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

**I. Construct and components**

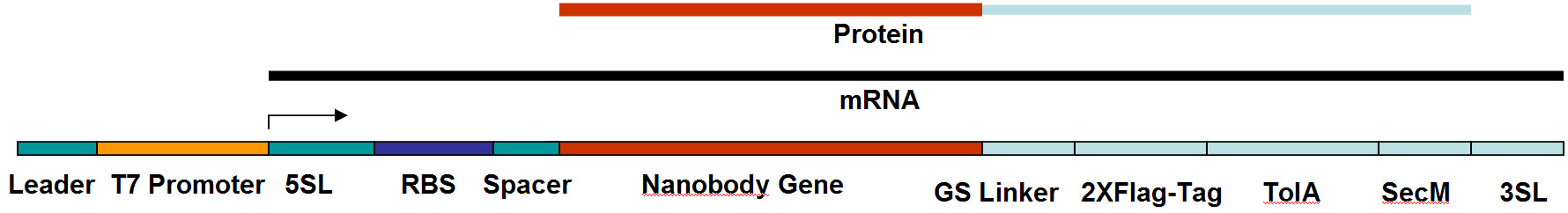
**II. Sequences of components**

**III. Residues to mutate**

**IV. Complete sequence of construct**

**V. Gene synthesis**

**I. Construct and Components**

****

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

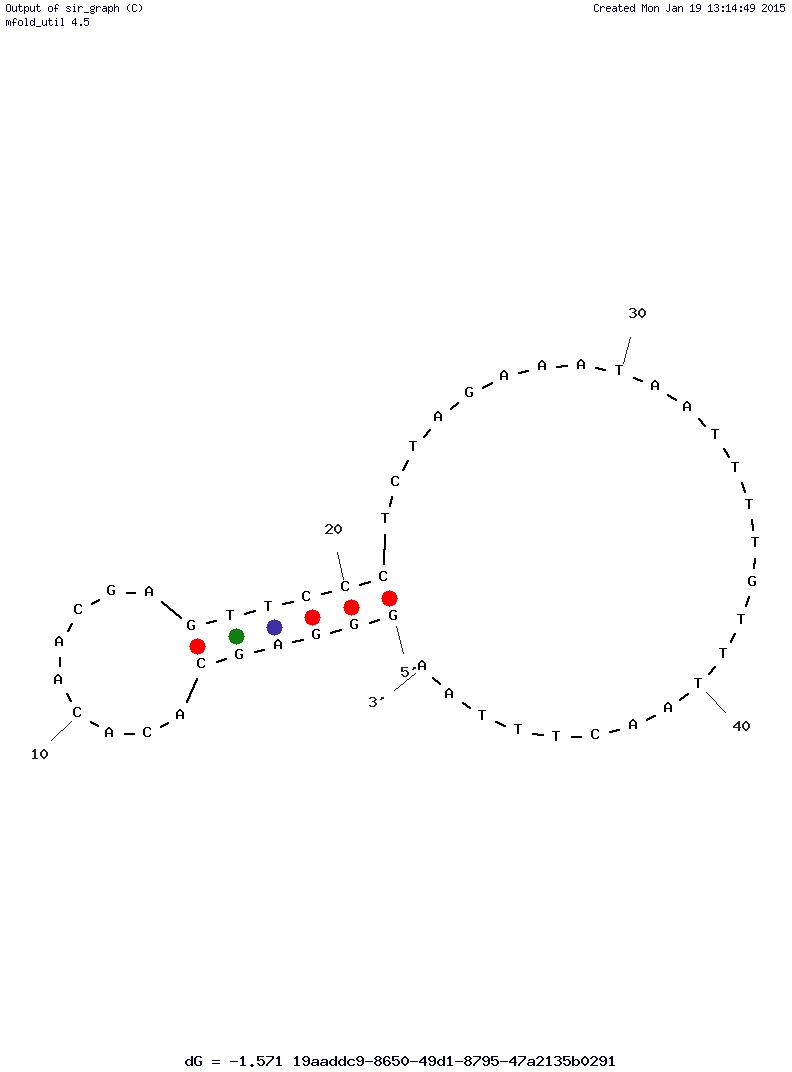
**5’-GTCTC GATCC CGCGA AA**

**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

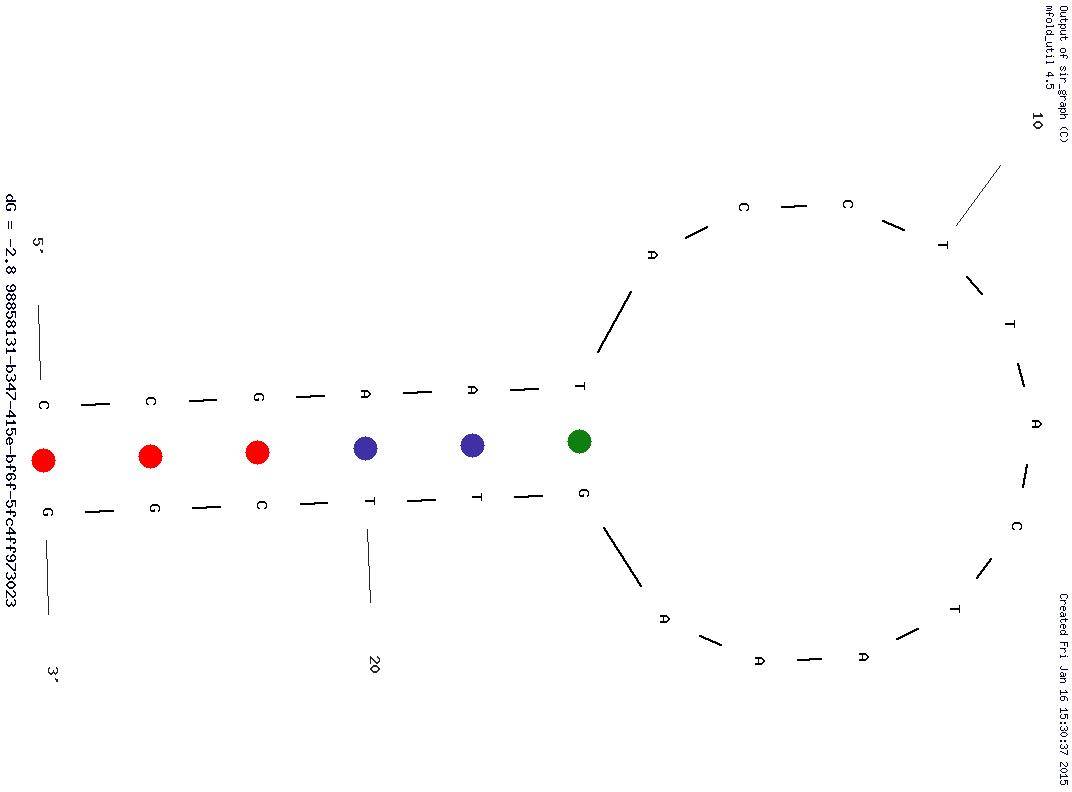
150-FSTPVWISQAQGIRAGP-166 + two rare codons (Arg)

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):**

Nt. BspQI Nt.BsmAI

**5’- ATCG GCCT AAGC GTCTCA GCTCTTCA GTCTCGATCC CGCGAAATAA TACGACTCAC TATAGGGAGC ACACAACGAG TTCCCTCTAG AAATAATTTT GTTTAACTTT**

**3’- TAGC CGGA TTCG CAGAGT CGAGAAGT 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**

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

**CGAGAAGT 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 CGAGAC ATCG CTAG GGCT AGAAGAGC**

**CCC TGG GTC CAA TGG CAA TCG TCG GCTCTG TAGC GATC CCGA TCTTCTCG**

001 010 020

**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 AGAAGAGC CGAGAC ATCG CTAG GGCT**

**GCT TAT GAT GCT TCT GGCTTA TGGAATGATTT CAAGCC TCTTCTCG GCTCTG TAGC GATC CCGA**

**V.A Gene Synthesis (4 piece assembly)**

**\**See accompanying file Pocket\_Nb\_gene\_ass.ppt for PCR modifications***

**Oligos to synthesize**

**NbPRBS1: 148 bases (Post PCR)**

**5’- ATCG GCCT AAGC GAGTC GCTCTTCA GTCTCGATCC CGCGAAATAA TACGACTCAC TATAGGGAGC ACACAACGAG TTCCCTCTAG AAATAATTTT GTTTAACTTT**

