BBF RFC 16: Synthetic Biology Open Language Visual (SBOLv) Specification

Cesar Rodriguez, Suzie Bartram, Anusuya Ramasubramanian, Drew Endy

15 October 2009

1. Purpose

Synthetic Biology Open Language Visual (SBOLv) is a graphical notation that supports biological device development. It provides a formal notation for describing the physical composition of basic parts into composite parts during the development of biological devices. It is targeted for use by biological engineers in forward engineering projects. It encourages and supports model-driven engineering.

2. Relation to other BBF RFCs

RFC 16 replaces RFC 4, RFC 17, and RFC 18.

3. Copyright Notice

4. Background

Synthetic Biology Open Language Visual (SBOLv) is a graphical notation for the formal specification of basic and composite parts used in the development of biological devices. It is targeted for use by biological engineers in model-driven, forward engineering projects. SBOLv is an open standard that balances the benefits of top-down control with the bottom-up needs to communicate. The notation is flourishing in an environment of open communication while rigorously defining symbols, syntax, and grammar. SBOLv is an abstract tool for solving concrete problems. It has a rigorous and consistent structure and is being used in concrete software tools while working on concrete Synthetic Biology projects.

SBOLv is a component of the Synthetic Biology Open Language (SBOL), an emerging open standard for structuring data and exchanging information between members of the Synthetic Biology Community. SBOL has four components; a relational data model (Core Data Model), a semantic model (SBOL Semantic), a scripting language (SBOL Script), and a graphical notation (SBOL Visual). SBOL promises to offer biological engineers a comprehensive set of standards and tools in support of data exchange and communication.

5. Specification

5.1 General Geometric Pattern of the Symbols

The canvas of all the symbols is a 2 inch square. The 3 point line centered along the y-axis and extending the full width of the x-axis represents a double-stranded DNA sequence. Geometric shapes above the line represent features on the forward strand while shapes below the line represent features on the reverse strand.
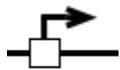
5.2 Central Dogma Symbols
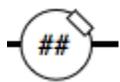
Origin of Replication
The circle represents a plasmid. The rectangle indicates the site of the replication origin on the plasmid. An indication of the plasmid copy number is OPTIONAL but MUST be located in the center of the circle. Molecules per cell is the RECOMMENDED scale.

Shorthand Origin of Replication
Shorthand version of the Origin of Replication. The rectangle is omitted. This symbol SHOULD be used for rapid drawing and communication. However, the Origin of Replication symbol is RECOMMENDED.
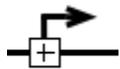
Forward Constitutive Promoter
Represents a DNA sequence that promotes RNA polymerase binding and transcription in the forward strand. The open square indicates constitutive transcription.
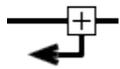
Reverse Constitutive Promoter
Represents a constitutive promoter initiating transcription in the reverse strand.
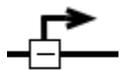
Forward Inducible Promoter
Represents a promoter that requires induction before promoting transcription in the forward strand. The "+" symbol in the square indicates that the promoter is inducible.
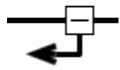
Reverse Inducible Promoter
Represents a promoter that requires induction before promoting transcription in the reverse strand. The "+" symbol in the square indicates that the promoter is inducible.

Forward Repressible Promoter
Represents a promoter where transcription in the forward strand can be repressed. The "-" symbol in the square indicates that the promoter is repressible.

Reverse Repressible Promoter
Represents a promoter where transcription in the reverse strand can be repressed. The "-" symbol in the square indicates that the promoter is repressible.
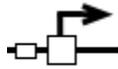
Forward Promoter with Downstream Operator
Represents a forward promoter with a downstream operator. The OPTIONAL "+" symbol MUST be located in the square and indicates an inducible promoter. The OPTIONAL "-" symbol MUST be located in the square and indicates a repressible promoter. The rectangle to the right of the square indicates a downstream operator.

Information about the specific location and function of the operator is OPTIONAL, but MUST be located within the rectangle.
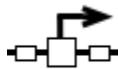
Forward Promoter with Upstream Operator

Represents a forward promoter with an upstream operator. The OPTIONAL "+" symbol MUST be located in the square and indicates an inducible promoter. The OPTIONAL "-" symbol MUST be located in the square and indicates a repressible promoter. The rectangle to the left of the square indicates an upstream operator. Information about the specific location and function of the operator is OPTIONAL, but MUST be located within the open rectangle.

