: Sequences in the *Thermococcus celer* genome would encode a product closely related to the TATA–binding protein of eukaryotes. *Proceedings of the National Academy of Sciences USA*, 91:4180–4184, 1994.

[194] W. F. Marzluff and R.C.C. Huang. *Transcription of RNA from isolated nuclei*, chapter 4, pages 89–129. IRL Press, Washington, D.C., 1984.

[195] W.F. Marzluff. Preparation of active nuclei. *Methods in Enzymology*, 181:30–36, 1990.

[196] I.A. Mastrangelo, P.V. C. Hough, and J.S. Wall. ATP–dependent assembly of double hexamers of SV40 T antigen at the viral origin of replication. *Nature*, 338:658–662, 1989.

[197] M.B. McAndrew, M. Read, P.F.G. Sims, and J.E. Hyde. Characterisation of the gene encoding an unusually divergent TATA–binding protein (TBP) from the extremmely A+T rich human malaria pararsite, *Plasmodium falciparum*. *Gene*, 124:165–171, 1993.

[198] R. McCarroll, G.J. Olsen, Y.D. Stahl, C.R. Woese, and M.L. Sogin. Nucleotide sequence of the *Dictyostelium discoideum* small-subunit ribosomal ribonucleic acid inferred from the gene sequence: evolutionary implications. *Biochemistry*, 22:5858–5868, 1983.

[199] J.R. McLachlin and L.K. Miller. Identification and characterization of *vlf–1*, a baculovirus gene involved in very late gene expression. *Journal of Virology*, 68:7746–7756, 1994.

[200] C.T. McMurray, W. D. Wilson, and J.O. Douglass. Hairpin formation within the enhancer region of the human enkephalin gene. *Proceedings of the National Academy of Sciences USA*, 88:666–670, 1991.

[201] J.C. Meade and J.R. Stringer. PCR amplification of DNA sequences from transcription factor IID and cation transporting ATPase genes in *Pneumocystis carinii*. *Journal of Protozoology.*, 38:66S–68S, 1991.

[202] M. Meisterernst, M. Horikoshi, and R.G. Roeder. Recombinant yeast TFIID, a general transcription factor, mediates activation by the gene–specific factor USF in a chromatin assembly assay. *Proceedings of the National Academy of Sciences USA*, 87:9153–7, 1990.

[203] M. Mizutani, T. Ohta, H. Watanabe, H. Handa, and S. Hirose. Negative supercoiling of DNA facilitates an interaction between transcription factor IID and the fibrion gene promoter. *Proceedings of the National Academy of Sciences USA*, 88:718–722, 1991.

[204] S.A. Monsma and G.W. Blissard. Identification of a membrane fusion domain and an oligmerization domain in the baculoviurs GP64 envelope fusion protein. *Journal of Virology*, 69:2583–2595, 1995.

[205] T.D. Morris and L.K. Miller. Promoter influence on baculovirus-mediated gene expression in permissive and nonpermissive insect cell lines. *Journal of Virology*, 66:7397–7405, 1992.

[206] T.D. Morris and L.K. Miller. Mutational analaysis of a baculovirus major late promoter. *Gene*, 140:147–153, 1994.

[207] T.D. Morris and L.K. Miller. Mutational analysis of a baculovirus major late promoter. *Gene*, 140:147–153, 1994.

[208] T.D. Morris, J.W. Todd, B. Fisher, and L.K. Miller. Identification of *lef-7*: A baculovirus gene affecting late gene expression. *Virology*, 200:360–369, 1994.

[209] M.L. Muhich, C.T. Iida, M. Horikoshi, R.G. Roeder, and C.S. Parker. cDNA clone encoding Drosophila transcription factor TFIID. *Proceedings of the National Academy of Sciences USA*, 87:9148–52, 1990.

[210] Y.M. Mul, C.P. Verrijzer, and P. van der Vliet. Transcription factors NFI and NFIII/Oct–1 function independently, employing different mechanisms to enhance adenovirus replication. *Journal of Virology*, 64:5510–5518, 1990.

[211] M.T. Muller. Binding of the herpes simplex virus type 1 gene product ICP4 to its own transcription start site. *Journal of Virology*, 61:858–865, 1987.

[212] R. Muller, M.N. Pearson, R.L.Q. Russell, and G.F. Rohrmann. A capsid-associated protein of the multicapsid nuclear polyhedrosis virus of *Orygia pseudotsugata*: genetic location, sequence, transcriptional mapping, and immunocytochemical characterization. *Virology*, 176:133–144, 1990.

[213] A.I.H. Murchie, R. Bowater, F. Aboul-ela, and D.M.J. Lilley. Helix opening transitions in supercoiled DNA. *Biochimica et Biophysicia Acta*, 1131:1–15, 1992.

[214] N. Nakajima, M. Horikoshi, and R.G. Roeder. Factors involved in specific transcription by mammalian RNA polymerase II: purification, genetic specificity, and TATA box–promoter interactions of TFIID. *Molecular and Cellular Biology*, 8:4028–40, 1988.

[215] H.A. Nash and A.E. Granston. Similarity between the DNA–binding domains of IHF protein and TFIID protein. *Cell*, 67:1037–1038, 1991.

[216] V.G. Nicoletti and D.F. Condorelli. Optimized PEG method for rapid plasmid DNA purification: high yield from midi-prep. *Biotechniques*, 14:532–533, 1993.

[217] D.B. Nikolov, S. Hu, J. Lanand A. Gasch, A. Hoffmann, M. Horikoshi, N. Chua, R.G. Roeder, and S.K. Burley. Crystal structure of TFIID TATA–box binding protein. *Nature*, 360:40–46, 1992.

[218] M.S. Nissen and P.D. Friesen. Molecular analysis of the transcriptional regulatory region of an early baculovirus gene. *Journal of Virology*, 63:493–503, 1989.

[219] E. Noguchi, T. Sekiguchi, Y. Nohiro, T. Hayashida, and E. Hirose. Genbank accession number D30051. 1994.

[220] V. G. Norton, B. S. Imai, P. Yau, and E. M. Bradbury. Histone acetylation reduces nucleosome core particle linking number change. *Cell*, 57:449–457, 1989.

[221] P. O'Hare and G.S. Hayward. Evidence for the direct role for both the 175,000– and 110,000–molecular–weight immediate–early proteins of herpes simplex virus in the transactivation of delayed–early promoters. *Journal of Virology*, 53:723–733, 1985.

[222] P. O'Hare and G.S. Hayward. Three trans–acting regulatory proteins of herpes simplex virus modulate immediate–early gene expression in a pathway involving positive and negative feedback regulation. *Journal of Virology*, 56:723–733, 1985.

[223] M. Ohresser, N. Morin, M. Cerutti, and C. Delsert. Temporal regulation of a complex and unconvential promoter by viral products. *Journal of Virology*, 68:2589–2597, 1994.

[224] G.J. Olsen and M.L. Sogin. Nucleotide sequence of *Dictyostelium discoideum* 5.8S ribosomal ribonucleic acid: evolutionary and secondary structural implications. *Biochemistry*, 21:2335–2343, 1982.

[225] B. G. Ooi and L.K. Miller. Regulation of host RNA levels during baculovirus infection. *Virology*, 166:515–523, 1988.

[226] B.G. Ooi, C. Rankin, and L.K. Miller. Downstream sequences augment transcription from the essential initiation site of the baculovirus polyhedrin gene. *Journal of Molecular Biology*, 210:721–736, 1989.

[227] D.R. O'Reilly, L.K. Miller, and V.A. Lucknow. *Baculovirus Expression Vectors: A laboratory manual.* W.H. Freeman and Company, New York, 1st edition, 1992.

[228] E. Orozco. Genbank accession number Z48307. 1995.

[229] C. Ouzounis and C. Sander. TFIIB, an evolutionary link between the transcription machineries of archaebacteria and eukaryotes. *Cell*, 71:189–190, 1992.

