Standardizing the Representation of Parts and Devices for Build Planning

Jacob Beal, Vinoo Selvarajah, Gael Chambonnier,
Traci Haddock-Angelli, Alejandro Vignoni, Gonzalo Vidal and
Nicholas Roehner

October 11, 2022

Standardizing the Representation of Parts and Devices for Build Planning

Jacob Beal[1], **Vinoo Selvarajah**[2], **Gael Chambonnier**[3], **Traci Haddock-Angelli**[2], **Alejandro Vignoni**[4], **Gonzalo Vidal**[5], **Nicholas Roehner**[1]

[1]Raytheon BBN, [2]iGEM Foundation, [3]Massachusetts Institute of Technology, [4]Universitat Politecnica de Valencia, [5]Newcastle University

jakebeal@ieee.org,vinoo@igem.org,gchambon@mit.edu,traci@igem.org,
vignoni@isa.upv.es,g.a.vidal-pena2@newcastle.ac.uk,nicholas.roehner@raytheon.com

1 MOTIVATION

One of the most common tasks in synthetic biology is building genetic constructs by assembling smaller parts. Despite this commonality, however, there is often much confusion when practitioners communicate about parts, sequences, and build plans. Parts often go through many stages during a build process, each with a different sequence. For example, a fragment of DNA may be synthesized as an insert into a vector backbone, then digested out of that backbone and assembled together with other fragments to produce a final construct. At present, without a shared standard for describing build plans, it is often difficult to tell which stage a given sequence is describing, leading to frequent confusion, errors, difficulty sharing information, and waste.

We address this problem with a standard vocabulary for describing build plans, which we have further mapped into a concrete representation using the SBOL 3 standard [4]. Specifically, we target representation of assembly based on digestion and ligation, supporting at least BioBricks Assembly [6] and Type IIS assemblies like GoldenGate [1], MoClo [7], and GoldenBraid [5]. The resulting vocabulary should be useful to practitioners no matter what tools or representations they may be using, while representation in SBOL 3 provides full details for use by software tool builders.

2 STANDARDIZING TERMINOLOGY

Our first target of standardization is the terminology used for describing DNA at different stages of build planning. Developing this vocabulary was motivated by challenges in developing the iGEM 2022 distribution, where we found many miscommunications between collaborators about how sequences related to our build plans (e.g., did a sequence already include flanking sequences, was this what should be synthesized or what it would look like after insertion into a backbone, etc.). To this end, we have proposed the following definitions, cleaving as closely as possible to pre-existing patterns in descriptions, and aligning with typical digestion/ligation build planning as shown in Figure 1:

- **Part:** Design for a single contiguous linear DNA construct with a completely specified sequence.
- **Unitary Part:** Any part that is not designed with reference to an assembly, often but not always having a well-defined role such as a CDS or promoter.
- **Composite Part:** A part designed as the composition of two or more other parts through an assembly plan.
- **Assembled Part:** A part, plus any 5' or 3' flanking scars, in the post-assembly context of a composite part.
- **Scar:** A sequence that is produced by the combination of flanking sequences in an assembly.
- **Backbone:** A DNA construct into which parts are intended to be inserted at one or more designated insertion sites (often, but not always, a circular plasmid).
- **Drop-Out Sequence:** A portion of a backbone at an insertion site that is removed when a part is inserted at that site. Some backbones include drop-out parts while others do not.
- **Part Insert**: A part, plus any 5' and 3' flanking sequences, that is intended to be placed into a designated insertion site of a backbone.
- **Part in Backbone:** A backbone with at least one insertion site where a part insert has been incorporated.
- **Part Extract:** A part, plus any 5' or 3' flanking sequences, that has been extracted from a part in backbone as part of an assembly process.
- **Assembly:** A plan for combining a set of parts in order to build one or more composite parts.

In the iGEM Engineering Committee, we found that agreeing on this common terminology greatly reduced the amount of confusion, and use of these terms has become commonplace in our multi-institution collaboration.

