



1 GGACCTGCAGGGCCTGAAATAACCTCTGAAAGAGAACTTGGTTAGGTACCTTCTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGTGTCAGTTAGGGTGTG  
101 GAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGAAAGTCCCCAGGCTCCCCAGCAGGCAG  
201 AAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCACTAGTCTCCGGTGCCCGTCACTGGGCGAGAGCGCACATCGCCACAGTCCCCGA  
301 GAAGTTGGGGGAGGGGTGGCAATTGAACGGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTATGTCGTGTACTGGCTCCGCCTTTTCC  
401 GAGGGTGGGGGAGAACCCTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTCCGCCAGAACACAGCTGAAGCTTCGAGGGCTC  
501 GCATCTCTCTTACGCGCCCGCCGCCCTACCTGAGGCGCCATCCACGCCGTTGAGTCGCGTTCTGCCGCTCCCGCTGTGGTGCCTCTGAACTGC  
601 GTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACCGGGCCTTTGTCCGGCGCTCCCTTGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTG

701 CTGACCCTGCTTCTCAACTCTACGCTTTTGTTCGTTTTCTGTTCTGCGCAAGTTACAGATCCAAGCTGTGACCGCGCCTACCTGAGATCACCGGTCAA

801 CATGTTTGTGTTCTTGGTGTGCTTCCACTGGTCAGTCCCAATGCCTTATGCTTCTCACTACCACCTCAATCTATACAAATTCCTTC  
1 M F V F L V L L P L V S S Q C V M P L F N L I T T T Q S Y T N S F

901 ACCAGAGGAGTGTACTATCCTGACAAAGTGTTCGGTCAAGTGCCTCCACTCAGGACCTCTTTCTGCCTTTCTTTTCTAACGTTACATGGTTTC  
34 T R G V Y Y P D K V F R S S V L H L T Q D L F L P F F S N V T W F

1001 ATGCAATCTCTGGGACAAACGGCACAAACGCTTCGACAACCTGTAATTGCCATTCAATGATGGGTGACTTTGCCTCCACAGAGAAATCCAACATCAT  
67 H A I S G T N G T K R F D N P V L P F N D G V Y F A S T E K S N I I

1101 TCGAGGATGGATTTTCCGGACTACTCTGGACTCAAAGACACAGAGCCTGCTGATCGTTAACAACGCCACAAACGTTTTCATCAAAGTGTGCGAATCCAG  
100 R G W I F G T T L D S K T Q S L L I V N N A T N V F I K V C E F Q

1201 TTTTGAATGATCCCTTCTGGACGTGTACCACAAGAATAACAAGTCTGGATGGAGAGCGAATCTGGGTCTACAGCAGCGCAAACAAGTGCACCTTCG  
134 F C N D P F L D V Y H K N N K S W M E S E S G V Y S S A N N C T F

1301 AGTACGTGAGTCAACCTTTCTGATGGACCTGGAAGGGAAAACAGGAAACTCAAGAACCTGAGAGAGTTTGTCTTTAAGAACATCGACGGCTATTTAA  
167 E Y V S Q P F L M D L E G K Q G N F K N L R E F V F K N I D G Y F K

1401 GATCTATAGTAAGCATACGCCTATCATCGGAAGGGATTTCCCCAGGGCTTTTCCAGCCCTGGAACCTTTGGTTGACTTGCCTATTGGTATCAATATCACC  
200 I Y S K H T P I I G R D F P Q G F S A L E P L V D L P I G I N I T

1501 AGATTTAGACCTTTCTGGCATTGAAICGGTCTTATCTTACTCCAAGTATTCTCTCCGGGTGGACTGCCGGCGCCGCTGACTACTATGTCGGCTATC  
234 R F Q T L L A L N R S Y L T P G D S S S G W T A G A A D Y Y V G Y

