BBF RFC 95: Open Sequence Initiative: a part submission standard to complement modern DNA assembly techniques

1. Purpose and Motivation

The discipline of synthetic biology emphasizes the application of engineering principles such as standardization, abstraction, modularity, and rational design to complex biological systems. The archetypical example of such standardization is BioBrick RFC[10], introduced in 2003 by Tom Knight at MIT[1]. BioBricks are stored on a standard plasmid, pSB1C3, which contains prefix and suffix sequences flanking the part. These sequences contain two pairs of 6 base-pair (bp) restriction sites (EcoRI+XbaI and SpeI+PstI), which can be used for both part assembly and quality control. BioBricks are intended to be well-characterized biological parts, such as genes or promoters, that function in a predictable fashion, are ready to use, and can be combined in unique ways. The rules of this assembly method also require that none of these sites are present in the parts themselves. This last requirement can be an onerous imposition for iGEM teams developing large, novel parts, such as genes or entire operons that are obtained by amplifying DNA sequences from environmental samples or microorganisms.

At the time BioBricks were introduced, restriction enzyme cloning was the best method for the assembly of multiple DNA sequences into a single construct. However, a decade of progress in molecular biology has led to the development of many powerful DNA assembly and synthesis alternatives. A great variety of assembly methods now exist, such as homology-based protocols using *in vitro* assembly (Gibson Cloning, Seamless Cloning), *in vivo* assembly (yeast recombination), and assembly using type II restriction enzymes (Golden Gate Assembly)[2–4]. Many of these methods can be employed on arbitrary DNA sequences; that is, they have no inherent requirement that specific base sequences must be present in the DNA specifying a part for it to be used in assembly.

While the requirement to remove the restriction sites present in the prefix and suffix may seem a minor inconvenience on first glance, calculating the frequency at which restriction sites occur reveals that compliance with the BioBrick RFC[10] standard likely burdens most teams submitting novel parts to the registry. The probability ***P*** of any given 6 base-pair restriction site occurring at a given site in a random sequence is $1/4^6 = 1/4096$. The probability ~***P*** of a restriction site ***not*** occurring in a sequence of length ***x*** is thus:

$$\sim P = (1 - \frac{1}{4096})^x = (\frac{4095}{4096})^x$$

For additional accuracy, the effect of average GC content can be taken into account, due to the occurrence of only 2 G/Cs in the recognition sites of EcoRI, XbaI, and SpeI, and 4

G/Cs in the PstI site. This bias results in restriction sites occurring more often in AT-rich sequences. The probability of a site occurring is therefore:

$$\sim P = (a)^x (b)^y$$

Where **a** and **b** are the probabilities of A/Ts and G/Cs occurring at a known GC%, while **x** and **y** are the number of A/T and G/C nucleotides in a given restriction site. As an example, the **~P** for EcoRI (GAATTC) in a region with 60% GC content is:

$$\sim P = (0.3)^4 (0.2)^2$$

From this equation, the probability ***Q*** of ***none*** of the four "illegal" restriction sites appearing in a sequence of length ***x*** can be determined:

$$Q = \sim P(EcoRI) \times \sim P(XbaI) \times \sim P(SpeI) \times \sim P(PstI)$$

Graphing the probability of at least one of the four restriction sites appearing in a sequence of a given length (**Fig. 1**) makes the problem clear:

