



Beal, J., Nguyen, T., Goroehowski, T., Goni-Moreno, A., Scott-Brown, J., McLaughlin, J., Madsen, C., Aleritsch, B., Bartley, B., Bhakta, S., Bissell, M., Castillo Hair, S., Clancy, K., Luna, A., Le Novere, N., Palchick, Z., Pocock, M., Sauro, H., Sexton, J., ... Wipat, A. (2019). Communicating Structure and Function in Synthetic Biology Diagrams. *ACS Synthetic Biology*, 8, 1818-1825. <https://doi.org/10.1021/acssynbio.9b00139>

Peer reviewed version

Link to published version (if available):  
[10.1021/acssynbio.9b00139](https://doi.org/10.1021/acssynbio.9b00139)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via American Chemical Society at <https://pubs.acs.org/doi/10.1021/acssynbio.9b00139#>. Please refer to any applicable terms of use of the publisher

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

# Communicating Structure and Function in Synthetic Biology Diagrams

Jacob Beal,<sup>\*,†</sup> Tramy Nguyen,<sup>‡</sup> Thomas E. Gorochowski,<sup>¶</sup> Angel Goñi-Moreno,<sup>§</sup>  
James Scott-Brown,<sup>||</sup> James Alastair McLaughlin,<sup>§</sup> Curtis Madsen,<sup>⊥</sup> Benjamin  
Aleritsch,<sup>#</sup> Bryan Bartley,<sup>†</sup> Shyam Bhakta,<sup>@</sup> Mike Bissell,<sup>△</sup> Sebastian Castillo  
Hair,<sup>@</sup> Kevin Clancy,<sup>▽</sup> Augustin Luna,<sup>††</sup> Nicolas Le Novère,<sup>‡‡</sup> Zach Palchick,<sup>¶¶</sup>  
Matthew Pocock,<sup>§§</sup> Herbert Sauro,<sup>|||</sup> John T. Sexton,<sup>@</sup> Jeffrey J. Tabor,<sup>@</sup>  
Christopher A. Voigt,<sup>⊥⊥</sup> Zach Zundel,<sup>‡</sup> Chris Myers,<sup>‡</sup> and Anil Wipat<sup>§</sup>

<sup>†</sup>*Raytheon BBN Technologies, USA*

<sup>‡</sup>*University of Utah, USA*

<sup>¶</sup>*University of Bristol, UK*

<sup>§</sup>*Newcastle University, UK*

<sup>||</sup>*University of Oxford, UK*

<sup>⊥</sup>*Boston University, USA*

<sup>#</sup>*Technical University of Munich, Germany*

<sup>@</sup>*Rice University, USA*

<sup>△</sup>*Amyris, Inc., USA*

<sup>▽</sup>*BioCoder Consulting*

<sup>††</sup>*Harvard Medical School, USA*

<sup>‡‡</sup>*aSciStance Ltd, UK*

<sup>¶¶</sup>*Zymergen, USA*

<sup>§§</sup>*Turing Ate My Hamster, Ltd., UK*

<sup>|||</sup>*University of Washington, USA*

<sup>⊥⊥</sup>*MIT, USA*

E-mail: [jakebeal@ieee.org](mailto:jakebeal@ieee.org)  
ACS Paragon Plus Environment

## Abstract

Biological engineers often find it useful to communicate using diagrams. These diagrams can include information both about the structure of the nucleic acid sequences they are engineering and about the functional relationships between features of these sequences and/or other molecular species. A number of conventions and practices have begun to emerge within synthetic biology for creating such diagrams, and the Synthetic Biology Open Language Visual (SBOL Visual) has been developed as a standard to organize, systematize, and extend such conventions in order to produce a coherent visual language. Here, we describe SBOL Visual version 2, which expands previous diagram standards to include new functional interactions, categories of molecular species, support for families of glyph variants, and the ability to indicate modular structure and mappings between elements of a system. SBOL Visual 2 also clarifies a number of requirements and best practices, significantly expands the collection of glyphs available to describe genetic features, and can be readily applied using a wide variety of software tools, both general and bespoke.

## Keywords

Synthetic Biology, Visualization, Diagrams, Synthetic Biology Open Language

In all fields of engineering, diagrams form a key tool for communication between practitioners. Diagrams are found at every stage of a system's development, from idea and requirements capture to refinement and realization, from analysis and debugging to deployment, operations, and maintenance. As an engineering field matures and operations become more routine, conventions often emerge and become standardized into a common language for displaying diagrams, e.g., circuit diagrams in electrical engineering<sup>1,2</sup> or schematic plans in architecture and mechanical engineering.<sup>3,4</sup> These shared languages for diagrams greatly simplify communication between practitioners and reduce the likelihood of mistakes and

1  
2  
3 misinterpretations. They also allow the development of software tools to graphically edit  
4 designs expressed in diagrammatic form.  
5  
6

7 The engineering of biological organisms is still relatively new as a field, but its practition-  
8 ers have already begun to establish conventions about how to communicate using diagrams.  
9 Accordingly, the Synthetic Biology Open Language (SBOL) community has been developing  
10 a standard diagram language, SBOL Visual, for the communication of synthetic biology de-  
11 signs. Prior versions of SBOL Visual<sup>5-7</sup> have focused on a standardized collection of glyphs  
12 (also known as symbols) for expressing commonly used features of nucleic acid sequences (as  
13 well as recently proposing glyphs for protein features<sup>8</sup>), but left many aspects of diagrams  
14 unspecified and did not define representations for other classes of chemical species (e.g., func-  
15 tional RNAs, small molecules), or functional relations (e.g., genetic production, repression,  
16 chemical reactions). Complementary to this effort, the Systems Biology Graphical Notation  
17 (SBGN)<sup>9</sup> and its many antecedents<sup>10-16</sup> have provided a way to visually express the func-  
18 tional relationships between chemical species, but not their structure or encoding into DNA.  
19 Moreover, SBGN has strict requirements that are at times incompatible with the *de facto*  
20 conventions of usage adopted by many practitioners.  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34

35 SBOL Visual 2 accordingly aims to provide a coherent language for expressing both the  
36 structure and the function of biological designs, by organizing, combining, and systematizing  
37 these prior works, and by incorporating emerging conventions across the field of synthetic  
38 biology<sup>17,18</sup>. This language is designed to be simultaneously simple and easy to use—either by  
39 hand or with a wide variety of software programs—and to allow a high degree of flexibility  
40 and freedom in how practitioners choose to organize, present, and style their diagrams.  
41 Finally, the standard also supports the use of custom and novel diagram elements, as well  
42 as providing a means for the adoption of useful new elements from such diagrams into the  
43 standard.  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Results

SBOL Visual 2 provides three main advancements over the prior SBOL Visual 1 standard. First, it expands the classes of glyphs to include molecular species and their interactions, as well as allowing alternatives for glyphs and giving precise specifications for what is and is not included in the definition of a glyph. Second, it defines a language for diagrams that can incorporate both structural information (nucleic acid sequence features, molecular species), and functional information (interactions, modular structure, and mappings), as well as labels and custom annotations. Finally, it significantly expands the collection of glyphs available to a designer, including glyphs for indicating genomic context, such as integration into a plasmid or the genome of a cell. We discuss each of these in turn, followed by examples illustrating the ease of representing diverse and complex systems with SBOL Visual 2. Full details of the specification summarized here are available at <http://sbolstandard.org/sbol-visual-specification/>