[230] A.L. Passarelli and L.K. Miller. Identification and characterization of *lef–1*, a baculovirus gene involved in late and very late gene expression. *Journal of Virology*, 67:3481–3488, 1993.

[231] A.L. Passarelli and L.K. Miller. Identification of genes encoding late expression factors located between 56.0 and 65.4 map units of the *Autographa californica* nuclear polyhedrosis virus genome. *Virology*, 197:704–714, 1993.

[232] A.L. Passarelli and L.K. Miller. Three baculovirus genes involved in late and very late gene expression: *ie–1, ie–n* and *lef–3*. *Journal of Virology*, 67:2149–2158, 1993.

[233] A.L. Passarelli and L.K. Miller. Identification and transcriptional regulation of the baculovirus *lef–6* gene. *Journal of Virology*, 68:4458–4467, 1994.

[234] 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. *Journal of Virology*, 68:4673–4678, 1994.

[235] M. Pearson, R. Bjornson, G. Pearson, and G. Rohrmann. The *Autographa californica* baculovirus genome: evidence for multiple replication origins. *Science*, 257:1382–1384, 1992.

[236] M.N. Pearson, R.M. Bjornson, C.H. Ahrensxsave-buffer, and G.F. Rohrmann. Identification and characterization of a putative origin of DNA replication in the genome of a baculovirus pathogenic for *Orygia pseudotsugata*. *Virology*, 197:715–725, 1993.

[237] M.N. Pearson and G.F. Rohrmann. *Lymantria dipsar* nuclear polyhedrosis virus homologous regions: characterization of their ability to function as replication origins. *Journal of Virology*, 69:213–221, 1995.

[238] M.N. Pearson, R.L.Q. Russell, G.F. Rohrmann, and G.S. Beaudreau. p39, a major baculovirus structural protein: immunoctyochemical characterization and genetic location. *Virology*, 167:407–413, 1988.

[239] K.W. Peden, C. Pipas, J.M. Pearson-White, and D. Nathans. Isolation of mutants of an animal virus in bacteria. *Science*, 209:1392–1396, 1980.

[240] D.S. Pederson and N. H. Heintz. *Transcription Factors and DNA replication.* CRC press, Boca Raton, Fl., 1994.

[241] M.G. Peterson, N. Tanese, B.F. Pugh, and R. Tjian. Functional domains and upstream activation properties of cloned human TATA binding protein. *Science*, 248:1625–30, 1990.

[242] J.R. Platt. Possible separation of intertwined nucleic acids chains by transfer–twist. *Proceedings of the National Academy of Sciences USA*, 41:181–183, 1955.

[243] D. Poon, S. Schroeder, C. K. Wang, T. Yamamoto, M. Horikoshi, R. G. Roeder, and A. P. Weil. The conserved carboxy-terminal domain of *Saccharomyces cerevisiae* TFIID is sufficient to support normal cell growth. *Molecular and Cellular Biology*, 11:4809–4821, 1991.

[244] B.F. Pugh and R. Tjian. Mechanism of transcriptional activation by Sp1: evidence for coactivators. *Cell*, 61:1187–97, 1990.

[245] B.F. Pugh and R. Tjian. Transcription from a TATA–less promoter requires a multisubunit TFIID complex. *Genes and Development*, 5:1935–1945, 1991.

[246] G. Puhler, H. Leffers, F. Gropp, P. Palm, H. Klenk, F. Lottspeich, R.A. Garrett, and W. Zillig. Archaebacterial DNA–dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. *Proceedings of the National Academy of Sciences USA*, 86:4569–4573, 1989.

[247] R.L. Quant, M.N. Pearson, G.F. Rohrmann, and G.S. Beaudreau. Production of polyhedrin monoclonal antibodies for distinguishing two *Orygia pseudotsugata* baculoviruses. *Applied and Environmental Microbiology*, 84:732–736, 1984.

[248] R.L. Quant-Russell, M.N. Pearson, G.F. Rohrmann, and G.S. Beaudreau. Characterization of baculovirus p10 synthesis using monoclonal antibodies. *Virology*, 160:9–19, 1987.

[249] C.A. Radebaugh, J.L. Matthews, G.K. Geiss, F. Liu, J.M. Wong, E. Bateman, S. Camier, A. Sentenac, and M. R. Paule. TATA box-binding protein (TBP) is a constituent of the polymerase I–specific transcription initiation factor TIF–IB (SL1) bound to the rRNA promoter and shows differential sensitivity to TBP–directed reagents in polymerase I, II and III transcription factors. *Molecular and Cellular Biology*, 14:597–605, 1994.

[250] C. Rankin, B.F. Ladin, and R.F. Weaver. Physical mapping of temporally regulated, overlapping transcripts in the region of the 10K protein gene in *Autographa californica* nuclear polyhedrosis virus. *Journal of Virology*, 57:18–27, 1986.

[251] C. Rankin, B.G. Ooi, and L. K. Miller. Eight base pairs encompassing the transcriptional start point are the major determinants for baculovirus polyhedrin expression. *Gene*, 70:39–49, 1988.

[252] C. Rasmussen and G.F. Rohrmann. Charaterization of the *Spodoptera frugiperda* TATA–Binding Protein: nucleotide sequence and response to baculovirus infection. *Insect Biochemistry and Molecular Biology*, 7:699–708, 1994.

[253] P. Reddy and S. Hahn. Dominant negative mutations in yeast TFIID define a bipartite DNA-binding region. *Cell*, 65:349–357, 1991.

[254] D. Reinberg and R.A. Roeder. Factors involved in specific transcription by mammalian RNA polymerase II: Purification and functional analysis of

initiation factors IIB and IIE. *Journal of Biological Chemistry*, 262:3310–3321, 1987.

[255] D. Reisman and B. Snugden. *Trans* activation of an Epstein–Barr viral transcriptional enhancer by the Epstein–Barr viral nuclear antigen 1. *Molecular and Cellular Biology*, 6:3838–3846, 1986.

[256] B.R. Ribeiro, K. Hutchinson, and L.K. Miller. A mutant baculovirus with a temperature–sensitive IE–1 transregulatory protein. *Journal of Virology*, 68:1075–1084, 1994.

[257] R. Rivera-Gonzalez, A.N. Imbalzano, B. Gu, and N.A. DeLuca. The role of ICP4 repressor activity in temporal expression of IE–3 and latency associated promoters during HSV–1 infection. *Virology*, 202:550–564, 1994.

[258] M. S. Roberts, A. Boundy, P. O'Hare, M. C. Pizzorno, D.M. Ciufo, and G.S. Hayward. Direct correlation bewteen a negative autoregulatory response element at the cap site of the herpes simplex virus type 1 IE175 promoter and a specific binding site for the IE175 (ICP4) protein. *Journal of Virology*, 62:4307–4320, 1988.

[259] S.M. Rodems and P.D. Friesen. Transcriptional enhancer activity of *hr*5 requires dual–palindrome half sites that mediate binding of a dimeric form of the baculovirus transregulator IE1. *Journal of Virology*, 69:5368–5375, 1995.

[260] P.W. Roelvink, M.M.M. van Meer, C.A.D. de Kort, R.D. Possee, B.D. Hammock, and J.M. Vlak. Dissimilar expression of *Autographa californica* multiple nucleocapsid nuclear polyhedrosis virus polyhedrin and p10 genes. *Journal of General Virology*, 73:1481–1489, 1992.

[261] G.F. Rohrmann. Polyhedrin structure. *Journal of General Virology*, 67:1499–1513, 1986.

[262] G.F. Rohrmann. Baculovirus structural proteins. *Journal of General Virology*, 73:749–761, 1992.

[263] G.F. Rohrmannn. Characterization of N–polyhedrin of two baculovirus strains pathogneic for *Orygia pseudotsugata*. *Biochemistry*, 1977:1631–1634, 1977.

