gm8100.hf
error report
(C. Crane-Robinson)

FBrf0129131

From [redacted] Sat May 27 19:17:38 2000
From: [redacted]
To: [redacted]
Subject: DSox14
Date: Sat, 27 May 2000 19:22:22 BST

I attach an annotated sequence of DSox14 (forget our DSox60B notation). It is an MS WORD file which I hope is OK for you. This gene is in FlyBase and is
annotated in GadFly but corrections are needed, as follows.
The genomic sequence at AE003462 corresponds to ours but the annotation at AE003462.1 is incorrect in that it acknowledges only the second (of 2) exons.
The entry at AJ252125.1 (Cremazy et al) is a cDNA that is only partly correct. There is a sequencing error at 1621 (deletion of 2 bp) so that their subsequent 16
amino acids are incorreect and their stop codon is premature: the protein is much longer.

Any replies to: [redacted]

Citation: Sparkes, A.C., Mumford, K., Crane-Robinson, C. & Newbury, S.F.
Biophysics Laboratories, University of Portsmouth, PO1 2DT, UK
To be published.

I would be glad if you confirmed receipt of this message and attachment.

Prof. Colyn Crane-Robinson,
Biophysics Laboratories,
St. Michael's Building,
University of Portsmouth,
Portsmouth, PO1 2DT, UK.

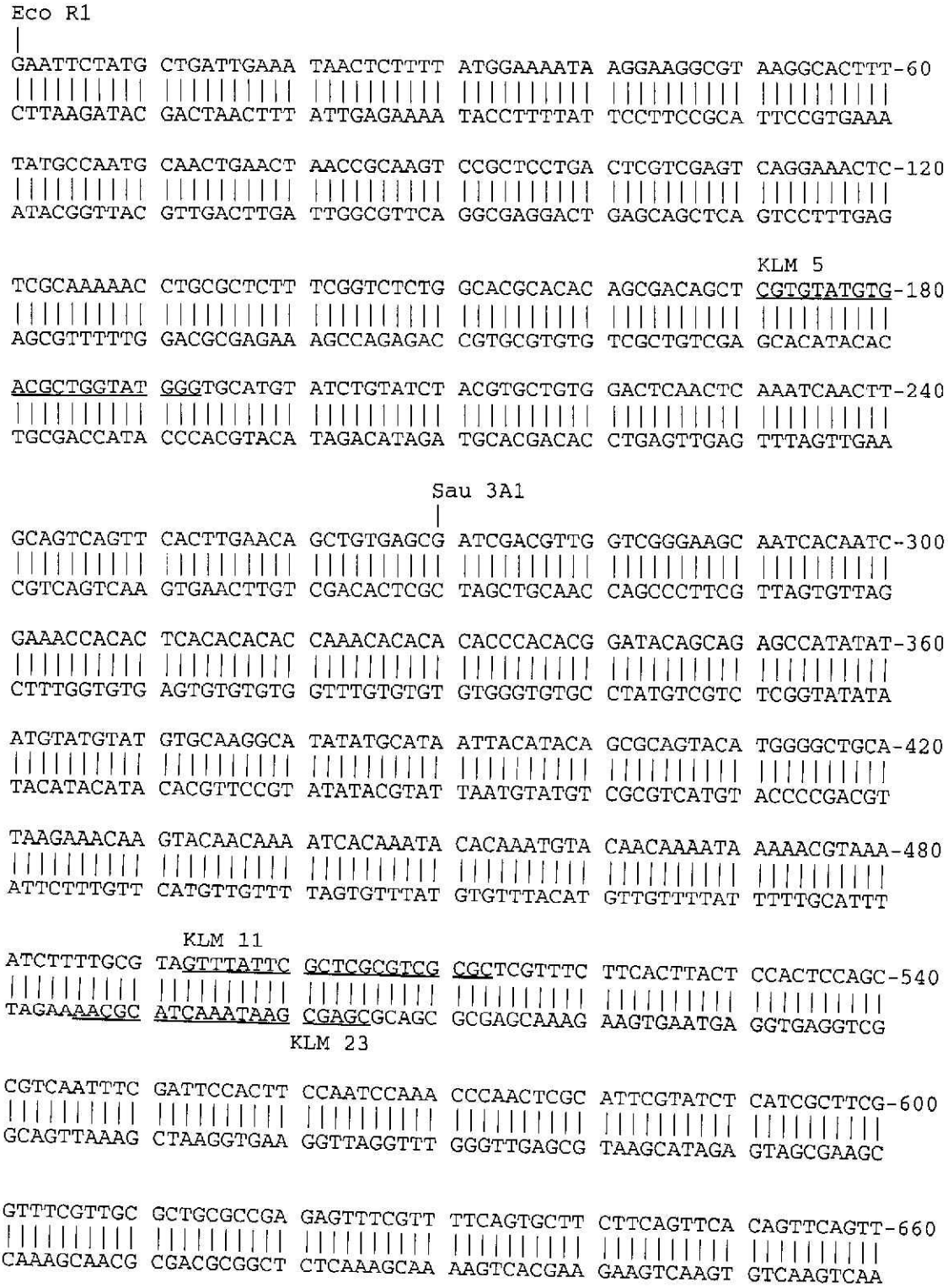
Tel: [redacted]
Fax:
e-mail: [redacted] or
[redacted]

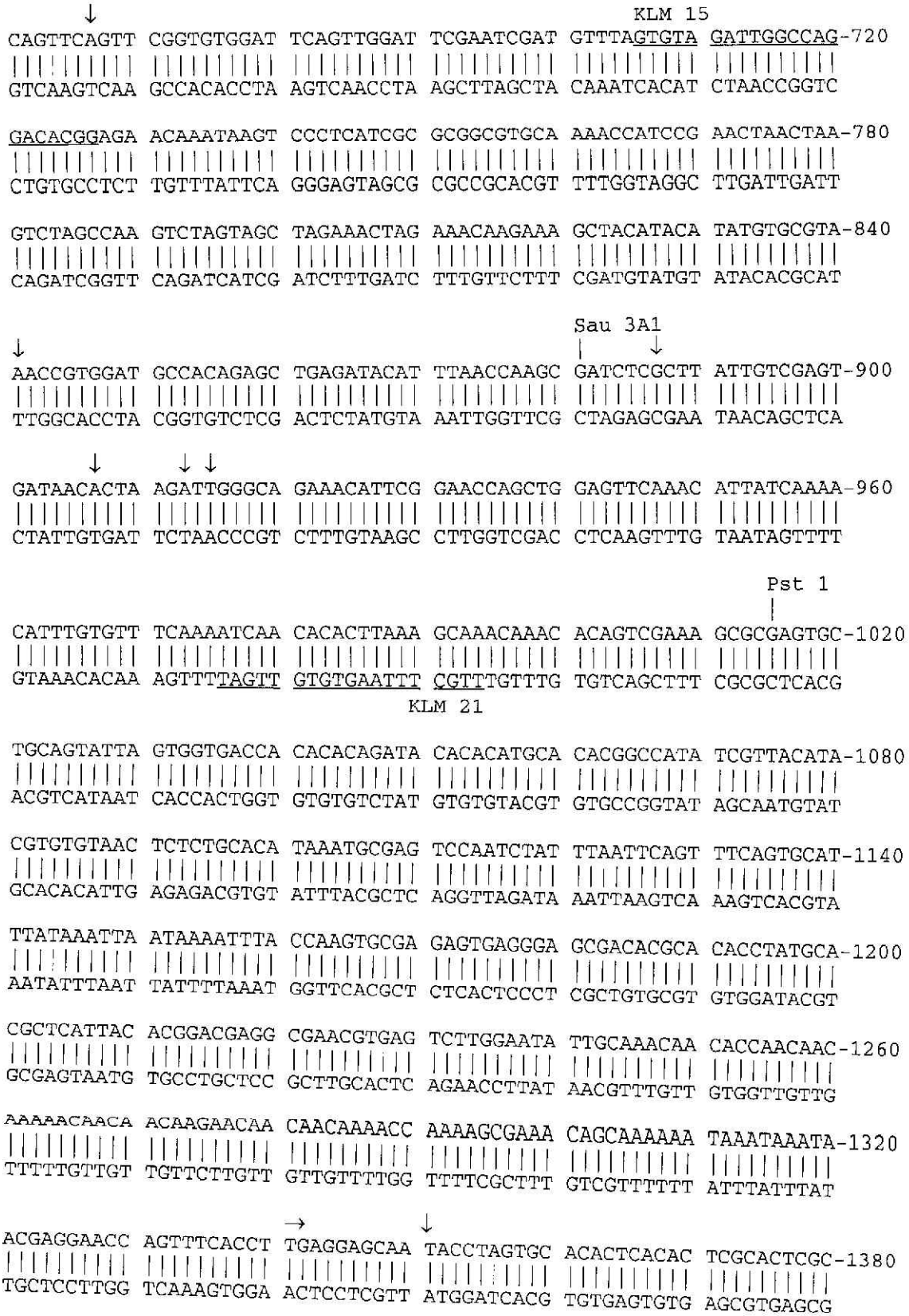
Get Your Private, Free E-mail from MSN Hotmail at http://www.hotmail.com

