

## Supplementary Note

pET11a-link-NGFP, 6181 bp

pET11a-link-NGFP is a plasmid vector designed for fusion of a target protein to the N-terminal fragment of GFP sg100 (residues 1-157). The construct contains a linker for the convenient subcloning of the target between XhoI and BamHI. There is an XmaI site between those two sites, as well. The protein has a hexahistidine tag fused to its N-terminus. With the linker, the protein sequence is:

```
MASHHHHHHGASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYG
KLTLKFICTTGKLPVPWPTLVTTLCYGVQCFSRYPDHMKRHDFFKSAMPE
GYVQERTIFFKDDGNYKTRAEVKFEGLTLVNRIELKGIDFKEDGNILGHK
LEYNYNSHNVPIMADKQGGSGSGSSNPGN*
```

The sequence GSGSGSS is intended as a spacer between the NGFP fragment and the target protein. The target must be cloned in frame for proper fusion.

```

                XhoI      XmaI      BamHI
...caaggtggcctctggcctctggCTCGAGcaatcccgggaattaaGGATCCattatt
...Q G G S G S G S S N P G N *
```

The first residue of the target protein should be inserted at the first AAT (Asn) codon above. TAA is a stop codon.

The NGFP fusion is under the control of the T7 promoter. Therefore, expression is only possible in a strain expressing T7 RNA Polymerase. We typically use BL21(DE3), but you could lysogenize any strain with lambdaDE3. For use with pMRBAD-link-CGFP, we recommend growth on LB agar supplemented with 100 ug/mL ampicillin, 35 ug/mL kanamycin, 10 uM IPTG and 0.02-0.2% arabinose. Grow the cells overnight at 30C and transfer them to room temperature for 2 days.

Plasmid features (approx)

```
1 - 543 NGFP+linker
1054 - 1914 ampicillin resistance
2144 - 2544 ColE1 origin
3008 - 2817 rop
5708 - 4626 lacI
```

Thomas J. Magliery, Ph.D., July 20, 2004

### References:

Magliery TJ, Wilson CGM, Pan W, Mishler D, Ghosh I, Hamilton AD, Regan L. 2004. Detecting protein-protein interactions with a GFP-fragment reassembly trap: scope and mechanism. *J Am Chem Soc*, submitted.

Ghosh I, Hamilton AD, Regan L. 2000. Antiparallel leucine zipper-directed protein reassembly: application to the green fluorescent protein. *J Am Chem Soc* 122: 5658.

References can be downloaded from:

