CHAPTER III

A NOVEL Sp-LIKE CIS-SEQUENCE IS PRESENT UPSTREAM OF THE AcNPV polh GENE PROMOTER

1. An upstream 766 bp region enhances transcription from Ppolh

2. AcSp binds to Sp family-like factor(s) in Sf9 cells

3. The Sp-like factors are distinct from PPBP, Sp1, Sp3 and Sp4

4. Human Sp1 protein is capable of binding AcSp and cSp

5. DNase I footprinting analyses with Sf9 nuclear extract show a weak protection of AcSp

6. Discussion
A NOVEL Sp-LIKE CIS-SEQUENCE IS PRESENT UPSTREAM OF THE AcNPV polh GENE PROMOTER

As described in Chapter I, an upstream ~766 bp region was shown to rescue basal levels of transcription even when the polh promoter was mutated (Fig. 1.4). Bal31 deletion analyses of the 4kb upstream region also identified this ~766 bp stretch (Region II) as playing a role in enhancement of transcription from Ppolh. Region II, which enhances basal transcription, contains an open reading frame, ORF603, encoding a putative 201 amino acid protein product. However, the enhancement is not mediated through its protein product as conclusively demonstrated by deletion experiments (Gearing and Possee, 1990) within this region. The enhancement caused by this sequence therefore pointed to the possibility of this sequence per se being essential and not the putative protein product, probably by binding certain transcription factors which could subsequently effect transcription enhancement. The 766 bp region (containing ORF603) was thus analysed to determine if there were binding sites for any known eukaryotic transcription factors and to demonstrate the existence of such factor(s) if any.

1. An upstream 766 bp region enhances transcription from Ppolh

The constructs pKNluc (Ppolh driving luc expression, with a 4 kb upstream region; Fig. 3.1), pKN603luc (Ppolh driving luc expression, with a 766 bp upstream region; Fig. 3.2), and pAJpolluc (Ppolh driving expression of the luc gene; Fig. 3.3) were used in transient expression assays. pAJpolluc displayed basal luciferase expression, pKN603luc consistently showed 2-3 fold increased expression above pAJpolluc, and pKNluc displayed about a 20-fold increase in basal expression (Fig. 3.4). The ORF603 containing upstream region was subsequently scanned for any putative transcription factor binding sites and a 32 bp sequence motif, termed AcSp (for Autographa californica Sp-like binding motif) was identified. The AcSp sequence (Fig. 3.5) contains a GGCG motif with loose homology to the classical GC box, and a classical CACCC motif (both on the non-coding strand), both of which may bind to members of the Sp family of proteins (Chapter I, section 5.1).
Fig. 3.1

**pVL1393 (~ 9.6 kb)**

- EcoRV
- BamHI
- ~ 4 kb upstream
- Ppolh

Digest with BamHI

Ligation

- Sall
- Sall
- EcoRV
- BamHI
- BamHI
- HindIII

**pKNluc (~11.5 kb)**

Fig. 3.1. Cloning strategy for pKNluc which carries the 92 bp Ppolh driving transcription of the luc gene, with ~4kb of upstream sequences. The 1.9 kb luc gene (BamHI fragment) was ligated into the baculovirus transfer vector pVL1393 which was linearized by digestion with BamHI, to construct pKNluc.
Fig. 3.3

A novel cis-sequence binds insect Sp-family proteins

**Fig. 3.3.** Cloning strategy for pAJpolluc which carries the 92 bp Ppolh driving transcription of the *luc* gene. The baculovirus transfer vector pVL1393 was digested with EcoRV and BamHI to release the 92 bp polh promoter fragment which was ligated into *HincII/BamHI* digested pUC18 plasmid, to construct the plasmid pAJpol. The 1.9 kb *luc* gene *BamHI* fragment (from plasmid pAcluc) was then ligated into pAJpol which was linearized by digesting with *BamHI*, to construct pAJpolluc.
Fig. 3.4

Transient expression of luciferase activity of pApolluc (bar 2), pKN603luc (bar 3) and pKNluc (bar 4). Bar 1 is the negative control with only AcNPV infection. Schematic representations of the three plasmids are depicted above.
Fig. 3.5. An Sp-like binding sequence (AcSp) is present upstream of the polh promoter. The AcSp sequence bears a consensus CACCC motif and a loose GC box, depicted in boldface, on the non-coding strand.
2. **AcSp binds to Sp family-like transcription factor(s) in Sf9 and HeLa cells**

The AcSp sequence was used in an EMSA using crude nuclear extracts prepared from AcNPV-infected (i) and uninfected (u) Sf9 cells respectively (Fig. 3.6 A). As a control, the cSp oligonucleotide, carrying a classical GC box which binds Sp family factors was also used. It is evident that at least one complex of similar mobility is obtained (shown by an arrowhead) with both uninfected (lanes 3, 4) and infected (lanes 5, 6) nuclear extracts using both probes. In addition, a faster mobility complex is also evident under these binding conditions with uninfected cell nuclear extract (lanes 3, 4) with both the probes. These results demonstrate the presence of a host factor(s) in insect cell nuclear extracts which can bind to the Sp1-like sequence motif present within ORF603 of region II (AcSp) as well as the consensus Sp1 sequence (cSp).

