Inverse PCR to identify DNA sequence upstream of the pea HMG I/Y open reading frame

Druffel, K., Carson, J.A., Hartney, S. and Hadwiger, L.A.
Dept. of Plant Path. Washington State Univ., Pullman, WA, USA

In the pursuit of the defense responses in pea plants, or in any induced response in eukaryotic organisms, the sequence of the open reading frame of the induced gene is attainable from a cDNA clone. Once the open reading frame transcription start site of the cDNA clone has been obtained, the promoter region is often of interest and is accessible by inverse PCR. Although inverse PCR is a standard technique, all PCR strategies do not necessarily result in success. The promoter of the pea gene HMG I/Y is of interest because of the possibility that its coded protein is influencing its own transcription. The HMG I/Y protein contains four AT-hooks (2, 5) that have affinity for AT-rich regions of DNA (4). The extensive research relative to this property has led to the recognition that HMG-I/Y is an architectural transcription factor (7). It has recently been determined that both its RNA transcript and its protein product are depleted as the pea defense response is initiated (4).

The pea HMG I/Y protein can efficiently bind to AT-rich segments of promoters from pea PR genes, genes that are activated as the pea tissue resists fungal pathogens (1, 4). Therefore it was of interest to determine if such AT-rich regions were also present in the HMG I/Y's 5' region. This manuscript describes in detail the strategy used to sequence this promoter.

Methods and Results

DNA extraction

One immature pea pod (3 cm length) from *Pisum sativum* was crushed between two sheets of weighing paper using pliers. The pod was then transferred to a 1.5 ml microfuge tube with 1 ml of extraction buffer (6) (100 mM Tris, pH 8.0, 50 mM EDTA, 500 mM NaCl, 10 mM 2-mercaptoethanol). 140 µl of 10% SDS was added. The tube was inverted to mix and incubated at 65 C for 10 minutes. 250 µl of 8M KOAc was added to the tube, inverted to mix, and placed on ice for 5 minutes. The tube was then centrifuged at 13,000 rpm for 8 minutes and 600 µl of the supernatant was transferred to a new tube. 300 µl of isopropanol was added and the contents were mixed and held at 4 C for 10 minutes. Following 10 minutes of centrifugation at 13,000 rpm the supernatant was discarded. 750 µl of 75% EtOH was added. The tube was gently mixed then centrifuged for 3 minutes. All supernatant was removed from the tube and the pellet resuspended in 50ul sterile ddH_20.

Inverse PCR

Pea genomic DNA was cleaved with the *Taq*I 4 bp cutter (BRL products/Life Technology, Grand Island, NY) in the following 100 µl reaction mix: 10µl~500 ng DNA, 10 µl *Taq*I Buffer 2 (BRL), 5 µl *Taq*I, and 75 µl water. Aliquots of the mix were incubated at 65° C for 30 min, 60 min, or 2 h. *Taq*I had the specificity to cleave the *Taq*I site located within the 3' end of HMG I/Y gene open reading frame and various sites upstream or down-stream of the gene of interest. The digested DNA was purified with a phenol/chloroform extraction and EtOH precipitation. The DNA was resuspended in 100 µl water. Ten µl of this DNA solution was ligated to itself (using T4 DNA ligase, BRL) to create a series of small circular DNAs representing the total genome, including the HMG-I/Y gene and its adjacent sequence. A primary PCR was performed as follows using the primer set indicated in Figs. 1 and 2:

