



SdaI (6)
1 CCTGCAGGCGTTACATAA...
NdeI (182)
101 CGCCAATAGGACTTTCCATTGACGTCAATGGTGGAGTATTTACGGTAAACTGCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCC
201 TATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATC
301 GCTATTACCATGATGATGCGGTTTTGGCAGTACATCAATGGCGTGGATAGCGGTTTGACTCACGGGATTTCGAAGTCTCCACCCATTGACGTCAATG
SpeI (414)
401 GGAGTTTGTGTTGACTAGTCAGGGCCCCAACCCCCAAGCCCCATTTCACAACACGCTGGCGCTACAGGCGGTGACTTCCCCTTGCTTTGGGGCGGG
BspEI (520)
501 GGGCTGAGACTCCTATGTGCTCCGATTGGTCAGGCACGGCCTTCGGCCCCCTCTGCCACCGAGATTGGCCGCTAGGCCTCCCCGAGCGCCCTGCC
HindIII (679)
601 TCCGAGGGCCGGCGCACCATAAAAGAAGCCGCTAGCCACGCTCCCTCGAGTTGGCGGTCCCGGGTCTGTCTCAAGCTTGCCGCCAGAACACAGg
XmnI (763)
701 taagtccgtgtgtggttcccgcgggctggcctctttacgggttatggccttgcgtgcctgaattacttccatgccctggctgcagtacgtgattc
801 ttgatcccagcctcgggttgaagtgggtgggagagttcgagccctgcgcttaaggagcccttcgctcgtgcttgagttgagccctggcttggggc
901 ctggggccgcccgtgctaactctggtggcaccttcgagcctgtctcgtgctttcgctaagtctctagccatttaaaattttgataaccagctgcgagc
BglIII (1039)
1001 ctttttttctggcgagatagcttctgtaaatgcgggccaagatctgcacactgggtatttcgggttttggggccgcgggcgagccgggcccgtgcgtccc
1101 agcgacatgctcggcgagggcggggcctgcgagcggccaccaggagaatcggacgggggtagctcdaaacggccgctgctggtgctggcctgcg
EcoNI (1290)
1201 gccgccgtgatcgccccctggcggaagcctggccggctggcaccagttgctgtagcggaagatggccgcttcccggcctgctgcagggagc
SmaI (1323)
1301 tcaaaatggaggacgcggcgcccgggagagcgggctggagtcaccacacaaaggaaaaggcccttctctcatcctcgcttcatgtgactcca
XhoI (1438)
1401 cggagtagcgggcccgtccaggcacctcgattagttctcgagcttttgagtagctcgtcttttaggttggggggaggggttttatgcatggagtttcc
1501 ccacactgagtggggtggagactgaagagttaggccagcttggcacttgatgtaattctccttggaaattgcccttttgagtttggatcttgcctcatc
NotI (1673)
1601 tcaagcctcagacagtggttcaaagttttttcttccatttcagGTGTCGTGAAAACACTACCCTAAAAGCCACCATGGGGGTTCTCATCATCATCA
MetGI yGI ySer Hi sHi sHi sHi
NheI (1711) **Acc65I (1767)**
1701 TCATGGTATGGCTAGCATGACTGGTGACAGCAAATGGTCCGGATCTGTACGACGATGACGATAAGGTACCTAAGGATCAGCTTGGAGTTGATCCCGTC
9> sHi sGI yMetAl aSer Met Thr GI yGI yGI nGI nMetGI yArgAspLeuTyrAspAspAspAspLysVal IP roLysAspGI nLeuGI yVal AspProVal
1801 GTTTTACAACGTCGTGACTGGGAAAACCTGGCGTTACCAACTTAATCGCCTTGACGACATCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCC
43> Val LeuGI nArgArgAspTrpGI uAsnProGI yVal Thr GI nLeuAsnArgLeuAl aAl aHi sProProPheAl aSer TrpArgAsnSer GI uGI uAl aA
FspI (1927)
1901 GCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCTTTCCTGGTTCCGGCACCAGAAGCGGTGCCGAAAGCTGGCTGGAGTG