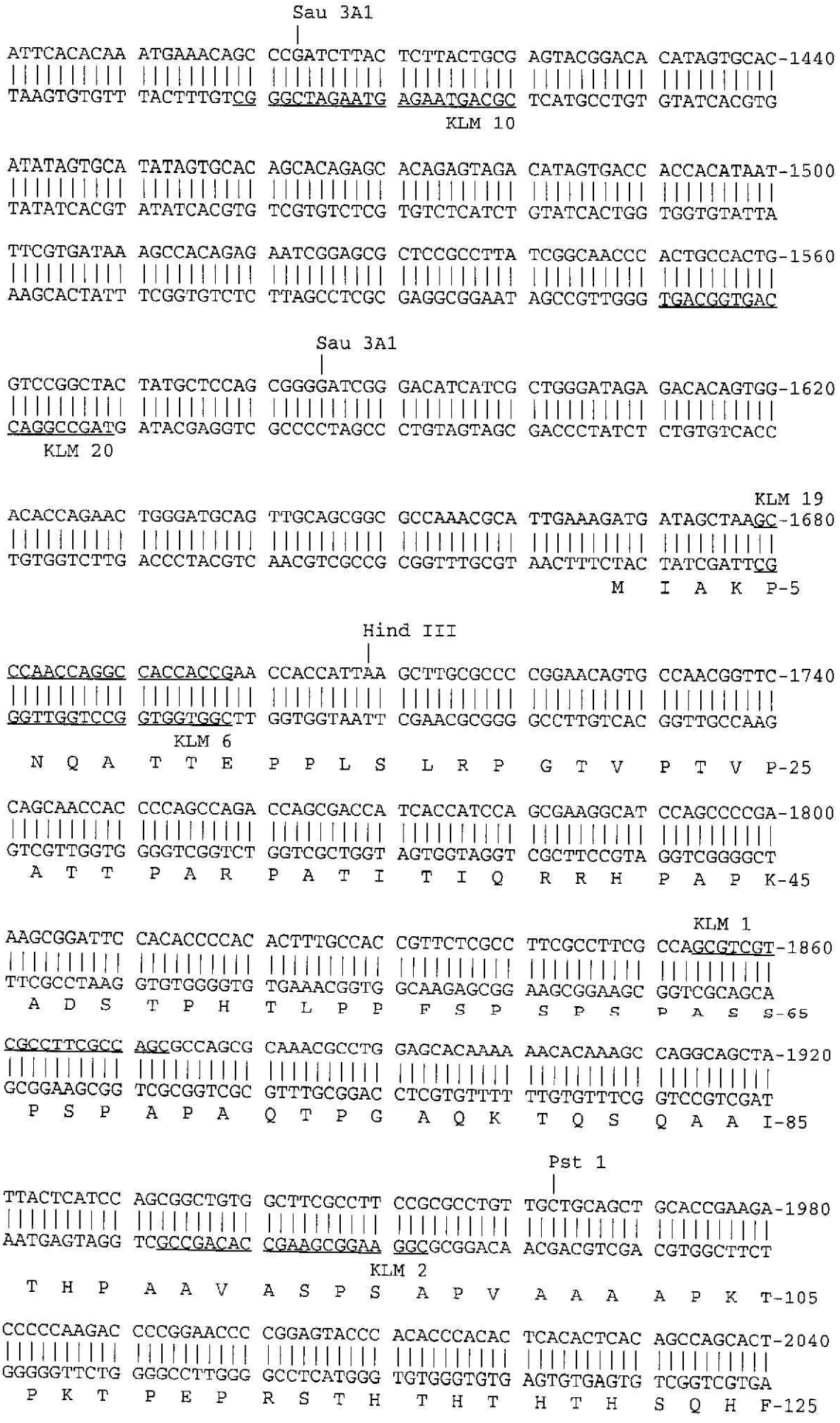
<This message contains 1 attachments (not printed)>