Ligated T*aq*I/cut genomic DNA	1 ul
10X PCR buffer (BRL)	2 µl
2.5 mM dNTPs	1.2 µl
50 mM MgCl2	0.8 µl
20 µm LeeInv367c (5' GAA CAA CCG AAT GGC CTT CT 3')	0.6 µl
20 µm LeeInv580F (5' CCA AAG GCT TCT GGA AGT GG 3')	0.6 µl
double distilled H_2O	13.6 µl
Taq Polymerase	0.2 µl

```
Alu1
AGCTTCATTGATGTATACAGATTATGAACAAAGTTTAACGGAAAATTCAAGATTAGTTGA
TCGAAGTAACTACATATGTCTAATACTTGTTTCAAATTGCCTTTTAAGTTCTAATCAACT
        10        20        30        40        50        60

TTTCTTAATAAGCAATTTGTGATAAAATGAATTGAAAAATAAAACGAGTATAACTAATCA
AAAGAATTATTCGTTAAACACTATTTTACTTAACTTTTTATTTTGCTCATATTGATTAGT
        70        80        90       100       110       120

TGTATGCATGTGGATATATTAGTAGCAACCTAACCCTAACATCTCTTTATCATCTAACAT
ACATACGTACACCTATATAATCATCGTTGGATTGGGATTGTAGAGAAATAGTAGATTGTA
       130       140       150       160       170       180

CTTTGTTATTCTCTTATTTTAATTTGTTATTGTTTTAATTTCTCAAACAAAACATCTTTC
GAAACAATAAGAGAATAAAATTAAACAATAACAAAATTAAAGAGTTTGTTTTGTAGAAAG
       190       200       210       220       230       240

AAAACAAATCCTAAATTGCTTAAAATAGTAACAATTACTTCAACAATAAAAAAACTTTTA
TTTTGTTTAGGATTTAACGAATTTTATCATTGTTAATGAAGTTGTTATTTTTTTGAAAAT
       250       260       270       280       290       300

TTTCAATCATATTTTGTACATGCATCTCATTACATCTTGAAATTCCACTTATTTTCTGTA
AAAGTTAGTATAAAACATGTACGTAGAGTAATGTAGAACTTTAAGGTGAATAAAAGACAT
       310       320       330       340       350       360

CATGCATCTCATTACATTTTGAAATTCCACTTATTTTCTGGACATGGATCTCATTACATT
GTACGTAGAGTAATGTAAAACTTTAAGGTGAATAAAAGACCTGTACCTAGAGTAATGTAA
       370       380       390       400       410       420

TTGAAATTCCACTTATTTTCCGACAATTATTTTAAATTACTTAAAATAGTAACAATTACT
AACTTTAAGGTGAATAAAAGGCTGTTAATAAAATTTAATGAATTTTATCATTGTTAATGA
       430       440       450       460       470       480

CCAACAATTTTTTTTAATTCTATCATATTTTATACATACATTTTATTACATCTTGAAATT
GGTTGTTAAAAAAAATTAAGATAGTATAAAATATGTATGTAAAATAATGTAGAACTTTAA
       490       500       510       520       530       540

TCACTTATTCTCTAACAATTATCTTAAATTACTTAAAATAGTAATAATTATCTATAAATT
AGTGAATAAGAGATTGTTAATAGAATTTAATGAATTTTATCATTATTAATAGATATTTAA
       550       560       570       580       590       600

GTATCGTAAGATGATAAAAACATACTAACGAATTGTAGTAGTTTATAACTTAATATTTTT
CATAGCATTCTACTATTTTTGTATGATTGCTTAACATCATCAAATATTGAATTATAAAAA
       610       620       630       640       650       660
    Taq1
TCTTTCGATTTTACTTTTATTATCTTAATTCAAAAAATTATATATTATTTAAATATATTT
AGAAAGCTAAAATGAAAATAATAGAATTAAGTTTTTTAATATATAATAAATTTATATAAA
       670       680       690       700       710       720

TTAAGTCATTTTATAATTATAAGTCATTTCATTTTATTCAACATTACAAATTTAATCAAT
AATTCAGTAAAATATTAATATTCAGTAAAGTAAAATAAGTTGTAATGTTTAAATTAGTTA
       730       740       750       760       770       780

TAATTTATTTTCAACCACCCATTACCAACTTATAAATTAAAAATAAATAAATTCATCAAC
ATTAAATAAAAGTTGGTGGGTAATGGTTGAATATTTAATTTTTATTTATTTAAGTAGTTG
       790       800       810       820       830       840

TATAAATTAATTTATCAATTATCCGTTATTTCTTTTAACAACAATGTCTATATACATATC
ATATTTAATTAAATAGTTAATAGGCAATAAAGAAAATTGTTGTTACAGATATATGTATAG
       850       860       870       880       890       900

GTATTAAAAATGATGATATAATTTCCACTTTTGATTTTTTAAAATCAAATTATGCAAAAA
CATAATTTTTACTACTATATTAAAGGTGAAAACTAAAAAATTTTAGTTTAATACGTTTTT
       910       920       930       940       950       960
                                                        Taq1
TATTTAAGAGGTCGGTGCGTCCCGACACTTAATATTTAGTATGAAAATTGTAATTATCGA
ATAAATTCTCCAGCCACGCAGGGCTGTGAATTATAAATCATACTTTTAACATTAATAGCT
       970       980       990      1000      1010      1020

AAATATACAAACCGAGTCAAACCGTTTCTTATTTTAGCAATAAAATTCACAGATACATTT
TTTATATGTTTGGCTCAGTTTGGCAAAGAATAAAATCGTTATTTTAAGTGTCTATGTAAA
      1030      1040      1050      1060      1070      1080
```

