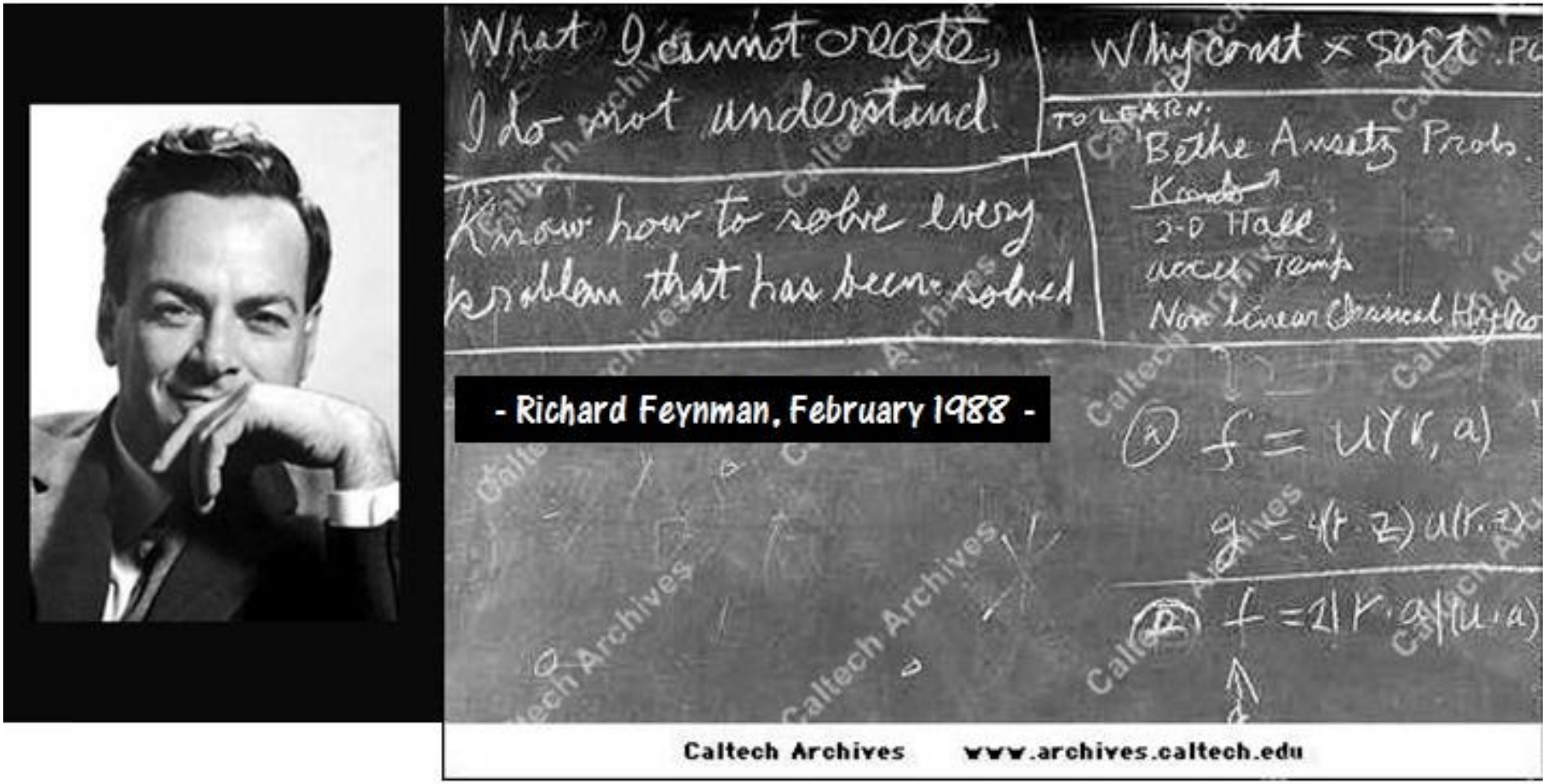
Engineering Genetic Circuits

Chris J. Myers

Lecture 2: Genetic Parts

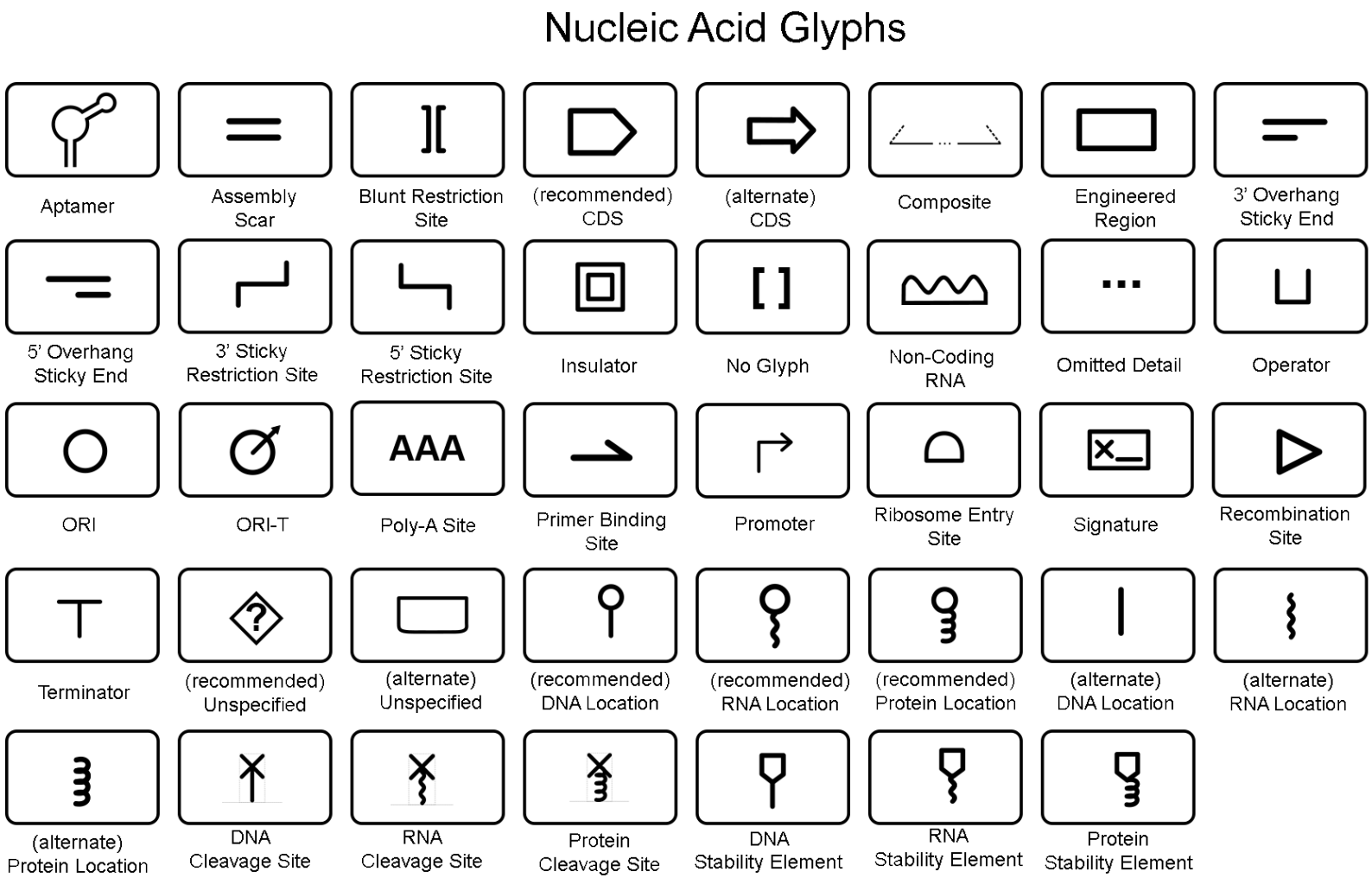
Richard Feynman



Genetic Parts

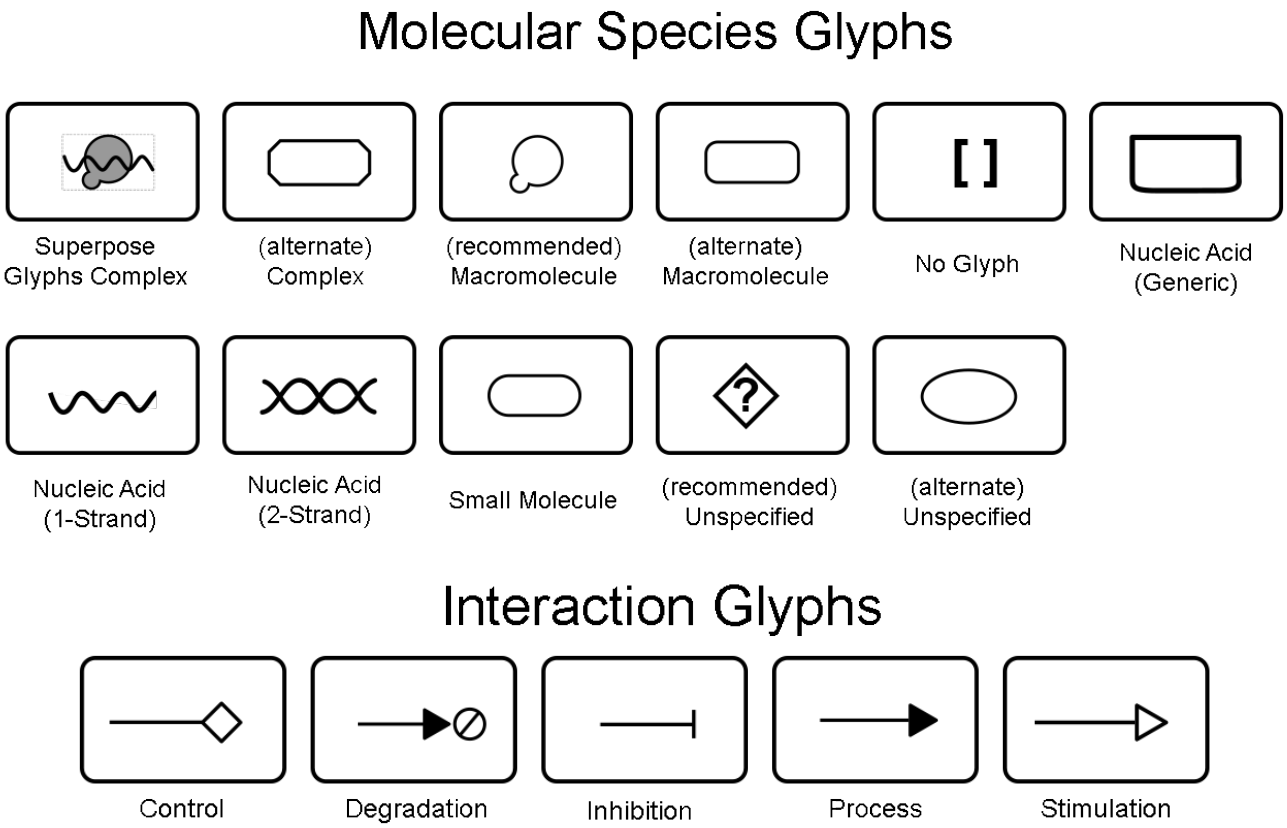
- Pieces of DNA that encode a biological function.
- Many correspond to well known sequence based genetic features.
  - Promoters
  - Ribosome binding sites
  - Protein coding sequences
  - Terminators
- The *Sequence Ontology* defines over 2800 types of sequence features.
  `http://www.sequenceontology.org`
- Parts can be combined to produce devices (see next lecture).

