Biochemical and structural bioinformatics studies of fungal CutA nucleotidyltransferases explain their unusual specificity towards CTP and increased tendency for cytidine incorporation at the 3'-terminal positions of synthesized tails

Kamil Kobyłecki, Krzysztof Kuchta, Andrzej Dziembowski, Krzysztof Ginalski & Rafał Tomecki

SUPPLEMENTAL MATERIAL

SUPPLEMENTAL FIGURE

```
AnCutA          ATGGTGACCGTGACCGATGGCAACGATCGCCTGAGCAGCCGCCCGGGCGAACAGAGCCCG
AnCutA_opt      ATGGTCACCGTCACCGATGGTAACGACCGCCTGTCCAGCCGCCCGGGCGAACAAAGTCCG
                ***** ***** ******** ***** ******  ****************** ** ***
protein          M  V  T  V  T  D  G  N  D  R  L  S  S  R  P  G  E  Q  S  P

AnCutA          ATTGCGAGCGCGAAACTGAGCACCATGCCGTGGCGCAGCAGCAACGAAGAAAGCAGCGAA
AnCutA_opt      ATTGCCTCAGCCAAACTGTCCACGATGCCGTGGCGTAGCTCTAACGAAGAAAGTTCCGAA
                *****    ** ******  *** *********** ***   ***********   ****
protein          I  A  S  A  K  L  S  T  M  P  W  R  S  S  N  E  E  S  S  E

AnCutA          ACCCCGAACCGCCCGGCGCTGACCAGCCAGCAGAGCAACAGCCTGCCGAGCACCCCGTAT
AnCutA_opt      ACCCCGAATCGCCCGGCACTGACGTCACAGCAATCGAACAGCCTGCCGTCGACCCCGTAT
                ******** ******** *****    *****    ************   *********
protein          T  P  N  R  P  A  L  T  S  Q  Q  S  N  S  L  P  S  T  P  Y

AnCutA          CAGCATGCGCGCAAAATTAGCTTTCATAGCCGCAGCCCGAGCCCGCGCCATGGCAGCACC
AnCutA_opt      CAGCATGCTCGTAAAATTAGCTTTCACTCTCGTAGTCCGTCCCCGCGTCATGGTTCAACG
                ******** ** **************    ** ** ***  ****** *****    ** 
protein          Q  H  A  R  K  I  S  F  H  S  R  S  P  S  P  R  H  G  S  T

AnCutA          AGCCCGCGCAGCACCCATAGCGAAAGCACCCATATTCCGCCGAGCGCGCGCAAACAGTTT
AnCutA_opt      TCGCCGCGTAGCACCCATAGCGAATCTACGCACATCCCGCCGTCTGCACGCAAACAGTTC
                   ***** ***************   ** ** ** ******   ** *********** 
protein          S  P  R  S  T  H  S  E  S  T  H  I  P  P  S  A  R  K  Q  F

AnCutA          ACCGGCTGCAAATATGAAACCGCGATGAGCTTTTTTCGCCGCCGCATGCCGTATAGCATT
AnCutA_opt      ACCGGCTGCAAGTATGAAACGGCTATGTCATTTTTCCGTCGCCGTATGCCGTACTCGATT
                *********** ******** ** ***   ***** ** ***** ********    ***
protein          T  G  C  K  Y  E  T  A  M  S  F  F  R  R  R  M  P  Y  S  I

AnCutA          GGCGCGGATGTGCTGCCGGAAGAAAAAGAAGGCCTGAAACAGCGCCTGCGCCCGGAAGAA
AnCutA_opt      GGTGCGGATGTGCTGCCGGAAGAAAAAGAAGGCCTGAAGCAGCGTCTGCGCCCGGAAGAA
                ** *********************************** ***** ***************
protein          G  A  D  V  L  P  E  E  K  E  G  L  K  Q  R  L  R  P  E  E

AnCutA          GAACGCAAACTGACCGCGGATATGCTGGAAGTGTATGAACGCCTGCTGCCGAGCGCGGAA
AnCutA_opt      GAACGTAAACTGACCGCGGATATGCTGGAAGTTTATGAACGCCTGCTGCCGAGCGCCGAA
                ***** ************************** *********************** ***
protein          E  R  K  L  T  A  D  M  L  E  V  Y  E  R  L  L  P  S  A  E

AnCutA          AGCGAAAGCCGCCGCCGCCGCCTGGTGCGCAAACTGGAAGATCTGTTTAACCGCCAGTGG
AnCutA_opt      AGTGAATCCCGCCGTCGCCGTCTGGTCCGTAAACTGGAAGACCTGTTTAACCGCCAGTGG
                ** ***  ****** ***** ***** ** *********** ******************
protein          S  E  S  R  R  R  R  L  V  R  K  L  E  D  L  F  N  R  Q  W

AnCutA          CCGGGCTGCGATATTAAAGTGCATGTGTTTGGCAGCAGCGGCAACAAACTGTGCAGCAGC
AnCutA_opt      CCGGGTTGCGACATTAAAGTGCATGTTTTCGGTTCATCGGGCAATAAGCTGTGTAGCTCT
                ***** ***** ************** ** **      ***** ** ***** ***   
protein          P  G  C  D  I  K  V  H  V  F  G  S  S  G  N  K  L  C  S  S

AnCutA          GATAGCGATGTGGATATTTGCATTACCACCACCTGCAAAGAACTGGAACATGTGTGCCTG
AnCutA_opt      GACAGTGATGTGGACATTTGCATCACCACGACCTGTAAAGAACTGGAACATGTTTGCCTG
                ** ** ******** ******** ***** ***** ***************** ******
protein          D  S  D  V  D  I  C  I  T  T  T  C  K  E  L  E  H  V  C  L

AnCutA          CTGGCGGATGTGCTGGCGAAAAACGGCATGGAACGCGTGGTGTGCATTAGCCATGCGAAA
AnCutA_opt      CTGGCGGATGTCCTGGCCAAGAACGGCATGGAACGTGTGGTTTGTATCTCTCACGCGAAA
                *********** ***** ** ************** ***** ** **    ** ******
protein          L  A  D  V  L  A  K  N  G  M  E  R  V  V  C  I  S  H  A  K

AnCutA          GTGCCGATTGTGAAAATTTGGGATCCGGAACTGCGCCTGGCGTGCGATATGAACGTGAAC
AnCutA_opt      GTGCCGATTGTTAAGATCTGGGATCCGGAACTGCGCCTGGCGTGTGACATGAACGTGAAC
                *********** ** ** ************************** ** ************
protein          V  P  I  V  K  I  W  D  P  E  L  R  L  A  C  D  M  N  V  N
```

```
AnCutA          AACACCATGGCGCTGGAAAACACCCGCATGGTGCGCACCTATGTGGAAATTGATGAACGC
AnCutA_opt      AATACCATGGCCCTGGAAAATACCCGTATGGTGCGCACGTATGTTGAAATTGATGAACGT
                ** ******** ******** ***** *********** ***** **************
protein          N  T  M  A  L  E  N  T  R  M  V  R  T  Y  V  E  I  D  E  R

AnCutA          GTGCGCCCGCTGGCGATGATTATTAAACATTGGACCAAACGCCGCATTCTGAACGATGCG
AnCutA_opt      GTTCGCCCGCTGGCGATGATTATCAAACATTGGACCAAGCGCCGTATCCTGAACGACGCC
                ** ******************** ************** ***** ** ******** **
protein          V  R  P  L  A  M  I  I  K  H  W  T  K  R  R  I  L  N  D  A

AnCutA          GGCCTGGGCGGCACCCTGAGCAGCTATACCTGGATTTGCCTGATTATTAACTTTCTGCAG
AnCutA_opt      GGTCTGGGCGGTACCCTGAGTTCCTACACGTGGATTTGCCTGATTATCAATTTTCTGCAG
                ** ******** ********   *** ** ***************** ** *********
protein          G  L  G  G  T  L  S  S  Y  T  W  I  C  L  I  I  N  F  L  Q

AnCutA          ACCCGCGAACCGCCGATTCTGCCGAGCCTGCAGGCGCGCCCGCATAAAAAACGCCTGACC
AnCutA_opt      ACCCGTGAACCGCCGATCCTGCCGAGCCTGCAAGCACGTCCGCACAAAAAGCGCCTGACG
                ***** *********** ************** ** ** ***** ***** ********
protein          T  R  E  P  P  I  L  P  S  L  Q  A  R  P  H  K  K  R  L  T

AnCutA          GCGGATGGCCTGGTGTGCAGCTTTGATGATGATCTGGATAGCCTGGTGGGCTATGGCAAA
AnCutA_opt      GCCGATGGCCTGGTCTGTAGTTTCGATGACGATCTGGACTCCCTGGTGGGTTATGGCAAA
                ** *********** ** ** ** ***** ********   ********* *********
protein          A  D  G  L  V  C  S  F  D  D  D  L  D  S  L  V  G  Y  G  K

AnCutA          CAGAACAAACAGAGCCTGGGCGAACTGTTTTTTCAGTTTTTTCGCTATTATGGCCATGAA
AnCutA_opt      CAGAATAAGCAATCTCTGGGTGAACTGTTTTTCCAGTTTTTCCGCTATTACGGCCATGAA
                ***** ** **    ***** *********** ******** ******** *********
