BBF RFC 44 - Bioscaffold-Linker

Mina Petros & Savery Nigel

18 October 2009

1. Purpose

Protein fusions are currently not supported by most assembly standards. This assembly standard aims to provide the basis of a solution to a biobrick-friendly protein fusion mechanism that supports the current favorite assembly standard 10 (RFC10) and hence allows utilization of most of the current biobricks database.

2. Relation to other BBF RFCs

BBF RFC10 & BBF RFC15. This RFC does not replace the aforementioned standards. It simply describes practices on how to achieve protein fusions that are consistent with RFC10 and draws inspiration and updates on the notion of RFC15.

3. Copyright Notice

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

4.1 Introduction

4.1.1 RFC10 and the protein fusion problem

There are various assembly standards that can be used to assemble biobrick parts together but the current favourite and most promoted is "Assembly Standard 10" (RFC10). In assembly standard 10 each biobrick is surrounded immediately upstream and downstream by strings of DNA known as a prefix (upstream) and suffix (downstream) each carrying DNA sequences recognised by two separate restriction enzymes, known as restriction sites. The prefix has an EcoRI and an XbaI restriction site, whilst the suffix has a SpeI and PstI site. This allows the easy removal of a part and its subsequent placement next to another part, either upstream or downstream, as one prefix restriction site (XbaI) creates compatible ends with a suffix restriction site (SpeI), after the DNA has been cleaved by the respective enzymes [1,2].

RFC10 regulations do not support protein fusions. Under RFC10 regulations when two protein encoding biobricks are joined, i.e. when the XbaI site ligates with SpeI the

resulting sequence, known as DNA scar, is 5- TACTAG-3. This poses a problem as TAG encodes for a stop codon. The problem however is not just that the DNA scar encodes for a stop codon but in addition each protein coding biobrick has to have two consecutive stop codons 5-TAA TAA-3 at the C-terminal tail [1,2].

4.1.2 Other Assembly Standards

No current assembly standard satisfactorily addresses the problem of protein fusion. Assembly standard 23 removes a base pair from the prefix and suffix respectively giving a 5-ACT AGA-3 scar, which replaces the DNA scar stop codon but instead inserts a problematic Arginine residue that is highly positively charged and also not preferred by the E.coli translational machinery. Assembly standard 25 adds an extra restriction consensus prior to the start codon of the upstream protein and this would upset the 5-6 base pair spacer rule that should exist between the RBS and the protein to ensure that translation is optimal [1]. As most biobricks in the database currently have two TAA stop codons at their 3' end, the problem remains even when the DNA scar is manipulated as it is the nature of protein encoding biobricks to carry stop codons.

A viable solution is offered partly by assembly standard 15 (RFC 15) [3], a concept known as a Bioscaffold. A bioscaffold is a biobrick that is placed between two protein coding biobricks. The bioscaffold carries consensus sequences recognised by atypical Type II enzymes that cleave DNA outside their recognition consensus with various mechanisms. This enables the reach beyond the consensus and the possibility of cleaving away the stop codon sequences of the protein coding biobricks. Although some notion of Bioscaffolds is provided by iGEM, the current scaffolds are not compatible as biobricks or not available in biobrick format and each assembly has to be carefully considered on a case by case basis, which is in conflict with the iGEM ethos of standardisation. What is required is an RFC10 upwards compatible Bioscaffold that is so versatile that can be used with almost all biobricks without extra considerations.

