A New Biobrick Assembly Strategy Designed for Facile Protein Engineering

Ira E. Phillips* and Pamela A. Silver*
April 18, 2006

Abstract

The existing biobrick assembly technique[1] provides a straightforward way to combine standardized biological components, termed biobricks. This system, however, is limited in that each protein-encoding biobrick must contain a complete translated region. Signal sequences or other protein domains often convey a specific function to the protein to which they are attached; hence, each domain should be considered an independent biological part. With the current assembly technique, assembling such parts is not possible. This paper presents a revised assembly strategy that is compatible with the current biobrick definition and permits the construction of fusion proteins.

Introduction

The standard biobrick assembly technique was created in an attempt to simplify the construction of long concatemeric pieces of DNA often used in synthetic biology [1]. Previously existing techniques were limiting in that one often needed to design a unique cloning strategy for each desired construct. Moreover, the intermediates in the cloning of one construct would be of little value in creating other constructs.

As a solution to these problems, T. Knight developed a novel cloning strategy that permits standardization of biological parts or biobricks [1]. Each biobrick can be placed in front of or behind another biobrick using a standard protocol that is independent of the content of the parts involved. (To insert a biobrick in front of another, simply digest the upstream biobrick with EcoRI and SpeI. Cut out this insert and ligate into a vector containing the second part cut with EcoRI and XbaI. Conversely, to insert a part behind another, digest the insert [downstream part] with XbaI and PstI. Then, ligate into a vector containing the upstream part that has been cut with SpeI and PstI.) Most importantly, the product of such a combination has the same combination properties as the

^{*}Department of Systems Biology, Harvard Medical School, Boston, MA 02115.

```
5' -GAATTC GCGGCCGC T TCTAGA G part T ACTAGT A GCGGCCG CTGCAG- 3'
3' -CTTAAG CGCCGGCG A ACATCT C part A TGATCA T CGCCGGC GACGTC- 5'
EcoRI NotI XbaI SpeI NotI PstI

5' -TCTAGA G Part_1 T ACTAGA G Part_2 T ACTAGT- 3'
3' -AGATCT C Part_1 A TGATCT C Part_2 A TGATCA- 5'
XbaI Mixed Site SpeI
```

Figure 1: Standard definition of the biobrick ends (top) and the result from biobrick assembly (bottom). The mixed site is eight base pairs, resulting in a frame shift if translated.

originals. Hence, component biobricks can, in turn, be assembled together, creating long concatemeric sequences.

There is, however, a limitation in the current biobrick assembly technique—it does not permit the creation of fusion proteins. This limitation, which is especially problematic for eukaryotic systems, stems from two problems: translation over the mixed site (that results from fusing two parts) creates a frame shift, and coding region parts end with stop codons.

Initially, this shortcoming was handled by creating various versions of each major coding region part. For example, at least four different versions of ECFP (BBa_E0020, BBa_E0022, BBa_E0024, BBa_E0026) were synthesized using non-biobrick assembly techniques [2]. While such a solution is manageable on a small scale, it would be advantageous to harness the power of the biobrick assembly strategy to permit combination of the basic ECFP reporter to the desired signal sequences. Such a technique becomes critical when multiple fusion protein sequences are desired from the combination of multiple protein domains.

Improvements

We wanted to create an improved assembly strategy that allows for the creation and assembly of protein domain parts and that is compatible with existing biobricks. Moreover, we desired a strategy that would require relatively little testing. Therefore, we decided to use the restriction enzyme system described by T. Knight [1], since this system had already been proven to work. We, however, needed to change the biobrick definition to allow translation over the mixed site, which results from the fusion of two parts. Previously, this mixed site consisted of eight base pairs, which would result in a frame shift when this region is translated.

While it is necessary to maintain the frame over the mixed site, the translation frame for the restriction-ligation scar (ACTAGA) can be changed by adding different numbers of spacer nucleotides on either side of the part sequence. All three frames were considered, but only two were viable (Figure 2). The first (left) frame was chosen due to its shorter length and higher hydrophilicity. (It is important to note that each of the two possible frames results in the incor-

ACT	AGA	xAC	TAG Axx	xxA	CTA	GAx
Thr	Arg		STOP		Leu	D/E

Figure 2: Possible translation frames over the mixed site. Two frames are viable.

```
5' -GAATTC GCGGCCGC T TCTAGA part ACTAGT A GCGGCCG CTGCAG- 3'
3' -CTTAAG CGCCGGCG A ACATCT part TGATCA T CGCCGGC GACGTC- 5'
ECORI NotI XbaI SpeI NotI PstI

5' -TCTAGA Part_1 ACTAGA Part_2 ACTAGT- 3'
3' -AGATCT Part_1 TGATCT Part_2 TGATCA- 5'
XbaI Mixed Site SpeI
```

Figure 3: New definition of the biobrick ends (top) and the result from biobrick assembly (bottom). The mixed site is six base pairs, thereby maintaining the reading frame when translated.

poration of a charged amino acid. Each could have unintended consequences if placed beside charge-sensitive sequences.)

