Robert Sidney Cox[1] / Curtis Madsen[2] / James McLaughlin[3] / Tramy Nguyen[4] / Nicholas Roehner[5] / Bryan Bartley[6] / Swapnil Bhatia[2] / Mike Bissell[7] / Kevin Clancy[8] / Thomas Gorochowski[9] / Raik Grünberg[10] / Augustin Luna[11] / Nicolas Le Novère[12] / Matthew Pocock[13] / Herbert Sauro[6] / John T. Sexton[14] / Guy-Bart Stan[15] / Jeffrey J. Tabor[14] / Christopher A. Voigt[16] / Zach Zundel[4] / Chris Myers[4] / Jacob Beal[17] / Anil Wipat[18]

Synthetic Biology Open Language Visual (SBOL Visual) Version 2.0

[1] Kobe University, Kobe, Japan
[2] Boston University, Boston, MA, USA
[3] Newcastle University, Newcastle, United Kingdom of Great Britain and Northern Ireland
[4] University of Utah, Salt Lake City, UT, USA
[5] Raytheon BBN Technologies, Cambridge, MA, USA
[6] University of Washington, Seattle, WA, USA
[7] Amyris, Inc., Emeryville, CA, USA
[8] Thermo Fisher Scientific, San Diego, CA, USA
[9] University of Bristol, Bristol, United Kingdom of Great Britain and Northern Ireland
[10] King Abdullah University of Science and Technology (KAUST), BESE, Thuwal 23955 - 6900, Saudi Arabia, E-mail: raik.grunberg@kaust.edu.sa
[11] Harvard Medical School, Boston, MA, USA
[12] Babraham Institute, Cambridge, Cambridgeshire, United Kingdom of Great Britain and Northern Ireland
[13] Turing Ate My Hamster, Ltd., Newcastle, United Kingdom of Great Britain and Northern Ireland
[14] Rice University, Houston, TX, USA
[15] Imperial College, London, United Kingdom of Great Britain and Northern Ireland
[16] Massachusetts Institute of Technology, Cambridge, MA, USA
[17] Raytheon BBN Technologies, Cambridge, MA, USA. http://orcid.org/0000-0002-1663-5102.
[18] Newcastle University, Newcastle, United Kingdom of Great Britain and Northern Ireland, E-mail: sbol-editors@googlegroups.com

Abstract:
People who are engineering biological organisms often find it useful to communicate in diagrams, both about the structure of the nucleic acid sequences that they are engineering and about the functional relationships between sequence features and other molecular species. Some typical practices and conventions have begun to emerge for such diagrams. The Synthetic Biology Open Language Visual (SBOL Visual) has been developed as a standard for organizing and systematizing such conventions in order to produce a coherent language for expressing the structure and function of genetic designs. This document details version 2.0 of SBOL Visual, which builds on the prior SBOL Visual 1.0 standard by expanding diagram syntax to include functional interactions and molecular species, making the relationship between diagrams and the SBOL data model explicit, supporting families of symbol variants, clarifying a number of requirements and best practices, and significantly expanding the collection of diagram glyphs.

Keywords: SBOL Visual, Standards, Diagrams
DOI: 10.1515/jib-2017-0074
Received: December 1, 2017; **Accepted:** February 1, 2018

Anil Wipat is the corresponding author.

BBF RFC 115: Synthetic Biology Open Language Visual (SBOL Visual) Version 2.0

Editors:

Robert Sidney Cox — *Kobe University, Japan*
Curtis Madsen — *Boston University, USA*
James McLaughlin — *Newcastle University, UK*
Tramy Nguyen — *University of Utah, USA*
Nicholas Roehner — *Raytheon BBN Technologies, USA*

Chair:

Anil Wipat — *Newcastle University, UK*

editors@sbolstandard.org

Additional authors, by last name:

Bryan Bartley — *University of Washington, USA*
Jacob Beal — *Raytheon BBN Technologies, USA*
Swapnil Bhatia — *Boston University, USA*
Mike Bissell — *Amyris, Inc., USA*
Kevin Clancy — *Thermo Fisher Scientific, USA*
Thomas Gorochowski — *University of Bristol, UK*
Raik Grunberg — *KAUST, Saudi Arabia*
Augustin Luna — *Harvard Medical School, USA*
Chris Myers — *University of Utah, USA*
Nicolas Le Novere — *Babraham Institute, UK*
Matthew Pocock — *Turing Ate My Hamster, Ltd., UK*
Herbert Sauro — *University of Washington, USA*
John T. Sexton — *Rice University, USA*
Guy-Bart Stan — *Imperial College, UK*
Jeffrey J. Tabor — *Rice University, USA*
Chris Voigt — *MIT, USA*
Zach Zundel — *University of Utah, USA*

Version 2.0

December 1, 2017

Contents

1 Purpose

People who engineer biological organisms often find it useful to communicate in diagrams, both about the structure of the nucleic acid sequences that they are engineering and about the functional relationships between sequence features and other molecular species. Some typical practices and conventions have begun to emerge for such diagrams. SBOL Visual aims to organize and systematize such conventions in order to produce a coherent language for expressing the structure and function of genetic designs. At the same time, we aim to make this language simple and easy to use, allowing a high degree of flexibility and freedom in how such diagrams are organized, presented, and styled—in particular, it should be readily possible to create diagrams both by hand and with a wide variety of software programs. Finally, means are provided for extending the language with new and custom diagram elements, and for adoption of useful new elements into the language.

1.1 Relation to Data Models

In order to ground SBOL Visual with precise definitions, we reference its visual elements to data models with well-defined semantics. In particular, glyphs in SBOL Visual are defined in terms of their relation to the SBOL 2 data model (as defined in BBF RFC 112) and terms in the Sequence Ontology (Eilbeck et al., 2005), the Systems Biology Ontology (Courtot et al., 2011), and BioPAX (Goldberg et al., 2010).

SBOL Visual is not intended to represent designs at the same level of detail as these data models. Effective visual diagrams are necessarily more abstract, focusing only on those aspects of a system that are the subject of the communication. Nevertheless, we take as a principle that it should be possible to transform any SBOL Visual diagram into an equivalent (if highly abstract) SBOL 2 data representation. Likewise, we require that SBOL Visual should be able to represent all of the significant structural or functional relationships in any GenBank or SBOL data representation.

2 Relation to other BBF RFCs and other Standards

BBF RFC 115 replaces BBF RFC 93 (SBOL Visual 1.0).

BBF RFC 115 also implicitly supersedes the previously replaced BBF RFC 16 (a prior version of SBOL Visual, replaced by BBF RFC 93).

Every glyph in SBOL Visual 2.0 corresponds to an element of the SBOL 2.1 data model, as defined in BBF RFC 112. BBF RFC 115 also defines many terms by reference to BBF RFC 112, or by reference to the Sequence Ontology (Eilbeck et al., 2005), the Systems Biology Ontology (Courtot et al., 2011), or BioPAX (Goldberg et al., 2010).

SBOL Visual is intended to be compatible with the Systems Biology Graphical Notation Activity Flow Language (SBGN AF) (Le Novère et al., 2009), and species and interaction glyphs have been imported from that language (see: Appendix A.2 and Appendix A.3). Some aspects are also imported from the Systems Biology Graphical Notation Process Description Language (SBGN PD).

3 Copyright and License Statement

4 SBOL Specification Vocabulary

4.1 Term Conventions

This document indicates requirement levels using the controlled vocabulary specified in IETF RFC 2119 and reiterated in BBF RFC 0. In particular, the key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119:

- The words "MUST", "REQUIRED", or "SHALL" mean that the item is an absolute requirement of the specification.
- The phrases "MUST NOT" or "SHALL NOT" mean that the item is an absolute prohibition of the specification.
- The word "SHOULD" or the adjective "RECOMMENDED" mean that there might exist valid reasons in particular circumstances to ignore a particular item, but the full implications need to be understood and carefully weighed before choosing a different course.
- The phrases "SHOULD NOT" or "NOT RECOMMENDED" mean that there might exist valid reasons in particular circumstances when the particular behavior is acceptable or even useful, but the full implications need to be understood and the case carefully weighed before implementing any behavior described with this label.
- The word "MAY" or the adjective "OPTIONAL" mean that an item is truly optional.

4.2 SBOL Class Names

The definition of SBOL Visual references several SBOL classes, which are defined as listed here. For full definitions and explanations, see BBF RFC 112, describing the SBOL 2.1 data model.

ComponentDefinition: Describes the structure of designed entities, such as DNA, RNA, and proteins, as well as other entities they interact with, such as small molecules or environmental properties.

- *Component*: Pointer class. Incorporates a child ComponentDefinition *by reference* into exactly one parent ComponentDefinition. Represents a specific occurrence or instance of an entity within the design of a more complex entity. Because the same definition might appear in multiple designs or multiple times in a single design, a single ComponentDefinition can have zero or more parent ComponentDefinitions, and each such parent-child link requires its own, distinct Component.
- *Location*: Specifies the base coordinates and orientation of a genetic feature on a DNA or RNA molecule or a residue or site on another sequential macromolecule such as a protein.
- *SequenceAnnotation*: Describes the Location of a notable sub-sequence found within the Sequence of a ComponentDefinition. Can also link to and effectively position a child Component.
- *SequenceConstraint*: Describes the relative spatial position and orientation of two Component objects that are contained within the same ComponentDefinition.

ModuleDefinition: Describes a “system” design as a collection of biological components and their functional relationships.

- *FunctionalComponent*: Pointer class. Incorporates a child ComponentDefinition *by reference* into exactly one parent ModuleDefinition. Represents a specific occurrence or instance of an entity within the design of a system. Because the same definition might appear in multiple designs or multiple times in a single design, a single ComponentDefinition can have zero or more parent ModuleDefinitions, and each such parent-child link requires its own, distinct FunctionalComponent.

