Creating standardized vectors for fluorescent microscopy

Research project for the Master program of Molecular Biology & Biotechnology

Faculty of Mathematics and Natural Sciences, Groningen, the Netherlands

Department of Molecular Genetics

Student number: 2368242

October 2012 – April 2013

Student:

Claudio Tiecher

Supervisors:

Mirjam Boonstra

O. P. Kuipers

Abstract

Fluorescent proteins (FPs) have been broadly employed during the last two decades in order to answer scientific questions. Since their great potential was unveiled by Prasher *et al.* in 1992, a lot has been done to improve the power of this tool. Its intrinsic natural fluorescence is given by the specific amino acid sequence. This makes the FPs suitable for *in vivo* cell imaging since they can be genetically engineered and fused to other proteins. Over the last decades several applications and set-ups have been developed to exploit these fluorescent probes. However the use of FPs to study biology has still a few drawbacks which sometimes are difficult to overcome. Often the properties of the probes are not suitable for certain experimental set-ups. For instance low brightness, maturation time or half-life of FPs often gives constraints in the use of live imaging studies. The creation of fusion proteins and its adaptation to a specific experimental set-up is time consuming. Therefore the goal of this project is the creation of a standardized fluorescent proteins collection for *Bacillus subtilis*, *Lactococcus lactis* and *Streptococcus pneumoniae.* This standardized collection aims to facilitate the creation of new fusion proteins.

1. INTRODUCTION

1.1 Biological molecules

The study of proteins is an essential step towards a deeper understanding of how living organisms function. Indeed, these macromolecules are the building blocks of all living cells, along with nucleic acids and lipids.

It is not surprising that the scientific community studies different aspects of those macromolecules, such as function, structure, expression and localization. For the investigation of these characteristics scientists have developed a lot of *in silico*, *in vitro* and *in vivo* techniques, which employ various tools. One of these is the fluorescent protein, which has shown its investigative power during the last two decades.

1.2 Fluorescent proteins

The fluorescent proteins was first discovered in the 60's when Osamu Shimomura purified for the first time the protein Aequorin from *Aequorea Victoria* (Shimomura et al. 1962). His pioneering work opened up the way towards the discovery and characterisation of the green fluorescent protein (GFP) (Shimomura 1979).

GFP was cloned (Prasher et al. 1992) and further investigated. The fluorescence of the GFP is an intrinsic property due to its amino acid sequence, specifically produced by the maturation of three amino acids (Ser-Tyr-Gly) in the presence of molecular oxygen (Cubitt et al. 1995). This, together with the evidence that the protein can correctly fold in the functional fluorescent protein within organisms other than the jellyfish, unlocked the potential of GFP (Chalfie et al. 1994). The scientific community recognized the striking value of the discovery and over the years put effort into engineering and characterizing a broad variety of fluorescent proteins.

Several mutations are shown to shift the emission spectrum of GFP towards the IR and the UV; although none of those engineered versions show an emission maximum longer than 529 nm. Only with the discovery of the drFP583, commercially known as DsRed and naturally present in a *Discosoma striata* species (Baird et al.

2000), the fluorescent toolbox finally includes proteins which span the light spectrum from deep blue to far-red.

From the two natural sources (GFP and DsRed) a lot of variants have been developed. They all feature different parameters; for instance they have slightly different excitation and emission wavelengths which provide the scientist with different colorful molecular labels. Moreover extinction coefficient, quantum yield and brightness are also important parameters which differ across the fluorescent protein palette (Day & Davidson 2009).

Many different protocols employing FPs have been set up to study *in vivo* system. Several fluorescence microscopy techniques have been shown to be useful and have enabled the researcher to investigate different biological questions (Ishikawa-Ankerhold et al. 2012). Some examples are: Fluorescence Resonance Energy Transfer (FRET), Fluorescence Microscopy After Photobleaching (FRAP) and Fluorescence Loss In Photobleaching (FLIP).

1.3 Fusion proteins

As mentioned before molecular biology widely uses fluorescence as a tool, in order to investigate the expression, localization and interaction of proteins *in vivo*.

One of the approach consists of linking a chromophore to the protein of interest. Applications have shown the power of this technique; fusion proteins have been used successfully to show the dynamics of tagged proteins within the different cell compartments (Lippincott-schwartz et al. 2000). They have also been used to study several protein-protein interactions (Suzuki et al. 1998).

Fusion proteins consist of creating a link between the FP and the protein of interest using recombinant DNA technology. Although any fusion is possible, in practice not all of them work as expected and only a few fit the chosen experimental set-up. The bulky body of the FPs can disrupt or hamper the natural function and behavior of the protein of interest. For this and other reasons, such as properties specific to the fluorescent protein, to know whether a fusion protein works as expected it has to be created and tested.

1.4 Standardized collection

The creation of new fusion proteins, which suit the experimental set-up follows trial and error approach. The need of handy tools to quickly create and test fusion proteins is necessary. The aim of this work is to create a new fluorescent protein collection which can be used to easily create new fusion proteins.

In the specific case the collection is designed to be compatible with most of the backbones used in the laboratory and the collection includes three different fluorescent proteins: two are derived from GFP (mGFP-dsm and mGFP-opt) and the other is derived from RFP (mKate2). mGFP-opt and mGFP-dsm have been shown to give optimal fluorescence signal when expressed, the first in *B. subtilis* and the second in both *L. lactis* and *S. pneumoniae* (Overkamp et al. 2013).