### Glyph Specifications

SBOL Visual 1 standardized a collection of glyphs for representing nucleic acid sequence features, such as promoters, coding sequences, and terminators, each defined by association to one or more terms from the Sequence Ontology.<sup>19</sup> In addition, SBOL Visual 2 adds two classes of glyphs, enabling diagrams to include other species and interactions between species (both of which are also largely outside of the scope of the Sequence Ontology). Molecular species glyphs represent any class of molecules whose detailed structure is not being shown (e.g., a protein, a non-coding RNA, a small molecule, etc.), and are defined by association with a term from the Systems Biology Ontology<sup>20</sup>. Interaction glyphs, on the other hand, are “arrows” indicating functional relationships between sequence features and/or molecular species (e.g., genetic production, inhibition, degradation, etc.). These too are defined by association with a term from the Systems Biology Ontology.<sup>20</sup>

1  
2  
3 Each of these classes is also associated with a class in the SBOL 2 data model<sup>21</sup>, en-  
4 abling automatic mapping from designs to diagrams: sequence feature glyphs represent an  
5 SBOL `Component` in a sequence specification, molecular species glyphs represent an SBOL  
6 `FunctionalComponent` in a module specification, and interaction glyphs represent an SBOL  
7 `Interaction` in a module with defined `Participation` roles for the elements at the glyph's  
8 head and tail.  
9  
10  
11  
12  
13  
14

15 SBOL Visual 2 also acknowledges that in some cases there are good reasons to allow  
16 more than one way to represent a particular concept with a glyph. For example, although it  
17 is recommended that coding sequences be represented as a pentagonal shape pointing in the  
18 direction of the sequence, there is a large community that prefers to use a block arrow instead,  
19 as illustrated in Figure 1(a). Accordingly, SBOL Visual 2 allows glyphs to form a family  
20 of variants: one of these must be designated as “recommended” in the standard, but the  
21 alternatives may still be used whenever there is good reason for doing so. Similarly, when a  
22 feature, species, or interaction could be represented by more than one glyph, the specification  
23 recommends always using the most specific applicable glyph, but the less specific alternative  
24 may nonetheless be used instead.  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34

35 While the standard continues to support broad stylistic variations, additional specifica-  
36 tion information has been added for each glyph, as illustrated in Figure 1(b). In addition to  
37 indicating the stroke outline of a glyph, each glyph's specification must now indicate which  
38 (if any) portions of the glyph are its “interior” for purposes of color fill (removing some pre-  
39 vious ambiguities, e.g., with respect to insulator glyphs). Sequence feature and molecular  
40 species glyphs must also include a bounding box indicating areas outside of the stroke that  
41 are still preferred to not overlap (e.g., the area underneath a promoter arrow). Sequence  
42 features also must provide a recommended alignment with the line representing the nucleic  
43 acid backbone, and all glyphs are defined as vector graphics on a standard canvas to enable  
44 determination of recommended relative scale.  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54

55 Counterintuitively, this additional specification information actually increases the free-  
56  
57  
58  
59  
60

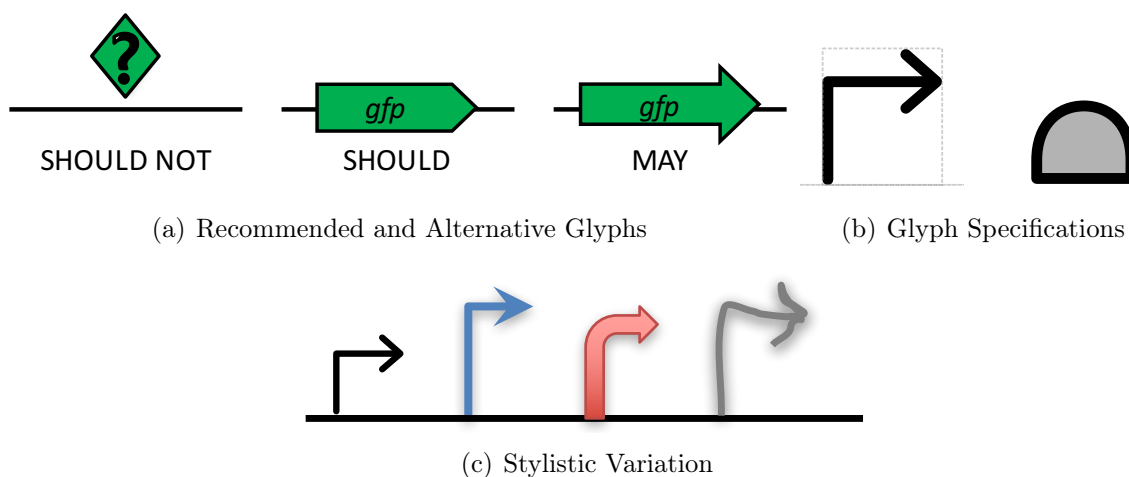


Figure 1: SBOL Visual 2 glyphs (a) support alternative representations, such as this protein coding sequence, which is best represented as part of a nucleic acid sequence by the recommended glyph in the middle, but can also use the less specific Unspecified glyph (left) or alternate arrow glyph (right). These recommendations have differing degrees of importance: less specific glyphs are strongly recommended against ("SHOULD NOT"), whereas alternative glyphs are less of an issue ("MAY"). (b) Glyphs also have specifications that include a preferred relative scale for the glyph outline (solid), fill (grey), bounding box (dashed box), and recommended alignment with the nucleic acid backbone (dashed horizontal line), as in these examples of the specification for a Promoter (left) and the specification for a Ribosome Entry Site (right), and (c) can be freely varied in line and fill style, scale, and minor "font-like" customizations.

1  
2  
3 dom for individual stylistic choices when drawing an SBOL Visual diagram, as illustrated in  
4 Figure 1(c), because anything not explicitly defined in the standard is allowed to be varied—  
5 whereas with SBOL Visual 1 it was unclear whether some aspects of glyphs were supposed to  
6 be fixed or flexible. Furthermore, the specification now specifically defines a set of reserved  
7 visual properties—line and fill color and styling—that are specifically disallowed from being  
8 constrained by glyph specifications. Glyph scaling is also allowed to be modified to encode  
9 additional information, such the length of the corresponding sequence, and minor “font-like”  
10 customizations (such as the addition of shadows and changes to corner styling) are explicitly  
11 endorsed.

12  
13 Finally, the specification provides means for defining and incorporating novel glyphs and  
14 glyph variants. There are a set of further recommended “best practices” for glyph design:  
15 in addition the requirements already described, a glyph should be easy to sketch by hand,  
16 should not risk confusion by looking similar to any other existing glyph (even when re-scaled  
17 or poorly sketched), and should not contain text (to avoid confusion between glyphs and  
18 labels). Sequence features are also recommended to be asymmetric (to indicate direction)  
19 and horizontally scalable (to represent feature size or complexity). Novel glyphs can thus  
20 be created and used freely on an *ad hoc* basis. When one proves useful, however, it is also  
21 recommended that practitioners submit the new glyph for approval, incorporation into the  
22 standard, and dissemination throughout the community of practitioners.