- *Interaction*: Describes a functional relationship between biological entities, such as regulatory activation or repression, or a biological process such as transcription or translation.
- *Participation*: Describes the role that a `FunctionalComponent` plays in an `Interaction`. For example, a transcription factor might participate in an `Interaction` as a repressor or as an activator.

5 SBOL Glyphs

A glyph is a visual symbol used to represent an element in an SBOL Visual diagram. All of the currently defined glyphs are collected in Appendix A. This section explains how glyphs are specified and how to add new glyphs.

Each SBOL glyph is defined by association with ontology terms, and can be used to represent any diagram element that is well-described by that term. Currently there are three classes of glyphs, each associated with a different ontology and different class in the SBOL 2 data model:

- **Sequence Feature Glyphs** describe features of nucleic acid sequences. They are associated with Sequence Ontology terms. For the SBOL 2 data model, this is formally defined as any `Component` with a compatible term within its associated `roles`, i.e., one that is equal to or a child of at least one term associated with the glyph.
- **Molecular Species Glyphs** represent any class of molecule whose detailed structure is not being shown using sequence feature glyphs. They are associated with BioPAX terms. For the SBOL 2 data model, this is formally defined as any `FunctionalComponent` with a compatible term within its associated `types`, i.e., one that is equal to or a child of at least one term associated with the glyph.
- **Interaction Glyphs** are "arrows" indicating functional relationships between sequence features and/or molecular species. They are associated with Systems Biology Ontology terms. For the SBOL 2 data model, this is formally defined as any `Interaction` with a compatible term within its `types`, i.e., one that is equal to or a child of at least one term associated with the glyph, and with a compatible `Participation` at the head and tail of the arrow.

More than one glyph may share the same definition: in this case, these glyphs form a family of variants, of which precisely one MUST be designated as the RECOMMENDED glyph, which is to be used unless there are strong reasons to prefer an alternative variant.

It will also frequently be the case that a diagram element could be represented by more than one glyph (e.g., a glyph for a specific term and a glyph for a more general term). In such cases, it is RECOMMENDED that the most specific applicable glyph be used.

For example, a protein coding sequence (CDS) is a sequence feature that may be represented either using the CDS glyph (Sequence Ontology term SO:0000316) or the Unspecified glyph (Sequence Ontology term SO:0000001). Since SO:0000316 is contained by SO:0000001, the preferred glyph is CDS, rather than Unspecified. Likewise, a CDS may be represented by either a pentagonal glyph or an arrow glyph, but the pentagon is the RECOMMENDED variant, and so it is likewise preferred. Figure 1 illustrates this example.

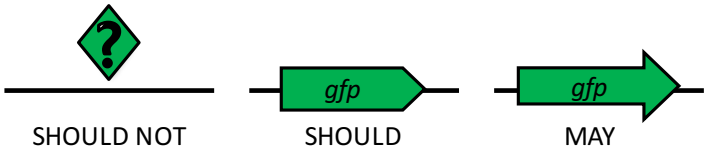


Figure 1: *A biological design element such as a protein coding sequence (CDS) is best represented by the most specific RECOMMENDED glyph (middle), but can be represented by a less specific glyph such as Unspecified (left) or an approved alternative glyph (right).*

5.1 Requirements for Glyphs

A number of requirements are placed on all SBOL Visual glyphs in order to ensure both the clarity of diagrams and the ease with which they can be constructed:

1. A glyph SHOULD have its meaning defined by associating the glyph with at least one ontology definition. Definitions are RECOMMENDED to be from the Sequence Ontology for nucleic acid components, from BioPAX for other components, and from the Systems Biology Ontology for interactions. If no applicable terms are available in the preferred ontology, proposal of a new glyph SHOULD be accompanied by a request to the ontology maintainers to add a term for the undefined entity.
2. A glyph SHOULD be relatively easy to sketch by hand (e.g., no high-complexity images or precise angles required).
3. A glyph specification MUST indicate which portions of the glyph are the "interior" for purposes of color fill.
4. A glyph specification SHOULD show the glyph in its preferred relative scale with respect to other glyphs.
5. A glyph SHOULD be specified using only solid black lines (leaving color and style to be determined by the user, as noted below).
6. A glyph SHOULD NOT be similar enough to be easily confused with any other glyph when written by hand, or when scaled either vertically, horizontally, or both.
7. A glyph SHOULD NOT include text (note that associated labels are not part of the glyph).

In addition, some requirements apply only to certain classes of glyphs:

8. A sequence feature or molecular species glyph specification MUST include a rectangular bounding box indicating its extent in space.
9. A sequence feature glyph specification MUST include exactly one horizontal rule for its RECOMMENDED vertical alignment with the nucleic acid backbone.
10. A sequence feature glyph SHOULD be asymmetric on the horizontal axis. Vertical asymmetry is also preferred when possible.
11. If a sequence feature glyph can represent components of highly variable size or structural complexity, the glyph SHOULD be able to be scaled horizontally to indicate relative property value.

Figure 2 shows examples of compliant glyph specification.

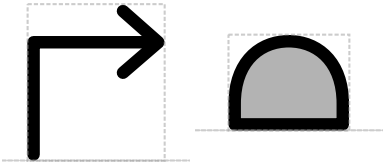


Figure 2: *Examples of glyph specification: this specification for the sequence feature glyphs for Promoter (left) and Ribosome Entry Site (right) include the glyph outline, fill (grey center of Ribosome Entry Site), bounding box (dashed box), and recommended alignment with the nucleic acid backbone (dashed horizontal line), all at a preferred relative scale.*

5.2 Reserved Visual Properties

SBOL Visual aims to allow as much flexibility and freedom as possible in how diagrams are organized, presented, and styled. To this end, a number of aspects of presentation are generally reserved for the communication of other types of information by the creator of a diagram. When using a glyph in a diagram, the following choices in glyph presentation are thus explicitly intended to be alterable:

1. The lines of a glyph MAY be given any line thickness and style
2. The interior of a glyph MAY be given any fill color, as long as the choice of fill does not interfere with recognizing the glyph.
3. The scale of glyphs are RECOMMENDED to be kept consistent with their specification and throughout a diagram, but can be altered if desired, particularly to convey additional information (e.g., length of a sequence).
4. Minor styling effects MAY be chosen (e.g., shadow, corner styling, other "font-level" customization)

Figure 3 shows some examples of acceptable style variation.

In certain special cases, the style of a glyph may be more constrained, but such cases are expected to be rare and strongly motivated.

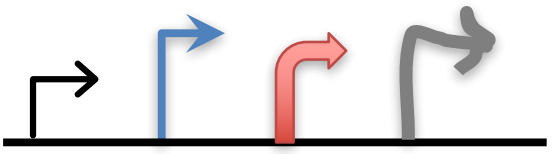


Figure 3: *Examples of acceptable style variation for a Promoter glyph.*

5.3 Extending the Set of Glyphs

The collection of SBOL Visual glyphs is not expected to provide complete coverage of all of the types of element that people will wish to include in genetic diagrams, particularly given the ongoing evolution of synthetic biology as an engineering discipline. As the need for new diagram elements or new practices of usage emerge, new glyphs or glyph definitions are expected to be added to SBOL Visual. In particular, the following three classes of changes are expected to occur regularly, and the SBOL development community will maintain clear processes for proposal and adoption of changes of this type:

- New glyphs, either representing a type of component that previously lacked a glyph or enabling a distinction between types of components previously represented by the same glyph.
- Additional glyph variants, accompanied by compelling use cases that cannot be adequately addressed by the existing glyph variants.
- Additional definitions for a glyph, capturing an alternate meaning that is useful to humans but existing within a disjoint branch of the relevant ontology.

In order to support the coherent extension of SBOL Visual, whenever a diagram creator uses a glyph not found in Appendix A, the creator SHOULD submit it to be considered for inclusion in an updated version of the standard following the processes for adding new glyphs found on the community website at `http://sbolstandard.org`

6 SBOL Visual Diagram Language

An SBOL Visual diagram represents information about the structure of a nucleic acid design and its associated molecular species and interactions. If desired, an SBOL Visual diagram may also be associated with a machine-interpretable model (e.g., in SBOL, GenBank, or SBML format). In this document we describe the association for the SBOL 2 data model, which provides a formal semantic grounding for all elements of an SBOL Visual diagram, but equivalent associations may be made between diagram elements and other models. In terms of the SBOL 2 data model, the description of a nucleic acid design is formally defined as a representation of a `ComponentDefinition` with a nucleic acid `type`, the `Component` and `SequenceAnnotation` objects describing the features and sub-structure of the design, and `SequenceConstraint` information on the relative positions of such elements. The description of interactions between some number of nucleic acid designs and other molecular species is formally defined as a representation of a `ModuleDefinition`, the `FunctionalComponent` objects describing the nucleic acid designs and other molecular species, and the `Interaction` and `Participation` objects describing their functional relationships.