<http://www.csb.yale.edu/people/regan/publications.html>

```
1 ATATGGCTAG CCATCACCAC CATCACCATG GCGCGAGCAA AGGAGAAGAA
51 CTCTTCACTG GAGTTGTCCC AATTCTTGTT GAATTAGATG GTGATGTTAA
101 CGGCCACAAG TTCTCTGTCA GTGGAGAGGG TGAAGGTGAT GCAACATACG
151 GAAAACCTAC CCTGAAGTTC ATCTGCACTA CTGGCAAACCT GCCTGTTCCA
201 TGGCCAACAC TAGTCACTAC TCTGTGCTAT GGTGTTCAAT GCTTTTCAAG
251 ATACCCGGAT CATATGAAAC GGCATGACTT TTTCAAGAGT GCCATGCCCG
301 AAGGTTATGT ACAGGAAAGG ACCATCTTCT TCAAAGATGA CGGCAACTAC
351 AAGACACGTG CTGAAGTCAA GTTTGAAGGT GATACCCTTG TTAATAGAAT
401 CGAGTTAAAA GGTATTGACT TCAAGGAAGA TGGCAACATT CTGGGACACA
451 AATTGGAATA CAACTATAAC TCACACAACG TTCCCATCAT GGCAGACAAA
501 CAAGGTGGCT CTGGCTCTGG CTCGAGCAAT CCCGGGAATT AAGGATCCGG
551 CTGCTAACAA AGCCCGAAAG GAAGCTGAGT TGGCTGCTGC CACCGCTGAG
601 CAATAACTAG CATAACCCCT TGGGGCCTCT AAACGGGTCT TGAGGGGTTT
651 TTTGCTGAAA GGAGGAACTA TATCCGGATA TCCCGCAAGA GGCCCGGCAG
701 TACCGGCATA ACCAAGCCTA TGCCTACAGC ATCCAGGGTG ACGGTGCCGA
751 GGATGACGAT GAGCGCATTG TTAGATTTCA TACACGGTGC CTGACTGCGT
801 TAGCAATTTA ACTGTGATAA ACTACCGCAT TAAAGCTTAT CGATGATAAG
851 CTGTCAAACA TGAGAATTCT TGAAGACGAA AGGGCCTCGT GATACGCCTA
901 TTTTATAGG TTAATGTCAT GATAATAATG GTTTCTTAGA CGTCAGGTGG
951 CACTTTTCGG GGAAATGTGC GCGGAACCCC TATTTGTTTA TTTTCTAAA
1001 TACATTCAA TATGTATCCG CTCATGAGAC AATAACCCTG ATAAATGCTT
1051 CAATAATATT GAAAAAGGAA GAGTATGAGT ATTCAACATT TCCGTGTCGC
1101 CTTATTCCC TTTTTGCGG CATTTTGCCT TCCTGTTTTT GCTACCCAG
1151 AAACGCTGGT GAAAGTAAAA GATGCTGAAG ATCAGTTGGG TGCACGAGTG
1201 GGTTACATCG AACTGGATCT CAACAGCGGT AAGATCCTTG AGAGTTTTCG
1251 CCCCGAAGAA CGTTTTCCAA TGATGAGCAC TTTTAAAGTT CTGCTATGTG
1301 GCGCGGTATT ATCCCGTGTT GACGCCGGGC AAGAGCAACT CGGTGCGCCG
1351 ATACACTATT CTCAGAATGA CTTGGTTGAG TACTCACCAG TCACAGAAAA
1401 GCATCTTACG GATGGCATGA CAGTAAGAGA ATTATGCAGT GCTGCCATAA
1451 CCATGAGTGA TAACACTGCG GCCAACTTAC TTCTGACAAC GATCGGAGGA
1501 CCGAAGGAGC TAACCGCTTT TTTGCACAAC ATGGGGGATC ATGTAACCTG
1551 CTTGATCGT TGGGAACCGG AGCTGAATGA AGCCATACCA AACGACGAGC
1601 GTGACACCAC GATGCCTGCA GCAATGGCAA CAACGTTGCG CAAACTATTA
1651 ACTGGCGAAC TACTTACTCT AGCTTCCCGG CAACAATTAA TAGACTGGAT
1701 GGAGGCGGAT AAAGTTGCAG GACCACTTCT GCGCTCGGCC CTTCCGGCTG
1751 GCTGGTTTAT TGCTGATAAA TCTGGAGCCG GTGAGCGTGG GTCTCGCGGT
1801 ATCATTGCAG CACTGGGGCC AGATGGTAAG CCCTCCCGTA TCGTAGTTAT
1851 CTACACGACG GGGAGTCAGG CAACTATGGA TGAACGAAAT AGACAGATCG
1901 CTGAGATAGG TGCCTCACTG ATTAAGCATT GGTAACGTGTC AGACCAAGTT
1951 TACTCATATA TACTTTAGAT TGATTTAAAA CTTCATTTTT AATTTAAAAG
2001 GATCTAGGTG AAGATCCTTT TTGATAATCT CATGACCAA ATCCCTTAAC
2051 GTGAGTTTTT GTTCCACTGA GCGTCAGACC CCGTAGAAAA GATCAAAGGA
2101 TCTTCTTGAG ATCCTTTTTT TCTGCGCGTA ATCTGCTGCT TGCAAACAAA
2151 AAAACCACCG CTACCAGCGG TGGTTTGTTT GCCGGATCAA GAGCTACCAA
2201 CTCTTTTTCC GAAGGTAACT GGCTTCAGCA GAGCGCAGAT ACCAAATACT
2251 GTCCTTCTAG TGTAGCCGTA GTTAGGCCAC CACTTCAAGA ACTCTGTAGC
2301 ACCGCCTACA TACCTCGCTC TGCTAATCCT GTTACCAGTG GCTGCTGCCA
2351 GTGGCGATAA GTCGTGTCTT ACCGGGTTGG ACTCAAGACG ATAGTTACCG
2401 GATAAGGCGC AGCGGTCGGG CTGAACGGGG GGTTCGTGCA CACAGCCAG
2451 CTTGGAGCGA ACGACCTACA CCGAACTGAG ATACCTACAG CGTGAGCTAT
2501 GAGAAAGCGC CACGCTTCCC GAAGGGAGAA AGGCGGACAG GTATCCGGTA
```