1601 TGCAACCAAGAAGTCTCTGCTCAAGTACAACGAAACGGCACTATTACGGATGCTGTTGATTGTGCCCTGGACCCTCTGTCTGAGACTAAATGCACCT  
267 L Q P R T F L L K Y N E N G T I T D A V D C A L D P L S E T K C T L

1701 CAAGAGCTTTACCGTTGAGAAGGGGATTTACCAAACAGTAATTTCCGGGTCCAACCCAGAAAGCATTGTGCGGTTCCCAAATGTCACCAATCTGTGT  
300 K S F T V E K G I Y Q T S N F R V Q P T E S I V R F P N V T N L C

1801 CCCTTTTCAAGAAGTGTCAATGCTACAAGGTTTCTTCTGTGTACGCATGGAATAGGACACGCATCTCCAATTGTGTGCTGATTACTCCGTGCTGTACA  
334 P F H E V F N A T R F A S V Y A W N R T R I S N C V A D Y S V L Y

1901 ATTTTGGCCCAATCTTCGCTTTCAAGTGTATGGCGTTTACCTACCAAACCTAACGACCTGTGCTTCACTAATGTGTATGCCGACTCTTTTGTGATAAA  
367 N F A P F F A F K C Y G V S P T K L N D L C F T N V Y A D S F V I K

2001 GGGCAATGAAGTAGCCAGATTGCACAGGGCAGACCGGCACATTGCCGACTACAACCTACAAGCTTCCAGATGACTTTACCGGATGTGTTATTGCATGG  
400 G N E V S Q I A P G Q T G N I A D Y N Y K L P D D F T G C V I A W

Protein sequence: MFVFLVLLPLVSSQCVMP LFNLI TTTQSYTNSF TRGVYYPDK VFRSSV LHLTQDLFLPFFSNVTWF HAISGTNGTKRFDNPVLPFNDGVYFASTEKSNII RGWIFGTTLD SKTQSLLI VNNATNVFIK VCEFQ TTTTGAATGATCCCTTCTGGACGTGTACCACAAGAATAACAAGTCTGGATGGAGAGCGAATCTGGGTCTACAGCAGCGCAAACAAGTGCACCTTCG 134 F C N D P F L D V Y H K N N K S W M E S E S G V Y S S A N N C T F AGTACGTGAGTCAACCTTTCTGATGGACCTGGAAGGGAAAACAGGAAACTCAAGAACCTGAGAGAGTTTGTCTTTAAGAACATCGACGGCTATTTAA 167 E Y V S Q P F L M D L E G K Q G N F K N L R E F V F K N I D G Y F K GATCTATAGTAAGCATACGCCTATCATCGGAAGGGATTTCCCCAGGGCTTTTCCAGCCCTGGAACCTTTGGTTGACTTGCCTATTGGTATCAATATCACC 200 I Y S K H T P I I G R D F P Q G F S A L E P L V D L P I G I N I T AGATTTAGACCTTTCTGGCATTGAAICGGTCTTATCTTACTCCAAGTATTCTCTCCGGGTGGACTGCCGGCGCCGCTGACTACTATGTCGGCTATC 234 R F Q T L L A L N R S Y L T P G D S S S G W T A G A A D Y Y V G Y TGCAACCAAGAAGTCTCTGCTCAAGTACAACGAAACGGCACTATTACGGATGCTGTTGATTGTGCCCTGGACCCTCTGTCTGAGACTAAATGCACCT 267 L Q P R T F L L K Y N E N G T I T D A V D C A L D P L S E T K C T L CAAGAGCTTTACCGTTGAGAAGGGGATTTACCAAACAGTAATTTCCGGGTCCAACCCAGAAAGCATTGTGCGGTTCCCAAATGTCACCAATCTGTGT 300 K S F T V E K G I Y Q T S N F R V Q P T E S I V R F P N V T N L C CCCTTTTCAAGAAGTGTCAATGCTACAAGGTTTCTTCTGTGTACGCATGGAATAGGACACGCATCTCCAATTGTGTGCTGATTACTCCGTGCTGTACA 334 P F H E V F N A T R F A S V Y A W N R T R I S N C V A D Y S V L Y ATTTTGGCCCAATCTTCGCTTTCAAGTGTATGGCGTTTACCTACCAAACCTAACGACCTGTGCTTCACTAATGTGTATGCCGACTCTTTTGTGATAAA 367 N F A P F F A F K C Y G V S P T K L N D L C F T N V Y A D S F V I K GGGCAATGAAGTAGCCAGATTGCACAGGGCAGACCGGCACATTGCCGACTACAACCTACAAGCTTCCAGATGACTTTACCGGATGTGTTATTGCATGG 400 G N E V S Q I A P G Q T G N I A D Y N Y K L P D D F T G C V I A W

