Appendix 1a

pSTA1 translation

```
Translation of DNAMAN1(938-3670)
Universal code
Total amino acid number: 910, MW=102412
Max ORF starts at AA pos 1(may be DNA pos 938) for 910 AA(2730 bases),
MW=102412
1          GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATG
61         CCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCG
121        CGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGC
181        TTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATT
241        GATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATA
301        TGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACC
361        CCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC
421        ATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGT
481        ATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATT
541        ATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCA
601        TCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTG
661        ACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACC
721        AAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCG
781        GTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCA
                                   T7 promoter
841        CTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGC
                     HinD3              BamH1     GFP
901        GTTTAAACTTAAGCTTGGTACCGAGCTCGGATCCACCATGGGGCACCACCACCACCACCA
1                                                M  G  H  H  H  H  H  H

961        CGGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGA
9            G  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D  G  D

1021       TGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAA
29           V  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y  G  K

1081       ACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGT
49           L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T  L  V

1141       CACTACTTTCTCTTATGGTGTTCAATGCTTTTCAAGATACCCAGATCATATGAAACGGCA
69           T  T  F  S  Y  G  V  Q  C  F  S  R  Y  P  D  H  M  K  R  H

1201       TGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAA
89           D  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F  F  K

1261       AGATGACGGGAACTACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAA
109          D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L  V  N

1321       TAGAATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAATT
129          R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H  K  L

1381       GGAATACAACTATAACTCACACAATGTATACATCATGGCAGACAAACAAAAGAATGGAAT
149          E  Y  N  Y  N  S  H  N  V  Y  I  M  A  D  K  Q  K  N  G  I

1441       CAAAGTTAACTTCAAAATTAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCA
169          K  V  N  F  K  I  R  H  N  I  E  D  G  S  V  Q  L  A  D  H

1501       TTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCT
189          Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H  Y  L
```

```
1561    GTCCACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCT
209      S  T  Q  S  A  L  S  K  D  P  N  E  K  R  D  H  M  V  L  L

1621    TGAGTTTGTAACAGCTGCTGGGATTACACATGGCATGGACGAACTATACAAGTCCGGGTC
229      E  F  V  T  A  A  G  I  T  H  G  M  D  E  L  Y  K  S  G  S

        Xba1 TaV 2A                                            Apa1
1681    TAGAGCGGCCGAGGGCAGGGGAAGTCTTCTAACATGCGGGGACGTGGAGGAAAATCCCGG
249      R  A  A  E  G  R  G  S  L  L  T  C  G  D  V  E  E  N  P  G

                GUS
1741    GCCCCACCACCACCACCACCACTTACGTCCTGTAGAAACCCCAACCCGTGAAATCAAAAA
269      P  H  H  H  H  H  H  L  R  P  V  E  T  P  T  R  E  I  K  K

1801    ACTCGACGGCCTGTGGGCATTCAGTCTGGATCGCGAAAACTGTGGAATTGATCAGCGTTG
289      L  D  G  L  W  A  F  S  L  D  R  E  N  C  G  I  D  Q  R  W

1861    GTGGGAAAGCGTGTTACAAGAAAGCCGGGCAATTGCTGTGCCAGGCAGTTTTAACGATCA
309      W  E  S  V  L  Q  E  S  R  A  I  A  V  P  G  S  F  N  D  Q

1921    GTTCGCCGATGCAGATATTCGTAATTATGCGGGCAACGTCTGGTATCAGCGCGAAGTCTT
329      F  A  D  A  D  I  R  N  Y  A  G  N  V  W  Y  Q  R  E  V  F

1981    TATACCGAAAGGTTGGGCAGGCCAGCGTGTCGTGCTGCGTTTCGATGCGGTCACTCATTA
349      I  P  K  G  W  A  G  Q  R  V  V  L  R  F  D  A  V  T  H  Y

2041    CGGCAAAGTGTGGGTCAATAATCAGGAAGTGATGGACCATCAGGGCGGCTATACGCCATT