2551 AGCGGCAGGG TCGGAACAGG AGAGCGCACG AGGGAGCTTC CAGGGGGAAA  
2601 CGCCTGGTAT CTTTATAGTC CTGTCCGGTT TCGCCACCTC TGACTTGAGC  
2651 GTCGATTTTT GTGATGCTCG TCAGGGGGGC GGAGCCTATG GAAAAACGCC  
2701 AGCAACGCGG CCTTTTTACG GTTCCTGGCC TTTTGCTGGC CTTTTGCTCA  
2751 CATGTTCTTT CCTGCGTTAT CCCCTGATTC TGTGGATAAC CGTATTACCG  
2801 CCTTTGAGTG AGCTGATACC GCTCGCCGCA GCCGAACGAC CGAGCGCAGC  
2851 GAGTCAGTGA GCGAGGAAGC GGAAGAGCGC CTGATGCGGT ATTTTCTCCT  
2901 TACGCATCTG TGCGGTATTT CACACCGCAT ATATGGTGCA CTCTCAGTAC  
2951 AATCTGCTCT GATGCCGCAT AGTTAAGCCA GTATACACTC CGCTATCGCT  
3001 ACGTGACTGG GTCATGGCTG CGCCCCGACA CCCGCCAACA CCCGCTGACG  
3051 CGCCCTGAGG GGCTTGTCTG CTCCCGGCAT CCGCTTACAG ACAAGCTGTG  
3101 ACCGTACTCCG GGAGCTGCAT GTGTCAGAGG TTTTCACCGT CATCACCGBAA  
3151 ACGCGCGAGG CAGCTGCGGT AAAGCTCATC AGCGTGGTCG TGAAGCGATT  
3201 CACAGATGTC TGCCTGTTCA TCCGCGTCCA GCTCGTTGAG TTTCTCCAGA  
3251 AGCGTTAATG TCTGGCTTCT GATAAAGCGG GCCATGTTAA GGGCGGTTTT  
3301 TTCCTGTTTG GTCACTGATG CCTCCGTGTA AGGGGGATTT CTGTTCATGG  
3351 GGGTAATGAT ACCGATGAAA CGAGAGAGGA TGCTCACGAT ACGGGTTACT  
3401 GATGATGAAC ATGCCCGGTT ACTGGAACGT TGTGAGGGTA AACAACTGGC  
3451 GGTATGGATG CGGCGGGACC AGAGAAAAAT CACTCAGGGT CAATGCCAGC  
3501 GCTTCGTTAA TACAGATGTA GGTGTTCCAC AGGGTAGCCA GCAGCATCCT  
3551 GCGATGCAGA TCCGGAACAT AATGGTGCAG GCGGCTGACT TCCGCGTTTT  
3601 CAGACTTTAC GAAACACGGA AACC GAAGAC CATT CATGTT GTTGCTCAGG  
3651 TCGCAGACGT TTTGCAGCAG CAGTCGCTTC ACGTTGCTC GCATATCGGT  
3701 GATTCATTCT GCTAACCAGT AAGGCAACCC CGCCAGCCTA GCCGGGTCTT  
3751 CAACGACAGG AGCACGATCA TGCGCACCCG TGGCCAGGAC CCAACGCTGC  
3801 CCGAGATGCG CCGCGTGCGG CTGCTGGAGA TGGCGGACGC GATGGATATG  
3851 TTCTGCCAAG GGTGGGTTTG CGCATT CACA GTTCTCCGCA AGAATTGATT  
3901 GGCTCCAATT CTTGGAGTGG TGAATCCGTT AGCGAGGTGC CGCCGGCTTC  
3951 CATT CAGGTC GAGTGGCCC GGCTCCATGC ACCGCGACGC AACCGGGGA  
4001 GGCAGACAAG GATATAGGCG CGCCTACAA TCCATGCCAA CCCGTTCCAT  
4051 GTGCTCGCCG AGGCGGCATA AATCGCCGTG ACGATCAGCG GTCCAGTGAT  
4101 CGAAGTTAGG CTGGTAAGAG CCGCGAGCGA TCCTTGAAGC TGTCCCTGAT  
4151 GGTCGTCATC TACCTGCCTG GACAGCATGG CCTGCAACGC GGGCATCCCG  
4201 ATGCCGCCGG AAGCGAGAAG AATCATAATG GGGAAGGCCA TCCAGCCTCG  
4251 CGTCGCGAAC GCCAGCAAGA CGTAGCCCAG CGCGTCGGCC GCCATGCCGG  
4301 CGATAATGGC CTGCTTCTCG CCGAAACGTT TGGTGGCGGG ACCAGTGACG  