G446S (2127) L455S (2154)  
 N440K (2109) V445H (2124) N450D+L452W (2139) N460K (2169)  
 2101 AACTCAAACAGCTGGATTCCAAGCACAGTGGCAACTATGACTACTGGTATAGAAGCTTCAGGAAATCCAGCTGAAACCATTTCGAGCGAGATATAAGCA  
 434▶ N S N K L D S K H S G N Y D Y W Y R S F R K S K L K P F E R D I S

---

F486P (2244) N501Y (2289)  
 N481K + Δ V483 (2232) S477N + T478K (2220) E484K (2238) Q498R (2280)  
 2201 CAGAAATCTACCAGGCTGGAACAAAACCTGCAGGGCAAAGGGCTAACTGCTACTTCCATTGCAGAGTTACGGATTCCAGCCTACATACGGGGTGGG  
 467▶ T E I Y Q A G N K P C K G K G P N C Y F P L Q S Y G F R P T Y G V G

---

Y505H (2301)  
 2301 TCAACCAACCCTATCGTGTCTGAGTCTCTGAGTTTTGAGCTCCTCCATGCCCCAGCCACAGTCTGTGGCCCCAAGAAAAGCACCAATCTGGTGAAGAACAAA  
 500▶ H Q P Y R V V V L S F E L L H A P A T V C G P K K S T N L V K N K

---

E554K (2448) A570V (2496)  
 2401 TCGTGAACTTTAACTTTAACGGACTCACAGGAACCGCGTATTGACAGAGTAACAAGAAGTTCTGCCATTCCAGCAGTTCCGGTCGCGATATTGTCG  
 534▶ C V N F N F N G L T G T G V L T K S N K K F L P F Q Q F G R D I V

---

D614G (2628) P621S (2649)  
 2601 CAGCAATCAGGTTGCCGTCTGTATCAGGGCGTAATTGCACCGAGGTAICTGTGCGCATCCACGCTGACCAACTTACACCCACATGGCGAGTATATTCC  
 600▶ S N Q V A V L Y Q G V N C T E V S V A I H A D Q L T P T W R V Y S

---

H655Y (2751) P681R (2829)  
 2701 ACCGGCTCAACGTCTTTCAGACACGTGTGGATGTCTGATCGGTGCAGAAATGTTAATAATAGCTACGAGTGTGATATCCCATCGGTGCTGGAATAT  
 634▶ T G S N V F Q T R A G C L I G A E Y V N N S Y E C D I P I G A G I

---

N679K (2823) Furin cleavage site  
 2801 GCGCCTTTATCAAACCTCAAACCAAATCTCGCAGGGGGCACTAGTGTAGCATCCCAAAGTATCATTGCCTACACAATGAGCCTCGGTGCTGAGAATTC  
 667▶ C A S Y Q T Q T K S R R R A R S V A S Q S I I A Y T M S L G A E N S  
 2901 TGTCGCCTACAGCAACAACCTCATTGCTATCCCTACTAACTTCAATCAGTGTGACAACCTGAAATTCGCCGATCTATGACCAAAAACAGCGTTGAC  
 700▶ V A Y S N N S I A I P T N F T I S V T T E I L P V S M T K T S V D