The Sp-like complex generated with uninfected Sf9 nuclear extract was subjected to cross-competition analyses in EMSAs using authentic Sp1 binding motifs. Fig. 3.6 B shows AcSp probe binding to (uninfected) Sf9 nuclear extract in the absence of any cold competitor (lane 2) and in the presence of AcSp, cSp and pUC18 cold competitors (lanes 3, 4, and 5, respectively). Lane 1 shows the free AcSp probe. It is interesting to note that in certain instances the consensus Sp1 sequence can compete even better than the homologous competitor for the AcSp probe. This is understandable as the factor(s) present could well have a higher affinity for the Sp1 cognate sequence defined by consensus (cSp) rather than the AcSp sequence. pUC18 did not compete for the binding, further pointing to the specificity of the complexes formed.

The reverse experiment was also carried out using radiolabeled cSp oligonucleotide as probe in EMSAs (Fig. 3.6 C). It is apparent that the binding (lane 3) was abolished after homologous cold competition (lane 4) but when cold AcSp was used as a competitor (lane 5), the complex with reduced mobility was competed out rather inefficiently. This is further evident after comparison of the competition with the non-specific competitor pUC18 (lane 6). Interestingly, the complex of higher mobility (lower shift) was better competed. Lanes 1 and 3 show cSp mobility without and with the nuclear extract, respectively. Lane 2 shows binding of labeled AcSp with the
Fig. 3.6 A. Electrophoretic mobility shift assay (EMSA) using oligonucleotides carrying the consensus Sp1 (cSp) and AcSp sequences. Sf9 nuclear extracts from uninfected (panel u) cells (lanes 3, 4) or virus-infected (panel i) cells (lanes 5, 6) were used in binding reactions. The arrowhead shows the complex obtained with the cSp/AcSp binding sequences and uninfected/infected cell nuclear extracts.
nuclear extract to compare the similar nature of the shifts obtained with both the probes.

At this stage, it was considered pertinent to investigate if the Sf9 Sp-family complex is the same as the well-characterized Sp family of factors present in HeLa cells. **Fig. 3.6 D** shows the binding of AcSp and cSp with Sf9 and HeLa cell nuclear extracts. Lanes 3 and 5 show binding of factor(s) present in Sf9 and HeLa extracts to radiolabeled AcSp, respectively. Lanes 4 and 6 show the binding of similar factors present in Sf9 and HeLa extracts to the cSp probe. Interestingly, a similar complex is obtained using both probes and either of the extracts. The HeLa cell factor(s) binds to the AcSp oligonucleotide to generate a complex with reduced intensity as compared to that obtained with cSp (compare lane 5 with lane 6), possibly because the AcSp sequence is not 100% identical to the Sp1 consensus. The fact that HeLa extracts containing the Sp-family of factors generates a similar complex as Sf9 extract with either of the probes directly points to the possibility of Sf9 cells also harboring Sp family-like transcription factors. Interestingly, the competition pattern seen with HeLa extracts mirrors that seen with Sf9 extracts (shown earlier in **Fig. 3.6 B, C**) in that cSp can compete very efficiently with the AcSp probe but AcSp competitor cannot compete as efficiently as cSp for binding to the cSp probe (compare lane 8 with lane 10). Homologous cross cold competitions expectedly abolished binding with both probes (lanes 7 and 9) whereas a non-specific competitor, pUC18, did not affect binding with either probe (lanes 11 and 12). These results convincingly demonstrate that while Sf9 cell nuclear extracts harbor Sp family-like transcription factors, the corresponding factors from HeLa cells can also bind to the viral AcSp motif.