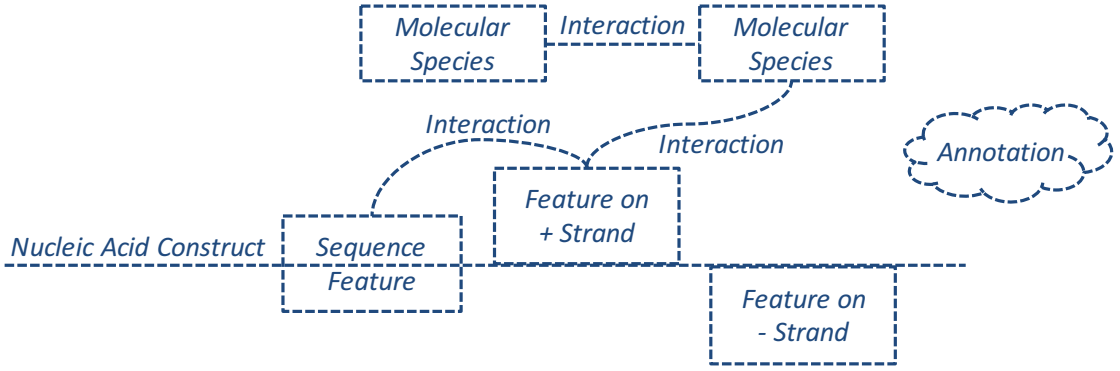


Figure 4: *Generic syntax of SBOL Visual 2: a diagram for a nucleic acid construct is based around a backbone line, its structure specified by the sequence of attached sequence feature glyphs. Strand can optionally be indicated by placing a glyph above or below the backbone. Other molecular species are indicated by glyphs not in contact with any backbone. Interactions are directed edges connecting sequence feature or molecular species glyphs. Any of these objects may have an associated label showing its name, and the diagram may further include any form of other annotations, including other types of text.*

Specifically, an SBOL Visual diagram consists of the classes of objects illustrated in Figure 4. Figure 5 shows an example of such a diagram, in a typical usage. Full details of this specification are provided in the remainder of this section.

6.1 Nucleic Acid Backbone

A diagram for a nucleic acid construct is based around a single or double line, representing the nucleic acid backbone. Information about features of the construct can then be represented by attaching nucleic acid glyphs to the backbone, as defined below in Section 6.2. In terms of the SBOL 2 data model, the backbone represents a `ComponentDefinition` with a nucleic acid `type` (e.g., DNA, RNA), and the features represent `Component` and `SequenceAnnotation` members of the `ComponentDefinition`.

1. Lines in some cases indicate strand count. A double-stranded region of the nucleic acid construct MAY use either a single or double line for the backbone. A single-stranded region of the nucleic acid construct MUST use a single line to indicate the backbone. When single and double lines are mixed within a single diagram, the single lines always indicate single-stranded regions. Examples are provided in Figure 6.

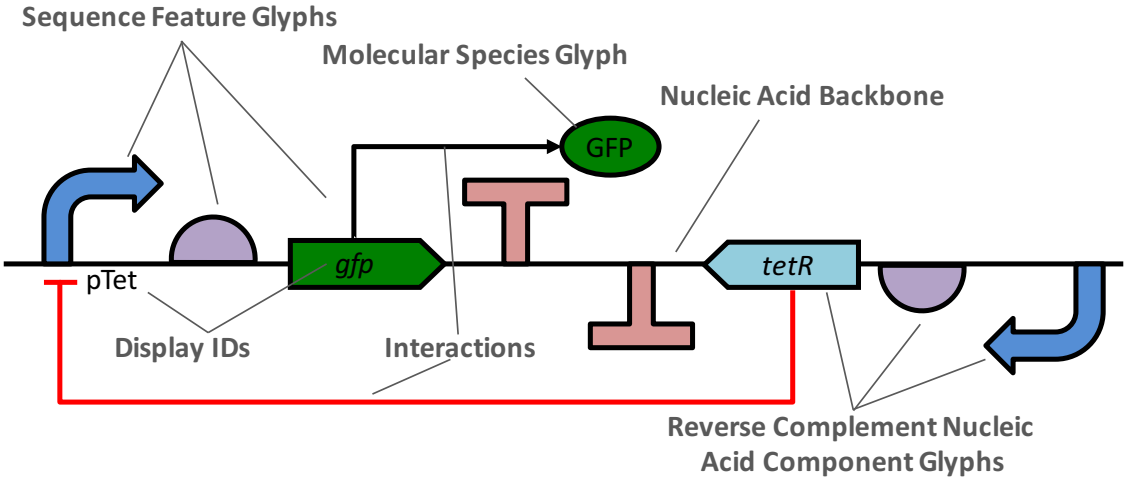


Figure 5: *Example illustrating the elements of an SBOL Visual 2 diagram, with nucleic acid sequence features on the forward and reverse strand of a backbone, other molecular species, and interactions between elements; the grey labels and indicator lines are annotations.*

(a) Single- or double-strand backbone (b) Double-strand backbone (c) Double-strand backbone with single-strand overhangs

Figure 6: *Examples of indicating strand count in nucleic acid backbones.*

2. A nucleic acid backbone SHOULD be horizontal in orientation, but MAY use non-horizontal structure to indicate important physical attributes (e.g., a closed loop to indicate a cyclic plasmid or more complex shapes for DNA nanotech structures). Examples are provided in Figure 7.

3. A nucleic acid backbone SHOULD have at least one associated feature glyph (else no structural information is being provided).

6.2 Nucleic Acid Sequence Features

A glyph in contact with a nucleic acid backbone indicates a feature of the nucleic acid sequence. In terms of the SBOL 2 data model, this is either a `SequenceFeature` or a `Component` with a nucleic acid `type` that is contained within the `ComponentDefinition` associated with that nucleic acid backbone. The `Component` may be contained either directly, as one of the `components` of the `ComponentDefinition`, or recursively through a sequence of such containments.

1. Every feature glyph MUST have its bounding box in contact with the backbone for the nucleic acid construct it describes. The placement of the glyph SHOULD follow the recommendation for backbone alignment in the glyph specification. Examples are provided in Figure 8.

2. The horizontal orientation of a glyph can be used to indicate the strand alignment of a feature, as shown in Figure 9. Any glyphs for a feature associated with the inline strand SHOULD be placed in the prototypical orientation given by the specification, while any glyph that is associated with the reverse complement strand SHOULD be inverted vertically and horizontally (i.e., rotated 180 degrees). Reverse complement MAY also be indicated by horizontal-only inversion. Finally, a glyph inverted only vertically still indicates inline strand, but it is RECOMMENDED NOT to use this orientation. Orientation SHOULD be used consistently throughout a diagram, rather than mixing conventions. Examples are provided in Figure 10.

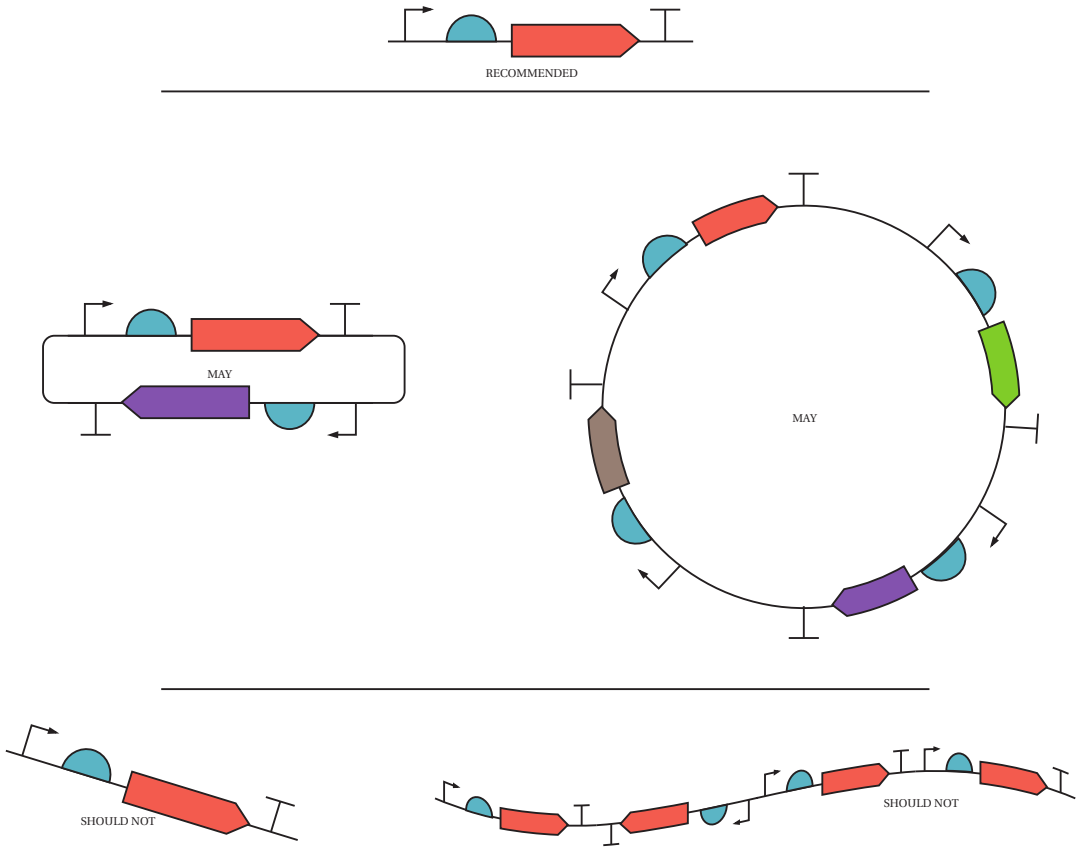


Figure 7: *Recommended, acceptable, and problematic examples of nucleic backbone orientation.*

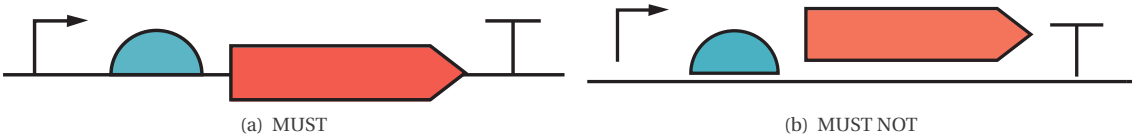


Figure 8: *Examples of correct and incorrect association of glyphs with a nucleic acid backbone.*

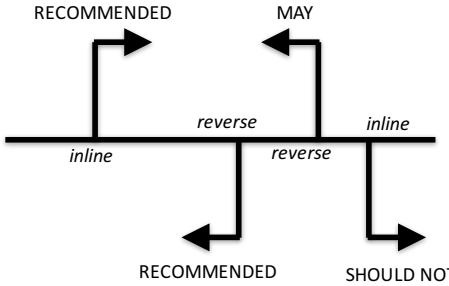


Figure 9: *Use of glyph orientation to indicate inline vs. reverse complement direction.*

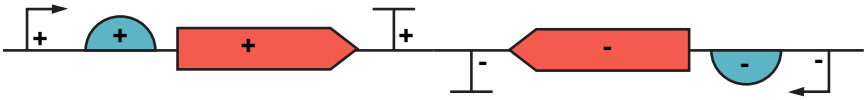


Figure 10: *Example construct incorporating both inline (+) and reverse complement (-) features.*

3. Nucleic acid features in a sequential relationship SHOULD be drawn from 5' left to 3' right on the inline strand and from 5' right to 3' left on the reverse complement strand. In terms of the SBOL 2 data model, this indicates a `SequenceConstraint` on the relative ordering of two features.
4. Nucleic acid features that do not overlap in their locations SHOULD NOT have glyphs whose bounding boxes overlap. An example is provided in Figure 11.