4351 AAGGCTTGAG CGAGGGCGTG CAAGATTCCG AATACCGCAA GCGACAGGCC  
4401 GATCATCGTC GCGCTCCAGC GAAAGCGGTC CTCGCCGAAA ATGACCCAGA  
4451 GCGCTGCCGG CACCTGTCTT ACGAGTTGCA TGATAAAGAA GACAGTCATA  
4501 AGTGCGGCGA CGATAGTCAT GCCCCGCGCC CACCGGAAGG AGCTGACTGG  
4551 GTTGAAGGCT CTCAAGGGCA TCGGTGCGAGA TCCCGGTGCC TAATGAGTGA  
4601 GCTAACTTAC ATTAATTGCG TTGCGCTCAC TGCCCGCTTT CCAGTCGGGA  
4651 AACCTGTCGT GCCAGCTGCA TTAATGAATC GGCCAACGCG CGGGGAGAGG  
4701 CGGTTTGCGT ATTGGGCGCC AGGGTGGTTT TTCTTTTAC CAGTGAGACG  
4751 GGCAACAGCT GATTGCCCTT CACCGCCTGG CCCTGAGAGA GTTGCAGCAA  
4801 GCGGTCCACG CTGGTTTGCC CCAGCAGGCG AAAATCCTGT TTGATGGTGG  
4851 TTAACGGCGG GATATAACAT GAGCTGTCTT CGGTATCGTC GTATCCCACT  
4901 ACCGAGATAT CCGACCAAC GCGCAGCCCG GACTCGGTAA TGGCGCGCAT  
4951 TGCGCCGAGC GCCATCTGAT CGTTGGCAAC CAGCATCGCA GTGGGAACGA  
5001 TGCCCTCATT CAGCATTGTC ATGGTTTGTT GAAAACCGGA CATGGCACTC  
5051 CAGTCGCCTT CCCGTTCCGC TATCGGCTGA ATTTGATTGC GAGTGAGATA  
5101 TTTATGCCAG CCAGCCAGAC GCAGACGCGC CGAGACAGAA CTTAATGGGC  
5151 CCGCTAACAG CGCGATTTGC TGGTGACCCA ATGCGACCAG ATGCTCCACG  
5201 CCCAGTCGCG TACCGTCTTC ATGGGAGAAA ATAATACTGT TGATGGGTGT  
5251 CTGGTCAGAG ACATCAAGAA ATAACGCCGG AACATTAGTG CAGGCAGCTT  
5301 CCACAGCAAT GGCATCCTGG TCATCCAGCG GATAGTTAAT GATCAGCCCA  
5351 CTGACGCGTT GCGCGAGAAG ATTTGTGACC GCCGCTTTAC AGGCTTCGAC

```

5401 GCCGCTTCGT TCTACCATCG ACACCACCAC GCTGGCACCC AGTTGATCGG
5451 CGCGAGATTT AATCGCCGCG ACAATTTGCG ACGGCGCGTG CAGGGCCAGA
5501 CTGGAGGTGG CAACGCCAAT CAGCAACGAC TGTTTGCCCG CCAGTTGTTG
5551 TGCCACGCGG TTGGGAATGT AATTCAGCTC CGCCATCGCC GCTTCCACTT
5601 TTTCCCGCGT TTTCGCAGAA ACGTGGCTGG CCTGGTTCAC CACGCGGGAA
5651 ACGGTCTGAT AAGAGACACC GGCATACTCT GCGACATCGT ATAACGTTAC
5701 TGGTTTCACA TTCACCACCC TGAATTGACT CTCTTCCGGG CGCTATCATG
5751 CCATACCGCG AAAGGTTTTG CGCCATTCGA TGGTGTCCGG GATCTCGACG
5801 CTCTCCCTTA TGCGACTCCT GCATTAGGAA GCAGCCCAGT AGTAGGTTGA
5851 GGCCGTTGAG CACCGCCGCC GCAAGGAATG GTGCATGCAA GGAGATGGCG
5901 CCCAACAGTC CCCC GCCAC GGGGCCTGCC ACCATACCCA CGCCGAAACA
5951 AGCGCTCATG AGCCCGAAGT GCGGAGCCCG ATCTTCCCCA TCGGTGATGT
6001 CGGCGATATA GGCGCCAGCA ACCGCACCTG TGGCGCCGGT GATGCCGGCC
6051 ACGATGCGTC CGGCGTAGAG GATCGAGATC TCGATCCCGC GAAATTAATA
6101 CGACTCACTA TAGGGGAATT GTGAGCGGAT AACAATTCCC CTCTAGAAAT
6151 AATTTTGT TT AACTTTAAGA AGGAGATATA C