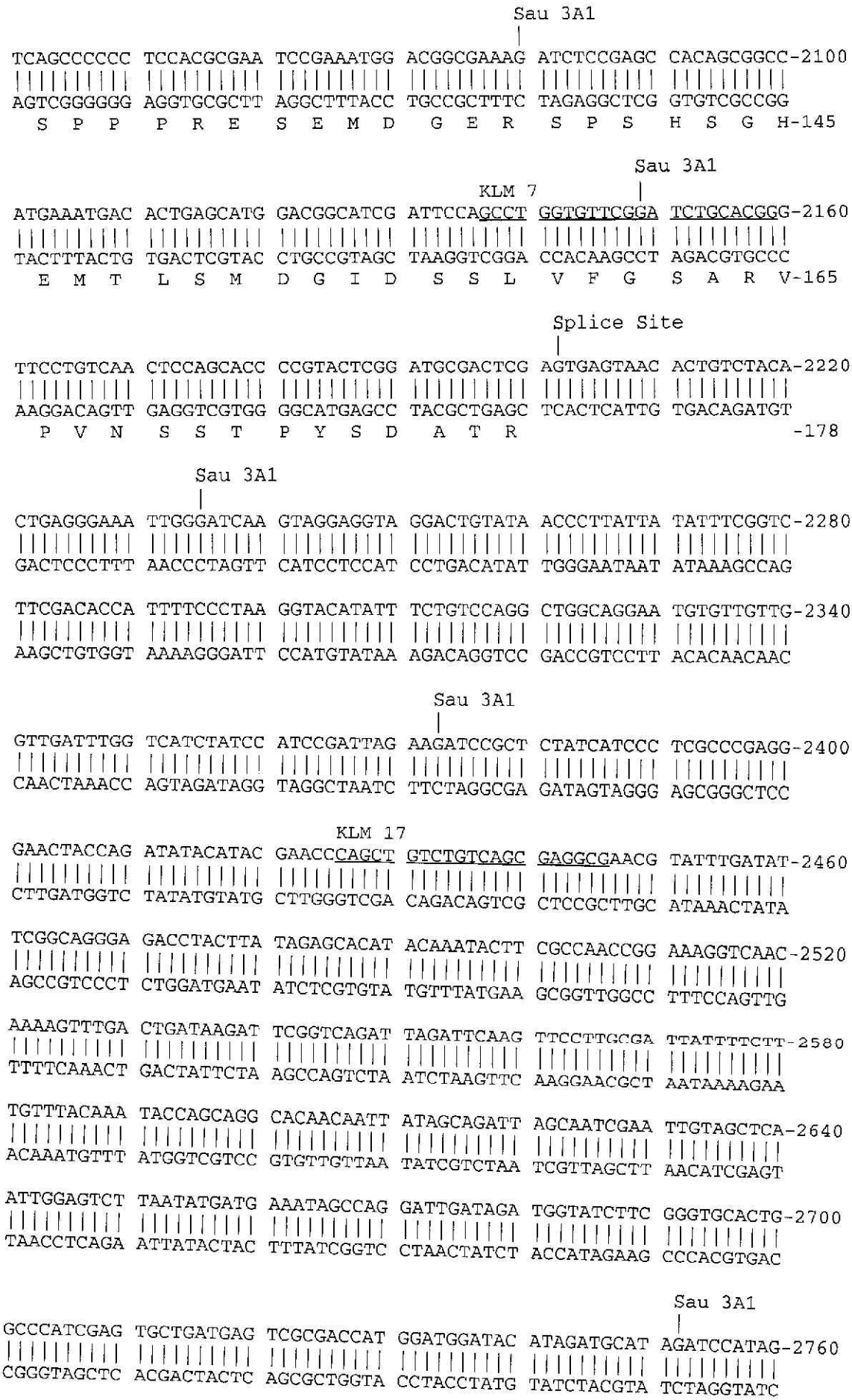
Appendix

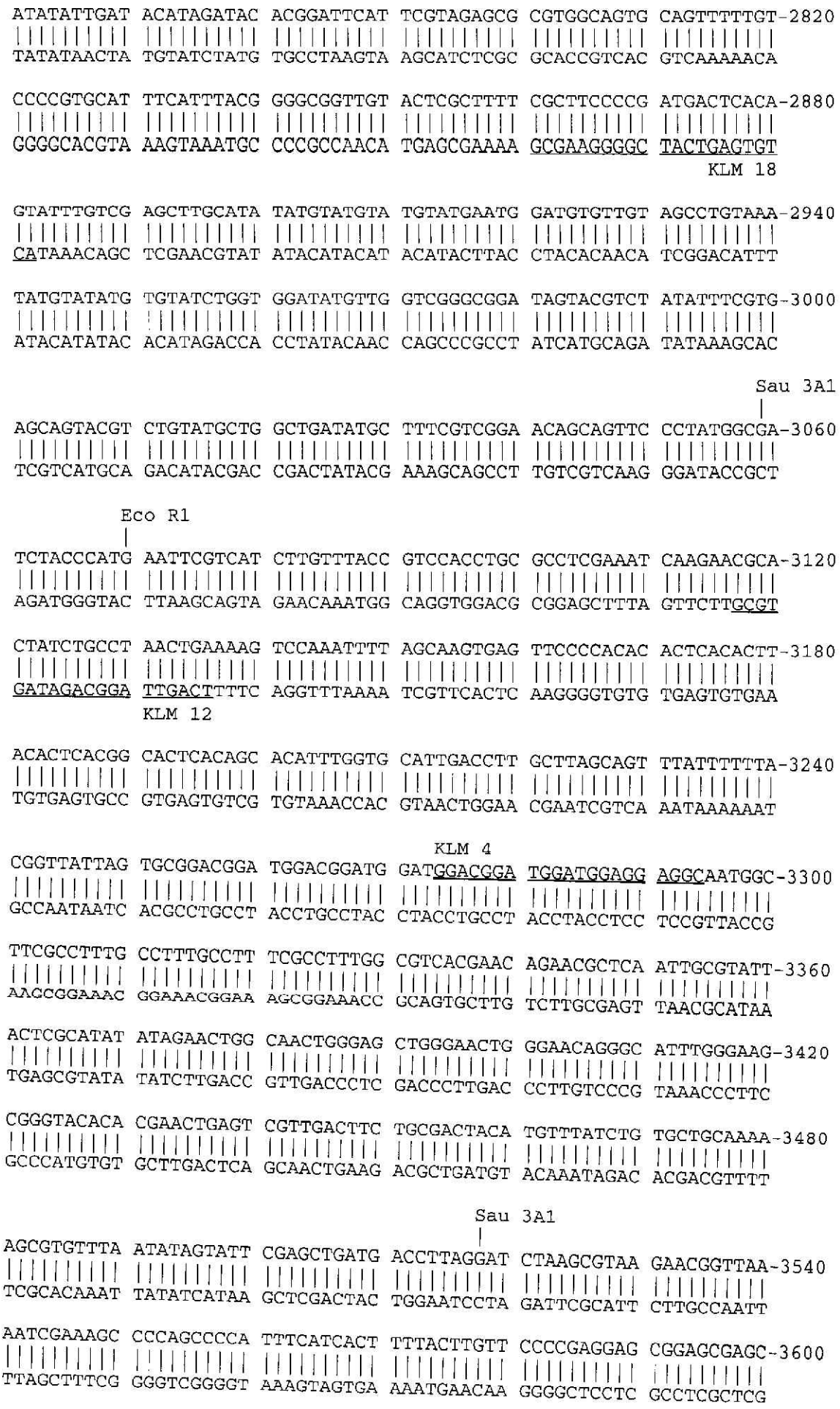
Annotated genomic sequence of the D*Sox60B* gene