4.2 The Bioscaffold-Linker overview

As figure 1 illustrates, the Bioscaffold-Linker shows a number of features. On the left and right hand sides two unique restriction sites are found. These sequences are recognised by non-conventional Type-IIS enzymes that digest DNA outside of these recognition sequences. On application they collapse the Bioscaffold and at the same time convert the upstream protein coding part stop codons into a single amino acid and maintain the Methionine start codon of the downstream protein coding part. It also adds a flexible Glycine-Serine-Glycine linker between the two protein coding regions, allowing each individual component to fold correctly and independently following translation. In addition a translational buffer exists within the Bioscaffold. A protein translational buffer is a series of stop codons found in different frames such that if the application of the Bioscaffold fails then the downstream protein coding region will not be expressed. It is important to note that everything is kept in frame in the DNA sequence after the application of the Bioscaffold.
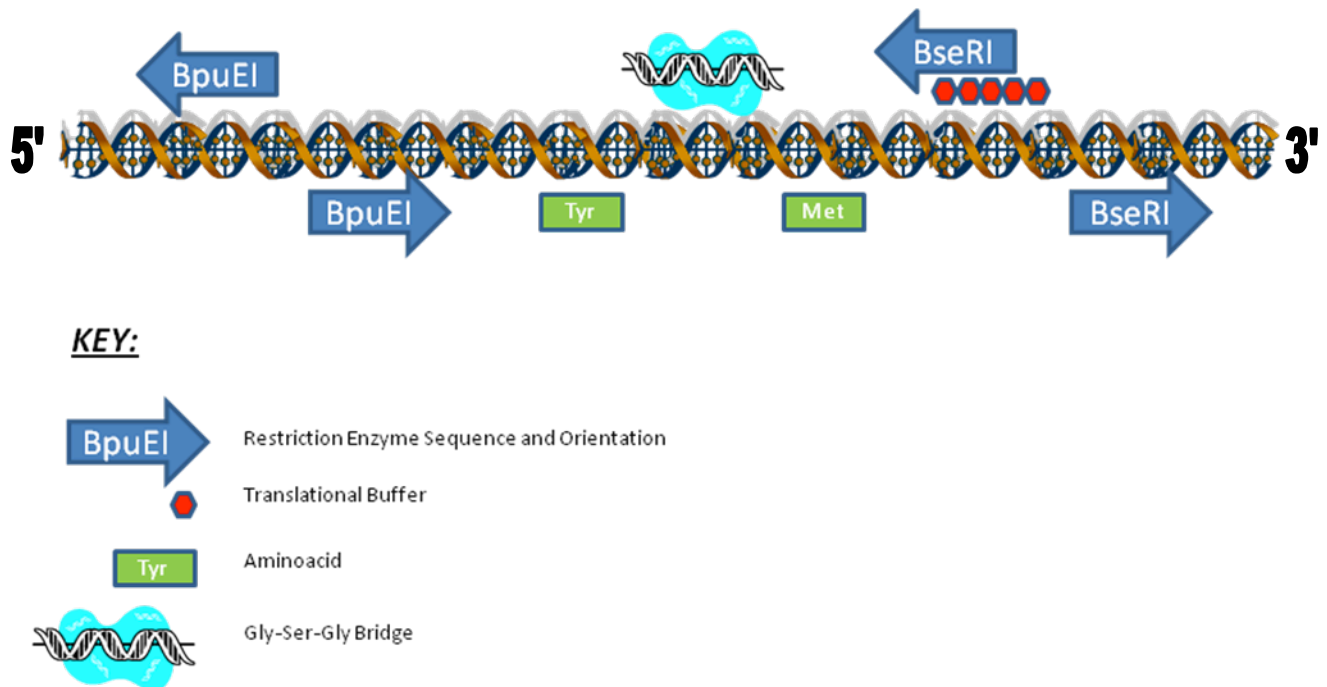
Figure 1: The structure of the Bioscaffold-Linker illustrating its features. The arrowhead pointing to the right indicates forward orientation and the left arrow head reverse orientation.

4.3 The Bioscaffold-Linker Specific Enzymes.

The Bioscaffold specific enzymes BpuEI and BseRI were chosen with a number of considerations in mind. BpuEI has an asymmetric six base pair recognition sequence, CTTGAG. Following from basic probability theory this means that the sequence occurs at random in approximately every $4^6/2$ base pairs (2048 base pairs; once every 4096bases on each strand of DNA). However the site frequency in the *E.coli* genome is even less than half the expected [4]. BseRI has a symmetric 6 base pair recognition sequence, GAGGAG, which means that the site appears every 4^6 at random (4096 base pairs) but the site frequency in the E.coli genome is about 1/4 of that expected [5]. Thus such sequences appear scarcely in the genome and can be used with most of the genes in the Biobrick database. Also neither sequence appears on the backbone plasmid pSB2K3 and hence this plasmid may be chosen as the preferred backbone for protein fusions using the Bioscaffold Linker. Many more enzymes have similar properties but are more expensive to buy and sometimes not commercially available.

Another consideration is that atypical Type II enzymes are very sensitive to buffer conditions and to the context that their recognition site is found. Such enzymes include CspCI (Type IIB) to which a Bioscaffold-Linker has also been developed but could not be tested as no assembly standard 10 compatible plasmid backbone is yet available (see application section later). A potential advantage of CspCI is that it has an asymmetric 7bp recognition sequence, hence it appears at random every $4^7/2$ bp (8192bp). A potential disadvantage is that CspCI has 1 base pair discrepancy in restricting DNA, base on the context of the recognition sequence, but the CspCI bioscaffold has been designed taking into consideration this. All three chosen enzymes have been recognised as "time-saver

qualified" by NEB and one unit can digest 1µg of DNA in 5min, saving time in the Bioscaffold application.

