**Nanobody Engineering –Based on Framework of Anti-RR6 (1QD0)**

**I. Construct and components**

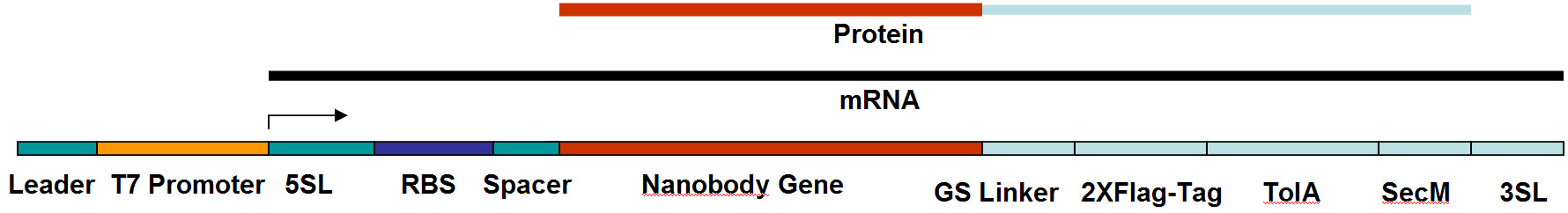
**II. Sequences of components**

**III. Residues to mutate**

**IV. Complete sequence of construct**

**V. Gene synthesis**

**I. Construct and Components**

****

**1. Leader Sequence (17 bases):**

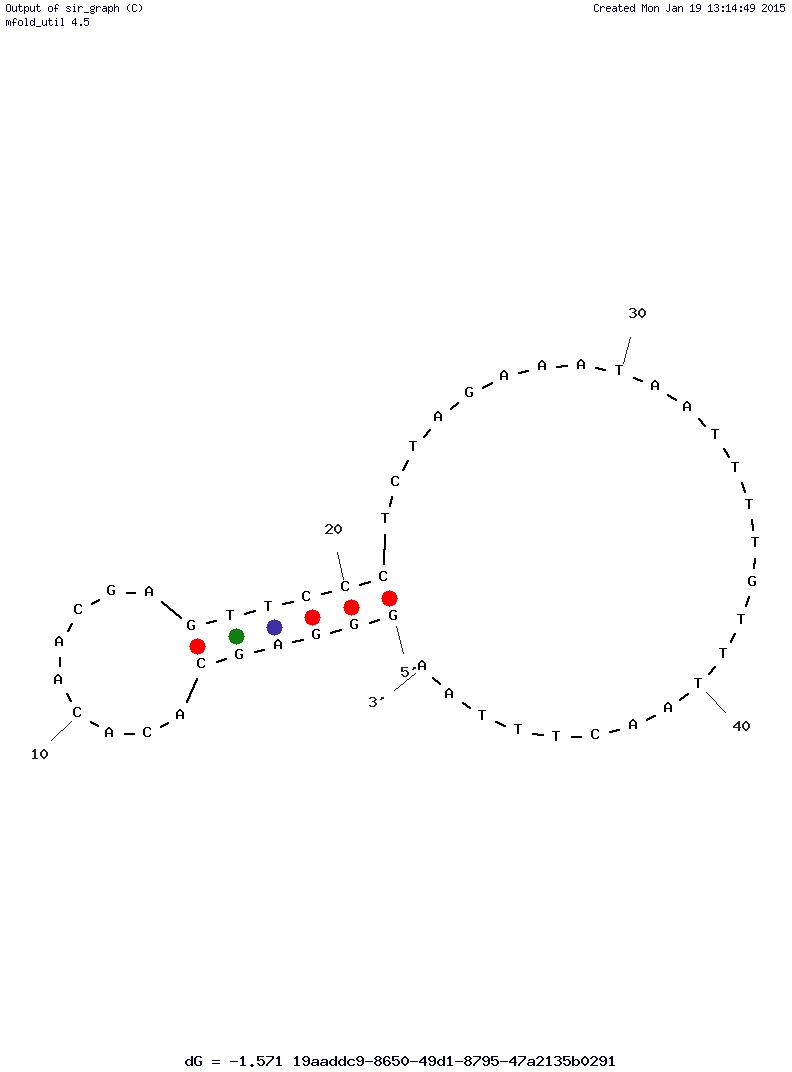
**5’-GTCTC GATCC CGCGA AA**

**2. T7 Promoter Sequence (17 bases):**

**5’-TAATA CGACT CACTA TA**

**3. 5’ Stemloop and Spacer Sequence: [not necessary actually, to prevent the degradation of the RNA as well as 3’]**

**5’-GGGAGCACACAACGAGTTCCCTCTAGAAATAATTTTGTTTAACTTTAA-3’**

****

**DNA: Tm = 39 °C (50 mM NaCl) /50 °C ( 10 mM Mg2+)**

**RNA: Tm = 78 °C (50 mM NaCl) /78 °C ( 10 mM Mg2+).**

**4. Ribosome Binding Site (RBS), 5-base Spacer and Start Codon:**

**5’-**GA**AGGAGAT** **ATACC** **ATG-3’**

**5. Nanobody Sequence (Anti-RR6, PDB ID: 2X6M)**

001 010 020

Q V Q L Q E S G G G L V Q A G G S L R L

CAG GTT CAG CTG CAA GAA AGC GGT GGT GGT CTG GTT CAG GCA GGC GGT AGC CTG CGT CTG

021 **027 028 029** 030 **032 033 034** **036** 040

S C A A S G **R** **A** **A** S G **H G H** Y **G** M G W F

AGC TGT GCA AGC AGC GGT CGT GCA GCA TCA GGT CAT GGT CAT TAT GGT ATG GGT TGG TTT

041 050  **055** **056**  060

R Q V P G K E R E F V A A I **R W** S G K E

CGT CAG GTT CCG GGT AAA GAA CGT GAA TTT GTT GCA GCA ATT CGT TGG AGC GGT AAA GAA

061 070 080