Synthetic Biology Open Language (SBOL) Visual



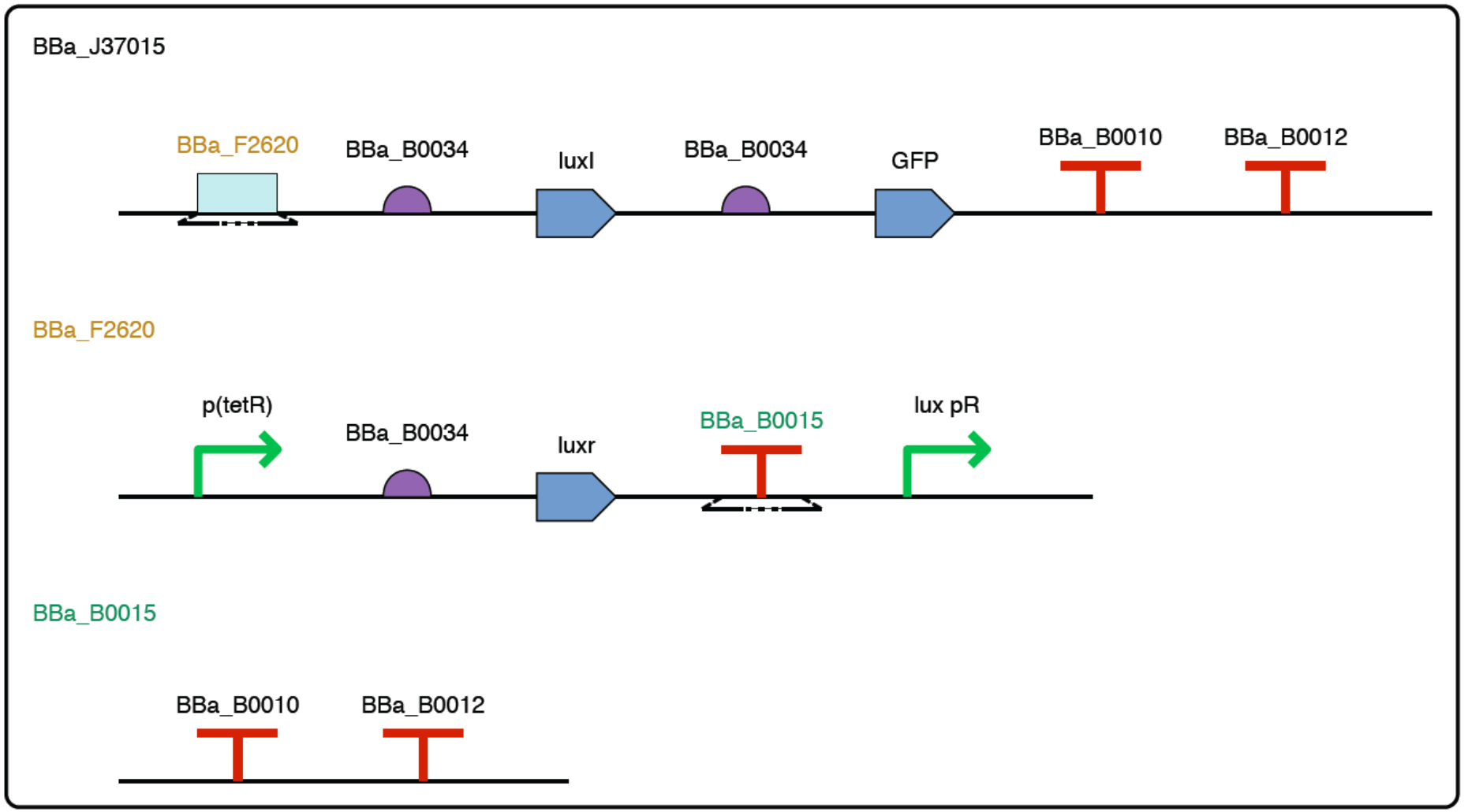
More information can be found at: `http://sbolstandard.org/visual/`

Synthetic Biology Open Language (SBOL) Visual



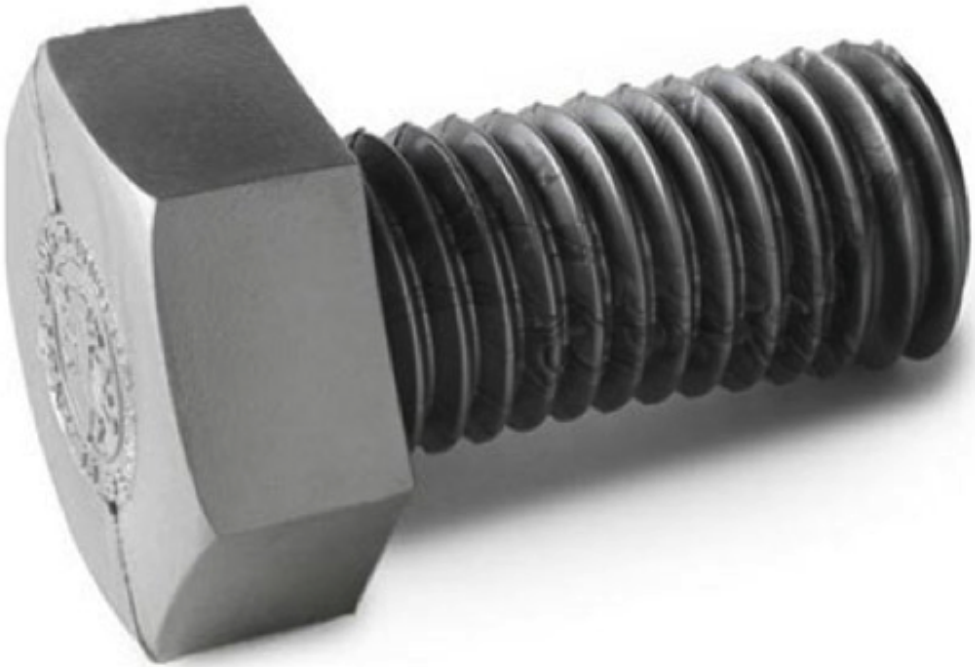
More information can be found at: `http://sbolstandard.org/visual/`

SBOL Visual Example



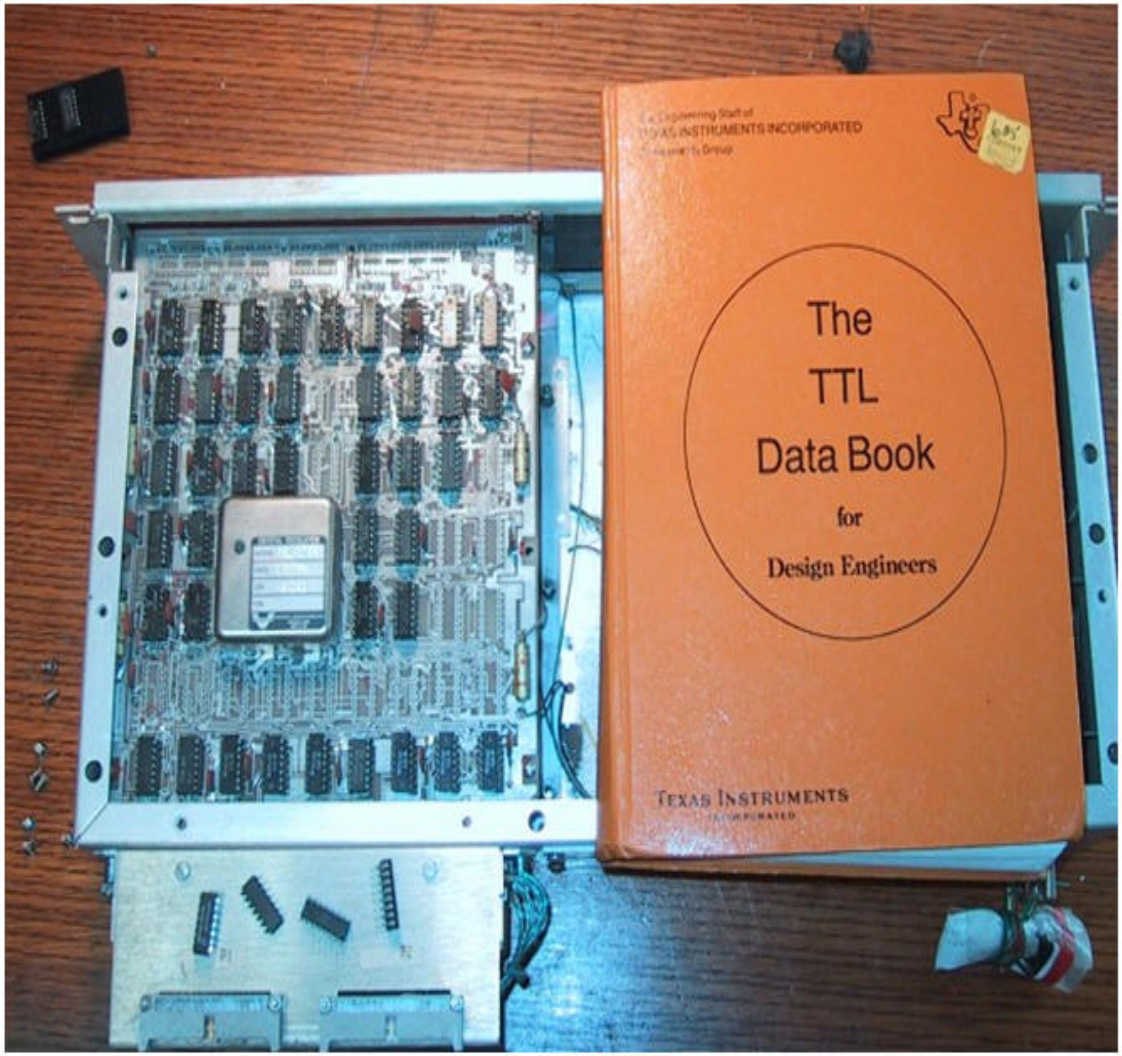
Rendered by VisBol (`http://visbol.org`)