```

pMRBAD-link-CGFP, 4778 bp

pMRBAD-link-GFP is a plasmid vector designed for fusion of a target protein to the C-terminal fragment of GFP sg100 (residues 158-238). The construct contains a linker for the convenient subcloning of the target between NcoI and AatII. There is an SphI site between those two sites, as well. With the linker, the protein sequence is:

```

MANACNGTSGGSGKNGIKVNFKTRHNIEDGSVQLADHYQQNTPIGDGPVL
LPDNHYLSTQSALS KDPNEKRDMVLLLEFVTAAGITHGMDELYN*

```

The target protein should be cloned between the initial MA and TSGGSG, which is intended as a spacer between the target protein and the CGFP. The target must be cloned in frame for proper fusion.

```

NcoI      SphI      AatII
CCATGGctaatgcatgcaatggGACGTCgggtggaagcggtagaatgg...
  M A N A C N G T S G G S G K N G ...

```

The first codon of the target protein should be inserted where the first AAT (Asn) codon is. The last codon of the target before the fusion point should either (1) replace the second AAT (Asn) codon or (2) replace the GGG (Gly) codon that overlaps with the AatII site if the codon ends with a G.

The CGFP fusion is under the control of the araBAD promoter. The required araC regulator is divergently transcribed in the same vector. For use with pET11a-link-NGFP, we recommend growth in BL21(DE3) on LB agar supplemented with 100 ug/mL ampicillin, 35 ug/mL kanamycin, 10 uM IPTG and 0.02-0.2% arabinose. Grow the cells overnight at 30C and transfer them to room temperature for 2 days.

Plasmid features (approx)

- 1 - 288 CGFP+linker
- 1816 - 1003 kanamycin resistance

2887 - 2162 p15A origin  
4468 - 3583 araC

Thomas J. Magliery, Ph.D., July 20, 2004

References:

Magliery TJ, Wilson CGM, Pan W, Mishler D, Ghosh I, Hamilton AD, Regan L. 2004. Detecting protein-protein interactions with a GFP-fragment reassembly trap: scope and mechanism. *J Am Chem Soc*, submitted.

Ghosh I, Hamilton AD, Regan L. 2000. Antiparallel leucine zipper-directed protein reassembly: application to the green fluorescent protein. *J Am Chem Soc* 122: 5658.

