BBF RFC 81: "BioSandwich", a homology-based assembly method using a library of standard parts

Yassen Abbas, Allan Crossman, Eugene Fletcher, Chris French, Clare Gibson,
Sylvia Ispasanie, Lukasz Kopec, Mun Ching Lee, Di Li, Fionn Tynan-O'Mahony

September 17, 2011

1. Purpose

This Request for Comments (RFC) describes a strategy for using homology-based assembly methods to assemble parts from a library in any order.

2. Relation to other BBF RFCs

BBF RFC 81 does not update or replace any earlier BBF RFC. It is compatible with most RFC10 parts, with the added requirement that they not contain BglII sites ("agatct"). RFC10 parts can be converted to BioSandwich parts by PCR with a forward primer that adds a BglII site.

3. Copyright Notice

4. Background

One of the main aspects of synthetic biology is the assembly of standard modular parts from a library to produce larger constructs. There are two main types of assembly strategy: *hierarchical* and *homology-based* (Ellis *et al*, 2010).

The RFC10 format, which is standard for iGEM projects, is based on hierarchical assembly methods such as Standard Assembly and Three Antibiotic Assembly. The main advantage of RFC10 is flexibility: any combination of parts can be assembled in any order. The main disadvantage is speed.

Many iGEM teams now use homology-based assembly strategies, such as Gibson Assembly. These methods use PCR to create parts that have homologous ends. They are then assembled using the method of Gibson *et al* (2009). This method is relatively fast. However, the order of the parts in the final construct is entirely determined by the primers chosen for the initial PCR reactions. If a different order is desired, parts have to be remade with new PCR primers.

BioSandwich is a hybrid that combines many of the benefits of hierarchical assembly and homology-based assembly. In particular, it is useful for creating different combinations of the same parts in different orders.

5. Outline

BioSandwich parts are reusable because they come in a standard format (much like ordinary BioBricks) with restriction sites flanking the part. The restriction sites are BglII ("agatct") in the prefix, and SpeI ("actagt") in the suffix. However, these restriction sites are not used directly for assembly; instead they are used to attach short (~35 bp) oligonucleotides (hereafter "spacers"). These spacers serve two purposes:

- They create homology between the end of one part and the start of another; this allows homology-based assembly.

- They can incorporate short meaningful sequences such as ribosome binding sites, linkers for fusion proteins, etc.

We note that many assemblies are, or could be, made using an alternating pattern of short and long sequences; for example, ribosome binding sites are placed between complete genes; or linker sequences are placed between domains to make a fusion protein. BioSandwich is suitable for such assemblies, whereas many other homology-based assembly methods have difficulty with small parts.

Any lab using BioSandwich will want to keep a small library of different spacers for different purposes. Once (carefully chosen) spacers have been attached to each part, homology-based assembly can be carried out in a single reaction. Precisely which spacers have been attached to which parts will determine the order of the parts in the final assembly.

6. Formats

6.1. Normal Parts

Parts MUST be free of internal BglII restriction sites ("agatct") and SpeI sites ("actagt"). Each part MUST be made with a BglII site at the start, and a SpeI site at the end. Parts SHOULD be made and stored as PCR products rather than minipreps.

For compliance with RFC10, parts SHOULD also be free of EcoRI sites ("gaattc"), XbaI sites ("tctaga") and PstI sites ("ctgcag").

For RFC10 compliance, the full part format becomes:

```
    gaattcgcggccgcttctagagatct NNN [...] NNt actagtagcggccgctgcag
```

After cutting with BglII and SpeI, we have the following (shown in frame, which is relevant for fusion proteins):

```
            5'     ga tct NNN [...] NNt a          3'
            3'         a nnn [...] nna tga tc    5'
```

When a part is used for fusion proteins, the "t" base at the start of the RFC10 suffix becomes the final base of the final codon in the part. Design of the part SHOULD take this into account. This

MAY be done by making this final codon "ggt", coding for an innocuous glycine residue. (If a fusion protein is not being made, no such consideration is needed. If RFC10-compliance is not required, this "t" base MAY be replaced with anything.)

6.2. Normal Spacers

Spacers are designed as a set of oligonucleotides that can be attached to any part via the part's restriction sites. Each type of spacer will be attached upstream of one part and downstream of another. Users MUST create three different single-stranded oligonucleotides (spacer oligos or "spoligos") for each spacer so as to allow annealing and ligation at the restriction sites. They SHOULD be designed so that the spacer's non-ligating ends are blunt.

The format for the upstream spacers is as follows (both strands shown):

```
5'    ct agc NNN [...] NNN g           3'           (forward spoligo)
3'    ga tcg nnn [...] nnn cct ag    5'       (reverse spoligo ONE)
```

The format for the downstream spacers is:

```
5'    ct agc NNN [...] NNN g           3'           (forward spoligo)
3'         g nnn [...] nnn c           5'       (reverse spoligo TWO)
```

Note that the forward spoligo is used for both upstream and downstream attachment.