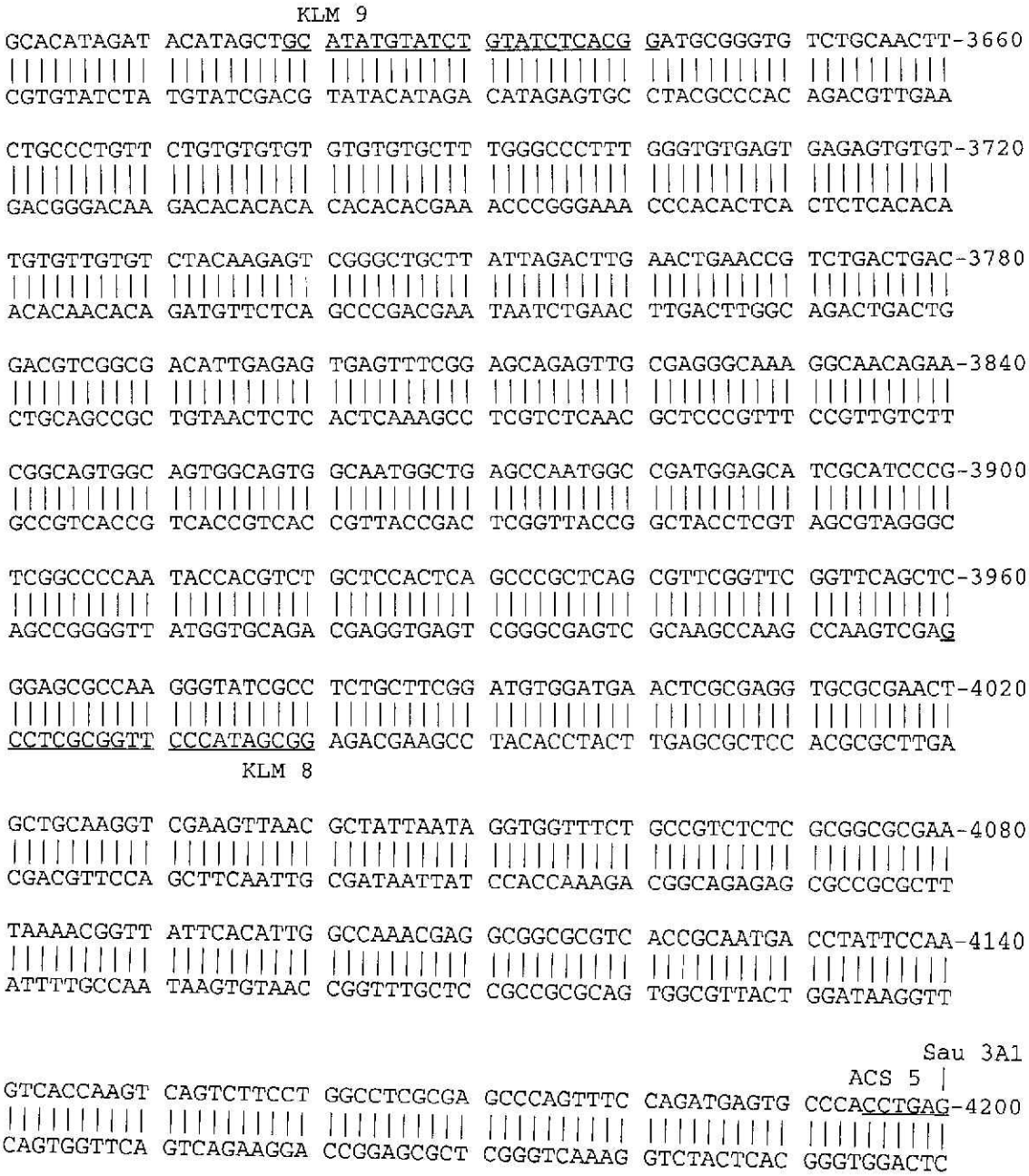


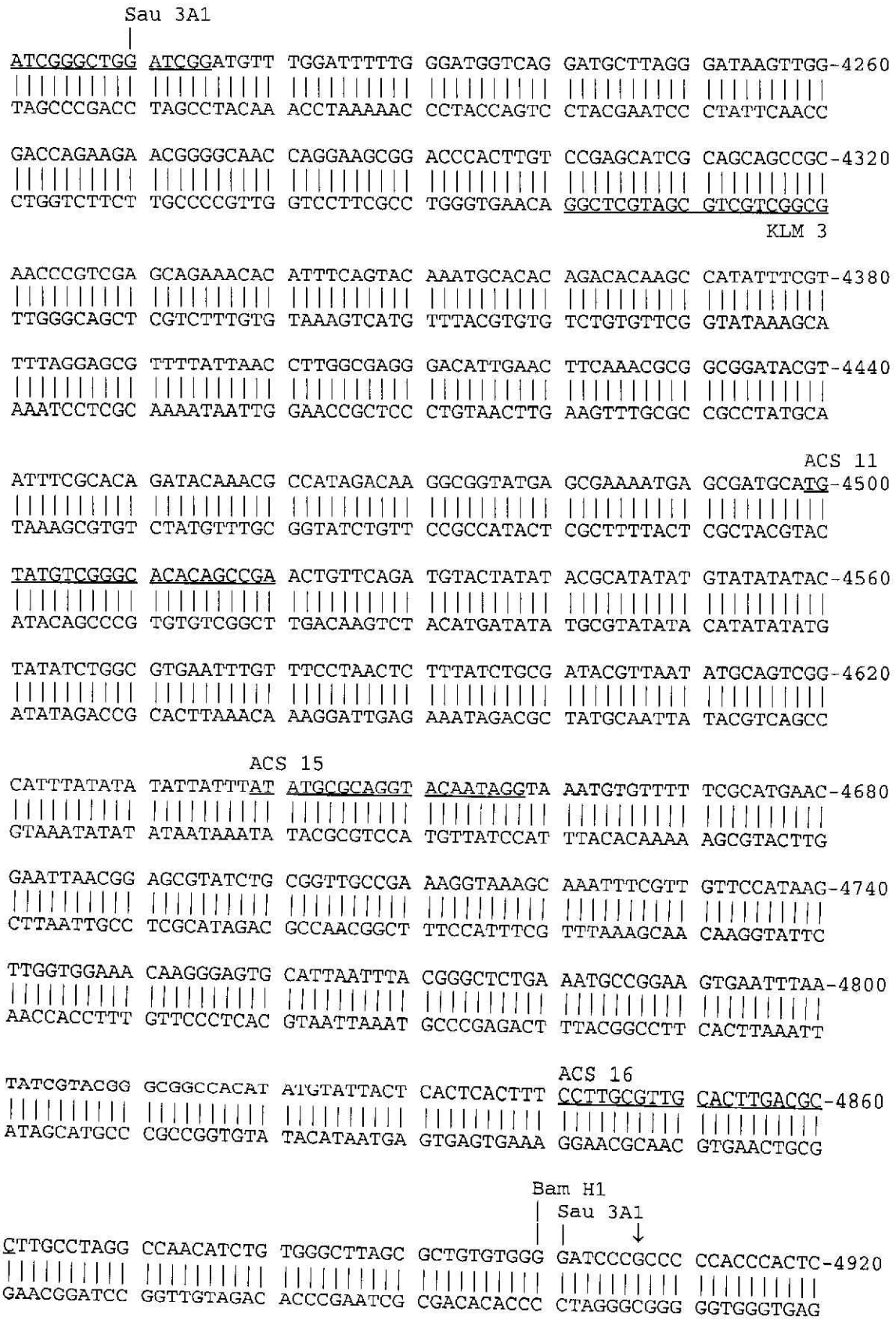

KLM 15
CAGTTCAGTT CGGTGTGGAT TCAGTTGGAT TCGAATCGAT GTTTAGTGTA GATTGGCCAG-720
GTCAAGTCAA GCCACACCTA AGTCAACCTA AGCTTAGCTA CAAATCACAT CTAACCGGTC
GACACGGAGA ACAAATAAGT CCCTCATCGC GCGGCGTGCA AAACCATCCG AACTAACTAA-780
CTGTGCCTCT TGTTTATTCA GGGAGTAGCG CGCCGCACGT TTTGGTAGGC TTGATTGATT
GTCTAGCCAA GTCTAGTAGC TAGAAACTAG AAACAAGAAA GCTACATACA TATGTGCGTA-840
CAGATCGGTT CAGATCATCG ATCTTTGATC TTTGTTCTTT CGATGTATGT ATACACGCAT
Sau 3A1
AACCGTGGAT GCCACAGAGC TGAGATACAT TTAACCAAGC GATCTCGCTT ATTGTCGAGT-900
TTGGCACCTA CGGTGTCTCG ACTCTATGTA AATTGGTTCG CTAGAGCGAA TAACAGCTCA
GATAACACTA AGATTGGGCA GAAACATTCG GAACCAGCTG GAGTTCAAAC ATTATCAAAA-960
CTATTGTGAT TCTAACCCGT CTTTGTAAGC CTTGGTCGAC CTCAAGTTTG TAATAGTTTT
Pst 1
CATTTGTGTT TCAAAATCAA CACACTTAAA GCAAACAAAC ACAGTCGAAA GCGCGAGTGC-1020
GTAAACACAA AGTTTTAGTT GTGTGAATTT CGTTTGTTTG TGTCAGCTTT CGCGCTCACG
KLM 21
TGCAGTATTA GTGGTGACCA CACACAGATA CACACATGCA CACGGCCATA TCGTTACATA-1080
ACGTCATAAT CACCACTGGT GTGTGTCTAT GTGTGTACGT GTGCCGGTAT AGCAATGTAT
CGTGTGTAAC TCTCTGCACA TAAATGCGAG TCCAATCTAT TTAATTCAGT TTCAGTGCAT-1140
GCACACATTG AGAGACGTGT ATTTACGCTC AGGTTAGATA AATTAAGTCA AAGTCACGTA
TTATAAATTA ATAAAATTTA CCAAGTGCGA GAGTGAGGGA GCGACACGCA CACCTATGCA-1200
AATATTTAAT TATTTTAAAT GGTTCACGCT CTCACTCCCT CGCTGTGCGT GTGGATACGT
CGCTCATTAC ACGGACGAGG CGAACGTGAG TCTTGGAATA TTGCAAACAA CACCAACAAC-1260
GCGAGTAATG TGCCTGCTCC GCTTGCACTC AGAACCTTAT AACGTTTGTT GTGGTTGTTG
AAAAACAACA ACAAGAACAA CAACAAAACC AAAAGCGAAA CAGCAAAAAA TAAATAAATA-1320
TTTTTGTTGT TGTTCTTGTT GTTGTTTTGG TTTTCGCTTT GTCGTTTTTT ATTTATTTAT
ACGAGGAACC AGTTTCACCT TGAGGAGCAA TACCTAGTGC ACACTCACAC TCGCACTCGC-1380
TGCTCCTTGG TCAAAGTGGA ACTCCTCGTT ATGGATCACG TGTGAGTGTG AGCGTGAGCG

