RFC 38: Building Blocks - Standard Large DNA/Genome Construction

Jef Boeke, James DiCarlo

7 October 2009

1. Purpose

The physical assembly of standard parts is currently a non-standard process, which can either come from direct genome PCR with restriction enzyme sites incorporated into the PCR primers, or overlap assembly PCR. Furthermore, current Biobrick construction standards [RFC 10,11,12 etc.] rely heavily on restriction enzyme based methods, which can be sequence, cost and time restrictive. This is especially true of large DNA assembly, where 100% control of over DNA sequence may be mandatory, such as in whole genome assembly. We propose a novel standard, The Building Block Method, for both the construction of standard parts and their assembly

This Building Block standard can be either A) an interchangeable or B) a non-interchangeable format, based on the desired use of the part.

2. Relations to other BBF RFCs

This RFC (Request for Comments) is a new proposal for standardized large DNA/genome construction. It describes a series of methods for construction and assembly that are unrelated to existing RFCs.

3. Copyright Notice

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

4. The Building Block

a. Definitions

- i. A Standard Build Block is a ~750 base pair sequence of DNA that is composed of annotated biological parts. This particular size represents a "sweet spot" for a currently optimized set of efficient assembly techniques (e.g. high accuracy sequencing in a single run) but need not be fixed in stone. The Building Block, however, MAY not be composed entirely of single biological parts (as in BioBricks) and the junction with neighboring Building Blocks MAY be essential for biological for function.
- ii. A Composite Building Block is composed of two or more Building Blocks. Composite Building Blocks MAY be constructed from non-interchangeable Building Blocks OR interchangeable Building Blocks.

b. Interchangeability

- i. An Interchangeable Building Block MUST contain standard sequences on both 3' and 5'ends
- ii. A non-interchangeable Building Block MAY contain non-standard 3' or 5' ends to aid in the construction of Composite Building Blocks.

5. Standard Building Block Ends

- a. Building blocks are assembled using (Uracil Specific Excision Reagent) USER reaction assembly, which relies on a uracil base within the end of the Building Block DNA. Both 5' ends of Building blocks must conform to the extremely degenerate consensus sequence of $A\{N_x\}T$ in which x=3,5,7,9 or 11. There are no other sequence requirements. The general format for a building block can be seen in **Figure 1.**
 - i. Non-interchangeable Building blocks MAY have unique and different sequences for the $\{N_x\}$ region. These sequences MAY be part of the Building Block functional contents
 - ii. Interchangeable Building blocks MUST have $A\{N_x\}T$ sequences that conform to a standard sequence agreed to by the community of researchers that decide to use it.

Figure 1. Here the Building block is diagrammed to conform to the sequence 5' $A\{N_x\}U$... $A\{N_x\}T3'$. Building blocks can be synthesized either in this format (by including terminal oligonucleotides each containing a single strategically placed deoxyuracil residue) or the more conventional 5' $A\{N_x\}T$... $A\{N_x\}T$ 3', in which the terminal oligonucleotides are composed of the four normal DNA bases. Either set of primers can be used to amplify a Building Block, depending on the stage in the synthesis procedure.

b. Sets of Universal Interchangeable $\{N_x\}$ Sequences MAY be based on the contents of the building block and the location of the end. One possible set of Interchangeable Building blocks follow the specific format seen in **Figure 2.** However please note that there is a large universe of possible $\{N_x\}$ sequences that could be chosen as standards. It is not the purpose of this RFC to explore all possible interchangeable $\{N_x\}$ sequences or even to propose particularly good ones. The focus of the remainder of this RFC is on the use of noninterchangeable $\{N_x\}$ sequences.