protein          Q  N  K  Q  S  L  G  E  L  F  F  Q  F  F  R  Y  Y  G  H  E

AnCutA          CTGGATTTTGAAAAATATGTGATTAGCGTGCGCGAAGGCCGCCTGATTAGCAAAGAAGGC
AnCutA_opt      CTGGATTTCGAAAAGTACGTCATTAGCGTGCGTGAAGGTCGCCTGATCTCTAAAGAAGGT
                ******** ***** ** ** *********** ***** ********    ********
protein          L  D  F  E  K  Y  V  I  S  V  R  E  G  R  L  I  S  K  E  G

AnCutA          AAAGGCTGGCATCTGCTGCAGAACAACCGCCTGTGCGTGGAAGAACCGTTTAACACCATT
AnCutA_opt      AAAGGTTGGCACCTGCTGCAGAACAATCGTCTGTGCGTGGAAGAACCGTTTAACACCATT
                ***** ***** ************** ** ******************************
protein          K  G  W  H  L  L  Q  N  N  R  L  C  V  E  E  P  F  N  T  I

AnCutA          CGCAACCTGGGCAACACCGCGGATGATACCAGCTTTCGCGGCGTGCATATGGAACTGCGC
AnCutA_opt      CGCAACCTGGGTAATACGGCGGACGATACCAGTTTCCGTGGCGTTCATATGGAACTGCGC
                *********** ** ** ***** ******** ** ** ***** ***************
protein          R  N  L  G  N  T  A  D  D  T  S  F  R  G  V  H  M  E  L  R

AnCutA          CGCGCGTTTAAAGCGCTGAGCGATGGCAACTTTGAACTGTGCTGCGAACAGTATGAATAT
AnCutA_opt      CGTGCATTTAAAGCTCTGTCCGATGGCAATTTCGAACTGTGCTGTGAACAGTATGAATAC
                ** ** ******** ***  ********* ** *********** **************
protein          R  A  F  K  A  L  S  D  G  N  F  E  L  C  C  E  Q  Y  E  Y

AnCutA          CCGCCGGAAGCGGAACGCAGCTGGGAACGCCCGGCGCCGCAGCCGCGCCCGGTGGTGACC
AnCutA_opt      CCGCCGGAAGCAGAACGTTCATGGGAACGCCCGGCTCCGCAACCGCGCCCGGTCGTGACG
                *********** *****    ************** ***** *********** *****
protein          P  P  E  A  E  R  S  W  E  R  P  A  P  Q  P  R  P  V  V  T

AnCutA          ACCGCGCCGCCGCCGAGCCGCGGCGGCCGCGGCGGCGGCCGCGGCGGCCGCCATAGCAAC
AnCutA_opt      ACCGCACCGCCGCCGTCCCGTGGCGGTCGCGGCGGTGGCCGTGGTGGCCGCCATAGTAAC
                ***** *********  *** ***** ******** ***** ** *********** ***
protein          T  A  P  P  P  S  R  G  G  R  G  G  G  R  G  G  R  H  S  N

AnCutA          CAGTATAGCCGCGGCGGCCATAGCGGCGGCCGCCGCAGCAGCAACGCGGCGAACAAAAAC
AnCutA_opt      CAGTATTCCCGTGGTGGCCACAGTGGTGGCCGCCGTTCATCGAATGCGGCCAACAAAAAT
                ******  *** ** ***** ** ** ********       ** ***** ********
protein          Q  Y  S  R  G  G  H  S  G  G  R  R  S  S  N  A  A  N  K  N
```

```
AnCutA          CATTTTCGCCAGGCGAACAACGGCATGAGCGCGAGCGAACTGAGCCTGCAGGCGCAGCAT
AnCutA_opt      CATTTTCGCCAGGCGAACAATGGCATGTCAGCCTCGGAACTGAGCCTGCAGGCGCAACAC
                ******************** ******   **    ******************** **
protein          H  F  R  Q  A  N  N  G  M  S  A  S  E  L  S  L  Q  A  Q  H

AnCutA          GCGCAGTATCTGCTGCATGATCATCTGTATCAGCAGATTCAGATTCTGCAGGCGCAGGAA
AnCutA_opt      GCCCAGTATCTGCTGCATGATCACCTGTACCAGCAAATTCAAATCCTGCAGGCACAAGAA
                ** ******************** ***** ***** ***** ** ******** ** ***
protein          A  Q  Y  L  L  H  D  H  L  Y  Q  Q  I  Q  I  L  Q  A  Q  E

AnCutA          CAGGAACTGCGCCTGCAGCTGCAGAACCAGGCGCTGATTACCGGCCGCCCGCCGCCGGTG
AnCutA_opt      CAGGAACTGCGTCTGCAACTGCAGAACCAAGCTCTGATTACCGGTCGTCCGCCGCCGGTT
                *********** ***** *********** ** *********** ** ***********
protein          Q  E  L  R  L  Q  L  Q  N  Q  A  L  I  T  G  R  P  P  P  V

AnCutA          CTGATTCGCCAGCCGTTTATTCAGTTTCCGGTGCCGGCGCAGGAAACCAGCGGCGATGAT
AnCutA_opt      CTGATTCGTCAGCCGTTTATCCAATTCCCGGTCCCGGCACAGGAAACCAGCGGTGACGAT
                ******** *********** ** ** ***** ***** ************** ** ***
protein          L  I  R  Q  P  F  I  Q  F  P  V  P  A  Q  E  T  S  G  D  D

AnCutA          AGCAGCCGCAGCCGCAGCGGCAGCCGCAGCGGCGCGGTGAACCATCAGCCGATTAGCGTG
AnCutA_opt      AGCTCTCGTAGCCGCTCTGGTAGTCGTTCCGGCGCGGTGAATCATCAGCCGATCAGTGTG
                ***   ** ******   ** ** **   ************ *********** ** ***
protein          S  S  R  S  R  S  G  S  R  S  G  A  V  N  H  Q  P  I  S  V

AnCutA          CCGAGCCGCCAGAACGTGTATTATAACAGCACCTATCTGCCGGTGGCGGTGCCGGGCGGC
AnCutA_opt      CCGTCCCGCCAAAACGTTTATTACAATTCAACCTATCTGCCGGTCGCAGTGCCGGGTGGC
                ***  ****** ***** ***** **    ************** ** ******** ***
protein          P  S  R  Q  N  V  Y  Y  N  S  T  Y  L  P  V  A  V  P  G  G

AnCutA          CCGGGCAGCAACACCAACCCGCCGAGCCCGAGCGCGGCGGCGGCGATGCCGGATCTGCGC
AnCutA_opt      CCGGGTTCGAACACGAATCCGCCGTCACCGTCGGCAGCTGCAGCAATGCCGGATCTGCGC
                *****    ***** ** ******   ***   ** ** ** ** ***************
protein          P  G  S  N  T  N  P  P  S  P  S  A  A  A  A  M  P  D  L  R

AnCutA          CGCAACCCGCGCCGCAGCAGCGTGGCGAACGGCAGCCCGAACGGCAGCCTGCGCGCGCAT
AnCutA_opt      CGTAACCCGCGCCGTAGTTCCGTGGCCAACGGTTCACCGAATGGCTCGCTGCGTGCACAT
                ** *********** **   ****** *****    ***** ***   ***** ** ***
protein          R  N  P  R  R  S  S  V  A  N  G  S  P  N  G  S  L  R  A  H

AnCutA          AGCCAGCCGGCGCGCCCGGTGAACAGCCTGCCGAACTTTGCGCCGCTGTATAGCATTAGC
AnCutA_opt      AGTCAGCCGGCACGTCCGGTTAACTCCCTGCCGAATTTTGCGCCGCTGTACAGCATTTCT
                ** ******** ** ***** ***  ********* ************** ******
protein          S  Q  P  A  R  P  V  N  S  L  P  N  F  A  P  L  Y  S  I  S

AnCutA          CAGCCGACCGAAACCAGCCAGAGCAGCAAACAGCGCCATCCGAGCGGCGCGAGCGATGAT
AnCutA_opt      CAACCGACGGAAACCAGCCAGTCATCGAAACAACGCCACCCGTCAGGTGCCTCGGACGAT
                ** ***** ************      ***** ***** ***   ** **    ** ***
protein          Q  P  T  E  T  S  Q  S  S  K  Q  R  H  P  S  G  A  S  D  D

AnCutA          AGCCTGAGCGAAGAAGATGGCACCAGCATGCCGAACAGCCTGCCGAACCGCACCCTGTTT
AnCutA_opt      AGTCTGTCCGAAGAAGATGGCACCAGCATGCCGAACTCTCTGCCGAATCGTACGCTGTTT
                ** ***  ****************************   ******** ** ** ******
protein          S  L  S  E  E  D  G  T  S  M  P  N  S  L  P  N  R  T  L  F

AnCutA          ATTGATGAAGCGCGCAGCAACGATTATATGACCTATTATTTTGGCGGCAGCCCGCAGCTG
AnCutA_opt      ATCGATGAAGCGCGCAGTAACGACTATATGACCTATTACTTCGGTGGCTCCCCGCAGCTG
                ** ************** ***** ************** ** ** ***  **********
protein          I  D  E  A  R  S  N  D  Y  M  T  Y  Y  F  G  G  S  P  Q  L

AnCutA          CATGCGTATCATCAGAACGCGATGCTGAGCCCGATGAGCGCGAGCCTGGGCCTGGCGCTG
AnCutA_opt      CATGCATACCACCAAAATGCTATGCTGTCTCCGATGTCAGCATCGCTGGGTCTGGCACTG
                ***** ** ** ** ** ** ******   ******   **    ***** ***** ***