References can be downloaded from:

<http://www.csb.yale.edu/people/regan/publications.html>

```
1  CCATGGCTAA  TGCATGCAAT  GGGACGTCGG  GTGGAAGCGG  TAAGAATGGA
51  ATCAAAGTGA  ACTTCAAGAC  CCGCCACAAC  ATTGAAGATG  GAAGCGTTCA
101  ACTAGCAGAC  CATTATCAAC  AAAATACTCC  AATTGGCGAT  GGCCCTGTCC
151  TTTTACCAGA  CAACCATTAC  CTGTCCACAC  AATCTGCCCT  TTCGAAAGAT
201  CCCAACGAAA  AGAGAGACCA  CATGGTCCTT  CTTGAGTTTG  TAACAGCTGC
251  TGGGATTACA  CATGGCATGG  ATGAACTGTA  CAACTAAGGA  TCCGGCTGCT
301  AACAAAGCCC  GAAAGGAAGC  TGAGTTGGCT  GCTGCCACCG  CTGAGCAATA
351  ACTAGCATAA  CCCCTTGGGG  CCTCTAAACG  GGTCTTGAGG  GGTTTTTTGC
401  TGAAAGGAGG  AACTATATCC  GGATATCCAC  AGGACGGGTG  TGGTCGCCAT
451  GATGGTGACC  ACCATTTGCC  CCATATCACC  GTCGGTACTG  ATCCCGTCGT
501  CAATAAACCG  AACCGCTACA  CCCTGAGCAT  CAAACTCTTT  TATCAGTTGG
551  ATCATGTCGG  CGGTGTCGCG  GCCAAGACGG  TCGAGCTTCT  TCACCAGAAT
601  GACATCACCT  TCCTCCACCT  TCATCCTCAG  CAAATCCAGC  CCTTCCCGAT
651  CTGTTGAACT  GCCGGATGCC  TTGTTCGGTAA  AGATGCGGTT  AGCTTTTACC
701  CCTGCATCTT  TGAGCGCTGA  GGTCTGCCTC  GTGAAGAAGG  TGTTGCTGAC
751  TCATACCAGG  CCTGAATCGC  CCCATCATCC  AGCCAGAAAG  TGAGGGAGCC
801  ACGGTTGATG  AGAGCTTTGT  TGTAGGTGGA  CCAGTTGGTG  ATTTTGAACT
851  TTTGCTTTGC  CACGGAACGG  TCTGCGTTGT  CGGGAAGATG  CGTGATCTGA
901  TCCTTCAACT  CAGCAAAAGT  TCGATTTATT  CAACAAAGCC  GCCGTCCCGT
951  CAAGTCAGCG  TAATGCTCTG  CCAGTGTTAC  AACCAATTAA  CCAATTCTGA
1001  TTAGAAAAAC  TCATCGAGCA  TCAAATGAAA  CTGCAATTTA  TTCATATCAG
1051  GATTATCAAT  ACCATATTTT  TGAAAAAGCC  GTTTCTGTAA  TGAAGGAGAA
1101  AACTCACCGA  GGCAGTTCCA  TAGGATGGCA  AGATCCTGGT  ATCGGTCTGC
1151  GATTCCGACT  CGTCCAACAT  CAATACAACC  TATTAATTTT  CCCTCGTCAA
1201  AAATAAGGTT  ATCAAGTGAG  AAATCACCAT  GAGTGACGAC  TGAATCCGGT
1251  GAGAATGGCA  AAAGCTTATG  CATTTCTTTC  CAGACTTGTT  CAACAGGCCA
1301  GCCATTACGC  TCGTCATCAA  AATCACTCGC  ATCAACCAAA  CCGTTATTCA
1351  TTCGTGATTG  CGCCTGAGCG  AGACGAAATA  CGCGATCGCT  GTTAAAAGGA
1401  CAATTACAAA  CAGGAATCGA  ATGCAACCGG  CGCAGGAACA  CTGCCAGCGC
1451  ATCAACAATA  TTTTCACTG  AATCAGGATA  TTCTTCTAAT  ACCTGGAATG
1501  CTGTTTTTCC  GGGGATCGCA  GTGGTGAGTA  ACCATGCATC  ATCAGGAGTA
1551  CGGATAAAAT  GCTTGATGGT  CGGAAGAGGC  ATAAATTCCG  TCAGCCAGTT
1601  TAGTCTGACC  ATCTCATCTG  TAACATCATT  GGCAACGCTA  CCTTTGCCAT
1651  GTTTCAGAAA  CAACTCTGGC  GCATCGGGCT  TCCCATACAA  TCGATAGATT
1701  GTCGCACCTG  ATTGCCCGAC  ATTATCGCGA  GCCCATTAT  ACCCATATAA
```