3. **The Sp-like factors are distinct from PPBP, Sp1, Sp3 and Sp4**

It was also important to show that the Sp family-like factor(s) present in Sf9 cells described above is distinct from the well characterized polh promoter binding protein (PPBP). **Fig. 3.7** shows the binding of PPBP to the polh promoter B domain oligonucleotide (Hasnain et al., 1996) carrying the basal promoter determinants. The complex generated by the binding of PPBP, present in uninfected Sf9 nuclear extract, to the labeled B domain oligo (lane 2) can be specifically competed out only by the presence of
Fig. 3.6 D. EMSA using HeLa (lanes 5-12) and Sf9 (lanes 3, 4) nuclear extracts. AcSp probe was used in the absence of any extract (lane 1) or with Sf9 nuclear extract (lane 3) and HeLa cell nuclear extract (lane 5). Lanes 2, 4 and 6 show the corresponding results using the cSp probe. Lanes 7 and 8 show competition with homologous cold AcSp and heterologous cSp competitors, respectively. Lanes 9 and 10 similarly show competition with homologous cold cSp and heterologous AcSp competitors, respectively. Competition using pUC18 is shown in lanes 11 and 12.
Fig. 3.7. EMSA was carried out with labeled polyhedrin promoter B domain with S/f9 nuclear extract (NE-S/f9) (lane 2). Competitions with 25-fold excess of unlabeled B domain (lane 3), cSp (lane 4) and AcSp (lane 5) oligonucleotides are shown. Lane 1 is the free probe, in the absence of any protein.

Fig. 3.8. EMSA with α-Sp1, α-Sp3 and α-Sp4 antibodies using cSp probe and S/f9 nuclear extract (NE). Lane 1 is the free probe. Lanes 2-6 and 7-11 are without and with NE respectively. Lanes 2 and 8 are controls pre-incubated 1 µl undiluted normal rabbit serum (NRS), lane 3 is with 1 µl of 1:500 diluted NRS. Lanes 4-6 (without NE) and 9-11 (with NE) represent binding reactions pre-incubated with 1 µl α-Sp1, α-Sp3 and α-Sp4 antibodies, respectively.
homologous cold competitor (lane 3) and not with cSp (lane 4) or AcSp (lane 5) indicating that the Sp-like factor(s) present in Sf9 extract are distinct from the host factor PPBP reported earlier (Ghosh et al., 1998) involved in polh gene transcription.

Antibody "supershifting" analyses were carried out using anti-Sp1/Sp3/Sp4 antibodies to determine if the AcSp/cSp binding proteins were classical Sp1, Sp3 or Sp4 (Fig. 3.8). The cSp oligo was chosen as the probe since it carries the consensus GC box which is the preferred binding motif for the Sp1, Sp3 and Sp4 proteins. Lane 1 is the free probe, lanes 2, 3 and 8 are controls using normal rabbit serum (undiluted and 1:500 diluted respectively) without nuclear extract (lanes 2 and 3) and with nuclear extract (lane 8) respectively. Lane 7 shows DNA-protein complexes with the probe in the presence of extract alone and no antibodies. Lanes 9, 10 and 11 show binding in the presence of nuclear extract and α-Sp1, α-Sp3, and α-Sp4 antibodies, respectively. Lanes 4, 5, 6 are the corresponding control reactions in the absence of nuclear extracts. No detectable "supershifting" or immunodepletion was observed, suggesting that the insect Sp-like factor(s) are not Sp1, Sp3, or Sp4.

4. Human Sp1 protein is capable of binding AcSp and cSp

In order to substantiate the authenticity of the cSp probe, EMSAs were carried out with pure recombinant human Sp1 (Fig. 3.9 A). cSp expectedly binds to pure hSp1 (lane 6), under the reaction conditions used. However, the mobility of the shift is less than that with the Sf9 extract (lane 5), indicating that the insect proteins binding to cSp are not classical Sp1 but a different member of the Sp family. AcSp probe also shows a weak shift with pure Sp1 (lane 3) as compared to the Sf9 extract (lane 2). These results demonstrate that the AcSp sequence is recognized by the human Sp1 protein and the insect cell Sp factor(s) is different from classical Sp1.

DNase I footprinting assays were carried out (Fig. 3.9 B) using a 140 bp viral polh upstream region as a probe containing the AcSp sequence motif in the centre, and pure Sp1 protein (Promega Inc., USA). The coding strand does not show any protection in the AcSp region. However, practically the entire AcSp sequence is significantly protected on the non-coding strand, confirming the ability of the AcSp element to bind to
Fig. 3.9 A. Binding of radiolabeled AcSp (lanes 1, 2, 3) and cSp (lanes 4, 5, 6) oligonucleotides with pure recombinant human Sp1 protein (rhSp1, lanes 3, 6) and Sf9 nuclear extract (NE-Sf9, lanes 2, 5). B. DNase I protection analysis using the radiolabeled 140 bp coding (lanes 1-3) strand of the AcNPV genome containing the AcSp sequence approximately in the centre. Lane 1 depicts the A+G sequencing ladder. Lanes 2 and 3 show DNase I treatment in the absence and presence respectively of recombinant human Sp1 (rhSp1) protein. Lanes 4-6 represent corresponding lanes using the labeled non-coding strand. The boundaries of the AcSp sequence are marked by arrowheads on both labeled strands.
members of the Sp family. The boundaries of the AcSp sequence are marked by arrowheads.