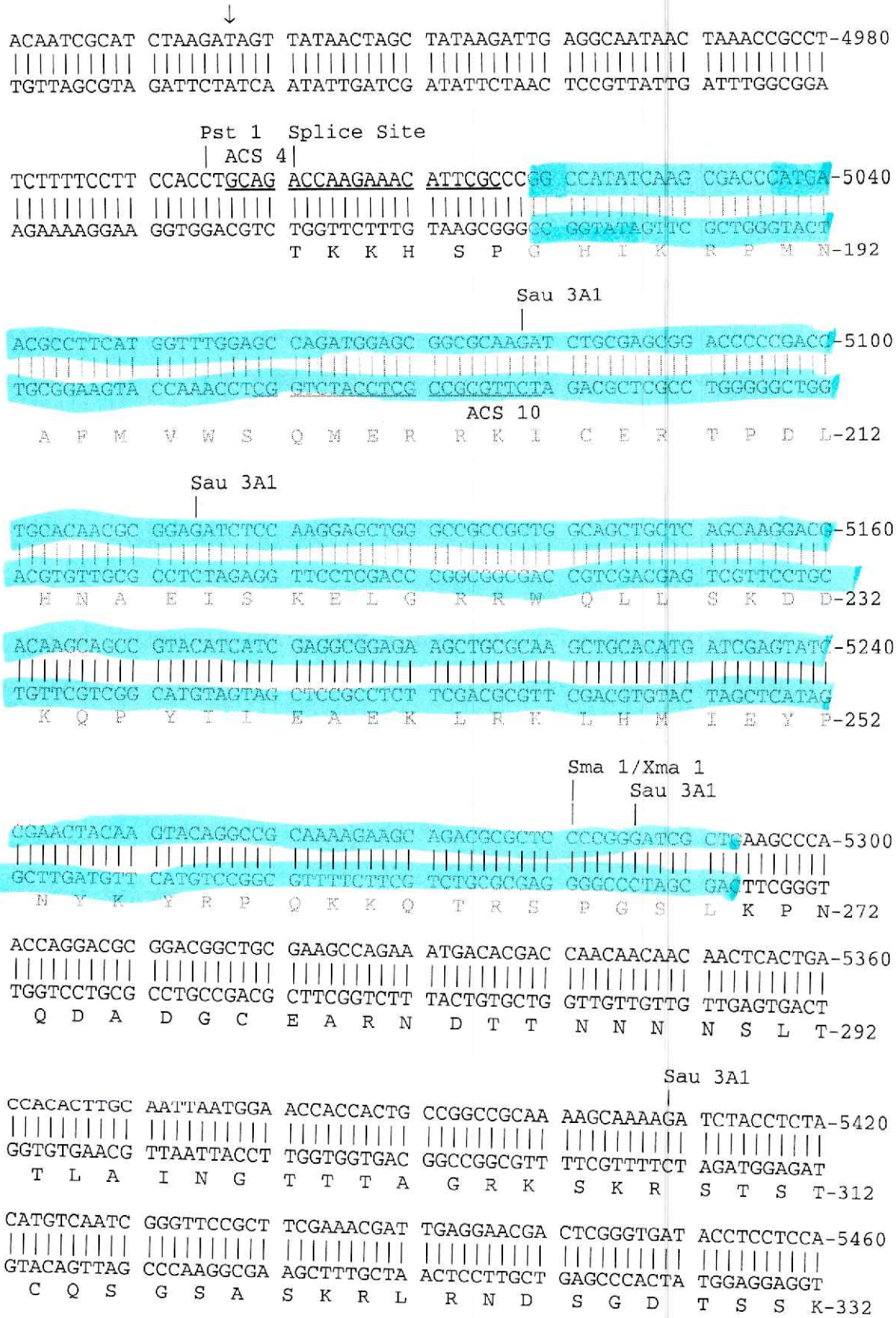

Sau 3A1
ATTCACACAA ATGAAACAGC CCGATCTTAC TCTTACTGCG AGTACGGACA CATAGTGCAC -1440
TAAGTGTGTT TACTTTGTCG GGCTAGAATG AGAATGACGC TCATGCCTGT GTATCACGTG
KLM 10
ATATAGTGCA TATAGTGCAC AGCACAGAGC ACAGAGTAGA CATAGTGACC ACCACATAAT -1500
TATATCACGT ATATCACGTG TCGTGTCTCG TGTCTCATCT GTATCACTGG TGGTGTATTA
TTCGTGATAA AGCCACAGAG AATCGGAGCG CTCCGCCTTA TCGGCAACCC ACTGCCACTG -1560
AAGCACTATT TCGGTGTCTC TTAGCCTCGC GAGGCGGAAT AGCCGTTGGG TGACGGTGAC
Sau 3A1
GTCCGGCTAC TATGCTCCAG CGGGGATCGG GACATCATCG CTGGGATAGA GACACAGTGG -1620
CAGGCCGATG ATACGAGGTC GCCCCTAGCC CTGTAGTAGC GACCCTATCT CTGTGTCACC
KLM 20
KLM 19
ACACCAGAAC TGGGATGCAG TTGCAGCGGC GCCAAACGCA TTGAAAGATG ATAGCTAAGC -1680
TGTGGTCTTG ACCCTACGTC AACGTCGCCG CGGTTTGCGT AACTTTCTAC TATCGATTCG
M I A K P-5
Hind III
CCAACCAGGC CACCACCGAA CCACCATTAA GCTTGCGCCC CGGAACAGTG CCAACGGTTC -1740
GGTTGGTCCG GTGGTGGCTT GGTGGTAATT CGAACGCGGG GCCTTGTCAC GGTTGCCAAG
KLM 6
N Q A T T E P P L S L R P G T V P T V P-25
CAGCAACCAC CCCAGCCAGA CCAGCGACCA TCACCATCCA GCGAAGGCAT CCAGCCCCGA -1800
GTCGTTGGTG GGGTCGGTCT GGTCGCTGGT AGTGGTAGGT CGCTTCCGTA GGTCGGGGCT
A T T P A R P A T I T I Q R R H P A P K-45
KLM 1
AAGCGGATTC CACACCCCAC ACTTTGCCAC CGTTCTCGCC TTCGCCTTCG CCAGCGTCGT -1860
TTCGCCTAAG GTGTGGGGTG TGAAACGGTG GCAAGAGCGG AAGCGGAAGC GGTCGCAGCA
A D S T P H T L P P F S P S P S P A S S-65
CGCCTTCGCC AGCGCCAGCG CAAACGCCTG GAGCACAAAA AACACAAAGC CAGGCAGCTA -1920
GCGGAAGCGG TCGCGGTCGC GTTTGCGGAC CTCGTGTTTT TTGTGTTTCG GTCCGTCGAT
P S P A P A Q T P G A Q K T Q S Q A A I-85
Pst 1
TTACTCATCC AGCGGCTGTG GCTTCGCCTT CCGCGCCTGT TGCTGCAGCT GCACCGAAGA -1980
AATGAGTAGG TCGCCGACAC CGAAGCGGAA GGCGCGGACA ACGACGTCGA CGTGGCTTCT
KLM 2
T H P A A V A S P S A P V A A A A P K T-105
CCCCCAAGAC CCCGGAACCC CGGAGTACCC ACACCCACAC TCACACTCAC AGCCAGCACT -2040
GGGGGTTCTG GGGCCTTGGG GCCTCATGGG TGTGGGTGTG AGTGTGAGTG TCGGTCGTGA
P K T P E P R S T H T H T H T H S Q H F-125

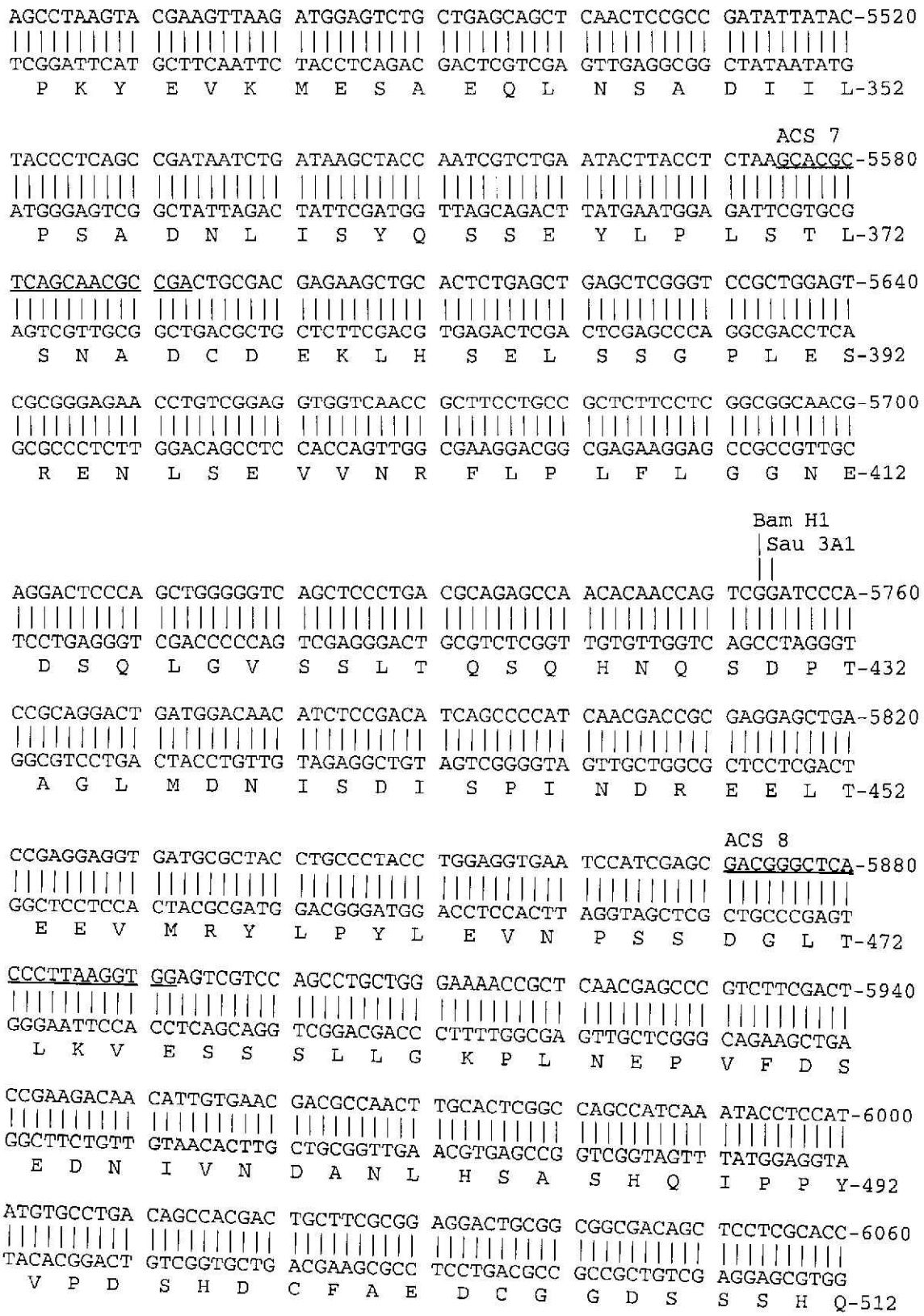

Sau 3A1
TCAGCCCCCC TCCACGCGAA TCCGAAATGG ACGGCGAAAG ATCTCCGAGC CACAGCGGCC-2100
AGTCGGGGGG AGGTGCGCTT AGGCTTTACC TGCCGCTTTC TAGAGGCTCG GTGTCGCCGG
S P P P R E S E M D G E R S P S H S G H-145
Sau 3A1
KLM 7
ATGAAATGAC ACTGAGCATG GACGGCATCG ATTCCAGCCT GGTGTTCGGA TCTGCACGGG-2160
TACTTTACTG TGACTCGTAC CTGCCGTAGC TAAGGTCGGA CCACAAGCCT AGACGTGCCC
E M T L S M D G I D S S L V F G S A R V-165
Splice Site
TTCCTGTCAA CTCCAGCACC CCGTACTCGG ATGCGACTCG AGTGAGTAAC ACTGTCTACA-2220
AAGGACAGTT GAGGTCGTGG GGCATGAGCC TACGCTGAGC TCACTCATTG TGACAGATGT
P V N S S T P Y S D A T R -178
Sau 3A1
CTGAGGGAAA TTGGGATCAA GTAGGAGGTA GGACTGTATA ACCCTTATTA TATTTCGGTC-2280
GACTCCCTTT AACCCTAGTT CATCCTCCAT CCTGACATAT TGGGAATAAT ATAAAGCCAG
TTCGACACCA TTTTCCCTAA GGTACATATT TCTGTCCAGG CTGGCAGGAA TGTGTTGTTG-2340
AAGCTGTGGT AAAAGGGATT CCATGTATAA AGACAGGTCC GACCGTCCTT ACACAACAAC
Sau 3A1
GTTGATTTGG TCATCTATCC ATCCGATTAG AAGATCCGCT CTATCATCCC TCGCCCGAGG-2400
CAACTAAACC AGTAGATAGG TAGGCTAATC TTCTAGGCGA GATAGTAGGG AGCGGGCTCC
KLM 17
GAACTACCAG ATATACATAC GAACCCAGCT GTCTGTCAGC GAGGCGAACG TATTTGATAT-2460
CTTGATGGTC TATATGTATG CTTGGGTCGA CAGACAGTCG CTCCGCTTGC ATAAACTATA
TCGGCAGGGA GACCTACTTA TAGAGCACAT ACAAATACTT CGCCAACCGG AAAGGTCAAC-2520
AGCCGTCCCT CTGGATGAAT ATCTCGTGTA TGTTTATGAA GCGGTTGGCC TTTCCAGTTG
AAAAGTTTGA CTGATAAGAT TCGGTCAGAT TAGATTCAAG TTCCTTGCGA TTATTTTCTT-2580
TTTTCAAACT GACTATTCTA AGCCAGTCTA ATCTAAGTTC AAGGAACGCT AATAAAAGAA
TGTTTACAAA TACCAGCAGG CACAACAATT ATAGCAGATT AGCAATCGAA TTGTAGCTCA-2640
ACAAATGTTT ATGGTCGTCC GTGTTGTTAA TATCGTCTAA TCGTTAGCTT AACATCGAGT
ATTGGAGTCT TAATATGATG AAATAGCCAG GATTGATAGA TGGTATCTTC GGGTGCACTG-2700
TAACCTCAGA ATTATACTAC TTTATCGGTC CTAACTATCT ACCATAGAAG CCCACGTGAC
Sau 3A1
GCCCATCGAG TGCTGATGAG TCGCGACCAT GGATGGATAC ATAGATGCAT AGATCCATAG-2760
CGGGTAGCTC ACGACTACTC AGCGCTGGTA CCTACCTATG TATCTACGTA TCTAGGTATC