The PCR temperature recycling program was: 94 C for 30 sec, 65 C for 20 sec, 72 C 1.5 min for 50 cycles. The product was then diluted 1:10 in preparation for secondary PCR.

A secondary PCR utilized the primer combination INV615F (5' GCC GAA GAA GAT TGC TAG GAC 3')/INV248c (5' TCA TTC AGT GAA TCA ATA GCC 3') as indicated in Fig. 1 with an annealing temperature of 61 C. This yielded an ~780 bp segment, which was cloned into a Topo PCR2.1 vector (Promega) and subsequently sequenced. The two new sequences generated from the clone were recognized as being within the region 5' of the HMG I/Y start site. The length of this sequence (780 bp) was insufficient for the complete analysis of a HMG I/Y 5' region but contained an *Alu*I site. Thus the pea DNA was cut with *Alu*I and the DNA was again circularized by ligation to obtain additional upstream sequence. Two new sets of primers were designed. Inv339F (5' ATC CTC ATC CAA AAG AAG 3')/ Inv292 (5' AAT TAA GGC TTT TTT GAC 3') was used for the primary PCR reaction. PCR was run under the following conditions; 94 C for 30 sec, 51 C for 20 sec, 72 C for 1 min. Secondary PCR using Inv187c (5' TAA TTG AAA AGG GTA TGC 3')/Inc609f (5' TCA ATC CTT AGT TCA TCC 3') was then performed (94 C for 30 sec, 51 C for 20 sec, 72 C for 1 min for 50 cycles). The subsequent PCR product was cloned and sequenced as before and increased the total sequence upstream of the HMG-I/Y gene to 1748bp (Fig. 1).