[264] T. Rowlands, P. Baumann, and S.P. Jackson. The TATA–Binding Protein: a general transcription factor in Eukaryotes and Archaebacteria. *Science*, 264:1325–1329, 1994.

[265] S.J. Rubinstein, T. Hammerle, E. Wimmer, and A. Dasgupta. Infection of HeLa cells with poliovirus results in modification of a complex that binds to the rRNA promoter. *Journal of Virology*, 66:3062–3068, 1992.

[266] R.L.Q. Russell, M.N. Pearson, and G.F. Rohrmann. Immunoelectron microscopic examination of *Orygia pseudotsugata* multicapsid nuclear polyhedrosis virus–infected *Lymantria dispar* cells: time course and localization of major polyhedron–associated proteins. *Journal of General Virology*, 72:275–283, 1991.

[267] R.L.Q. Russell and G.F. Rohrmann. The p6.5 gene region of a nuclear polyhedrosis virus of *Orygia pseudotsugata*: DNA sequence and transcriptional analysis of four late genes. *Journal of General Virology*, 71:551–560, 1990.

[268] N. Saitou and M. Nei. The neighbor–joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4:406–425, 1987.

[269] J. Sambrook, E.F. Fritsch, and T. Maniatis. *Molecular Cloning: a laboratory manual*, volume 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1989.

[270] M.A. Samuels, A. Fire, and P.A. Sharp. Separartion and characterization of factors mediating accurate transcription by RNA polymerase II. *Journal of Biological Chemistry*, 257:14419–14427, 1982.

[271] M. Sawadogo and R.G. Roeder. Interaction of a gene-specific transcription factor with the Adenovirus major late promoter upstream of the TATA box region. *Cell*, 43:165–175, 1985.

[272] M. Sawadogo and A. Sentenac. RNA polymerase B (II) and general transcription factors. *Annual Review of Biochemistry*, 59:711–754, 1990.

[273] M.P. Schafer, G. Rohrmann, U. Heine, and G.S. Beaudreau. DNA from two *Orygia psuedotsugata* baculoviruses: molecular weight determination

by means of electron microscopy and restriction endonuclease analysis. *Virology*, 95:176–184, 1979.

[274] T. D. Schneider, G. D. Stromo, L. Gold, and A. Ehrenfeucht. Information content of binding sites on nucleotide sequences. *Journal of Molecular Biology*, 188:415–431, 1985.

[275] T.D. Schneider and R.M. Stephens. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Research*, 18:6097–6100, 1990.

[276] G.P. Schroth, J. M. Gottesfeld, and E. M. Bradbury. TFIIIA induced DNA bending: effect of low ionic strength electrophoresis buffers. *Nucleic Acids Research*, 19:511–516, 1991.

[277] G.P. Schroth, J. S. Sino, C. A. Cooney, J. P.H. Th'ng, P. S. Ho, and E. M. Bradbury. Intrinsically bent DNA flanks both sides of an RNA polymerase I transcription start site. *Journal of Biological Chemistry*, 267:9958–9964, 1992.

[278] D. Sen and W. Gilbert. Formation of parallel four–stranded complexes by guanine–rich motifs in DNA and its implications for meiosis. *Nature*, 334:364–334, 1988.

[279] P.A. Sharp. TATA–binding protein is a classless factor. *Cell*, 68:819–821, 1992.

[280] G. Shaw and R. Kamen. A conserved AU sequence from the 3' untranslated region of GM–CSF mediates selective mRNA degradation. *Cell*, 46:659–667, 1986.

[281] K.A. Simmen, J. Bernues, H.D. Parry, H.G. Stunnenberg, A. Berkenstam, B. Cavallini, J.M. Egly, and I.W. Mattaj. TFIID is required for *in vitro* transcription of the human U6 gene by RNA polymerase III. *EMBO Journal*, 10:1853–1862, 1991.

[282] S.T. Smale, M.C. Schmidt, A.J. Berk, and D. Baltimore. Transcriptional activation by Sp1 as directed through TATA or initiator: specific requirement for mammalian transcription factor IID. *Proceedings of the National Academy of Sciences USA*, 87:4509–4513, 1990.

[283] C.A. Smith, P.Bates, R. Rivera-Gonzalez, B. Gu, and N. A. DeLuca. ICP4, the major transcriptional regulatory protein of herpes simplex virus type 1, forms a tripartite complex with TATA–binding protein and TFIIB. *Journal of Virology*, 67:4676–4687, 1993.

[284] G.E. Smith and M.D. Summers. Analysis of baculovirus genomes with restriction endonucleases. *Virology*, 89:517–527, 1978.

[285] C. Spiro, J.P. Richards, S. Chandrasekaran, R.G. Brennan, and C.T. McMurray. Secondary structure creates mismatched base pairs required for high–affinity binding of cAMP response element–binding protein to the human enkephalin enhancer. *Proceedings of the National Academy of Sciences USA*, 90:4606–4610, 1993.

[286] K. Sprague. personal communication. 1994.

[287] H. Stahl, P. Droge, and R. Knippers. DNA helicase activity of SV40 large tumor antigen. *EMBO Journal*, 5:1939–1944, 1986.

[288] L.A. Stargell and M.A. Gorovsky. TATA–binding protein and nuclear differentiation in *Tetrahymena thermophila*. *Molecular and Cellular Biology*, 14:723–734, 1994.

[289] D.B. Starr and D.K. Hawley. TFIID binds in the minor groove of the TATA box. *Cell*, 67:1231–1240, 1991.

[290] K.F. Stringer, C.J. Ingles, and J. Greenblatt. Direct and selective binding of an acidic transcriptional activation domain to TATA–box factor TFIID. *Nature*, 345:783–86, 1990.

[291] S.L. Stringer, J.R. Stringer, M.A. Blase, P.D. Walzer, and M.T. Cushion. *Pneumocystis carinii*: sequence from ribosomal RNA implies a close relationship with fungi. *Experimental Parasitology*, 68:450–461, 1989.

[292] M. Strubin and K. Struhl. Yeast and human TFIID with altered DNA–binding specificty for TATA elements. *Cell*, 68:721–730, 1992.

[293] K. Struhl. Duality of TBP, the universal transcription factor. *Science*, 263:1103–1104, 1994.

[294] K. Sumita, Y. Makino, K. Katoh, T. Kishimoto, M. Muramatsu, K. Mikoshiba, and T. Tamura. Structure of a mammanlian TBP (TATA–binding protein) gene: isolation of the mouse TBP genome. *Nucleic Acids Research*, 21:2769, 1993.

[295] M.D. Summers and G.E. Smith. *A manual for methods for baculovirus vectors and insect cell culture procedures.* Texas Agricultural Experiment Station Bulletin no. 1555, 1987.

[296] D.L. Swofford and G.J. Olsen. *Molecular Systematics*, chapter 11, pages 411–501. Sinauer Associates Inc., 1990.

[297] T. Tamura, K. Sumita, I. Fujino, A. Aoyama, M. Horikoshi, A. Hoffman, R. Roeder, M. Muramatsu, and K. Mikoshiba. Striking homology of the variable N–terminal as well as the conserved core domains of the mouse and human TATA–factors (TFIID). *Nucleic Acids Research*, 19:3861–3865, 1991.

[298] I. Tanaka, K. Appelt, J. Dijk, S. White, and K. Wilson. 3–A resolution structure of a protein with histone–like properties in prokaryotes. *Nature*, 310:376–381, 1984.

[299] D. A. Theilmann and S. Stewart. Identification and characterization of the IE–1 gene of *Orygia psuedotsugata* multicapsid nuclear polyhedrosis virus. *Virology*, 180:492–508, 1991.