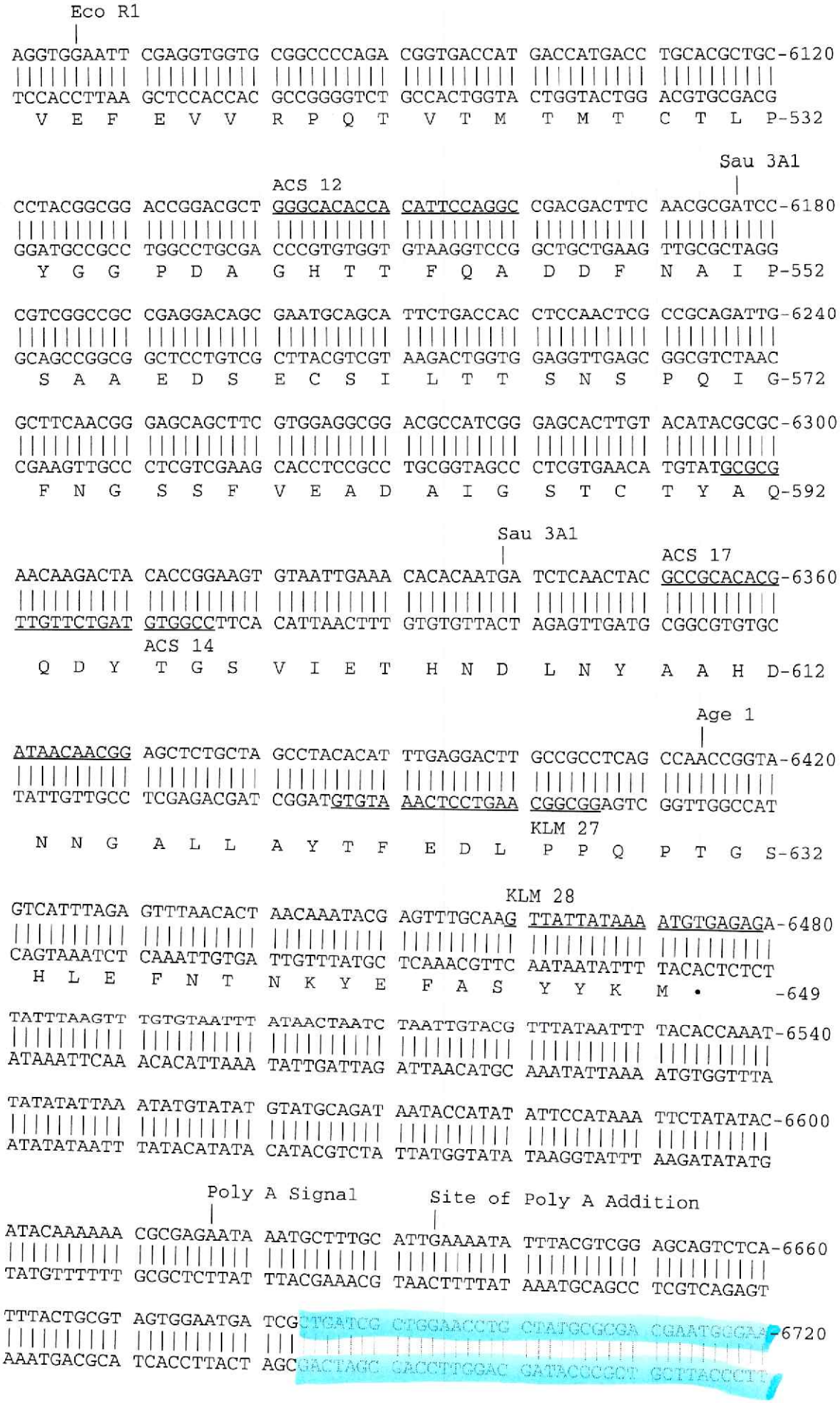

ATATATTGAT ACATAGATAC ACGGATTCAT TCGTAGAGCG CGTGGCAGTG CAGTTTTTGT-2820
TATATAACTA TGTATCTATG TGCCTAAGTA AGCATCTCGC GCACCGTCAC GTCAAAAACA
CCCCGTGCAT TTCATTTACG GGGCGGTTGT ACTCGCTTTT CGCTTCCCCG ATGACTCACA-2880
GGGGCACGTA AAGTAAATGC CCCGCCAACA TGAGCGAAAA GCGAAGGGGC TACTGAGTGT
KLM 18
GTATTTGTCG AGCTTGCATA TATGTATGTA TGTATGAATG GATGTGTTGT AGCCTGTAAA-2940
CATAAACAGC TCGAACGTAT ATACATACAT ACATACTTAC CTACACAACA TCGGACATTT
TATGTATATG TGTATCTGGT GGATATGTTG GTCGGGCGGA TAGTACGTCT ATATTTCGTG-3000
ATACATATAC ACATAGACCA CCTATACAAC CAGCCCGCCT ATCATGCAGA TATAAAGCAC
Sau 3A1
AGCAGTACGT CTGTATGCTG GCTGATATGC TTTCGTCGGA ACAGCAGTTC CCTATGGCGA-3060
TCGTCATGCA GACATACGAC CGACTATACG AAAGCAGCCT TGTCGTCAAG GGATACCGCT
Eco R1
TCTACCCATG AATTCGTCAT CTTGTTTACC GTCCACCTGC GCCTCGAAAT CAAGAACGCA-3120
AGATGGGTAC TTAAGCAGTA GAACAAATGG CAGGTGGACG CGGAGCTTTA GTTCTTGCGT
CTATCTGCCT AACTGAAAAG TCCAAATTTT AGCAAGTGAG TTCCCCACAC ACTCACACTT-3180
GATAGACGGA TTGACTTTTC AGGTTTAAAA TCGTTCACTC AAGGGGTGTG TGAGTGTGAA
KLM 12
ACACTCACGG CACTCACAGC ACATTTGGTG CATTGACCTT GCTTAGCAGT TTATTTTTTA-3240
TGTGAGTGCC GTGAGTGTCG TGTAAACCAC GTAACTGGAA CGAATCGTCA AATAAAAAAT
KLM 4
CGGTTATTAG TGCGGACGGA TGGACGGATG GATGGACGGA TGGATGGAGG AGGCAATGGC-3300
GCCAATAATC ACGCCTGCCT ACCTGCCTAC CTACCTGCCT ACCTACCTCC TCCGTTACCG
TTCGCCTTTG CCTTTGCCTT TCGCCTTTGG CGTCACGAAC AGAACGCTCA ATTGCGTATT-3360
AAGCGGAAAC GGAAACGGAA AGCGGAAACC GCAGTGCTTG TCTTGCGAGT TAACGCATAA
ACTCGCATAT ATAGAACTGG CAACTGGGAG CTGGGAACTG GGAACAGGGC ATTTGGGAAG-3420
TGAGCGTATA TATCTTGACC GTTGACCCTC GACCCTTGAC CCTTGTCCCG TAAACCCTTC
CGGGTACACA CGAACTGAGT CGTTGACTTC TGCGACTACA TGTTTATCTG TGCTGCAAAA-3480
GCCCATGTGT GCTTGACTCA GCAACTGAAG ACGCTGATGT ACAAATAGAC ACGACGTTTT
Sau 3A1
AGCGTGTTTA ATATAGTATT CGAGCTGATG ACCTTAGGAT CTAAGCGTAA GAACGGTTAA-3540
TCGCACAAAT TATATCATAA GCTCGACTAC TGGAATCCTA GATTCGCATT CTTGCCAATT
AATCGAAAGC CCCAGCCCCA TTTCATCACT TTTACTTGTT CCCCGAGGAG CGGAGCGAGC-3600
TTAGCTTTCG GGGTCGGGGT AAAGTAGTGA AAATGAACAA GGGGCTCCTC GCCTCGCTCG