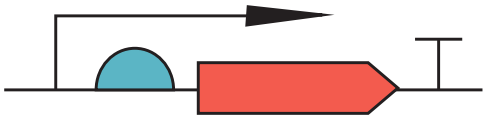


Figure 11: *Example of incorrect glyph overlap: promoter (arrow) does not overlap in sequence with the ribosome entry site and CDS, so SHOULD NOT overlap visually with them.*

5. Nucleic acid features that overlap in their locations SHOULD have glyphs whose bounding boxes overlap. Overlap size MAY be used to indicate relative position. Examples are provided in Figure 12.

(a) Restriction site in a CDS (b) 3'-side operator in a promoter (c) 5'-side operator in a promoter

Figure 12: *Examples where glyphs SHOULD overlap, but might not if it is more clear, e.g., with an operator site located within the 5' portion of a promoter.*

6. A nucleic acid feature SHOULD be represented using a glyph defined in Appendix A.1. In this case, the feature MUST be contained within at least one of the glyph's associated terms. In terms of the SBOL 2 data model, this means the glyph is equal to or a parent of at least one of the `roles` for the `Component` or its associated `ComponentDefinition`. Moreover, the glyph used SHOULD be the RECOMMENDED variant of the most specific applicable glyph. Note that novel glyphs not defined in Appendix A.1 MAY be used, but SHOULD be proposed for adoption as described in Section 5.3. Examples are provided in Figure 13.

6.3 Molecular Species

A glyph that is not in contact with any backbone represents any class of molecule whose detailed structure is not being shown using sequence feature glyphs. In other words, either not a nucleic acid (e.g., proteins, small molecules) or else an "uninteresting" nucleic acid (e.g., showing a transcribed mRNA, but not the features of its sequence). In terms of the SBOL 2 data model, this is a `FunctionalComponent` that is contained within a `ModuleDefinition` implicit in the diagram.

1. A molecular species glyph MUST NOT contact any nucleic acid backbone with any part of its bounding box.

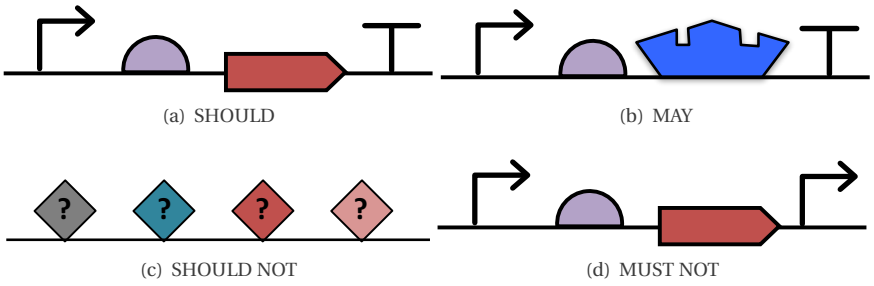


Figure 13: *Examples of recommended, allowed, and forbidden representation of a* `ComponentDefinition` *comprising a sequence of promoter, ribosome entry site, CDS, and terminator: (a) is RECOMMENDED because it uses the preferred variant of the most specific defined glyphs, (b) is allowed because it uses some novel custom non-conflicting symbol, not matching any glyph defined in this document, to encode more specific information about the particular CDS, (c) is recommended against because it uses less specific glyphs, and (d) is forbidden because it use a promoter symbol to represent the terminator.*

2. A molecular species SHOULD be represented using a glyph defined in Appendix A.2. In this case, the species MUST be contained within at least one of the glyph's associated terms. In terms of the SBOL 2 data model, this means the glyph is equal to or a parent of at least one of the `types` for the associated `ComponentDefinition`. Moreover, the glyph used SHOULD be the RECOMMENDED variant of the most specific applicable glyph. Note that novel glyphs not defined in Appendix A.2 MAY be used, but SHOULD be proposed for adoption as described in Section 5.3.

6.4 Interaction

A directed edge "arrow" attached to one or more glyphs indicates a functional interaction involving those elements. The roles of the elements is indicated by their position at the head or tail of the edge. In terms of the SBOL 2 data model, this is an `Interaction`, with either one or two `Participation` relationships, their `role` set by position at the head or tail of the edge. An example is provided in Figure 14.

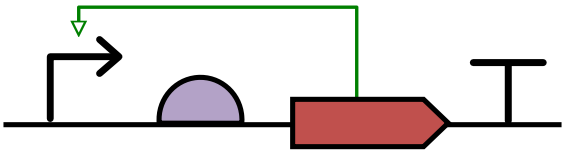


Figure 14: *Example of an interaction indicating a promoter stimulated by the CDS that it regulates.*

1. Two interaction edges SHOULD NOT cross one another. When edges cross, they MUST indicate the distinction between arrows with a crossover pattern, in which one edge "diverts" at the intersection (see Figure 15). Examples are provided in Figure 16.

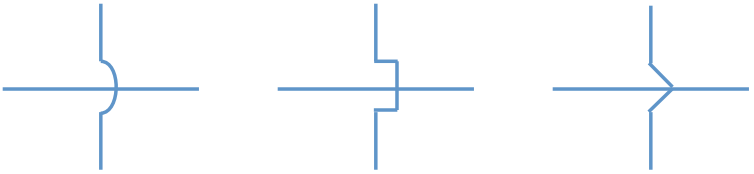
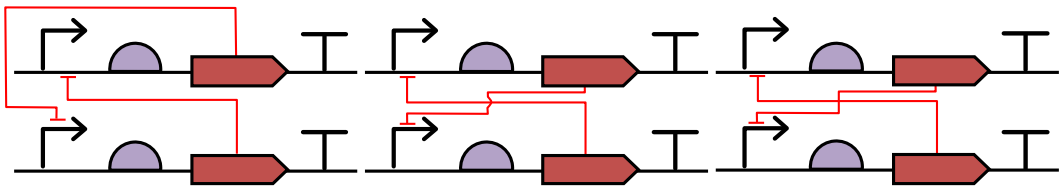


Figure 15: *Examples of* `Interaction` *crossover patterns.*



(a) SHOULD (b) MAY (c) MUST NOT

Figure 16: *Examples of recommended, allowed, and forbidden relationships between two interactions in a mutual repression system: (a) non-crossing is recommended, (b) using a crossover pattern is allowed, but (c) crossing without a crossover pattern is forbidden, since the relationship between the two edges is ambiguous.*

2. An interaction SHOULD be represented using a glyph defined in Appendix A.3. In this case, the interaction type MUST be contained within at least one of the glyph's associated terms. In terms of the SBOL 2 data model, this means the glyph is equal to or a parent of at least one of the `types` for the `Interaction`, and that each associated `Participation` object has a `role` compatible with its position on the head or tail of the edge. Moreover, the glyph used SHOULD be the RECOMMENDED variant of the most specific applicable glyph. Note that novel glyphs not defined in Appendix A.3 MAY be used, but SHOULD be proposed for adoption as described in Section 5.3.

6.5 Labels

The name of any object in a diagram is RECOMMENDED to be displayed as text within, adjacent to, or otherwise clearly visually connected to the object's associated glyph. In terms of the SBOL 2 data model, this is the `name` property, and if no `name` is supplied then the `displayId` MAY be used instead. Examples are provided in Figure 17.

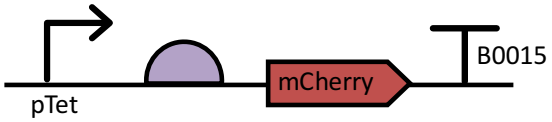


Figure 17: *Examples of labels on glyphs.*

6.6 Annotations

Other text or graphics may be included as annotations with no constraint on their syntax or semantics.

1. Annotations SHOULD NOT be displayed in a way that allows them to be confused with other SBOL Visual elements.

2. Annotations SHOULD NOT be used to display information that can be displayed using other SBOL Visual elements.

6.7 Criteria for Compliance with SBOL Visual

A diagram of a biological system is compliant with SBOL Visual if it complies with all MUST and MUST NOT requirements as specified above. A diagram is compliant with SBOL Visual best practices if it also complies with all RECOMMENDED, SHOULD, and SHOULD NOT statements as specified above.

Importantly, note that a non-SBOL glyph can be used in a compliant diagram when its definition is a subset or superset of a definition that does have an SBOL Visual glyph. For example, a diagram that creates a new glyph for a special type of promoter can be SBOL Visual compliant even though there is an SBOL Visual glyph for a general

promoter.

A piece of software or other system for producing diagrams is compliant with SBOL Visual under the following conditions:

1. The system MUST be capable of producing diagrams that are compliant with SBOL Visual.
2. If the system can also produce diagrams that are *not* compliant with SBOL Visual, it MUST clearly distinguish to the user between compliant and non-compliant usage and diagrams.