The collection has the aim to fulfill the following goals. It is compatible with most frequent restriction sites found in *B. subtilis, L. lactis* and *S. pneumoniae.* The fluorescent proteins are monomeric, thus suitable for live imaging study. Furthermore the collection can be used to create both, C-terminal and N-terminal fusion.

The plasmid collection is created by the use of TA cloning which it does not require restriction steps.

2. MATERIALS AND METHODS

2.1 Polymerase chain reaction

Two sets of primers were designed (Appendix E) for the amplification of the C-terminal fusion and the N-terminal fusion constructs, respectively.

The C-terminal fusion inserts were generated through two PCR rounds: for the first 1-F and 2-R were employed and for the second 3-F together with 2-R was used. This resulted in the addition of the linker sequence and several restriction sites.

The N-terminal fusion inserts were generated after three PCR rounds: in the first 2-F and 1-R1 were used, for the second 2-F together with 2-R2 was used, lastly the final product was obtained using 2-F and 3-R3. This resulted in the addition of the ribosome binding site, the linker sequence and several restriction sites.

For each primer, the version which anneals the desired template was picked (Appendix G). Each fluorescent protein coding sequence was amplified starting from three templates: pJWV102-GFP-opt, pNZ-MS2S-mGFPdsm and pNZ-mKATE2. Phusion High-Fidelity DNA Polymerase (ThermoScientific) was used to perform all the PCR reactions except for the last PCR round where only *Taq* polymerase was employed, since it produces products with 3'-A overhang necessary for the cloning into pGEM-T Easy vector (see Appendix D for technical details). The insert can enter the backbone in two different orientations resulting in the creation of two slightly different constructs, since the PCR products are not symmetrical. Different ratios (from 3:1 to 1:3) between vector and insert were tested until the desired ligation product was obtained.

Gel electrophoresis was used to check the sizes of all the PCR products. The samples were prepared mixing 1 part of PCR reaction with 1 part of loading dye (Fermentas™ 6X DNA Loading Dye). The samples were then loaded on 1% agarose gel (Agarose SERVA for DNA electrophoresis), dissolved in buffer TAE 1x, beside DNA Ladder (GeneRuler 1 kb DNA Ladder). The gel was run at 80 mA for 40 min.

All the PCR products were purified using High Pure PCR Product Purification Kit (Roche Applied Science).

2.2 *Escherichia coli* competent cells

The glycerol stock of *E. Coli* competent cells (DH5α and BL21 (DE3)) was stored at -80°C and prepared as follow.

The strain was streaked out on LB agar (without antibiotic) and incubated overnight at 37°C. Then one colony was picked, inoculated in 5 mL LB broth and grown overnight at 37°C in the shaker (~220 rpm). Afterwards 1:100 of the inoculum was transferred in 200 mL LB and was grown to an OD_{600} in between 0.3-0.4. Once the desired OD was reached the cells were pellet for 15 min at 4°C and 2700 rpm. The cells were gently resuspended in 16 mL chill RF1 buffer (RbCl 12 g/L, CH_3CO_2K 4.9 g/L, $CaCl_2$ $2{\cdot}H_2O$ 1.5 g/L, Glycerol 150 g/L, pH 5.8), vortexed and incubated in ice for 15 min. They were then centrifuged again (see previous step). This time the cells were resuspended in 4 mL RF2 buffer (0.5 M MOPS 20mL/L, RbCl 1.2 g/L, $CaCl_2$ $2{\cdot}H_2O$ 11g/L, Glycerol 150 g/L, pH 6.8) and incubated in ice for 15 min. Eventually the cells can be stored in 200 µL aliquots at -80°C ready to be used.

2.3 Transformation in *E. Coli*

All the ligation products were transformed in *E. Coli* competent cells (see 2.2) following the protocol below.

The ligation product was added to 50 µL of competent cells which were incubated in ice for 20 min. Afterwards the cells were heat shocked at 42°C for 50 sec and returned to ice for 2 min. LB broth was added up to 1 mL and the tubes were placed in the shaker at 37°C for 1,5 hrs for recovering. Eventually the cells were plated onto LB agar, containing the appropriate selection marker. The plates were incubated at 37°C overnight.

2.4 Screening of transformants

The plates showing transformants were screened for the presence of positive colonies containing the plasmid with the desired insert.

In the first place this is done, when transforming into DH5α strain, at a glance, by discerning between white and blue colonies which was possible since 45 µg/mL of X-Gal were added to the growth medium. White colonies were promising candidates; in

those the sequence coding for the α-peptide of the enzyme ß-galactosidase (lacking in the genome of the DH5α strain) was disrupted by the insert, therefore X-Gal was not processed and the colonies remained white. Although blue colonies could also be positive candidates, in the case the insert would not disrupt the *lacZ* gene, which could be due to an in-frame insertion.

Secondly colonies were picked and the size of the insert was checked using Colony PCR (see 2.5) and restriction analysis. The latter was done picking meaningful restriction endonucleases and following the ThermoScientific FastDigest® protocol.

2.5 Colony PCR

For each transformation, depending on the total number of grown colonies, 1 to 10 were picked and the size of the insert was checked using the Colony PCR primers (Appendix E).

The PCR products were checked using gel electrophoresis (see 2.1).

Eventually the transformants, which resulted to be positive after both checks were sequenced by MacroGen. The samples which showed the correct sequences were included in the collection.