## 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 **Diagram Language**

44  
45 In addition to defining glyphs, SBOL Visual defines a language for combining these glyphs  
46 into diagrams that can be readily and consistently interpreted. An SBOL Visual 2 diagram  
47 centers around representations of nucleic acid constructs and molecular species. A diagram  
48 for a nucleic acid construct is based around a backbone line, its primary structure specified  
49 by the sequence of attached sequence feature glyphs, with strand information optionally  
50 indicated by placing a glyph above or below the backbone. Molecular species, on the other  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



hand, are indicated by glyphs not in contact with any backbone. Interactions involving sequence features and molecular species may then be represented with a network of directed edges and nodes. Finally, any of these objects may have an associated label showing its name, may be grouped together into modules and mappings, and the diagram may further include any form of other annotations, including other uses of text. Figure 2 shows a simple example illustrating all of these types of diagram elements. As with individual glyphs, a diagram as a whole may also be associated with an SBOL 2 data class if desired (a `ComponentDefinition` for each nucleic acid backbone and a `ModuleDefinition` for any diagram including molecular species or interactions). Typically, however, the translation from data model to diagram will omit or compress significant amounts of information.

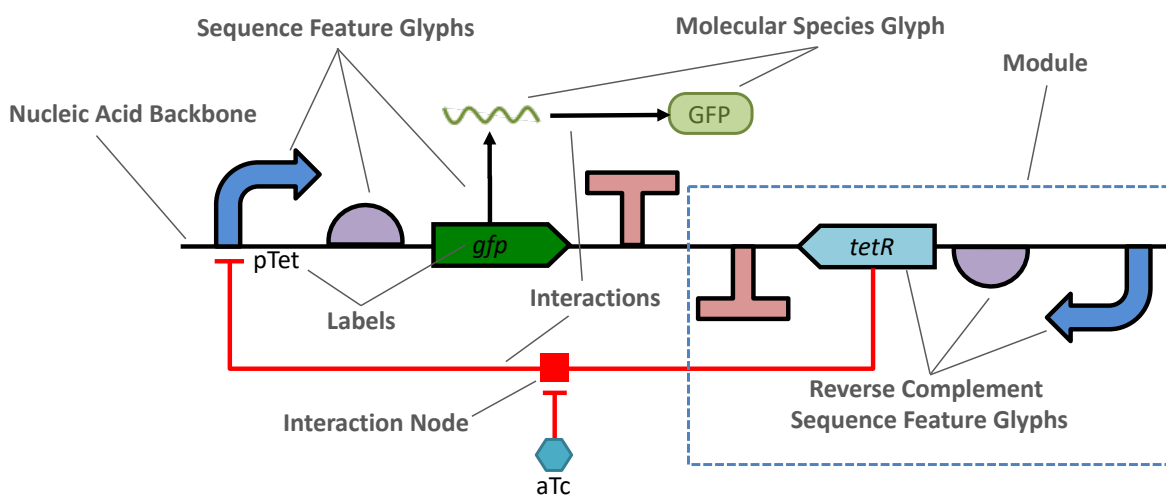


Figure 2: Example illustrating the elements of an SBOL Visual 2 diagram, with nucleic acid sequence features on the forward and reverse strand of a backbone, other molecular species, interactions and interaction nodes indicating the functional relationships between various elements of the system, and a module boundary grouping together one set of elements; the grey labels and indicator lines are annotations.

Nucleic acid backbones can be drawn using either a single or double line, with double lines being an optional means of explicitly indicating double-stranded regions of the backbone. They are typically horizontal in orientation, but can use other orientations and topologies to indicate information such as circularity or DNA nanotech structures. Capturing some common use cases, certain stylized backbone shapes have further been defined into glyphs

1  
2  
3 to indicate the context of a construct, such as integration into a genomic locus or inclusion  
4 in a circular plasmid.  
5

6  
7 Sequence features are indicated with glyphs placed in contact with the nucleic acid back-  
8 bone, following the vertical alignment recommendation for the glyph where possible and  
9 optionally using direction and/or inversion to represent strand information. The ordering of  
10 features on the positive (inline) strand goes from 5' left to 3' right, and the opposite for the  
11 negative (reverse complement) strand. Overlap between glyphs (or their bounding boxes)  
12 indicates overlap in locations, and the horizontal scaling of glyphs can be used to indicate  
13 relative size of features.  
14  
15  
16  
17  
18  
19  
20

21 Away from backbones, molecular species glyphs represent anything whose structure is not  
22 being described in terms of sequence features. This encompasses small molecules, proteins  
23 and other macromolecules, as well as nucleic acids that are of interest but whose structure  
24 is “uninteresting” for the diagram (e.g., a transcribed mRNA that is just being shown as an  
25 intermediate product). The interactions between molecular species and/or sequence features  
26 are shown as directed edges (e.g., arrows), with the meaning of the arrow determined by  
27 its head. For diagrammatic clarity, interaction edges should not cross, but when there is  
28 no good alternative and edges must cross, they are required to use “crossover” patterns  
29 (like in electronic wiring diagrams) to clearly disambiguate which edge is which. Otherwise,  
30 crossing edges might be mistaken for arrows with multiple heads and/or tails, which are used  
31 to represent superposition (e.g., production of a protein from two different coding sequences,  
32 or a repressor acting on two different promoters. Biochemical processes, such as association,  
33 dissociation, or catalysis, are indicated by edges that come together at a node glyph, whose  
34 type indicates the type of process. Critically, these requirements enable clear distinctions  
35 between superpositions and biochemical processes, which are currently often ambiguous in  
36 system diagrams.  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52

53 So far, all of these diagram elements directly represent the biochemical objects and pro-  
54 cesses of a system. Notions of engineering intention and abstraction can be represented as  
55  
56  
57  
58  
59  
60

1  
2  
3 well, as modules represented by closed visual boundaries. Similarly, identity mappings be-  
4 tween elements in different modules can be represented with undirected edges. Modules can  
5 be either “white box” modules that show their contents or “black box” modules that hide  
6 their contents, thereby simplifying a diagram through abstraction, and can indicate intended  
7 interactions with rectangular “ports” on the boundary, similarly to both electronics diagrams  
8 and SBGN diagrams.<sup>9</sup>