protein          H  A  Y  H  Q  N  A  M  L  S  P  M  S  A  S  L  G  L  A  L
```

```
AnCutA          CCGAACGGCGGCATTGTGCCGTGGGTGGGCAACCCGCAGGAATATATGAGCGCGGCGCCG
AnCutA_opt      CCGAACGGTGGCATTGTTCCGTGGGTCGGCAATCCGCAGGAATATATGAGTGCAGCTCCG
                ******** ******** ******** ***** ***************** ** ** ***
protein          P  N  G  G  I  V  P  W  V  G  N  P  Q  E  Y  M  S  A  A  P

AnCutA          GGCGCGGAAAGCTATGGCCATAGCTTTGATAACGGCAGCATTACCAGCCCGAAAAGCGGC
AnCutA_opt      GGTGCGGAAAGTTACGGCCATTCCTTTGATAATGGCAGCATCACCTCTCCGAAAAGTGGC
                ** ******** ** ******  ********* ******** ***   ******** ***
protein          G  A  E  S  Y  G  H  S  F  D  N  G  S  I  T  S  P  K  S  G

AnCutA          CAGAGCCAGCAGCGCATGGGCAGCCGCACCGCGGCGAGCAGCGATCGCGGCGGCCCGCTG
AnCutA_opt      CAGTCCCAGCAACGTATGGGTAGCCGTACGGCAGCAAGCTCTGATCGTGGTGGCCCGCTG
                ***  ****** ** ***** ***** ** ** ** ***   ***** ** *********
protein          Q  S  Q  Q  R  M  G  S  R  T  A  A  S  S  D  R  G  G  P  L

AnCutA          ATTGTGGATGGCAGCGTGCCGCCGAGCGAACAGCGCATGAGCATTCCGGCGGATCATTAT
AnCutA_opt      ATTGTCGACGGTTCAGTGCCGCCGTCGGAACAGCGCATGAGCATCCCGGCAGATCATTAT
                ***** ** **    *********   ***************** ***** *********
protein          I  V  D  G  S  V  P  P  S  E  Q  R  M  S  I  P  A  D  H  Y

AnCutA          GAACATTATGGCCTGATGAGCCATTGCCCGAGCAACAGCGATGATCAGAACACCGATACC
AnCutA_opt      GAACACTACGGCCTGATGTCTCACTGTCCGAGCAACTCTGACGATCAGAATACGGATACC
                ***** ** *********   ** ** *********   ** ******** ** ******
protein          E  H  Y  G  L  M  S  H  C  P  S  N  S  D  D  Q  N  T  D  T

AnCutA          CCGGCGAGCGTGAGCGATAGCTTTAGCCAGGATTTTCAGGATAGCAGCAGCGTGGAAATT
AnCutA_opt      CCGGCCAGTGTCTCCGACTCATTTTCGCAGGATTTCCAAGACAGTTCCTCAGTGGAAATT
                ***** ** **   ***    ***   ******** ** ** **   *   *********
protein          P  A  S  V  S  D  S  F  S  Q  D  F  Q  D  S  S  S  V  E  I

AnCutA          GATCCGAGCCCGTTTTTTAGCCGCCAGGGCTTTGATAGCCATGCGATGGGCGGCCCGGCG
AnCutA_opt      GATCCGAGCCCGTTTTTCTCTCGTCAGGGTTTTGATAGCCATGCAATGGGTGGCCCGGCT
                *****************    ** ***** ************** ***** ******** 
protein          D  P  S  P  F  F  S  R  Q  G  F  D  S  H  A  M  G  G  P  A

AnCutA          GATGTGTTTGTGAACGGCCATAGCGGCAAACAGAGCCTGCTGAGCAGCCGCCTGCAGGGC
AnCutA_opt      GACGTTTTCGTCAACGGTCACAGTGGCAAACAGTCCCTGCTGTCGAGCCGCCTGCAAGGT
                ** ** ** ** ***** ** ** *********  *******   *********** ** 
protein          D  V  F  V  N  G  H  S  G  K  Q  S  L  L  S  S  R  L  Q  G

AnCutA          CTGCATCTGAGCAACGGCGAAAAAGTGGATATTGCGAGCAAAAGCGCGCAGGATAAACCG
AnCutA_opt      CTGCATCTGTCTAATGGCGAAAAAGTGGATATTGCAAGCAAGTCTGCTCAGGACAAACCG
                *********   ** ******************** *****    ** ***** ******
protein          L  H  L  S  N  G  E  K  V  D  I  A  S  K  S  A  Q  D  K  P

AnCutA          AAAGCGGTGGTGCCGCATAACGATGCGGTGGAACGCGAAGCGCCGCGCGCGAAAAACCAT
AnCutA_opt      AAGGCAGTTGTCCCGCACAACGATGCTGTGGAACGTGAAGCACCGCGCGCTAAAAATCAT
                ** ** ** ** ***** ******** ******** ***** ******** ***** ***
protein          K  A  V  V  P  H  N  D  A  V  E  R  E  A  P  R  A  K  N  H

AnCutA          CATGCGGTGGATAAAGGCCCGAGCGCGAGCAGCTTTAGCCATGCGGCGGATAGCCATATT
AnCutA_opt      CACGCGGTTGATAAGGGCCCGAGCGCCTCTAGTTTTTCTCATGCAGCTGACTCACACATC
                ** ***** ***** ***********    ** ***   ***** ** **    ** ** 
protein          H  A  V  D  K  G  P  S  A  S  S  F  S  H  A  A  D  S  H  I

AnCutA          GCGAGCCCGAACGGCCGCCGCCGCCATCATGGCAGCGATACCAGCGAAAAAGTGAACGGC
AnCutA_opt      GCATCGCCGAACGGTCGCCGTCGCCATCACGGCTCAGATACCTCGGAAAAAGTTAACGGT
                **    ******** ***** ******** ***   ******   ******** ***** 
protein          A  S  P  N  G  R  R  R  H  H  G  S  D  T  S  E  K  V  N  G

AnCutA          ACCGCGCATAAAGGCAAACCGCGCAACCAGACCAGCAGCACCGCGAGCAACGATCAGAAA
AnCutA_opt      ACGGCACATAAAGGCAAGCCGCGTAATCAGACGTCCTCAACCGCAAGCAACGATCAAAAA
                ** ** *********** ***** ** *****   *   ***** *********** ***
protein          T  A  H  K  G  K  P  R  N  Q  T  S  S  T  A  S  N  D  Q  K
```

```
AnCutA          GAACGCAACGCGGGCGGCCCGCGCAAACCGAACGCGGCGGGCCTGACCCCGCAGGATAGC
AnCutA_opt      GAACGTAATGCGGGTGGCCCGCGTAAGCCGAATGCAGCAGGTCTGACCCCGCAGGACAGC
                ***** ** ***** ******** ** ***** ** ** ** ************** ***
protein          E  R  N  A  G  G  P  R  K  P  N  A  A  G  L  T  P  Q  D  S

AnCutA          AACCATAGCCATGCGAACAGCGGCTGGCAGACCACCAAACGCAAAAACAAAAAAAACAAC
AnCutA_opt      AACCATTCTCACGCGAATTCAGGCTGGCAAACGACCAAGCGTAAGAACAAAAAGAACAAC
                ******   ** *****    ******** ** ***** ** ** ******** ******
protein          N  H  S  H  A  N  S  G  W  Q  T  T  K  R  K  N  K  K  N  N

AnCutA          AAACTGTTTAGCGATCATCGCCATGGCCTGCAGAACGGCGCGGAACCGCTGCCGATTGAT
AnCutA_opt      AAGCTGTTCTCGGATCATCGCCACGGTCTGCAGAATGGCGCCGAACCGCTGCCGATTGAC
                ** *****    *********** ** ******** ***** *****************
protein          K  L  F  S  D  H  R  H  G  L  Q  N  G  A  E  P  L  P  I  D

AnCutA          GAAAGCCTGCGCAAAGGCGGC
AnCutA_opt      GAAAGCCTGCGCAAAGGTGGC
                ***************** ***
protein          E  S  L  R  K  G  G
```

```
TtCutA          ATGGATAGCCAGCCGGCGACCGCGGAAAGCAAACATCATTATCAGAGCTATCCGTGGCCG
TtCutA_opt      ATGGACTCGCAACCGGCAACCGCAGAATCAAAGCATCACTACCAATCATATCCGTGGCCG
                *****    ** ***** ***** ***    ** ***** ** **    ************
protein          M  D  S  Q  P  A  T  A  E  S  K  H  H  Y  Q  S  Y  P  W  P

TtCutA          GGCGCGCCGAGCGCGGGCACCACCAGCCGCGTGAAACCGGCGCCGGTGCCGGAAAGCAGC
TtCutA_opt      GGCGCTCCGTCAGCAGGCACCACCAGCCGCGTGAAACCGGCACCGGTTCCGGAAAGCTCT
                ***** ***    ** ************************** ***** *********
protein          G  A  P  S  A  G  T  T  S  R  V  K  P  A  P  V  P  E  S  S

TtCutA          CCGGGCTATTTTCCGGCGCAGCTGGATCTGCTGCCGATTCTGTTTCCGCAGCAGCTGTAT
TtCutA_opt      CCGGGTTATTTCCCGGCCCAGCTGGATCTGCTGCCGATTCTGTTTCCGCAGCAACTGTAT
                ***** ***** ***** *********************************** ******
protein          P  G  Y  F  P  A  Q  L  D  L  L  P  I  L  F  P  Q  Q  L  Y

TtCutA          CAGGATAAACTGCTGCAGTATAACAAACTGGTGAGCGCGGGCCGCGGCGGCGGCGGCGGC