T W Y K D S V K G R F T I S R D N A K T

ACC TGG TAT AAA GAT AGC GTG AAA GGT CGT TTT ACC ATC AGC CGT GAT AAT GCA AAA ACC

081 090 100

T V Y L Q M N S L K G E D T A V Y Y C A

ACC GTT TAC CTG CAG ATG AAT AGT CTG AAA GGT GAA GAT ACG GCA GTG TAT TAT TGT GCA

101 **103 104 106 107** 110 120

A R **P V** R **V A** D ISLP V G F D Y W G Q

GCA CGT CCG GTT CGT GTT GCA GAT ATT AGC CTG CCG GTT GGT TTT GAT TAT TGG GGC CAG

121 128

G T Q V T V S S

GGG ACC CAG GTT ACC GTT AGC AGC

**6. (G4S)4 LINKER:**

G G G G S G G G G S G G G G S G G G G S

ggt ggt ggt ggt tct ggt ggt ggt ggt tct ggc ggc ggc ggc tcc agt ggt ggt gga tcc-

**7. 2x Flag Tag:**

D Y K D D D D K G G G S G G G S D Y K D

GAT TAC AAA GAC GAT GAT GAT AAA GGT GGT GGT AGC GGT GGT GGT TCA GAT TAT AAA GAT

D D D K G G G S

GAC GAC GAC AAA GGT GGC GGT TCA

**Note: 1X Flag-Tag (Sigma)**: -Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys – (OR –DYKDDDDK-)

**8. TolA Spacer (E. coli): [no structure, just makes it even longer]**

Sequence (Residues 131-214 of Tal A): QKQAEEAAAKAAADAKAKAEADAKAAEEAAKKAAADAKKKAEAEAAKAAAEAQKKAEAAAAALKKKAEAAEAAAAEARKKAATE

* Length: 84 residues
* D&E (Asp and Glu): 15 residues
* K (Lysine)/R (Arg): 18/1 residues
* Net charge: 3+.

**9. SecM Translation Stalling Sequence: [stalls translation without stop codon; stops the translation]**

Sequence (Residue 150 to 166): FSTPVWISQAQGIRAGP

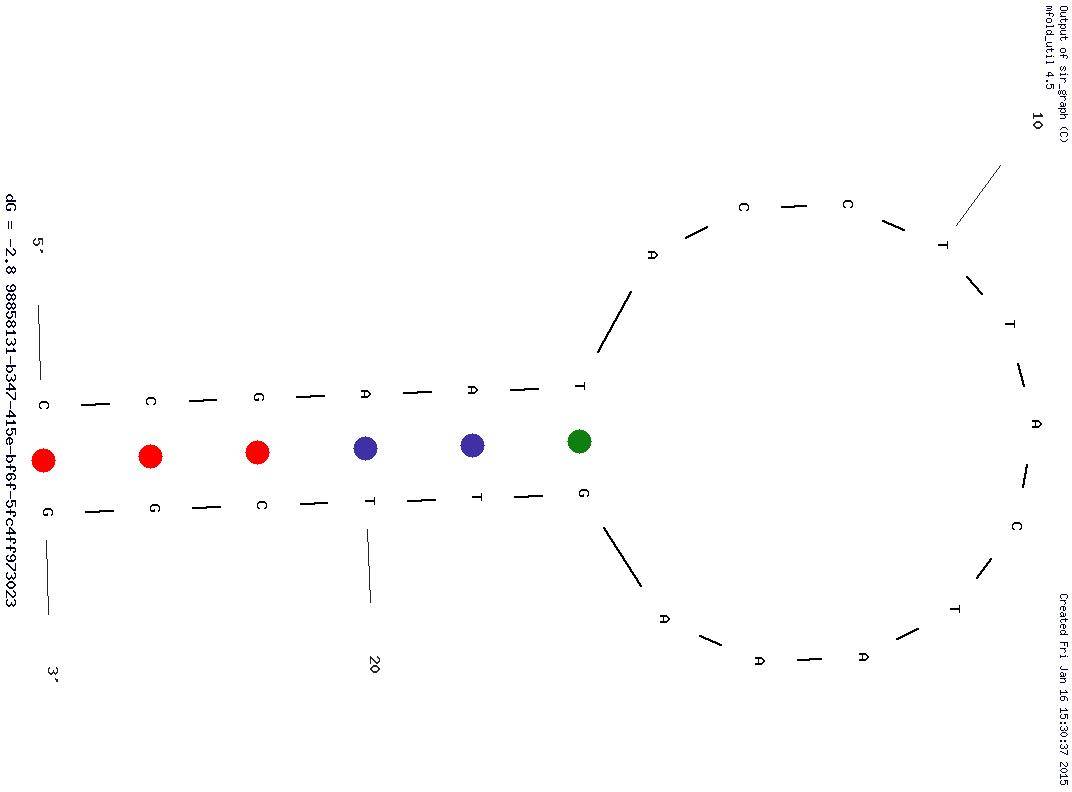
150-FSTPVWISQAQGIRAGP-166 + two rare codons (Arg)