[300] D. A. Theilmann and S. Stewart. Tandemly repeated sequence at the 3' end of the IE–2 gene of the baculovirus *Orygia pseudotsugata* multicapsid nuclear polyhedrosis virus is an enhancer element. *Virology*, 187:97–106, 1992.

[301] S. Thiem and L.K. Miller. Differential gene expression mediated by late, very late and hybrid baculovirus promoters. *Gene*, 91:87–94, 1990.

[302] S.M. Thiem and L.K. Miller. Identification, sequence and transcriptional mapping of the major capsid protein gene of the baculovirus *Autographa californica* nuclear polyhedrosis virus. *Journal of Virology*, 63:2008–2018, 1989.

[303] J.W. Todd, A.L. Passarelli, and L.K. Miller. Eighteen baculovirus genes including *lef-11*, *p35*, *39k*, and *p47* support late gene expression. *Journal of Virology*, 69:968–974, 1995.

[304] M.D. Tomalski, J. Wu, and L.K. Miller. The location, sequence, transcription, and regulation of a baculovirus DNA polymerase gene. *Virology*, 167:591–600, 1988.

[305] F. Toneguzzo, S. Glynn, E. Levi, S. Mjolsness, and A. Hayday. Use of a chemically modified T7 DNA polymerase for manual and automated sequencing of supercoiled DNA. *Biotechniques*, 6:460–469, 1988.

[306] A. Usheva, E. Maldonado, A. Goldring, H. Lu, C. Houbavi, D. Reinberg, and Y. Aloni. Specific interaction between the non–phosphorylated form of RNA polymerase II and the TATA–binding protein. *Cell*, 69:871–881, 1992.

[307] M.M. van Oers, D. Malarme, J.M.P. Jore, and J.M. Vlak. Expression of the *Autographa californica* nuclear polyhedrosis virus p10 gene: effect of polyhedrin gene expression. *Archives of Virology*, 123:1—11, 1992.

[308] C.P. Verrijzer, J. A.W.M. van Oosterhout, W.W. van Weperen, and P. C. van der Vliet. POU proteins bend DNA via POU–specific domains. *EMBO Journal*, 10:3007–3014, 1991.

[309] J.M. Vlak, F. A. Klinkenberg, K. J.M. Zaal, M. Usmay, E. C. Kling-Roode, J.B.F. Geervliet, J. Roosien, and J.W.M. Van Lent. Functional studies on the p10 gene of *Autographa californica* nuclear polyhedrosis virus using a recombinant expressing p10–B–galactosidase fusion gene. *Journal of General Virology*, 69:765–776, 1988.

[310] J.M. Vlak and G.E. Smith. Orientation of the genome of *Autographa californica* nuclear polyhedrosis virus: a proposal. *Journal of Virology*, 41:1118–1121, 19982.

[311] J.M. Vogel, B. Roth, M. Cigan, and M. Freeling. Expression of the two maize TATA binding protein genes and function of the encoded TBP proteins by complementation in yeast. *Plant Cell*, 5:1627–1638, 1993.

[312] B. Vogelstein and D. Gillespie. Preparative and analytical purification of DNA from agarose. *Proceedings of the National Academy of Sciences USA*, 76:615–619, 1979.

[313] L.E. Volkman, M.D. Summers, and C. Hsieh. Occluded and nonoccluded nuclear polyhedrosis virus grown in *Trichoplusia ni*: comparative neutralization, comparative infectivity, and in vitro growth studies. *Journal of Virology*, 19:820–832, 1976.

[314] A.V. Vologodskii and M.D. Frank-Kamenetskii. Theoretical study of cruciform states in superhelical DNAs. *FEBS Letters*, 143:257–260, 1982.

[315] P.O. Wainright, G. Hinkle, M.L. Sogin, and S.K. Stickel. Monophyletic origins of the metazoa: An evolutionary link with fungi. *Science*, 260:340–342, 1993.

[316] A. H. J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel, and A. Rich. Molecular structure of a left–handed double helical DNA fragment at atomic resolution. *Nature*, 282:680–686, 1979.

[317] R.O.J. Weinzierl, S. Ruppert, B.D. Dynlacht, N. Tanese, and R. Tjian. Cloning and expression of Drosophila TAFII60 and human TAFII70 reveal conserved interactions with other subunits of TFIID. *EMBO Journal*, 12:5303–5309, 1993.

[318] L. Weis and D. Reinberg. Transcription by RNA polymerase II: initiator–directed formation of transcription–competent complexes. *FASEB Journal*, 6:3300–3309, 1992.

[319] U. Weyer and R. D. Possee. Analysis of the promoter of *Autographa californica* nuclear polyhedrosis virus p10 gene. *Journal of General Virology*, 70:203–208, 1989.

[320] R.J. White and S. P. Jackson. The TATA-binding protein: a central role in transcription by RNA polymerase I, II and III. *Trends in Genetics*, 8:284–288, 1992.

[321] M.A. Whitt and J. S. Manning. A phosporylated 34–kDa protein and a subpopulation of polyhedrin are thiol linked to the carbohydrate layer surrounding a baculovirus occlusion body. *Virology*, 163:33–42, 1988.

[322] D.L. Williams and D. Kowalski. Easily unwound DNA sequences and hairpin structures in the epstein—barr virus origin of plasmid replication. *Journal of Virology*, 67:2707–2715, 1993.

[323] G.V. Williams, D.Z. Rohel, J. Kuzio, and P. Faulkner. A cytopathological investigation of *Autographa californica* nuclear polyhedrosis virus p10 gene function using insertion/deletion mutants. *Journal of General Virology*, 70:187–202, 1989.

[324] M.E. Wilson, T.H. Mainprize, P. D. Friesen, and L.K. Miller. Location, transcription and sequence of a baculovirus gene encoding a small arginine-rich polypeptide. *Journal of Virology*, 61:661–666, 1987.

[325] J. Wong, F. Liu, and E. Bateman. Cloning and expression of the *Acanthameoba castellani* gene encoding transcription factor TFIID. *Gene*, 117:91–97, 1992.

[326] R. J. Wootton. Palaeozoic insects. *Annual Review of Entomology*, 26:319–344, 1981.

[327] J.L. Workman, S. Abmayr, W. A. Cromlish, and R.G. Roeder. Transcriptional regulation by the immediate early protein of psuedorabies virus during in vitro nucleosome assembly. *Cell*, 55:211–219, 1988.

[328] J.L. Workman, R.G. Roeder, and R.E. Kingston. An upstream transcription factor, USF (MLTF), facilitates the formation of preinitiation complexes during in vitro chromatin assembly. *EMBO Journal*, 9:1299–1308, 1990.

[329] N. Woychik, S.M. Liao, P. Kolodzeij, and R.A. Young. Subunits shared by eukaryotic nuclear RNA polymerases. *Genes and Development*, 4:313–323, 1990.

[330] L. Wu, D.S.E. Rosser, M. Schmidt, and A.J. Berk. A TATA–box implicated in E1a transcriptional activation of a simple adenovirus 2 promoter. *Nature*, 326:512–515, 1987.

[331] B. Xu, S. Yoo, and L. A. Guarino. Differential transcription of baculovirus late and very late promoters: Fractionation of nuclear extracts by phosphcellulose chromatography. *Journal of Virology*, 69:2912–2917, 1995.

[332] C.C. Yang and H.A. Nash. The interaction of *E. coli* IHF protein with its specific binding sites. *Cell*, 57:869–880, 1989.

[333] C.L. Yang, D.A. Stetler, and R.F. Weaver. Structural comparison of the *Autographa californica* nuclear polyhedrosis virus–induced RNA polymerase and three nuclear RNA polymerases from the host, *Spodoptera frugiperda*. *Virus Research*, 20:251–264, 1991.

[334] S. Yoo and L. A. Guarino. The *Autographa californica* nuclear polyhedrosis virus *ie2* gene encodes a transcriptional regulator. *Virology*, 202:746–753, 1994.