A SBOL Visual Glyphs

The following pages present all current glyphs for SBOL Visual, organized by glyph families. Each entry lists:

- Glyph family name
- Associated ontology terms
- Recommended and alternate glyphs
- At least one example of when this glyph would be used
- Any additional notes

A.1 Sequence Feature Glyphs

These glyphs represent features of nucleic acid sequences, and include a bounding box (grey dashed box) and a recommended alignment to the nucleic acid backbone (grey dashed horizontal line).

Aptamer

Associated SO term(s)

SO:0000031: Aptamer

Recommended Glyph and Alternates

The aptamer glyph is a cartoon diagram of a prototypical nucleic acid secondary structure for an aptamer:

Prototypical Example

theophylline aptamer

Notes

this section deliberately blank

Assembly Scar

Associated SO term(s)

SO:0001953

Recommended Glyph and Alternates

The assembly scar glyph is an "equal sign" image, the pattern produced by the union of a 5' sticky end and 3' sticky end glyph. The scar will cover the backbone, creating a visual break suggesting the potential disruption associated with a scar:

With a double-stranded backbone:

Prototypical Example

Ligated sticky ends following BioBrick assembly.

Notes

this section deliberately blank

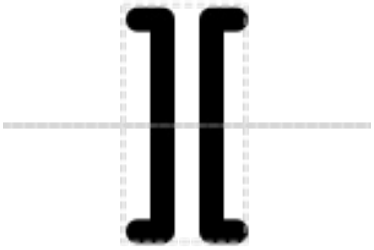
Blunt Restriction Site

Associated SO term(s)

SO:0001691

Recommended Glyph and Alternates

The blunt restriction site glyph is an image of two brackets facing away from one another to make a smooth-edged gap:



Prototypical Example

EcoRV restriction site

Notes

this section deliberately blank

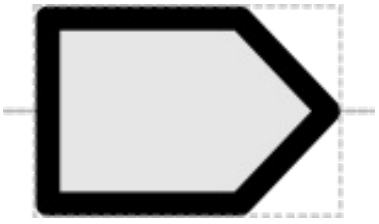
CDS

Associated SO term(s)

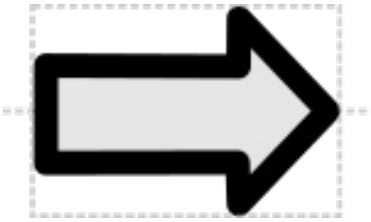
SO:0000316

Recommended Glyph and Alternates

The coding sequence glyph is a "box" with one side bent out arrow-like to show direction:



Alternately, CDS may be represented as a block arrow:



Prototypical Example

α-Hemoglobin coding sequence

Notes

this section deliberately blank

Cleavage Site

Associated SO term(s)

SO:0001956 (Protease Site)

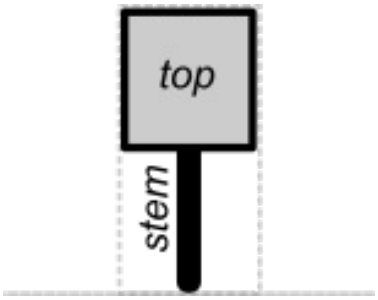
SO:0001977 (Ribonuclease Site)

SO:0001688 (Restriction Enzyme Cleavage Junction), SO:0001687 (Restriction Enzyme Recognition Site)

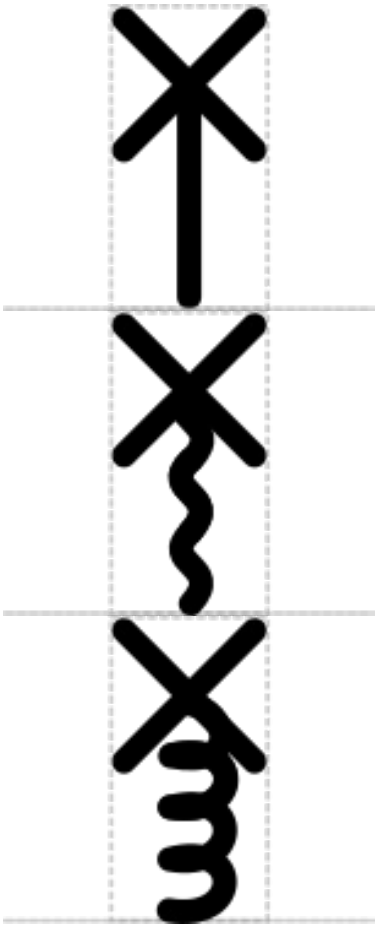
Recommended Glyph and Alternates

Cleavage Site is a "stem-top" glyph for describing small sites. In this system:

- the top glyph indicates the type of site (e.g., Cleavage Site)
- the stem glyph indicates whether the site affects DNA, RNA, or protein (respectively: straight, wavy, or looped)



The Cleavage Site top is an "X" suggesting slicing on top of a stem connecting to the backbone at the point where cleavage will occur (in order: DNA, RNA, Protein):



Prototypical Example

RNAse E site, BamHI

Notes

SO:0000061 (which was previously associated with Restriction Enzyme Recognition Site in SBOL Visual 1.0) is no longer associated with the DNA Cleavage glyph in SBOL Visual 2, as SO:0000061 refers to the binding site and not the location of cleavage.

The Ribonuclease Site, Protease Site, and Restriction Enzyme Recognition Site glyphs from SBOL Visual 1.0 are now replaced by the Cleavage Site glyph with the appropriate stem.

Describing a Restriction Enzyme Cleavage Site with a vertical line glyph on a DNA backbone (as done previously in SBOL Visual 1.0 via the Restriction Enzyme Recognition Site glyph) can persist in a SBOL Visual 2 diagram and still be considered compliant with SBOL Visual 2, where it is now classified as a Biopolymer Location (which is a superclass of cleavage sites). Thus, the Biopolymer Location glyph from SBOL Visual 2.0 is backwards compatible with the Restriction Enzyme Recognition Site glyph from SBOL Visual 1.0.

The 5' Sticky Restriction Site, 3' Sticky Restriction Site, and Blunt Restriction Site glyphs remain unchanged, and are more specific children/derivatives of the DNA-Stem Cleavage-Top glyph.

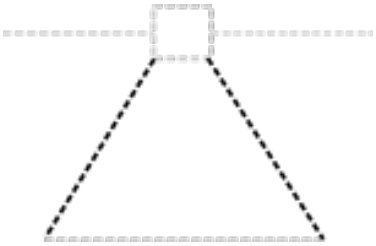
Composite

Associated SO term(s)

Composite does not have an associated SO term, as it merely links a base glyph (with its own SO term) to a sub-diagram (comprising glyphs with their own associated SO terms).

Recommended Glyph and Alternates

The glyph for Composite is dashed "expanding lines" connecting any "base" glyph representing the more abstract composite (e.g., Omitted Detail, or Terminator, or Promoter) to a backbone diagramming the contents of the composite. Note the bounding box is indicating the location of the base glyph, and would scale with that glyph.



Prototypical Example

An "expression cassette" containing a ribosome entry site, coding sequence, and terminator.

In this case, the recommended "base" glyph would be Engineered Region.

Notes

An "abbreviated" representation of composite, simply indicating that more structure is available, can be made by using short lines and placing only an Omitted Detail glyph in the secondary backbone. For example, here is an example of an abbreviated composite promoter:

and a composite with an Engineered Region of otherwise unspecified content:

Engineered Region

Associated SO term(s)

SO:0000804 (Engineered Region)

Recommended Glyph and Alternates

Engineered Region is represented by a plain rectangle suggesting a blank slate to be written upon:

Prototypical Example

An "expression cassette" containing a ribosome entry site, coding sequence, and terminator.

Notes

this section deliberately blank

5' Overhang Site

Associated SO term(s)

SO:0001932: 5' Overhang Site

SO:0001933: 3' Overhang Site

Recommended Glyph and Alternates

The 5' overhang site glyph is an image of a strand of DNA extended on the 5' edge of its forward strand:

With a double-stranded backbone:

Prototypical Example

EcoRI site after cleavage.

Notes

The complementary 3' Overhang Site glyph is a reflection of the 5' Overhang Site.

5' Sticky Restriction Site

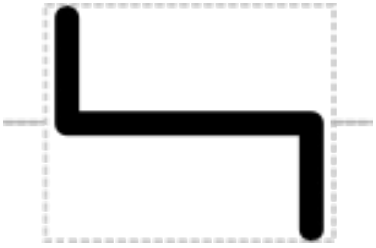
Associated SO term(s)

SO:0001975 (5' Sticky Restriction Site)

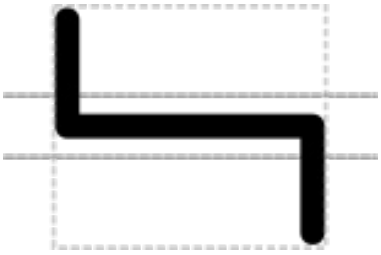
SO:0001976 (3' Sticky Restriction Site)

Recommended Glyph and Alternates

The 5' sticky restriction site glyph is an image of the lines along which two strands of DNA will be cut into 5' sticky ends. Vertical position with respect to the backbone is in a break in a single backbone:



and between strands of a double backbone:



Prototypical Example

EcoRI restriction site.

Notes

The complementary 3' Sticky Restriction Site glyph is a reflection of the 5' Sticky Restriction Site.