F S T P V W I S Q A Q G I R A G P *R R*

TTT AGC ACA CCG GTT TGG ATT AGC CAG GCA CAG GGT ATT CGT GCA GGT CCG AGA CGA

**10. 3’ Spacer and Stemloop:**

**5’-CCGAAT ACCTTACTAAA GTTCGG**



**DNA: Tm = 44 °C (50 mM NaCl) /53 °C ( 10 mM Mg2+)**

**RNA: Tm = 70 °C (50 mM NaCl) /70 °C ( 10 mM Mg2+)**

**II. Sequences**

**Length of DNA: 99 + 384 + 60 + 84 + 252 + 57 + 24 + 23 = 983**

**Length of Polypeptide [Nb to TolA] 831/3 = 277 amino acid residues (128 + 20 + 28 + 84 + 17)**

**-------------------------------------------------------------------------------**

**GTCTC GATCC CGCGA AA**

**TAATA CGACT CACTA TA**

**GGGAGCACACAACGAGTTCCCTCTAGAAATAATTTTGTTTAACTTTAA**

GA**AGGAGGT** **ATACC** **ATG**

**CAG GTT CAG CTG CAA GAA AGC GGT GGT GGT CTG GTT CAG GCA GGC GGT AGC CTG CGT CTG**

**AGC TGT GCA AGC AGC GGT CGT GCA GCA TCA GGT CAT GGT CAT TAT GGT ATG GGT TGG TTT**

**CGT CAG GTT CCG GGT AAA GAA CGT GAA TTT GTT GCA GCA ATT CGT TGG AGC GGT AAA GAA**

**ACC TGG TAT AAA GAT AGC GTG AAA GGT CGT TTT ACC ATC AGC CGT GAT AAT GCA AAA ACC**

**ACC GTT TAC CTG CAG ATG AAT AGT CTG AAA GGT GAA GAT ACG GCA GTG TAT TAT TGT GCA**

**GCA CGT CCG GTT CGT GTT GCA GAT ATT AGC CTG CCG GTT GGT TTT GAT TAT TGG GGC CAG**

**GGG ACC CAG GTT ACC GTT AGC AGC**

**ggt ggt ggt ggt tct ggt ggt ggt ggt tct ggc ggc ggc ggc tcc agt ggt ggt gga tcc**

**GAT TAC AAA GAC GAT GAT GAT AAA GGT GGT GGT AGC GGT GGT GGT TCA GAT TAT AAA GAT**

**GAC GAC GAC AAA GGT GGC GGT TCA**

CAG AAA CAG GCA GAA **GAA** **GCA GCA GCA** AAA GCA GCC GCA GAT GCA AAA GCA AAA GCC GAA

GCC GAT GCC AAA GCA GCG GAA GAA GCC GCA AAA AAA GCG GCA GCG GAT GCG AAA AAA AAA

GCT GAA GCA GAA GCA GCC AAA GCC GCA GCA GAG GCA CAG AAA AAA GCC GAG GCA GCA GCG

GCA GCA CTG AAA AAA AAA GCA GAG GCT GCA GAA GCA GCT GCA GCA GAA GCC CGT AAA AAA

GCA GCA ACC GAA

**TTT AGC ACA CCG GTT TGG ATT AGC CAG GCA CAG GGT ATT CGT GCA GGT CCG**

**AGA CGA CTA CGA ATA CTA CGA AGA**

**CCGAAT ACCTTACTAAA GTTCGG**

-------------------------------------------------------------------------------

**III. Residues on Nanobody to Mutate**

|  |
| --- |
| (R) Arginine – 27 |
| (A) Alanine – 28 |
| (A) Alanine – 29 |
| (H) Histidine – 32 |
| (G) Glycine – 33 |
| (H) Histidine – 34 |
| (G) Glycine – 36 |
| (R) Arginine – 55 |
| (W) Tryptophan – 56 |
| (P) Proline – 103 |
| (V) Valine – 104 |

**Codon Table**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **2nd Base** | | | |  | |
| **T** | C | **A** | **G** |  | |
| **1st**  **base** | **T** | **Phe** | **Ser** | **Tyr** | **Cys** | **T**  C  **A**  **G** | **3rd**  **base** |
| **Phe** | **Ser** | **Tyr** | **Cys** |
| **Leu** | **Ser** | **Stop** | **Stop** |
| **Leu** | **Ser** | **Stop** | **Trp** |
| C | **Leu** | ***Pro*** | ***His*** | **Arg** | **T**  C  **A**  **G** |
| **Leu** | ***Pro*** | ***His*** | **Arg** |
| **Leu** | ***Pro*** | **Gln** | **Arg** |
| **Leu** | ***Pro*** | **Gln** | **Arg** |
| **A** | **Ile** | **Thr** | **Asn** | **Arg** | **T**  C  **A**  **G** |
| **Ile** | **Thr** | **Asn** | **Ser** |
| **Ile** | **Thr** | **Lys** | **Arg** |
| **Met** | **Thr** | **Lys** | **Arg** |
| **G** | **Val** | **Ala** | **Asp** | **Gly** | **T**  C  **A**  **G** |
| **Val** | **Ala** | **Asp** | **Gly** |
| **Val** | **Ala** | **Glu** | **Gly** |
| **Val** | **Ala** | **Glu** | **Gly** |