2.6 Strains and plasmids

The strains and plasmids used to perform the experiments are listed in Appendix G. The cells were grown overnight in lysogeny broth (10 g/L Tryptone, 5 g/L Yeast Extract, 5 g/L NaCl) at 37°C and selection marker was added whether necessary (Ampicillin 100 µg/mL).

All the stocks were prepared mixing 2 parts of overnight culture with 1 part of 85% glycerol and were stored at -80°C.

All the plasmids were isolated using High Pure Plasmid Isolation Kit (Roche Applied Science).

3. RESULTS

3.1 Molecular features of the collection

All the constructs show two common features: linker and restriction sites. The N-terminal fusion constructs show a ribosome binding site (RBS) and do not have stop codon. The C-terminal constructs contain a stop codon and no RBS.

The linker sequence is made out 15 amino acids (Figure 1C). It is an hybrid between flexible linker sequence and helical one. The helix forming unit (EAAAK) confers solubility due to the presence of glutamic acid (E) and lysine (K) residues. Additionally the presence of arginine (R), serine (S) and threonine (T) also increases the solubility of the linker. The helical unit in the middle effectively controls the distance between the two fused proteins, thus avoiding inadequate interactions. Moreover the flexible portion of the linker provides flexibility to the dimer (Arai et al. 2001).

All the constructs feature several restriction sites which are native from the cloning vector (Appendix D) and some which are introduced during the PCR steps next to the linker. Those customized restriction sites can be visualized in Figure 1A-B. The pool includes well-known endonucleases, which are generally used in molecular genetic laboratories and specifically in the laboratory where the research is carried out. Moreover the natural occurrence of the restriction sites within the genome of *B. subtilis*, *L. lactis* and *S. pneumoniae* is taken into account. So the likelihood that the restriction sites will be contained in the sequence of the gene of interest is reduced.

All the N-terminal constructs contain the RBS (AGGAGG) which is cloned 7 base pairs upstream the starting codon (Figure 1C). This makes sure that any fusion protein construct will contain the site that can enable transcription of the gene downstream the RBS. Its sequence is optimized for three bacteria: *B. subtilis*, *L. lactis* and *S. pneumoniae.*

All the C-terminal constructs have a stop codon (TAA) and do not show any RBS. The stop codon is essential for the correct termination of transcription which ensures the mRNA is efficiently transcribed.