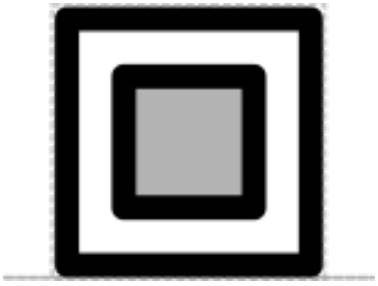
Insulator

Associated SO term(s)

SO:0000627

Recommended Glyph and Alternates

The insulator glyph is a box inside another box that isolates it from its environment:



Prototypical Example

RiboJ

Notes

this section deliberately blank

Biopolymer Location

Associated SO term(s)

SO:0000699 (Junction, Boundary, Breakpoint)

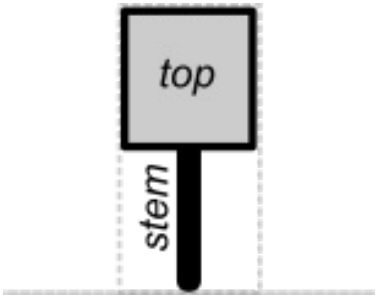
SO:0001236 (Base)

SO:0001237 (Amino Acid)

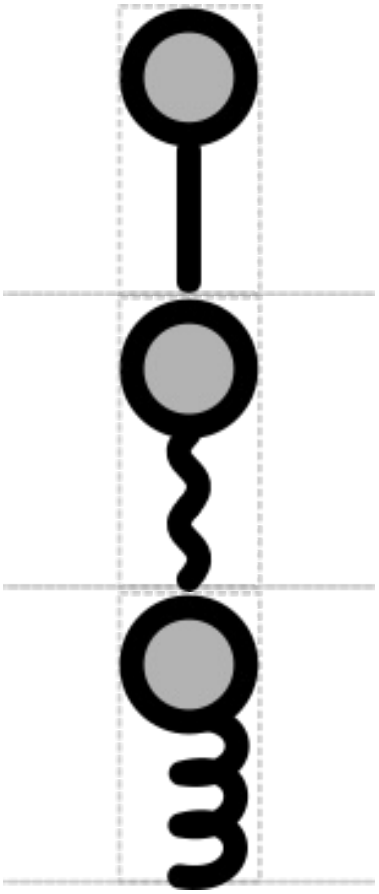
Recommended Glyph and Alternates

Biopolymer Location is a "stem-top" glyph for describing small sites. In this system:

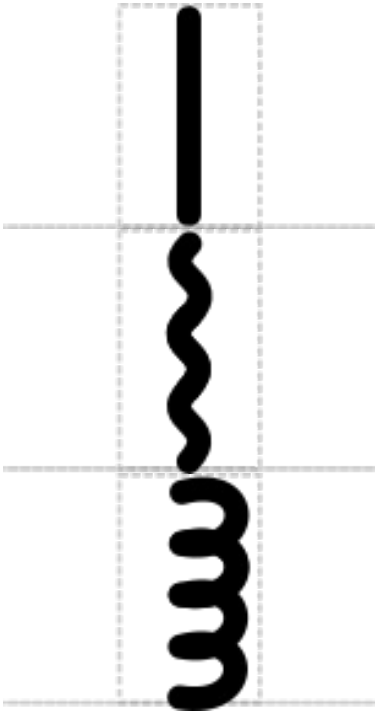
- the top glyph indicates the type of site (e.g., Biopolymer Location)
- the stem glyph indicates whether the site affects DNA, RNA, or protein (respectively: straight, wavy, or looped)



The RECOMMENDED top for Biopolymer Location is a circle, reminiscent of a pin stuck into a location (in order: DNA, RNA, Protein):



An alternative is to have "nothing" for the top, just an extended version of the stem itself (in order: DNA, RNA, Protein):



Prototypical Example

CRISPR-targeted insertion site, protease site, mutation site

Notes

Biopolymer Location is a general glyph for all zero- and one-length sequence features, including insertion and deletion sites and X-ase cut sites.

Note also that Biopolymer Location does not cover stability elements, since their length is typically multiple bases / amino acids.

Describing a Restriction Enzyme Cleavage Site with a vertical line glyph on a DNA backbone (as done previously in SBOL Visual 1.0 via the Restriction Enzyme Recognition Site glyph) can persist in a SBOL Visual 2 diagram and still be considered compliant with SBOL Visual 2, where it is now classified as a Biopolymer Location (which is a superclass of cleavage sites). Thus, the Biopolymer Location glyph from SBOL Visual 2.0 is backwards compatible with the Restriction Enzyme Recognition Site glyph from SBOL Visual 1.0.

No Glyph Assigned

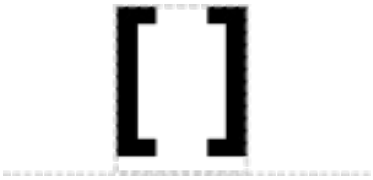
Associated SO term(s)

Any SO term that is not covered by any glyph besides the root Sequence Feature

Recommended Glyph and Alternates

When a part has no assigned glyph it is RECOMMENDED that a user provide their own glyph. The user is also encouraged to submit the new glyph for possible adoption into the SBOLv standard.

An alternative is brackets, suggesting information that needs to be filled in:



As a best practice, it is RECOMMENDED that the name of the term be put in between the brackets.

Prototypical Example

No Glyph Assigned is intended to be used for any Component that is not covered by other SBOL Visual glyphs.

For example, at present there is no glyph recommended for representing a transposon.

Notes

No Glyph Assigned is intended for constructs with a defined specific role that happens to not yet be covered by available approved glyphs (other than the root "Sequence Feature"). It is more likely to appear in machine-generated diagrams than in human-generated diagrams, since humans are likely to invent and use their own glyph for the purpose.

Non-Coding RNA Gene

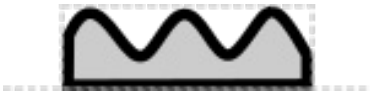
Associated SO term(s)

SO:0001263: Non-Coding RNA Gene

SO:0000834: Mature Transcript Region

Recommended Glyph and Alternates

The non-coding RNA glyph is a rectangular box whose top is a single-stranded RNA "wiggle":



Prototypical Example

gRNA sequence for targeting a dCas9 repressor

Notes

This section left deliverately blank

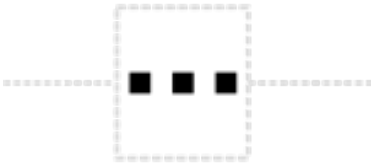
Omitted Detail

Associated SO term(s)

No SO term is associated with Omitted Detail, as it is indicating that something is *not* being represented.

Recommended Glyph and Alternates

The Omitted Detail glyph is a break in the backbone with an ellipsis to indicate that material would normally be in that location:



Prototypical Example

A diagram in which a sequence feature is not drawn.

Notes

This glyph actually places a "break" in the nucleic acid backbone.

Operator / Binding Site

Associated SO term(s)

SO:0000057 Operator

SO:0000409 Binding Site

Recommended Glyph and Alternates

The operator glyph is an open "cup" suggesting a binding location:

Prototypical Example

Gal4 binding site in an activatable promoter.

Notes

This glyph puts a "dent" in the backbone line.

Origin of Replication

Associated SO term(s)

SO:0000296

Recommended Glyph and Alternates

The origin of replication glyph is a circle suggesting the "bulge" opened in a piece of circular DNA when replication is beginning:

Prototypical Example

human herpesvirus-6 OOR

Notes

this section deliberately blank

Origin of Transfer

Associated SO term(s)

SO:0000724: Origin of Transfer

Recommended Glyph and Alternates

The origin of transfer glyph is circular like origin of replication, but also includes an outbound arrow:

Prototypical Example

oriT

Notes

This section left deliberately blank

PolyA Site

Associated SO term(s)

SO:0000553: polyA Site

Recommended Glyph and Alternates

The polyA site glyph is a sequence of As sitting atop the backbone:

Prototypical Example

polyA tail on mammalian coding sequence

Notes

This section left deliberately blank

Primer Binding Site

Associated SO term(s)

SO:0005850

Recommended Glyph and Alternates

The primer binding site glyph is a line with a bent end suggesting a partially complementary strand of nucleic acid attaching to the backbone:

Prototypical Example

seq-F

Notes

this section deliberately blank

Promoter Site

Associated SO term(s)

SO:0000167

Recommended Glyph and Alternates

The promoter glyph is a bent arrow pointing forward, suggesting the action of transcription from its transcription start site:

Prototypical Example

The lacYZA promoter

Notes

this section deliberately blank

Ribosome Entry Site

Associated SO term(s)

SO:0000139: Ribosome Entry Site

Recommended Glyph and Alternates

The ribosome entry promoter glyph is a half-ovoid sitting on the backbone, suggesting an attached ribosome beginning transcription:

Prototypical Example

T7g10 ribosome binding site

Notes

this section deliberately blank

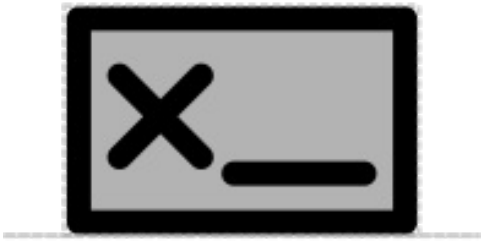
Signature

Associated SO term(s)

SO:0001978

Recommended Glyph and Alternates

The signature glyph is a box sitting atop the backbone with an X and line inside it, suggesting a signature on a form:



Prototypical Example

DNA Barcode

Notes

this section deliberately blank

Specific Recombination Site

Associated SO term(s)

SO:0000299: Specific Recombination Site

Recommended Glyph and Alternates

The specific recombination site glyph is a triangle, centered on the backbone, as has appeared in a number of recombinase circuit papers:

Prototypical Example

flippase recognition target (FRT) site

Notes

This section left deliberately blank

Stability Element

Associated SO term(s)

SO:0001955, SO:0001546 (Protein Stability Element)

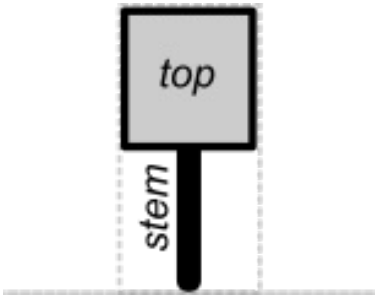
SO:0001979 (RNA Stability Element)

No SO term is currently associated with DNA stability.