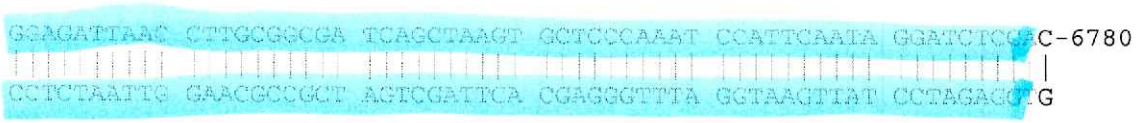

KLM 9
GCACATAGAT ACATAGCTGC ATATGTATCT GTATCTCACG GATGCGGGTG TCTGCAACTT-3660
CGTGTATCTA TGTATCGACG TATACATAGA CATAGAGTGC CTACGCCCAC AGACGTTGAA
CTGCCCTGTT CTGTGTGTGT GTGTGTGCTT TGGGCCCTTT GGGTGTGAGT GAGAGTGTGT-3720
GACGGGACAA GACACACACA CACACACGAA ACCCGGGAAA CCCACACTCA CTCTCACACA
TGTGTTGTGT CTACAAGAGT CGGGCTGCTT ATTAGACTTG AACTGAACCG TCTGACTGAC-3780
ACACAACACA GATGTTCTCA GCCCGACGAA TAATCTGAAC TTGACTTGGC AGACTGACTG
GACGTCGGCG ACATTGAGAG TGAGTTTCGG AGCAGAGTTG CGAGGGCAAA GGCAACAGAA-3840
CTGCAGCCGC TGTAACTCTC ACTCAAAGCC TCGTCTCAAC GCTCCCGTTT CCGTTGTCTT
CGGCAGTGGC AGTGGCAGTG GCAATGGCTG AGCCAATGGC CGATGGAGCA TCGCATCCCG-3900
GCCGTCACCG TCACCGTCAC CGTTACCGAC TCGGTTACCG GCTACCTCGT AGCGTAGGGC
TCGGCCCCAA TACCACGTCT GCTCCACTCA GCCCGCTCAG CGTTCGGTTC GGTTCAGCTC-3960
AGCCGGGGTT ATGGTGCAGA CGAGGTGAGT CGGGCGAGTC GCAAGCCAAG CCAAGTCGAG
GGAGCGCCAA GGGTATCGCC TCTGCTTCGG ATGTGGATGA ACTCGCGAGG TGCGCGAACT-4020
CCTCGCGGTT CCCATAGCGG AGACGAAGCC TACACCTACT TGAGCGCTCC ACGCGCTTGA
KLM 8
GCTGCAAGGT CGAAGTTAAC GCTATTAATA GGTGGTTTCT GCCGTCTCTC GCGGCGCGAA-4080
CGACGTTCCA GCTTCAATTG CGATAATTAT CCACCAAAGA CGGCAGAGAG CGCCGCGCTT
TAAAACGGTT ATTCACATTG GCCAAACGAG GCGGCGCGTC ACCGCAATGA CCTATTCCAA-4140
ATTTTGCCAA TAAGTGTAAC CGGTTTGCTC CGCCGCGCAG TGGCGTTACT GGATAAGGTT
Sau 3A1
ACS 5
GTCACCAAGT CAGTCTTCCT GGCCTCGCGA GCCCAGTTTC CAGATGAGTG CCCACCTGAG-4200
CAGTGGTTCA GTCAGAAGGA CCGGAGCGCT CGGGTCAAAG GTCTACTCAC GGGTGGACTC


Sau 3A1
ATCGGGCTGG ATCGGATGTT TGGATTTTTG GGATGGTCAG GATGCTTAGG GATAAGTTGG-4260
TAGCCCGACC TAGCCTACAA ACCTAAAAAC CCTACCAGTC CTACGAATCC CTATTCAACC
GACCAGAAGA ACGGGGCAAC CAGGAAGCGG ACCCACTTGT CCGAGCATCG CAGCAGCCGC-4320
CTGGTCTTCT TGCCCCGTTG GTCCTTCGCC TGGGTGAACA GGCTCGTAGC GTCGTCGGCG
KLM 3
AACCCGTCGA GCAGAAACAC ATTTCAGTAC AAATGCACAC AGACACAAGC CATATTTCGT-4380
TTGGGCAGCT CGTCTTTGTG TAAAGTCATG TTTACGTGTG TCTGTGTTCG GTATAAAGCA
TTTAGGAGCG TTTTATTAAC CTTGGCGAGG GACATTGAAC TTCAAACGCG GCGGATACGT-4440
AAATCCTCGC AAAATAATTG GAACCGCTCC CTGTAACTTG AAGTTTGCGC CGCCTATGCA
ACS 11
ATTTCGCACA GATACAAACG CCATAGACAA GGCGGTATGA GCGAAAATGA GCGATGCATG-4500
TAAAGCGTGT CTATGTTTGC GGTATCTGTT CCGCCATACT CGCTTTTACT CGCTACGTAC
TATGTCGGGC ACACAGCCGA ACTGTTCAGA TGTACTATAT ACGCATATAT GTATATATAC-4560
ATACAGCCCG TGTGTCGGCT TGACAAGTCT ACATGATATA TGCGTATATA CATATATATG
TATATCTGGC GTGAATTTGT TTCCTAACTC TTTATCTGCG ATACGTTAAT ATGCAGTCGG-4620
ATATAGACCG CACTTAAACA AAGGATTGAG AAATAGACGC TATGCAATTA TACGTCAGCC
ACS 15
CATTTATATA TATTATTTAT ATGCGCAGGT ACAATAGGTA AATGTGTTTT TCGCATGAAC-4680
GTAAATATAT ATAATAAATA TACGCGTCCA TGTTATCCAT TTACACAAAA AGCGTACTTG
GAATTAACGG AGCGTATCTG CGGTTGCCGA AAGGTAAAGC AAATTTCGTT GTTCCATAAG-4740
CTTAATTGCC TCGCATAGAC GCCAACGGCT TTCCATTTCG TTTAAAGCAA CAAGGTATTC
TTGGTGGAAA CAAGGGAGTG CATTAATTTA CGGGCTCTGA AATGCCGGAA GTGAATTTAA-4800
AACCACCTTT GTTCCCTCAC GTAATTAAAT GCCCGAGACT TTACGGCCTT CACTTAAATT
ACS 16
TATCGTACGG GCGGCCACAT ATGTATTACT CACTCACTTT CCTTGCGTTG CACTTGACGC-4860
ATAGCATGCC CGCCGGTGTA TACATAATGA GTGAGTGAAA GGAACGCAAC GTGAACTGCG
Bam H1
Sau 3A1
CTTGCCTAGG CCAACATCTG TGGGCTTAGC GCTGTGTGGG GATCCCGCCC CCACCCACTC-4920
GAACGGATCC GGTTGTAGAC ACCCGAATCG CGACACACCC CTAGGGCGGG GGTGGGTGAG