Forward Promoter with Upstream and Downstream Operators

Represents a forward promoter with upstream and downstream operators. The OPTIONAL "+" symbol MUST be located in the square and indicates an inducible promoter. The OPTIONAL "-" symbol MUST be located in the square and indicates a repressible promoter. The rectangles to the left and right of the square indicate upstream and downstream operators, respectively. Information about the specific location and function of the operators are OPTIONAL, but MUST be located within the rectangles.

Forward Open Reading Frame

Represents a forward open reading frame (ORF). The polygon indicates the single continuous DNA fragment of the ORF. The angled right wall indicates the forward direction of the reading frame.

Reverse Open Reading Frame

Represents a reverse open reading frame (ORF). The polygon indicates the single continuous DNA fragment of the ORF. The angled left wall indicates the reverse direction of the reading frame.

Forward Translation Start Site

Represents a 5' to 3' Shine-Delgarno sequence in the forward strand of prokaryotes, IRES in viruses and ribosome loading structure in eukaryotes.

Reverse Translation Start Site

Represents a 5' to 3' Shine-Delgarno sequence on the reverse strand in prokaryotes and equivalent sites or structures in eukaryotes.
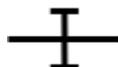
Forward Terminator

Represents the location on a DNA sequence where transcription ends in the forward strand.

Reverse Terminator

Represents the location on a DNA sequence where transcription ends in the reverse strand.

Bidirectional Terminator

Represents the location on a DNA sequence where transcription ends in both strands.

Ribonuclease Site

Represents the location on a DNA sequence that codes for a ribonuclease cleavage site on the RNA. The dashed line indicates a functional element located on the RNA. The "x" symbol indicates cleavage of the RNA.

RNA Stability Element, "Stable"

Represents a location on a DNA sequence that codes for an RNA secondary structure that effects the stability of the RNA. The dashed line indicates a functional element located on the RNA. The diameter of the circle SHOULD indicate the half life of the RNA. A log scale is RECOMMENDED. The larger circle indicates qualitatively an RNA molecule with a long half-life.
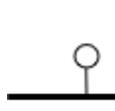
RNA Stability Element, "Intermediate"

Represents a location on a DNA sequence that codes for an RNA secondary structure that effects the stability of the RNA. The dashed line indicates a functional element located on the RNA. The diameter of the circle SHOULD indicate the half life of the RNA. A log scale is RECOMMENDED. The smaller circle indicates qualitatively an RNA molecule with an intermediate half-life.

RNA Stability Element, "Unstable"

Represents the location on a DNA sequence that codes for an RNA secondary structure that effects the stability of the RNA. The dashed line indicates a functional element located on the RNA. The diameter of the circle SHOULD indicate the half-life of the RNA. A log scale is RECOMMENDED. The smallest circle indicates qualitatively an RNA molecule with a short life span.
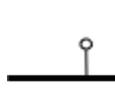
Protease Site

Represents the location on a DNA sequence that codes for a protease cleavage site. The solid line indicates a functional element located on a peptide/protein. The "x" symbol indicates cleavage of a peptide or protein.

Protein Degradation Element "Stable"

Represents a location on a DNA sequence that codes for a degradation-related structure in a peptide/protein . The solid line indicates a functional element located on the peptide/protein. The diameter of the circle SHOULD indicate the half-life of the peptide/protein. A log scale is RECOMMENDED. The larger circle indicates qualitatively a peptide/protein with a long half-life.

Protein Degradation Element "Intermediate"

Represents a location on a DNA sequence that codes for a degradation-related structure in a peptide/protein . The solid line indicates a functional element located on the peptide/protein. The diameter of the circle SHOULD indicate the half-life of the peptide/protein. A log scale is RECOMMENDED. The smaller circle indicates qualitatively a peptide/protein with an intermediate half-life.

Protein Degradation Element "Unstable"

Represents a location on a DNA sequence that codes for a degradation-related structure in a peptide/protein . The solid line indicates a functional element located on the peptide/protein. The diameter of the circle SHOULD indicate the half-life of the

peptide/protein. A log scale is RECOMMENDED. The smallest circle indicates qualitatively a peptide/protein with a short half-life.

5.2 Genetic Engineering Symbols

Forward Primer Site
Represents a site for primer binding in the forward strand.