5. **DNase I footprinting analysis with Sf9 nuclear extract shows a weak protection of AcSp**

DNase I footprinting assays were also carried out with the AcSp sequence-containing probe in the presence of Sf9 nuclear extract. There was no protection observed on the noncoding strand (Fig. 3.10 A), although some protection of AcSp was observed on the coding strand (Fig. 3.10 B). However, the footprint was not as clean as with the pure Sp1 protein, probably due to nuclease contamination of the nuclear extract since when larger amounts of extract was used (containing up to 64 µg protein) there was significant degradation of the probe. In both figures, lane 1 represents the A+G sequencing ladder (Maxam and Gilbert, 1977), lane 2 is a control with the DNase I reaction carried out to generate a ladder in the absence of any protein, and lanes 3-6 show the DNase I digestion pattern obtained after binding with 8, 16, 32, and 64 µg of Sf9 nuclear extract, respectively.

6. **Discussion**

The presence of an unusual upstream 766 bp sequence, which both enhanced transcription from Ppolh (Fig. 3.4) and rescued basal levels of transcription from the mutated promoter (Fig. 1.4), called for a more detailed analysis of the ORF603-containing region. A cis-sequence, AcSp, with binding motifs resembling those of the ubiquitous Sp1 family of mammalian transcription factors (Fig. 3.5) was identified. It was therefore considered necessary to document the existence, if any, of novel Sp family-like host factor(s) in Sf9 cells which could bind to this motif.

As mentioned earlier (Chapter I, section 5.1), no members of the Sp family of proteins have been demonstrated so far in adult insect cells. Protein products of the huckebein (hkb) (Brönner et al., 1994), buttonhead (bth) (Wimmer et al., 1995), and D-Sp1 (Wimmer et al., 1996) genes which are expressed in Drosophila embryos during the blastoderm stage show structural and functional homology to Sp1, but are not present in adult tissue. As a consequence, Drosophila Schneider or SL2 cells are commonly used as an Sp-negative background to carry out functional studies on Sp-
Lanes:

1. A+G ladder
2. Buffer control
3. 1 µl NE (8 µg protein)
4. 2 µl NE (16 µg protein)
5. 4 µl NE (32 µg protein)
6. 8 µl NE (64 µg protein)

**Fig. 3.10 A.** DNase I footprinting analysis of the non-coding strand. Lane 1 is the A+G DNA sequencing ladder of the 140 bp probe. Lane 2 shows the DNase I digestion pattern in the absence of Sj9 nuclear extract (NE). Lanes 3-6 depict DNase I digestion patterns after binding with 8, 16, 32 and 64 µg protein, respectively. The boundaries of the AcSp sequence are demarcated by arrows.
Fig. 3.10 B

Lanes:
1. A+G ladder
2. Buffer control
3. 1 μl NE (8 μg protein)
4. 2 μl NE (16 μg protein)
5. 4 μl NE (32 μg protein)
6. 8 μl NE (64 μg protein)

Fig. 3.10 B. DNase I footprinting analysis of the coding strand. Lane 1 is the A+G DNA sequencing ladder of the 140 bp probe. Lane 2 shows the DNase I digestion pattern in the absence of S9 nuclear extract (NE). Lanes 3-6 depict DNase I digestion patterns after binding with 8, 16, 32 and 64 μg protein, respectively. The boundaries of the AcSp sequence are demarcated by arrows.
A novel cis-sequence binds insect Sp-family proteins (Courey and Tjian, 1988). However, the list of Sp family proteins is growing at a rapid rate, and it is quite probable that insect cells may harbour novel members of the family. Recently, Scohy and coworkers (2000) identified two putative new members of the Sp family of proteins, KLF13 and KLF 14, by screening a mouse EST database with the Sp1 C-terminal DNA binding domain sequence. Reverse transcription-PCR revealed that the KLF 13 and -14 mRNAs were ubiquitously expressed. However, no studies have been carried out yet in order confirm the presence of or characterize the putative protein(s) if any.