[335] P.M.A. Zanotto, M.J.A. Sampaio, D.W. Johnson, T.L. Rocha, and J.E. Maruniak. The *Anticarsia gemmatalis* nuclear polyhedrosis virus polyhedrin gene region: sequence analysis, gene product and structural comparisons. *Journal of General Virology*, 73:1049–1056, 1992.

[336] L. Zawel and D. Reinberg. Advances in RNA polymerase II transcription. *Current Opinion in Cell Biology*, 4:488—495, 1992.

[337] B. Zenzi-Gregory, A. Khachi, I.P. Garraway, and S.T. Smale. Mechanism of initiator–mediated transcription: evidence for a functional interaction between TATA–binding protein and DNA in the absence of a specific recognition sequence. *Molecular and Cellular Biology*, 13:3841–3849, 1993.

[338] J. Zhu, P.W.Rice, and M. Chamberlain. Mapping the transcriptional transactivation function of SV40 large T antigen. *Journal of Virology*, 65:2778–2790, 1991.

[339] M. Zucker. On finding all suboptimal folding of an RNA molecule. *Science*, 244:48–52, 1989.

APPENDICES

Appendix A
Alignment of TBP Coding Regions

```
                    1         11         21         31         41         50
                    |         |          |          |          |          |
M.auratus           MDQNNSLPPY AQGLASPQGA MTPGIPIFSP MMPYGTGLTP QPIQNTNSLS
M.musculus          MDQNNSLPPY AQGLASPQGA MTPGIPIFSP MMPYGTGLTP QPIQNTNSLS
H.sapiens           MDQNNSLPPY AQGLASPQGA MTPGIPIFSP MMPYGTGLTP QPIQNTNSLS
X.laevis            MDQNNSIPPF -QGLASPQGS LTPGINIFSP LITYGTGLTP QPVQTTNSLS
B.mori              MDHMLPSPYN IPGIGTPLHQ P-------EE DQQILP---- ----------
S.frugiperda        MDQMLPSPYN IPGIDTPLHQ P-------EE DQQILP---- ----------
D.melanogaster      MDQMLSPNFS IPSIGTPLHQ -----MEADQ QIVANPVYHP PAVSQPDSLM
O.volvulus          ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
C.elegans           MN-LNSPAVS MLGGDTPAHG GPNSVLGGQG PSSILTGHGP NSVMGPNSIL
A.cliftonii         MTG------- ---------- ---------- ---------- ----------
Z.mays              MAE------- ---------- --PG------ ---------- ----------
S.tuberosum         MAD------- ---------- --QG------ ---------- ----------
A.thaliana          MAD------- ---------- --QG------ ---------- ----------
T.aestivum          MAA------- ---------- --AAVDPM-- VL-------- -------GLG
S.pombe             MD-------- ---------- ---------- ---------- ----------
S.cerevisiae        MAD------- ---------- --E--ERLKE FK-------- ----EANKIV
P.carinii           MA-------- ---------- ---------- ---------- ----------
A.castellanii       MSG------- ---------- --ITLPSLTN VL-------- ----QSAGMA
D.discoideum        MST------- ---------- --AT------ ---------- ----------
T.thermophila       MSSDKTSQQT FK-------- ---------- -------LAP NNSVASNSI-
E.histolytica       MS-------- ---------- -TPGDFSLSP FI-LGGAVDP RSMSQ-----
P.falciparum        MNFLEQDQLF LENINQDN-- ---------- ---------- ----------
T.celer             M---SNVKL- ---------- ---------- ---------- ----------
P.woesei            MVDMSKVKL- ---------- ---------- ---------- ----------

                   51         61         71         81         91        100
                    |         |          |          |          |          |
M.auratus           ILEEQQREQQ QQQ------- ---------- ------QQQQ QQQQQAVATA
M.musculus          ILEEQQR--Q QQQ------- ---------- ------QQQQ QQQQQAVATA
H.sapiens           ILEEQQRQQQ QQQQQQQQQQ QQQQ--QQQQ QQQQQQQQQQ QQQQQQQAVA
X.laevis            ILEEQQRQQQ Q--------- ---------- ---------- ----------
B.mori              ----NAMQQQ QLQQQQSQ-- -------AQP SLAALG---S SPIVGFGA-I
S.frugiperda        ----NAMQQQ H-QHQQQQ-- -------QQH ALAAMG---S SPLVGFGASL
D.melanogaster      PAPGSSSVQH QQQQQQSDAS GGSGLFGHEP SLPLAHKQMQ SYQPS-ASYQ
O.volvulus          ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
C.elegans           -GPGSVLNPQ SIQPMQSQ-- -------QMH SLQGSSMQMH SHLANSNLNL
A.cliftonii         ---------- ---------- ---------- ---------- ----------
Z.mays              ---------- ---------- ---------- ---------- ----------
S.tuberosum         ---------- ---------- ---------- ---------- ----------
```

```
A.thaliana      ---------- ---------- ---------- ---------- ----------
T.aestivum      T--------- ---------- ---------- ---------- ----SGGASG
S.pombe         ---------- ---------- ---------- ---------- ----------
S.cerevisiae    F--------- ---------- ---------- ---------- ----D-----
P.carinii       ---------- ---------- ---------- ---------- ----------
A.castellanii   V--------- ---------- ---------- ---------- ----HGHPSA
D.discoideum    ---------- ---------- ---------- ---------- ----------
T.thermophila   ---------- ------DQNK NKNNILSTIE TMDKSI---- ----------
E.histolytica   ---------- ---------- ---------- ---------- ----------
P.falciparum    VVSAHYTSEY DNNEKEKSDD LKNKLVHKNI SLN------- ----------
T.celer         ---------- ---------- ---------- ---------- ----------
P.woesei        ---------- ---------- ---------- ---------- ----------

                101        111        121        131        141       150
                 |          |          |          |          |         |
M.auratus       AASVQQSTSQ QSTQGASGQT PQLFHSQT-- --LTTAPLPG TTPLY-----
M.musculus      AASVQQSTSQ QPTQGASGQT PQLFHSQT-- --LTTAPLPG TTPLY-----
H.sapiens       AAAVQQSTSQ QATQGTSGQA PQLFHSQT-- --LTTAPLPG TTPLY-----
X.laevis        ---TQQSTLQ QGNQG-SGQT PQLFHPQT-- --LTTAPLPG NTPLY-----