To verify the sequence 1748bp upstream of the HMG-I/Y gene, a primer set 13f (5' ACA GAT TAT GAA CAA AGT TTA ACG 3')/754 (5' TCA CTT GTG TCA ACT GAG GC 3') was used to amplify an approximate 2500 bp segment off of uncut genomic DNA. PCR conditions were 94 C 30sec, 61 C 20sec, 72 C 2:4 min for 50 cycles. The PCR product was cloned and completely sequenced to confirm that the inverse sequence reactions were assembled correctly.

```
ATCATTTTATTTTCCTATTGAATAAAATACAATGTTTTTTCATTTATTTAATCTTTAAAT
TAGTAAAATAAAAGGATAACTTATTTTATGTTACAAAAAAGTAAATAAATTAGAAATTTA
      1090      1100      1110      1120      1130      1140
                                                   Inv187c
AATTTTTCTTGTTTATTTTATCACATTTTGATAACTATGAATTTGAAAAGCATACCCTTT
TTAAAAAGAACAAATAAAATAGTGTAAAACTATTGATACTTAAACTTTTCGTATGGGAAA
                                                        1200

TCAATTAAAAAATCAATTTATTTATTTCATTTCATAAATAATATTCATAAAATTAAATAC
AGTTAATTTTTTAGTTAAATAAATAAAGTAAAGTATTTATTATAAGTATTTTAATTTATG
      1210      1220      1230      1240      1250      1260
                                   Inv2923c
AATGAGTAGAATTTCAAACTCTCAATAAATTTTAGTCAAAAAAGCCTTAATTTAAAAATA
TTACTCATCTTAAAGTTTGAGAGTTATTTAAAATCAGTTTTTTCGGAATTAAATTTTTAT
      1270      1280      1290      1300      1310      1320
                                         Inv339f
AATAAAAATATTTAAAATTGAGATAGTCTACATCACAAATCCTCATCCAAAAGAACAAAG
TTATTTTTATAAATTTTAACTCTATCAGATGTAGTGTTTAGGAGTAGGTTTTCTTGTTTC
      1330      1340      1350      1360      1370      1380

AATACAAAAAACAGTAGGTACCTCCAAATATTTCTGTGAACTAACACATTTTTGCCATGT
TTATGTTTTTTGTCATCCATGGAGGTTTATAAAGACACTTGATTGTGTAAAAACGGTACA
      1390      1400      1410      1420      1430      1440

CATCAATCCATGTGAGATTCTCCATATTATAATATCAACCCTTGGATCATCATCATTCTA
GTAGTTAGGTACACTCTAAGAGGTATAATATTATAGTTGGGAACCTAGTAGTAGTAAGAT
      1450      1460      1470      1480      1490      1500

TTGATTCCTAGCCGTCCATTGTCTTGTTCAGACAAACACAAGATATATCTTGGGAAAAAG
AACTAAGGATCGGCAGGTAACAGAACAAGTCTGTTTGTGTTCTATATAGAACCCTTTTTC
      1510      1520      1530      1540      1550      1560

AAGAGCAAACTTTTTTTAATATATTTAATTTCTTTCCAATCTTTTAATACATTTATCTCC
TTCTCGTTTGAAAAAAATTATATAAATTAAAGAAAGGTTAGAAAATTATGTAAATAGAGG
      1570      1580      1590      1600      1610      1620
           Inv609F
CTTTAAATTCAATCCTTAGTTCATCCATCAATTCACTCACAATCTCATTTCTCATAACAA
GAAATTTAAGTTAGGAATCAAGTAGGTAGTTAAGTGAGTGTTAGAGTAAAGAGTATTGTT
      1630      1640      1650      1660      1670      1680

AATTTCTATCTCCCTCAGATTTTTTATCTCAATTTTAAAGCTTTTTTCCTCACTCTTTCG
TTAAAGATAGAGGGAGTCTAAAAAATAGAGTTAAAATTTCGAAAAAAGGAGTGAGAAAGC
      1690      1700      1710          Alu1      1730      1740
 Alu1   ORF  HMG1F
CAGCTTCAATGGCAACAAGAGAGGTTAATAAGCCTCTGTCACTTCCTCCTTACCCTGAGG
GTCGAAGTTACCGTTGTTCTCTCCAATTATTCGGAGACAGTGAAGGAGGAATGGGACTCC
      1750      1760      1770      1780      1790      1800

TAAACACAAACCCCAATTTTTACTTTTCTCATGGATATTTATTCTGTACTATTTTCTTAG
ATTTGTGTTTGGGGTTAAAAATGAAAAGAGTACCTATAAATAAGACATGATAAAAGAATC
      1810      1820      1830      1840      1850      1860

TAAAGTTGAAATTTTTTCACTGATCTGTTTGATCTCAAAATTTTCACTTACTTAGTTTTT
ATTTCAACTTTAAAAAAGTGACTAGACAAACTAGAGTTTTAAAAGTGAATGAATCAAAAA
      1870      1880      1890      1900      1910      1920

TTTTTCTCATCAAGTTTTGTTTGTTTTGGGTTTTTGTGTTTGTTTTGGGTTTTTGTTATT
AAAAAGAGTAGTTCAAAACAAACAAAACCCAAAAACACAAACAAAACCCAAAAACAATAA
      1930      1940      1950      1960      1970      1980

TTGATGGAAAAGATTGATTATGTGTTTTGTTGCATGTTTTTGTTGTAGTTGATACTGAAG
AACTACCTTTTCTAACTAATACACAAAACAACGTACAAAAACAACATCAACTATGACTTC
      1990      2000      2010      2020      2030      2040
   Inv248c
GCTATTGATTCACTGAATGAACCAAATGGATCAAACAAATCAGCAATATCAAACTACATA
CGATAACTAAGTGACTTACTTGGTTTACCTAGTTTGTTTAGTCGTTATAGTTTGATGTAT
      2050      2060      2070      2080      2090      2100
                           Inv367c
GAATCAGTTTACGGTGAACTACCAGAAGGCCATTCGGTTGTTCTTTTATATCATCTGAAC
CTTAGTCAAATGCCACTTGATGGTCTTCCGGTAAGCCAACAAGAAAATATAGTAGACTTG
      2110      2120      2130      2140      2150      2160
```

