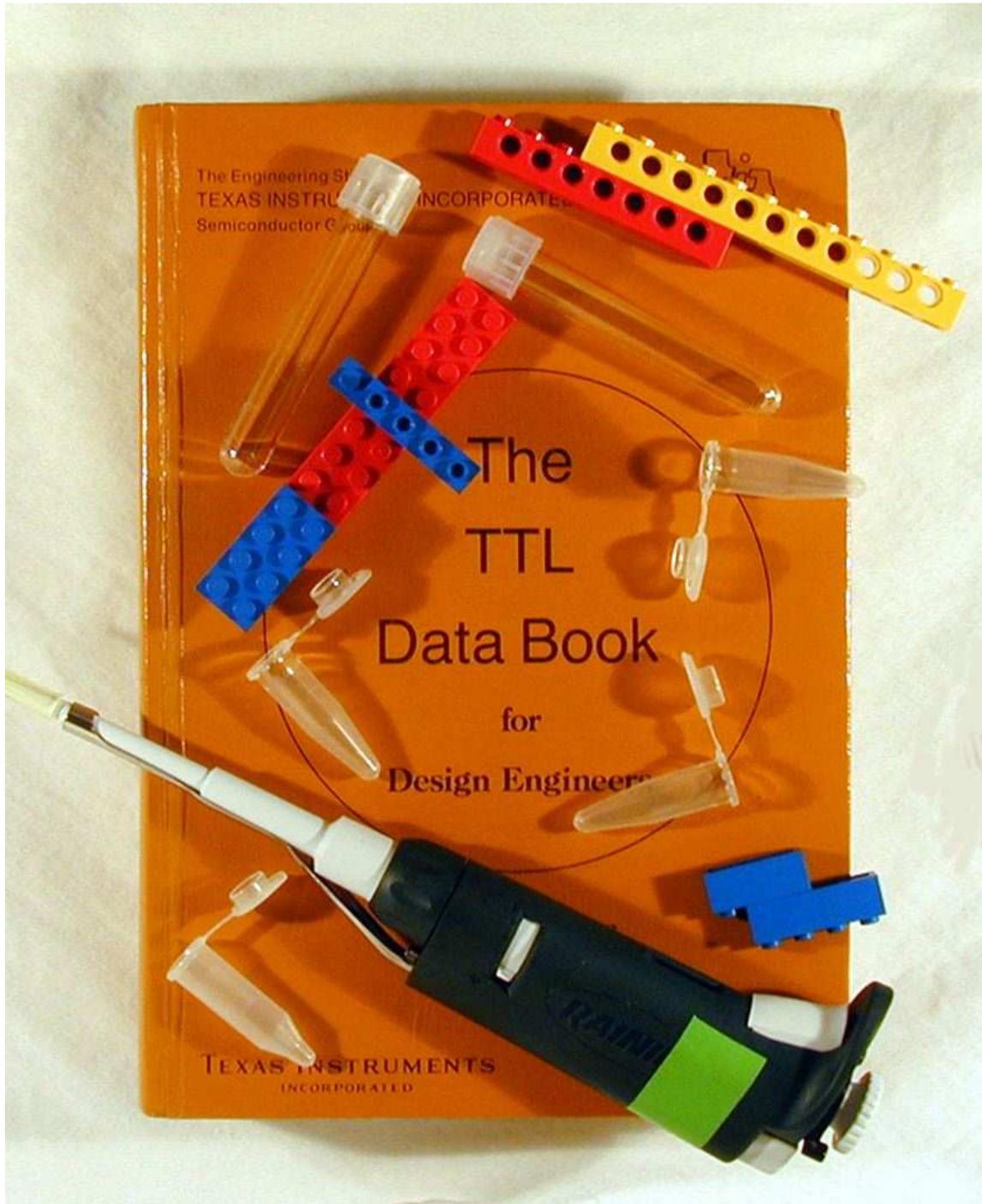


Standard Sequence Assembly of Biobricks

Tom Knight
MIT Artificial Intelligence Laboratory



0.1 Motivation

The lack of standards in the assembly of DNA sequences leads to a situation in which each assembly reaction is, in addition to being an experimental tool for addressing the current research topic, an experiment in and of itself. One of our goals is to replace this *ad hoc* experimental design with a set of standard and reliable engineering mechanisms to remove much of the tedium and surprise in assembling genetic components into larger systems. We anticipate advantages similar to those which accompany the standardization of screw threads in mechanical design – the widespread ability to interchange parts, to assemble sub-components, to outsource assembly to others, and to rely extensively on previously manufactured components. Here, we present a simple sequence standard as part of an experiment in how far this idea of standardized interface technology can be applied.