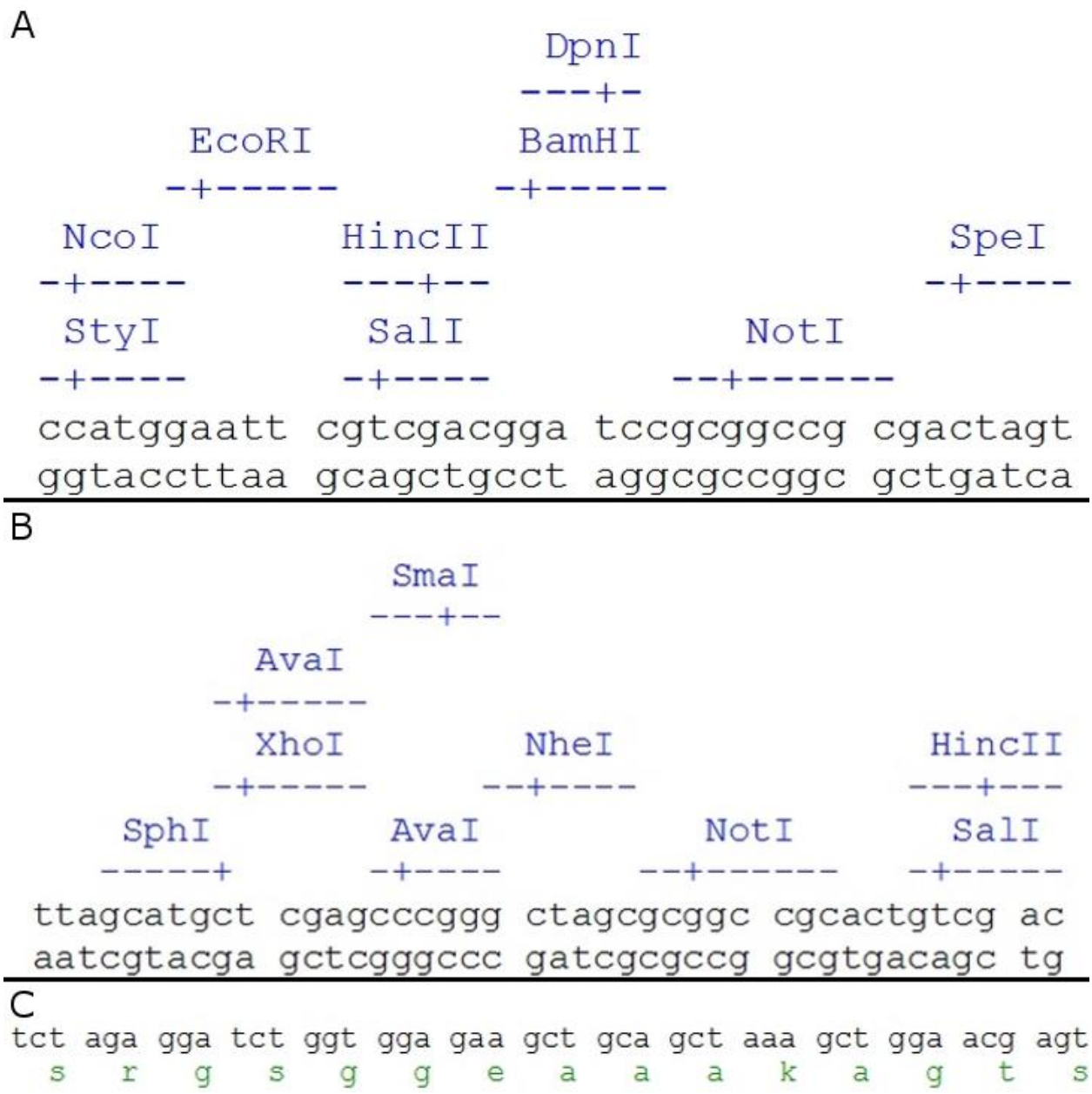


Figure 1 Collection features. (A) The region adjacent to the linker and upstream the fluorescent protein gene of the C-terminal fusion constructs is shown. (B) The region adjacent to the linker and downstream the fluorescent protein gene of the N-terminal fusion constructs is shown. (C) The linker sequence. In green the peptide sequence is shown. The nucleotide sequence is also displayed, in black and it is taken from the construct containing mGFP-dsm..

3.2 Fluorescent protein collection for fusion protein creation

The constructs included in the collection are listed in Table 1.

For mKate2 all the expected combinations are obtained.

For mGFP-dsm half of the expected constructs are created: one construct for C-terminal fusions and one for the N-terminal fusions.

For mGFP-opt none of the desired combination are obtained, neither C-terminal nor N-terminal fusion.

mGFP-dsm plates always show colonies. The positive ones are sequenced, but all of them only show one orientation: Rev and Fwd for the N-terminal and C-terminal construct, respectively. Different ratios between the insert and the vector during the ligation step are tested. This seems not to affect the orientation obtained.

mGFP-opt plates always show colonies. When these are tested none of them result positive. The Colony PCR check shows the presence of empty plasmid and plasmid containing an insert which is roughly half the expected size.

Table 1 The summary of the constructs which were created.

Fluorescent protein	Type	Result	Orientation obtained	Construct name
mKate2	N-terminal fusion	OK	2	pGEM-rbs-mKate2-linker_Fwd/Rev
	C-terminal fusion	OK	2	pGEM-linker-mKate2_Fwd/Rev
mGFP-dsm	N-terminal fusion	OK	1(*)	pGEM-rbs-mGFPdsm-linker_Rev
	C-terminal fusion	OK	1(**)	pGEM-linker-mGFPdsm_Fwd
mGFP-opt	N-terminal fusion	NO	0	
	C-terminal fusion	NO	0	

In the first column the fluorescent protein which was cloned is shown. In the following columns the construct type, name and the results are shown. (*) The orientation of the fluorescent protein open reading frame is antisense compared to the ß-lactamase gene (Appendix F). (**)The orientation of the fluorescent protein open reading frame is sense compared to the ß-lactamase gene (Appendix F).