TtCutA_opt      CAAGACAAGCTGCTGCAGTACAACAAACTGGTCAGCGCAGGTCGTGGCGGTGGCGGTGGC
                ** ** ** *********** *********** ***** ** ** ***** ***** ***
protein          Q  D  K  L  L  Q  Y  N  K  L  V  S  A  G  R  G  G  G  G  G

TtCutA          CTGCAGACCCCGCCGCATCCGCCGGTGCTGAGCAGCCCGCAGAGCAAAAGCAGCACCCGC
TtCutA_opt      CTGCAGACGCCGCCGCACCCGCCGGTTCTGAGTTCCCCGCAGAGTAAATCATCGACCCGC
                ******** ******** ******** *****   ********* ***      ******
protein          L  Q  T  P  P  H  P  P  V  L  S  S  P  Q  S  K  S  S  T  R

TtCutA          AGCCGCAGCAGCAGCAAAGTGAGCAACAAAGATAACAGCGCGCAGGCGAAAAGCGCGCCG
TtCutA_opt      TCCCGTAGCTCTAGTAAGGTGTCTAACAAAGATAATTCCGCCCAAGCAAAAAGCGCGCCG
                  *** ***   ** ** ***   ***********   *** ** ** ************
protein          S  R  S  S  S  K  V  S  N  K  D  N  S  A  Q  A  K  S  A  P

TtCutA          CAGAGCTGCCATGGCAACAAACCGGCGAAAAACAGCGAACCGGCGGAACGCGCGCTGATG
TtCutA_opt      CAGTCTTGCCATGGTAACAAGCCGGCCAAAAATTCAGAACCGGCAGAACGCGCTCTGATG
                ***   ******** ***** ***** *****    ******** ******** ******
protein          Q  S  C  H  G  N  K  P  A  K  N  S  E  P  A  E  R  A  L  M

TtCutA          AAAAGCGGCAAAGAACCGAGCAGCAAAATGAGCGCGAAAAAACCGAGCGAACAGGCGGTG
TtCutA_opt      AAGTCGGGCAAAGAACCGTCCTCAAAAATGTCAGCAAAAAAGCCGTCGGAACAGGCTGTC
                **    ************   *   ******   ** ***** ***   ******** **
protein          K  S  G  K  E  P  S  S  K  M  S  A  K  K  P  S  E  Q  A  V
```

```
TtCutA         GATGGCGATGCGGCGGTGCGCCAGCCGCCGGCGAGCCGCCAGGGCGCGCCGGCGACCGCG
TtCutA_opt     GATGGTGACGCGGCCGTGCGTCAACCGCCGGCATCACGTCAGGGTGCTCCGGCAACCGCC
               ***** ** ***** ***** ** ********    ** ***** ** ***** *****
protein         D  G  D  A  A  V  R  Q  P  P  A  S  R  Q  G  A  P  A  T  A

TtCutA         ACCCAGAGCAGCAGCGTGCCGAGCACCCCGCATCAGCATCCGCGCAAATTTAGCTTTGAA
TtCutA_opt     ACGCAATCGAGCTCTGTTCCGTCGACCCCGCATCAGCACCCGCGCAAATTTTCCTTCGAA
               ** **    ***   ** ***   ************** ************  *** ***
protein         T  Q  S  S  S  V  P  S  T  P  H  Q  H  P  R  K  F  S  F  E

TtCutA         AGCCGCGAACAGAGCCCGGGCGCGGCGCAGAACCATAGCCCGCGCAGCGCGTATAGCGAA
TtCutA_opt     TCACGTGAACAAAGTCCGGGTGCAGCTCAGAACCATTCGCCGCGCAGCGCCTATTCTGAA
                  ** ***** ** ***** ** ** *********   *********** ***   ***
protein         S  R  E  Q  S  P  G  A  A  Q  N  H  S  P  R  S  A  Y  S  E

TtCutA         ACCAACGGCAACGTGCCGAGCCTGCGCCCGCTGCCGCCGCGCCTGGGCGGCTGCCGCTTT
TtCutA_opt     ACCAACGGTAATGTGCCGAGCCTGCGTCCGCTGCCGCCGCGTCTGGGTGGCTGTCGCTTC
               ******** ** ************** ************** ***** ***** *****
protein         T  N  G  N  V  P  S  L  R  P  L  P  P  R  L  G  G  C  R  F

TtCutA         GAAACCGCGATTCCGTTTAGCCGCCGCCGCATGCCGTATAGCCTGGGCACCGATAAACTG
TtCutA_opt     GAAACGGCAATTCCGTTTAGCCGTCGCCGTATGCCGTACTCTCTGGGCACCGATAAGCTG
               ***** ** ************** ***** ********    ************** ***
protein         E  T  A  I  P  F  S  R  R  R  M  P  Y  S  L  G  T  D  K  L

TtCutA         GAAAAAGTGGATCCGGATAAAATTAAAAGCAAACTGAGCGAAGATGTGGAACGCAAACTG
TtCutA_opt     GAAAAAGTGGATCCGGACAAAATCAAGAGTAAACTGTCCGAAGACGTTGAACGTAAACTG
               ***************** ***** ** ** ******  ****** ** ***** ******
protein         E  K  V  D  P  D  K  I  K  S  K  L  S  E  D  V  E  R  K  L

TtCutA         GAAACCGATATGCGCGAACTGTATGATCGCCTGCTGCCGACCGAAGCGATTGAAGTGAAC
TtCutA_opt     GAAACGGATATGCGTGAACTGTATGACCGCCTGCTGCCGACCGAAGCGATTGAAGTTAAC
               ***** ******** *********** ***************************** ***
protein         E  T  D  M  R  E  L  Y  D  R  L  L  P  T  E  A  I  E  V  N

TtCutA         CGCCGCGAACTGGTGAGCAAACTGGAACGCCTGTTTAACACCGAATGGCCGGGCCATGAT
TtCutA_opt     CGCCGTGAACTGGTCAGCAAACTGGAACGCCTGTTCAATACCGAATGGCCGGGTCATGAT
               ***** ******** ******************** ** ************** ******
protein         R  R  E  L  V  S  K  L  E  R  L  F  N  T  E  W  P  G  H  D

TtCutA         ATTCGCGTGCATCTGTTTGGCAGCAGCGGCAACCTGCTGTGCAGCGATGATAGCGATGTG
TtCutA_opt     ATCCGTGTTCACCTGTTTGGTAGTTCCGGCAATCTGCTGTGCTCAGATGACTCGGATGTT
               ** ** ** ** ******** **   ****** *********   *****    *****
protein         I  R  V  H  L  F  G  S  S  G  N  L  L  C  S  D  D  S  D  V

TtCutA         GATATTTGCATTACCACCCCGTGGCGCGAACTGGAAAGCGTGTGCATGATTGCGGAACTG
TtCutA_opt     GACATTTGTATCACCACGCCGTGGCGCGAACTGGAAAGCGTCTGCATGATTGCAGAACTG
               ** ***** ** ***** *********************** *********** ******
protein         D  I  C  I  T  T  P  W  R  E  L  E  S  V  C  M  I  A  E  L

TtCutA         CTGGATCGCCATGGCATGGAAAAAGTGGTGTGCGTGAGCAGCGCGAAAGTGCCGATTGTG
TtCutA_opt     CTGGATCGTCATGGCATGGAAAAGGTGGTTTGTGTGTCATCGGCGAAGGTGCCGATTGTT
               ******** ************** ***** ** ***      ***** ***********
protein         L  D  R  H  G  M  E  K  V  V  C  V  S  S  A  K  V  P  I  V

TtCutA         AAAATTTGGGATCCGGAACTGAAACTGGCGTGCGATATGAACGTGAACAACACCCTGGCG
TtCutA_opt     AAAATCTGGGATCCGGAACTGAAACTGGCATGCGACATGAACGTTAACAATACCCTGGCT
               ***** *********************** ***** ******** ***** ********
protein         K  I  W  D  P  E  L  K  L  A  C  D  M  N  V  N  N  T  L  A

TtCutA         CTGGAAAACACCCGCATGGTGCGCACCTATGTGAGCATTGATGATCGCGTGCGCCCGCTG
TtCutA_opt     CTGGAAAATACCCGCATGGTCCGTACGTATGTGAGCATTGATGACCGCGTGCGTCCGCTG
               ******** *********** ** ** ***************** ******** ******
protein         L  E  N  T  R  M  V  R  T  Y  V  S  I  D  D  R  V  R  P  L
```

```
TtCutA          GCGATGATTATTAAATATTGGACCCGCCGCCGCGTGGTGAACGATGCGGCGTTTGGCGGC
TtCutA_opt      GCGATGATTATCAAATACTGGACCCGCCGTCGCGTCGTGAACGATGCAGCCTTCGGTGGC
                *********** ***** *********** ***** *********** ** ** ** ***
protein          A  M  I  I  K  Y  W  T  R  R  R  V  V  N  D  A  A  F  G  G

TtCutA          ACCCTGAGCAGCTATACCTGGATTTGCATGATTATTGCGTTTCTGCAGCTGCGCGATCCG
TtCutA_opt      ACCCTGAGCTCTTACACGTGGATCTGTATGATTATCGCATTTCTGCAGCTGCGTGATCCG
                *********   ** ** ***** ** ******** ** ************** ******
protein          T  L  S  S  Y  T  W  I  C  M  I  I  A  F  L  Q  L  R  D  P

TtCutA          CCGGTGCTGCCGGCGCTGCATCAGCAGCATGATCTGAAACTGGTGAAACAGGATGGCGCG
TtCutA_opt      CCGGTCCTGCCGGCTCTGCATCAGCAACACGACCTGAAGCTGGTGAAACAGGATGGTGCG