0.2 Standard Biobrick Sequence Interface

We have chosen a very simple but powerful sequence standard for biobrick components. Each component consists of a fragment of double stranded DNA containing the regulatory sequence, flanked on the upstream end by a nonspecific overhang and a set of restriction sites, EcoRI, NotI, and BglII (5' tctc GAATTC GCGGCCG AGATCT 3'), and on the downstream side by a nonspecific overhang and a BamHI restriction site (5' GGATCC gaga 3'). We require that the component fragment not contain any of these restriction sites. These sites must be removed by point mutation prior to assembly with this technique. Techniques for inserting the flanking cut sites and performing point mutations are described below, but we will assume this sort of manipulation for the time being.

A typical component strand, thus would be fabricated with a sequence of the following form:

```

5' tctc GAATTC GCGGCCG AGATCT -----insert----- GGATCCgaga 3'
3' gaga CTTAAG CGCCGGCG TCTAGA ----- CCTAGGctct 5'
      EcoRI  NotI  BglII                               BamHI
  
```

As part of the component fabrication, the sequence is double digested with EcoRI and BamHI enzymes, to yield the cut, 5' phosphorylated sequence:

```

5' *AATTC GCGGCCG AGATCT -----insert----- G      3'
3'      G CGCCGGCG TCTAGA ----- CCTAG* 5'
      EcoRI  NotI  BglII                               BamHI
  
```

An example of such a component is the DNA sequence for the transcriptional terminator T1, with restriction site ends meeting the assembly standard, as shown in figure 1.

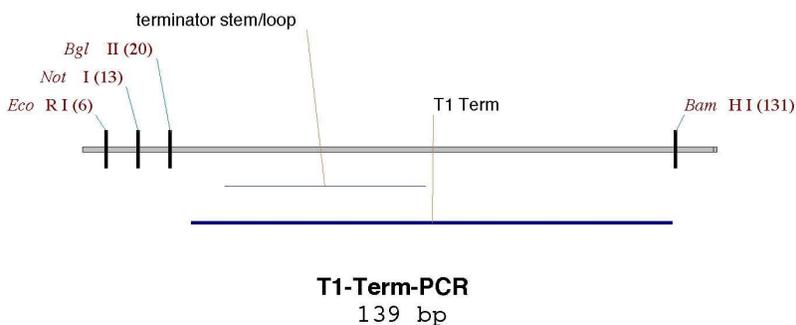


Figure 1: Standard Component form of the T1 terminator

A second example is the DNA coding sequence for the fluorescent protein ECFP, shown below, with a ribosomal binding site and engineered restriction sites meeting the standard as shown in figure 2.

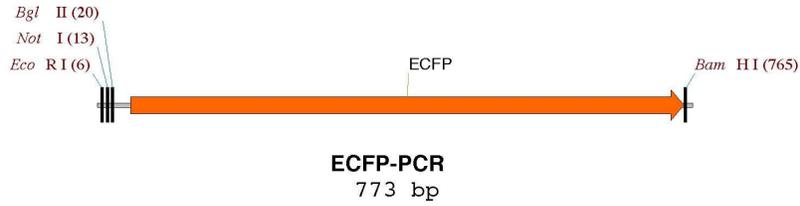


Figure 2: Standard Component form of the ECFP fluorescent protein coding sequence

0.3 Standard Assembly Vector Sequence Interface

We have constructed an assembly vector, pSB102, which is a high copy plasmid, derivative of pUC18, missing the *LacZ* gene, promoter, and MCS, which are replaced by a cloning site of the following sequence:

```

5' ---- GAATTC GCGGCCGC AGATCT GAGACG ---- 3'
3' ---- CTTAAG CGCCGGCG TCTAGA CTCTGC ---- 5'
      EcoRI  NotI   BglIII BsmBI
  