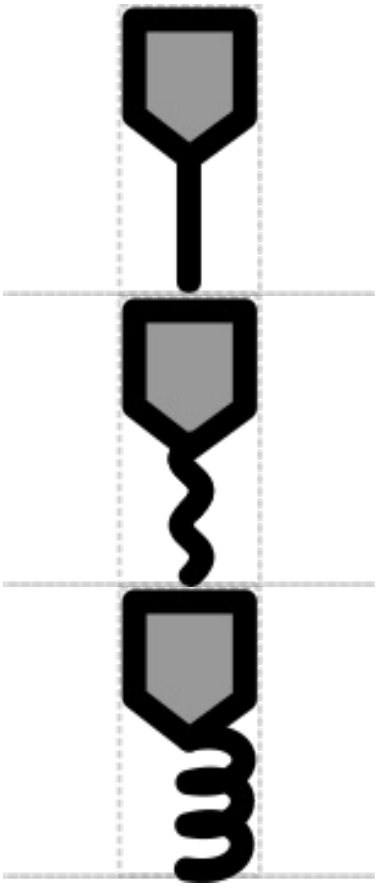
Recommended Glyph and Alternates

Stability Element is a "stem-top" glyph for describing small sites. In this system:

- the top glyph indicates the type of site (e.g., Stability Element)
- the stem glyph indicates whether the site affects DNA, RNA, or protein (respectively: straight, wavy, or looped)



The top for a Stability Element is a pentagon suggesting the shape of a shield, on top of a stem connecting to the backbone at the point where the stability element is located (in order: DNA, RNA, Protein):



Prototypical Example

PEST tag, 3' Hairpin

Notes

RNA Stability Element glyph was previously also associated with SO:0001957, but that SO term has been declared obsolete in Sequence Ontology.

This glyph is not backwards compatible with SBOL Visual 1.0.

Despite both being stem-top glyphs, Biopolymer Location is not a parent to Stability Element, since the length of a Stability Element is typically multiple bases / amino acids.

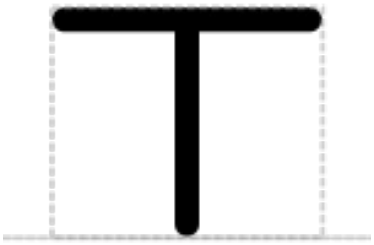
Terminator

Associated SO term(s)

SO:0000141: Terminator

Recommended Glyph and Alternates

The terminator is a T sitting atop the backbone:



Prototypical Example

T1 terminator

Notes

this section deliberately blank

Unspecified

Associated SO term(s)

Unspecified: SO:0000110 Sequence Feature

Recommended Glyph and Alternates

Unspecified is represented by the unicode "replacement character" glyph, indicating a missing or invalid symbol, is RECOMMENDED:

A half-rounded rectangle, the SBGN glyph for a nucleic acid, is an alternative:

Prototypical Example

An anonymous sequence that is missing any information about its nature or intended purpose.

Notes

The Unspecified glyph is intended for showing where a sequence's role is missing (or, equivalently, given only the uninformative "Sequence Feature" root role). It should never appear with well-curated designs or diagrams.

A.2 Molecular Species Glyphs

These glyphs represent molecular species in a diagram, and include a bounding box (grey dashed box) but are not connected to any nucleic acid backbone.

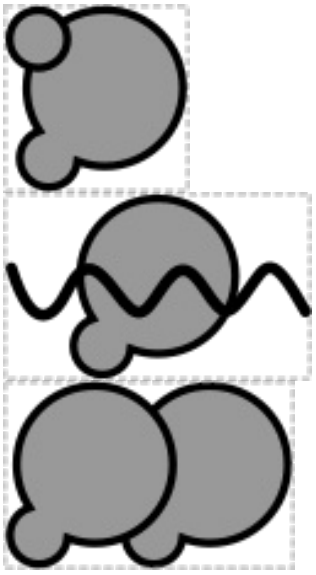
Complex

Associated BioPAX term(s)

Complex: http://www.biopax.org/release/biopax-level3.owl#Complex

Recommended Glyph and Alternates

The RECOMMENDED glyph for a complex is a composite of the glyphs for the molecules of comprising the complex. For example, a protein bound to a small molecule, a guide RNA, or another protein:



An alternative is the SBGN "cornered rectangle" glyph for a complex:

Prototypical Example

Arabinose bound to AraC

Notes

This section left intentionally blank

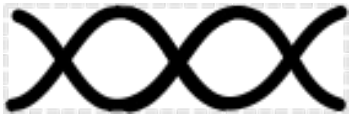
Double-Stranded Nucleic Acid

Associated BioPAX term(s)

Dna: http://www.biopax.org/release/biopax-level3.owl#Dna

Recommended Glyph and Alternates

The RECOMMENDED glyph for dsNA is a double-helix:



An alternative is the SBGN "nucleic acid" half-round rectangle:

Prototypical Example

DNA fragment during assembly

Notes

This section left intentionally blank

Macromolecule

Associated BioPAX term(s)

Protein: http://www.biopax.org/release/biopax-level3.owl#Protein

Recommended Glyph and Alternates

The macromolecule glyph is a diagonally offset union of a large and small circle, intended to invoke the complex shapes of proteins:

An alternative is the SBGN macromolecule glyph, a rounded rectangle:

Prototypical Example

AraC

Notes

It is unclear whether this should be just "Protein" or whether we also want it to be able to repesent multi-component elements like a protein composed of multiple sub-units or a complex polymer.

No Glyph Assigned

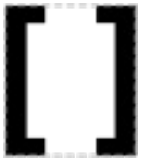
Associated BioPAX term(s)

Any BioPAX type that is not covered by any glyph besides the root

Recommended Glyph and Alternates

When a species has no assigned glyph it is RECOMMENDED that a user provide their own glyph. The user is also encouraged to submit the new glyph for possible adoption into the SBOLv standard.

An alternative option is to have a bracket, suggesting information that needs to be filled in:



Prototypical Example

No Glyph Assigned is intended to be used for any chemical species whose type is not covered by other SBOL Visual glyphs.

Notes

No Glyph Assigned is intended for molecular species with a defined specific type that happens to not yet be covered by available approved glyphs (other than the root). It is more likely to appear in machine-generated diagrams than in human-generated diagrams, since humans are likely to invent and use their own glyph for the purpose.

Small Molecule

Associated BioPAX term(s)

Small Molecule: http://www.biopax.org/release/biopax-level3.owl#SmallMolecule

Recommended Glyph and Alternates

The small molecule glyph is a circle that stretches sideways into a "stadium" to accomodate longer names:

Prototypical Example

Arabinose

Notes

This section left intentionally blank

Single-Stranded Nucleic Acid

Associated BioPAX term(s)

Rna: http://www.biopax.org/release/biopax-level3.owl#Rna

Recommended Glyph and Alternates

The RECOMMENDED glyph for ssNA is a wiggly line:

An alternative is the SBGN "nucleic acid" half-round rectangle:

Prototypical Example

mRNA, gRNA, siRNA

Notes

This section left intentionally blank

Unspecified

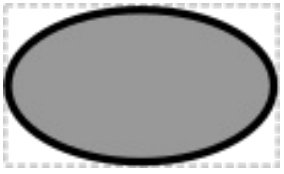
Associated BioPAX term(s)

PhysicalEntity: http://www.biopax.org/release/biopax-level3.owl#PhysicalEntity

Recommended Glyph and Alternates

Unspecified is RECOMMENDED to be represented by the unicode "replacement character" glyph, indicating a missing or invalid symbol:

An alternative is the SBGN "generic species" glyph, which is an ellipse:



Prototypical Example

An anonymous chemical species that is missing any information about its nature or intended purpose.

Notes

The Unspecified glyph is intended for showing where a chemical species' type is missing (or, equivalently, given only the uninformative root role). It should never appear with well-curated designs or diagrams.

A.3 Interaction Glyphs

These glyphs are different forms of “arrow” representing interactions between sequence features and/or molecular species. As arrows, they are extensible and do not have a separately identified bounding box.

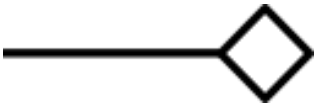
Control

Associated SBO term(s)

SBO:0000168 Control

Recommended Glyph and Alternates

An arrow with a diamond head:



Prototypical Example

Inversion of a sequence flanked by FRT sites by FLP recombinase

Notes

This section left intentionally blank

Degradation

Associated SBO term(s)

SBO:0000179 Degradation

Recommended Glyph and Alternates

Identical to the Process glyph, but with an empty set at the sink of the arrowhead:

Prototypical Example

Cellular recycling of mRNA

Notes

This section left intentionally blank

Inhibition

Associated SBO term(s)

SBO:0000169 Inhibition

Recommended Glyph and Alternates

An arrow whose head is a bar, suggesting blocking:

Prototypical Example

Repression of pTAL14 promoter by TAL14

Notes

This section left intentionally blank

Process

Associated SBO term(s)

SBO:0000375 Process

Recommended Glyph and Alternates

An arrow with a filled head the same color as the line:

Prototypical Example

Production of Green Fluorescent Protein (GFP) from the gfp Coding Sequence

Notes

The assocated SBO term also covers:

- SBO:0000176 Biochemical Reaction
- SBO:0000589 Genetic Production (source is DNAcomponent, sink is usually RNA or Macromolecule)
- SBO:0000177 Non-covalent Binding (sink is a Complex)

Stimulation

Associated SBO term(s)

SBO:0000170 Stimulation

Recommended Glyph and Alternates

An arrow with an head that is empty or of a different color than the line:

Prototypical Example

Activation of pTAL14 promoter by Gal4VP16 activator

Notes

This section left intentionally blank

B Examples

This section contains prototypical examples, including use of all current glyphs to attempt to ensure that their use is clear.