Standard Parts



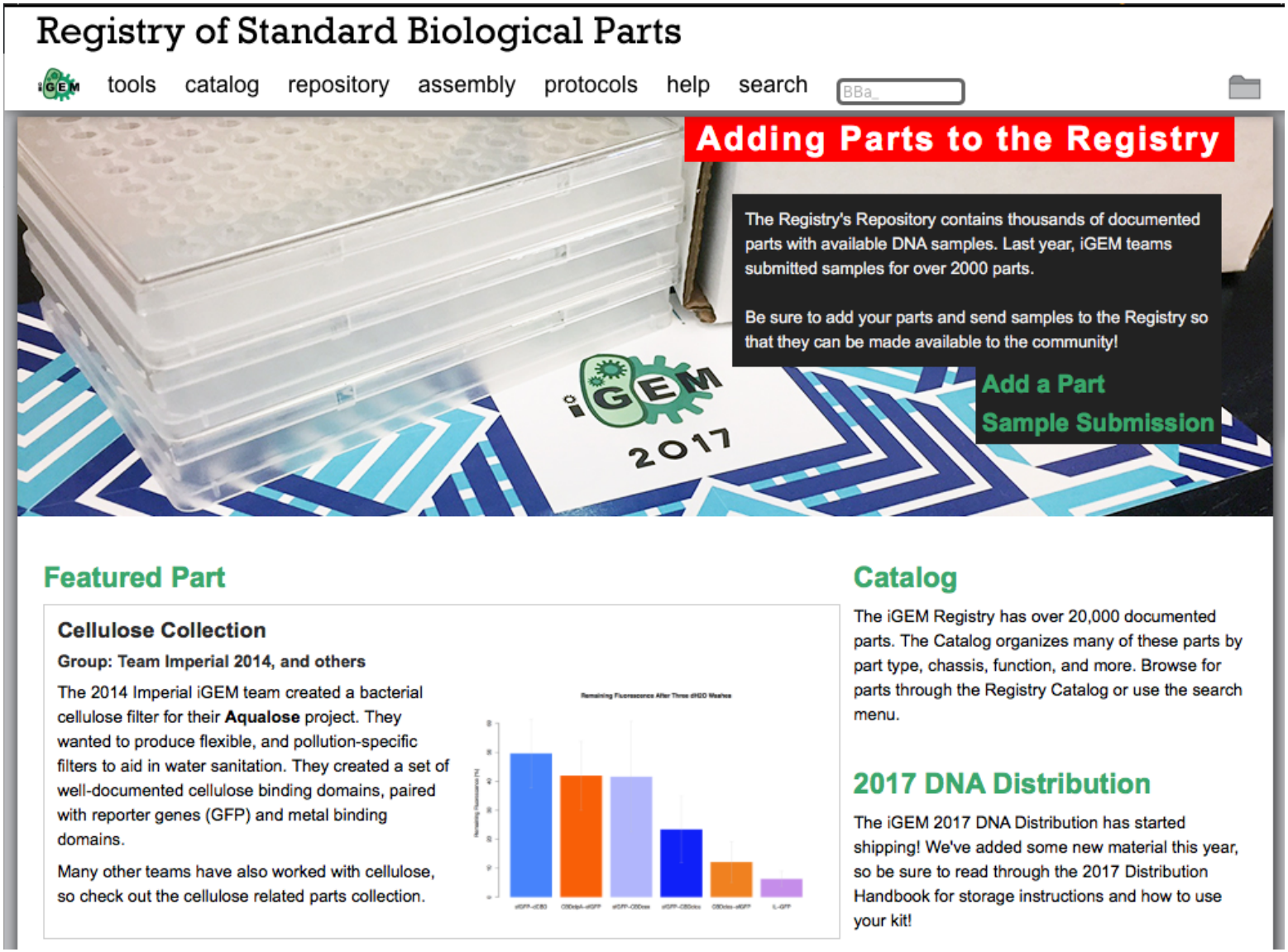
Standard Parts

Standard Parts



http://ee.hawaii.edu/~sasaki/EE260/Labs/Datasheets/7400.pdf

iGEM Registry of Standard Biological Parts (BioBricks)