```

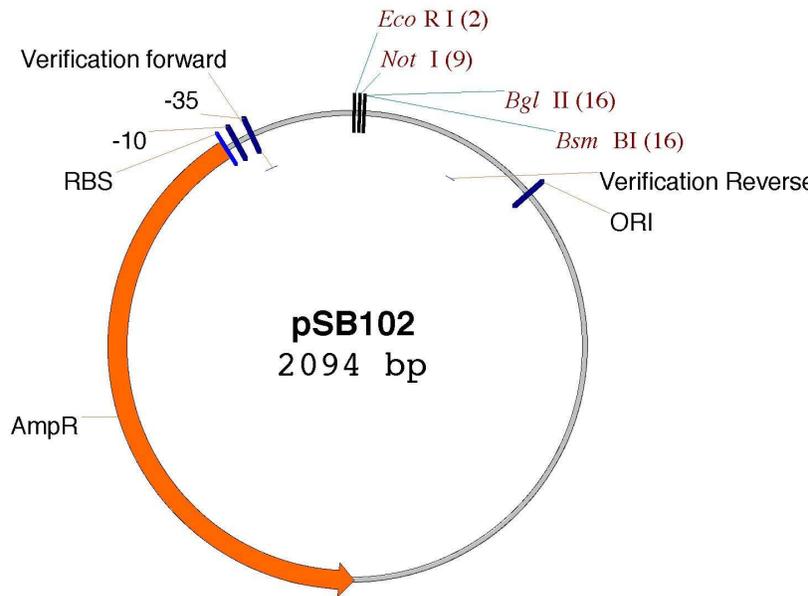


Figure 3: Structure of the pSB102 vector

The vector is designed to be cut in two different ways. For linear assembly, we cut with *EcoRI* and *BglIII*, discarding the *NotI* restriction site fragment. Cutting with *EcoRI* and *BsmBI* yields an identical fragment in this vector, but is the foundation of more complex cluster assembly strategies which are discussed below. *BsmBI* is an “offset” cutter, which is located in just the correct location in the pSB102 vector to cut at the same locations as the *BglIII* enzyme cuts the unmodified vector.

The *EcoRI* and *BglIII* cut vector has ends of this form:

```

5' ---- G                               *GATCT GAGACG ---- 3'
3' ---- CTTAA*                          A CTCTGC ---- 5'
      EcoRI                               BglIII BsmBI
  
```

0.4 Biobrick Composition Techniques

0.4.1 Linear Assembly

Our composition technique relies on the details of the interface standards discussed above. Components are assembled, one at a time, in downstream to upstream order, on a host vector.

Cutting the insert with EcoRI/BamHI and the vector with EcoRI/BglII yields fragments as described above. The BamHI and BglII sites have compatible GATC overhangs, and ligate effectively, yielding a new vector with sequence:

```

5' ---- GAATTC GCGGCCGC AGATCT ---insert1----- GGATCT GAGACG ---- 3'
3' ---- CTTAAG CGCCGGCG TCTAGA ----- CCTAGA CTCTGC ---- 5'
          EcoRI  NotI   BglII                mixed  BsmBI
    
```

Note that this vector has similar restriction site structure as the original vector, with the exception of the insert and a mixed site which can be cut by neither BamHI nor BglII.

The similarity of this vector to the parent pSB102 vector is the key to the linear assembly strategy. Cutting the newly made vector again with EcoRI and BglII yields a cut vector of this form:

```

5' ---- G                * GATCT ---insert1----- GGATCT GAGACG ---- 3'
3' ---- CTTAA*          A ----- CCTAGA CTCTGC ---- 5'
          EcoRI                BglII                mixed  BsmBI
    
```