3 REPRESENTING ASSEMBLY PLANS IN SBOL

To facilitate better tool support for planning and communicating build information, we mapped the vocabulary and build plans shown in Figure 1 onto the SBOL 3 standard [4], which we found to provide all of the concepts necessary for a succinct representation. Here we present a summary of key points; full details are available as SBOL Best Practice Proposal (BPP) 001 in the SBOL Examples collection at https://github.com/SynBioDex/SBOL-examples/pull/4.
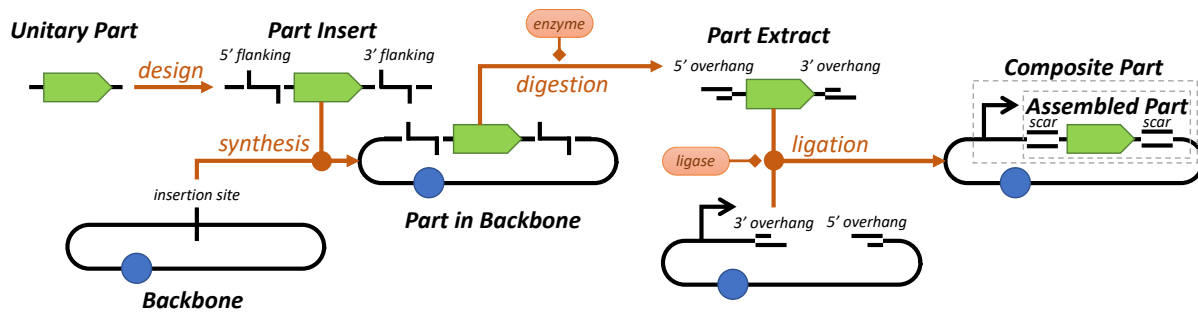
Figure 1: Proposed build terminology, illustrated on a typical digestion/ligation build workflow: a *unitary part* is extended with flanking sequences needed for assembly to create a *part insert* that can be synthesized or assembled into an insertion point on a *backbone* to produce a *part in backbone* ready for assembly. Digestion produces a *part extract* that can be ligated together with other part extracts to produce a *composite part* in backbone, including the original part as an *assembled part* in its final context.

In this representation, each part is an SBOL `Component`, and the distinction between a unitary part and a composite part can be made by using the `prov:wasGeneratedBy` property to link any composite part to a `prov:Activity` representing an assembly plan, as described below. An assembled part within the composite part is represented by an appropriate `Feature` (typically a `SubComponent`), and similarly a scar is a `SequenceFeature` with its role set to the Sequence Ontology (SO) term for assembly scar.

A backbone is also represented by an SBOL `Component`, but has a role indicating its use as a backbone, such as `SO:plasmid_vector`. An insertion site or drop-out sequence is indicated using a `Feature` with the corresponding role, respectively `SO:insertion_site` and `SO:deletion`. Part in backbone and part insert are much the same, represented with a `Component` that includes a `SequenceFeature` for each restriction site (with SO restriction site terms), while a part extract will typically have features for overhangs.

Finally, an assembly plan is represented by a `prov:Activity` with appropriate typing and a link to an SBOL `Component` describing the network of digestion and ligation reactions for the assembly. Each reaction can be described by an `Interaction`, with each reactant, enzyme, and product a `Participant`. Digestion uses type `SBO:cleavage`, with the part in backbone and enzyme having role `SBO:reactant` and the part extract having role `SBO:product`. Ligation uses type `SBO:conversion`, with the part extracts and ligase being the reactants and the composite part in backbone being the product. Many composite parts will be described with just one digestion/ligation stage, but a more complex assembly may have multiple digestion and ligation stages and may have multiple products.

4 FUTURE DIRECTIONS

The proposal has met with general consensus during community review and is currently in process of being adopted as a best practice officially endorsed by the SBOL community. A full supporting Python API is currently being implemented for the SBOL Utilities library (https://github.com/SynBioDex/ SBOL-utilities). This implementation is intended to form the basis for integration of these representations with laboratory automation. Finally, while the current proposal has been worked out specifically with regards to Type IIS and Bio-Bricks assembly methods, we believe it is likely to extend well to other assembly methods as well, such as Gibson Assembly [2] or Ligase Cycling Reaction Assembly [3], though certain details will likely need to be adjusted.

5 ACKNOWLEDGEMENTS

REFERENCES

[1] Engler, C., Kandzia, R., and Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PloS one 3*, 11 (jan 2008), e3647.

[2] Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. Enzymatic assembly of dna molecules up to several hundred kilobases. *Nature methods 6*, 5 (2009), 343–345.

[3] Kok, S. d., Stanton, L. H., Slaby, T., Durot, M., Holmes, V. F., Patel, K. G., Platt, D., Shapland, E. B., Serber, Z., Dean, J., et al. Rapid and reliable dna assembly via ligase cycling reaction. *ACS synthetic biology 3*, 2 (2014), 97–106.

[4] McLaughlin, J. A., Beal, J., Misirli, G., Grünberg, R., Bartley, B. A., Scott-Brown, J., Vaidyanathan, P., Fontanarrosa, P., Oberortner, E., Wipat, A., et al. The synthetic biology open language (sbol) version 3: simplified data exchange for bioengineering. *Frontiers in Bioengineering and Biotechnology 8* (2020), 1009.

[5] Sarrion-Perdigones, A., Falconi, E. E., Zandalinas, S. I., Juárez, P., Fernández-del Carmen, A., Granell, A., and Orzaez, D. Goldenbraid: an iterative cloning system for standardized assembly of reusable genetic modules. *PloS one 6*, 7 (2011), e21622.

[6] Shetty, R., Lizarazo, M., Rettberg, R., and Knight, T. F. Assembly of biobrick standard biological parts using three antibiotic assembly. In *Methods in enzymology*, vol. 498. Elsevier, 2011, pp. 311–326.

[7] Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. A modular cloning system for standardized assembly of multigene constructs. *PloS one 6*, 2 (2011), e16765.