

BBF RFC 25: Fusion Protein (Freiburg) Biobrick assembly standard

Kristian M. Müller, Katja M. Arndt, iGEM 2007 Team Freiburg, Raik Grünberg

31.3.2009

1. Purpose

This Request for Comments (RFC) describes an extension to the original BioBrick assembly standard (BBF RFC 10). The Fusion Assembly strategy described here is fully backwards compatible with RFC 10 and supports the construction of in frame fusion proteins.

2. Relation to other BBF RFCs

RFC 25 extends RFC 10. RFC 25 can be used instead of RFC 23.

3. Copyright Notice

Copyright (C) The BioBricks Foundation (2009). All Rights Reserved.

4. Motivation

The original BioBrick prefix and suffix (BBF RFC 10) with its easy cloning strategy is an excellent and universal way to combine various parts, e.g. promoter region, gene of interest, terminator etc. However, also individual proteins can be composed of different parts and the "mix-and-match" combination of fusion proteins is thus an important task in Synthetic Biology. However, the original assembly strategy does not allow for the cloning of protein fusions due to the generation of a stop codon at the SpeI/Xba scar.

According to RFC 10 the suffix of the first part (part in gray, PstI, NotI, SpeI):

...NNtactagtagcggccgctgcag

is combined with prefix of the second part (EcoRI, NotI, XbaI, part in gray):

gaattcgcggccgcttctagagNN...

resulting in the SpeI/Xba combination/scar:

...NNtactagagNN...

This scar encodes for the amino acids Tyr (codon: tac), STOP (codon: tag) and Ser or Arg (codon agn).

Consequently, according to RFC 10 proteins can only be designed as one part but not in a modular fashion. This request for Comments proposes a backward compatible extension to RFC 10 which allows for the construction of fusion proteins from modular BioBrick "FusionParts".

5. Detailed Description

5.1 Formal requirements

FusionParts MUST have the following **Prefix**:

gaattcgcggccgcttctagatggccggcNN...

EcoRI, NotI, XbaI, NgoMIV, part in gray. The underlined sequence corresponds to the original (RFC 10) BioBrick prefix for protein coding sequences (i.e. 'expression parts').

Fusion Parts MUST have the following **Suffix**:

...NNaccggttaatactagtagcggccgctgcag

part in gray, AgeI, SpeI, NotI, PstI. The underlined sequence corresponds to the original (RFC 10) BioBrick suffix.

According to RFC 25, a complete FusionPart has thus the following format (part sequence is given as nnnnnn):

```
EcoRI   NotI   XbaI   NgoMIV           AgeI   SpeI   NotI   PstI
|       |       |       |               |       |       |       |
GAATTCgcggccgctTCTAGAtgGCCGGCnnnnnnACCGGTtaataACTAGTagcggccgCTGCAG
5'-----+-----+-----+-----+-----+-----+-----+-----+-----3'
CTTAAGcgccggcgaAGATCTacCGGCCGnnnnnnTGGCCAattaTGATCAtcgccggcGACGTC
c      I R G R F * M A G ? ? T G * Y * * R P L Q
```

Fusion parts MUST NOT contain recognition sites for AgeI (ACCGGT) or NgoMIV (GCCGGC) anywhere in their sequence (outside prefix and suffix). The previous restrictions of RFC 10 apply as well. That means, fusion parts, furthermore, MUST NOT contain EcoRI, XbaI, SpeI, or PstI restriction sites. It is RECOMMENDED to also remove further restriction sites. See RFC 10 for details.

5.2 Application of the FusionPart Standard

Fusion parts can be used like standard Biobrick parts according to RFC 10. In addition, fusion proteins can be constructed using the restriction sites NgoMIV and AgeI following the same cloning principles. The NgoMIV and AgeI recognition sites are placed in frame with the protein coding sequence. The NgoMIV overhang (G'CCGG,C) is compatible with the AgeI overhang (A'CCGG,T) and ligation results in the scar ACCGCC coding for Thr and Gly.

For example, in order to fuse Part A N-terminally to part B, in the classic serial strategy, Part A SHOULD be digested with EcoRI and AgeI, and the vector containing Part B SHOULD be digested with EcoRI and NgoMIV. Ligation of the Part A into the vector B results in the FusionPart AB.