369      G  K  V  W  V  N  N  Q  E  V  M  D  H  Q  G  G  Y  T  P  F

2101    TGAAGCCGATGTCACGCCGTATGTTATTGCCGGGAAAAGTGTACGTATCACCGTTTGTGT
389      E  A  D  V  T  P  Y  V  I  A  G  K  S  V  R  I  T  V  C  V

2161    GAACAACGAACTGAACTGGCAGACTATCCCGCCGGGAATTGTGATTACCGACGAAAACGG
409      N  N  E  L  N  W  Q  T  I  P  P  G  I  V  I  T  D  E  N  G

2221    CAAGAAAAAGCAGTCTTACTTCCATGATTTCTTTGACTATGCCGGAATCCATCGCAGCGT
429      K  K  K  Q  S  Y  F  H  D  F  F  D  Y  A  G  I  H  R  S  V

2281    AATGCTCTACACCACGCCGAACACCTGGGTGGACGATATCACCGTGGTGACGCATGTCGC
449      M  L  Y  T  T  P  N  T  W  V  D  D  I  T  V  V  T  H  V  A

2341    GCAAGACTGTAACCACGCGTCTGTTGACTGGCAGGTGGTGGCCAATGGTGATGTCAGCGT
469      Q  D  C  N  H  A  S  V  D  W  Q  V  V  A  N  G  D  V  S  V

2401    TGAACTGCGTGATGCGGATCAACAGGTGGTTGCAACTGGACAAGGCACTAGCGGGACTTT
489      E  L  R  D  A  D  Q  Q  V  V  A  T  G  Q  G  T  S  G  T  L

2461    GCAAGTGGTGAATCCGCACCTCTGGCAACCGGGTGAAGGTTATCTCTATGAACTGTGCGT
509      Q  V  V  N  P  H  L  W  Q  P  G  E  G  Y  L  Y  E  L  C  V

2521    CACAGCCAAAAGCCAGACAGAGTGTGATATCTACCCGCTTCGCGTCGGCATCCGGTCAGT
529      T  A  K  S  Q  T  E  C  D  I  Y  P  L  R  V  G  I  R  S  V

2581    GGCAGTGAAGGGCCAACAGTTCCTGATTAACCACAAACCGTTCTACTTTACTGGCTTTGG
549      A  V  K  G  Q  Q  F  L  I  N  H  K  P  F  Y  F  T  G  F  G

2641    TCGTCATGAAGATGCGGACTTACGTGGCAAAGGATTCGATAACGTGCTGATGGTGCACGA
569      R  H  E  D  A  D  L  R  G  K  G  F  D  N  V  L  M  V  H  D

2701    CCACGCATTAATGGACTGGATTGGGGCCAACTCCTACCGTACCTCGCATTACCCTTACGC
```

```
589          H  A  L  M  D  W  I  G  A  N  S  Y  R  T  S  H  Y  P  Y  A

2761       TGAAGAGATGCTCGACTGGGCAGATGAACATGGCATCGTGGTGATTGATGAAACTGCTGC
609          E  E  M  L  D  W  A  D  E  H  G  I  V  V  I  D  E  T  A  A

2821       TGTCGGCTTTAACCTCTCTTTAGGCATTGGTTTCGAAGCGGGCAACAAGCCGAAAGAACT
629          V  G  F  N  L  S  L  G  I  G  F  E  A  G  N  K  P  K  E  L

2881       GTACAGCGAAGAGGCAGTCAACGGGGAAACTCAGCAAGCGCACTTACAGGCGATTAAAGA
649          Y  S  E  E  A  V  N  G  E  T  Q  Q  A  H  L  Q  A  I  K  E

2941       GCTGATAGCGCGTGACAAAAACCACCCAAGCGTGGTGATGTGGAGTATTGCCAACGAACC
669          L  I  A  R  D  K  N  H  P  S  V  V  M  W  S  I  A  N  E  P

3001       GGATACCCGTCCGCAAGTGCACGGGAATATTTCGCCACTGGCGGAAGCAACGCGTAAACT
689          D  T  R  P  Q  V  H  G  N  I  S  P  L  A  E  A  T  R  K  L

3061       CGACCCGACGCGTCCGATCACCTGCGTCAATGTAATGTTCTGCGACGCTCACACCGATAC
709          D  P  T  R  P  I  T  C  V  N  V  M  F  C  D  A  H  T  D  T

3121       CATCAGCGATCTCTTTGATGTGCTGTGCCTGAACCGTTATTACGGATGGTATGTCCAAAG
729          I  S  D  L  F  D  V  L  C  L  N  R  Y  Y  G  W  Y  V  Q  S

3181       CGGCGATTTGGAAACGGCAGAGAAGGTACTGGAAAAAGAACTTCTGGCCTGGCAGGAGAA
749          G  D  L  E  T  A  E  K  V  L  E  K  E  L  L  A  W  Q  E  K

3241       ACTGCATCAGCCGATTATCATCACCGAATACGGCGTGGATACGTTAGCCGGGCTGCACTC
769          L  H  Q  P  I  I  I  T  E  Y  G  V  D  T  L  A  G  L  H  S

3301       AATGTACACCGACATGTGGAGTGAAGAGTATCAGTGTGCATGGCTGGATATGTATCACCG
789          M  Y  T  D  M  W  S  E  E  Y  Q  C  A  W  L  D  M  Y  H  R

3361       CGTCTTTGATCGCGTCAGCGCCGTCGTCGGTGAACAGGTATGGAATTTCGCCGATTTTGC
809          V  F  D  R  V  S  A  V  V  G  E  Q  V  W  N  F  A  D  F  A

3421       GACCTCGCAAGGCATATTGCGCGTTGGCGGTAACAAGAAAGGGATCTTCACTCGCGACCG
829          T  S  Q  G  I  L  R  V  G  G  N  K  K  G  I  F  T  R  D  R

3481       CAAACCGAAGTCGGCGGCTTTTCTGCTGCAAAAACGCTGGACTGGCATGAACTTCGGTGA
849          K  P  K  S  A  A  F  L  L  Q  K  R  W  T  G  M  N  F  G  E
                                                  Xho1
3541       AAAACCGCAGCAGGGAGGCAAACAATCTAGTGCGGCCGCTCGAGGCCCGTTTAAACCCGC
869          K  P  Q  Q  G  G  K  Q  S  S  A  A  A  R  G  P  F  K  P  A

3601       TGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTG
889          D  Q  P  R  L  C  L  L  V  A  S  H  L  L  F  A  P  P  P  C

3661       CCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATT
909          L  P  *

3721       GCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGC
3781       AAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCT
3841       TCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGC
3901       GCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCC
3961       CTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCC
4021       CGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTC
4081       GACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACG
4141       GTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACT
4201       GGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATT
4261       TCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGT
4321       GGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGC
```

4381 AAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAG
4441 GCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTC
4501 CGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAA
4561 TTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGT
4621 GAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCA
4681 TTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGAT
4741 TGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAAC
4801 AGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTC
4861 TTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGC
4921 TATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAG
4981 CGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACC
5041 TTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTG
5101 ATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTC
5161 GGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGC
5221 CAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGA
5281 CCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCA
5341 TCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTG
5401 ATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCG
5461 CCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGG
5521 GACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGA
5581 TTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTG
5641 GATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTAT
5701 TGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATT
5761 TTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTG
5821 TATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTG
5881 AAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGC
5941 CTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTT
6001 CCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGG
6061 CGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGT
6121 TCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATC
6181 AGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAA
6241 AAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAA
6301 TCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCC
6361 CCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTC
6421 CGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAG
6481 TTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGA
6541 CCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATC
6601 GCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTAC
6661 AGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTG
6721 CGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACA
6781 AACCACCGCTGGTAGCGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGG
6841 ATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTC
6901 ACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAA
6961 TTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTA
7021 CCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGT
7081 TGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAG
7141 TGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCA
7201 GCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTC
7261 TATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGT
7321 TGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAG
7381 CTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGT
7441 TAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCAT
7501 GGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGT
7561 GACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTC
7621 TTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCAT
7681 CATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAG
7741 TTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGT
7801 TTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACG
7861 GAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTA
7921 TTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCC
7981 GCGCACATTTCCCCGAAAAGTGCCACCTGACGTC