76> rgThrAspArgProSer GI nGI nLeuArgSer LeuAsnGI yGI uTrpArgPheAl aTrpPheProAl aProGI uAl aVal IP roGI uSer TrpLeuGI uCy
2001 CGATCTTCTGAGCCGATCTGCTGCTCCCTCAAACCTGGCAGATGCACGTTACGATGCGCCCATACACCAACGTAACCTATCCCATTACGGTC
109> sAspLeuProGI uAl aAspThr Val Val Val IP roSerAsnTrpGI nMetHi sGI yTyrAspAl aProI l eTyrThrAsnVal Thr TyrProI l eThr Val
2101 AATCCGCGTGTGTTCCACGAGAAATCCGACGGTGTACTCGCTCACATTAATGTTGATGAAAGCTGGCTACAGGAAGGCCAGACGCGAATTATTT
143> AsnProProPheVal IP roThr GI uAsnProThr GI yCysTyrSer LeuThr PheAsnVal AspGI uSer TrpLeuGI nGI uGI yGI nThr ArgI l eI l eP
2201 TTGATGGCGTTAACTCGGCTTTCATCTGTGGTCAACGGCGCTGGGTACGGCCAGGACAGTCTGTTGCGCTGAATTTGACCTGAGCGCATT
176> heAspGI yVal lAsnSer Al aPheHi sLeuTrpCysAsnGI yArgTrpVal GI yTyrGI yGI nAspSer ArgLeuProSer GI uPheAspLeuSer Al aPh
2301 TTTACGCGCCGGAGAAAACCGCTCGCGGTGATGGTGTGCGTTGGAGTGACGGCAGTTATCTGGAAGATCAGGATATGTGGCGATGAGCGCATTTTC
209> eLeuArgAl aGI yGI uAsnArgLeuAl aVal MetVal LeuArgTrpSerAspGI ySer TyrLeuGI uAspGI nAspMetTrpArgMetSer GI yI l ePhe
2401 CGTGACGTCTGTTGCTGCATAAACCGACTACACAAATCAGCGATTTCCATGTTGCCACTCGCTTAAATGATGATTTACGCCGCTGACTGGAGGCTG
243> yGI uI l eI l eAspGI uArgGI yGI yTyrAl aAspArgVal Thr LeuArgLeuAsnVal GI uAsnProLysLeuTrpSer Al aGI uI l eP roAsnLeuTyr
2501 AAGTTCAGATGTGCGGCGAGTTGCGTGACTACCTACGGTAACAGTTTCTTTATGGCAGGTTGAAACGCAGGTCGCCAGCGGCACCGCCTTTCGGCGG
276> l uVal GI nMetCysGI yGI uLeuArgAspTyrLeuArgVal Thr Val Ser LeuTrpGI nGI yGI uThr GI nVal lAl aSer GI yThr Al aProPheGI yGI
2601 TGAATATCGATGAGCGTGGTGTATGCCGATCGCGTCACTACGCTGAACGTCGAAAACCCGAAACTGTGGAGCGCCGAAATCCCGAATCTCTAT
309> yGI uI l eI l eAspGI uArgGI yGI yTyrAl aAspArgVal Thr LeuArgLeuAsnVal GI uAsnProLysLeuTrpSer Al aGI uI l eP roAsnLeuTyr
2701 CGTGCGGTGGTTGAACTGCACCCGCGACGGCAGCTGATTGAAGCAGAAGCTGCGATGTCGGTTCCGCGAGGTGCGGATTGAAATGGTCTGCTGC
343> ArgAl aVal Val GI uLeuHi sThr Al aAspGI yThr LeuI l eGI uAl aGI uAl aCysAspVal GI yPheArgGI uVal ArgI l eGI uAsnGI yLeuLeu
EcoRV (2)
2801 TGCTGAACGGCAAGCCGTTGCTGATTGAGGCGTTAACCGTCACGAGCATCATCTCTGCATGGTCAGGTCATGGATGAGCAGACGATGGTGCAGGAT
376> euLeuAsnGI yLysProLeuLeuI l eArgGI yVal lAsnArgHi sGI uHi sHi sProLeuHi sGI yGI nVal MetAspGI uGI nThr MetVal GI nAspI l
2901 CCTGCTGATGAAGCAGAACAACCTTAAACCGCTGCGCTGTTCCGATTATCGAACCATCCGCTGTGGTACACGCTGTGCGACCGCTACGGCCTGTATGTG
409> eLeuLeuMetLysGI nAsnAsnPheAsnAl aVal lArgCysSerHi sTyrProAsnHi sProLeuTrpTyrThr LeuCysAspArgTyrGI yLeuTyrVal
3001 GTGGATGAAGCAATATTGAAACCCACGGCATGGTGCCAAATGAATCGTCTGACCGATGATCCGCGCTGGCTACCGCGATGAGCGAACCGGTAACCGGAA