4.4 Bioscaffold-Linker Application

4.4.1 Required parts

The following parts MUST be used when applying the Bioscaffold-Linker:

1. *Plasmid Backbone:* pSB2K3 (hereafter called the plasmid backbone).
2. *Bioscaffold-Linker:* BBa_K259002 (hereafter called the Bioscaffold).
3. *Upstream Protein:* The protein encoding gene MUST be compatible with RFC10.
4. *Downstream Protein:* The protein encoding gene MUST be compatible with RFC10.
5. *Bioscaffold Specific Enzymes:* Currently BpuEI and BseRI. Supplied from NEB and used under NEB guidelines. Avoid incubation times greater than one hour (recommended times 5-50 min).

The usage of pSB2K3 is essential as it has no Bioscaffold conflicting sequences with respect to the restriction enzymes used. Also the downstream/upstream protein encoding genes MUST NOT carry the recognition consensus of the Bioscaffold specific enzymes.

Also standard assembly refers to the assembly created with RFC10 rules.

4.4.2 Bioscaffold-Linker Application

The application of the Bioscaffold is straightforward. Firstly a standard assembly is created on a pSB2K3 plasmid backbone with the Bioscaffold in the middle of two protein coding parts as shown in figure 2. Then Bioscaffold specific enzymes are used to collapse the Bioscaffold, creating scarless fusions.
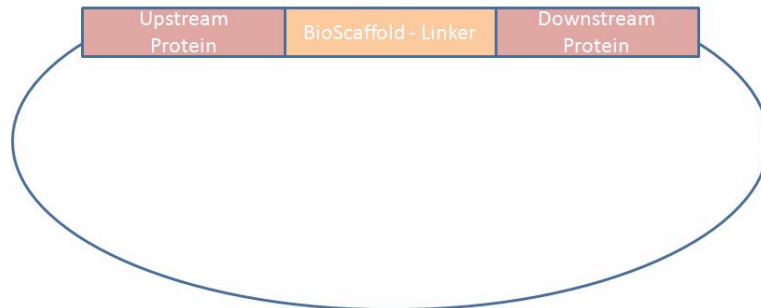
Figure 2: The first step in applying the Bioscaffold is to create an assembly standard 10 where the bioscaffold is placed between two protein encoding genes. The destination plasmid must be pSB2K3.

The left hand side features a Type-IIS enzyme recognition sequence, BpuEI, that will cleave a DNA sequence a few base pairs after the recognition site. Two such sites have been incorporated in inverse orientation, allowing cleavage that removes the stop codons from the upstream part and a small sequence from the Bioscaffold. Complementary overhangs are created between the partly cleaved stop codon and the Bioscaffold sequence, which when ligated convert the stop codon to a Tyrosine aminoacid. The first step in applying the Bioscaffold after it has been incorporated between two protein encoding genes is to restrict with BpuEI and ligate the restricted product.

On the right hand site there exist a different Type IIS enzyme recognition sequence, BseRI, again in two copies with inverse orientation. BseRI cleaves both upstream and downstream from the sequence and in doing so collapses the Bioscaffold and creates complementary DNA sequence ends between the Bioscaffold and the downstream coding gene that upon ligation will give a Methionine start codon (see figure 3). As the Met start codon is maintained this leads to a scarless downstream fusion. Hence for correct usage of the scaffold after the first round of restrictrion and ligation with BpuEI and ligase enzyme is to incubate the product with BseRI, and ligate the complementary ends.

In the middle of the Bioscaffold sequence there exists a region encoding a Gly-Ser-Gly linker bridge. This is maintained if the application of the Bioscaffold is done in the correct order as outlined in the two paragraphs above.