**AA**GA**AGGAGG TATACCATG CAG GTT CAG CTG CAA GAA AGC GGT GGT GGT CTG G -3’**

**NbCDS1: (132 bases)**

**5’ -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 DKT DMA DMA TCA GGT MWK DRT MWK TAT DKT ATG GGT TGG TTT**

**CGT CAG GTT CCG -3’**

**NbCDS2: (120 bases)**

**3’ TAC CCA ACC AAA**

**GCA GTC CAA GGC CCA TTT CTT GCA CTT AAA CAA CGT CGT TAA ZMA ZMM TCG CCA TTT CTT**

**TGG ACC ATA TTT CTA TCG CAC TTT CCA GCA AAA TGG TAG TCG GCA CTA-5’**

**Synthesize:**

**5’ ATC ACG GCT GAT**

**GGT AAA ACG ACC TTT CAC GCT ATC TTT ATA CCA GGT TTC TTT ACC GCT MMZ AMZ AAT TGC**

**TGC AAC AAA TTC ACG TTC TTT ACC CGG AAC CTG ACG AAA CCA ACC CAT -3’**

**NbCDS3: (102 bases)**

**5’ -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 -3’**

**NbCDS4: (102 bases) + NE Site (8 bases) = 110 bases**

**3’- CGT CAC ATA ATA ACA CGT**

**CGT GCA KKM ZWA GCA ZWA ZKT** **CTA TAA TCG GAC GGC CAA CCA AAA CTA ATA ACC CCG GTC**

**CCC TGG GTC CAA TGG CAA TCG TCG NCTTCTCG -5’**

**Synthesize:**

**5’- GCTCTTCN GCT GCT AAC GGT AAC CTG GGT CCC CTG GCC CCA ATA ATC AAA ACC AAC CGG**

**CAG GCT AAT ATC TKZ** **AWZ ACG AWZ MKK ACG TGC TGC ACA ATA ATA CAC TGC -3’**

**NbCDS5: (123 bases) – Already have; will be modified via PCR**

**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’**

**NbCDS5: (129 bases)(Post PCR)**

**3’-CCG GTC CCC TGG GTC CAA TGG CAA TCG TCG 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’**

**Reverse:**

**5’- GCC GCT ACC ACC ACC TTT ATC ATC ATC GTC TTT GTA ATC GGA TCC ACC ACC ACT GGA GCC GCC GCC GCC AGA ACC ACC ACC ACC AGA ACC ACC ACC ACC GCT GCT AAC GGT AAC CTG GGT CCC CTG GCC’-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: (128 bases) (Post PCR)**

**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 GGC TCTTCTCG GCTCTG TAGC GATC CCGA -5’**

**Reverse:**

**5’- AGCC CTAG CGAT GTCTCG GCTCTTCT CGG TCT TCG TAG TAT TCG TAG TCG TCT CGG ACC TGC ACG AAT ACC CTG TGC CTG GCT AAT CCA AAC CGG TGT GCT AAA TTC GGT TGC TGC TTT TTT ACG GGC -3’**

**Gene Synthesis (Replicate and Ligate)**

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

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

**5’ - AACTTT AA** GA**AGGAGG TATACCATG CAG GTT CAG CTG CAA GAA AGC GGT GGT CTG GTT CAG GCA GGC GGT AGC CTG**

*3’ - ttgaaa tt* **CTTCCTCC ATATGGTAC CAG GTT CAG CTG CAA GAA AGC** *cca cca gac caa gtc cgt ccg cca tcg gac*