443. Val Asp Gl uAl aAsn l l eGl uThr Hi sGl yMet Val l Pro Me tAsn Arg Leu Thr Asp Asp P ro Arg T rp Leu P ro Al a Me tSer Gl uArg Val l Thr Arg M
3101 TGGTGCAGCGCATCGTAATCACCCGAGTGTGATCATCTGGTCGCTGGGAAATGAATCAGGCCACGGCGCTAATCACGACGCGTGTATCGCTGGATCAA
476▶ e tVal l nArg Asp Arg Asn Hi s P ro Ser Val l l e l l e T rp Ser Leu Gl yAsn Gl uSer Gl yHi sGl yAl aAsn Hi sAsp Al aLeu Tyr Arg T rp l l eLy
3201 ATCTGTGCATCCTCCCGCCCGGTGCAGTATGAAGGCGCGGAGCCGACACCGCCACCGATATTATTGGCCGATGTACGCGCGGTGGATGAAGAC
509▶ sSer Val Asp P ro Ser Arg P ro Val l Gl nTyr Gl uGl yGl yAl aAsp Thr Thr Al aThr Asp l l e l l e Cys P ro Me tTyr Al aArg Val l Asp Gl uAsp
3301 CAGCCCTCCCGGCTGTGCCGAAATGGTCCATCAAAAAATGGCTTTCGCTACCTGGAGAGACGCGCCCGTGTATCCTTTGCGAATACGCCACGCGATGG
543▶ Gl nP roPhe P ro Al aVal l P ro Lys T rp Ser l l eLys Lys T rp Leu Ser Leu P ro Gl yGl uThr Arg P ro Leu l l eLeu Cys Gl uTyr Al aHi sAl aMe tG
3401 GTAACAGTCTTGGCGTTCGCTAAATAC TGGCAGGCGTTCCTGTCAGTATCCCGTTTACAGGGCGGCTTCTGCTGGGACTGGGTGGATCAGTCGCTGAT
576▶ l yAsn Ser Leu Gl yGl yPhe Al aLys Tyr T rp Gl nAl aPhe Arg Gl nTyr P ro Arg Leu Gl nGl yGl yPhe Val l T rp Asp T rp Val l Asp Gl nSer Leu l l
3501 TAAATATGATGAAAACGGCAACCCGTGGTCGGCTTACGGCGGTGATTTGGCGATACGCCAACGATCGCCAGTCTGTATGAACGGTCTGGTCTTTGCC
609▶ eLys Tyr Asp Gl uAsn Gl yAsn P ro T rp Ser Al aTyr Gl yGl yAsp Phe Gl yAsp Thr P ro Asn Asp Arg Gl nPhe Cys Me tAsn Gl yLeu Val l Phe Al a

Eco47III (3620)