In case the Bioscaffold has failed to perform as expected the translational buffer (for explanation see section 4.2) will prevent the translational machinery of the cell translating the region encoding the downstream protein.
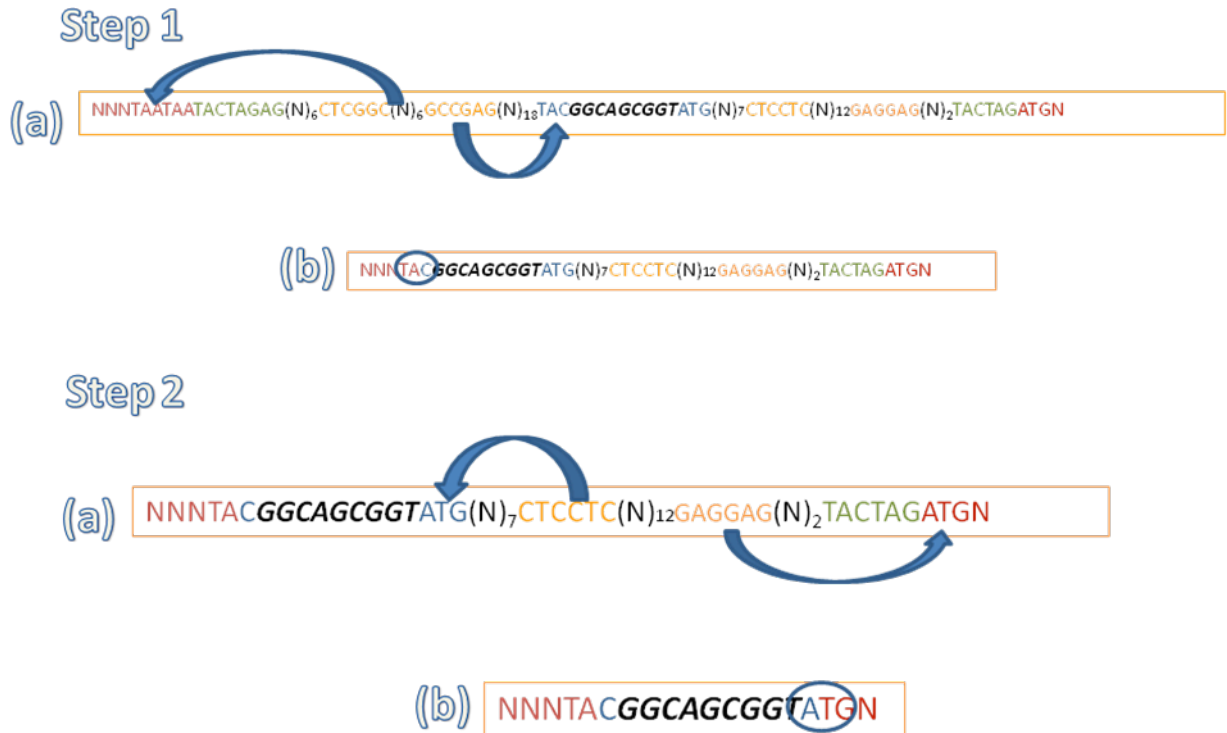
Figure 3: The manipulation of the Bioscaffold in order to create a protein fusion. The sequences in red are protein encoding genes. The sequences in orange are the Bioscaffold specific restriction sites and the sequences in blue are aminoacids within the bioscaffold involved in restriction/ligation. The arrows indicate the reach of the restriction enzymes in each step. In Step 1 part (a) we illustrate the reach of BpuEI and in (b) the subsequent product after ligation giving a TAC codon. In Step 2 part (a) we illustrate the reach of BseRI restriction enzyme and in (b) the product after ligation giving an ATG codon.

4.4.3 Recommended Assembly

Although 3-way ligation is promoted by the iGEM organisation we recommend that when creating an assembly of the kind *5'-Upstream Protein-Bioscaffold-Downstream Protein-3'* you use a 2-way ligation and take advantage of the fact that the Bioscaffold is already found on a pSB2K3 backbone plasmid. As a first step to insert a biobrick upstream of the bioscaffold digest both parts with EcoRI and XbaI and then proceed with a ligation reaction. Secondly, to insert a biobrick part downstream of the Bioscaffold digest both parts with SpeI and PstI and then proceed with a ligation reaction. This sequence of events maximises the correct ligation products to be obtained with respect to a more complex 3-way ligation reaction that will need to be carried out twice (once to obtain *5'-Upstream Protein-Bioscaffold-3'*, and once to obtain *5'-Upstream Protein-Bioscaffold-Downstream Protein-3'*). Do make sure that after each restriction and ligation enzyme denaturation and DNA clean-up is performed to remove any enzymes that might impede on subsequent steps.

4.5 Associated Biobricks developed for Bioscaffold-Linker

The following biobricks have been designed and developed to be used as Bioscaffold-Linkers for protein fusions:

- K259002 - BpuEI BseRI Bioscaffold
- K259004 - BpuEI CspCI Standard Bioscaffold
- K259009 - BpuEI CspCI Version 2 Bioscaffold
- K259010- BpuEI CspCI Version 3 Bioscaffold

4.6 Requests for future updates on plasmid backbones

The authors kindly request the iGEM team to consider mutating the single CspCI site within the pSB2K3 plasmid backbone. As CspCI has a 7bp asymmetric recognition sequence this means that it can provide a very good alternative bioscaffold recognition sequence in cases where conflicts with the protein encoding genes occur.

5. Author's Contact Information

Petros Mina - petros.mina.08@bristol.ac.uk

References

[1] BBF RFC Technical standards, http://bbf.openwetware.org/Standards/Technical/Formats.html, True on October 2009.

[2]BBF RFC10, http://openwetware.org/wiki/The_BioBricks_Foundation:BBFRFC10, True on October 2009.

[3]BBF RFC15, http://openwetware.org/wiki/The_BioBricks_Foundation:BBFRFC15, True on October 2009.

[4]Cut Frequency of CTTGAG, http://tools.neb.com/~vincze/cutfreq/CTTGAG.html, True on October 2009.

[4]Cut Frequency of GAGGAG, http://tools.neb.com/~vincze/cutfreq/GAGGAG.html, True on October 2009.