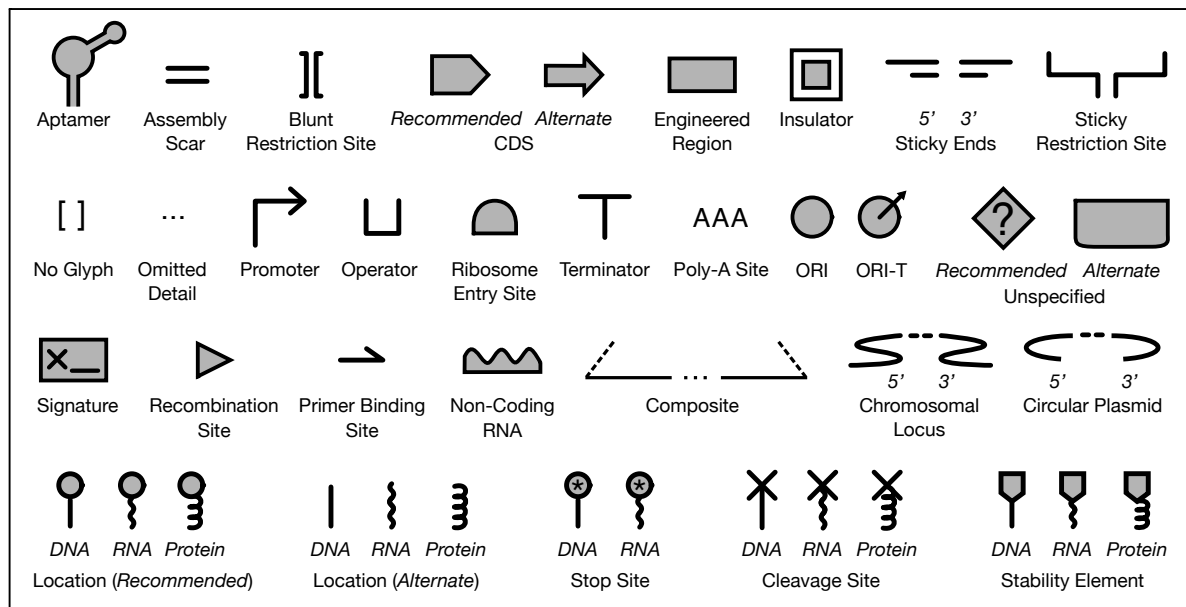
9  
10  
11 Finally, these diagrams may be freely annotated with labels—text giving the names  
12 of objects—and also any other textual or graphical annotations that are visually distinct  
13 from SBOL Visual elements and do not needlessly reinvent their functions. Notice that  
14 these requirements provide minimal constraint on how diagrams are laid out, their contents,  
15 choices of which parts are detailed and which parts simplified, etc. Moreover, most of the  
16 diagram language is either identical or close to how diagrams are already created by many  
17 practitioners. Therefore, it should be quite simple for most practitioners to create diagrams  
18 compliant with SBOL Visual 2.  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30

## 31 32 33 **Expanded Glyph Collection**

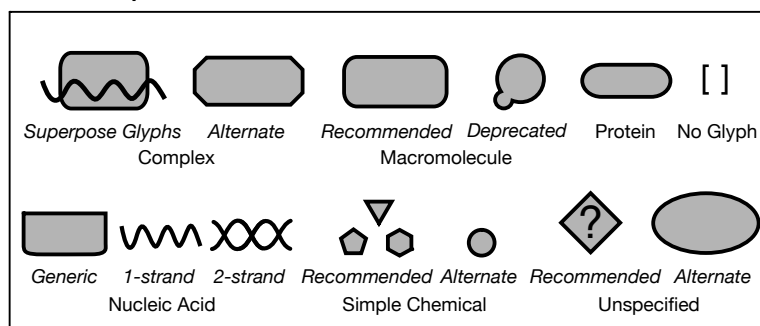
34  
35 SBOL Visual 2 expands on the collection of glyphs provided by SBOL Visual 1 in three main  
36 ways: it extends the collection of glyphs representing nucleic acid sequence features, it adds  
37 a category of glyphs representing *Molecular Species*, and it adds a set of glyphs representing  
38 *Interactions* and *Interaction Nodes*. The full collection of current SBOL Visual 2 glyphs are  
39 shown in Figure 3.  
40  
41  
42  
43  
44

45 With respect to the prior collection of sequence feature glyphs in SBOL Visual 1, one  
46 major change is that the **User Defined** glyph has been replaced by four separate glyphs,  
47 each representing one of the often-confused prior common usages of **User Defined**. The  
48 **Unspecified** glyph typically indicates missing information in the specification of a sequence  
49 and is thus recommended to be represented by a Unicode “replacement character” glyph (but  
50 can be alternatively represented by an SBGN half-rounded rectangle glyph for nucleic acids).  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Genetic Elements



## Molecular Species



## Interactions

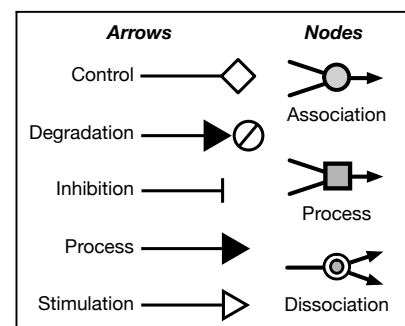


Figure 3: Expanded glyph collection available in SBOL Visual 2. Note that genetic elements are shown here with their glyph and fill only, omitting bounding box and backbone alignment.

No Glyph Assigned, on the other hand, is represented by brackets, suggesting information that needs to be filled in: it is recommended that instead of using this glyph, users provide their own glyph, and submit it for possible adoption into the SBOL Visual standard. The third, Engineered Region, is represented by a plain rectangle that is suggestive of a blank slate to be written upon, and the fourth, Composite is drawn as a pair of dashed “expanding lines” linking any base glyph to an “inset” backbone diagramming the contents of the composite. Complementary to Composite, there is also now an Omitted Detail glyph allows users to explicitly indicate that features are not being shown with an ellipsis in a break in the backbone.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
New glyphs have also been added for a number of types of biological features that were not previously represented: **Aptamer** (a cartoon of the secondary structure of a prototypical aptamer), **Non-Coding RNA Gene** (a rectangular box whose top is a single-stranded RNA ‘wiggle’), **Origin of Transfer** (a circle like the origin of replication, but with an outbound arrow), **PolyA Site** (a sequence of As sitting atop the backbone), and **Specific Recombination Site** (a triangle centered on the backbone). Two other newly introduced glyph pairs stylize the shape of a backbone to indicate that it is part of a **Circular Plasmid** (a C-shaped curve) or integrated into a chromosome at a **Chromosomal Locus** (an S-shaped curve).

21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
A consistent framework of *stem-top glyphs* for indicating sites has been created that extends beyond the **RNA Stability Element**, **Protein Stability Element**, and **Protease Site** provided in SBOL Visual 1. In this system the *top glyph* indicates the type of site, and the vertical *stem* linking this to backbone indicates whether it affects DNA (straight line), RNA (wavy line), or protein (looped line). This system includes a **DNA/RNA/protein Stability Element**, for which the top glyph is a pentagon suggestive of a shield; a **transcription/translation Stop Site**, for which the top glyph is a circle containing an asterisk; a **DNA/RNA/protein Cleavage Site**, for which the top glyph is a cross suggestive of a cut; and a **DNA/RNA/protein Biopolymer Location**, which represents sequence features of length zero or one, and for which the top glyph is a circle suggestive of a pin stuck into a location (or, alternatively, no top at all, as in many plasmid or genomic diagrams).

43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
Finally, a number of other small changes have been made to glyphs that previously existed in SBOL Visual 1. The most notable is that the top edge of the **Operator** glyph has been removed, changing it from a square to an “open cup” in order to make it asymmetrical and better distinguish it from **Engineered Region**. The rest of the changes address prior ambiguities in how glyphs should be positioned on the backbone, where their bounding boxes are, or which portions should be filled when the glyph is colored—all of which can now be explicitly specified.