When two parts are combined using the new strategy, the reading frame is maintained, and a threonine and arginine are inserted between the coding parts (Figure 3).

It is important to note that the spacer nucleotides between the part and the XbaI and SpeI sites were originally incorporated in order to prevent DNA methylation and its consequential inhibition of the required restriction enzymes. These spacer nucleotides are not present in the assembly strategy presented here. Hence, it is required to prescreen parts to ensure that methylation sites are not incorporated. Dam methylation of the XbaI site is the most critical. Parts beginning with the sequences TCx create a site capable of inhibiting XbaI. If a desired part naturally begins with this sequence (coding for serine), it is necessary to change the first codon to either AGT or AGC (both also encoding serine).

Additionally, several modifications to the part specifications are required. First, each coding region should start and end in frame and not contain a stop codon. Also, it may be desirable to omit the start codon as well in order to reduce aberrant translation starting points [3].

Since parts containing protein domains have neither a start or stop codon, it is necessary to create a part category for each of these. Stop codon parts are straightforward. They should contain one or more stop codons to halt translation.

Start codon containing parts, however, prove more interesting. A limitation of the current biobrick assembly strategy is that ribosome binding sites contain only untranslated regions and a fraction of the start codon. It is known in eukaryotic systems, however, that the first nucleotide following the ATG of the start codon can have a significant effect on translational efficiency [4]. It

is possible with the improved biobrick strategy to incorporate another codon downstream of the start codon to control this variable. This permits the creating of a part of the Kozak consensus sequence, which is the consensus translation start codon sequence for highly expressed eukaryotic genes. Doing so should make the translation initiation site strengths more predictable.

In summary, the following modifications have been made to the original biobrick definition.

- The biobrick ends have changed so that the XbaI and SpeI restriction sites are adjacent to the biobrick part.
- Coding parts begin and end "in frame" and do not contain start or stop codons.
- Coding parts should be fused to both promoter and translation initiator parts and stop codon and transcriptional terminator parts.
- Parts must not begin with the nucleotides TC.
- Non-translation-related parts created with T. Knight's definition can be mixed with the parts proposed here.
- Previously built coding parts can be easily converted by PCR to meet the fusion biobrick specification.

Possible Protein Domain Parts

This new assembly technique permits the creation of a wide new list of part categories, including some interesting eukaryotic-specific parts. An incomplete list of such possible new parts is included below.

- Degradation Signals
- DNA Binding Domains
- Introns: Increase nuclear export; possibly delay translation.
- Localization Signals: Export signals, nuclear import signals, etc.
- Protein Domain Linkers
- Reporters
- Start Codons: Ribosome binding sites, consensus Kozak sequence, etc.
- Stop Codons
- Tags: His, FLAG, maltose binding protein, etc.
- Transcriptional regulatory domains



Figure 4: Example of a yeast fusion protein created with the described assembly method. The CTS1 promoter is followed by the Kozak sequence, two YFP proteins fused to an SV40 NLS, a stop codon, and a terminator from ADH1.



Figure 5: Fluorescent image overlaid with Nomarski of yeast with the construct in Figure 4 integrated in the genome. The YFP fusion protein is expressed in the daughter cell and localizes to the nucleus as expected.

Results

This modified cloning strategy has been used in our lab to produce approximately 100 protein fusion constructs. An example of such a fusion construct is shown in Figure 4. This construct contains a yeast promoter from CTS1, a gene expressed specifically in daughter cells [5], followed by the Kozak consensus sequence, two copies of yeast-codon optimized YFP, an SV40 nuclear localization signal, and finally a stop codon and transcriptional terminator from the yeast ADH1 gene. If properly expressed, this fusion protein should be present primarily in the nucleus of daughter cells. The predicted result is confirmed in Figure 5, where the presence of the fluorescent YFP in the nucleus confirms that the full-length fusion protein has been produced.

In summary, this modified version of the biobrick assembly technique permits the creation and assembly of a more diverse set of biological parts.

Acknowledgements

This work was supported by funds from the Keck Foundation.

References

- [1] Knight T. "Idempotent Vector Design for Standard Assembly of Biobricks." DSpace http://hdl.handle.net/1721.1/21168 (2003).
- [2] Registry of Standard Biological Parts. MIT. Release 11.21.05.
- [3] Kozak M. "Regulation of translation via mRNA structure in prokaryotes and eukaryotes." Gene. 361: 13-37 (2005).
- [4] Kozak M. "An analysis of 5-noncoding sequences from 699 vertebrate messenger RNAs." Nucleic Acids Res. 15(20): 8125-48 (1987).
- [5] Colman-Lerner A, Chin TE, and Brent R. "Yeast Cbk1 and Mob2 activate daughter-specific genetic programs to induce asymmetric cell fates." Cell. 107: 739-750 (2001).