---

N764K (3078) D796Y (3174)  
 3001 TGCACCATGTACATCTGTGGCGATTCTACCGAATGTAGCAATCTCCTCCTGCAATACGGATCATTCTGCACTCAGCTGAGCGTGCCTCACAGGTATTG  
 734▶ C T M Y I C G D S T E C S N L L L Q Y G S F C T Q L K R A L T G I

---

S939F (3603) Q954H (3648) N969K (3693)  
 3101 CAGTTGAGCAGGACAAGAATACGCAGGAAGTGTGGCCAGGTGAAGCAAATCTACAAAACCTCCACCATAAAAATCTTTGGCGGATTCAATTTCTCACA  
 767▶ A V E Q D K N T Q E V F A Q V K Q I Y K T P P I K Y F G G F N F S Q  
 3201 GATCCTGCCCGATCCCTCAAACCCCTCAAGCGTAGCTTTATCGAGGATCTGCTCTTCAACAAGGTAACCCCTCGCAGATGCCGGTTTCATCAAGCAGTAT  
 800▶ I L P D P S K P S K R S F I E D L L F N K V T L A D A G F I K Q Y  
 3301 GCGGATTGTCTGGGAGACATCGCCGCTCGGGACCTGATCTGTGCACAGAAGTTCAATGGACTGACCGTGTGCTCCCTTGTGACCGACGAGATGATAG  
 834▶ G D C L G D I A A R D L I C A Q K F N G L T V L P P L L T D E M I  
 3401 CCAATACACTAGCGCCTGTGGCCGGCACCATCACTTCTGGGTGGACATTCGGAGCTGGCGTGCCTTCAGATTCTTTTGTATGCAGATGGCCTA  
 867▶ A Q Y T S A L L A G T I T S G W T F G A G A A L Q I P F A M Q M A Y  
 3501 CCGCTTAAACGGCATCGGTGTGACACAAAACGTTCTGTATGAAAACAGAACTCATCGCAACCAAGTTCAACAGTGTCTATCGGTAAGATAACAGGATAGC  
 900▶ R F N G I G V T Q N V L Y E N Q K L I A N Q F N S A I G K I Q D S

---

P1143L (4215)  
 4201 TACGATCCTTTGCGAGCTGAGCTCCTCAAGGAGGAGCTTGACAAATATTTAAGAATCACACATCACCTGACGTCGACCTCGGAGATATTTTCAG  
 1134▶ Y D P L Q L E L D S F K E E L D K Y F K N H T S P D V D L G D I S  
 4301 GAATCAATGCTCCGTTGGTCAATATTCAGAAGGAGATAGACAGGCTGAATGAGGTTGCCAAGAACCCTCAACGAGTCTCTGATCGATCTCGAGGAGTTGGG  
 1167▶ G I N A S V V N I Q K E I D R L N E V A K N L N E S L I D L Q E L G  
 4401 CAAGTACGAACAGTATATCAATGGCCITGGTACATTTGGCTTGGGTTTCATTGCTGGGCTGATAGCTATCGTCATGGTGACAATTATGTTGTGTTGCATG  
 1200▶ K Y E Q Y I K W P W Y I W L G F I A G L I A I V M V T I M L C C M

---

NheI (4555)  
 4501 ACATCCTGCTGAGTTGTCTGAAGGGCTGCTGCTCATGCGGCAGCTGTTGCTAAAGCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAA  
 1234▶ T S C C S C L K G C C S C G S C C •