Reverse Primer Site
Represents a site for primer binding in the reverse strand.
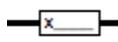
Forward Barcode
Represents a location on a DNA sequence where developer information is encoded on the forward strand. The encoded information can, for example, include the name of the individual or team that developed the standard biological part.
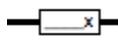
Reverse Barcode
Represents a location on a DNA sequence where developer information is encoded on the reverse strand. The encoded information can, for example, include the name of the individual or team that developed the standard biological part.
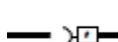
Prefix
Represents the prefix sequence of an assembly standard. The identifier of the specific prefix sequence used is OPTIONAL, but it MUST be located in the rectangle.

Shorthand Prefix
Represents the prefix sequence of an assembly standard. The rectangle found in the Prefix symbol is excluded.
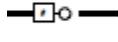
Suffix
Represents the suffix sequence of an assembly standard. The identifier of the specific prefix sequence used is OPTIONAL, but it MUST be located in the rectangle.

Shorthand Suffix
Represents the prefix sequence of an assembly standard. The rectangle found in the Suffix symbol is excluded.
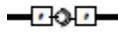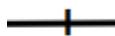
Scar
Represents where a suffix and prefix sequence were ligated.
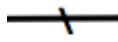
3' Overhang Restriction Site
Represents a restriction endonuclease site that results in 3' overhangs. The diagonal line indicates the geometry of the cut site and the 3' sticky ends.
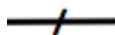
5' Overhang Restriction Site
Represents a restriction endonuclease site that results in 5' overhangs. The diagonal line indicates the geometry of the cut site and the 5' sticky ends.

Blunt Restriction Site
Represents a restriction endonuclease site that results in no overhangs. The straight line indicates the geometry of the cut site.

6. Documentation, Extension, and Maintenance of SBOLv

Documentation, support material for the open extension process, and maintenance-related information are located on OpenWetWare at the following location:

http://openwetware.org/wiki/Endy:Notebook/Synthetic_Biology_Open_Language

Any registered member of the OpenWetWare community can submit a new symbol or modify an existing one. There is an area reserved for comments. The submitted symbols and comments are reviewed by the authors of this RFC. Each symbol is given a status:

- Request for Proposals
- Under Development
- Proposed
- Consensus
- Included in v. #
- Deprecated

"Request for Proposals" indicates that a symbol is needed but their isn't consensus on the appearance of the symbol. "Under Development" indicates that a rough draft of a symbol was submitted. "Proposed" indicates that a final version of a symbol was submitted and the author is seeking comments. "Consensus" indicates that a rough consensus has been reached that the symbol is a candidate for inclusion in an official release of SBOLv. "Included in v. #" indicates the version of SBOLv where the symbol was included. "Deprecated" indicates that a symbol is not being considered for inclusion in SBOLv.

A new version of SBOLv is released every 3-6 months. A release includes, 300 dpi JPG images, 300 dpi PNG images, and PDF vector images of all the symbols included in the new version of SBOLv.

7. Acknowledgements

It takes many individuals to nurture the birth of a new language. We are indebted to our colleagues; Lesia Bilitchenko, Jerome Bonnet, Barry Canton, Deepak Chandran, Douglas Densmore, Joanna Chen, Eric Fernandez, Tim Ham, Raik Gruenberg, Jason Kelly, Adam Liu, Richard Mar, Lance Martin, Alec Nielsen, Robert Ovadia, Randy Rettberg, Herbert Sauro,

8. Author's Contact Information

Suzie Bartram: sbartram@stanford.edu
Drew Endy: endy@stanford.edu
Anusuya Ramasubramanian: anusuyar@stanford.edu
Cesar Rodriguez: cesarr@stanford.edu

9. References

Novere, et al. Systems Biology Graphical Notation. Nature Biotechnology 2009 vol. 27 (8) pp. 735-41

Matsuoka, et al. Consistent design schematics for biological systems: standardization of representation in biological engineering. Journal of the Royal Society Interface. EPub June 3 2009

Shetty, et al. Engineering BioBrick vectors from BioBrick parts. Journal of Biological Engineering. 2008; 2: 5.

Voigt. Genetic parts to program bacteria. Current Opinion in Biotechnology, 17(5):548–57, Oct 2006.

Andrianantoandro, et al. Synthetic biology: new engineering rules for an emerging discipline. Molecular Systems Biology. 2006 vol. 2 pp. 2006.0028.

Endy. Foundations for engineering biology. Nature. 2005 vol. 438 (7067) pp. 449-53.