http://parts.igem.org

SynBioHub Data Repository

Newcastle University

James McLaughlin

Anil Wipat

THE UNIVERSITY OF UTAH®

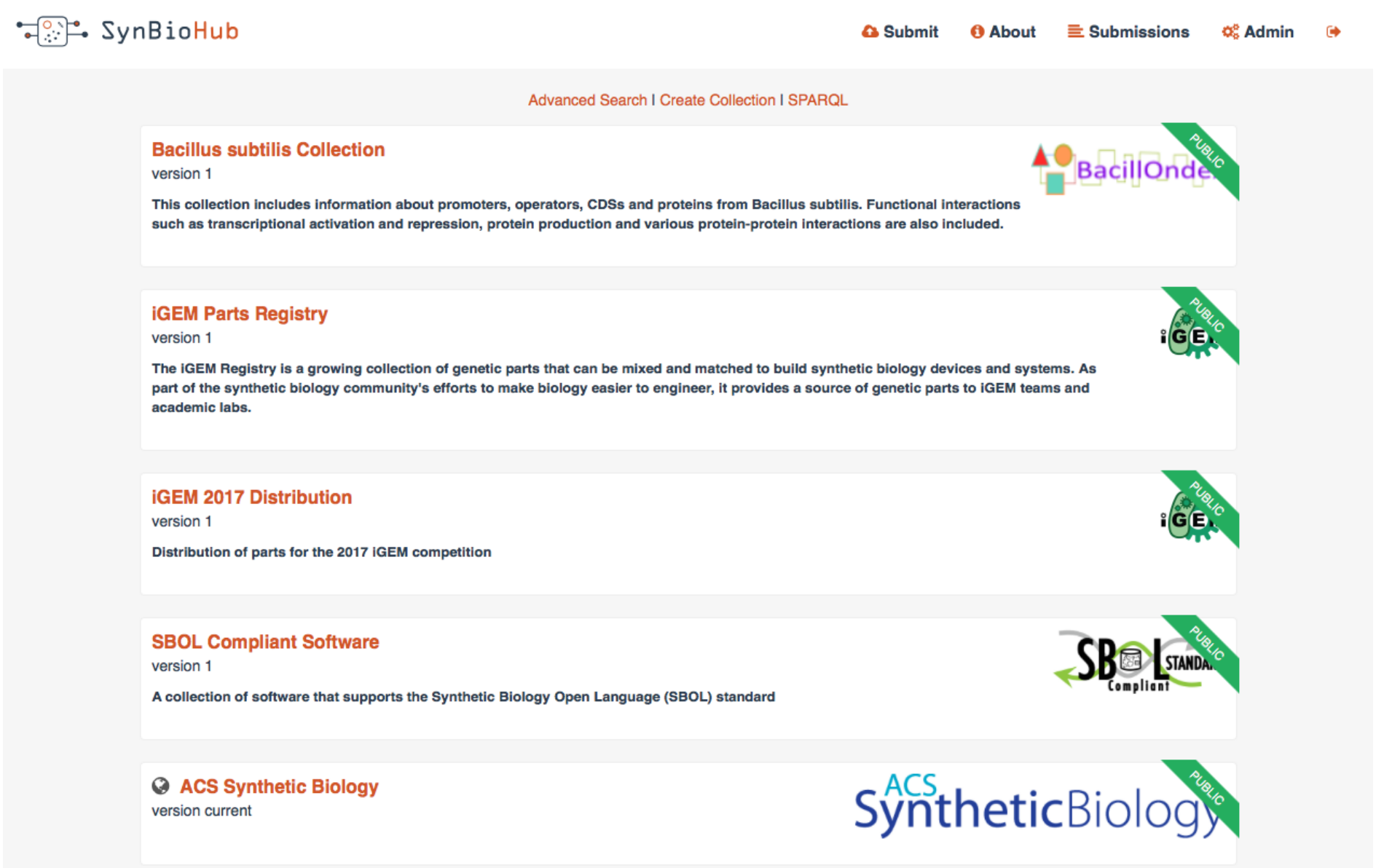
Zach Zundel

Chris Myers

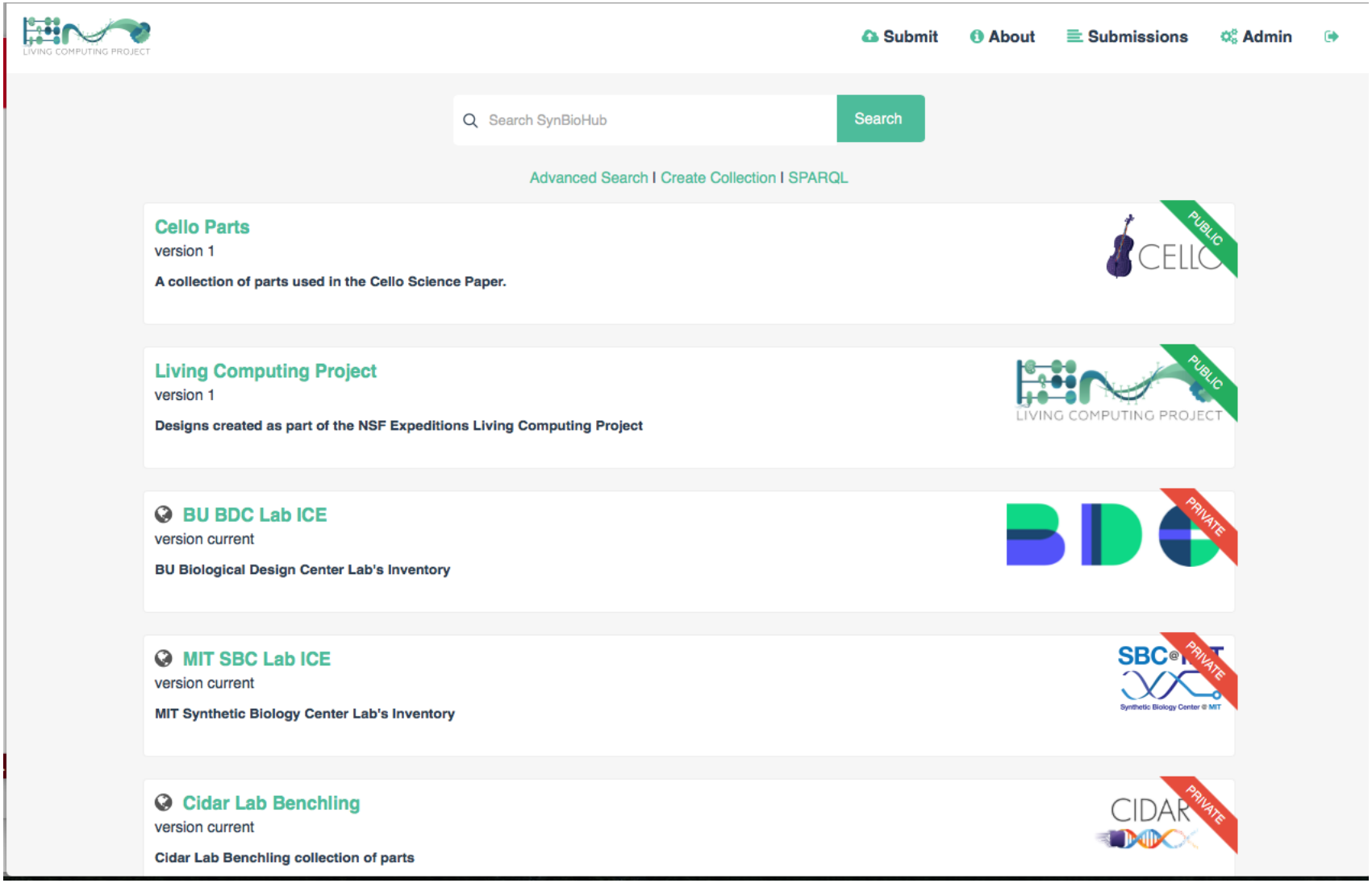
Version 1.0 released

June 14, 2017

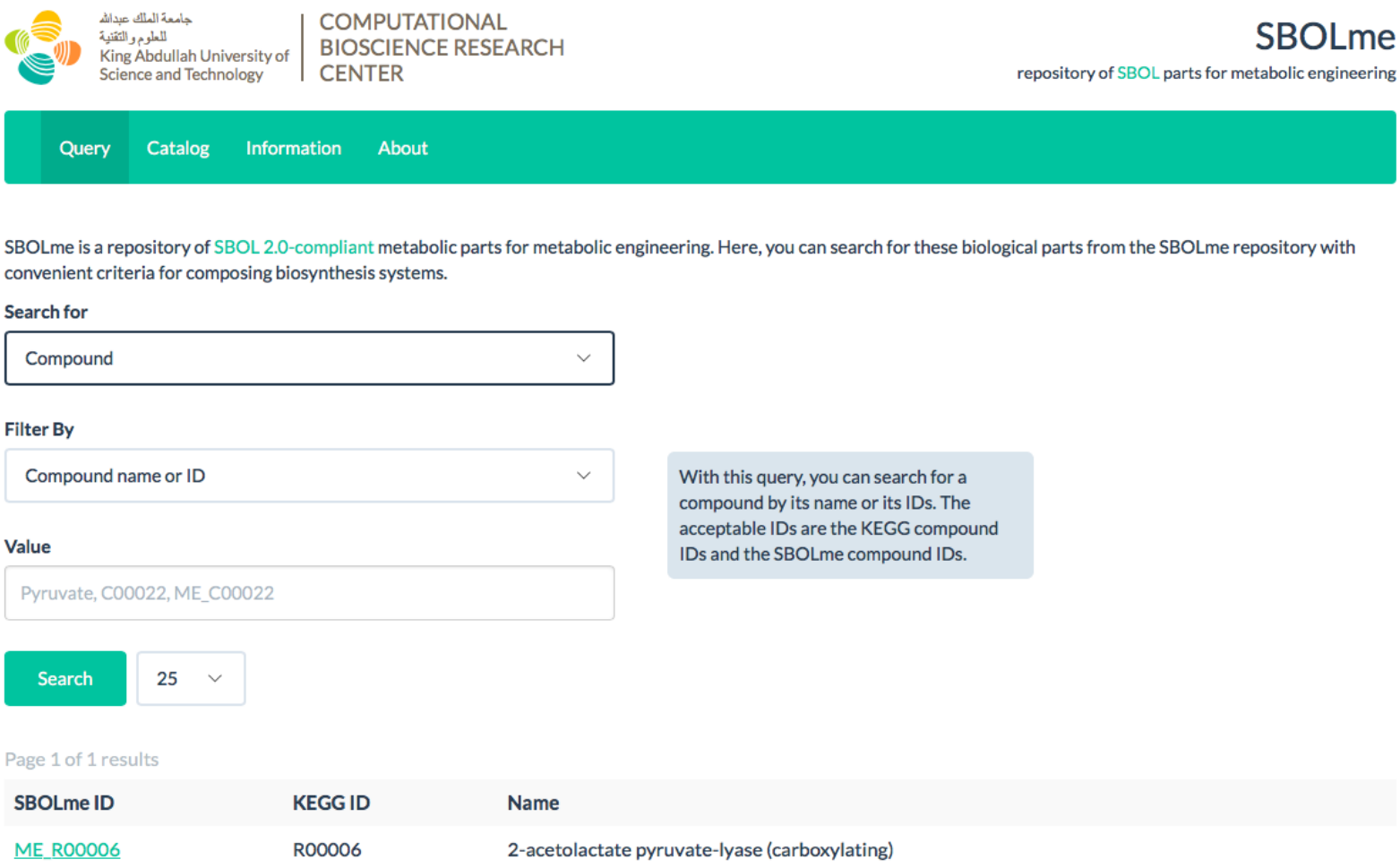
Reference Instance (https://synbiohub.org)



BU/MIT Living Computing Project (https://synbiohub.programmingbiology.org)



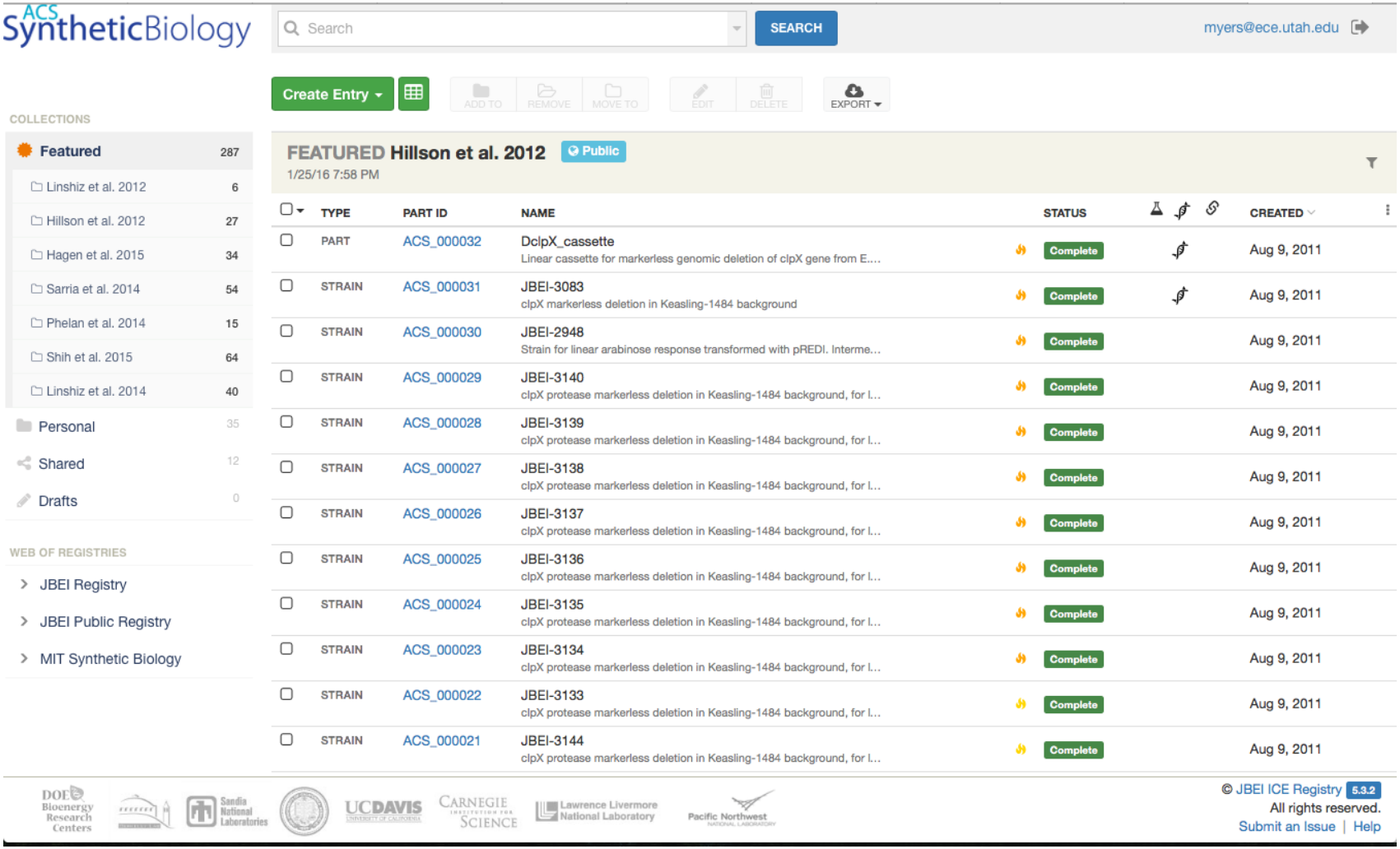
SBOLme (http://www.cbrc.kaust.edu.sa/sbolme/)



Kuwahara et al., ACS Synthetic Biology (2017)

Includes 28,437 chemical compounds, 6,883 enzyme classes, 9,909 metabolic reactions, and 3,173,238 proteins from 3,908 organisms.

ICE (https://acs-registry.jbei.org)



Ham et al., Nucleic Acid Research (2012)

Other Useful Repositories

- NCBI - https://www.ncbi.nlm.nih.gov
  - PubMed - https://www.ncbi.nlm.nih.gov/pubmed
  - GeneBank - https://www.ncbi.nlm.nih.gov/genbank
  - Protein - https://www.ncbi.nlm.nih.gov/protein
  - Many others - https://www.ncbi.nlm.nih.gov/guide/all/
- EMBL-EBI - http://www.ebi.ac.uk
  - BioModels - http://www.ebi.ac.uk/biomodels-main/
  - ChEBI - http://www.ebi.ac.uk/chebi/
  - Reactome - http://reactome.org
  - UniProt - http://www.uniprot.org
  - Many others - http://www.ebi.ac.uk/services/all
- addgene - https://www.addgene.org
- Pathway Commons - http://www.pathwaycommons.org

Data Standards: FASTA

```
>BBa_B0015 sequence 1 (129 bp)
ccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgt
tgtttgtcggtgaacgctctctactagagtcacactggctcaccttcgggtgggcctttt
ctgcgtttata
```

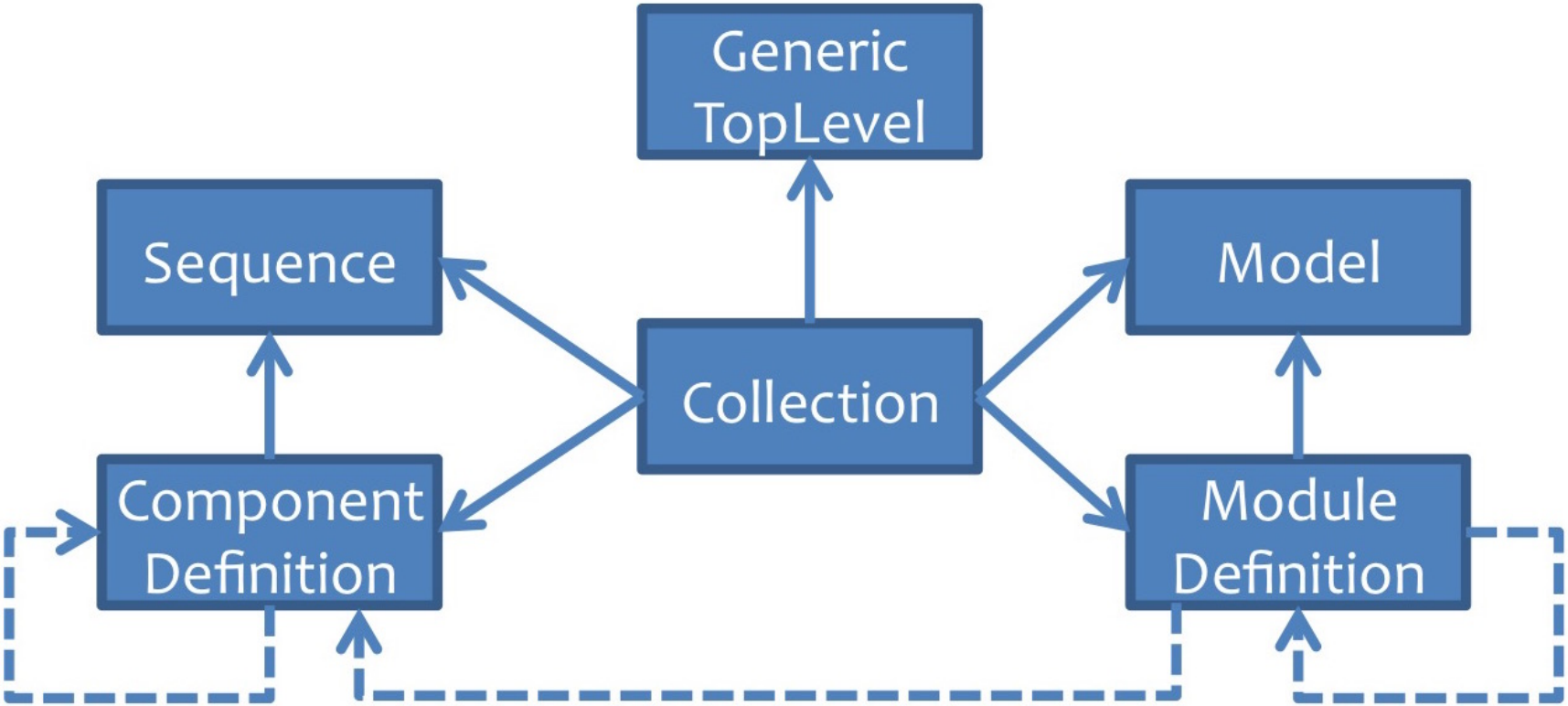
Data Standards: GenBank

```
LOCUS       BBa_B0015                 129 bp    DNA     linear   UNK 23-Aug-20
DEFINITION  double terminator (B0010-B0012)
ACCESSION   BBa_B0015
VERSION     BBa_B0015.1
FEATURES             Location/Qualifiers
     misc_feature    12..55
     misc_feature    1..80
     terminator      1..80
     misc_feature    96..115
     misc_feature    122..122
     polyA_site      116..129
     misc_feature    89..129
     terminator      89..129
ORIGIN
        1 ccaggcatca aataaaacga aaggctcagt cgaaagactg ggcctttcgt tttatctgtt
       61 gtttgtcggt gaacgctctc tactagagtc acactggctc accttcgggt gggcctttct
      121 gcgtttata
//
```