**A: Small polar; B: Large polar/charged; C: Small nonpolar; D: Large non-polar; E: Aromatic**

**Hydropilic: white background ; Hydrophobic: light blue background**

**Mixed bases:**

**D: T, A, and G**

**Z: A, T, and C**

**K: T and G**

**M: A and C**

**W: A and T**

**S: G and C**

**R: G and A**

**Mutations to Be Introduced:**

001 010 020

**Q V Q L Q E S G G G L V Q A G G S L R L**

**CAG GTT CAG CTG CAA GAA AGC GGT GGT GGT CTG GTT CAG GCA GGC GGT AGC CTG CGT CTG**

021 **027 028 029** 030 **032 033 034** **036** 040

**S C A A S G R A A S G H G H Y G M G W F**

**AGC TGT GCA AGC AGC GGT CGT GCA GCA TCA GGT CAT GGT CAT TAT GGT ATG GGT TGG TTT**

**AGC TGT GCA AGC AGC GGT DKT DMA DMA TCA GGT MWK DRT MWK TAT DKT ATG GGT TGG TTT**

041 050  **055** **056**  060

**R Q V P G K E R E F V A A I R W S G K E**

**CGT CAG GTT CCG GGT AAA GAA CGT GAA TTT GTT GCA GCA ATT CGT TGG AGC GGT AAA GAA**

**CGT CAG GTT CCG GGT AAA GAA CGT GAA TTT GTT GCA GCA ATT DKT DKK AGC GGT AAA GAA**

061 070 080

**T W Y K D S V K G R F T I S R D N A K T**

**ACC TGG TAT AAA GAT AGC GTG AAA GGT CGT TTT ACC ATC AGC CGT GAT AAT GCA AAA** ACC

081 090 100

**T V Y L Q M N S L K G E D T A V Y Y C A**

**ACC GTT TAC CTG CAG ATG AAT AGT CTG AAA GGT GAA GAT ACG GCA GTG TAT TAT TGT GCA**

101 **103 104 106 107** 110 120

**A R P V R V A D I S L P V G F D Y W G Q**

**GCA CGT CCG GTT CGT GTT GCA GAT ATT AGC CTG CCG GTT GGT TTT GAT TAT TGG GGC CAG**

**GCA CGT MMK DWT CGT DWT DMA GAT ATT AGC CTG CCG GTT GGT TTT GAT TAT TGG GGC CAG**

121 128

**G T Q V T V S S**

**GGG ACC CAG GTT ACC GTT AGC AGC**

**IV. Complete Sequence of Construct (both strands and amino acid sequence):**

**5’-GTCTCGATCC CGCGAAATAA TACGACTCAC TATAGGGAGC ACACAACGAG TTCCCTCTAG AAATAATTTT GTTTAACTTT**

**3’-cagagctagg gcgctttatt atgctgagtg atatccctcg tgtgttgctc aagggagatc tttattaaaa caaattgaaa**

**AA**GA**AGGAGG TATACCATG-3’**

**ttCTTCCTCC ATATGGTAC-5’**

001 010 020

**Q V Q L Q E S G G G L V Q A G G S L R L**

**CAG GTT CAG CTG CAA GAA AGC GGT GGT GGT CTG GTT CAG GCA GGC GGT AGC CTG CGT CTG**

**gtc caa gtc gac gtt ctt tcg cca cca cca gac caa gtc cgt ccg cca tcg gac gca gac**

021 **027 028 029** 030 **032 033 034** **036** 040

**S C A A S G R A A S G H G H Y G M G W F**

**AGC TGT GCA AGC AGC GGT DKT DMA DMA TCA GGT MWK DRT MWK TAT DKT ATG GGT TGG TTT**

**tcg aca cgt tcg tcg cca cga gct gct agt cca cat gga cat ata gga TAC CCA ACC AAA**

041 050  **055** **056**  060

**R Q V P G K E R E F V A A I R W S G K E**

**CGT CAG GTT CCG ggt aaa gaa cgt gaa ttt gtt gca gca att cga cgg agc ggt aaa gaa**

**GCA GTC CAA GGC CCA TTT CTT GCA CTT AAA CAA CGT CGT TAA DKT DKK ATC GCC ATT TCT**

061 070 080

**T W Y K D S V K G R F T I S R D N A K T**

**acc tgg tat aaa gat agc gtg aaa GGT CGT TTT ACC ATC AGC CGT GAT AAT GCA AAA** **ACC**

**TTG GAC CAT ATT TCT ATC GCA CTT TCC AGC AAA ATG GTA GTC GGC ACT att acg ttt TTG**

081 090 100

**T V Y L Q M N S L K G E D T A V Y Y C A**

**ACC GTT TAC CTG CAG ATG AAT AGT CTG AAA GGT GAA GAT ACG GCA GTG TAT TAT TGT GCA**

**gtg gca aat gga cgt cta ctt atc aga ctt tcc act tct** **tgc** **CGT CAC ATA ATA ACA CGT**

101 **103 104 106 107** 110 120

**A R P V R V A D I S L P V G F D Y W G Q**

**GCA CGT ccg gtt cgt gtt gca gat att agc ctg ccg gtt ggt ttt gat tat tgg ggc cag**

**CGT GCA MMK DWA GCA DWA DMT** **CTA TAA TCG GAC GGC CAA CCA AAA CTA ATA ACC CCG GTC**

121 128

**G T Q V T V S S**

**ggg acc cag gtt acc gtt agc agc**

**CCC TGG GTC CAA TGG CAA TCG TCG**

001 010 020

**G Q L V E S G G G S V Q A G G S L R L S**

**GGT CAG CTG gtt gaa agc ggt ggt ggt agc gtt cag gca ggc ggt agc ctg cgt ctg agc**

**CCA GTC GAC CAA CTT TCG CCA CCA CCA TCG CAA GTC CGT CCG CCA TCG GAC GCA GAC TCG**

021 030 040