**5’ - CGT CTG AGC TGT GCA AGC AGC GGT DKT DMA DMA TCA GGT MWK DRT MWK TAT DKT ATG GGT TGG TTT CGT CAG**

*3’ - gca gac tcg aca cgt tcg tcg cca xxa xxt xxt agt cca xxx xxa xxx ata xxa* **3’<-- TAC CCA ACC AAA GCA GTC**

**5’ - GTT CCG -->3’***ggt aaa gaa cgt gaa ttt gtt gca gca att xxt xxx agc ggt aaa gaa acc tgg tat aaa gat agc*

**3’ - CAA GGC CCA TTT CTT GCA CTT AAA CAA CGT CGT TAA ZMA ZMM TCG CCA TTT CTT TGG ACC ATA TTT CTA TCG**

*5’ - 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**

**3’ - CAC TTT CCA GCA AAA TGG TAG TCG GCA CTA** *tta cgt ttt tgg tgg caa atg gac gtc tac tta tca gac ttt cca*

**5’ - GAA GAT ACG GCA GTG TAT TAT TGT GCA GCA CGT -->3’** *ttt ttt cgt ttt ttt gat att agc ctg ccg gtt ggt*

**3’ –** *ctt cta tgc* **3’<-- CGT CAC ATA ATA ACA CGT CGT GCA KKM ZWA GCA ZWA ZKT** **CTA TAA TCG GAC GGC CAA CCA**

**5’ - TTT GAT TAT TGG GGC CAG GGG ACC CAG GTT ACC GTT AGC AGC NGAAGAGC** *ggt ggt ggt ggt tct ggt ggt ggt ggt*

**3’ - AAA CTA ATA ACC CCG GTC CCC TGG GTC CAA TGG CAA TCG TCG ^NCTTCTCG CCA CCA CCA CCA AGA CCA CCA CCA CCA**

*5’ - tct ggc ggc ggc ggc tcc agt ggt ggt gga tcc gat tac aaa gac gat* **GAT GAT AAA GGT GGT GGT AGC GGC**