The content of different spacers MUST be varied to avoid homology. It is RECOMMENDED that non-coding spacers have an in-frame stop codon, in case they are to be used after a coding part which lacks one. Non-coding spacers MAY be designed to contain useful sequences such as ribosome binding sites.

6.3. Vector Part

A vector MUST be made as a PCR product with a BglII site ("agatct") at the 5' end and an XbaI site ("tctaga") at the 3' end. Since XbaI produces sticky-ends compatible with SpeI, the vector is compatible with standard spacers.

If one wishes the final products (after insertion into the vector) to be RFC10-compliant, the format for a vector SHOULD be as follows:

```
agatct [...] tactagtagcggccgctgcag [ori, cmlR, etc] gaattcgcggccgcttctaga
```

Note that a few additional bases MUST also be present at the 5' and 3' ends to give the restriction enzymes space to work. When spacers are to be annealed, the vector MUST be cut with BglII and XbaI (not SpeI).

6.4. Start Spacer

For compliance with RFC10, the spacer that connects the vector to the first part SHOULD be in the following format:

Spoligos attaching upstream of first part:

```
5'      ctagag NNN [...] NNN g        3'            (forward spoligo)
3'      gatctc nnn [...] nnn cctag    5'        (reverse spoligo ONE)
```

Spoligos attaching downstream of vector:

```
5'      ctagag NNN [...] NNN g        3'            (forward spoligo)
3'          tc nnn [...] nnn c        5'        (reverse spoligo TWO)
```

This format ensures that the final product contains a standard RFC10 prefix and suffix.

7. Assembly

7.1. Digestion of Parts and Vector

Each part MUST be digested separately.

- Parts: digest with BglII and SpeI in NEB Buffer 2
- Vector: digest with BglII and XbaI in NEB Buffer 2 or 3

As an example protocol, each digestion reaction might contain:

```
36 uL      Water
 5 uL      DNA
 5 uL      Buffer
 2 uL      Enzyme 1
 2 uL      Enzyme 2
```

In this example, the digestions could be left for 2 hours at 37 °C. Afterwards they would be purified (e.g. with 5 uL glass beads, and eluted to 10 uL elution buffer).

7.2. Ligation of Parts and Vector to Spacers

The parts and vector MUST now be ligated to the correct spacers.

As an example protocol, each ligation reaction might contain:

```
10 uL      Water
 5 uL      DNA
 1 uL      Spacer pair #1
               (upstream spacer forward spoligo +
                upstream spacer reverse spoligo ONE)
 1 uL      Spacer pair #2
               (downstream spacer forward spoligo +
                downstream spacer reverse spoligo TWO)
 2 uL      Ligase buffer
 1 uL      T4 DNA ligase
```

In this example, reactions could be left for 9 hours at 16 °C. Afterwards they would be purified.

7.3. Homology-Based Assembly

The parts MUST now be assembled using a homology-based assembly method.

Users can choose from a number of different methods, including Gibson Assembly (Gibson *et al*, 2009); Overlap Extension PCR (Horton *et al*, 1989); Circular Polymerase Extension Cloning (Quan *et al*, 2009); and others.

8. Authors' Contact Information

Yassen Abbas:	`s0817711@sms.ed.ac.uk`
Allan Crossman:	`igem@nothinginbiology.com`
Eugene Fletcher:	`E.K.A.Fletcher@sms.ed.ac.uk`
Chris French:	`C.French@ed.ac.uk`
Clare Gibson:	`claregibson58@gmail.com`
Sylvia Ispasanie:	`s0834270@sms.ed.ac.uk`
Lukasz Kopec:	`L.Kopec@sms.ed.ac.uk`
Mun Ching Lee:	`leemunching@gmail.com`
Di Li:	`D.li-7@sms.ed.ac.uk`
Fionn Tynan-O'Mahony:	`fionntom@yahoo.ie`

References

Ellis T, Adie T, Baldwin GS (2010) DNA assembly for synthetic biology: from parts to pathways and beyond. *Integrative Biology* **3**: 109-118 (doi: 10.1039/c0ib00070a).

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**: 343-345 (doi: 10.1038/nmeth.1318).

Horton R, Hunt H, Ho S, Pullen J, Pease L (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* **77**(1): 61-68. (doi: 10.1016/0378-1119(89)90359-4).

Quan J, Tian J (2009) Circular Polymerase Extension Cloning of complex gene libraries and pathways. *PLoS ONE* **4**(7): e6441 (doi: 10.1371/journal.pone.0006441).