3601 GACCGCACGCCGCATCCAGCGCTGACGGAAAGCAAACACCAGCAGCAGTTTTTCCAGTTCGGTTTATCCGGCAAACCATCGAAGTGACCAGCGAATACC
643▶ Asp Arg Thr P ro Hi s P ro Al aLeu Thr Gl uAl aLys Hi sGl nGl nPhe Phe Gl nPhe Arg Leu Ser Gl yGl nThr l l eGl uVal l Thr Ser Gl uTyr L
3701 TGTCCGTCATAGCGATAACGAGCTCTGCACTGGATGGTGGCGTGGTAAGCCGCTGGCAAGCGGTGAAGTCCCTCTGGATGTGCTCCACAAGG
676▶ euPhe Arg Hi sSer Asp Asn Gl uLeu Leu Hi sT rp Met Val l Al aLeu Asp Gl yLys P ro Leu Al aSer Gl yGl uVal l P ro Leu Asp Val l Al aP ro Gl nGl
3801 TAAACAGTTGATTGAACTGCCTGAACTACCGCAGCCGAGAGCGCCGGCAACTCTGGCTCAGATACGGTAGTGCAACCGAACGCGACCGCATGGTCA
709▶ yLys Gl nLeu l l eGl uLeu P ro Gl uLeu P ro Gl nP ro Gl uSer Al aGl yGl nLeu T rp Leu Thr Val l Arg Val l Val l Gl nP ro Asn Al aThr Al aT rp Ser
3901 GAAGCCGGGCACATCAGCGCTGGCAGCAGTGGCGTCTGGCGAAAACCTCAGTGTGACGCTCCCGCCCGTCCACGCCATCCCGCATCTGACCACA
743▶ Gl uAl aGl yHi s l l eSer Al aT rp Gl nGl nT rp Arg Leu Al aGl uAsn Leu Ser Val l Thr Leu P ro Al aAl aSer Hi sAl l eP ro Hi sLeu Thr S
4001 GCGAAATGGATTTTGCATCGAGCTGGTAATAAGCGTGGCAATTTAACCGCCAGTCAGGCTTTCTTTCACAGATGTGGATTGGCGATAAAAAACA
776▶ er Gl uMe tAsp Phe Cys l l eGl uLeu Gl yAsn Lys Arg T rp Gl nPhe Asn Arg Gl nSer Gl yPhe Leu Ser Gl nMe tT rp l l eGl yAsp Lys Lys Gl nLe
4101 GCTGACGCCGCTGCGCATCAGTTACCCGTCACCGCTGGATAACGACATTGGCGTAAGTGAAGCGACCCGATTGACCCTAACGCCTGGTCCGAACGC
809▶ uLeu Thr P ro Leu Arg Asp Gl nPhe Thr Arg Al aP ro Leu Asp Asn Asp l l eGl yVal l Ser Gl uAl aThr Arg l l eAsp P ro Asn Al aT rp Val l Gl uArg
4201 TGGAAAGCGCGGGCCATTACAGCCGAAGCAGCGTGTGTCAGTGCACGGCAGATACACTTGCTGATGCGGTGCTGATTACGACCGCTCACGCGTGGC
843▶ T rp Lys Al aAl aGl yHi sTyr Gl nAl aGl uAl aAl aLeu Leu Gl nCys Thr Al aAsp Thr Leu Al aAsp Al aVal l Leu l l eThr Thr Al aHi sAl aT rp G
4301 AGCATCAGGGGAAAACCTTATTTATCAGCCGAAACCTACCGGATTGATGGTAGTGGTCAAATGGCGATTACCGTTGATGTTGAAGTGGCGAGCGATAC
876▶ l nHi sGl nGl yLys Thr Leu Phe l l eSer Arg Lys Thr Tyr Arg l l eAsp Gl ySer Gl yGl nMe tAl a l l eThr Val l Asp Val l Gl uVal l Al aSer Asp Th
4401 ACCGCATCCGGCGCGGATTGGCTGAACTGCCAGCTGGCGCAGGTAGCAGAGCGGGTAAACTGGCTCGGATTAGGGCCGCAAGAAAACCTATCCCGACCGC
909▶ r P ro Hi s P ro Al aArg l l eGl yLeu Asn Cys Gl nLeu Al aGl nVal l Al aGl uArg Val l Asn T rp Leu Gl yLeu Gl yP ro Gl nGl uAsn Tyr P ro Asp Arg
4501 CTTACTGCCGCTGTTTTGACCGCTGGGATCTGCCATTGTCAGACATGTATACCCCGTACGCTCTCCGAGCGAAAACCGTCTGCGCTGCGGGACGCGCG
943▶ Leu Thr Al aAl aCys Phe Asp Arg T rp Asp Leu P ro Leu Ser Asp Me tTyr Thr P ro Tyr Al l Phe P ro Ser Gl uAsn Gl yLeu Arg Cys Gl yThr Arg G
4601 AATTGAATTATGGCCACACCGTGGCGGCGACTTCCAGTCAACCTCAGCCGCTACAGTCAAGTCAAGCAACTGATGGAALCAGCCATCGCCATCTGCT
976▶ l uLeu Asn Tyr Gl yP ro Hi sGl nT rp Arg Gl yAsp Phe Gl nPhe Asn l l eSer Arg Tyr Ser Gl nGl nLeu Me tGl uThr Ser Hi sArg Hi sLeu Le

NdeI (4742)