**C A A S G I D S S S Y C M G W F R Q R P**

**tgt gca gca agc ggt att gat AGC AGC AGC TAT TGT ATG GGT TGG TTT CGT CAG CGT CCG**

**ACA CGT CGT TCG CCA TAA CTA TCG TCG TCG ATA ACA TAC CCA ACC aaa gca gtc gca ggc**

041 050 **051** **057** **058** 060

**G K E R E G V A R I N G L G G V K T A Y**

**GGT AAA GAA CGT GAA GGT GTT GCA CGT ATT DMT GGT CTG GGT GGT DWT DMK ACC GCA TAT**

**cca ttt ctt gca ctt cca caa cgt gca taa tta cca gac cca cca caa ttt TGG CGT ATA**

061 070 080

**A D S V K D R F T I S R D N A E N T V Y**

**GCA GAT AGC GTT AAA gat cgc ttt acc att agc cgt gat aat gcc gaa aat acc gtt tac**

**CGT CTA TCG CAA TTT CTA GCG AAA TGG TAA TCG GCA CTA TTA CGG CTT TTA TGG CAA ATG**

081 090 **100**

**L Q M N S L K P E D T A I Y Y C A A K F**

**ctg cag atg aat agc ctg aaa ccg gaa gat ACC GCA ATC TAT TAT TGT GCA GCC AAA DWT**

**GAC GTC TAC TTA TCG GAC TTT GGC CTT CTA TGG CGT TAG ATA ATA ACA CGT CGG ttt aaa**

**101** **103 104** **108 109 110 111**  120

**S P G Y C G G S W S N F G Y W G Q G T Q**

**DMT CCG GGT DWT DST GGT GGT DMT DRK DMT DMT TTT GGT TAT TGG GGT CAG GGC ACC CAG**

**tca ggc cca ata aca cca cca tca acc aga tta aaa cca ata acc cca gtc ccg TGG GTC**

121 126

**V T V S S H**

**GTT ACC GTT AGC AGC CAT-3’**

**CAA TGG CAA TCG TCG GTA-5’**

**G G G G S G G G G S G G G G S G G G G S**

**ggt ggt ggt ggt tct ggt ggt ggt ggt tct ggc ggc ggc ggc tcc agt ggt ggt gga tcc**

**CCA CCA CCA CCA AGA CCA CCA CCA CCA AGA CCG CCG CCG CCG AGG TCA CCA CCA CCT AGG**

**D Y K D D D D K G G G S G G G S D Y K D**

**gat tac aaa gac gat GAT GAT AAA GGT GGT GGT AGC GGC GGT GGT TCA GAT TAT AAA GAT**

**CTA ATG TTT CTG CTA CTA CTA TTT CCA CCA CCA TCG CCG cca cca agt cta ata ttt cta**

**D D D K G G G S**

**GAC GAC GAC AAA GGT GGC GGT TCA**

**ctg ctg ctg ttt cca ccg cca agt**

Q K Q A E E A A A K A A A D A K A K A E

**CAG AAA CAG GCA GAA GAA GCA GCA GCA AAA GCA GCC GCA GAT GCA AAA GCA AAA GCC GAA**

**gtc ttt gtc cgt ctt ctt cgt cgt cgt ttt CGT CGG CGT CTA CGT TTT CGT TTT CGG CTT**

A D A K A A E E A A K K A A A D A K K K

**gcc gat gcc aaa gca gcg gaa gaa gcc gca aaa aaa gcg gca gcg gat gcg aaa aaa aaa**

**CGG CTA CGG TTT CGT CGC CTT CTT CGG CGT TTT TTT CGC CGT CGC CTA CGC TTT TTT TTT**

A E A E A A K A A A E A Q K K A E A A A

**gct gaa gca GAA GCA GCC AAA GCC GCA GCA GAG GCA CAG AAA AAA GCC GAG GCA GCA GCG**

**CGA CTT CGT CTT CGT CGG TTT CGG CGT CGT CTC cgt gtc ttt ttt cgg ctc cgt cgt cgc**

A A L K K K A E A A E A A A A E A R K K

**GCA GCA CTG AAA AAA AAA GCA GAG GCT GCA GAA GCA GCT GCA GCA GAA GCC CGT AAA AAA**

**cgt cgt gac ttt ttt ttt cgt ctc cga cgt ctt cgt cga cgt cgt ctt CGG GCA TTT TTT**

A A T E

**GCA GCA ACC GAA**

**CGT CGT TGG CTT**

**F S T P V W I S Q A Q G I R A G P *R R L***

**ttt agc aca ccg gtt tgg att agc cag gca cag ggt att cgt gca ggt ccg aga cga cta**

**AAA TCG TGT GGC CAA ACC TAA TCG GTC CGT GTC CCA TAA GCA CGT CCA GGC TCT GCT GAT**

**R I L R R**

**CGA ATA CTA CGA AGA CCGAAT ACCTTACTAAA GTTCGG**

**GCT TAT GAT GCT TCT GGCTTA TGGAATGATTT CAAGCC**

**V. Gene Synthesis**

**Oligos to synthesize**

**NbPRBS1: (108 bases)**

**5’-GTCTCGATCC CGCGAAATAA TACGACTCAC TATAGGGAGC ACACAACGAG TTCCCTCTAG AAATAATTTT GTTTAACTTT**

**AA**GA**AGGAGG TATACCATG GGT CAG CTG-3’**

**NbCDS1: (122 bases)**

**3’-CTTCCTCC ATATGGTAC**

**CCA GTC GAC CAA CTT TCG CCA CCA CCA TCG CAA GTC CGT CCG CCA TCG GAC GCA GAC TCG**

**ACA CGT CGT TCG CCA TAA CTA TCG TCG TCG ATA ACA TAC CCA ACC-5’**

**5’-CCAACCCATACAATAGCTGCTGCTATCAATACCGCTTGCTGCACA**

**GCTCAGACGCAGGCTACCGCCTGCCTGAACGCTACCACCACCGCTTTCAACCAGCTGACC**