                ***** ******** *********** ** ** ***** ***************** ***
protein          P  V  L  P  A  L  H  Q  Q  H  D  L  K  L  V  K  Q  D  G  A

TtCutA          CTGAGCGATTTTGCGGATGATATTCCGAAACTGCGCGGCTTTGGCGCGAAAAACAAAGAT
TtCutA_opt      CTGTCTGACTTCGCCGATGACATTCCGAAACTGCGTGGTTTTGGCGCGAAGAATAAAGAT
                ***    ** ** ** ***** ************** ** *********** ** ******
protein          L  S  D  F  A  D  D  I  P  K  L  R  G  F  G  A  K  N  K  D

TtCutA          AGCCTGGCGGTGCTGCTGTTTCAGTTTTTTCGCTTTTATGCGCATGAATTTGATTATGAT
TtCutA_opt      AGTCTGGCCGTGCTGCTGTTTCAGTTTTTCCGCTTCTATGCACACGAATTTGATTATGAC
                ** ***** ******************** ***** ***** ** **************
protein          S  L  A  V  L  L  F  Q  F  F  R  F  Y  A  H  E  F  D  Y  D

TtCutA          AAATATACCCTGAGCATTCGCATGGGCACCCTGCTGACCAAAGCGGAAAAAAACTGGCAG
TtCutA_opt      AAATACACCCTGAGCATCCGTATGGGCACCCTGCTGACGAAGGCAGAAAAGAACTGGCAG
                ***** *********** ** ***************** ** ** ***** *********
protein          K  Y  T  L  S  I  R  M  G  T  L  L  T  K  A  E  K  N  W  Q

TtCutA          TATCTGGTGAACAACGCGCTGTGCGTGGAAGAACCGTTTAACGATGGCCGCAACCTGGGC
TtCutA_opt      TACCTGGTGAACAATGCTCTGTGCGTTGAAGAACCGTTTAATGATGGTCGCAACCTGGGC
                ** *********** ** ******** ************** ***** ************
protein          Y  L  V  N  N  A  L  C  V  E  E  P  F  N  D  G  R  N  L  G

TtCutA          AACACCGCGGATGAAACCAGCTTTCGCGGCCTGCATATGGAACTGCGCCGCGCGTTTGAT
TtCutA_opt      AATACCGCCGACGAAACGAGCTTCCGTGGTCTGCATATGGAACTGCGTCGCGCATTTGAT
                ** ***** ** ***** ***** ** ** ***************** ***** ******
protein          N  T  A  D  E  T  S  F  R  G  L  H  M  E  L  R  R  A  F  D

TtCutA          CTGATTGCGGAAGGCAAACTGGAAGAATGCTGCGAACAGTATGTGTTTCCGAAAGAAGAA
TtCutA_opt      CTGATTGCTGAAGGCAAACTGGAAGAATGCTGTGAACAGTATGTGTTTCCGAAAGAAGAA
                ******** ********************** ***************************
protein          L  I  A  E  G  K  L  E  E  C  C  E  Q  Y  V  F  P  K  E  E

TtCutA          GAACGCGTGTGGCAGAAACCGCCGCCGGCGCCGCGCCCGATTCTGCTGCGCAGCAGCAGC
TtCutA_opt      GAACGCGTTTGGCAGAAACCGCCGCCGGCACCGCGTCCGATCCTGCTGCGCAGTTCCTCA
                ******** ******************** ***** ***** ***********   *
protein          E  R  V  W  Q  K  P  P  P  A  P  R  P  I  L  L  R  S  S  S

TtCutA          CAGCAGCATGGCAACAGCGGCAACCGCCTGCGCGGCGGCGGCTTTCGCGGCAACCGCCAG
TtCutA_opt      CAGCAACACGGTAACAGCGGCAATCGCCTGCGTGGTGGCGGTTTCCGCGGTAACCGTCAG
                ***** ** ** *********** ******** ** ***** ** ***** ***** ***
protein          Q  Q  H  G  N  S  G  N  R  L  R  G  G  G  F  R  G  N  R  Q

TtCutA          TTTCATCGCAACGGCGGCGCGGGCCGCCGCAGCAGCAGCAGCGTGACCTATGATCCGAAC
TtCutA_opt      TTTCATCGTAATGGCGGTGCAGGTCGTCGCTCGAGCTCTAGTGTCACGTATGATCCGAAC
                ******** ** ***** ** ** ** ***   ***   ** ** ** ************
protein          F  H  R  N  G  G  A  G  R  R  S  S  S  S  V  T  Y  D  P  N

TtCutA          AGCGTGTATAGCCAGGCGGGCATGCCGCCGACCATGAGCCCGCAGGATCTGCTGTGGTAT
TtCutA_opt      AGCGTGTACTCTCAAGCCGGTATGCCGCCGACCATGAGCCCGCAGGATCTGCTGTGGTAT
                ********    ** ** ** ***************************************
protein          S  V  Y  S  Q  A  G  M  P  P  T  M  S  P  Q  D  L  L  W  Y
```

```
TtCutA          CAGGCGCAGAGCCCGCAGCTGAGCGTGCCGCAGGAAATTCTGACCACCAGCCTGAACGCG
TtCutA_opt      CAAGCTCAGAGTCCGCAACTGTCCGTGCCGCAGGAAATTCTGACCACGTCACTGAACGCG
                ** ** ***** ***** ***  ************************    *********
protein          Q  A  Q  S  P  Q  L  S  V  P  Q  E  I  L  T  T  S  L  N  A

TtCutA          CTGGCGGCGCAGGAAAGCATGCGCTTTCAGCTGTATACCCAGCTGAACCAGCAGCAGGTG
TtCutA_opt      CTGGCAGCTCAGGAATCGATGCGCTTTCAACTGTACACCCAGCTGAATCAGCAACAGGTT
                ***** ** ******   *********** ***** *********** ***** *****
protein          L  A  A  Q  E  S  M  R  F  Q  L  Y  T  Q  L  N  Q  Q  Q  V

TtCutA          CTGGCGCATGCGCAGCGCGTGCAGCAGGGCGCGAGCACCACCGATCGCAGCCGCACCAAC
TtCutA_opt      CTGGCACACGCTCAGCGTGTCCAACAGGGCGCCAGCACCACGGATCGCTCACGTACGAAC
                ***** ** ** ***** ** ** ******** ******** ******   ** ** ***
protein          L  A  H  A  Q  R  V  Q  Q  G  A  S  T  T  D  R  S  R  T  N

TtCutA          AGCCTGGATAACCCGCCGCTGACCGCGCCGGTGCGCCCGGAACTGATGTATGGCATTGGC
TtCutA_opt      TCGCTGGACAATCCGCCGCTGACCGCACCGGTTCGTCCGGAACTGATGTACGGTATTGGC
                   ***** ** ************** ***** ** ************** ** ******
protein          S  L  D  N  P  P  L  T  A  P  V  R  P  E  L  M  Y  G  I  G

TtCutA          TTTCCGATTCAGCATGCGCCGTATTTTCATCCGGGCTTTACCACCTATCCGAGCAGCCCG
TtCutA_opt      TTCCCGATCCAGCATGCGCCGTATTTTCACCCGGGTTTCACCACGTACCCGTCCTCACCG
                ** ***** ******************** ***** ** ***** ** ***  *   ***
protein          F  P  I  Q  H  A  P  Y  F  H  P  G  F  T  T  Y  P  S  S  P

TtCutA          GCGACCGCGGCGCCGCCGAGCAACGGCCAGCCGGAATTTCGCCGCAGCCTGCATCGCACC
TtCutA_opt      GCAACCGCAGCACCGCCGAGCAACGGCCAGCCGGAATTTCGTCGCTCTCTGCATCGTACC
                ** ***** ** ***************************** ***   ******** ***
protein          A  T  A  A  P  P  S  N  G  Q  P  E  F  R  R  S  L  H  R  T

TtCutA          GCGGCGGCGAGCGATAGCGGCGTGAGCCCGGGCAGCGGCGCGCTGCGCAGCCAGAGCCAG
TtCutA_opt      GCAGCTGCGTCAGATTCGGGTGTGAGTCCGGGTTCCGGTGCACTGCGTAGCCAATCTCAG
                ** ** ***   ***   ** ***** *****   *** ** ***** *****    ***
protein          A  A  A  S  D  S  G  V  S  P  G  S  G  A  L  R  S  Q  S  Q

TtCutA          CCGGCGAGCCGCACCACCCTGCCGCCGCATAGCATGGCGGTGTATAGCACCAGCAGCCAG
TtCutA_opt      CCGGCTTCTCGTACCACGCTGCCGCCGCATAGCATGGCGGTTTATTCTACCTCGAGCCAG
                *****    ** ***** *********************** ***   ***   ******
protein          P  A  S  R  T  T  L  P  P  H  S  M  A  V  Y  S  T  S  S  Q

TtCutA          AGCCATACCGGCATTGGCGCGAGCCAGAGCCGCCAGGTGCATGATGTGGCGGTGCCGAGC
TtCutA_opt      AGTCACACGGGTATCGGCGCGAGTCAATCCCGCCAGGTGCATGATGTTGCCGTCCCGAGC
                ** ** ** ** ** ******** **   ****************** ** ** ******
protein          S  H  T  G  I  G  A  S  Q  S  R  Q  V  H  D  V  A  V  P  S

TtCutA          TTTATGCCGGATGAACGCGGCGATCTGGGCTTTGATGATGTGCCGACCAAACCGATTAGC
TtCutA_opt      TTCATGCCGGATGAACGTGGTGACCTGGGCTTTGATGACGTTCCGACCAAACCGATTTCA