The identification of AcSp, an Sp-like DNA binding cis-sequence, prompted an investigation on the possibility of Sj9 cells having Sp family-like proteins. As a control, an oligonucleotide (cSp) carrying the classical GC box, the consensus binding motif for most of the Sp family members, was used in parallel, in EMSAs. Interestingly, identical DNA-protein complexes using both AcSp and cSp, and both uninfected and virus-infected nuclear extracts (Fig. 3.6 A) were observed, which could be competed out with either oligonucleotide (Fig. 3.6 B, C). In fact, the cSp sequence was observed to act as a better cold competitor than AcSp (Fig. 3.6 B, compare lanes 3 and 4) for binding to the AcSp probe. Conversely, unlabeled AcSp cold competitor could not compete as efficiently as cSp for binding to the cSp probe (Fig. 3.6 C, compare lanes 4 and 5). This could be a reflection of a difference in binding affinities of the factor(s) to AcSp and cSp. The Sp family-like factor(s) may have a higher affinity for binding to cSp which carries the consensus GC box. The presence of homologous DNA-protein complexes with identical cross competition patterns observed with HeLa cell nuclear extracts (Fig. 3.6 D), which are a known source of the Sp family of proteins, further strengthen the Sp-like characteristics of the Sj9 factors.

The Sp family factor(s) was also found to be distinct from PPBP, the other host factor discovered in our laboratory (Burma et al., 1994; Mukherjee et al., 1995b; Hasnain et al., 1996; Hasnain et al., 1997; Ghosh et al., 1998) which binds to the polh promoter (Fig. 3.7 A). This was essential because, in view of the role played by PPBP in regulation of Ppolh transcription, it was important to rule out the possibility that PPBP itself may be binding to upstream sequences to bring about enhancement of transcription. Supershifting assays were also carried out using α-Sp1, α-
Sp3, and α-Sp4 antibodies (Fig. 3.8). Sp1, Sp3 and Sp4 are known to bind CACCC motifs and GC boxes very strongly (unlike Sp2, which binds strongly to GT boxes) and hence it was pertinent to check if any of these proteins were present in insect cells and capable of binding to AcSp. No "supershifting" or immunodepletion of the protein-DNA complexes could be seen with any of the antibodies, indicating that the insect protein(s) are in all probability distinct from Sp1, Sp3 and Sp4.

The authenticity of the cSp probe was confirmed by the observation that it could indeed bind to pure recombinant human Sp1 protein under the conditions used for EMSAs (Fig. 3.9 A). These results also reveal that the AcSp-binding protein(s) is probably not classical Sp1, since the mobility of the complexes with rhSp1 in EMSAs was different from those obtained with Sj9 nuclear extracts. However, AcSp could bind to rhSp1 only very weakly in EMSAs (Fig. 3.9 A), which prompted analyses by DNase I footprining (Fig. 3.9 B) to establish that the AcSp sequence does carry sequence motifs which could bind to Sp-like proteins. There was a strong protection with rhSp1 on the non-coding strand, confirming it's ability to bind to AcSp. DNase I protection studies to check the binding of Sj9 nuclear proteins to AcSp were not as conclusive. No protection of the non-coding strand was seen (Fig. 3.10 A) and the coding strand was only weakly protected (Fig. 3.10 B). This is surprising, since EMSAs show very strong DNA-protein complexes with Sj9 nuclear extract. This could probably be due to the Sp-like proteins being present in very small amounts in the extract, or an inability of the protein(s) to bind to the probe under the conditions used for the DNase I reaction.

The results presented here thus reveal the functional importance of the 766 bp upstream sequence in enhancing transcription from Ppolh. This region harbours a sequence, AcSp, showing sequence similarity to the binding sequences of the Sp family of proteins. AcSp was used in EMSAs along with an oligonucleotide, cSp, carrying the consensus Sp-binding sequence. Both AcSp and cSp showed similar DNA-protein complexes with Sj9 (uninfected and virus-infected) and HeLa cell nuclear extracts and also showed comparable cross-cold competition patterns, indicating that the proteins belong to the Sp family of transcription factors. The insect Sp-like proteins were shown to be distinct from the other insect host factor, PPBP.
They were also demonstrated not to be Sp1, Sp3 or Sp4, by EMSA, UV crosslinking analyses and antibody-"supershifting" EMSAs, although the pure Sp1 protein was shown to be capable of binding to cSp and AcSp using EMSA and DNase I footprinting analyses. DNase I footprinting also revealed that while pure Sp1 protein bound to the non-coding strand of AcSp, the insect Sp-like factor(s) probably bound to the coding strand.