

# **BBF RFC 64: Building Protein Domain Based Composite Biobricks for Mammalian Expression Systems**

Balint L. Balint, Ophir Keret, Peter Brazda, Mate Demeny and iGEM Debrecen-Hungary 2010 team.

October 26, 2010

## **1. Purpose**

The purpose of this RFC is to describe a method that allows the design of protein domain based parts, starting with gene centered information and translate these informations into BBF RFC 25 compatible part. The method is designed to be used in mammalian expression systems.

## **2. Relation to other BBF RFCs**

The parts described with this procedure SHALL be compatible with BBF RFC 25 and BBF RFC 10.

## **3. Copyright Notice**

Copyright (C) The BioBricks Foundation (2010). All Rights Reserved.

## **4. A method to design biological parts that correspond to protein domains and are compatible with RFC 25**

Domains are structural features of proteins that do not need to have sequence simillarities, but have functional simillarities. Domains of proteins represent functions of the proteins. By connecting domains one can connect functions in a modular way. The method described below is RECOMMENDED to be used for the design of mamalian domains that are planned to be used in mammalian expression vectors.

## **5. Detailed description of the design algorithm:**

In order to design an RFC 25 compatible domain it is RECOMMENDED to follow the following steps:

01. Identify the protein from which you would like to isolate a domain with the Uniprot ID:  
<http://www.uniprot.org> .
02. Use the Uniprot ID to start a domain search at the SMART webtool:  
<http://smart.embl-heidelberg.de> . Click into the search query „PFAM domains“ .
03. Proceed to data analysis by „Sequence SMART“ button. SMART should give a graphical representation of the identified domains. Click on the graphical representation of the domain you would like to extract and you will receive the peptide sequence of your domain. Copy, paste and save this sequence for further reference with a predefined identifier.
04. Reverse translate the amino acid sequence with the reverse translation tool at the <http://genedesign.thruhere.net/gd/> by selecting codon usage table „H. sapiens“ and pressing „Reverse translate“ button.

05. Remove with the „RE sites substraction“ button restriction sites named in RFC10 [1] and RFC25 [2] as being suggested to be removed as follows.: EcoRI, XbaI, SpeI, PstI, NotI, PvuII, XhoI, AvrII, NheI, SapI, AgeI, NgoMIV. See further details and exact sequences in the RFC 10 and RFC 25.
06. Copy, paste the designed sequence into a separate file with the appropriate predefined name.
07. Add the required prefixes and suffixes to your sequence in order to make them compatible with RFC 25.
08. RECOMMENDED N terminal prefix: „GAA TTC GCG GCC GCT TCT AGA TGG CCG GCG CCG CCG CCA **TGG** CC“. This sequence contains RFC25 prefix, Kozak sequence [3] and ATG in frame.
09. RECOMMENDED C terminal suffix is “**TAA TAA** ACC GGT TAA TAC TAG TAG CGG CCG CTG CAG”. This sequence contains two stop codons and the requested suffix. A RECOMMENDED eukaryotic polyA linker is **Bba\_K364302**. The PolyA tail of mRNA has multiple adenilates which is important for the nuclear export, translation and stability of mRNA in eukaryotes.
10. A RECOMMENDED eukaryotic tetracycline inducible promoter is the TRE-CMV promoter: **Bba\_K364304**.
11. If you would like to put expression or purification tags at the N or C terminal ends of your protein do not use the prefixes and suffixes suggested at point 08 and 09, but use the original RFC prefixes and suffixes as they follow:
  - RFC 25 prefix: “GAA TTC CGC GGC CGC TTC TAG ATG GCC GGC”
  - RFC 25 suffix: “ACC GGT TAA TAC TAG TAG CGG CCG CTG CAG”.

## 6. Experimental experience

The suffix mentioned at point 09 was tested by the iGEM 2010 Debrecen-Hungary team on several parts as being functional. This suffix **MUST** be on the C terminal end of a composite part. One **MAY** use the RFC 25 suffix described at point 12, which contains a single stop codon after the AgeI site for inserting the part as N- terminal or C- terminal domain of a protein.

## 7. Author’s Contact Information

Balint L. Balint ([balintblaszlo@yahoo.com](mailto:balintblaszlo@yahoo.com)) primary contact  
 Ophir Keret ([ophir\\_keret@hotmail.com](mailto:ophir_keret@hotmail.com))  
 Peter Brazda ([peter.brazda@gmail.com](mailto:peter.brazda@gmail.com))  
 Mate Demeny ([mate.demeny@gmail.com](mailto:mate.demeny@gmail.com))  
 iGEM team Debrecen-Hungary 2010 ([igem\\_2010\\_hungary@googlegroups.com](mailto:igem_2010_hungary@googlegroups.com))

## References

[1] BBF RFC10: <http://hdl.handle.net/1721.1/45138>, doi: 1721.1/45138

[2] BBF RFC 25: <http://hdl.handle.net/1721.1/45140>, doi:1721.1/45140

[3] Kozak M (October 1987). "An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs". *Nucleic Acids Res.* 15 (20): 8125–8148. doi:10.1093/nar/15.20.8125. PMID 3313277. PMC 306349. <http://nar.oxfordjournals.org/cgi/pmidlookup?view=long&pmid=3313277>