1  
2  
3 Shifting to the new glyph classes, a new set of *Molecular Species* glyphs have been added  
4 to represent molecular species in a diagram. When drawn, they must not be connected  
5 to a nucleic acid backbone (or else they might be mistaken for sequence features). For  
6 each of these glyphs, the glyphs from SBGN are either used or included as alternatives to  
7 ensure compatibility with that existing standard, but the recommended glyphs have been  
8 chosen to better follow common diagrammatic practices and to be more visually distinctive.  
9 Double-Stranded Nucleic Acid is represented by a double-helix, and Single-Stranded  
10 Nucleic Acid by a single helix; alternatively, either can be represented by the half-round  
11 rectangle SBGN glyph for nucleic acids. Macromolecules are represented by the rounded  
12 rectangle SBGN macromolecule glyph (or alternately, by a diagonally offset union of a large  
13 and small circle). Proteins can be represented either using the macromolecule glyph of by  
14 the more specific protein glyph (a “pill” or “stadium” shape); there is no specific representaiton  
15 for proteins in SBGN, but this choice of shape is consistent with the visual protein language  
16 in<sup>8</sup>. A Simple Chemical is represented either by a small polygon (e.g., triangle, pentagon,  
17 hexagon), or by a small circle, which is compatible with SBGN. A Complex is represented  
18 by a composite of the glyphs for the molecules comprising the complex (or, alternatively,  
19 the corner-cut rectangle used in SBGN). As for sequence features, No Glyph Assigned is  
20 represented by a pair of brackets, and unspecified by a Unicode “replacement character”  
21 glyph (with the elliptical SBGN “generic species” glyph as an alternative).  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40

41 Several kinds of arrows are defined as *Interaction Glyphs* representing interactions be-  
42 tween sequence features and/or molecular species. As with molecular species, these are all  
43 compatible with existing SBGN conventions: all are defined as either the same term or a  
44 parent term in the Systems Biology Ontology, Their names also differ from SBGN in some  
45 cases, as SBOL Visual in all cases uses the name of the associated SBO term. Specifically,  
46 a diamond arrowhead represents **Control** (a generic interaction, including such activities  
47 as recombinase inversion of a sequence flanked by specific recombination sites, equivalent to  
48 SBGN **Modulation**), an arrowhead filled with the same color as the line represents a **Process**  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 (such as production of a protein from a coding sequence, a superset of SBGN **Production**),  
4 an arrowhead that is empty or filled with a different color to the line represents **Stimulation**  
5 (such as activation of a promoter by an activator protein), a bar (or “T-shaped”) arrowhead  
6 represents **Inhibition** (such as repression of a promoter by a repressor protein), and an  
7 arrowhead pointing at an empty-set symbol represents **Degradation** (such as the recycling  
8 of mRNAs),  
9  
10

11  
12  
13  
14  
15 Finally, *Interaction Node* glyphs represent biochemical processes, and can be drawn at  
16 the ends of *Interaction Glyphs*: **Association** or non-covalent binding is represented by a  
17 circular node; **Dissociation** is represented by a circular node nested inside a second circle;  
18 and a generic **Process** is represented by a square node. As with interaction glyphs, these  
19 three glyphs are based on corresponding glyphs from SBGN.  
20  
21  
22  
23  
24  
25  
26

## 27 **Examples of Usage**

28  
29  
30 This section provides several examples to illustrate how SBOL Visual 2 can be used to  
31 produce clear diagrams describing a broad range of systems, all of which use consistent  
32 symbols despite being drawn in a range of distinct visual styles. All make heavy use of  
33 both structural information about nucleic acid sequences and/or other molecules of interest,  
34 as well as functional information about regulation, production, binding, decay, or other  
35 interactions of interest.  
36  
37  
38  
39  
40  
41  
42

- 43 • **CRISPR/Cas9-based circuits** – Two genetic Boolean logic gates from Gander *et*  
44 *al.*,<sup>22</sup> NOR and AND, are shown in Figure 4. Complex formation between dCas9 and  
45 various gRNAs is indicated using superposition of glyphs. Note that each gRNA variant  
46 is identified both explicitly with a textual label and implicitly with color, including a  
47 color match with its associated binding site.  
48  
49  
50  
51  
52
- 53 • **Large gene regulatory circuit** – A large genetic circuit, taken from Nielsen *et al.*,<sup>23</sup>  
54 is drawn in Figure 5. Genetic details are hidden within *black-box* modules, and only  
55  
56  
57  
58  
59  
60

1  
2  
3 the molecular connectivity and degradations of the circuit is made explicit, showing  
4 an example of how an author can choose which details to communicate.  
5  
6  
7

- 8 • **Merged metabolic and regulatory network** – Biological circuits often make use  
9 of both metabolic and regulatory nodes to process information.<sup>24,25</sup> This is the case of  
10 the TOL network of *Pseudomonas putida*, which is drawn in Figure 6. As in previous  
11 examples, modules are used to abstract away those details that are not considered  
12 fundamental to communicate the function of the device. Note that the small molecules  
13 are not represented by SBOL Visual glyphs, allowing their chemical structures to be  
14 communicated directly. Importantly, catalytic connections are consistent with SBGN  
15 notation. The *meta* pathway is zoomed in (Figure 6b) to show the whereabouts of the  
16 *Pm* promoter, which is a crucial part of a new genetic design<sup>26</sup> shown in Figure 6c.  
17  
18 • **Genetic constructs for metabolic engineering** – Figure 7 shows a large metabolic  
19 system developed by Li *et al.*<sup>27</sup> The diagram demonstrates the ability to show genetic  
20 integration of constructs at nine separate modules into different chromosomal loci.  
21 Note the use of non-SBOL Visual graphics to show the chromosomal position of each  
22 locus, and use of color and shading to indicate the source of each coding sequence and  
23 whether it has been modified (e.g., codon optimization of genes).  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

40 It should be noted that across all diagrams the use of color and line style is used to com-  
41 municate many different types of information, e.g., the integration of non-SBOL materials  
42 in Figure 5 and Figure 7. Note also the major differences in organization and graphical style  
43 applied across the various examples.  
44  
45  
46

47 Finally, we note that four different standard graphics editing tools were used to produce  
48 the example figures: Microsoft PowerPoint for Figure 4, Inkscape for Figure 5, Adobe Il-  
49 lustrator for Figure 6, and OmniGraffle for Figure 7. Each was independently chosen by  
50 an author as their preferred illustration tool, indicating how readily SBOL Visual 2 dia-  
51 grams can be created in a diverse set of tools. SBOL Visual 2 is also supported by more  
52  
53  
54  
55  
56  
57  
58  
59  
60