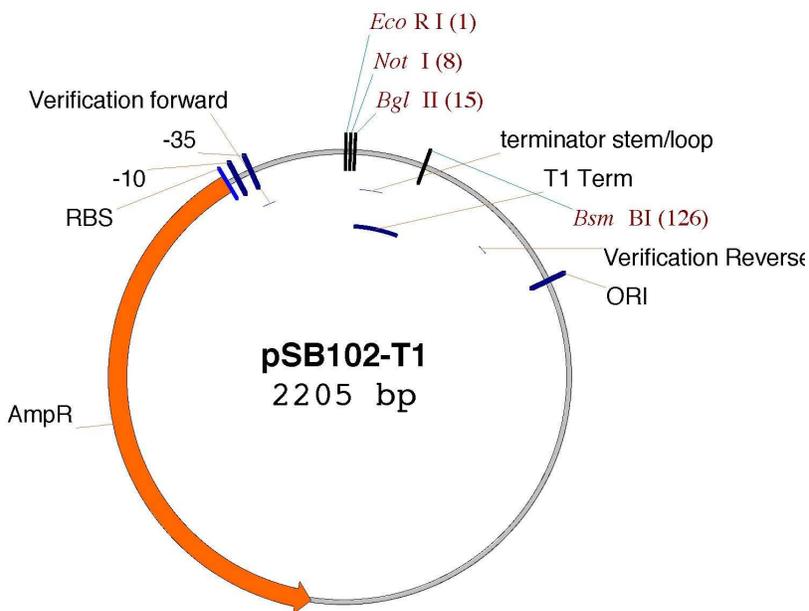


Figure 4: Structure of the pSB102-T1 vector

Cutting another insert with EcoRI and BglII and ligating it into place yields a vector of the form:

```

5' -- GAATTC GCGGCCGC AGATCT --in2-- GGATCT --in1-- GGATCT GAGACG -- 3'
3' -- CTTAAG CGCCGGCG TCTAGA ----- CCTAGA ----- CCTAGA CTCTGC -- 5'
          EcoRI  NotI   BglII                mixed                mixed  BsmBI
    
```

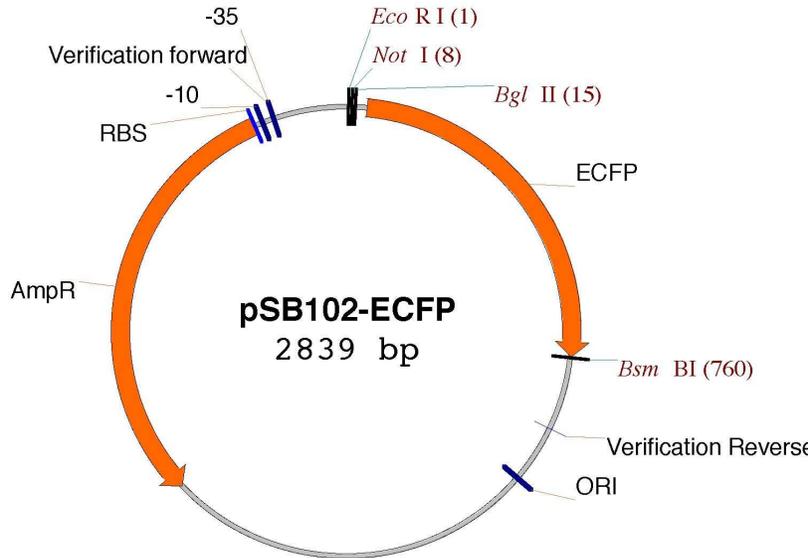


Figure 5: Structure of the pSB102-ECFP vector

leaving a second “mixed” restriction site. This linear process of incrementally adding one component after another can be continued until the desired structure is assembled. Intermediate vectors holding partially complete structures may be used to shorten the assembly process.