Type of Building Block	5' Standard End	3' Standard End
(based on its contents)		
Promoter ONLY	A <u>AATAA</u> T	A <u>AAA</u> T
Protein Coding Region ONLY	A <u>AAA</u> Tg	tA <u>AGG</u> T
Terminator ONLY	<i>tAAGG</i> T	A <u>AATAA</u> T
Promoter + Protein Coding Region	A <u>AATAA</u> T	tA <u>AGG</u> T
Protein Coding Region + Terminator	A <u>AAA</u> Tg	A <u>AATAA</u> T

Figure 2. Underlined sequences are $\{N_x\}$ sequences. Lowercase sequences are optional flanking nucleotides. In the examples shown here, the universal initiation and stop codons ATG and TAA are included in *italics* to demonstrated how such signals can facilitate "hot swapping" of biological cassettes and form a series of interchangeable parts much like BioBricks but without the liabilities of restriction enzyme use or leaving any scars behind.

- 6. Building Block synthesis.
 - a. A general method of building block synthesis based on overlap extension PCR is an efficient and inexpensive way to construct Building Blocks.
 - Building Blocks are assembled by overlap assembly PCR from standard 5'-OH, 3'-OH 60-80 base pair oligonucleotides purchased from a DNA synthesis company. The overlap extension assembly process is broken into two polymerase chain reactions:
 - 1. "Templateless" PCR- Oligos are mixed and standard overlap extension assembly PCR program is utilized. Mixed length products are formed.
 - 2. "Finish" PCR Outer two primers are mixed with the Templateless PCR product, such that a higher concentration of outer primer is present in the mixture, and a second round of PCR is completed to enrich for full length products
 - ii. Cloning Finish PCR product is cloned using PGEM-T easy vector, taking advantage of Taq polymerase single 3' adenylation.
 - iii. Sequence verification Sequencing is completed by outside sources and sequence alignment using programs such as "CloneQC" aids with correct clone verification
 - b. Templateless/Finish PCR assembly trouble-shooting
 - PCR program conditions can be altered to optimize construction of Building Blocks. Varying the anneal and extension temperature MAY assist in the production f the Building Block
 - ii. Ligase or Ligase Chain Reaction is an alternative for Building Block synthesis instead of Templateless. Complete overlap between oligonucleotides must be utilized, such that ends can be directly ligated to form a full length Building Block. The ordered oligonucleotides must have 5' phosphate groups, and then a DNA ligase is utilized to ligate the pieces.
 - c. Substrate production for USER reaction
 - i. PCR is performed with special primers containing a single U as described in Figure 1. A high-fidelity polymerase is preferred at this step.
- 7. USER Assembly Method. This is the method used to combine multiple Building blocks into a single composite building block.

- a. The Uracil Specific Excision Reagent is composed of two enzymes: Uracil DNA glycosylase (UDG) and the DNA glycosylase/lyase Endo VII. The Uracil DNA glycosylase targets the pyrimidine and ribose components of the uracil base and cleaves them from the phosodiester backbone. The DNA glycosylase/lyase Endo VII then cleaves the exposed phosphodiester backbone, which allows the release of the 3-11 bases upstream of the uracil base, thus creating a nonpalindromic sticky end. The enzyme is commercially available and works well.
- b. This sticky end that is released is utilized for the construction of Composite Building Blocks; when these ends are designed so as to be complementary only to that of the adjacent Building Block, the Building Blocks can be ligated only in a single, predestined order and in a single, desired orientation.
- c. See **Figure 3** for an example in which the Promoter, Coding sequence and Terminator above are joined into a Composite Building Block.

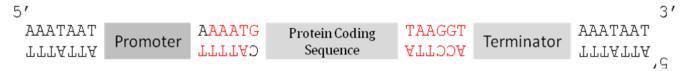


Figure 3. Example of an assembled Composite Building Block. The red bases represent the sticky ends that annealed together to form the Composite Building Block.

8. Summary

- a. RFC 38 is proposes a new standard for large DNA/genome assembly and defines a new DNA part standard, the Building Block.
- b. The plasticity of the assembly ends proposed allows for various DNA sequences to be used as assembly junctions.
- c. RFC 38 proposes a standard set of assembly ends for interchangeable Building Blocks and allows for nonstandard ends for non-interchangeable Building Blocks. A standard workflow for the construction of Building Blocks is also proposed.

9. Authors' Contact

Jef Boeke jboeke@jhmi.edu

James DiCarlo jdicarlo@jhu.edu