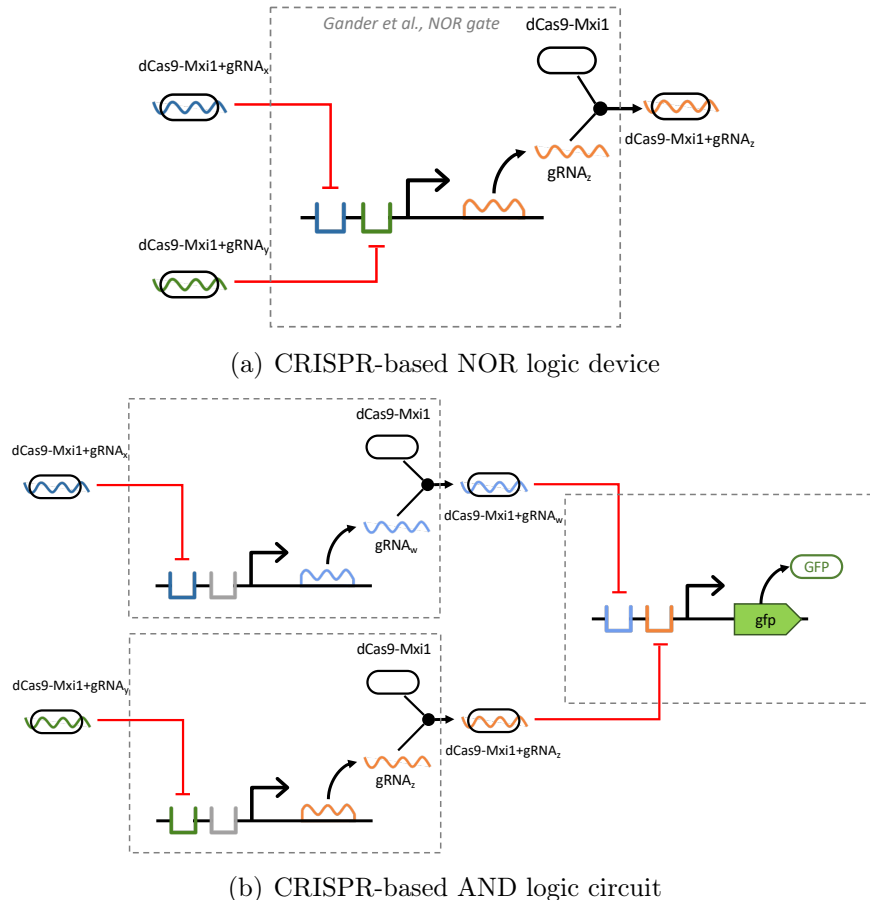


Figure 4: CRISPR/Cas9-based devices and circuits drawn using SBOL Visual 2. (a) NOR device from Gander *et al.*,<sup>22</sup> in which gRNA/dCas9 complexes repress operators upstream of a promoter that regulates the production of gRNA, which in turn binds with dCas9-Mxi1 to complete implementation of a digital NOR logic device. Note the use of color coding to distinguish the  $x/y/z$  gRNAs, and the dashed module boundary identifying device inputs and outputs. (b) Interconnection of three NOR devices to implement an AND circuit from Gander *et al.*<sup>22</sup> Expression of green fluorescence protein (GFP) is used as output and module boundary crossings show how devices are interconnected to form a circuit.

specialized tools for genetic circuit illustration and editing, including updated versions of DNAplotlib<sup>28,29</sup>, VisBOL,<sup>30</sup> and SBOLDesigner.<sup>31</sup>

## Discussion

We have shown how SBOL Visual 2 synthesizes and extends prior means of expressing both structural and functional diagrams of biological designs. The resulting language is succinct,

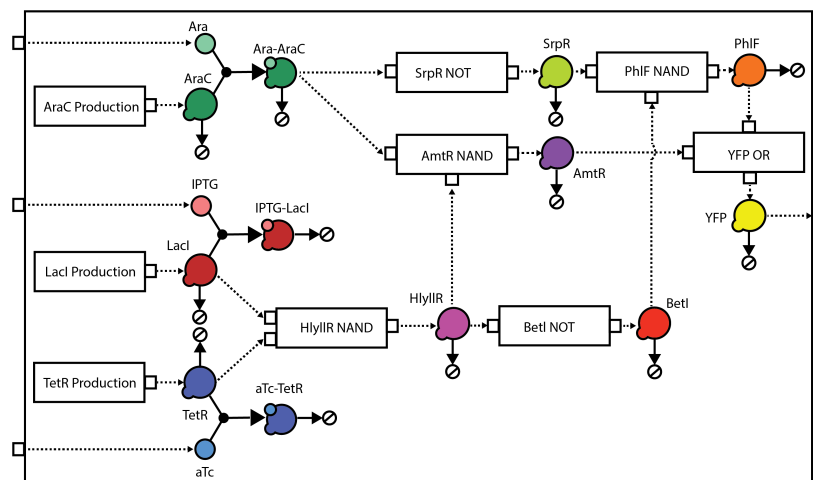


Figure 5: Complex gene regulatory circuit implementing Wolfram's Rule 30 drawn using SBOL Visual 2. Circuit taken from Nielsen *et al.*<sup>23</sup> and displayed using abstract modules and ports to highlight the species interconnecting the devices and their binding and decay relationships.

distinct, and flexible, enabling the construction of highly lucid and readily interpretable diagrams. As SBOL Visual 2 draws on prior diagrammatic conventions wherever possible (including SBGN<sup>9</sup>), widespread adoption of SBOL Visual should be relatively simple. For individual practitioners, there is a clear benefit to adoption, improving the ease with which others can understand and build upon their work. Likewise, software tools will benefit from adoption by making it simpler for users to learn their interface and produce diagrams that can be widely understood. We also argue that journals and funding bodies should strongly consider requiring use of this standard in order to improve the clarity and impact of works that they publish or fund (see Hillson *et al.*<sup>32</sup> for a step in this direction).

Anticipated future directions for evolution of the standard include continued expansion of the glyph collection and standardization of languages for proteins and functional RNA. There are also questions to be addressed regarding how to best represent overlapping features, how to diagram variants and combinatorial libraries, and refinement of the recommendations for diagramming interactions. Ultimately, however, SBOL Visual is an open standard driven by the needs and contributions of the synthetic biology community. The community maintains a public website at [sbolstandard.org](http://sbolstandard.org) and the SBOL Visual project is hosted publicly on

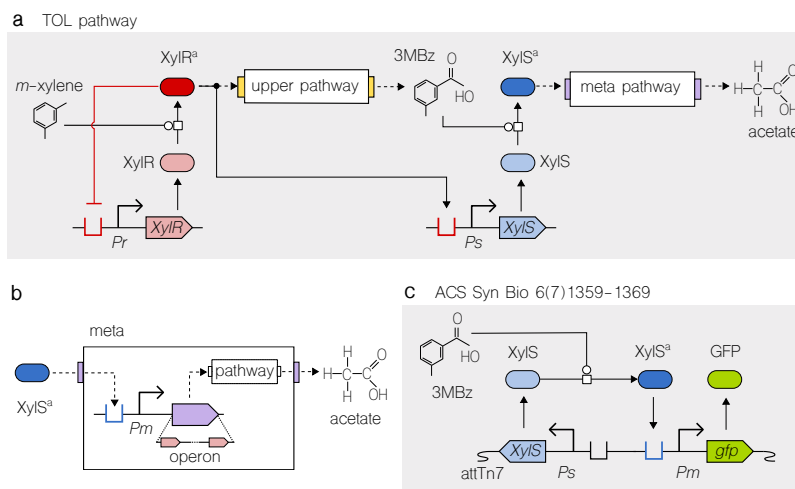


Figure 6: The TOL metabolic and regulatory network drawn using SBOL Visual 2. (a) Structure of the TOL network of *Pseudomonas putida* with focus on the two master regulators, XylR and XylS, along with their cognate expression (*Pr*:XylR and *Ps*:XylS) and activation (*m*-xylene and 3MBz) systems. Modules are used to abstract away specifics of the *upper* and *meta* pathways. (b) Detail of *meta* pathway where the XylS' target, promoter *Pm*, is explicitly drawn. *Pm*'s downstream operon is abstracted using a composite glyph. (c) Regulatory circuit from Goni-Moreno *et al.*<sup>26</sup> inserted into the chromosome (at the *attTn7* site) using components of the TOL network. For all panels, structural formulae of small molecules (i.e., *m*-xylene, 3MBz and acetate) do not follow SBOL Visual glyphs.