                ** ************** ** ** ************** ** ***************
protein          F  M  P  D  E  R  G  D  L  G  F  D  D  V  P  T  K  P  I  S

TtCutA          GATAGCCCGCCGGAAAACGATGGCCCGCGCTTTATTGGCTATTATGTGAACGGCCATAGC
TtCutA_opt      GATTCGCCGCCGGAAAATGACGGTCCGCGCTTCATCGGCTATTACGTGAACGGTCACAGC
                ***   *********** ** ** ******** ** ******** ******** ** ***
protein          D  S  P  P  E  N  D  G  P  R  F  I  G  Y  Y  V  N  G  H  S

TtCutA          GGCAGCAGCCGCCCGGCGAACGGCTTTCCGAGCGTGGGCGTGGCGAGCGCGGAAAGCAGC
TtCutA_opt      GGCTCTAGTCGTCCGGCCAATGGTTTTCCGTCCGTGGGCGTTGCGTCAGCCGAATCCTCA
                ***   ** ** ***** ** ** ******  ********* ***   ** ***  *
protein          G  S  S  R  P  A  N  G  F  P  S  V  G  V  A  S  A  E  S  S

TtCutA          CAGAGCGGCCAGGGCGGCGGCCGCCGCCGCCTGAGCACCGATCAGCTGCCGCAGAGCGTG
TtCutA_opt      CAATCTGGTCAGGGCGGTGGCCGTCGCCGTCTGAGTACGGATCAACTGCCGCAGTCCGTG
                **    ** ******** ***** ***** ***** ** ***** *********  ****
protein          Q  S  G  Q  G  G  G  R  R  R  L  S  T  D  Q  L  P  Q  S  V
```

```
TtCutA        CTGGATCGCCGCATGAAACGCACCAGCCGCAGCCCGAGCCCGATGGGCCATGCGCGCGCG
TtCutA_opt    CTGGACCGCCGTATGAAACGCACCAGCCGTTCTCCGAGTCCGATGGGTCATGCACGTGCT
              ***** ***** *****************    ***** ******** ***** ** ** 
protein        L  D  R  R  M  K  R  T  S  R  S  P  S  P  M  G  H  A  R  A 

TtCutA        CTGAGCGTGGGCGTGAACAGCGCGCCGCTGCTGAGCGCGCCGTTTGCGCAGAGCAACGGC
TtCutA_opt    CTGTCGGTCGGCGTGAACAGCGCACCGCTGCTGTCTGCACCGTTTGCACAGTCAAACGGC
              ***   ** ************** *********   ** ******** ***   ******
protein        L  S  V  G  V  N  S  A  P  L  L  S  A  P  F  A  Q  S  N  G 

TtCutA        AAACTGCCGAGCAAACCGCTGGTGGTGAACGGCACCGCGCCGAAAACCGCGAGCACCCCG
TtCutA_opt    AAGCTGCCGTCGAAACCGCTGGTTGTCAATGGTACCGCACCGAAAACCGCAAGCACGCCG
              ** ******   *********** ** ** ** ***** *********** ***** ***
protein        K  L  P  S  K  P  L  V  V  N  G  T  A  P  K  T  A  S  T  P 

TtCutA        GGCAGCGGCCGCCCGACCGTGGTGACCGAAAGCCTGGTGAGCGATGATAGCCTGCATGAT
TtCutA_opt    GGTTCTGGTCGTCCGACCGTGGTTACGGAATCACTGGTGTCGGATGACAGCCTGCATGAT
              **   ** ** *********** ** ***   ******   ***** ************
protein        G  S  G  R  P  T  V  V  T  E  S  L  V  S  D  D  S  L  H  D 

TtCutA        AACGCGCTGCATATTCATCAGGGCGTGAGCCCGAGCAGCAGCTGGCCGGAACAGCAGCCG
TtCutA_opt    AATGCCCTGCATATTCACCAGGGTGTTTCCCCGTCGAGCTCTTGGCCGGAACAACAGCCG
              ** ** *********** ***** **   ****   ***   *********** ******
protein        N  A  L  H  I  H  Q  G  V  S  P  S  S  S  W  P  E  Q  Q  P 

TtCutA        ACCCCGCCGAGCAGCGCGGCGGAAAGCGTGGCGACCCGCGTGCCGGAACGCCCGGTGATT
TtCutA_opt    ACGCCGCCGAGTTCCGCAGCAGAAAGCGTGGCAACCCGTGTCCCGGAACGTCCGGTTATC
              ** ********   *** ** *********** ***** ** ******** ***** ** 
protein        T  P  P  S  S  A  A  E  S  V  A  T  R  V  P  E  R  P  V  I 

TtCutA        GTGAACGGCAGCAGCACCAACCGCAGCCCGAGCGCGACCGTGCCGAACCTGCTGGATACC
TtCutA_opt    GTCAACGGCTCATCGACCAATCGCTCCCCGTCAGCCACGGTTCCGAACCTGCTGGATACC
              ** ******      ***** ***  ****   ** ** ** ******************
protein        V  N  G  S  S  T  N  R  S  P  S  A  T  V  P  N  L  L  D  T 

TtCutA        AGCTTTCAGCAGCGCGTGGCGATGGGCGCGGGCCTGCATCTGCCGTATAGCGCGGTGGTG
TtCutA_opt    AGCTTCCAACAGCGTGTCGCAATGGGTGCTGGCCTGCACCTGCCGTATAGTGCGGTCGTG
              ***** ** ***** ** ** ***** ** ******** *********** ***** ***
protein        S  F  Q  Q  R  V  A  M  G  A  G  L  H  L  P  Y  S  A  V  V 

TtCutA        GGCGATCCGAACGGCCTGGCGGGCTTTAGCCGCGCGCATACCCGCCAGCAGCTGGCGCCG
TtCutA_opt    GGTGATCCGAATGGTCTGGCGGGCTTTTCCCGCGCCCATACGCGTCAACAGCTGGCACCG
              ** ******** ** ************  ****** ***** ** ** ******** ***
protein        G  D  P  N  G  L  A  G  F  S  R  A  H  T  R  Q  Q  L  A  P 

TtCutA        CTGGATCTGGCGACCGGCGATTATGCGGCGCATCAGGATCTGCCGCATCTGAGCCCGGTG
TtCutA_opt    CTGGATCTGGCTACCGGCGACTATGCTGCGCATCAGGATCTGCCGCACCTGAGTCCGGTC
              *********** ******** ***** ******************** ***** ***** 
protein        L  D  L  A  T  G  D  Y  A  A  H  Q  D  L  P  H  L  S  P  V 

TtCutA        TATGAACATCGCACCCCGAGCCCGACCGTGCCGCGCAAATTTGAACCGAGCCCGATTAGC
TtCutA_opt    TACGAACACCGTACCCCGTCCCCGACGGTGCCGCGTAAGTTTGAACCGTCTCCGATTAGT
              ** ***** ** ******  ****** ******** ** *********   ******** 
protein        Y  E  H  R  T  P  S  P  T  V  P  R  K  F  E  P  S  P  I  S 

TtCutA        CAGCTGAGCGGCGCGGGCGCGAAAGATTTTCGCAGCGATTTTAGCCCGAAAGCGACCCAG
TtCutA_opt    CAGCTGTCCGGTGCGGGCGCCAAAGATTTTCGCTCAGACTTCTCGCCGAAGGCAACCCAG
              ******  *** ******** ************   ** **   ***** ** ******
protein        Q  L  S  G  A  G  A  K  D  F  R  S  D  F  S  P  K  A  T  Q 

TtCutA        AAAGCGCCGAGCAGCCGCAGCCCGAACGGCGGCGCGCCGACCCGCCAGCAGGATAGCGCG
TtCutA_opt    AAAGCTCCGAGCTCTCGTAGCCCGAATGGTGGCGCACCGACGCGTCAACAGGATTCTGCA
              ***** ******   ** ******** ** ***** ***** ** ** ******   ** 
protein        K  A  P  S  S  R  S  P  N  G  G  A  P  T  R  Q  Q  D  S  A 
```

```
TtCutA        ACCAGCCGCAGCCCGGCGCAGGATCAGCGCATTAACGGCCTGGCGCGCGAAAACGGCCAT
TtCutA_opt    ACCTCGCGTAGCCCGGCACAAGACCAGCGTATCAACGGTCTGGCGCGCGAAAATGGCCAT
              ***   ** ******** ** ** ***** ** ***** ************** ******
protein        T  S  R  S  P  A  Q  D  Q  R  I  N  G  L  A  R  E  N  G  H

TtCutA        GTGCGCGCGGCGAAAAGCCAGACCGAAAGCTTTAGCGGCTGGCAGAAAACCAAAACCCGC
TtCutA_opt    GTGCGTGCCGCAAAAAGTCAAACCGAATCTTTCAGTGGTTGGCAGAAGACCAAAACGCGC
              ***** ** ** ***** ** ******   ** ** ** ******** ******** ***
protein        V  R  A  A  K  S  Q  T  E  S  F  S  G  W  Q  K  T  K  T  R

TtCutA        AAAAAAGGCGGCGCGGCGGCGGCGGAACTGAAACAGAGCAGCGCGGCGAACGGCTTTCCG
TtCutA_opt    AAAAAGGGTGGCGCTGCAGCAGCAGAACTGAAACAAAGTTCCGCTGCGAACGGCTTTCCG
              ***** ** ***** ** ** ** *********** **   *** ***************
protein        K  K  G  G  A  A  A  A  E  L  K  Q  S  S  A  A  N  G  F  P

TtCutA        CAGAGCGAACCGCCGCCGAAAAACGAAGCGGATCGCAAAGGCGGC
TtCutA_opt    CAGAGCGAACCGCCGCCGAAGAATGAAGCGGATCGTAAAGGTGGC
              ******************** ** *********** ***** ***
protein        Q  S  E  P  P  P  K  N  E  A  D  R  K  G  G
```