**CATGGTATACCTCCTTC-3’**

**NbCDS2: (114 bases)**

**5’- AGC AGC AGC TAT TGT ATG GGT TGG TTT CGT CAG CGT CCG**

**GGT AAA GAA CGT GAA GGT GTT GCA CGT ATT DMT GGT CTG GGT GGT DWT DMK ACC GCA TAT**

**GCA GAT AGC GTT AAA-3'**

**NbCDS3: (123 bases)**

**3'-TGG CGT ATA**

**CGT CTA TCG CAA TTT CTA GCG AAA TGG TAA TCG GCA CTA TTA CGG CTT TTA TGG CAA ATG**

**GAC GTC TAC TTA TCG GAC TTT GGC CTT CTA TGG CGT TAG ATA ATA ACA CGT CGG-5'**

**5’-GGCTGCACAATAATAGATTGCGGTATCTTCCGGTTTCAGGCTATTCATCTGCAG**

**GTAAACGGTATTTTCGGCATTATCACGGCTAATGGTAAAGCGATCTTTAACGCTATCTGC**

**ATATGCGGT-3’**

**NbCDS4: (108 bases)**

**5'-ACC GCA ATC TAT TAT TGT GCA GCC AAA DWT**

**DMT CCG GGT DWT DST GGT GGT DMT DRK DMT DMT TTT GGT TAT TGG GGT CAG GGC ACC CAG**

**GTT ACC GTT AGC AGC CAT-3'**

**NbCDS5: (123 bases)**

**3'-TGG GTC CAA TGG CAA TCG TCG GTA**

**CCA CCA CCA CCA AGA CCA CCA CCA CCA AGA CCG CCG CCG CCG AGG TCA CCA CCA CCT AGG**

**CTA ATG TTT CTG CTA CTA CTA TTT CCA CCA CCA TCG CCG-5’**

**5’-GCCGCTACCACCACCTTTATCATCATCGTCTTTGTAATC**

**GGATCCACCACCACTGGAGCCGCCGCCGCCAGAACCACCACCACCAGAACCACCACCACC**

**ATGGCTGCTAACGGTAACCTGGGT-3’**

**NbCDS6: (123 bases)**

**5’-GAT GAT AAA GGT GGT GGT AGC GGC GGT GGT TCA GAT TAT AAA GAT**

**GAC GAC GAC AAA GGT GGC GGT TCA**

**CAG AAA CAG GCA GAA GAA GCA GCA GCA AAA GCA GCC GCA GAT GCA AAA GCA AAA-3’**

**NbCDS7: (123 bases)**

**3’-CGT CGG CGT CTA CGT TTT CGT TTT CGG CTT**

**CGG CTA CGG TTT CGT CGC CTT CTT CGG CGT TTT TTT CGC CGT CGC CTA CGC TTT TTT TTT**

**CGA CTT CGT CTT CGT CGG TTT CGG CGT CGT CTC-5’**

**5’-CTCTGCTGCGGCTTTGGCTGCTTCTGCTTCAGC**

**TTTTTTTTTCGCATCCGCTGCCGCTTTTTTTGCGGCTTCTTCCGCTGCTTTGGCATCGGC**

**TTCGGCTTTTGCTTTTGCATCTGCGGCTGC-3’**

**NbCDS8: (123 bases)**

**5’-GAA GCA GCC AAA GCC GCA GCA GAG GCA CAG AAA AAA GCC GAG GCA GCA GCG**

**GCA GCA CTG AAA AAA AAA GCA GAG GCT GCA GAA GCA GCT GCA GCA GAA GCC CGT AAA AAA**

**GCA GCA ACC GAA-3’**

**NbCDS9: (122 bases)**

**3’-CGG GCA TTT TTT CGT CGT TGG CTT**

**AAA TCG TGT GGC CAA ACC TAA TCG GTC CGT GTC CCA TAA GCA CGT CCA GGC TCT GCT GAT**

**GCT TAT GAT GCT TCT GGCTTA TGGAATGATTT CAAGCC-5’**

**5’-CCGAACTTTAGTAAGGTATTCGGTCTTCGTAGTATTCG**

**TAGTCGTCTCGGACCTGCACGAATACCCTGTGCCTGGCTAATCCAAACCGGTGTGCTAAA**

**TTCGGTTGCTGCTTTTTTACGGGC-3’**

**Gene Synthesis (Replicate and Ligate)**

**5’-GTCTCGATCC CGCGAAATAA TACGACTCAC TATAGGGAGC ACACAACGAG TTCCCTCTAG AAATAATTTT GTTT-3’**

*3’-cagagctagg gcgctttatt atgctgagtg atatccctcg tgtgttgctc aagggagatc tttattaaaa caaa-5’*

**AACTTT AA** GA**AGGAGG TATACCATG GGT CAG CTG-->3’***gtt gaa agc ggt ggt ggt agc gtt cag gca*

*TTGAAA TT* **3’<--CTTCCTCC ATATGGTAC CCA GTC GAC CAA CTT TCG CCA CCA CCA TCG CAA GTC CGT**

*ggc ggt agc ctg cgt ctg agc*

**CCG CCA TCG GAC GCA GAC TCG-5’**

*tgt gca gca agc ggt att gat* **AGC AGC AGC TAT TGT ATG GGT TGG TTT CGT CAG CGT CCG-3’**

**3’-ACA CGT CGT TCG CCA TAA CTA TCG TCG TCG ATA ACA TAC CCA ACC** *aaa gca gtc gca ggc****-*5'**

**5’-GGT AAA GAA CGT GAA GGT GTT GCA CGT ATT DMT GGT CTG GGT GGT DWT DMK-3’**

*3’-cca ttt ctt gca ctt cca caa cgt gca taa xxa cca gac cca cca xxa xxx-5’*

**5’- ACC GCA TAT GCA GAT AGC GTT AAA-->***gat cgc ttt acc att agc cgt gat aat gcc gaa aat-***3'**

**3'<--TGG CGT ATA CGT CTA TCG CAA TTT CTA GCG AAA TGG TAA TCG GCA CTA TTA CGG CTT TTA-5’**

**5'***-acc gtt tac ctg cag atg aat agc ctg aaa ccg gaa gat* **ACC GCA ATC TAT TAT TGT GCA GCC-3’**

**3’-TGG CAA ATG GAC GTC TAC TTA TCG GAC TTT GGC CTT CTA TGG CGT TAG ATA ATA ACA CGT CGG-5'**

**5’-AAA DWT DMT CCG GGT DWT DST GGT GGT DMT DRK DMT DMT TTT GGT TAT TGG GGT CAG GGC-3’**

*3’-ttt xxa xxa ggc cca xxa xxa cca cca xxa xxx xxa xxa aaa cca ata acc cca gtc ccg-5’*