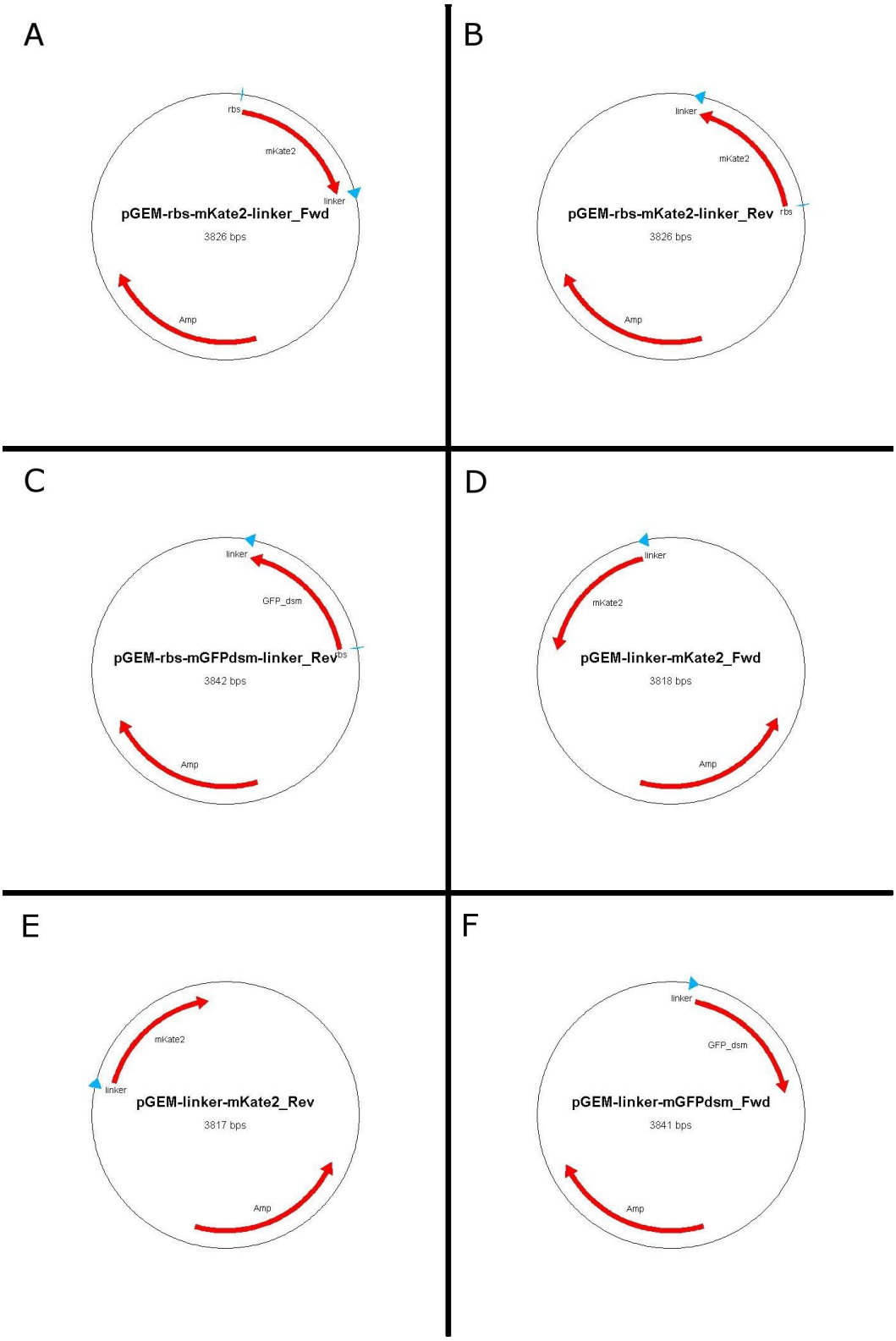


Figure 2 The summary of the constructs. The ampicillin antibiotic resistance (Amp), the RBS, the linker and the name of the FP are shown. (A) pGEM-rbs-mKate2-linker_Fwd, (B) pGEM-rbs-mKate2-linker_Rev, (C) pGEM-rbs-mGFPdsm-linker_Rev, (D) pGEM-linker-mKate2_Fwd, (E) pGEM-linker-mKate2_Rev and (F) pGEM-linker-mGFPdsm_Fwd.

4. DISCUSSION AND CONCLUSION

4.1 The fluorescent protein collection

The collection is a handy tool which can be used in order to build desired fusion protein constructs. It is versatile due to the presence of a rich multiple cloning site. The fusion proteins can be created into *E. coli*, used as host cloning organism and then easily transferred into other plasmid backbones of interest.

Important is that the constructs have never been tested for the creation of fusion proteins. Moreover the N-terminal constructs contain a ribosome biding site which may be optimal for enhancing transcription in some organisms, such as *B. subtilis*, *L. lactis* and *S. pneumoniae*, but may have other effects in different species.

4.2 Cloning issues

The reason why none of the constructs containing mGFP-opt are obtained is not clear. The flanking regions of the inserts are identical across the same type of constructs. This leads to the conclusion that the extra restriction sites and PCR editing is not the reason of the cloning failure.

The homology between the two green fluorescent protein nucleotide sequences is relatively high (match percentage: 79%), thus excluding also that the different DNA sequences may lead to unexpected recombination events. Furthermore the nucleotide sequences of the inserts are checked for homology with the *E. Coli* genome, which results to be none.

The cloning is also attempted in different strains (DH5α and BL21 (DE3)). In the DH5α strain the T7 promoter may be leaky and cause the over production of the fluorescent protein. Whether the protein product is toxic for the cells cloning in DE3 strains should be successful because of the very little T7-mediated transcription. Unfortunately no relevant differences are observed during the cloning into different strains, both in the presence and absence of IPTG. Other strains, with a genotype profile which tightly represses the T7 and SP6 promoters may be tested. However the combination vector/construct is most likely the cause of the cloning issue, therefore another vector featuring different promoters may be used for the cloning.

In the case of mGFP-dsm only one orientation is obtained: this may be due to the stability of the insert which would be higher in one conformation. This hypothesis is supported by the results obtained from the experiments. The insert always enters the multiple cloning site with the same annealing behavior. It seems that the linker region finds its most stable conformation when annealing to one end of the open vector (Appendix F).

4.3 Future

The constructs should be tested trying to create some fusion proteins. Once a few constructs are obtained the expression and the functionality of the protein fused to the fluorescent probe may be tested. The creation of a FRET couple using this collection is possible.

Another aspect that could be tested is the role of the promoter and its strength when interchanged with different ones. The synergy of the ribosome binding site when coupled to a different promoter than IPTG inducible could also be investigated.

Overall this collection retains a lot of potential with several applications in different directions. These could lead to answer other biological questions and confirm once more the inestimable value of the fluorescent protein tool.

References

Arai, R. et al., 2001. Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Engineering*, 14(8), pp.529–532. Available at: http://peds.oxfordjournals.org/content/14/8/529.abstract.

Baird, G.S., Zacharias, D.A. & Tsien, R.Y., 2000. Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proceedings of the National Academy of Sciences*, 97(22), pp.11984–11989. Available at: http://www.pnas.org/content/97/22/11984.abstract.

Chalfie, M. et al., 1994. Green fluorescent protein as a marker for gene expression. *Science*, 263(5148), pp.802–805. Available at: http://www.sciencemag.org/content/263/5148/802.abstract.

Cubitt, A.B. et al., 1995. Understanding, improving and using green fluorescent proteins. *Trends in Biochemical Sciences*, 20(11), pp.448–455. Available at: http://www.sciencedirect.com/science/article/pii/S0968000400890994.

Day, R.N. & Davidson, M.W., 2009. The fluorescent protein palette: tools for cellular imaging. *Chem. Soc. Rev.*, 38(10), pp.2887–2921. Available at: http://dx.doi.org/10.1039/B901966A.