SUPPLEMENTAL FIGURE 1. Alignments of the native (AnCutA/TtCutA) and recoded (AnCutA_opt/TtCutA_opt) open reading frames corresponding to AnCutA and TtCutA proteins. Nucleotides unchanged during recoding are marked with asterisks. Amino acid sequences of the proteins are shown in bold.

SUPPLEMENTAL TABLES

SUPPLEMENTAL TABLE 1. Frequency of nucleotide incorporation into the tails synthesized by full-length wild-type AnCutA and TtCutA depending on the identity of 1 or 2 preceding nucleotides at the primer's end. Higher and lower frequencies, as compared to the average values (marked in grey), which are discussed in the main text, are indicated with green and red colors, respectively.

Primer's end	AnCutA								TtCutA							
	ss22-A_4				ss22-U_4				ss22-A_4				ss22-U_4			
	A	G	C	U	A	G	C	U	A	G	C	U	A	G	C	U
Average frequency of nucleotide incorporation [%]																
*	**30.9**	**0.5**	**52.8**	**15.9**	**20.2**	**1.0**	**65.7**	**13.1**	**48.3**	**0.5**	**43.5**	**7.6**	**36.7**	**1.1**	**56.3**	**5.9**
Frequency of nucleotide incorporation depending on <u>1 preceding nucleotide</u>																
*A	42.2	0.3	47.0	10.5	28.0	0.6	63.0	8.4	58.5	0.4	36.1	4.9	45.9	0.7	50.1	3.3
*G	28.0	4.0	50.3	17.7	18.6	1.5	68.3	11.6	39.4	2.2	45.8	12.5	30.9	2.8	59.8	6.5
*C	25.1	0.6	52.5	21.8	22.6	1.4	55.7	20.3	39.5	0.7	48.6	11.3	33.7	1.3	56.9	8.0
*U	14.6	0.4	76.6	8.4	11.1	0.7	80.8	7.4	18.4	0.5	76.0	5.1	19.2	0.9	75.4	4.5
Frequency of nucleotide incorporation depending on <u>2 preceding nucleotides</u>																
*AA	48.9	0.3	39.4	11.5	31.3	0.4	61.8	6.5	60.4	0.4	33.9	5.3	47.9	0.7	48.5	2.9
*AG	28.7	2.9	53.4	15.1	15.7	3.1	73.1	8.1	42.5	1.4	43.8	12.2	30.7	2.7	59.7	6.9
*AC	25.9	0.3	56.0	17.8	18.6	0.6	65.5	15.4	39.2	0.5	50.9	9.4	29.9	0.9	62.6	6.6
*AU	20.3	0.2	71.5	7.9	11.4	0.3	83.9	4.4	21.2	0.4	73.7	4.7	19.2	0.8	76.6	3.4
*GA	26.5	2.4	62.7	8.4	18.1	2.0	72.7	7.2	55.2	1.5	36.8	6.5	53.4	0.9	42.4	3.4
*GG	19.2	32.7	9.6	38.5	34.9	14.0	41.9	9.3	23.2	27.0	29.5	20.4	27.2	6.6	57.9	8.3
*GC	24.5	0.8	48.4	26.3	20.6	1.6	56.4	21.4	37.7	1.5	47.8	13.0	41.0	1.5	52.9	4.6
*GU	14.7	6.1	63.2	16.0	15.8	3.7	66.6	13.9	20.5	3.7	68.3	7.4	24.2	3.0	67.1	5.7
*CA	32.1	0.3	59.6	8.1	28.8	0.7	61.5	9.0	53.2	0.4	42.4	4.1	44.0	0.8	51.8	3.4
*CG	26.7	1.4	55.6	16.3	24.2	1.1	62.7	12.0	37.2	1.3	49.7	11.8	31.6	2.6	60.2	5.6
*CC	25.1	0.6	49.5	24.8	17.7	1.5	53.7	27.0	39.1	0.7	47.7	12.4	29.2	1.6	60.4	8.7
*CU	12.1	0.4	79.4	8.1	10.5	0.7	81.3	7.5	16.7	0.4	77.7	5.2	18.9	0.9	75.4	4.8
*UA	31.4	0.4	58.2	10.0	23.6	0.6	67.5	8.3	51.6	0.4	42.7	5.3	47.6	0.7	47.5	4.2
*UG	25.6	11.0	43.9	19.5	15.0	1.9	68.4	14.8	33.0	7.3	43.2	16.5	31.1	3.4	58.3	7.2
*UC	22.9	0.9	55.1	21.2	22.9	1.8	53.9	21.5	40.6	0.9	44.8	13.7	44.1	1.5	42.9	11.5
*UU	13.9	0.8	76.6	8.8	11.7	1.0	78.4	8.9	17.8	0.9	75.2	6.1	17.2	1.0	77.3	4.5

SUPPLEMENTAL TABLE 2. Plasmids used in this study.

Plasmid	Source or genotype
pET28M N-6xHis-SUMOTag	(Lebreton et al, 2008; Malakhov et al, 2004)
pET28 C-6xHisTag	pET-28b(+) from Novagen
pUC57-AnCutA	this study; recoded AnCutA ORF synthesized by Genscript
pUC57-TtCutA	this study; recoded TtCutA ORF synthesized by Genscript
pCC1	[pET28 C-6xHisTag] AnCutA^{WT}, full-length
pCC2	[pET28 C-6xHisTag] AnCutA$^{D203A/D205A}$ (DADA), full-length
pCC3	[pET28M N-6xHis-SUMOTag] AnCutA_tr^{WT}
pCC4	[pET28 C-6xHisTag] TtCutA^{WT}, full-length
pCC5	[pET28M N-6xHis-SUMOTag] TtCutA_tr1^{WT}
pCC6	[pET28 C-6xHisTag] TtCutA_tr2^{WT}
pCC7	[pET28 C-6xHisTag] TtCutA_tr3^{WT}
pCC8	[pET28 C-6xHisTag] TtCutA_tr3^{R557A} (PAP1)
pCC9	[pET28 C-6xHisTag] TtCutA_tr3^{R557Q} PAP2 (PAP2)
pCC10	[pET28 C-6xHisTag] TtCutA_tr3$^{E551D/N554E/R557A}$ (PAP3)
pCC11	[pET28 C-6xHisTag] TtCutA_tr3$^{E551T/R557A/L559S}$ (PAP4)
pCC12	[pET28 C-6xHisTag] TtCutA_tr3$^{R557N/N561R}$ (PAP5)
pCC13	[pET28 C-6xHisTag] TtCutA_tr3 $^{E551D/R557N/N561R}$ (PAP6)
pCC14	[pET28 C-6xHisTag] TtCutA_tr3^{R557H} (PUP1)
pCC15	[pET28 C-6xHisTag] TtCutA_tr3$^{E551D/R557H}$ (PUP2)
pCC16	[pET28 C-6xHisTag] TtCutA_tr3$^{E551D/N554E/R557H}$ (PUP3)
pCC17	[pET28M N-6xHis-SUMOTag] Cid1^{WT}
pCC18	[pET28M N-6xHis-SUMOTag] Cid1$^{D330E/E333N/H336R}$ (PCP)

SUPPLEMENTAL TABLE 3. Oligonucleotides used in this study.

Oligonucleotide	Sequence (5'-3')	Purpose
Oligonucleotides for cloning and site-directed mutagenesis		
278F	ttttgtttaactttaagaaggagatata**ccATGG**TCACCGTCACCGATGG	cloning of AnCutA WT ORF into pET28 C-6xHisTag
278R	atctcagtggtggtggtggtggtg**ctcgag**GCCACCTTTGCGCAGGCTTTC	cloning of AnCutA WT ORF into pET28 C-6xHisTag
280F	ttttgtttaactttaagaaggagatata**ccATGG**ACTCGCAACCGGCAAC	cloning of TtCutA WT ORF into pET28 C-6xHisTag
280R	atctcagtggtggtggtggtggtg**ctcgag**GCCACCTTTACGATCCGCTTC	cloning of TtCutA WT ORF into pET28 C-6xHisTag
CUCUmutF	CTCTGAC**agcgct**GTGGCCATTTGCATCACCACGACCTGTAAAGAACTGG	site-directed mutagenesis (AnCutA D203A D205A)
CUCUmutR	TGCAAATGGCCAC**agcgct**GTCAGAGCTACACAGCTTATTGCCCGATGAACC	site-directed mutagenesis (AnCutA D203A D205A)
278sumoF	gaagtctaccaggaacaaaccggt**ggatcc**AGCACCCATAGCGAATCTAC	cloning of AnCutA_tr into pET28M N-6xHis-SUMOTag
278sumo2	tcagtggtggtggtggtggtg**ctcgag***tta*ACGGGACGGCGGCGGTGCGG	cloning of AnCutA_tr into pET28M N-6xHis-SUMOTag
280sumoF	gaagtctaccaggaacaaaccggt**ggatcc**AGCGCCTATTCTGAAACCAAC	cloning of TtCutA_tr1 into pET28M N-6xHis-SUMOTag
280sumo2	tcagtggtggtggtggtggtg**ctcgag***tta*GCGCAGCAGGATCGGACGCG	cloning of TtCutA_tr1 into pET28M N-6xHis-SUMOTag
12.280 S_F	ttttgtttaactttaagaaggagatata***ccatg***AGCGCCTATTCTGAAACC	cloning of TtCutA_tr2 into pET28 C-6xHisTag
12.280 sI_R	atctcagtggtggtggtggtggtg**ctcgag**GCGACCATCATTAAACGGTTC	cloning of TtCutA_tr2 into pET28 C-6xHisTag
280sIIIF	ttttgtttaactttaagaaggagatata***ccatg***AAGCCGGCCAAAAATTC	cloning of TtCutA_tr3 into pET28 C-6xHisTag
280sIIIR	atctcagtggtggtggtggtggtg**ctcgag**CAGCGTGGTACGAGAAGCCG	cloning of TtCutA_tr3 into pET28 C-6xHisTag
280PAP1F	GTTTAATGATGGTGCCAA**cctagg**CAATACCGCCGACGAAACG	site-directed mutagenesis TtCutA WT→PAP1 (R557A)
280PAP1R	CGGCGGTATTG**cctagg**TTGGCACCATCATTAAACGGTTCTTC	site-directed mutagenesis TtCutA WT→PAP1 (R557A)
280PAP2F	GTTTAATGATGGTCAGAA**cctagg**CAATACCGCCGACGAAACG	site-directed mutagenesis TtCutA WT→PAP2 (R557Q)
280PAP2R	CGGCGGTATTG**cctagg**TTCTGACCATCATTAAACGGTTCTTC	site-directed mutagenesis TtCutA WT→PAP2 (R557Q)
280PAP3F	GCGTTGAAGACCCGTTTGAGGATGGTGCCAAcct***c***ggCAATACCGCCGAC	Site-directed mutagenesis TtCutA PAP1→PAP3 (E551D N554E R557A)
280PAP3R	ATTGcc***g***aggTTGGCACCATCCTCAAACGGGTCTTCAACGCACAGAGCATTG	site-directed mutagenesis TtCutA PAP1→PAP3 (E551D N554E R557A)
280PAP4F	TTGAAACACCGTTTAATGATGGTGCCAAc***agc***ggCAATACCGCCGACGAAAC	site-directed mutagenesis TtCutA PAP1→PAP4 (E551T R557A L559S)
280PAP4R	TTGcc***gct***gTTGGCACCATCATTAAACGGTGTTTCAACGCACAGAGCATTGTTC	site-directed mutagenesis TtCutA PAP1→PAP4 (E551T R557A L559S)
280PAP5F	TTTAATGATGGTAACAA**cctagg**CAGGACCGCCGACGAAACGAGCTTCC	site-directed mutagenesis TtCutA WT→PAP5 (R557N N561R)