B.mori          MGTPQRSMHT YAPTASYATP QQMMQPQTP- QNMMSPMIAA G-NLSSQ---
S.frugiperda    MGTPQRSVHT YAPAASYATP QQMMQPQTP- QNLMSPMITS G-SLAGQ---
D.melanogaster  QQQQQQQLQS QAPGGGGSTP QSMMQPQTP- QSMMAHMMPM SERSVGGSGA
O.volvulus      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
C.elegans       NINPASVGPD RNPGSVMHHN LDINPPSVAY QNLTVPMTPL AYSVYDR---
A.cliftonii     ---------- ---------- ---------- ---------- ----------
Z.mays          ---------- ---------- ---------- -------LED SQP-------
S.tuberosum     ---------- ---------- ---------- -------LEG SQP-------
A.thaliana      ---------- ---------- ---------- -------TEG SQP-------
T.aestivum      SGVVG----- ----GGVGRA GGGG------ -----AVMEG AQP-------
S.pombe         ---FALPTTA SQASAFMNNS SLTF------ -----PVLPN ANNEA-----
S.cerevisiae    ------PNTR QVWENQNRDG TKPA------ -----TTFQS EEDIK-----
P.carinii       ------PSSL SFPSSHILMS GAMY------ -----PGSRD EKGME-----
A.castellanii   PGSTQLPPLH QLNISSQPSS QPPQ------ -----PSLQY SEPAQ-----
D.discoideum    ---------- ---------- ---------- ---------- ----------
T.thermophila   ---------- ---------- -----SEDLY PKL------- ----------
E.histolytica   ---------- ---------- ---------- ---------- ----------
P.falciparum    ---------- ---------- ---------- ---------- ----------
T.celer         ---------- ---------- ---------- ---------- ----------
P.woesei        ---------- ---------- ---------- ---------- ----------

                151        161        171        181        191       200
                 |          |          |          |          |         |
M.auratus       ---------- PSPMTPMTPI TPATPASESS GIVPQLQNIV STVNLGCKLD
M.musculus      ---------- PSPMTPMTPI TPATPASESS GIVPQLQNIV STVNLGCKLD
H.sapiens       ---------- PSPMTPMTPI TPATPASESS GIVPQLQNIV STVNLGCKLD
```

```
X.laevis          ---------- PSPITPMTPI SPATPASESS GIVPQLQNIV STVNLGCKLD
B.mori            ----QMLSQA SPAP-----M TPLTPLSADP GILPQLQNIV STVNLDCKLD
S.frugiperda      ----QMLSQA SPAP-----M TPMTPHSADP GIVPQLQNIV STVNLNCKLD
D.melanogaster    GGGGDALSNI HQTMGPSTPM TPATPGSADP GIVPQLQNIV STVNLCCKLD
O.volvulus        ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ VPSPALQNIV STVNLGVPLD
C.elegans         ----DALTHQ APASNIAATM VPATPASQLD IPMPALQNIV STVNLGVQLD
A.cliftonii       ---------- ---------- DVDMSL-HPS GIIPELQNVV STVNLGCTLE
Z.mays            ---------- ---------- -VDLS-KHPS GIVPTLQNIV STVNLDCKLD
S.tuberosum       ---------- ---------- -VDLT-KHPS GIVPTLQNIV STVNLDCKLD
A.thaliana        ---------- ---------- -VDLT-KHPS GIVPTLQNIV STVNLDCKLD
T.aestivum        ---------- ---------- -VDLA-RHPS GIVPVLQNIV STVNLDCRLD
S.pombe           ---------- TNETADSGDA EVSKN-EGVS GIVPTLQNIV ATVNLDCRLD
S.cerevisiae      ---------- RAAPESEKDT SAT------S GIVPTLQNIV ATVTLGCRLD
P.carinii         ---------- HGVVSTSLNQ TATNTFAGVS GIVPTLQNIV ATVNLDCRLD
A.castellanii     ---------- STAASDDMDS DVDRT-KHPS GIVPTLQNIV STVNLGCKLD
D.discoideum      ---------- ---TTSTPAQ NVDLS-KHPS GIIPTLQNIV STVNMATELY
T.thermophila     ---------- ---------- ---------- -------NIV STVNLSTKLD
E.histolytica     ------LGNI CHAVICQLQL SHKKVLIIQT ITHPEIVNVV SRFQLGVKLE
P.falciparum      ---------- ---------- ---------- -----IHNII SSANLCIDIN
T.celer           ---------- ---------- ---------- ----RIENIV ASVDLFTQLN
P.woesei          ---------- ---------- ---------- ----RIENIV ASVDLFAQLD

                  201        211        221        231        241        250
                    |          |          |          |          |          |
M.auratus           LKTIALRARN AEYNPKRFAA VIMRIREPRT TALIFSSGKM VCTGAKSEEQ
M.musculus          LKTIALRARN AEYNPKRFAA VIMRIREPRT TALIFSSGKM VCTGAKSEEQ
H.sapiens           LKTIALRARN AEYNPKRFAA VIMRIREPRT TALIFSSGKM VCTGAKSEEQ
X.laevis            LKTIALRARN AEYNPKRFAA VIMRIREPRT TALIFSSGKM VCTGAKSEEQ
B.mori              LKKIALHARN AEYNPKRFAA VNMRIREPRT TALIFSSGKM VCTGAKSEED
S.frugiperda        LKKIALHARN AEYNPKRFAA VIMRIREPRT TALIFSSGKM VCTGAKSEED
D.melanogaster      LKKIALHARN AEYNPKRFAA VIMRIREPRT TALIFSSGKM VCTGAKSEDD
O.volvulus          LKKIALHARN AEYNPKRFAA VIMRIREPRT TALIFSSGKM VCTGAKSEES
C.elegans           LKKIALHARN AEYNPKRFAA VIMRIREPRT TALIFSSGKM VCTGAKSEEA
A.cliftonii         LKEIAMQARN AEYNPKRFAA VIMRIRDPKT TALIFGSGKM VCTGAKSED-
Z.mays              LKAIALQARN AEYNPKRFAA VIMRIREPKT TALIFASGKM VCTGAKSEQQ
S.tuberosum         LKAIALQARN AEYNPKRFAA VIMRIREPKT TALIFASGKM VCTGAKSEQQ
A.thaliana          LKAIALQARN AEYNPKRFAA VIMRIREPKT TALIFASGKM VCTGAKSEHL
T.aestivum          LKQIALQARN AEYNPKRFAA VIMRIRDPKT TALIFASGKM VCTGAKSEEH
S.pombe             LKTIALHARN AEYNPKRFAA VIMRIREPKS TALIFASGKM VVLGGKSEDD
S.cerevisiae        LKTVALHARN AEYNPKRFAA VIMRIREPKT TALIFASGKM VVTGAKSEDD
P.carinii           LKTIALHARN AEYNPKRFAA VIMRIREPKT TALIFASGKM VVTGAKSEDD
A.castellanii       LKNIALHARN AEYNPKRFAA VIMRIREPKT TALIFASGKM VCTGAKSEEA
D.discoideum        LKAIALGARN AEYNPKRFAA VIMRIREPKT TALIFKSGKM VCTGAKSEDA
T.thermophila       LKQIALRARN AEYNPKRFAA VIMRLRDPKT TALIFASGKM VCTGAKTEED
E.histolytica       LRKIVQKAIN AIYNPKRFAG AIMRISSPKS TALIFQTGKI VCTGTRSIEE
```

```
P.falciparum     LRLVAVSIRN AEYNPSKINT LIIRLNKPQC TALIFKNGRI MLTGTRTKKD
T.celer          LERVIEMCPH SKYNPEEFPG IICRFDEPKV ALLIFSSGKL VVTGAKSVED
P.woesei         LEKVLDLCPN SKYNPEEFPG IICHLDDPKV ALLIFSSGKL VVTGAKSVQD

                251        261        271        281        291       300