1751 ATCAGCATCC ATGTTGGAAT TTAATCGCGG CCTCGAGCAA GACGTTTCCC  
1801 GTTGAATATG GCTCATAACA CCCCTTGTAT TACTGTTTAT GTAAGCAGAC  
1851 AGTTTTTATTG TTCATGATGA TATATTTTTTA TCTTGTGCAA TGTAACATCA  
1901 GAGATTTTGA GACACAACGT GGCTTTGTTG AATAAATCGA ACTTTTGCTG  
1951 AGTTGAAGGA TCAGATCACG CATCTTCCCG ACAACGCAGA CCGTTCCGTG  
2001 GCAAAGCAAA AGTTCAAAAT CACCAACTGG TCCACCTACA ACAAAGCTCT  
2051 CATCAACCGT GGCTCCCTCA CTTTCTGGCT GGATGATGGG GCGATTGAGG  
2101 CCTGGTATGA GTCAGCAACA CTTTCTTCCAC GAGGCAGACC TCAGCGCTAG  
2151 CGGAGTGTAT ACTGGCTTAC TATGTTGGCA CTGATGAGGG TGTCAGTGAA  
2201 GTGCTTCATG TGGCAGGAGA AAAAAGGCTG CACCGGTGCG TCAGCAGAAT  
2251 ATGTGATACA GGATATATTC CGCTTCCCTCG CTCACTGACT CGCTACGCTC  
2301 GGTTCGTTTCA CTGCGGCGAG CGGAAATGGC TTACGAACGG GGCGGAGATT  
2351 TCCTGGAAGA TGCCAGGAAG ATACTTAACA GGGAAAGTGAG AGGGCCGCGG  
2401 CAAAGCCGTT TTTCCATAGG CTCCGCCCCC CTGACAAGCA TCACGAAATC  
2451 TGACGCTCAA ATCAGTGGTG GCGAAACCCG ACAGGACTAT AAAGATACCA  
2501 GGCGTTTCCC CTGGCGGCTC CCTCGTGCGC TCTCCTGTTC CTGCCTTTCG  
2551 GTTTACCGGT GTCATTCCGC TGTTATGGCC GCGTTTGTCT CATTCCACGC  
2601 CTGACACTCA GTTCCGGGTA GGCAGTTCGC TCCAAGCTGG ACTGTATGCA  
2651 CGAACCCCCC GTTCAGTCCG ACCGCTGCGC CTTATCCGGT AACTATCGTC  
2701 TTGAGTCCAA CCCGGAAAGA CATGCAAAAG CACCACTGGC AGCAGCCACT  
2751 GGTAATTGAT TTAGAGGAGT TAGTCTTGAA GTCATGCGCC GGTTAAGGCT  
2801 AAAGTAAAG GACAAGTTTT GGTGACTGCG CTCCTCCAAG CCAGTTACCT  
2851 CGGTTCAAAG AGTTGGTAGC TCAGAGAACC TTCGAAAAAC CGCCCTGCAA  
2901 GGCGGTTTTT TCGTTTTTTCAG AGCAAGAGAT TACGCGCAGA CCAAAACGAT  
2951 CTCAAGAAGA TCATCTTATT AAGGGGTCTG ACGCTCAGTG GAACGAAAAC  
3001 TCACGTAAAG GGATTTTGGT CATGAGATTA TCAAAAAGGA TCTTACCTA  
3051 GATCCTTTTTA AATTAATAAAT GAAGTTTTTAA ATCAATCTAA AGTATATATG  
3101 AGTAAACTTG GTCTGACAGT TACCAATGCT TAATCAGTGA GGCACCTATC  
3151 TCAGCGATCT GTCTATTTTCG TTCATCCATA GTTGCCTGAC TCCCCGTCGT  
3201 GTAGATAACT ACGATACGGG AGGGCTTACC ATCTGGCCCC AGTGCTGCAA  