280PAP5R	TTTCGTCGGCGGTCCTG**cctagg**TTGTTACCATCATTAAACGGTTCTTC	site-directed mutagenesis TtCutA WT→PAP5 (R557N N561R)
280PAP6F	TTTAATGATGGTAACAAcct***t***ggCAGGACCGCCGACGAAACGAGCTTCC	site-directed mutagenesis TtCutA PUP2→PAP6 (E551D R557N N561R)
280PAP6R	TTTCGTCGGCGGTCCTGcc***a***aggTTGTTACCATCATTAAACGGATCTTC	site-directed mutagenesis TtCutA PUP2→PAP6 (E551D R557N N561R)
280PUP1F	GTTTAATGATGGTCACAA**cctagg**CAATACCGCCGACGAAACG	site-directed mutagenesis TtCutA WT→PUP1 (R557H)
280PUP1R	GCGGTATTG**cctagg**TTGTGACCATCATTAAACGGTTCTTC	site-directed mutagenesis TtCutA WT→PUP1 (R557H)
280PUP2F	GCGTTGAAGATCCGTTTAATGATGGTCACAA**cctagg**CAATACCGCCGACGAAAC	site-directed mutagenesis TtCutA WT→PUP2 (E551D R557H)
280PUP2R	GCGGTATTG**cctagg**TTGTGACCATCATTAAACGGATCTTCAACGCACAGAGCATTG	site-directed mutagenesis TtCutA WT→PUP2 (E551D R557H)
280PUP3F	AGATCCGTTTGAAGATGGTCACAAcct***c***ggCAATACCGCCGACGAAACG	site-directed mutagenesis TtCutA PUP2→PUP3 (E551D N554E R557H)
280PUP3R	CGGTATTGcc***g***aggTTGTGACCATCTTCAAACGGATCTTCAACGCACA	site-directed mutagenesis TtCutA PUP2→PUP3 (E551D N554E R557H)
Cid1SumoF	tgattgaagtctaccaggaacaaaccggt**ggatcc**ATGAACATTTCTTCTGCACAA	cloning of Cid1 WT ORF into pET28M N-6xHis-SUMOTag
Cid1SumoR	cggatctcagtggtggtggtggtggtg**ctcgag***tca*CTCAGAATTGTCACCATCGG	cloning of Cid1 WT ORF into pET28M N-6xHis-SUMOTag
Cid1PCPF	GCGATTGAA**gagccc**TTCAATATTTCACGTAATGTGGGTAGGACAGTTAG	site-directed mutagenesis Cid1 WT→PCP (D330E E333N H336R)
Cid1PCPR	TACCCACATTACGTGAAATATTGAA**gggctc**TTCAATCGCAAGAATATAC	site-directed mutagenesis Cid1 WT→PCP (D330E E333N H336R)
Sequencing primers		
T7promoter	TAATACGACTCACTATAGGG	Sequencing of inserts in pET28 C-6xHisTag
SumoF	TCATACTGTCAAAGACAGGG	Sequencing of inserts in pET28M N-6xHis-SUMOTag
T7terminator	TGCTAGTTATTGCTCAGCGG	Sequencing of inserts in pET28 C-6xHisTag and pET28M N-6xHis-SUMOTag
278 seq 1	GAACAATACCATGGCCCTGG	Sequencing of AnCutA
278 seq 2	AGTAACCAGTATTCCCGTGG	Sequencing of AnCutA
12_107_s1	ACATTTGCATCACCACGACC	Sequencing of AnCutA
12_107_s2	AACACCATTCGCAACCTGGG	Sequencing of AnCutA
12_107_s3	CAATTCAACCTATCTGCCGG	Sequencing of AnCutA
12_107_s4	GCTGATTGTCGACGGTTCAG	Sequencing of AnCutA
280seq5	CAGCATTCTTCCAGTTTGCC	Sequencing of TtCutA
12_109_s1	GCAGCGCCTATTCTGAAACC	Sequencing of TtCutA
12_109_s2	TATCAAATACTGGACCCGCC	Sequencing of TtCutA
12_109_s3	CCGCGGTAACCGTCAGTTTC	Sequencing of TtCutA
12_109_s4	GAGCTTCATGCCGGATGAAC	Sequencing of TtCutA

RNA oligonucleotides		
5'-FAM-ss17-A_{35}	FAM-r(**CCCCACCACCAUCACUU**AAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAA)	RNA substrate used in biochemical assays
5'-FAM-ss17-$A_{31}U_4$	FAM-r(**CCCCACCACCAUCACUU**AAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAUUUU)	RNA substrate used in biochemical assays
5'-FAM-ss17-$A_{31}C_4$	FAM-r(**CCCCACCACCAUCACUU**AAAAAAAAAAAAAAAA AAAAAAAAAAAAAAACCCC)	RNA substrate used in biochemical assays
5'-FAM-ss22-A_4	FAM-r(**CGACUGGAGCACGAGGACACUG**AAAA)	RNA substrate used in biochemical assays
5'-FAM-ss22-U_4	FAM-r(**CGACUGGAGCACGAGGACACUG**UUUU)	RNA substrate used in biochemical assays
5'-FAM-ss22-C_4	FAM-r(**CGACUGGAGCACGAGGACACUG**CCCC)	RNA substrate used in biochemical assays
Oligonucleotides used in the 3'-RACE-seq procedure		
RA3	TGGAATTCTCGGGTGCCAAGG	Adapter oligonucleotide
RTP	*GCCTTGGCACCCGAGAATTCCA*	Reverse transcription primer
RA5_ss22	*GTTCAGAGTTCTACAGTCCGA*CGATC**CGACTGGAGCACGAGGA CACTG**	First-round PCR primer (forward)
RP1	AATGATACGGCGACCACCGAGATCTACAC*GTTCAGAGTTCTAC AGTCCGA*	Second-round PCR primer (forward)
RPI_X	CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTT *CCTTGGCACCCGAGAATTCCA*	PCR primer with RNA-seq index (reverse)

SUPPLEMENTAL TABLE 4. Summary of protein purification strategies.

CutA variant	Epitope	Purification scale	SUMO cleavage + desalting + second nickel column	Buffering agent	[NaCl]	Additional purification	Figure
An FL WT An FL DADA Tt FL WT	C-6xHisTag	Small	-	Na-phosphate pH=8.0	300 mM	-	Figs. 2-4
An_tr WT Tt_tr1 WT	N-6xHis-SUMOTag	Large	Yes	Na-phosphate pH=4.5	300 mM	Gel filtration (Superdex S200) 20 mM Na-phosphate; 150 mM NaCl	Fig. 5
Tt_tr3 WT	C-6xHisTag	Large	-	Na-phosphate pH=4.5	300 mM	Gel filtration (Superdex S200) 20 mM Na-phosphate; 150 mM NaCl	Fig. 5
An_tr WT Tt_tr1 WT	N-6xHis-SUMOTag	Small	No	Na-phosphate pH=4.5	300 mM	-	Fig. 6
An FL WT Tt FL WT Tt_tr3 WT	C-6xHisTag	Small	-	Na-phosphate pH=4.5	300 mM	-	Fig. 6
Tt_tr3 (WT, PAP1-PAP6, PUP1-PUP3)	C-6xHisTag	Large	-	Na-phosphate pH=4.5	300 mM	Gel filtration (S200) 20 mM Na-phosphate; 150 mM NaCl	Figs. 8-13
Cid1 (WT, PCP)	N-6xHis-SUMOTag	Large	Yes	Tris-HCl pH=8.0	150 mM	Gel filtration (S200) 20 mM Tris-HCl 150 mM NaCl	Fig. 14