                 |          |          |          |          |         |
M.auratus        SRLAARKYAR VVQKLGFP-A KFL---DFKI QNMVGSCDVK FPIRLEGLVL
M.musculus       SRLAARKYAR VVQKLGFP-A KFL---DFKI QNMVGSCDVK FPIRLEGLVL
H.sapiens        SRLAARKYAR VVQKLGFP-A KFL---DFKI QNMVGSCDVK FPIRLEGLVL
X.laevis         SRLAARKYAR VVQKLGFP-A KFL---DFKI QNMVGSCDVK FPIRLEGLVL
B.mori           SRLAARKYAR IIQKLGFT-A KFL---DFKI QNMVGSCDVK FPIRLEGLVL
S.frugiperda     SRLAARKYAR IIQKLGFT-A KFL---DFKI QNMVGSCDVK FPIRLEGLVL
D.melanogaster   SRLAARKYAR IIQKLGFP-A KFL---DFKI QNMVGSCDVK FPIRLEGLVL
O.volvulus       SRLAARKYAR IVQKLGFN-A KFT---EFKV QNMVGSCDVR FPIQLEGLCL
C.elegans        SRLAARKYAR IVQKLGFQ-A KFT---EFMV QNMVGSCDVR FPIQLEGLCI
A.cliftonii      SRTAARKYAK IVQKLGFP-A KFT---EFKI ---------- ----------
Z.mays           SKLAARKYAR IIQKLGFP-A KFK---DFKI QNIVGSCDVK FPIRLEGLAY
S.tuberosum      SKLAARKYAR IIQKLGFP-A KFK---DFKI QNIVGSCDVK FPIRLEGLAY
A.thaliana       SKLAARKYAR IVQKLGFP-A KFK---DFKI QNIVGSCDVK FPIRLEGLAY
T.aestivum       SKLAARKYAR IVQKLGFP-A TFK---DFKI QNIVASCDVK FPIRLEGLAY
S.pombe          SKLASRKYAR IIQKLGFN-A KFT---DFKI QNIVGSCDVK FPIRLEGLAY
S.cerevisiae     SKLASRKYAR IIQKIGFA-A KFT---DFKI QNIVGSCDVK FPIRLEGLAF
P.carinii        SKLASRKYAR IIQKLGFN-A KFT---DFKI QNIVGSCDVK FPIRLEGLAY
A.castellanii    SRLAARKYAR IIQKLGFA-A KFL---DFKI QNIVGSCDVR FPIRLEGLAF
D.discoideum     SRFAARKYAR IIQKLDFP-A RFT---DFKI QNIVGSCDVK FPIKLELLHN
T.thermophila    SNRAARKYAK II-------- ---------- ---------- ----------
E.histolytica    SKIASKKYAK IIKKIGYP-I HYS---NFNV QNIVGSCDVK FQIALRTLVD
P.falciparum     SIMGCKKIAK IIKIVTKDKV KFC---NFKI ENIIASANCN IPIRLEVLAH
T.celer          IERAVNKLIQ MLKKIG---A KFSRAPQIDI QNMVFSGDIG MEFNLDAVAL
P.woesei         IERAVAKLAQ KLKSIG---V KFKRAPQIDV QNMVFSGDIG REFNLDVVAL

                301        311        321        331        341       350
                 |          |          |          |          |         |
M.auratus        THQQFSSYEP ELFPGLIYR- --MIKPRIVL LIFVSGKVVL TGAKVRAEIY
M.musculus       THQQFSSYEP ELFPGLIYR- --MIKPRIVL LIFVSGKVVL TGAKVRAEIY
H.sapiens        THQQFSSYEP ELFPGLIYR- --MIKPRIVL LIFVSGKVVL TGAKVRAEIY
X.laevis         THQQFSSYEP ELFPGLIYR- --MIKPRIVL LIFVSGKVVL TGAKVRAEIY
B.mori           THGQFSSYEP ELFPGLIYR- --MVKPRIVL LIFVSGKVVL TGAKVREEIY
S.frugiperda     THGQFSSYEP ELFPGLIYR- --MVKPRIVL LIFVSGKVVL TGAKVREEIY
D.melanogaster   THCNFSSYEP ELFPGLIYR- --MVRPRIVL LIFVSGKVVL TGAKVRQEIY
O.volvulus       THTQFSTYEP ELFPGLIYR- --MVKPRVVL LIFVSGKVVI TGAKYKKDID
C.elegans        THSQFSTYEP ELFPGLIYR- --MVKPRVVL LIFVSGKVVI TGAKTKRDID
A.cliftonii      ----FCSYEP ELFPGLIYR- --MLQPKIVL LIFVSGKVVL TGAKERTEIY
Z.mays           SHGAFSSYEP ELFPGLIYR- --MKQPKIVL LIFVSGKIVL TGAKVREETY
S.tuberosum      AHGAFSSYEP ELFPGLIYR- --MKQPKIVL LIFVSGKIVI TGAKVRDETY
```

```
A.thaliana        SHSAFSSYEP ELFPGLIYR- --MKLPKIVL LIFVSGKIVI TGAKMREETY
T.aestivum        SHGAFSSYEP ELFPGLIYR- --MKQPKIVL LVFVSGKIVL TGAKVRDEIY
S.pombe           SHGTFSSYEP ELFPGLIYR- --MVKPKVVL LIFVSGKIVL TGAKVREEIY
S.cerevisiae      SHGTFSSYEP ELFPGLIYR- --MVKPKIVL LIFVSGKIVL TGAKQREEIY
P.carinii         SHGTFSSYEP ELFPGLIYR- --MVKPKIVL LIFVSGKIVL TGAKVREEIY
A.castellanii     AHNHYCSYEP ELFPGLIYR- --MVQPKIVL LIFVSGKIVL TGAKVREEIY
D.discoideum      AHTSFTNYEP EIFPGLIYK- --MIQPKVLL LIFVSGKIVL TGAKVREYIY
T.thermophila     -------YEP EIFPGKIYR- --EFNTKIVL LIFVSGKIVL TGAKTRENIN
E.histolytica     SDLAFCQYEP EVFPGLVYR- --MASPKVTL LVFSTGKVVL TGAKDEESLN
P.falciparum      DHKEYCNYEP ELFAGLVYRY KPTSNLKSVI LIFVSGKIII TGCKSVNKLY
T.celer           SLPN-CEYEP EQFPGVIYR- --VKEPRAVI LLFSSGKIVC SGAKSEHDAW
P.woesei          TLPN-CEYEP EQFPGVIYR- --VKEPKSVI LLFSSGKIVC SGAKSEADAW

                351        361        371        381
                  |          |          |          |
M.auratus         EAFENIYPIL KGFRKTTWLP CPASPTHLFF KASQFWYHW*
M.musculus        EAFENIYPIL KGFRKTT--- ---------- ---------*
H.sapiens         EAFENIYPIL KGFRKTT--- ---------- ---------*
X.laevis          EAFENIYPIL KGFRKTT--- ---------- ---------*
B.mori            EAFDNIYPIL KSFKK----- ---------- -------Q-*
S.frugiperda      EAFDNIYPIL KSFKK----- ---------- -------Q-*
D.melanogaster    DAFDKIFPIL KKFKK----- ---------- -------QS*
O.volvulus        DAFNQIYPIL KGFKK----- ---------- ---------*
C.elegans         EAFGQIYPIL KGFKK----- ---------- ---------*
A.cliftonii       RAFEQIYPVL TQFRK----- ---------- --------R*
Z.mays            TAFENIYPVL AEFRKV---- ---------- -------QQ*
S.tuberosum       TAFENIYPVL TEFRKN---- ---------- -------QQ*
A.thaliana        TAFENIYPVL REFRKV---- ---------- -------QQ*
T.aestivum        AAFENIYPVL TEYRKS---- ---------- -------QQ*
S.pombe           QAFEAIYPVL SEFRKH---- ---------- ---------*
S.cerevisiae      QAFEAIYPVL SEFRKM---- ---------- ---------*
P.carinii         QAFEAIYPVL SEFRKSS--- ---------- ---------*
A.castellanii     EAFENIYPVL TEYKKT---- ---------- ---------*
D.discoideum      EAFENIYPVL SAFKKVN--- ---------- -AIT----Q*
T.thermophila     KAF------- ---------- ---------- ---------*
E.histolytica     LAYKNIYPIL LANRKED--- ---------- ------ISN*
P.falciparum      TVFQDIYNVL IQYKN----- ---------- ---------*
T.celer           EAVRKLLREL EKY------- ------DLIG EGEEEW---*
P.woesei          EAVRKLLREL DKY------- ------GLLE EEEEE----*
```