Ishikawa-Ankerhold, H.C., Ankerhold, R. & Drummen, G.P.C., 2012. Advanced Fluorescence Microscopy Techniques—FRAP, FLIP, FLAP, FRET and FLIM. *Molecules*, 17(4), pp.4047–4132. Available at: http://www.mdpi.com/1420-3049/17/4/4047.

Lippincott-schwartz, J., Roberts, T.H. & Hirschberg, K., 2000. Secretory Protein Trafficking and Organelle Dynamics in Living Cells. *Cell*, 16, pp.557–89.

Overkamp, W. et al., 2013. Benchmarking various green fluorescent protein variants in Bacillus subtilis, Streptococcus pneumoniae, and Lactococcus lactis for live cell imaging. *Applied and environmental microbiology*, 79(20), pp.6481–90. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23956387 [Accessed December 5, 2013].

Pedelacq, J.-D. et al., 2006. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotech*, 24(1), pp.79–88. Available at: http://dx.doi.org/10.1038/nbt1172.

Prasher, D.C. et al., 1992. Primary structure of the Aequorea victoria green-fluorescent protein. *Gene*, 111(2), pp.229–233. Available at: http://www.sciencedirect.com/science/article/pii/037811199290691H.

Shcherbo, D. et al., 2009. Far-red fluorescent tags for protein imaging in living tissues. *Biochem J*, 418(3), pp.567–574. Available at: http://www.biochemj.org/bj/418/bj4180567.htm.

Shimomura, O., 1979. Structure of the chromophore of Aequorea green fluorescent protein. *{FEBS} Letters*, 104(2), pp.220–222. Available at: http://www.sciencedirect.com/science/article/pii/0014579379808182.

Shimomura, O., Johnson, F.H. & Saiga, Y., 1962. Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea. *Journal of Cellular and Comparative Physiology*, 59(3), pp.223–239. Available at: http://dx.doi.org/10.1002/jcp.1030590302.

Suzuki, Y. et al., 1998. Swing of the lever arm of a myosin motor at the isomerization and phosphate-release steps. *Nature*, 396, pp.380–383.

Zacharias, D.A. et al., 2002. Partitioning of Lipid-Modified Monomeric GFPs into Membrane Microdomains of Live Cells. *Science*, 296(5569), pp.913–916. Available at: http://www.sciencemag.org/content/296/5569/913.abstract.

Appendix A

Nucleotide sequence used as source for mGFP-opt: GenBank KF410613.1 (sfGFP(Sp)).

Vector name used as source (Molecular Genetics laboratory): pJWV102-GFP-opt.

Appendix B

Nucleotide sequence used as source for mGFP-dsm: derived from GenBank KF410617.1 (sfGFP(Bs)).

Vector name used as source (Molecular Genetics laboratory): pNZ-MS2S-mGFPdsm.

Appendix C

Nucleotide sequence used as source for mKate2. The sequence below shows only the CDS excluding the base pairs coding for the starting codon (ATG) and the stop codon (TAA). The protein sequence matches the GenBank entry AFM44941.1.

Vector name used as source: pNZ-mKATE2.

```
  1  agctcagaac ttatcaagga aaatatgcac atgaaattgt acatggaagg
     >>.....................mKate2.....................>
       s  s  e   l  i  k   e  n  m  h   m  k  l   y  m  e

 51  aacagtaaat aatcaccact ttaaatgtac ctcagaagga gaaggaaaac
     >.....................mKate2.....................>
     g  t  v  n   n  h  h   f  k  c   t  s  e  g   e  g  k

101  catatgaagg tactcaaacc atgcgtatta aggccgttga aggtggacca
     >.....................mKate2.....................>
      p  y  e   g  t  q  t   m  r  i   k  a  v   e  g  g  p

151  ttgccttttg cctttgatat tcttgctaca tcttttatgt acggatcaaa
     >.....................mKate2.....................>
       l  p  f   a  f  d   i  l  a  t   s  f  m   y  g  s

201  gacttttatc aatcataccc aaggtatccc agatttcttt aaacagtcat
     >.....................mKate2.....................>
     k  t  f  i   n  h  t   q  g  i   p  d  f  f   k  q  s

251  ttcctgaagg atttacatgg gaacgtgtca caacttatga agatggtgga
     >.....................mKate2.....................>
      f  p  e   g  f  t  w   e  r  v   t  t  y   e  d  g  g

301  gtattgacag caactcaaga tacatctttg caagatggtt gtcttatcta
     >.....................mKate2.....................>
       v  l  t   a  t  q   d  t  s  l   q  d  g   c  l  i

351  caatgtaaag atccgtggag ttaattttcc aagtaatggt cctgttatgc
     >.....................mKate2.....................>
     y  n  v  k   i  r  g   v  n  f   p  s  n  g   p  v  m

401  agaaaaagac ccttggatgg gaagcatcta ccgaaacatt atatcctgcc
     >.....................mKate2.....................>
      q  k  k   t  l  g  w   e  a  s   t  e  t   l  y  p  a

451  gatggtggat tggaaggtcg tgctgatatg gcattgaaac ttgtcggtgg
     >.....................mKate2.....................>
       d  g  g   l  e  g   r  a  d  m   a  l  k   l  v  g

501  aggtcacctt atctgtaatt tgaagaccac ataccgttct aaaaagccag
     >.....................mKate2.....................>
     g  g  h  l   i  c  n   l  k  t   t  y  r  s   k  k  p

551  ctaagaatct taagatgcct ggtgtttact acgtcgatcg tcgtttagaa
```

```
     >.....................mKate2.....................>
      a  k  n   l  k  m  p   g  v  y   y  v  d   r  r  l  e

601  cgtatcaaag aagcagataa ggaaacttat gttgaacagc acgaagtagc
     >.....................mKate2.....................>
       r  i  k   e  a  d   k  e  t  y   v  e  q   h  e  v

651  cgtcgcacgt tattgtgatt tgcctagtaa attgggacac cgt
     >.................mKate2.................>>
     a  v  a  r   y  c  d   l  p  s   k  l  g  h   r
```

Appendix D

Technical manual source: pGEM®-T Easy Vector System (http://www.promega.co.uk/resources/protocols/technical-manuals/0/pgem-t-and-pgem-t-easy-vector-systems-protocol/).