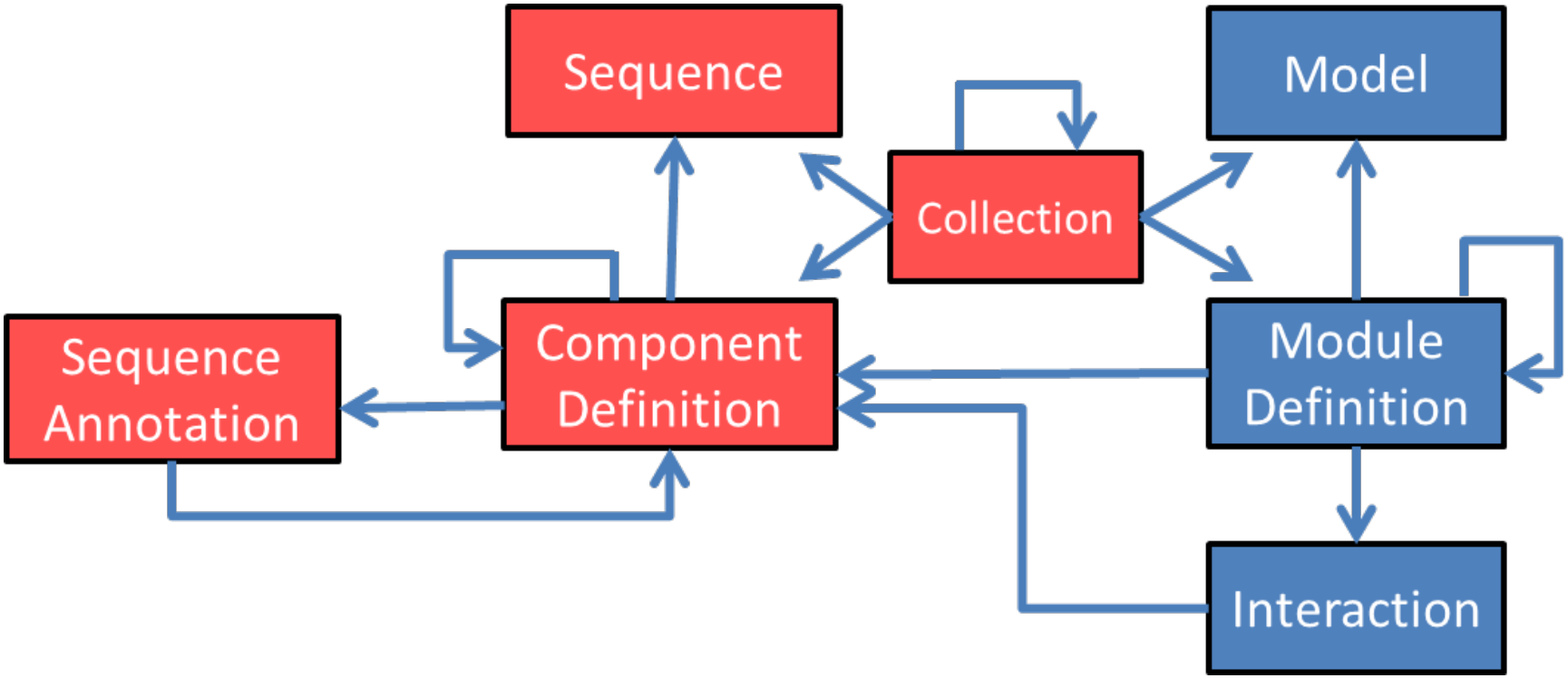
Data Standards: SBOL

- Provides support for:
  - Incomplete sequences.
  - Hierarchical designs.
  - Non-DNA components.
  - Functional interactions.
  - External models.
  - Arbitrary annotations.
- Serialized in XML/RDF format.
- Leverages *semantic web* database technology.
- There are several library implementations of the SBOL data structure, which provide an *application programmers interface* (API) for tool developers to interact with SBOL data objects.

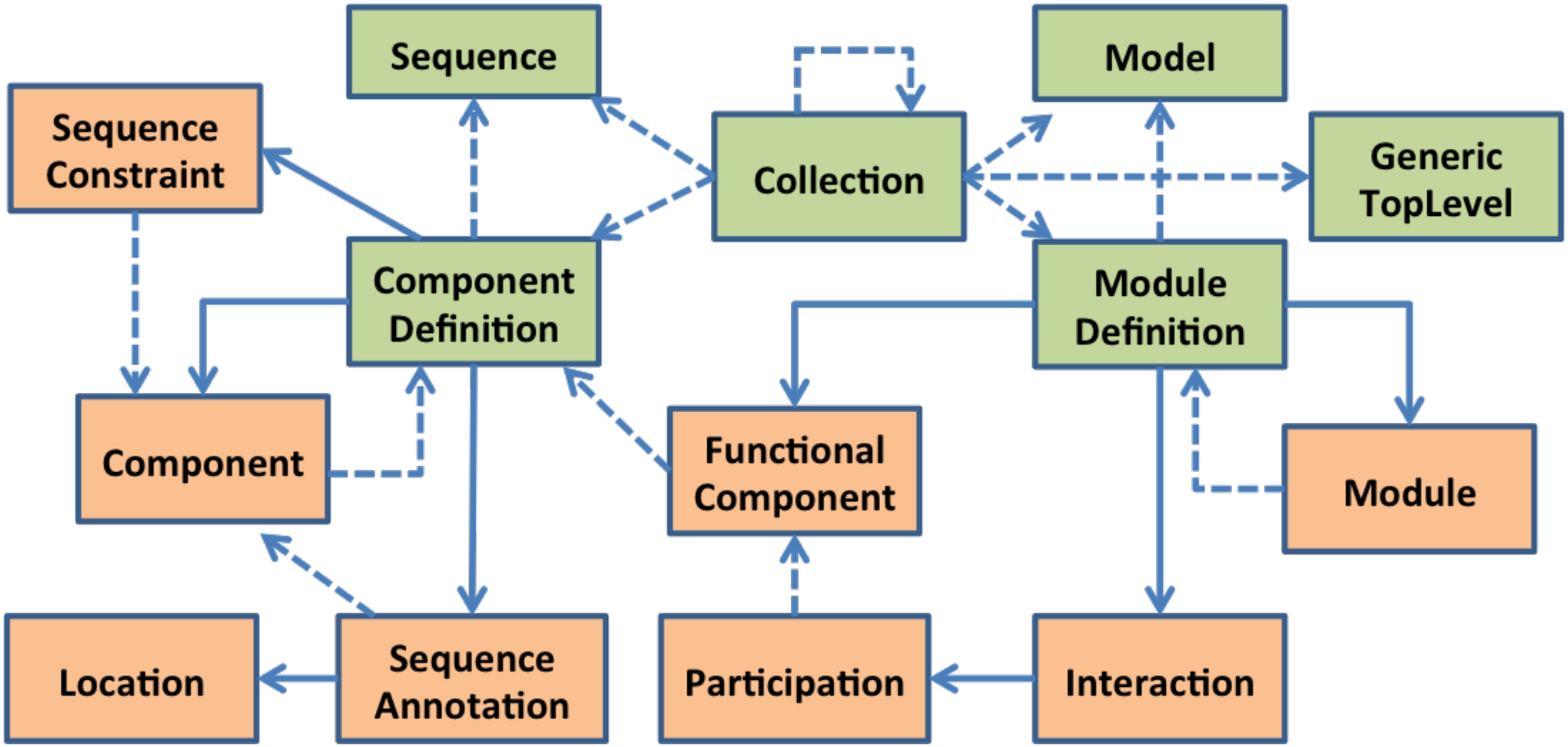
Data Standards: SBOL



Data Standards: SBOL



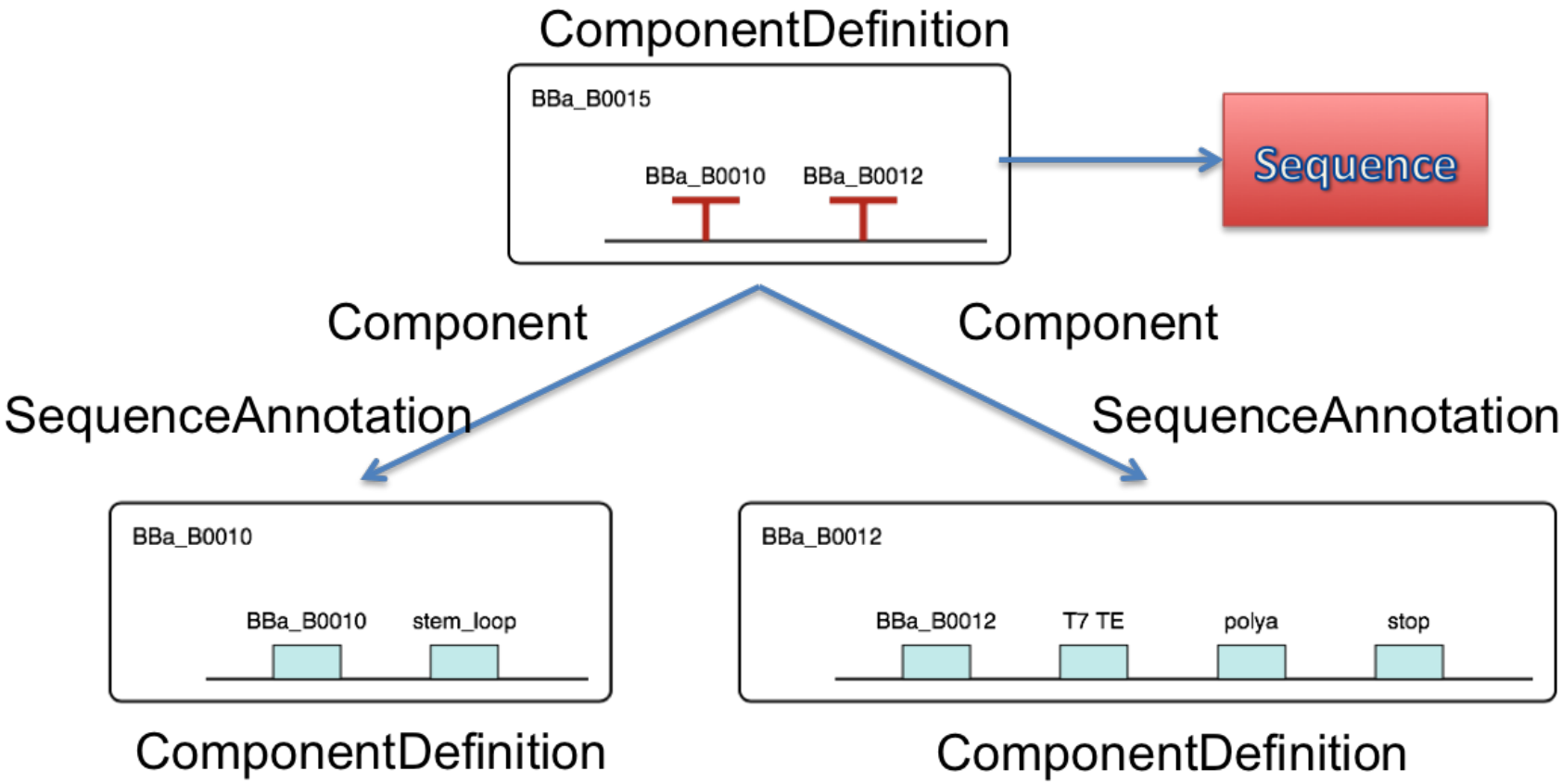
Data Standards: SBOL



SBOL 2.2 Updates

- Attachment class - serves as a general container for data files.
- Implementation class - represents a physical instance of a synthetic biological construct.
- CombinatorialDerivation class - specifies combinatorial genetic designs without having to specify every possible design variant.
- Provenance ontology (Prov-O):
  - Activity class - something that occurs over a period of time and acts upon or with entities.
  - Agent class - something that bears some form of responsibility for an activity taking place.
  - Plan class - an entity that represents a set of actions or steps intended by one or more agents to achieve some goals.