GitHub at <https://github.com/SynBioDex/SBOL-visual>. All practitioners are encouraged to participate, whether by expressing needs or by directly involving themselves in development of the standard and its supporting instantiations and tools, in order to help assure the standard continues to develop in ways that will best suit their needs.

## Acknowledgement

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agencies. This document does not contain technology or technical data controlled under either the U.S. International Traffic in Arms Regulations or the U.S. Export Administration Regulations.

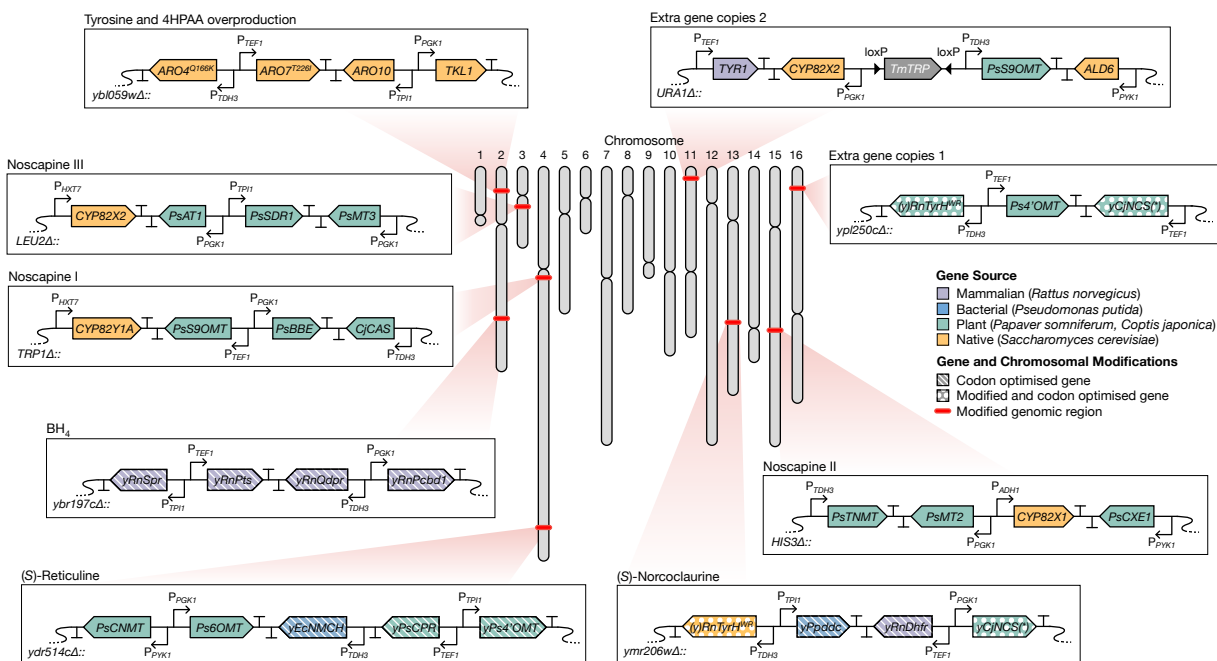


Figure 7: Metabolic engineering of the yeast genome to produce noscapine and halogenated alkaloids from Li *et al.*<sup>27</sup> shown using SBOL Visual 2. Source organism of each gene is denoted by color (purple: mammalian, blue: bacterial, green: plant, and orange: native) and modifications including codon optimization of coding regions shown using patterning (stripes: codon optimized, dots: modified and codon optimized).

## Funding Sources

J.B. is supported in part by the National Science Foundation Expeditions in Computing Program Award #1522074 as part of the Living Computing Project. T.E.G. is supported by a Royal Society University Research Fellowship (grant UF160357) and BrisSynBio, a BBSRC/EPSRC Synthetic Biology Research Centre (grant BB/L01386X/1). G.-B.S. is supported by an EPSRC Fellowship for Growth in Synthetic Biology (grant EP/M002187/1). A.G-M. is supported by the EPSRC grant SynBio3D (EP/R019002/1) and the BioRo-boost contract of the European Union (EU-H2020-BIOTEC-820699). J.S-B acknowledges support from the EPSRC & BBSRC Centre for Doctoral Training in Synthetic Biology (EP/L016494/1) and from DSTL. H.M.S acknowledges support from NIH grant GM123032. T.N., Z.Z., and C.J.M. are supported by the National Science Foundation under Grant No., 1522074, CCF-1748200, DBI-1356041, and DARPA FA8750-17-C-0229. J.A.M. is sup-

ported by FUJIFILM DioSynth Technologies. A.W. is supported by the Engineering and Physical Sciences Research Council (EPSRC) grants EP/N031962/1, EP/R003629/1 and EP/K039083/1. A.L. is supported by National Institute of General Medical Sciences (NIGMS) grant P41 GM103504 and National Human Genome Research Institute (NHGRI) grant U41 HG006623.

## References

- (1) IEEE, IEEE Graphic Symbols for Logic Functions (Includes IEEE Std 91A-1991 Supplement, and IEEE Std 91-1984). IEEE Std. 91a-1991, 1991.
- (2) IEEE, IEEE Standard American National Standard Canadian Standard Graphic Symbols for Electrical and Electronics Diagrams (Including Reference Designation Letters). IEEE Std. 315-1975 (Reaffirmed 1993), 1993.
- (3) Schley, M.; Buday, R.; Sanders, K.; Smith, D. *AIA CAD layer guidelines*; The American Institute of Architects Press: Washington, DC, 1997.
- (4) British Standards Institution, Collaborative production of architectural, engineering and construction information. BS 1192:2007, 2007.
- (5) Quinn, J. Y. et al. SBOL Visual: A Graphical Language for Genetic Designs. *PLoS Biol.* **2015**, *13*, e1002310.
- (6) Quinn, J.; Beal, J.; Bhatia, S.; Cai, P.; Chen, J.; Clancy, K.; Hillson, N.; Galdzicki, M.; Maheshwari, A.; Pocock, M.; Rodriguez, C.; Stan, G.-B.; Endy, D. *Synthetic Biology Open Language Visual (SBOL Visual), version 1.0.0*; 2013.
- (7) Rodriguez, C.; Bartram, S.; Ramasubramanian, A.; Endy, D. *BBF RFC 16: Synthetic biology open language visual (SBOLv) specification*; 2009.