3251 TGATACCGCG AGACCCACGC TCACCGGCTC CAGATTTATC AGCAATAAAC  
3301 CAGCCAGCCG GAAGGGCCGA GCGCAGAAGT GGTCCCTGCAA CTTTATCCGC  
3351 CTCCATCCAG TCTATTAATT GTTGCCGGGA AGCTAGAGTA AGTAGTTCGC  
3401 CAGTTAATAG TTTGCGCAAC GTTGTGCCA TTGCTGCAGA TCGATGCATA  
3451 ATGTGCCTGT CAAATGGACG AAGCAGGGAT TCTGCAAACC CTATGCTACT  
3501 CCGTCAAGCC GTCAATTGTC TGATTCGTTA CCAATTATGA CAACTTGACG  
3551 GCTACATCAT TCACTTTTTTTC TTCACAACCG GCACGGAAC CTGCTCGGGCT  
3601 GGCCCCGGTG CATTTTTTTAA ATACCCGCGA GAAATAGAGT TGATCGTCAA  
3651 AACCAACATT GCGACCGACG GTGGCGATAG GCATCCGGGT GGTGCTCAA  
3701 AGCAGCTTCG CCTGGCTGAT ACGTTGGTCC TCGCGCCAGC TTAAGACGCT  
3751 AATCCCTAAC TGCTGGCGGA AAAGATGTGA CAGACGCGAC GGCGACAAGC  
3801 AAACATGCTG TGCGACGCTG GCGATATCAA AATTGCTGTC TGCCAGGTGA  
3851 TCGCTGATGT ACTGACAAGC CTCGCGTACC CGATTATCCA TCGGTGGATG  
3901 GAGCGACTCG TTAATCGCTT CCATGCGCCG CAGTAACAAT TGCTCAAGCA  
3951 GATTTATCGC CAGCAGCTCC GAATAGCGCC CTTCCCCTTG CCCGGCGTTA  
4001 ATGATTTGCC CAAACAGGTC GCTGAAATGC GGCTGGTGCG CTTTATCCGG  
4051 GCGAAAGAAC CCCGTATTGG CAAATATTGA CGGCCAGTTA AGCCATTTCAT  
4101 GCCAGTAGGC GCGCGGACGA AAGTAAACCC ACTGGTGATA CCATTGCGGA  
4151 GCCTCCGGAT GACGACCGTA GTGATGAATC TCTCCTGGCG GGAACAGCAA  
4201 AATATCACCC GGTTCGGCAA CAAATTCTCG TCCCTGATTT TTCACCACCC  
4251 CCTGACCGCG AATGGTGAGA TTGAGAATAT AACCTTTCAT TCCCAGCGGT  
4301 CCGTTCGATA AAAAATCGAG ATAACCGTTG GCCTCAATCG GCGTTAAACC  
4351 CGCCACCAGA TGGGCATTAA ACGAGTATCC CGGCAGCAGG GGATCATTTT  
4401 GCGCTTCAGC CATACTTTTTC ATACTCCCGC CATTTCAGAGA AGAAACCAAT  
4451 TGTCCATATT GCATCAGACA TTGCCGTCAC TGCGTCTTTT ACTGGCTCTT  
4501 CTCGCTAACC AAACCGGTAA CCCCCTTAT TAAAAGCATT CTGTAACAAA  
4551 GCGGGACCAA AGCCATGACA AAAACGCGTA ACAAAGTGT CTATAATCAC

4601 GGCAGAAAAG TCCACATTGA TTATTTGCAC GCGGTCACAC TTTGCTATGC  
4651 CATAGCATT TATCCATAA GATTAGCGGA TCCTACCTGA CGCTTTTTAT  
4701 CGCAACTCTC TACTGTTTCT CCATACCCGT TTTTTTGGGC TAGAAATAAT  
4751 TTTGTTTAAC TTTAAGAAGG AGATATAC