4701 GCACGCGGAAGAAGGCACATGGCTGAATATCGACGGTTTCCATATGGGGATTGGTGGCGCAGACTCCTGGAGCCCGTCACTATCGGGGAATTACAGCTG
1009▶ uHi sAl aGl uGl uGl yThr T rp Leu Asn l l eAsp Gl yPhe Hi sMe tGl y l l eGl yGl yAsp Asp Ser T rp Ser P ro Ser Val l Ser Al aGl uLeu Gl nLeu

NheI (4864)

EcoRI (4858)

4801 AGCGCCGGTCTGCTACCATTACAGTTGGTCTGGTGTCAAAAAATAATAATCTAGTCGAGAATTCGCTAGCTCGACATGATAAGATACATTGATGAGTTTGG
1043▶ Ser Al aGl yA rg Tyr Hi sTyr Gl nLeu Val l T rp Cys Gl nLys ●●●
4901 ACAAAACCAACTAGAATGCAGTGAAAAAATGCTTTATTTGGAAATTTGTGATGCTATTGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTG

MfeI (5038)

5001 TAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTATGTTTTAGGTTTCAGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTA

SwaI (5129)

5101 AAACCTCTACAATGTGGTAGATCCATTTAAATGTTAATTAAGTACGATGACCAAATCCCTTAACGTGAGTTTTCTGTTCCACTGAGCGTCAGACCCCG
5201 TAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGGCAACAAAAAACCCGCTACCAGCGGTGGTTTGTGTTGCC
5301 GGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAGTGGCTTCCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCAC
5401 TTCAAGAACTCTGTAGACCCGCTACATACCTCGCTCTGCTAATCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCTGTCTTACCGGGTTGGACT
5501 CAAGACGATAGTTACCGGATAAAGCGCAGCGGTGGGCTGAACGGGGGTTCTGTCACACAGCCAGCTTGGAGCGAACGACTACACCGAACTGAGATA
5601 CCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGGAAAGCGGCAGAGGTATCCGGTAAGCGGCAGGGTCCGAACAGGAGAGCGCACGAGG
5701 GAGCTTCCAGGGGAAACGCCTGGTATCTTTATAGTCTGCTGCGGTTTCCGACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAAGGGGGCGGA
5801 GCCTATGGA AAAACCGCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGTGCGCTTTTGTCTACATGTTCTTAATTAATTTTTCAAAGTAGTTG

AseI (5905)

MseI (5967)

5901 ACAATTAATCATCGGCATAGTATATCGGCATAGTATAATACGACTCACTATAGGAGGGCCATCATGGCCAAGTTGACCGAGTGTGCCAGTGCTCACAG
6001 CCAGGGATGTGGCTGGAGCTGTTGAGTTCTGGACTGACAGGTTGGGTTCTCCAGAGATTTGTGGAGGATGACTTTGCAGGTGTGGTCAGAGATGATGT
13▶ l aArg Asp Val l Al aGl yAl aVal l Gl uPhe T rp Thr Asp Arg Leu Gl yPhe Ser Arg Asp Phe Val l Gl uAsp Asp Phe Al aGl yVal l Val l Arg Asp Asp Va
6101 CACCTGTTCATCTCAGCAGTCCAGGACAGTGGTGCCTGACAACCCCTGGCTGGGTGGGTGAGAGGACTGGATGAGCTGATGCTGAGTGGAGT
46▶ l Thr Leu Phe l l eSer Al aVal l Gl nAsp Gl nVal l Val l P ro Asp Asn Thr Leu Al aT rp Val l T rp Val l Arg Gl yLeu Asp Gl uLeu Tyr Al aGl uT rp Ser
6201 GAGGTGGTCTCCACCAACTTCAGGATGCGAGTGGCCCTGCCATGACAGAGATTGGAGAGCAGCCCTGGGGGAGAGAGTTTGGCTGAGAGACCCAGCAG
80▶ Gl uVal l Val l Ser Thr Asn Phe Arg Asp Al aSer Gl yP ro Al aMe tThr Gl u l l eGl yGl uGl nP ro T rp Gl yA rg Gl uPhe Al aLeu Arg Asp P ro Al aG
6301 GCAACTGTGTCACTTTGTGGCAGAGGAGCAGGACTGAGGATAAGAATTGAGTTTCAGAAAAGGGGCGCTGAGTGGCCCTTTTTTCACTTAATTA
113▶ l yAsn Cys Val l Hi sPhe Val l Al aGl uGl nAsp ●●●