Appendix E

Here is the list of primers used during this project.

The starting codon (ATG) is bold and the ribosome binding site (AGGAGG) is in italic. The linker sequence is underlined.

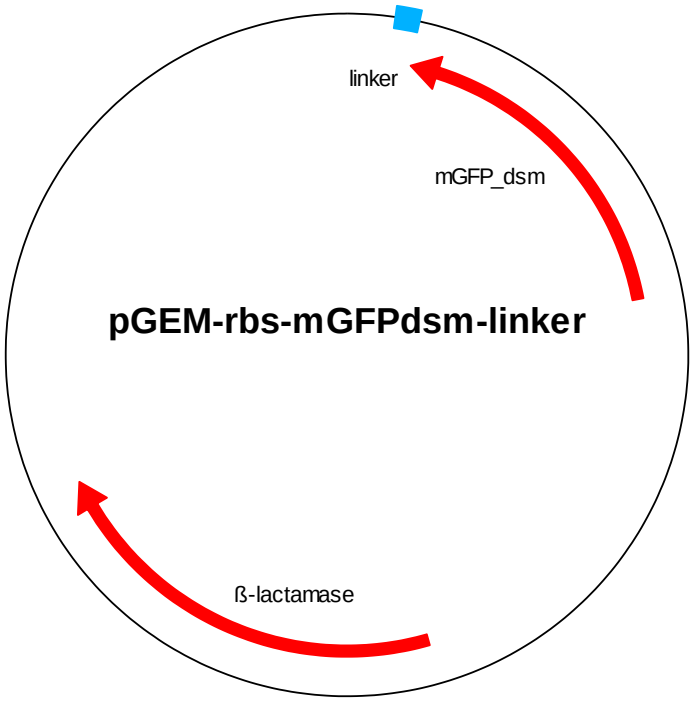
Primer name	Oligonucleotide sequence (5'→3')
C-terminal fusion	
1-C-mKate2-F	*TAGAGGATCTGGTGGAGAAGCTGCAGCTAAAGCTGGAACG AGT*TCAGAACTTATCAAGGAAAATATGCACATG
1-C-GFPdsm-F	*TAGAGGATCTGGTGGAGAAGCTGCAGCTAAAGCTGGAACG AGT*TCAAAAGGAGAAGAGCTGTT
1-C-GFPopt-F	*TAGAGGATCTGGTGGAGAAGCTGCAGCTAAAGCTGGAACG AGT*GTTTCTAAAGGTGAAGAATTGTTTACAG
2-C-mKate2-R	GCATGCTCGAGCCCGGGCTAGCTTAACGGTGTCCCAATTTA CTAGGCAAATC
2-C-GFPdsm-R	GCATGCTCGAGCCCGGGCTAGCTTACTTATAAAGCTCATCC ATGCCGTG
2-C-GFPopt-R	GCATGCTCGAGCCCGGGCTAGCTTATTTATACAATTCATCC ATACCATGTGTAATAC
3-C-Terminal-F	CCATGGAATTCGTCGACGGATCCGCGGCCGCGACTAGT*TC TAGAGGATCTGGTGGAGAAG*
N-terminal fusion	
1-N-mKate2-R1	AGTTCCAGCTTTAGCTGC*AGCTTCTCCACCAGATCCTCTTG A*ACGGTGTCCCAATTTACTAG
1-N-GFPdsm-R1	TAGTTCCAGCTTTAGCTGC*AGCTTCTCCACCAGATCCTCTT GA*CTTATAAAGCTCATCCATGC
1-N-GFPopt-R1	TAGTTCCAGCTTTAGCTGC*AGCTTCTCCACCAGATCCTCTT GA*TTATTTATACAATTCATCCATACCATGTG
2-N-mKate2-F	CTAGTCTAG*AGGAGG*CATATCCC**ATG**AGCTCAGAACTTATC AAGGAAAATATGCACATG

2-N-GFPdsm-F	CTAGTCTAG*AGGAGG*CATATCCC**ATG**TCAAAAGGAGAAGAGCTGTTCACAGGTGTTGTGC
2-N-GFPopt-F	CTAGTCTAG*AGGAGG*CATATCCC**ATG**GTTTCTAAAGGTGAAGAATTGTTTACAGG
1-N-Treminal-R2	TAGCGCGGCCGCACTGTCGAC*AGCTTCTCCACCAGATCCTCTTGA*
3-N-Treminal-R	TTAGCATGCTCGAGCCCGGGCTAGCGCGGCCGCACTGTC
Colony PCR	
Seq pUC Fw	GGAATTGTGAGCGGATAAC
Seq pUC Rv	AAGGCGATTAAGTTGGGTAACG