Discussion

The availability of this sequence enables the identification of potential promoter elements 5' of the HMG-I/Y opening reading frame and assists the development of nested series of promoter/reporter elements to evaluate those sequences vital for activation (1). Once these regulatory elements have been found it is possible to identify and characterize the sequences involved in HMG-I/Y transcription. Stretches of alternating A and T sequences are known to bind the AT hooks of the HMG-I/Y protein (7). There are fifteen 4 bp stretches, three 5 bp stretches, four 6 bp stretches and one 7 bp stretch of alternating A and T found in the region 5' of the pea HMG-I/Y gene open reading frame. The availability of pure HMG-I/Y protein and the 5' sequence information enables gel mobility assays. These assays can determine if any of the regions 5' of HMG-I/Y associate with this architectural transcription factor that is its own coded gene product. Many other predicted transcription factor binding sites can be derived from the 5' sequence. For example, the MYB attachment site, AACCG, is found twice in the pea 5' region and once in the *Arabidopsis* promoter region (3).

Acknowledgment: The GenBank Accession number for the pea HMG-I/Y sequence reported in this manuscript is AY864056. We thank the Washington Sea Grant program for support. Ag. Research Center ms. no.0399.

1. Choi, J.J., Klosterman, S.J. and Hadwiger, L.A. 2004. Phytopath. 94: 651-660.

```
CAGATGAAAGAGAGTGGGGACCTTGTTTTTGCAAAGAACAACTACTTGAGGCCTGATCCA
GTCTACTTTCTCTCACCCCTGGAACAAAAACGTTTCTTGTTGATGAACTCCGGACTAGGT
      2170      2180      2190      2200      2210      2220

AATGCTCCACCGAAGAGAGGGCGCGGTAGGCCTCCTAAGGCGAAGGATCCGTTGGCCTCA
TTACGAGGTGGCTTCTCTCCCGCGCCATCCGGAGGATTCCGCTTCCTAGGCAACCGGAGT
      2230      2240      2250      2260      2270      2280

CCGCCTTCAGGTGCTGTGTCCACACCGAGGCCAAGGGGTCGTCCGCCTAAGGATCCTAAT
GGCGGAAGTCCACGACACAGGTGTGGCTCCGGTTCCCCAGCAGGCGGATTCCTAGGATTA
      2290      2300      2310      2320      2330      2340
                Inv580F                              Inv 615F
GCGCCACCGAAGACTCCAAAGGCTTCTGGAAGTGGTAGGCCAAGGGGTAGGCCGAAGAAG
CGCGGTGGCTTCTGAGGTTTCCGAAGACCTTCACCATCCGGTTCCCCATCCGGCTTCTTC
      2350      2360      2370      2380      2390      2400

ATTGCTAGGACCGAGGATGTTGATGCTTCAACTCCTAGTCCTGTGAGTGTTGCTGCTGTT
TAACGATCCTGGCTCCTACAACTACGAAGTTGAGGATCAGGACACTCACAACGACGACAA
      2410      2420      2430      2440      2450      2460
                                          Taq1
AATGTTGATGTTGTTGTTCCATGTGTTGCTGCTGTTCCTACTTCGAGTGGGAGACCAAGG
TTACAACTACAACAACAAGGTACACAACGACGACAAGGATGAAGCTCACCCTCTGGTTCC
      2470      2480      2490      2500      2510      2520


GGTAGGCCTCCTAAGGTGAAGCCTCAGTTGACACAAGTGA
CCATCCGGAGGATTCCACTTCGGAGTCAACTGTGTTCACT HMG754
      2530      2540      2550      2560
```