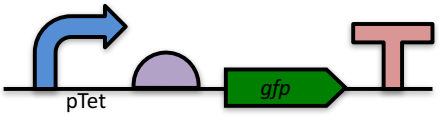


Figure 18: *DNA sequence for a functional unit in which the pTet promoter and an anonymous ribosome entry site regulate expression of a coding sequence for GFP, ended by a terminator.*

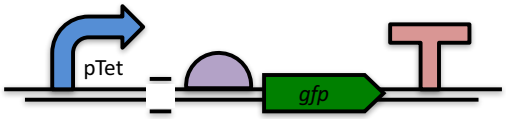


Figure 19: *The same functional unit as in Figure 18, with additional assembly-focused information: there is a 5' overhang before the promoter, a 3' overhand after the terminator, and an assembly scar between the promoter and the ribosome entry site left over from a prior step of assembly.*

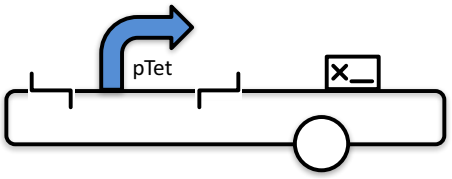


Figure 20: *Promoter pTet stored in a circular plasmid. The promoter is prepared for being cut out of the plasmid: it is preceded by a 5' sticky end restriction site and followed by a 3' stick end restriction site. In addition, the plasmid has been bar-coded with a signature and has its origin of replication marked.*

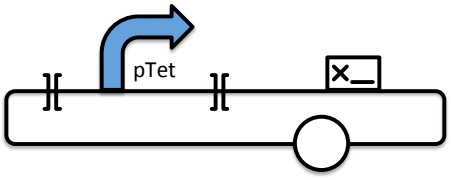


Figure 21: *Promoter stored in a plasmid as in Figure 20, except that the restriction sites before and after the promoter are blunt-end.*

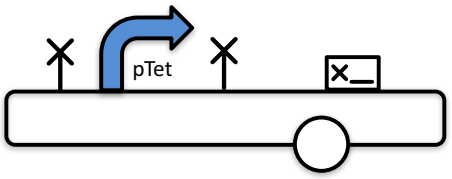


Figure 22: *Promoter stored in a plasmid as in Figure 20, except that the cut structure of the restriction sites before and after the promoter is not specified.*

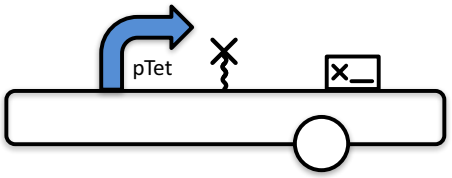


Figure 23: *Promoter stored in a plasmid as in Figure 20, except that there is a ribonuclease site after the promoter rather than restriction sites flanking it.*

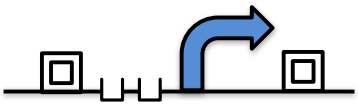


Figure 24: *Detailed design of a promoter, in which the transcription start site is preceded by two operator sites where regulators bind, and the whole is flanked by insulators.*

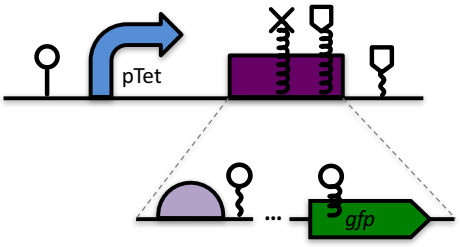


Figure 25: *Promoter regulating the production of an engineered composite sequence that includes RNA and protein stability elements at its 3' end, as well as an internal site for protease cleavage, as well as the expansion of the composite to show it contains a ribosome entry site. coding sequence, and other omitted details. Single residue locations of interest are indicated for the DNA (before the promoter), RNA (after the ribosome entry site), and protein (in the CDS).*

Figure 26: *DNA sequence with three primer binding sites.*

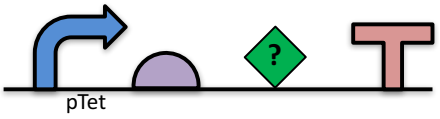


Figure 27: *The same functional unit as in Figure 18, except that information about the CDS is missing, leaving it to fall back on the default unspecified glyph.*

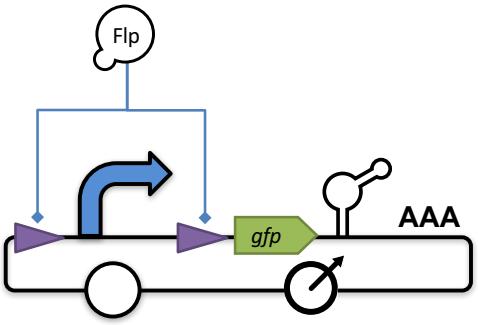


Figure 28: *Promoter regulating the expression of GFP, which is also regulated by an aptamer between it and the poly-A tail of the transcript. The promoter can be cut out by a pair of recombinase target sites, which are acted on by the Flp protein. The whole construct is stored in a circular plasmid with an origin of replication and also an origin of transfer.*

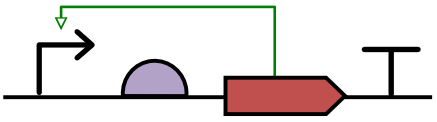


Figure 29: *Promoter stimulated by the CDS that it regulates.*

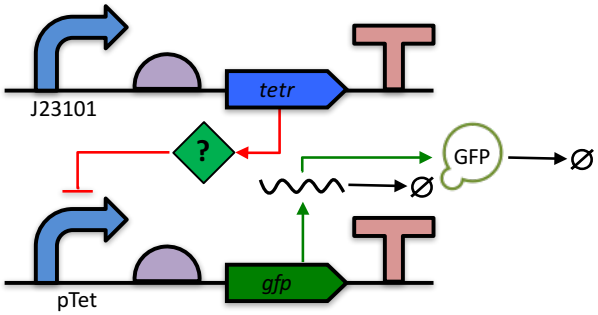


Figure 30: *Constitutive production of TetR, except that information about the protein is missing, leaving it as the default unspecified glyph. TetR represses the pTet promoter, which is regulating production of GFP. The diagram of GFP production explicitly includes the intermediate mRNA and the degradation of both the mRNA and protein products.*

C Relationship to SBOL Visual 1.0

SBOL Visual 2.0 differs from SBOL Visual 1.0 in the following major ways:

- Diagram syntax is expanded to include functional interactions and other molecular species.
- The relationship between diagrams and the SBOL data model is made explicit.
- A number of requirements and best practices are specified for glyphs and diagrams, including:
  - Glyphs include information on interior, bounding box, and recommended backbone alignment.
  - Sequence feature glyphs are required to have their bounding boxes contact the nucleic acid backbone.
  - Nucleic acid diagrams now require the nucleic acid backbone line, and the number of lines allowed in various circumstances is constrained.
  - Explicit statement of when a glyph can and cannot be used to represent a particular element of a diagram.
- Labels that name objects are distinguished from other types of textual annotation.
- Explicit statement of which aspects of a symbol are *not* controlled.
- Symbol variants are now supported.

In addition, the collection of sequence feature glyphs have been expanded and modified in the following ways:

- All non-ambiguous glyphs have been provided with bounding box, interior, and recommended backbone alignment.
- The User Defined glyph has been split into Unspecified, No Glyph Assigned, Engineered Region, and Composite.
- Glyphs have been added for Aptamer, Omitted Detail, Biopolymer Location, Non-Coding RNA Gene, Origin of Transfer, PolyA Site, and Specific Recombination Site.
- The following ontology terms have been assigned or adjusted:
  - Ribonuclease Site has been assigned SO:0001977.
  - 5' Sticky End Restriction Site has been assigned SO:0001975.
  - 3' Sticky End Restriction Site has been assigned SO:0001976.
  - Signature has been assigned SO:0001978.
  - RNA Stability Element has been updated from the obsolete SO:0001957 to the current SO:0001979
  - Restriction Enzyme Recognition Site, in addition to SO:0000139 has a second definition as SO:0000061.
  - 5' Overhang Site and 3' Overhang Site were erroneously listed with their ontology terms exchanged; this has been fixed.

References

Courtot, M., Juty, N., Knüpfer, C., Waltemath, D., Zhukova, A., Dräger, A., Dumontier, M., Finney, A., Golebiewski, M., Hastings, J., et al. (2011). Controlled vocabularies and semantics in systems biology. *Molecular systems biology*, 7(1):543.

Eilbeck, K., Lewis, S. E., Mungall, C. J., Yandell, M., Stein, L., Durbin, R., and Ashburner, M. (2005). The Sequence Ontology: a tool for the unification of genome annotations. *Genome biology*, 6(5):R44.

Goldberg, R. N., Cary, M., and Demir, E. (2010). BioPAX: A community standard for pathway data sharing. *Nature Biotechnology*, 28(9).

Le Novère, N., Hucka, M., Mi, H., Moodie, S., Schreiber, F., Sorokin, A., Demir, E., Wegner, K., Aladjem, M. I., Wimalaratne, S. M., Bergman, F. T., Gauges, R., Ghazal, P., Kawaji, H., Li, L., Matsuoka, Y., Villéger, A., Boyd, S. E., Calzone, L., Courtot, M., Dogrusoz, U., Freeman, T. C., Funahashi, A., Ghosh, S., Jouraku, A., Kim, S., Kolpakov, F., Luna, A., Sahle, S., Schmidt, E., Watterson, S., Wu, G., Goryanin, I., Kell, D. B., Sander, C., Sauro, H., Snoep, J. L., Kohn, K., and Kitano, H. (2009). The systems biology graphical notation. *Nat. Biotechnol.*, 27(8):735–741.