The 3A Assembly strategy is also supported. In this case, Part A SHOULD be digested with EcoRI and AgeI and Part B SHOULD be digested with NgoMIV and PstI and the construction vector SHOULD be digested with EcoRI and PstI followed by triple ligation. In the latter case, OPTIONALLY, either EcoRI can be replaced by XbaI or PstI by SpeI.

5.3 N-part

Some protein parts do not tolerate modifications of their N-terminus. However, combining Fusion Parts with existing RFC 10 RBS parts adds the sequence Met-Ala-Gly to the N-terminus of the protein. In cases where the native N-terminus needs to be preserved, FusionParts MAY be constructed in a hybrid “N-part” format. The Fusion N-part Prefix is identical to the BioBrick RFC 10 prefix for coding sequences (EcoRI, NotI, XbaI, part in gray):

```
gaattcgcggccgcttctagATG...
```

An N-part MUST have the following format:

```
EcoRI   NotI   XbaI           AgeI           SpeI           NotI           PstI
|       |       |           |           |           |           |
GAATTCgcggccgctTCTAGAtgnnnnnnACCGGTtaatACTAGTagcggcgcCTGCAG
-----+-----+-----+-----+-----+-----+-----+-----+-----+
CTTAAGcgcggcgaAGATCTacnnnnnnTGGCCAattaTGATCATcgccggcGACGTC
  I  R  G  R  F  *  M  ?  ?  T  G  *  Y  *  *  R  P  L  Q
```

The underlined sequence corresponds to the RFC 10 proposal.

The classic RFC 10 assembly is used to combine Fusion N-parts with RFC 10 compliant

RBS parts. That means, the N-part SHOULD be digested with XbaI + PstI, and the RBS part with EcoRI and SpeI. FusionParts can be fused to the N-part by digesting the N-part with AgeI/SpeI and the FusionPart with NgoMIV/SpeI. Any number of FusionParts can be combined and optionally fused to the N-part.

6. Discussion

The Fusion part format meets the following design criteria:

(1) It enables protein fusions.

(2) RFC 25 is an extension rather than modification of RFC 10. Both FusionParts and Fusion N-parts are fully compatible with standard (RFC 10) BioBricks. They feature the RFC 10 BioBrick prefix for coding sequences and the standard BioBrick suffix. A FusionPart is equivalent to a RFC 10 stand-alone protein part as it has a start codon after the XbaI site (as defined for RFC 10 coding sequences) with two additional amino acids (Ala-Gly) encoded before the sequence of the protein of interest. The Fusion suffix sequence contains 3 stop codons to prevent accidental read-through.

(2) These fusion constructs are biochemically well behaved – the “scar sequence” between protein parts does not contain a frame-shift or stop codon. It does not contain any codon that is rarely used in E. coli. It translates to the sequence Thr-Gly, both of which are small uncharged amino acids that are often found in natural and synthetic linker peptides.

(3) The Ala-Gly sequence coded by the NgoMIV site satisfies the N-end rule [1]. That means that that this sequence does not promote degradation of the protein. Transcription start Biobricks (for example, RBS+start codon) can therefore also be designed as FusionParts and assembled without switching back to RFC 10 assembly. This improves over the BioFusion format described in BBF RFC 23. (4) Where needed, RFC 25 allows for preserving a protein’s natural N-terminus. This can be important for some proteins..

(4) The Fusion format is compatible with the established BioBrick cloning protocols. All enzymes used can be inactivated by heating.

(5) In addition, the authors provide 3A (three antibiotic resistance) construction vectors that adhere to the presented format and allow for three fragment ligations with selection for insertion.

RFC 25 improves over RFC 23, in particular, regarding scar composition, N-end rule safety, and compatibility. This comes at the expense of two additional restriction sites which lengthen prefix and suffix and put additional constraints on part sequences.

There are different application scenarios for the Fusion Biobrick format. “Casual” protein

engineers can first assemble their full-length protein from FusionParts using the Fusion assembly, and then switch back to the classic RFC 10 assembly in order to embed this synthetic protein into larger circuits. In this way, novel protein fusions can be combined with the large set of transcription regulators or other devices that are already available in the Registry of Standard Biological Parts. Note though, that the assembly **MUST** be finished with RFC 10 enzymes, once the switch has been made.

Users with a focus on protein engineering may prefer to construct all of their parts, including non-coding ones, in Fusion format. This avoids the need to switch assembly strategies.

BioFusion parts (RFC 23) can be coupled to FusionParts as long as we are dealing with independent proteins or non-coding elements. On the other hand, it is not possible to create protein fusions directly from the two different formats. However, BBF RFC 24 describes an indirect strategy that involves one additional sub-cloning step.