Appendix B
IUB Codes

IUB / GCG	Meaning	Complement
A	A	T
C	C	G
G	G	C
T / U	T	A
M	A or C	K
R	A or G	Y
W	A or T	W
S	C or G	S
Y	C or T	R
K	G or T	M
V	A or C or G	B
H	A or C or T	D
D	A or G or T	H
B	C or G or T	V
X / N	G or A or T or C	X
.	not (G or A or T or C)	.

Appendix C

AcMNPV *Hr* Alignment

```
        1          11         21         31         41         51     58
        |          |          |          |          |          |      |
 hr5-1  atgatgtcat tt-gtttt-- -------t-- aaaattga-a ctggctttac gagtagaa
 hr2-4  atgatgtcat tt-gtttt-- -------tt- aaaattga-a ctggctttac gagtagaa
 hr2-5  atgatgtcat tt-gtttt-- -------tt- aaaattga-a ctggctttac gagtagaa
 hr2-8  atgatgtcat tt-gtttt-- -------tc- aaaactaa-a ctcgctttac gagtagaa
 hr2-6  atga-gtcat tt-gtttt-- -------tc- aaaactaa-a ctcgctttac gagtagaa
hr4b-1  atgatgtcat tt-g-ttt-- -------tc- aaaattga-a ctggctttac gagtagaa
hr1a-1  atga-gtt-t tt-gtcgt-- ---------- aaaaatgcca cttgttttac gagtagaa
hr4b-4  atgatgtcat tt-gtttt-- -------tc- aaaaccga-a ctcgctttac gagtagaa
 hr2-7  atga-ctcat tt-gtttt-- -------tc- aaaactga-a ctcgctttac gagtagaa
 hr2-3  atgatgtcat tt-gtttt-- -------tc- aaaaccga-a ctcgctttac gagtagaa
hr4b-3  atgatgtcat tt-gtttt-- -------ttt aaaattca-a ctcgctttac gagtagaa
 hr5-5  atgatgtcat tt-gtttt-- -------tc- aaaactga-a ctcgctttac gagtagaa
 hr2-2  atgatgtcat tt-gtttt-- -------tc- aaaactga-a ctggctttac gagtagaa
 hr5-4  atga-ctcat tt-gtttt-- -------tc- aaaaccga-a cttgatttac gggtagaa
hr1a-2  atga-ctcat tt-gtttt-- ---------- aaaacag--a cttgttttac gagtagaa
 hr2-1  atgtcatc-- ---gtt---- ---------- ----cta--a ctcgctttac gagtagaa
hr4b-2  atgtcatc-- ---gta---- ---------- ----c-aa-a ctcgctttac gagtagaa
hr4b-5  atgtcatc-- ---gta---- ---------- ----c-aa-a ctcgctttac gagtagaa
 hr5-2  atgtcat--- ---gttttgt acacggctc- ataaccga-a ctggctttac gagtagaa
 hr5-3  atgtcat--- ---gttttgc acacggctc- ataa-----a ctcgctttac gagtagaa
 hr5-6  atgtcat--- ---gttttgc acatggctc- ataactaa-a ctcgctttac gggtagaa
 hr1-5  a-gtcata-- ---att---- ---------- ------aa-t cgtgcgttac aagtagaa
 hr3-1  atgacatcat tcc------- ---------- ----g-gat- catgatttac gcgtagaa
 hr1-3  atgacatcat cca------- ---------- ----ctgat- cgtgcgttac aagtagaa
 hr1-4  atgac-tcat act------- ---------- -----tgat- tgtgttttac gcgtagaa
hr4a-1  atgac-tcat taa------- ---------- ----tcgat- cgtgcgttac aagtagaa
hr4a-2  atgacatcat ccg------- ---------- ----acgat- tgtgttttac aagtagaa
 hr3-3  atgacatcat cca------- ---------- ----ctaat- cgtgcgttac aagtagaa
 hr3-5  atgacatcat ttc------- ---------- ----ttgat- tatgttttac aagtagaa
 hr3-6  atgacatcat ctc------- ---------- ----ttgat- tatgttttac aagtagaa
 hr3-4  atgacatcat ttc------- ---------- ----ttgat- tgtgttttac acgtagaa
 hr1-1  atgacattat ccc------- ---------- ----tcgat- tgtgttttac aagtagaa
 hr3-2  atgac-taat aa-------- ---------- ----ttgat- cgtgcgttac aagtagaa
 hr1-2  atgacatcat ccc------- ---------- ----ctgat- tgtgttttac aagtagaa
 hr3-7  atgacatcat ccc------- ---------- ----ttgat- catgcgttac aagtagaa

       59         69         79         89         99        109    116
        |          |          |          |          |          |      |
 hr5-1  ttctacgcgt aaaacacaat c-aagt---- ---------- ---------- -atga-gt
```

```
 hr2-4  ttctacttgt aaaacacaat c-gagag--- ---------- ---------- -atgatgt
 hr2-5  ttctacttgt aaaacacaat c-gagag--- ---------- ---------- -atgatgt
 hr2-8  ttctacgtgt aaaacacaat c-aaggg--- ---------- ---------- -atgatgt
 hr2-6  ttctacttgt aacgcacgcc c-aaggg--- ---------- ---------- -atgatgt
hr4b-1  ttttacttgt aaaacacaat c-aagaa--- ---------- ---------- -atgatgt
hr1a-1  ttctacgtgt aacacacgat ctaaaag--- ---------- ---------- -atgatgt
hr4b-4  ttctacttgt aacgcaagat c-ggtgg--- ---------- ---------- -atgatgt
 hr2-7  ttctacttgt aaaacacaat c-aagcg--- ---------- ---------- -atgatgt
 hr2-3  ttctacttgt aaaacataat c-gaaag--- ---------- ---------- -atgatgt
hr4b-3  ttctacttgt aaaacacaat c-gaggg--- ---------- ---------- -atgatgt
 hr5-5  ttctacgtgt aaaacacaat c-aagaa--- ---------- ---------- -atgatgt
 hr2-2  ttctacttgt aacgcatgat c-aaggg--- ---------- ---------- -atgatgt
 hr5-4  ttctacttgt aaagcacaat caaaaag--- ---------- ---------- -atgatgt
hr1a-2  ttctacgtgt aaagcatgat c
 hr2-1  ttctacgtgt aaaacataat c-aagag--- ---------- ---------- -atgatgt
hr4b-2  ttctacttgt aacgcatgat c-aaggg--- ---------- ---------- -atgatgt
hr4b-5  ttctacgtgt aaaacacgat
 hr5-2  ttctacttgt aatgcacgat c-agtgg--- ---------- ---------- -atgatgt
 hr5-3  ttctacgtgt aacgcacgat c-gattg--- ---------- ---------- -atga-gt
 hr5-6  ttctacgcgt aaaacatgat
 hr1-5  ttctactcgt aaagcgagtt g-aagg
 hr3-1  ttctacttgt aaagcaagtt a-aa---ata ag------cc gtgtgcaaaa -atga---
 hr1-3  ttctactcgt aaagccagtt c-gg---tta tga----gcc gtgtgcaaaa catga---
 hr1-4  ttctactcgt aaagcgagtt c-gg---tta tga----gcc gtgtgcaaaa catga---
hr4a-1  ttctactggt aaagcaagtt c-gg---ttg tga----gcc gtgtgcaaaa catga---
hr4a-2  ttctactcgt aaagcgagtt
 hr3-3  ttctactcgt aaagcgagtt c-gg---ttt tgaaaaa--- -----caa-- -atga---
 hr3-5  ttctactcgt aaagcaagtt t-ag---ttt taaaaaa--- -----caa-- -atga---
 hr3-6  ttctactcgt aaagcgagtt t-ag---ttt tgaaaaa--- -----caa-- -atga---
 hr3-4  ttctactcgt aaagtatgtt c-ag---ttt -aaaaaa--- -----caa-- -atga---
 hr1-1  ttctacccgt aaagcgagtt t-ag---ttt tgaaaaa--- -----caa-- -atga---
 hr3-2  ttctactcgt aaagcgagtt t-ag---ttt tgaaaaa--- -----caa-- -atga-gt
 hr1-2  ttctatccgt aaagcgagtt c-ag---ttt tgaaaac--- ------aa-- -atga-gt
 hr3-7  ttctactcgt aaagcgagtt g-aa---ttt tg
```