Fig. 1: DNA sequence of the 5' region and the open reading frame of the pea gene HMG-I/Y. The restriction enzyme sequences and the primers utilized in developing the 5' sequence are underlined and labeled.

2. Gupta, R., Webster, C.I. and Gray, J.C. 1997. Plant Mol. Biol. 35: 987- 992.
3. Gupta, R., Webster, C.I. and Gray, J.C. 1998. Plant Mol. Biol. 36: 897- 907.
4. Klosterman, S. J., Choi, J. J., Hadwiger, L. A. 2003. Mol. Plant Path. 4: 249-258.
5. Klosterman, S.J. and Hadwiger, L.A. 2002. Plant Sci. 162: 855-866.
6. Presting, G.G., Smith, O.P. and Brown, C.R. 1995. Phytopath. 85: 436-442.
7. Reeves, R. 2001. Gene 277: 63-81.

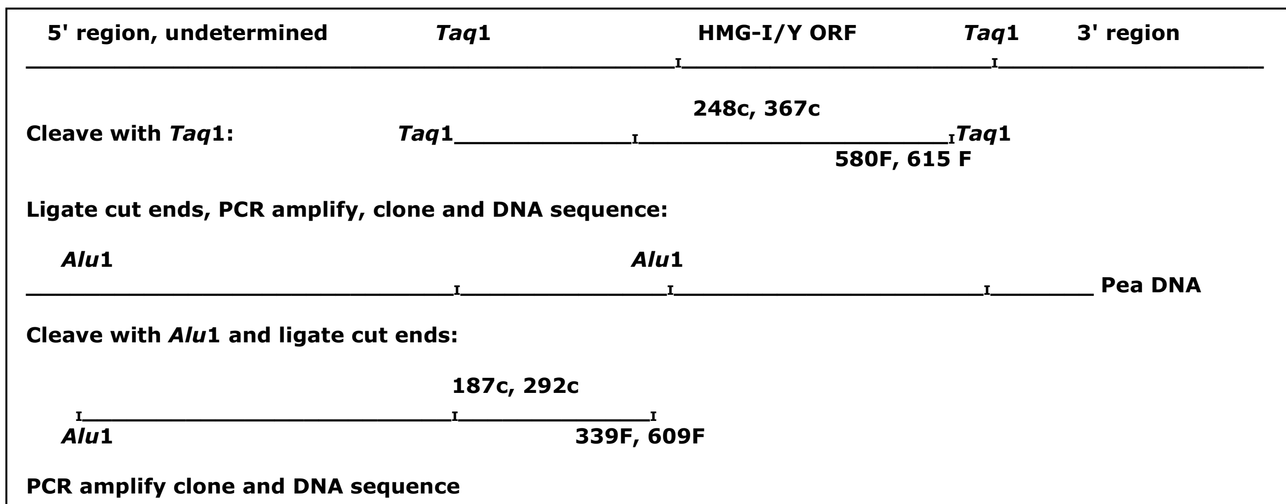

Fig. 2 Cartoon of the sequence of restriction digestion, primer development and DNA sequencing analyses of the pea HMG-I/Y 5' region (See Methods).