**5’-ACC CAG GTT ACC GTT AGC AGC CAT-->3'**

**3'<--TGG GTC CAA TGG CAA TCG TCG GTA-5’**

*5’-ggt ggt ggt ggt tct ggt ggt ggt ggt tct ggc ggc ggc ggc tcc agt ggt ggt gga tcc-3’*

**3’-CCA CCA CCA CCA AGA CCA CCA CCA CCA AGA CCG CCG CCG CCG AGG TCA CCA CCA CCT AGG-5’**

*5’-gat tac aaa gac gat* **GAT GAT AAA GGT GGT GGT AGC GGC-3’**

**3’-CTA ATG TTT CTG CTA CTA CTA TTT CCA CCA CCA TCG CCG-5’**

**5’-GGT GGT TCA GAT TAT AAA GAT GAC GAC GAC AAA GGT GGC GGT TCA-3’**

*3’-CCA CCA AGT CTA ATA TTT CTA CTG CTG CTG TTT CCA CCG CCA AGT-5’*

**5’-CAG AAA CAG GCA GAA GAA GCA GCA GCA AAA GCA GCC GCA GAT GCA AAA GCA AAA-->3’**

*3’-GTC TTT GTC CGT CTT CTT CGT CGT CGT TTT* **3’<-CGT CGG CGT CTA CGT TTT CGT TTT-5’**

*5’-GCC GAA GCC GAT GCC AAA GCA GCG GAA GAA GCC GCA AAA AAA GCG GCA GCG GAT*

**3’-CGG CTT CGG CTA CGG TTT CGT CGC CTT CTT CGG CGT TTT TTT CGC CGT CGC CTA-5’**

5’-*GCG AAA AAA AAA GCT GAA GCA* **GAA GCA GCC AAA GCC GCA GCA GAG-3’**

**CGC TTT TTT TTT CGA CTT CGT CTT CGT CGG TTT CGG CGT CGT CTC-5’**

**5’-GCA CAG AAA AAA GCC GAG GCA GCA GCG**

*CGT GTC TTT TTT CGG CTC CGT CGT CGC*

**5’-GCA GCA CTG AAA AAA AAA GCA GAG GCT GCA GAA GCA GCT GCA GCA GAA**

*3’-CGT CGT GAC TTT TTT TTT CGT CTC CGA CGT CTT CGT CGA CGT CGT CTT*

**5’- GCC CGT AAA AAA GCA GCA ACC GAA-->3’**

**3’<--CGG GCA TTT TTT CGT CGT TGG CTT**

*5’-TTT AGC ACA CCG GTT TGG ATT AGC CAG GCA CAG GGT ATT CGT GCA GGT CCG AGA CGA CTA-3’*

**3’-AAA TCG TGT GGC CAA ACC TAA TCG GTC CGT GTC CCA TAA GCA CGT CCA GGC TCT GCT GAT**

*CGA ATA CTA CGA AGA CCGAAT ACCTTACTAAA GTTCGG-3’*

**GCT TAT GAT GCT TCT GGCTTA TGGAATGATTT CAAGCC-5’**

**Overlapping regions Tm (100 nM, 50 mM NaCl, 5 mM MaCl2):**

**1. NbPRBS1-NbCSD1: 67 °C**

**2. NbCSD1-NbCSD2: 67 °C**

**3. NbCSD2-NbCSD3: 65 °C**

**4. NbCSD3-NbCSD4: 66 °C**

**5. NbCSD4-NbCSD5: 69 °C**

**6. NbCSD5-NbCSD6: 67 °C**

**7. NbCSD6-NbCSD7: 69 °C**

**8. NbCSD7-NbCSD8: 71 °C**

**9. NbCSD8-NbCSD9: 68 °C**

**NbPRFW:**

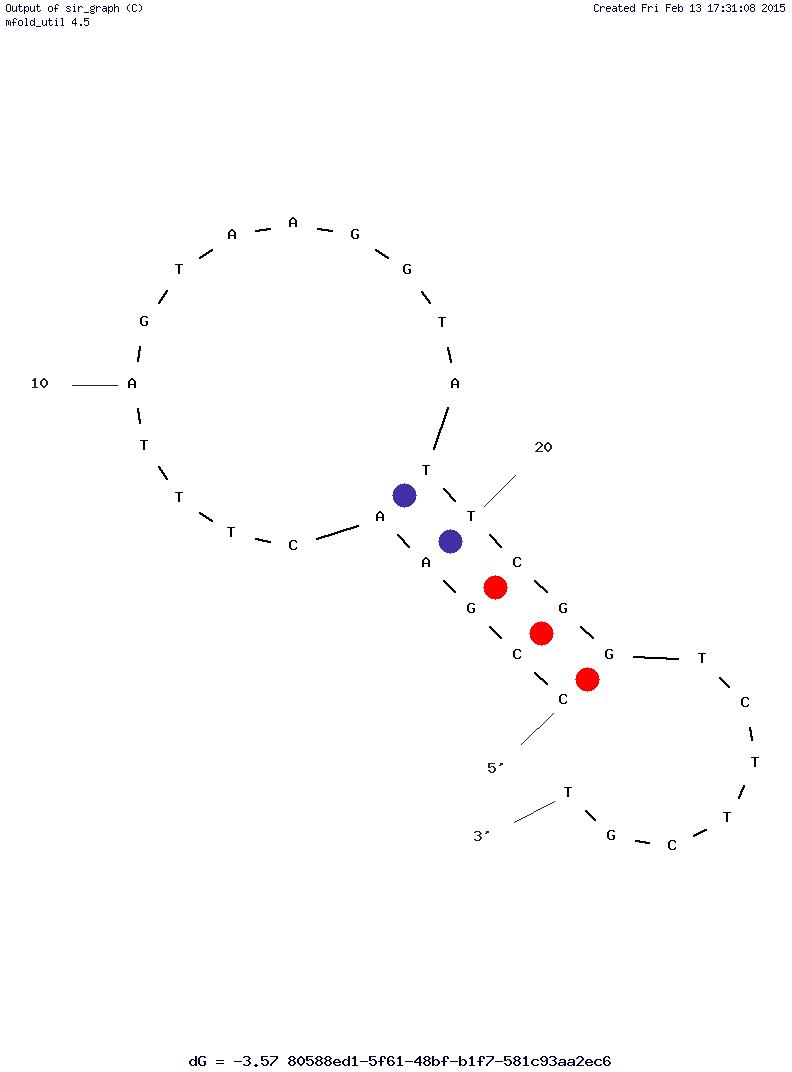
5'- GTC TCG ATC CCG CGA AAT AAT ACG ACT CAC -3'