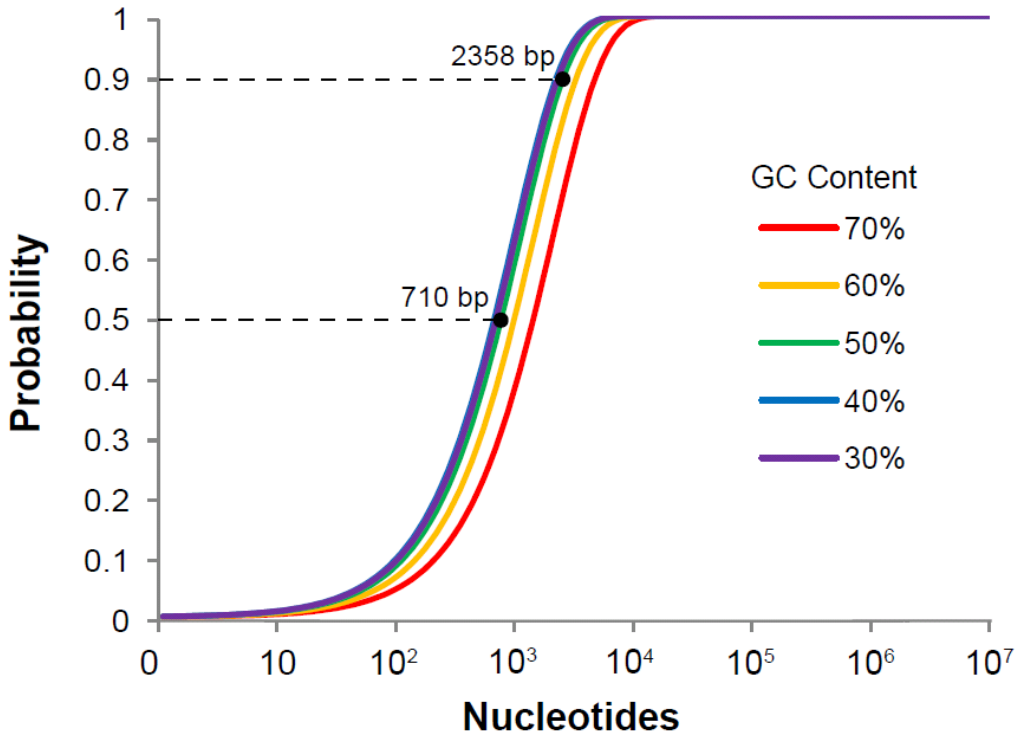


Fig. 1: The probability of at least one restriction site appearing in random DNA sequences of different lengths. The lengths at which sequences will have a 50% and 90% likelihood of containing a site are highlighted for the 50% GC-content line.

The probability of a random part containing at least one of the four restriction sites rapidly increases with sequence length, such that a majority of parts >710 bp will contain a BioBrick restriction site, and more than 90% of those >2360 bp will. Given that the length of a typical bacterial gene is roughly 1000 bp, an iGEM team would be lucky to find a novel or naturally occurring gene-sized part, let alone an operon, that did not contain an illegal restriction site.

While iGEM teams may use methods such as site-directed mutagenesis to remove illegal restriction sites, it is certainly possible that this mutation will alter the functionality of the part – a very undesirable outcome. In addition, the mutagenesis of illegal restriction sites is an unnecessary burden on teams, given the limited time and resources available to teams during each year's competition. Efforts spent mutagenizing sites would be better spent characterizing and improving parts. This RFC proposes an alternative standard submission format, which eliminates these problems, and suggests strategies for improving the quality of parts submitted to the registry.

2. Relation to other BBF RFCs

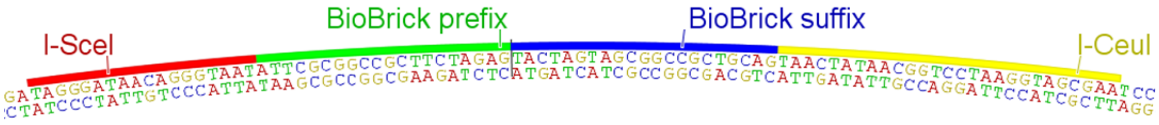
This RFC is intended as an alternative or a replacement to the current Biobrick standard RFC[10] in order to reduce restrictions on parts submitted to the Registry of Standard Biological Parts. Crucially, this RFC removes the sequence restrictions imposed by RFC[10] regarding the removal of the restriction enzymes sites EcoRI, XbaI, SpeI, and PstI from parts before they are eligible for inclusion in the Parts Registry.

3. Copyright Notice

4. Submission Standard Format and Quality Control

Quality control is critical for any serious engineering endeavor. While teams SHOULD sequence verify their parts to detect point mutations and small indels, the BioBricks Foundation needs a simple and rapid method for determining that parts are at least the correct size when submitted to The Registry. Currently this quality control is achieved through the use of the restriction sites present in the BioBrick prefix and suffix sequences, resulting in restrictive rules for part submission. Therefore, we propose an alternative submission plasmid, **pSB1C95**, featuring the homing endonucleases I-SceI and I-CeuI and their corresponding restriction enzyme sites for quality control purposes. Parts MUST be submitted between the BioBrick prefix and the BioBrick suffix that are present in the **pSB1C95** backbone**,** and the homing endonuclease sites I-SceI and I-CeuI MUST NOT be contained within the part itself (**Fig. 2A**). There are no other rules or restrictions for part submission. Note that, if a part that does not contain internal BioBrick restriction sites is cloned into **pSB1C95** it remains compatible with RFC[10] assembly, otherwise it must be combined with other parts using other assembly methods.

A



B

Fig. 2: (A) Order of BioBrick prefix, suffix, and homing endonuclease sites on the RFC[95] standard submission plasmid pSB1C95 . The rest of the pSB1C3-derived backbone remains the same. (B) Recognition sequences and corresponding cut sites of the homing endonucleases I-SceI and I-CeuI.

Unlike typical 6 bp restriction sites, the I-SceI and I-CeuI homing endonuclease sites are 18 bp and 27 bp in length (**Fig. 2B**), respectively, with a tolerance of sequence degeneracy corresponding to a normal restriction site 10 to 12 bp long **[NEB]**. However, the probability of even a 10 bp sequence randomly occurring in a submitted part is orders of magnitude lower than a 6 bp site, thus effectively eliminating the problem of illegal sites for the purposes of rapid quality control on gene- and operon-sized parts.