**3’ - 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’-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’**

**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’**

**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-NbCDS1: 69.3 °C**

**2. NbCSD1-NbCSD2: 68.7 °C**

**3. NbCSD2-NbCSD3: 67.0 °C**

**4. NbCSD3-NbCSD4: 66.1 °C**

**5. NbCSD4-NbCSD5: 66.3 °C**

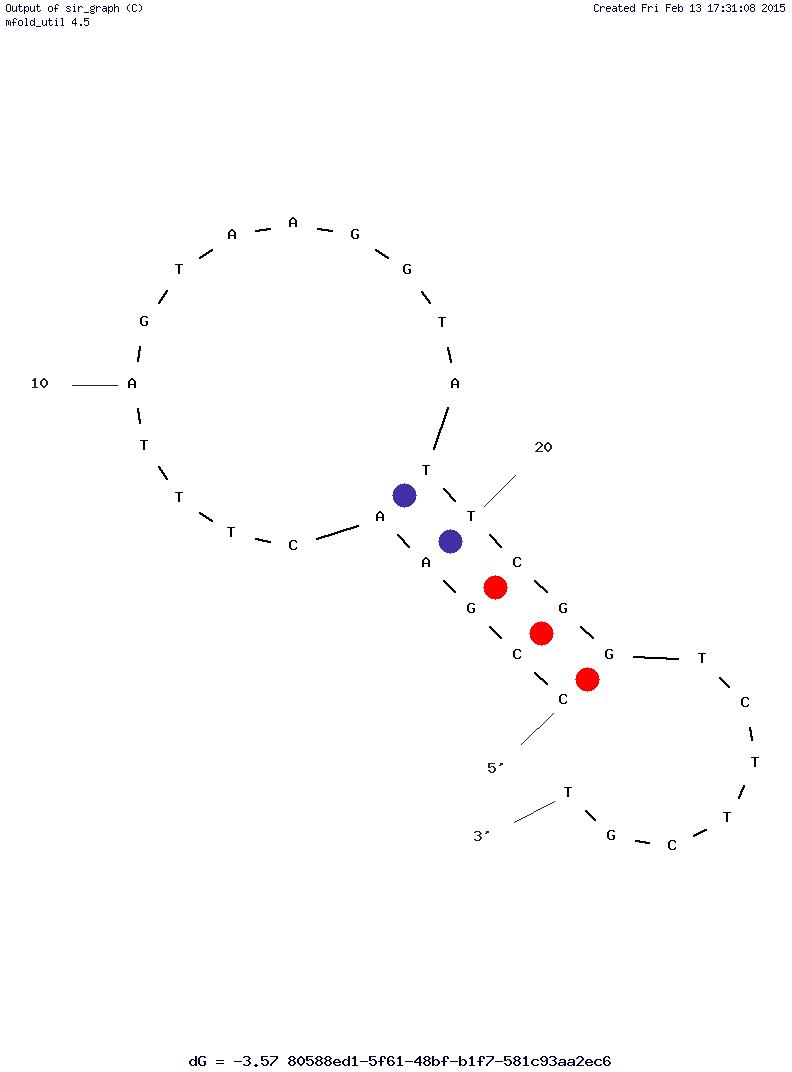
**6. NbCSD5-NbCSD6: 66.8 °C**

**7. NbCSD6-NbCSD7: 69.4 °C**

**8. NbCSD7-NbCSD8: 71.3 °C**

**9. NbCSD8-NbCSD9: 68.5 °C**

**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:

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

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

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

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

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

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

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

**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   
QVQLQESGGG LVQAGGSLRL SCAASGRAAS GHGHYGMGWF RQVPGKEREF VAAIRWSGKE   
  
 70 80 90 100 110 120   
TWYKDSVKGR FTISRDNAKT TVYLQMNSLK GEDTAVYYCA ARPVRVADIS LPVGFDYWGQ   
  
 130 140 150 160 170 180   
GTQVTVSSGG GGSGGGGSGG GGSGGGGSDY KDDDDKGGGS GGGSDYKDDD DKGGGSQKQA   
  
 190 200 210 220 230 240   
EEAAAKAAAD AKAKAEADAK AAEEAAKKAA ADAKKKAEAE AAKAAAEAQK KAEAAAAALK   
  
 250 260 270 280   
KKAEAAEAAA AEARKKAATE FSTPVWISQA QGIRAGPRRL RILRR

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

**Number of amino acids:** 285

**Molecular weight:** 29067.0

**Theoretical pI:** 9.27

Top of Form

**Amino acid composition:**   
Ala (A) 58 20.4%

Arg (R) 16 5.6%

Asn (N) 2 0.7%

Asp (D) 18 6.3%

Cys (C) 2 0.7%

Gln (Q) 13 4.6%

Glu (E) 18 6.3%

Gly (G) 44 15.4%

His (H) 2 0.7%

Ile (I) 6 2.1%

Leu (L) 10 3.5%

Lys (K) 28 9.8%

Met (M) 2 0.7%

Phe (F) 5 1.8%

Pro (P) 5 1.8%

Ser (S) 21 7.4%

Thr (T) 9 3.2%

Trp (W) 5 1.8%

Tyr (Y) 8 2.8%

Val (V) 13 4.6%

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):** 44

**Atomic composition:**

Carbon C 1246

Hydrogen H 1989

Nitrogen N 385

Oxygen O 411

Sulfur S 4

**Formula:** C1246H1989N385O411S4

**Total number of atoms:** 4035

**Extinction coefficients:**

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

Ext. coefficient 39545

Abs 0.1% (=1 g/l) 1.360, assuming all pairs of Cys residues form cystines

Ext. coefficient 39420

Abs 0.1% (=1 g/l) 1.356, assuming all Cys residues are reduced

**Estimated half-life:**

The N-terminal of the sequence considered is Q (Gln).

The estimated half-life is: 0.8 hours (mammalian reticulocytes, in vitro).

10 min (yeast, in vivo).

10 hours (Escherichia coli, in vivo).

**Instability index:**

The instability index (II) is computed to be 37.97

This classifies the protein as stable.

**Aliphatic index:** 55.47

**Grand average of hydropathicity (GRAVY):** -0.642