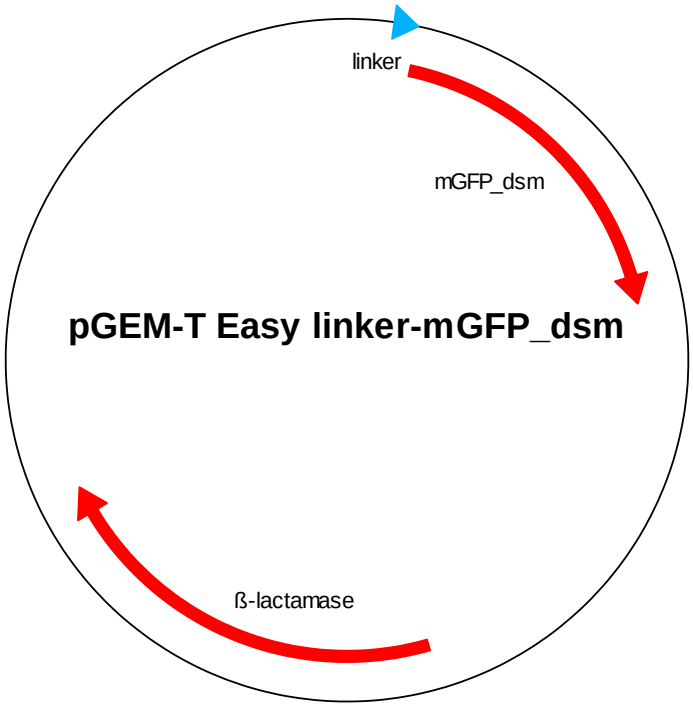
Appendix F

The schematic representation shows the orientation of the fluorescent protein gene mGFP-dsm within the experimental constructs obtained.

(*) N-terminal fusion construct containing mGFP-dsm.



(**) C-terminal fusion construct containing mGFP-dsm.



Appendix G

Table 2 *Escherichia coli* strains.

Species	Strain	Genotype
Escherichia coli	DH5α	F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK- mK+), λ–
Escherichia coli	BL21 (DE3)	F- ompT hsdSB(rB–, mB–) gal dcm

Table 3 Source of the fluorescent protein gene and the cloning vector.

Plasmid or FP source	Relevant features	Properties	Reference
mGFP-opt	Changes to *A. Victoria* GFP (GenBank: AAA27721.1): M1MV, S65A, V68L, S72A, A206K.	Monomeric (Zacharias et al. 2002)	Appendix A
mGFP-dsm	Changes to *A. Victoria* GFP (GenBank: AAA27721.1): S30R, Y39N, F64L, S65T, Q80R, F99S, N105T, Y145F, M153T,V163A, I171V, A206K.	Monomeric (Zacharias et al. 2002), superfolder (Pedelacq et al. 2006)	Appendix B
mKate2	Changes to TagRFP (GenBank: ABR08320.1): M1MS, V45A, R67K, N143S, M146T, S158A, F174L, H197R, K231R.	Monomeric (Shcherbo et al. 2009)	Appendix C
pGEM®-T Easy	TA cloning vector from Promega.	T-overhang	Appendix D

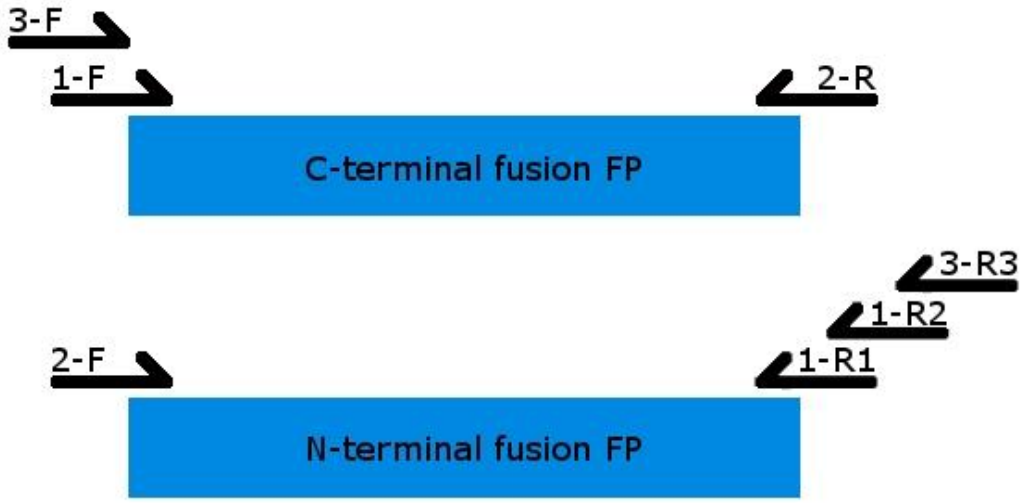


Figure 3 The sketch shows the PCR scheme. In the upper part the primers (arrows), which were used in order to obtain the C-terminal insert are shown. In the lower part the primers, which were used in order to obtain the N-terminal insert, are shown. The primers are labeled according to Appendix E. The final PCR products used for the C-terminal constructs required two PCR rounds, whereas the one for the N-terminal products required three PCR rounds.

Table 4 Colony PCR reaction protocol.

	Final concentration	Volume (µl)
Dream Taq buffer 5x	1x	2
dNTPs (10 mM)	1.25 mM	2.5
Seq pUC Fw (10 µM)	0.2 µM	0.4
Seq pUC Rv (10 µM)	0.2 µM	0.4
HomeTaq		1
H_2O		13.7
TOTAL VOLUME		20