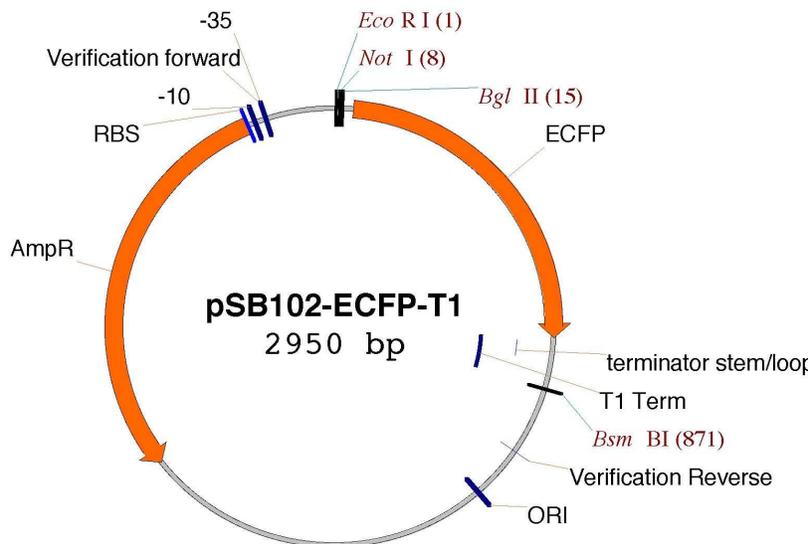


Figure 6: Linearly assembled vector with ECFP followed by T1 transcriptional terminator, created by cutting pSB102-T1 and inserting the ECFP standard component

We chose to fabricate our operons in the downstream to upstream direction. This choice was made because the transcriptional terminator is seldom changed, is almost always wanted, and can be essentially inserted once, and becomes part of the base vector sequence. Similarly, the coding sequence of some genes, such as the GFP reporter genes, is often required, and can be inserted once, viewed as part of the base vector, and used as a base level component. The promoters and other genes of the operon, however, will change frequently.

0.4.2 Recursive Assembly: Cutting with BsmBI

The process of linear assembly described above could be used to build complex, multi-operon structures. But, because operons, once constructed, are useful compositional objects in their own right, we provide an additional compositional tool. The use of the BsmBI cut site allows the cutting of entire operons out of a constructed vector, yielding a long fragment which can then be used as part of the assembly process. Specifically, consider the assembled linear vector above,

```

5' -- GAATTC GCGGCCGC AGATCT --in2-- GGATCT --in1-- GGATCT GAGACG -- 3'
3' -- CTTAAG CGCCGGCG TCTAGA ----- CCTAGA ----- CCTAGA CTCTGC -- 5'
      EcoRI  NotI  BglII           mixed           mixed  BsmBI
  
```

which can be cut with EcoRI and BsmBI to yield the following insert:

```

5' *AATTC GCGGCCGC AGATCT --in2-- GGATCT --in1-- G      3'
3'      G CGCCGGCG TCTAGA ----- CCTAGA ----- CCTAG* 5'
      EcoRI  NotI  BglII           mixed           BamHI
  
```

which obeys the same compositional rules as our standard biobrick component.

As an example, we can cut the pSB102-ECFP-T1 vector with EcoRI and BsmBI to give a standard component fragment, as shown in figure 7.

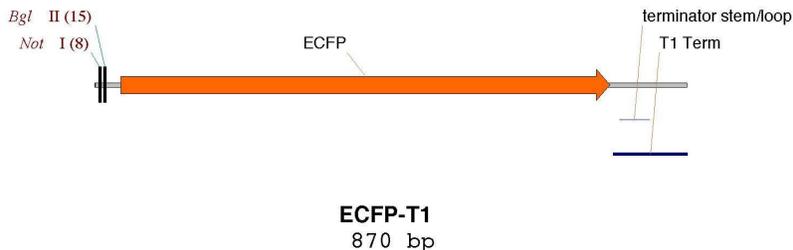


Figure 7: Standard component form of the assembled ECFP-T1 pair

Thus, this assembled operon can itself be used as a complex component in a larger assembly reaction. The recursive nature of the assembly process can, of course, be used to make yet more complex biobrick components.

0.5 Standards Summary

0.5.1 Component End Sequences

After cutting, the insert must have 5' phosphorylated ends of the following form:

```
5' *AATTC GCGGCCGC AGATCT -----insert----- G      3'
3'      G CGCCGGCG TCTAGA ----- CCTAG* 5'
      EcoRI   NotI   BglII                               BamHI
```

The insert region must not contain EcoRI, NotI, BglII, BamHI, or BsmBI sites. This excludes the following hexamer and octomer sequences:

EcoRI: GAATTC

NotI: GCGGCCGC

BglII: AGATCT

BamHI: GGATCC

BsmBI: GAGACG
AGTCTC

0.5.2 Vector Cloning Site Sequence

After cutting, the vector must have 5' phosphorylated ends of the following form:

```
5' ---- G                               *GATCT GAGACG ---- 3'
3' ---- CTTAA*                          A CTCTGC ---- 5'
      EcoRI                               BglII BsmBI
```