SBOL Example



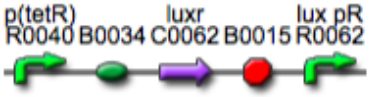
Conversion of the iGEM Registry to SBOL

Conversion of the iGEM Registry to SBOL

Part:BBa_F2620:Design

Designed by: Barry Canton [bcanton@mit.edu] and An

3OC$_6$HSL -> PoPS Receiver



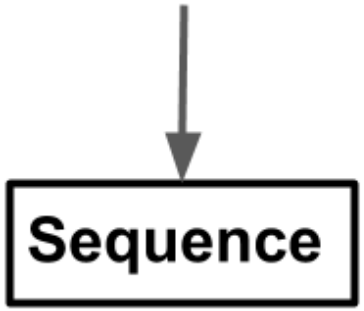
ComponentDefinition

Conversion of the iGEM Registry to SBOL

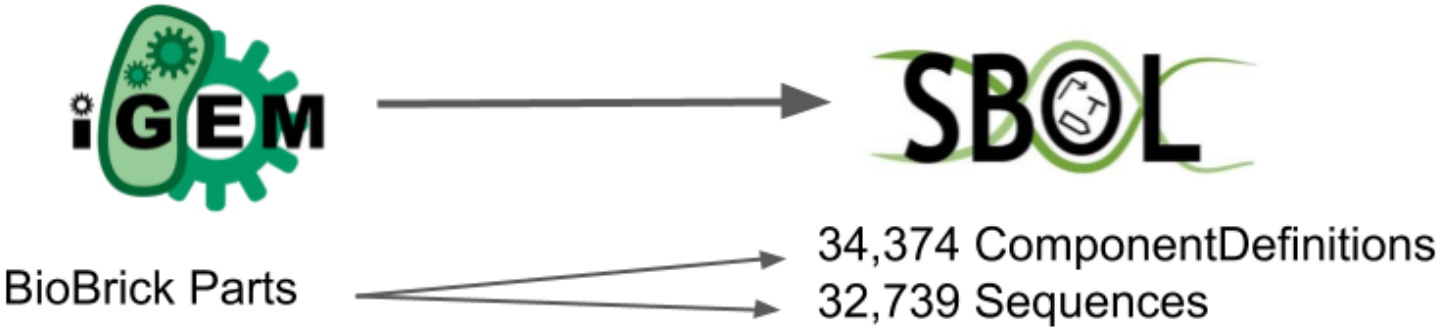
Conversion of the iGEM Registry to SBOL

34,374 ComponentDefinitions

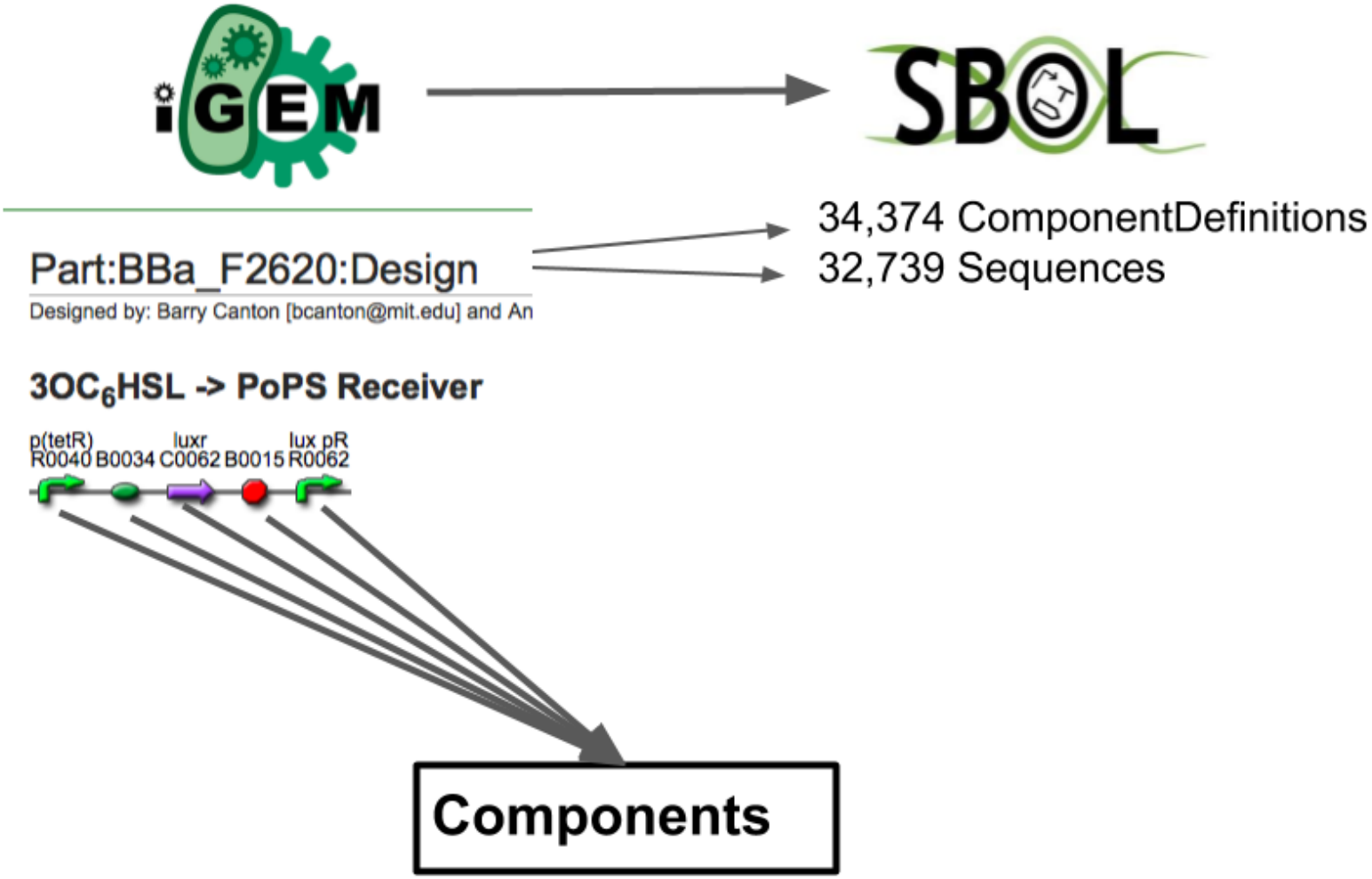
```
>BBa_F2620 Part-only sequence (1061 bp)
tccctatcagtgatagagattgacatccctatcagtgatagagatactgagcactactagagaaagaggagaaatactagatgaaaaacataaatgccga
cgacacatacagaataattaataaaattaaagcttgtagaagcaataatgatattaatcaatgcttatctgatatgactaaaatggtacattgtgaatat
tatttactcgcgatcatttatcctcattctatggttaaatctgatatttcaatcctagataattaccctaaaaaatggaggcaatattatgatgacgcta
atttaataaaatatgatcctatagtagattattctaactccaatcattcaccaattaattggaatatatttgaaaacaatgctgtaaataaaaaatctcc
aaatgtaattaaagaagcgaaaacatcaggtcttatcactgggtttagtttccctattcatacggctaacaatggcttcggaatgcttagttttgcacat
tcagaaaaagacaactatatagatagtttatttttacatgcgtgtatgaacataccattaattgttccttctctagttgataattatcgaaaaataaata
tagcaaataataaatcaaacaacgatttaaccaaaagagaaaaagaatgtttagcgtgggcatgcgaaggaaaaagctcttgggatatttcaaaaatatt
aggttgcagtgagcgtactgtcactttccatttaaccaatgcgcaaatgaaactcaatacaacaaaccgctgccaaagtatttctaaagcaattttaaca
ggagcaattgattgcccatactttaaaaattaataacactgatagtgctagtgtagatcactactagagccaggcatcaaataaaacgaaaggctcagtc
gaaagactgggcctttcgttttatctgttgtttgtcggtgaacgctctctactagagtcacactggctcaccttcgggtgggcctttctgcgtttatata
ctagagacctgtaggatcgtacaggtttacgcaagaaaatggtttgttatagtcgaataaa
```