ACAATCGCAT CTAAGATAGT TATAACTAGC TATAAGATTG AGGCAATAAC TAAACCGCCT-4980
TGTTAGCGTA GATTCTATCA ATATTGATCG ATATTCTAAC TCCGTTATTG ATTTGGCGGA
Pst 1 Splice Site
ACS 4
TCTTTTCCTT CCACCTGCAG ACCAAGAAAC ATTCGCCCGG CCATATCAAG CGACCCATGA-5040
AGAAAAGGAA GGTGGACGTC TGGTTCTTTG TAAGCGGGCC GGTATAGTTC GCTGGGTACT
T K K H S P G H I K R P M N-192
Sau 3A1
ACGCCTTCAT GGTTTGGAGC CAGATGGAGC GGCGCAAGAT CTGCGAGCGG ACCCCCGACC-5100
TGCGGAAGTA CCAAACCTCG GTCTACCTCG CCGCGTTCTA GACGCTCGCC TGGGGGCTGG
ACS 10
A F M V W S Q M E R R K I C E R T P D L-212
Sau 3A1
TGCACAACGC GGAGATCTCC AAGGAGCTGG GCCGCCGCTG GCAGCTGCTC AGCAAGGACG-5160
ACGTGTTGCG CCTCTAGAGG TTCCTCGACC CGGCGGCGAC CGTCGACGAG TCGTTCCTGC
H N A E I S K E L G R R W Q L L S K D D-232
ACAAGCAGCC GTACATCATC GAGGCGGAGA AGCTGCGCAA GCTGCACATG ATCGAGTATC-5240
TGTTCGTCGG CATGTAGTAG CTCCGCCTCT TCGACGCGTT CGACGTGTAC TAGCTCATAG
K Q P Y I I E A E K L R K L H M I E Y P-252
Sma 1/Xma 1
Sau 3A1
CGAACTACAA GTACAGGCCG CAAAAGAAGC AGACGCGCTC CCCGGGATCG CTCAAGCCCA-5300
GCTTGATGTT CATGTCCGGC GTTTTCTTCG TCTGCGCGAG GGGCCCTAGC GAGTTCGGGT
N Y K Y R P Q K K Q T R S P G S L K P N-272
ACCAGGACGC GGACGGCTGC GAAGCCAGAA ATGACACGAC CAACAACAAC AACTCACTGA-5360
TGGTCCTGCG CCTGCCGACG CTTCGGTCTT TACTGTGCTG GTTGTTGTTG TTGAGTGACT
Q D A D G C E A R N D T T N N N N S L T-292
Sau 3A1
CCACACTTGC AATTAATGGA ACCACCACTG CCGGCCGCAA AAGCAAAAGA TCTACCTCTA-5420
GGTGTGAACG TTAATTACCT TGGTGGTGAC GGCCGGCGTT TTCGTTTTCT AGATGGAGAT
T L A I N G T T T A G R K S K R S T S T-312
CATGTCAATC GGGTTCCGCT TCGAAACGAT TGAGGAACGA CTCGGGTGAT ACCTCCTCCA-5460
GTACAGTTAG CCCAAGGCGA AGCTTTGCTA ACTCCTTGCT GAGCCCACTA TGGAGGAGGT
C Q S G S A S K R L R N D S G D T S S K-332


AGCCTAAGTA CGAAGTTAAG ATGGAGTCTG CTGAGCAGCT CAACTCCGCC GATATTATAC-5520
TCGGATTCAT GCTTCAATTC TACCTCAGAC GACTCGTCGA GTTGAGGCGG CTATAATATG
P K Y E V K M E S A E Q L N S A D I I L-352
ACS 7
TACCCTCAGC CGATAATCTG ATAAGCTACC AATCGTCTGA ATACTTACCT CTAAGCACGC-5580
ATGGGAGTCG GCTATTAGAC TATTCGATGG TTAGCAGACT TATGAATGGA GATTCGTGCG
P S A D N L I S Y Q S S E Y L P L S T L-372
TCAGCAACGC CGACTGCGAC GAGAAGCTGC ACTCTGAGCT GAGCTCGGGT CCGCTGGAGT-5640
AGTCGTTGCG GCTGACGCTG CTCTTCGACG TGAGACTCGA CTCGAGCCCA GGCGACCTCA
S N A D C D E K L H S E L S S G P L E S-392
CGCGGGAGAA CCTGTCGGAG GTGGTCAACC GCTTCCTGCC GCTCTTCCTC GGCGGCAACG-5700
GCGCCCTCTT GGACAGCCTC CACCAGTTGG CGAAGGACGG CGAGAAGGAG CCGCCGTTGC
R E N L S E V V N R F L P L F L G G N E-412
Bam H1
Sau 3A1
AGGACTCCCA GCTGGGGGTC AGCTCCCTGA CGCAGAGCCA ACACAACCAG TCGGATCCCA-5760
TCCTGAGGGT CGACCCCCAG TCGAGGGACT GCGTCTCGGT TGTGTTGGTC AGCCTAGGGT
D S Q L G V S S L T Q S Q H N Q S D P T-432
CCGCAGGACT GATGGACAAC ATCTCCGACA TCAGCCCCAT CAACGACCGC GAGGAGCTGA-5820
GGCGTCCTGA CTACCTGTTG TAGAGGCTGT AGTCGGGGTA GTTGCTGGCG CTCCTCGACT
A G L M D N I S D I S P I N D R E E L T-452
ACS 8
CCGAGGAGGT GATGCGCTAC CTGCCCTACC TGGAGGTGAA TCCATCGAGC GACGGGCTCA-5880
GGCTCCTCCA CTACGCGATG GACGGGATGG ACCTCCACTT AGGTAGCTCG CTGCCCGAGT
E E V M R Y L P Y L E V N P S S D G L T-472
CCCTTAAGGT GGAGTCGTCC AGCCTGCTGG GAAAACCGCT CAACGAGCCC GTCTTCGACT-5940
GGGAATTCCA CCTCAGCAGG TCGGACGACC CTTTTGGCGA GTTGCTCGGG CAGAAGCTGA
L K V E S S S L L G K P L N E P V F D S
CCGAAGACAA CATTGTGAAC GACGCCAACT TGCACTCGGC CAGCCATCAA ATACCTCCAT-6000
GGCTTCTGTT GTAACACTTG CTGCGGTTGA ACGTGAGCCG GTCGGTAGTT TATGGAGGTA
E D N I V N D A N L H S A S H Q I P P Y-492
ATGTGCCTGA CAGCCACGAC TGCTTCGCGG AGGACTGCGG CGGCGACAGC TCCTCGCACC-6060
TACACGGACT GTCGGTGCTG ACGAAGCGCC TCCTGACGCC GCCGCTGTCG AGGAGCGTGG
V P D S H D C F A E D C G G D S S S H Q-512


Eco R1
AGGTGGAATT CGAGGTGGTG CGGCCCCAGA CGGTGACCAT GACCATGACC TGCACGCTGC-6120
TCCACCTTAA GCTCCACCAC GCCGGGGTCT GCCACTGGTA CTGGTACTGG ACGTGCGACG
V E F E V V R P Q T V T M T M T C T L P-532
Sau 3A1
ACS 12
CCTACGGCGG ACCGGACGCT GGGCACACCA CATTCCAGGC CGACGACTTC AACGCGATCC-6180
GGATGCCGCC TGGCCTGCGA CCCGTGTGGT GTAAGGTCCG GCTGCTGAAG TTGCGCTAGG
Y G G P D A G H T T F Q A D D F N A I P-552
CGTCGGCCGC CGAGGACAGC GAATGCAGCA TTCTGACCAC CTCCAACTCG CCGCAGATTG-6240
GCAGCCGGCG GCTCCTGTCG CTTACGTCGT AAGACTGGTG GAGGTTGAGC GGCGTCTAAC
S A A E D S E C S I L T T S N S P Q I G-572
GCTTCAACGG GAGCAGCTTC GTGGAGGCGG ACGCCATCGG GAGCACTTGT ACATACGCGC-6300
CGAAGTTGCC CTCGTCGAAG CACCTCCGCC TGCGGTAGCC CTCGTGAACA TGTATGCGCG
F N G S S F V E A D A I G S T C T Y A Q-592
Sau 3A1
ACS 17
AACAAGACTA CACCGGAAGT GTAATTGAAA CACACAATGA TCTCAACTAC GCCGCACACG-6360
TTGTTCTGAT GTGGCCTTCA CATTAACTTT GTGTGTTACT AGAGTTGATG CGGCGTGTGC
ACS 14
Q D Y T G S V I E T H N D L N Y A A H D-612
Age 1
ATAACAACGG AGCTCTGCTA GCCTACACAT TTGAGGACTT GCCGCCTCAG CCAACCGGTA-6420
TATTGTTGCC TCGAGACGAT CGGATGTGTA AACTCCTGAA CGGCGGAGTC GGTTGGCCAT
KLM 27
N N G A L L A Y T F E D L P P Q P T G S-632
KLM 28
GTCATTTAGA GTTTAACACT AACAAATACG AGTTTGCAAG TTATTATAAA ATGTGAGAGA-6480
CAGTAAATCT CAAATTGTGA TTGTTTATGC TCAAACGTTC AATAATATTT TACACTCTCT
H L E F N T N K Y E F A S Y Y K M • -649
TATTTAAGTT TGTGTAATTT ATAACTAATC TAATTGTACG TTTATAATTT TACACCAAAT-6540
ATAAATTCAA ACACATTAAA TATTGATTAG ATTAACATGC AAATATTAAA ATGTGGTTTA
TATATATTAA ATATGTATAT GTATGCAGAT AATACCATAT ATTCCATAAA TTCTATATAC-6600
ATATATAATT TATACATATA CATACGTCTA TTATGGTATA TAAGGTATTT AAGATATATG
Poly A Signal
Site of Poly A Addition
ATACAAAAAA CGCGAGAATA AATGCTTTGC ATTGAAAATA TTTACGTCGG AGCAGTCTCA-6660
TATGTTTTTT GCGCTCTTAT TTACGAAACG TAACTTTTAT AAATGCAGCC TCGTCAGAGT
TTTACTGCGT AGTGGAATGA TCG
AAATGACGCA TCACCTTACT AGC



ATATGTTATT-6790
||||||||||
TATACAATAA

KEY

↓	Transcriptional start point determined by 5′ primer extension of mRNA
→	Start site of the LD30105 cDNA cosmid clone
Green	HMG box
Cyan	3'-UTR of the PHM gene
Underlined	Site of a sequencing primer