4601 ACCACAAC TAGAATGCAGTGA AAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACA  
4701 ACAACAATTGCATTCATTTTATGTTTCAGGTT CAGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATGGAATTCTAAAA  
4801 TACAGCATAGCAAAACTTTAACCTCCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGC  
4901 ATTAGCTGTTTGCAGCCTCACCTTCTTTTCATGGAGTTAAGATATAGTGTATTTTCCAAGGTTTGAAGTACTGCTTTCATTTCTTTATGTTTTAAATGCA  
5001 CTGACCTCCCACATTCCCTTTTTAGTAAAATATTCAGAAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAGGCAGAATCCAGATGC  
5101 TCAAGGCCCTTCATAATATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGAAATGGACAGCAAGAAAGCGAGCTTCTAGCTTT  
5201 AGTTCCTGGTGTACTTGAGGGGATGAGTTTCTCAATGGTGGTTTTGACCAGCTTGCATTTCATCTCAATGAGCACAAAGCAGTCAGGAGCATAGTCAGA  
140 N R T Y K L P I L E E I T T K V L K G N M E I L V F C D P A Y D S  
5301 GATGAGCTCTCTGCACATGCCACAGGGGCTGACCACCTGATGGATCTGTCCACCTCATCAGAGTAGGGGTGCCTGACAGCCACAATGGTGTCAAAGTCC  
107 I L E R C M G C P S V V R I S R D V E D S Y P H R V A V I T D F D  
5401 TTCTGCCCGTTGCTCACAGCAGACCCAATGGCAATGGCTT CAGCACAGACAGTGACCCTGCCAATGTAGGCCTCAATGTGGACAGCAGAGATGATCTCCC  
73 K Q G N S V A S G I A I A E A C V T V R G I Y A E I H V A S I I E G  
5501 CAGTCTTGGTCTGATGGCCGCCCCGACATGGTGTCTGTTGCTCATAGAGCATGGTGTCTTCTCAGTGGCGACCTCCACCAGCTCCAGATCTGCTG  
40 T K T R I A A G V H H K N D E Y L M T I K E T A V E V L E L D Q Q  
5601 AGAGATGTTGAAGGTCTTCATGATGGCCCTCTATAGTGAGTCGTATTATACTATGCCGATATACTATGCCGATGATTAATTGTCAAACAGCGTGGATG  
7 S I N F T K M  
5701 GCGTCTCCAGCTTATCTGACGGTTCACATAACGAGCTCTGCTTATATAGACCTCCCACCGTACACGCCTACCGCCATTTGCGTCAATGGGGCGGAGTTG  
5801 TTACGACATTTTGAAAGTCCCGTTGATTTACTAGTCAAACAAACTCCATTGACGTCAATGGGTGGAGACTTGAAATCCCGTGAGTCAAACCGT  
5901 ATCCACGCCATTGATGTA CTGCAAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTA CTGCAAGTAGGAAAGTCCCAT AAGGTGAT  
6001 GTA CTGGGCATAATGCCAGGCGGGCCATTTACCGTCA TTGACGTCAATAGGGGGCTACTTGGCATATGATACACTTGATGTA CTGCAAGTGGGCAGTT  
6101 TACCGTAAATACTCCACCCATTGACGTCAATGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCAATTATTGACGTCAATGGGCGGGGTCGTTGG  
6201 GCGGTACGCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCCTGCAGGTTAATTAAAGAACATGTGAGCAAAGGCCAGCAAAGGCCAGGAACCGTAAA  
BspLU111 (6259)  
6301 AAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTAT  
6401 AAAGATACCAGGCGTTTCCCGTGAAGCTCCCTCGTGCCTCTCCTGTTCCGACCCTGCCGTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAG  
6501 CGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCCGCTCCAAGCTGGGCTGTGTGCACGAACCCCGTT CAGCCGAC  
6601 CGCTGCGCTTATCCGGTAACTATCGTCTT GAGTCCAACCCGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGA  
6701 GGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTAC  
6801 CTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAA  
6901 GGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGGAAACGAAACTACGTTAAGGGATTTTGGTATGGCTAGTTAATTAACAT  
7001 TTAATCAGCGCCGCAATAAAATATCTTTATTTTATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAACAAA  
7101 ACGAAACAAACAAACTAGCAAATAGGCTGTCCCAGTGCAAGTGCAAGTGCCAGAGTCCAGAACATTTCTCTATCGAA