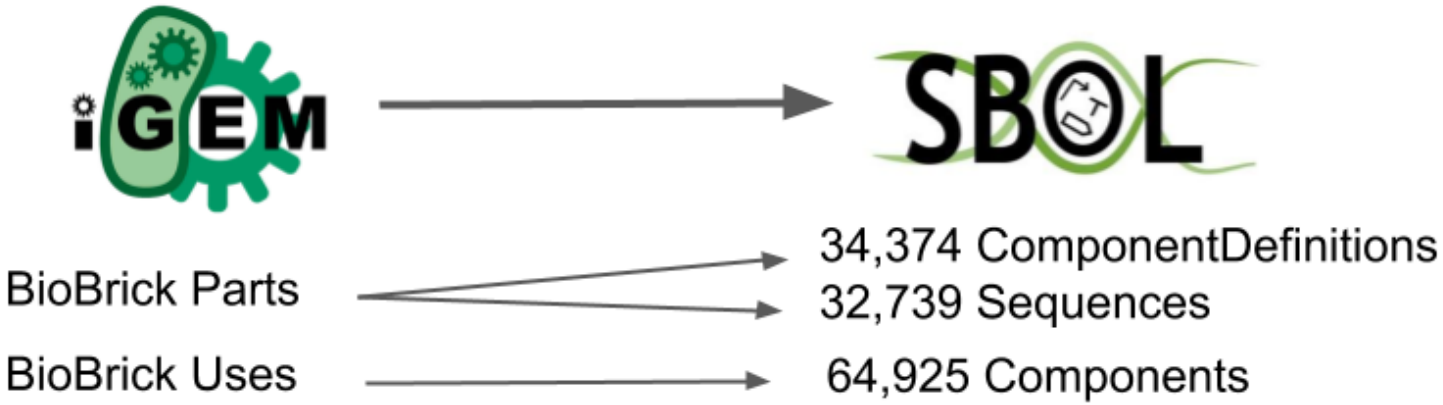
Conversion of the iGEM Registry to SBOL



Conversion of the iGEM Registry to SBOL

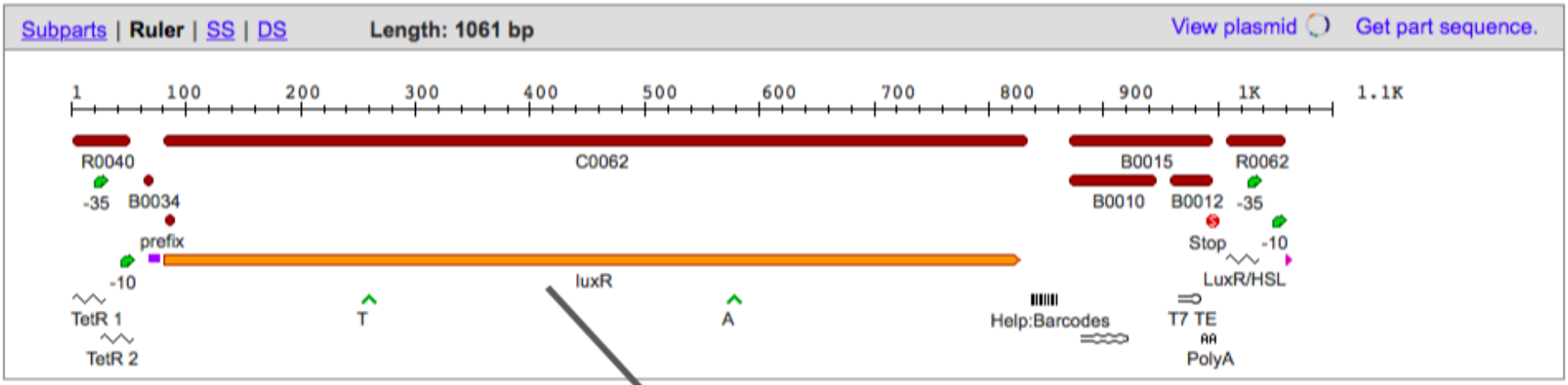


Conversion of the iGEM Registry to SBOL



Conversion of the iGEM Registry to SBOL

34,374 ComponentDefinitions



SequenceAnnotations

Conversion of the iGEM Registry to SBOL

BioBrick Parts → 34,374 ComponentDefinitions
BioBrick Parts → 32,739 Sequences

BioBrick Uses → 64,925 Components

BioBrick Features → 102,226 SequenceAnnotations

Conversion of the iGEM Registry to SBOL

iGEM Part/Feature Type	SequenceOntology (SO) Term
Coding	CDS (SO:0000316)
Device	Engineered Region (SO:0000804)
Primer	Primer (SO:0000112)
Protein Domain	Polypeptide Domain (SO:0000417)
RBS	Ribosome Entry Site (SO:0000139)
Regulatory	Promoter (SO:0000167)
Tag	Tag (SO:0000324)
Terminator	Terminator (SO:0000141)
etc.	etc.

Conversion of the iGEM Registry to SBOL

BioBrick Parts → 34,374 ComponentDefinitions
BioBrick Parts → 32,739 Sequences

BioBrick Uses → 64,925 Components

BioBrick Features → 102,226 SequenceAnnotations

27 Part / 20 Feature Types → Sequence Ontology (SO) (includes 2817 terms)

Conversion of the iGEM Registry to SBOL

Page Header

Login to edit

Part Name	**BBa_F2620**
Short Description	**3OC$_6$HSL -> PoPS Receiver**
Part Type	Signalling
Nickname	
Designer(s)	**Barry Canton [bcanton@mit.edu] and Anna Labno [labnoa@mit.edu]**
DNA Status	Available
Qualitative Experience	Works
Group Favorite	No
Star Rating	1
Delete This Part	Not Deleted

Dublin Core/iGEM Annotations

Conversion of the iGEM Registry to SBOL

BioBrick Parts → 34,374 ComponentDefinitions
BioBrick Parts → 32,739 Sequences

BioBrick Uses → 64,925 Components

BioBrick Features → 102,226 SequenceAnnotations

27 Part / 20 Feature Types → Sequence Ontology (SO) (includes 2817 terms)

Additional metadata → Dublin Core / iGEM annotations

Conversion of the iGEM Registry to SBOL

BioBrick Parts → 34,374 ComponentDefinitions, 32,739 Sequences

BioBrick Uses → 64,925 Components

BioBrick Features → 102,226 SequenceAnnotations

27 Part / 20 Feature Types → Sequence Ontology (SO) (includes 2817 terms)

Categories
//classic/signalling/receiver
//function/cellsignalling

→ Dublin Core / custom annotations

→ **Collections**

Conversion of the iGEM Registry to SBOL

BioBrick Parts → 34,374 ComponentDefinitions
BioBrick Parts → 32,739 Sequences

BioBrick Uses → 64,925 Components

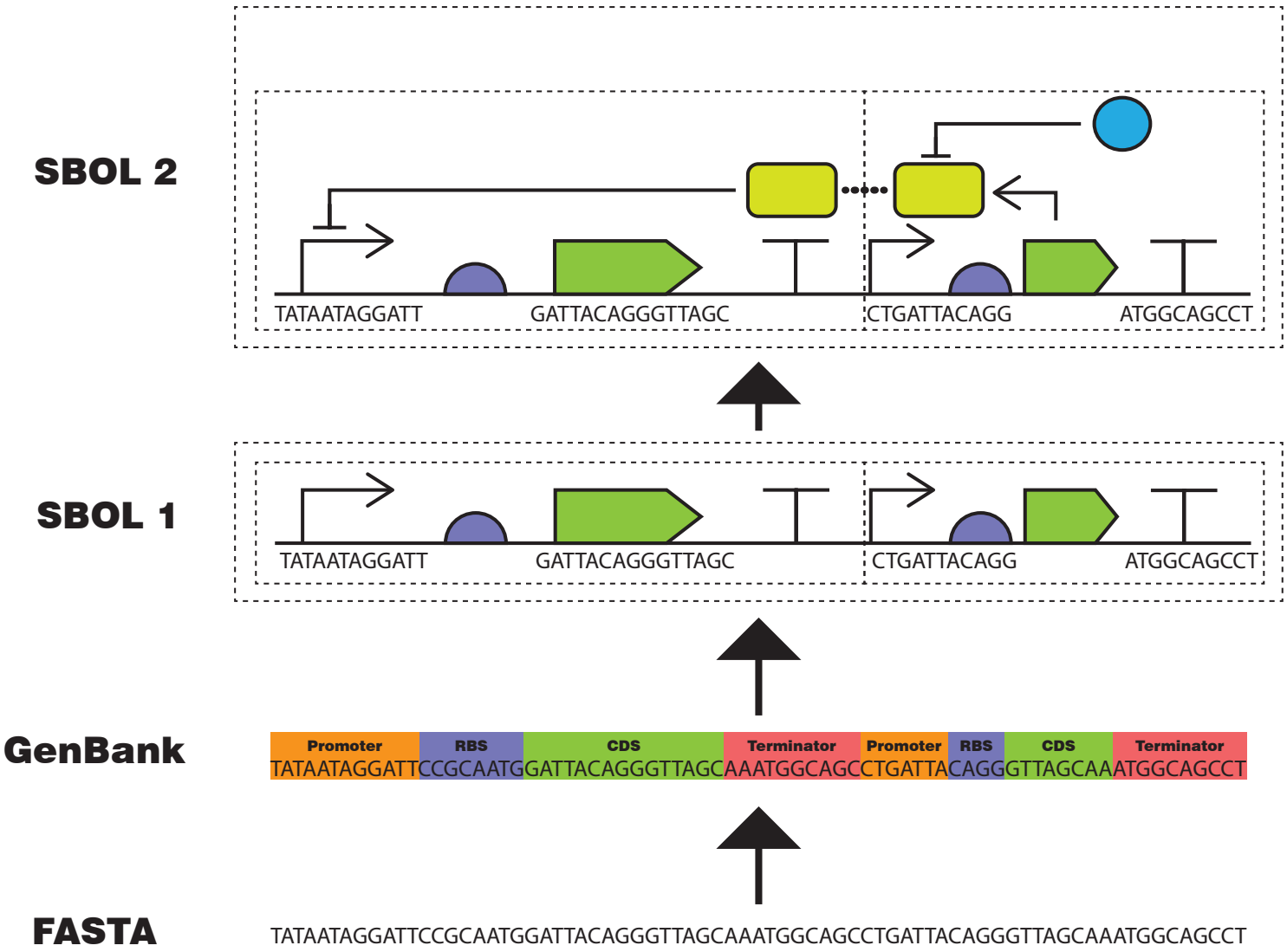
BioBrick Features → 102,226 SequenceAnnotations

27 Part / 20 Feature Types → Sequence Ontology (SO) (includes 2817 terms)

Additional metadata → Dublin Core / custom annotations

370 Categories → Collections

Comparison of Standard Data Formats



Promoters

- Recruit transcriptional machinery.
  - RNA polymerase
  - Other accessory proteins
- Lead to transcription of the downstream DNA sequence producing RNA.
- Can be thought of as a part that produces a certain number of RNA molecules per unit time.
- Constitutive
  - Not regulated
  - Always on
- Regulated
  - Activity is modulated by the presence of other protein factors.
  - Positive regulation
  - Negative regulation
- Strength of promoters varies - sequence and spacing of -10/-35 matters
- See `http://parts.igem.org/Promoters`

λ Promoters

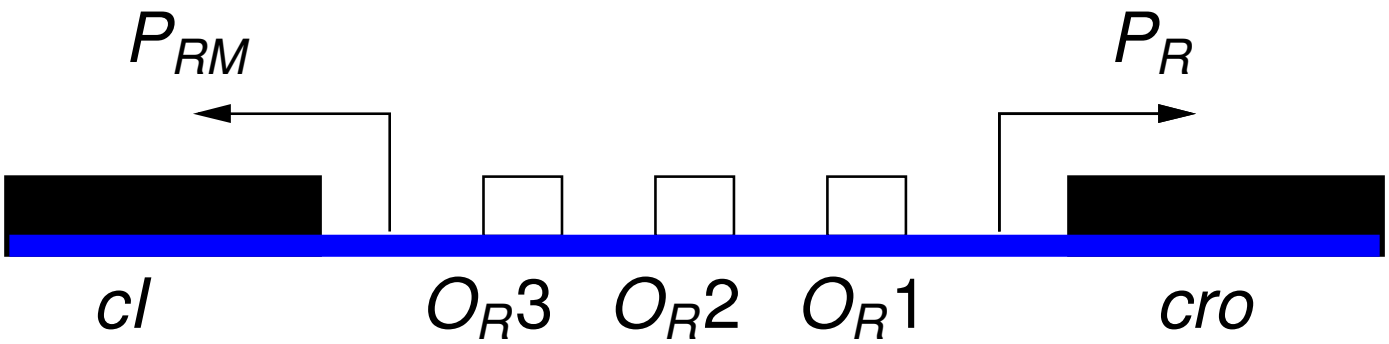
The registry symbol for a promoter is shown above a typical sequence for a bacterial promoter. The lavender shaded boxes indicate the two most conserved regions of a bacterial promoter and are located at -10 and -35 bases from the transcriptional start site (shaded in green). There are, on average, 17bp between the -10 and -35 sites and 7bp between the -10 site and the transcriptional start site [1][2][3][4].