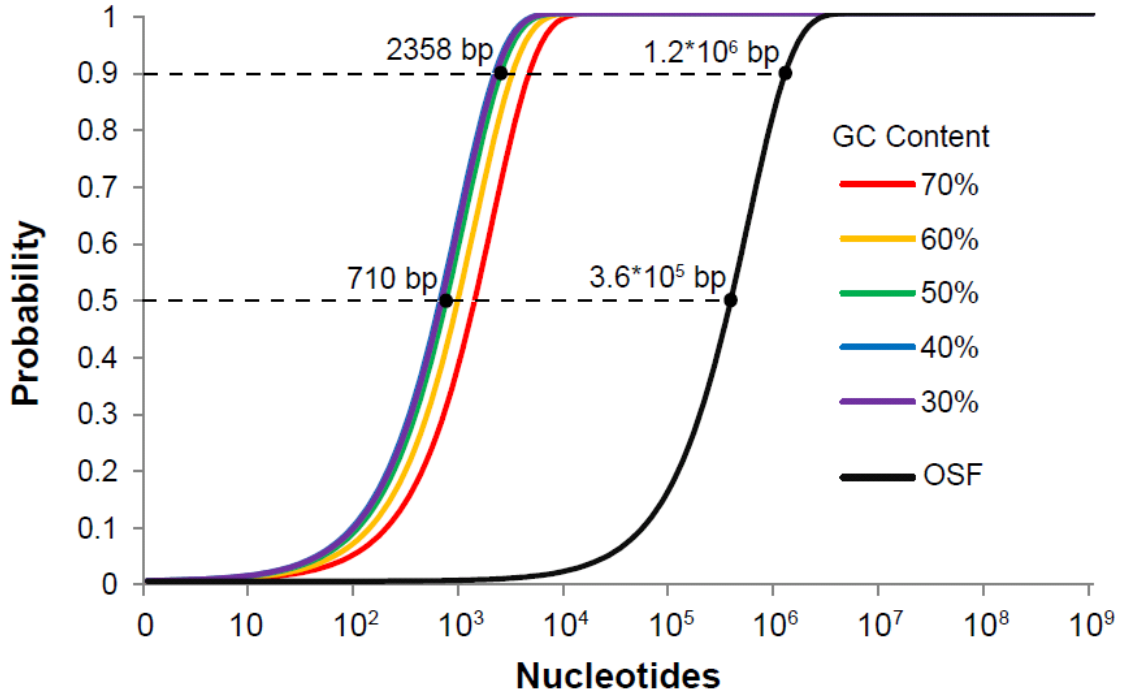


Fig. 3: Comparison between the probability of the occurrence of homing endonuclease sites RFC[95] and the current BioBrick RFC[10] restriction sites in a sequence of a given length.

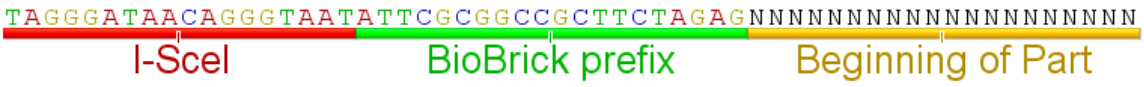
These two homing endonucleases are both available from New England Biolabs, and function at 100% efficiency in NEB Cutsmart™ buffer. All other aspects of the quality control process would remain unaltered, drastically reducing the probability of illegal sites or erroneous restriction analysis results while preserving the majority of the present quality control workflow, and supporting any desired assembly method.

5. Example Assembly Method

This RFC decouples the assembly standard proposed by RFC[10] from the submission standards currently used by the Parts Registry. Modern assembly methods require less time, can combine many parts at once, work well with novel parts, and leave no scar. One such method, Gibson assembly, can combine multiple genes or gene fragments in the desired order in a one-hour, one-pot reaction.

This process can be used to submit parts compatible with the RFC[95] submission standard by designing overlap regions between the part to be submitted and the plasmid backbone that include the SceI and CeuI sites. Examples of primers to amplify parts are shown below:

Forward Part Primer



Reverse Part Primer

Fig. 4: A 58 bp forward primer used to amplify a part in preparation for Gibson assembly using the OSF plasmid, and the corresponding 58 bp reverse primer. Note that only part of the I-CeuI homing endonuclease site is needed to achieve a homology region of sufficient length.

Using a reverse primer to the Biobrick prefix, and a forward primer to the Biobrick suffix, teams may produce a linear plasmid backbone fragment. By design of the part amplification primers, this backbone will share ~40 bp of sequence identity with the amplified part. Many DNA fragments may easily be assembled by introducing sequence overlap in the order in which they are to be assembled, allowing the rapid assembly of many parts into the plasmid backbone. These fragments are assembled by combining them in Gibson Master Mix and following the assembly protocol. Full details on Gibson assembly protocols can be found in RFC[57][5], or the original publication[2].

#. Author's Contact Information

References

1. Knight, T. Idempotent Vector Design for Standard Assembly of Biobricks Idempotent Vector Design for Standard Assembly of Biobricks.

2. Gibson, D. G. Enzymatic assembly of overlapping DNA fragments. *Methods in enzymology* **498,** 349–61 (2011).

3. Gibson, D. G. Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic acids research* **37,** 6984–90 (2009).

4. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PloS one* **3,** e3647 (2008).

5. Collins, B. *et al.* BBF RFC 57 : Assembly of BioBricks by the Gibson Method Relation to other BBF RFCs. 1–7 (2010).