- 1  
2  
3 (8) Cox III, R. S.; McLaughlin, J. A.; Grunberg, R.; Beal, J.; Wipat, A.; Sauro, H. M. A  
4 Visual Language for Protein Design. *ACS Synth. Biol.* **2017**, *6*, 1120–1123.  
5  
6  
7  
8 (9) Le Novere, N. et al. The systems biology graphical notation. *Nat. Biotechnol.* **2009**,  
9 *27*, 735–741.  
10  
11  
12 (10) Pirson, I.; Fortemaison, N.; Jacobs, C.; Dremier, S.; Dumont, J. E.; Maenhaut, C.  
13 The visual display of regulatory information and networks. *Trends Cell Biol.* **2000**, *10*,  
14 404–408.  
15  
16  
17  
18 (11) Cook, D. L.; Farley, J. F.; Tapscott, S. J. A basis for a visual language for describing,  
19 archiving and analyzing functional models of complex biological systems. *Genome Biol.*  
20 **2001**, *2*, research0012–1.  
21  
22  
23  
24  
25 (12) Kitano, H. A graphical notation for biochemical networks. *Biosilico* **2003**, *1*, 169–176.  
26  
27  
28 (13) Kitano, H.; Funahashi, A.; Matsuoka, Y.; Oda, K. Using process diagrams for the  
29 graphical representation of biological networks. *Nat. Biotechnol.* **2005**, *23*, 961.  
30  
31  
32  
33 (14) Kohn, K. W.; Aladjem, M. I.; Weinstein, J. N.; Pommier, Y. Molecular interaction  
34 maps of bioregulatory networks: a general rubric for systems biology. *Mol. Biol. Cell*  
35 **2006**, *17*, 1–13.  
36  
37  
38  
39 (15) Moodie, S. L.; Sorokin, A.; Goryanin, I.; Ghazal, P. A graphical notation to describe  
40 the logical interactions of biological pathways. *Journal of Integrative Bioinformatics*  
41 **2006**, *3*, 177–187.  
42  
43  
44  
45 (16) Longabaugh, W. J.; Davidson, E. H.; Bolouri, H. Visualization, documentation, anal-  
46 ysis, and communication of large-scale gene regulatory networks. *Biochim. Biophys.*  
47 *Acta, Gene Regul. Mech.* **2009**, *1789*, 363–374.  
48  
49  
50  
51 (17) Cox, R. S. et al. Synthetic Biology Open Language Visual (SBOL Visual) Version 2.0.  
52 *Journal of Integrative Bioinformatics* **2018**, *15*.  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 (18) Madsen, C. et al. Synthetic Biology Open Language Visual (SBOL Visual) Version 2.1.  
4 *Journal of Integrative Bioinformatics* **2019**,  
5  
6  
7  
8 (19) Eilbeck, K.; Lewis, S. E.; Mungall, C. J.; Yandell, M.; Stein, L.; Durbin, R.; Ash-  
9 burner, M. The Sequence Ontology: a tool for the unification of genome annotations.  
10 *Genome Biol.* **2005**, *6*, R44.  
11  
12  
13  
14 (20) Courtot, M. et al. Controlled vocabularies and semantics in systems biology. *Mol. Syst.*  
15 *Biol.* **2011**, *7*, 543.  
16  
17  
18  
19 (21) Madsen, C. et al. Synthetic Biology Open Language (SBOL) Version 2.3. *Journal of*  
20 *Integrative Bioinformatics* **2019**,  
21  
22  
23  
24 (22) Gander, M. W.; Vrana, J. D.; Voje, W. E.; Carothers, J. M.; Klavins, E. Digital logic  
25 circuits in yeast with CRISPR-dCas9 NOR gates. *Nat. Commun.* **2017**, *8*, 15459.  
26  
27  
28  
29 (23) Nielsen, A. A.; Der, B. S.; Shin, J.; Vaidyanathan, P.; Paralanov, V.; Strychalski, E. A.;  
30 Ross, D.; Densmore, D.; Voigt, C. A. Genetic circuit design automation. *Science* **2016**,  
31 *352*, aac7341.  
32  
33  
34  
35  
36 (24) Chavarría, M.; Goñi-Moreno, Á.; de Lorenzo, V.; Nikel, P. I. A metabolic widget adjusts  
37 the phosphoenolpyruvate-dependent fructose influx in *Pseudomonas putida*. *mSystems*  
38 **2016**, *1*, e00154–16.  
39  
40  
41  
42  
43 (25) Goni-Moreno, A.; Nikel, P. I. High-Performance Biocomputing in Synthetic Biology–  
44 Integrated Transcriptional and Metabolic circuits. *Frontiers in Bioengineering and*  
45 *Biotechnology* **2019**, *7*, 40.  
46  
47  
48  
49  
50 (26) Goni-Moreno, A.; Benedetti, I.; Kim, J.; de Lorenzo, V. Deconvolution of gene expres-  
51 sion noise into spatial dynamics of transcription factor–promoter interplay. *ACS Synth.*  
52 *Biol.* **2017**, *6*, 1359–1369.  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 (27) Li, Y.; Li, S.; Thodey, K.; Trenchard, I.; Cravens, A.; Smolke, C. D. Complete biosyn-  
4 thesis of noscapine and halogenated alkaloids in yeast. *Proc. Natl. Acad. Sci. U. S. A*  
5 **2018**, *115*, E3922–E3931.  
6  
7  
8  
9  
10 (28) Bartoli, V.; Dixon, D. O. R.; Goroehowski, T. E. In *Synthetic Biology: Methods and*  
11 *Protocols*; Braman, J. C., Ed.; Springer New York: New York, NY, 2018; pp 399–409.  
12  
13  
14 (29) Der, B. S.; Glassey, E.; Bartley, B. A.; Enghuus, C.; Goodman, D. B.; Gordon, D. B.;  
15 Voigt, C. A.; Goroehowski, T. E. DNAPlotlib: Programmable Visualization of Genetic  
16 Designs and Associated Data. *ACS Synth. Biol.* **2017**, *6*, 1115–1119.  
17  
18  
19  
20  
21 (30) McLaughlin, J. A.; Pocock, M.; Mısırlı, G.; Madsen, C.; Wipat, A. VisBOL: web-based  
22 tools for synthetic biology design visualization. *ACS Synth. Biol.* **2016**, *5*, 874–876.  
23  
24  
25  
26 (31) Zhang, M.; McLaughlin, J. A.; Wipat, A.; Myers, C. J. SBOLDesigner 2: an intuitive  
27 tool for structural genetic design. *ACS Synth. Biol.* **2017**, *6*, 1150–1160.  
28  
29  
30  
31 (32) Hillson, N. J.; Plahar, H. A.; Beal, J.; Prithviraj, R. Improving Synthetic Biology  
32 Communication: Recommended Practices for Visual Depiction and Digital Submission  
33 of Genetic Designs. *ACS Synth. Biol.* **2016**, *5*, 449–451.  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



# Graphical TOC Entry