**NbPRRV:**

5’-CCG AAC TTT AGT AAG GTA TTC GGT CTT CGT-3’

**DNA: Tm = 51.4 °C( 1 uM oligo, 50 mM NaCl, 2 mM Mg2+)/ 53 °C(50 mM NaCl, 6 mM Mg2+)**

**Secondary structure (Hairpin):**

****

**DNA: Tm = 51.4 °C( 50 mM NaCl, 2 mM Mg2+)/ 53 °C(50 mM NaCl, 6 mM Mg2+)**

Complete amino acid sequence of protein:

**G Q L V E S G G G S V Q A G G S L R L S**

**C A A S G I D S S S Y C M G W F R Q R P**

**G K E R E G V A R I N G L G G V K T A Y**

**A D S V K D R F T I S R D N A E N T V Y**

**L Q M N S L K P E D T A I Y Y C A A K F**

**S P G Y C G G S W S N F G Y W G Q G T Q**

**V T V S S H**

**G G G G S G G G G S G G G G S G G G G S**

**D Y K D D D D K G G G S G G G S D Y K D**

**D D D K G G G S**

Q K Q A E E A A A K A A A D A K A K A E

A D A K A A E E A A K K A A A D A K K K

A E A E A A K A A A E A Q K K A E A A A

A A L K K K A E A A E A A A A E A R K K

A A T E

**F S T P V W I S Q A Q G I R A G P *R R L***

**R I L R R.**

# ProtParam

## User-provided sequence:

10 20 30 40 50 60   
MGQLVESGGG SVQAGGSLRL SCAASGIDSS SYCMGWFRQR PGKEREGVAR INGLGGVKTA   
  
 70 80 90 100 110 120   
YADSVKDRFT ISRDNAENTV YLQMNSLKPE DTAIYYCAAK FSPGYCGGSW SNFGYWGQGT   
  
 130 140 150 160 170 180   
QVTVSSHGGG GSGGGGSGGG GSGGGGSDYK DDDDKGGGSG GGSDYKDDDD KGGGSQKQAE   
  
 190 200 210 220 230 240   
EAAAKAAADA KAKAEADAKA AEEAAKKAAA DAKKKAEAEA AKAAAEAQKK AEAAAAALKK   
  
 250 260 270 280   
KAEAAEAAAA EARKKAATEF STPVWISQAQ GIRAGPRRLR ILRR

[References](http://web.expasy.org/tools/protparam/protpar-ref.html) and [documentation](http://web.expasy.org/tools/protparam/protparam-doc.html) are available.

http://web.expasy.org/images/others/new_flash_anim.gifPlease note the [modified algorithm for extinction coefficient](http://web.expasy.org/tools/protparam/protparam-doc.html).

**Number of amino acids:** 284

**Molecular weight:** 28712.4

**Theoretical pI:** 8.85

Top of Form

**Amino acid composition:**   
Ala (A) 56 19.7%

Arg (R) 14 4.9%

Asn (N) 5 1.8%

Asp (D) 18 6.3%

Cys (C) 4 1.4%

Gln (Q) 11 3.9%

Glu (E) 18 6.3%

Gly (G) 46 16.2%

His (H) 1 0.4%

Ile (I) 7 2.5%

Leu (L) 9 3.2%

Lys (K) 27 9.5%

Met (M) 3 1.1%

Phe (F) 5 1.8%

Pro (P) 5 1.8%

Ser (S) 25 8.8%

Thr (T) 8 2.8%

Trp (W) 4 1.4%

Tyr (Y) 9 3.2%

Val (V) 9 3.2%

Pyl (O) 0 0.0%

Sec (U) 0 0.0%

(B) 0 0.0%

(Z) 0 0.0%

(X) 0 0.0%

Bottom of Form

**Total number of negatively charged residues (Asp + Glu):** 36

**Total number of positively charged residues (Arg + Lys):** 41

**Atomic composition:**

Carbon C 1219

Hydrogen H 1939

Nitrogen N 375

Oxygen O 415

Sulfur S 7

**Formula:** C1219H1939N375O415S7

**Total number of atoms:** 3955

**Extinction coefficients:**

Extinction coefficients are in units of M-1 cm-1, at 280 nm measured in water.

Ext. coefficient 35660

Abs 0.1% (=1 g/l) 1.242, assuming all pairs of Cys residues form cystines

Ext. coefficient 35410

Abs 0.1% (=1 g/l) 1.233, assuming all Cys residues are reduced

**Estimated half-life:**

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).

>20 hours (yeast, in vivo).

>10 hours (Escherichia coli, in vivo).

**Instability index:**

The instability index (II) is computed to be 36.77

This classifies the protein as stable.

**Aliphatic index:** 50.88

**Grand average of hydropathicity (GRAVY):** -0.658