Request for comments

Biobrick assembly standard modifications

8 July 2008
Tom Knight

Background:

Over the past several years the original Biobrick assembly standard has proven to be a useful DNA assembly technique. Despite this, a significant flaw has been the composition of the mixed base scar, T ACTAGA G. Since this scar is 8 bp long, it makes protein fusions, aligned on three base codon boundaries, quite difficult. Ira Phillips at the Silver Lab worked around this problem by ignoring the flanking T and G sites (inserted for protection against methylation issues) and using the mixed site ACTAGA. This resulted in the amino acid sequence Thr-Arg inserted into fusion protein designs, two amino acids with significant chemical difficulties in many contexts.

Chris Anderson at Berkeley worked around this problem in a different way, by adopting a new restriction enzyme set, BglII (prefix, AGATCT site) and BamHI (suffix, GGATCC site). These enzymes are insensitive to methylation, and produce a scar GGATCT (Gly-Ser). The Gly-Ser amino acids are near ideal for most protein fusion work, and the enzymes are cheap and effective. Unfortunately, neither of these enzymes can be heat inactivated, making automated assembly with them substantially more difficult.

Another difficulty with these enzymes is the frequency of the BamHI and BglII sites in many natural DNA sequences. For example, in the *E. coli* genome the BamHI average fragment length is 9,000, while the average fragment length of XbaI fragment is 120,000. This reflects the relative rarity of the CTAG sequence in *E. coli* genomic DNA (for reasons poorly understood). The high frequency of sites causes two problems. First, making new Biobricks from existing genomic DNA becomes substantially more difficult. Second, the frequent occurrence of these sites in contaminating genomic DNA in minipreps results in short fragments which can replace desirable parts in assembly reactions, yielding incorrect products.

Proposal:

Two additional restriction enzymes exist with a CTAG overhang: AvrII (CCTAGG site) and NheI (GCTAGC site). AvrII cannot be heat killed, and produces poorer codon choices than NheI. I propose that we restructure the cloning site and flanking sites of Biobrick parts with the following structure:

……<EcoRI>……<SpeI> Part <NheI>…..<PstI>…..

The part would be flanked by bare SpeI and NheI sites. The mixed site formed by assembly of these fragments, using standard approaches, would be GCTAGT, coding for

Ala-Ser. The Ala-Ser amino acids are almost as fusion-friendly as the Gly-Ser of the Anderson fusion technique.

The NheI enzyme can be heat killed, and thus is more amenable to automated assembly processes.

The rarity of the NheI site in *E. coli* genomic DNA means that many fewer fragments accidentally cut from genomic DNA contamination of minipreps will clone in place of the desired part.

Transition issues:

We would need to construct new cloning vectors with the new cloning site. Parts would need to be recloned into the new vectors, probably using PCR with new primers. Manual assembly of parts mixed between old and new formats would likely be possible in many cases as an interim solution, since the parts retain a common CTAG overhang.

We should rethink the use of the EcoRI enzyme for the prefix outside cutter. There are likely more robust enzymes usable.

We should rethink the need/desirability of the NotI sites between the outside and inside restriction enzyme sites. Some DNA fragment is necessary there, but it need not be that sequence, and the two sequences need not be identical.

Plan:

1) Circulate this document for comments and blunder stopping
2) Analyze the frequency of NheI sites in existing registry parts
3) Test the efficiency of NheI and any other recommended enzymes
4) Test for the ability to heat kill the enzymes
5) Design automated programs to assist in the primer design for transition
6) Choose which parts are worth transitioning
7) Design desirable part collections for protein fusion work