Experience

At the time of writing, the Fusion format has been thoroughly tested by two student teams [2, 3] as well as several more senior investigators in different labs. Fusion parts perform robustly both in classic as well as in 3A assemblies without gel-purification. Customized protocols are available on OpenWetware.org [4].

7. Parts and Materials

As of 31.3.2009 the following construction vectors are fully compatible with the FusionPart proposal. Many other vectors work as well.

- [pSB1AC3F \(BBa_J18901\)](#)
- [pSB1AK3F \(BBa_J18902\)](#)
- [pSB1AT3F \(BBa_J18903\)](#)

Several FusionParts in the registry are shipped in the following vector (provided by a gene synthesis company).

- [pMA \(BBa_K157000\)](#)

Note: 3A assemblies starting from parts in this vector backbone show consistently low yields that require extensive screening of transformants. The reason for this remains unclear. We advise to sub-clone parts into pSB1* vectors before channeling them into 3A assemblies or to gel purify the insert before ligation.

The following restriction enzymes are used:

NgoMIV (also named NgoMI)

- As of spring 2009 this enzyme is available from the companies New England Biolabs and Promega
- Isoschizomers are NaeI, MroNI, NgoAIV, and PdiI.

- Activity in NEB buffers as of spring 2009: 1 - 100%, 2 - 50%, 3 – 10%, 4 – 100%.
- Heat Inactivation: 80°C for 20 minutes
- Methylation sensitivity: not sensitive to dam methylation and dcm methylation, blocked by CpG methylation.
- Number of recognition sites in DNA molecules are (adapted from Fermentas):

Lambda	PhiX 174	M13 mp18/19	pBR 322	pUC 18/19	pUC 57	pTZ19R/U	pBluescript IKS(-/+)	pBluescript IISK(-/+)	pACYC 177	pACYC 184
1	0	1	4	0	0	1	1	1	0	5

AgeI

- As of spring 2009 this enzyme is available from New England Biolabs, Nippon Gene Co, and Promega.
- Isoschizomers are AsiAI, AsiGI, BshTI, CspAI, PinAI
- Activity in NEB buffers as of spring 2009: 1 - 100%, 2 - 50%, 3 – 10%, 4 – 75%.
- Methylation sensitivity: not sensitive to dam methylation and dcm methylation, blocked by CpG methylation.
- Heat inactivation: 65°C for 20 minutes
- Number of recognition sites in DNA molecules are (adapted from Fermentas):

Lambda	PhiX 174	M13 mp18/19	pBR 322	pUC 18/19	pUC 57	pTZ19R/U	pBluescript IKS(-/+)	pBluescript IISK(-/+)	pACYC 177	pACYC 184
13	0	0	0	0	0	0	0	0	2	4

8. Author's Contact Information

Dr. Kristian M. Müller
 Assistant Professor for Molecular Biology and Protein Engineering
 Institute for Biology III
 University of Freiburg
 Schaezlestr. 1
 D-79104 Freiburg
 Germany
 e-mail kristian@biologie.uni-freiburg.de
 Tel. ++49-(0)761-203-2748 (office)
 Tel. ++49-(0)761-203-2763 (lab)
 Fax. ++49-(0)761-203-2745

Dr. Katja M. Arndt
 Fellow of the Freiburg Institute for Advanced Studies
 Institute for Biology III
 University of Freiburg
 Schaezlestr. 1

D-79104 Freiburg
Germany
email Katja@biologie.uni-freiburg.de
Tel. ++49-(0)761-203-2748 (office)

Dr. Raik Grünberg
EMBL-CRG Systems Biology Unit
CRG – Centre de Regulacio Genomica
Dr. Aiguader 88
08003 Barcelona
e-mail raik.gruenberg@crg.es

References

[1] Varshavsky A "The N-end rule: functions, mysteries, uses." Proc Natl Acad Sci U S A. 1996 Oct 29;93(22):12142-9.

[2] Original Proposal for Fusion parts at iGEM 2007:
<http://parts.mit.edu/igem07/index.php/Freiburg> .

[3] Use of the standard for synthetic receptors at iGEM 2008:
<http://2008.igem.org/Team:Freiburg> .

[4] The PrbbBioBrick community: Biobrick assembly protocols:
<http://openwetware.org/wiki/PrbbBB:Protocols> .