	-35		-10
Consensus	T T G A C A	- 17bp -	T A T A A T
λP_{RM}	T A G A T A	- 17bp -	T A G A T T
λP_{R}	T T G A C T	- 17bp -	G A T A A T

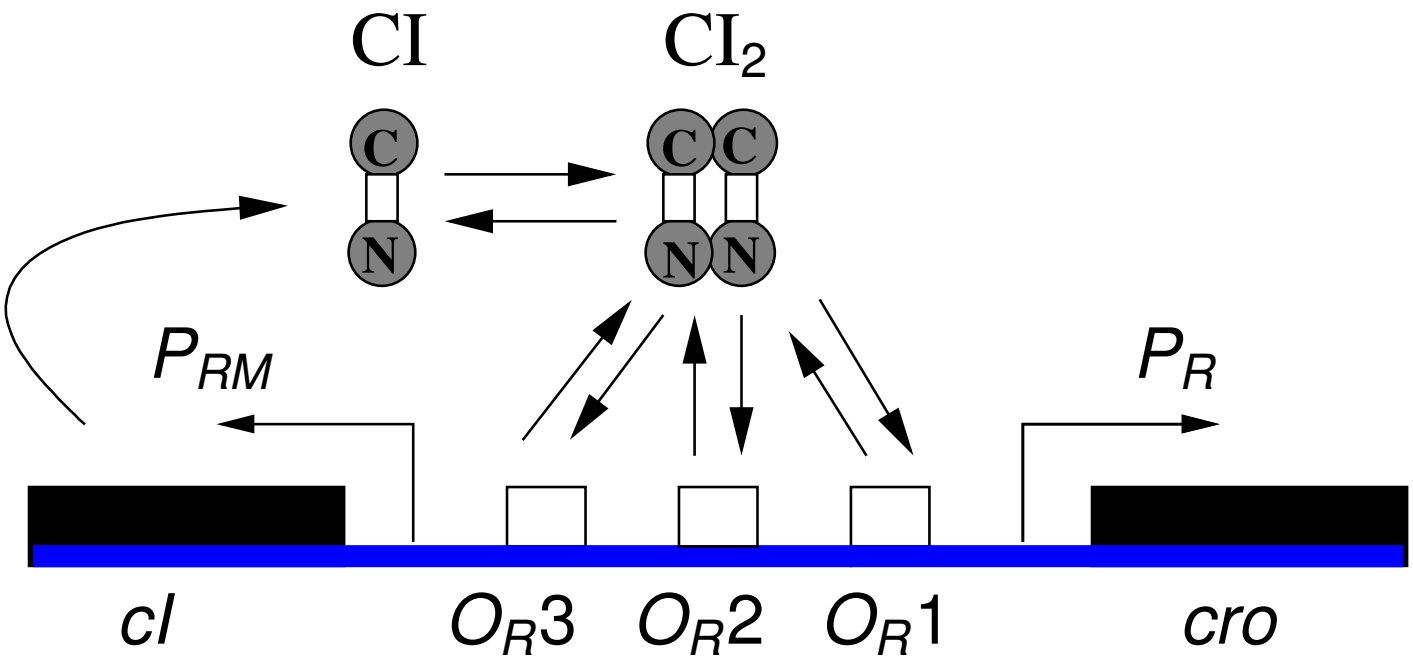
Operators

- Sequences that recruit the binding of *transcription factors* that regulate promoter activity.
- Classically operators recruit repressors to negatively regulate promoters.
- Also some operators bind activators (positive regulators).
- Often not considered as separate parts - included within the promoter.
- Strength of regulation mediated by the sequence of the operator.
- Can be manipulated to tune the level of repression or activation.

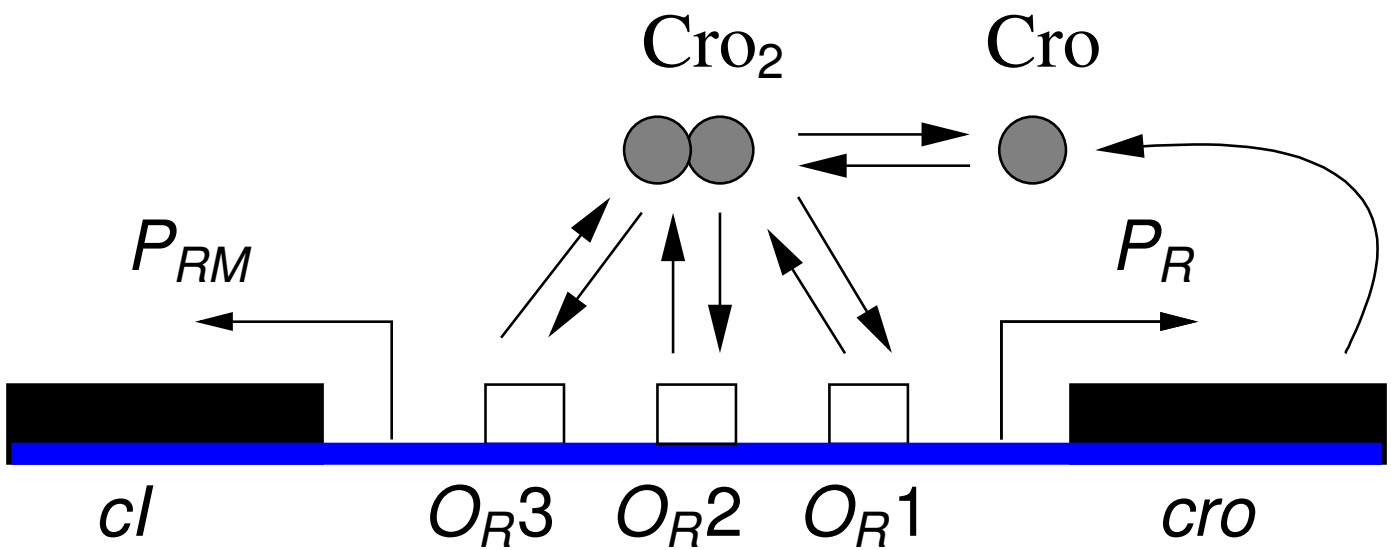
The O_R Operator



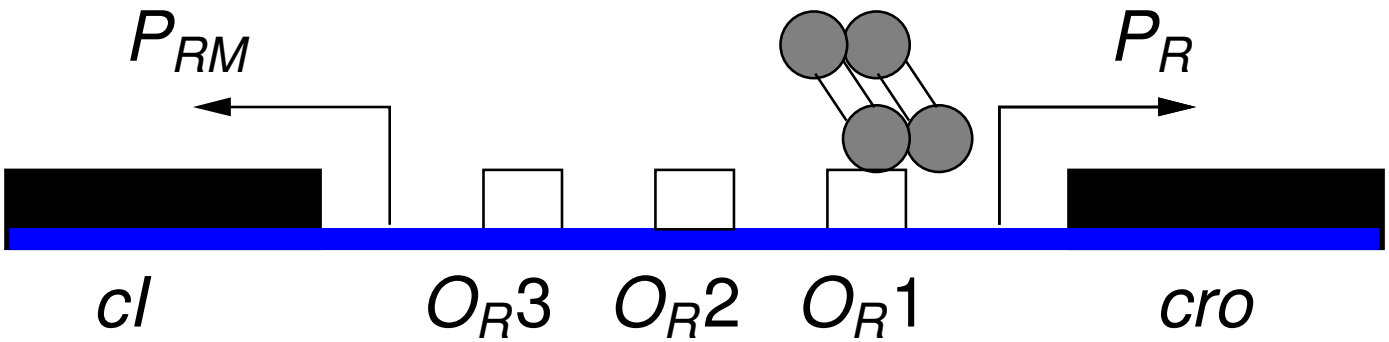
CI, the λ Repressor



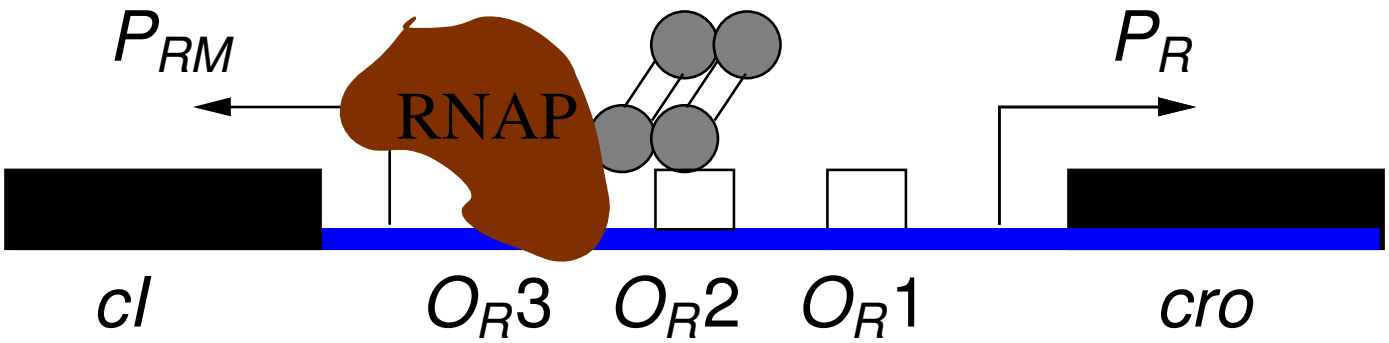
λ's Cro Molecule



CI_2 Bound to O_R1 Turns Off P_R



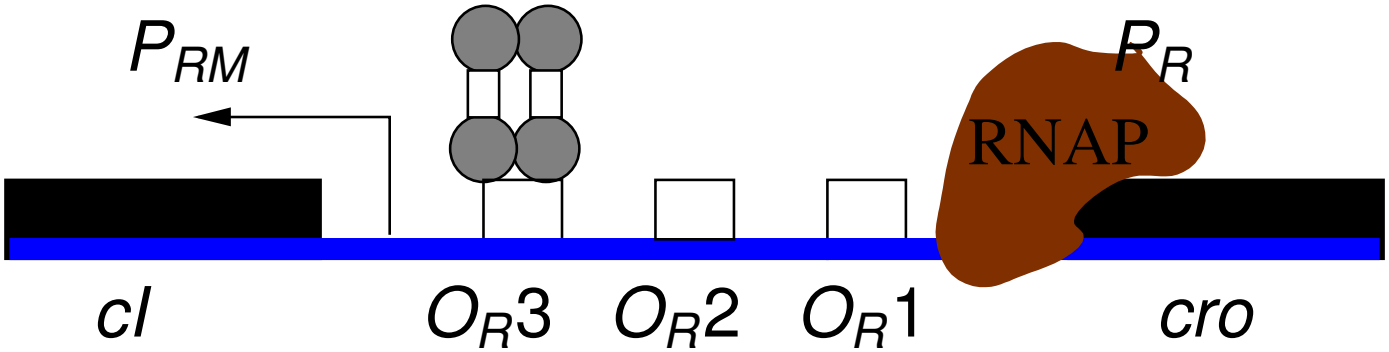
CI_2 Bound to O_R2 Turns On P_{RM}



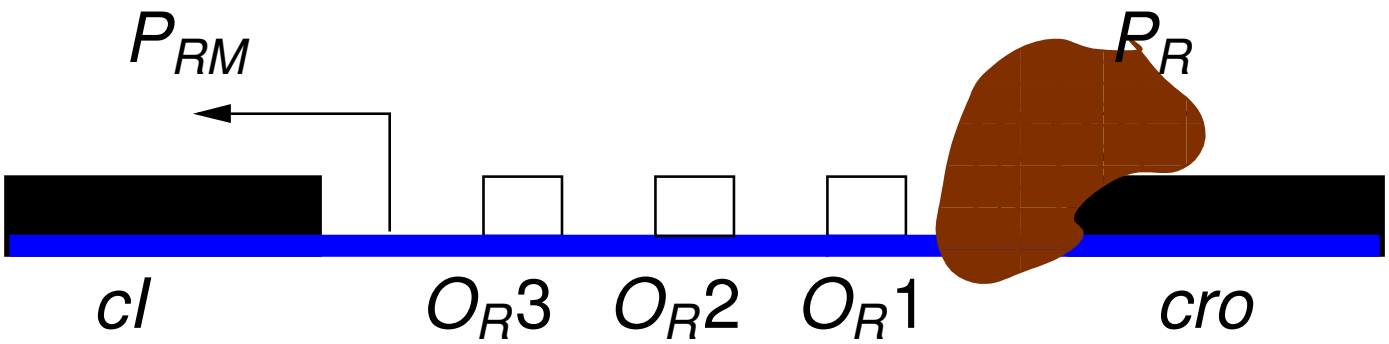
Basal Versus Activated Rate

- Nothing in biology is clear-cut.
- Without CI_2 bound to O_R2, RNAP can still bind to P_{RM} and initiate transcription of CI at a reduced *basal rate*.
- With CI_2 bound to O_R2, transcription occurs at enhanced *activated rate*.

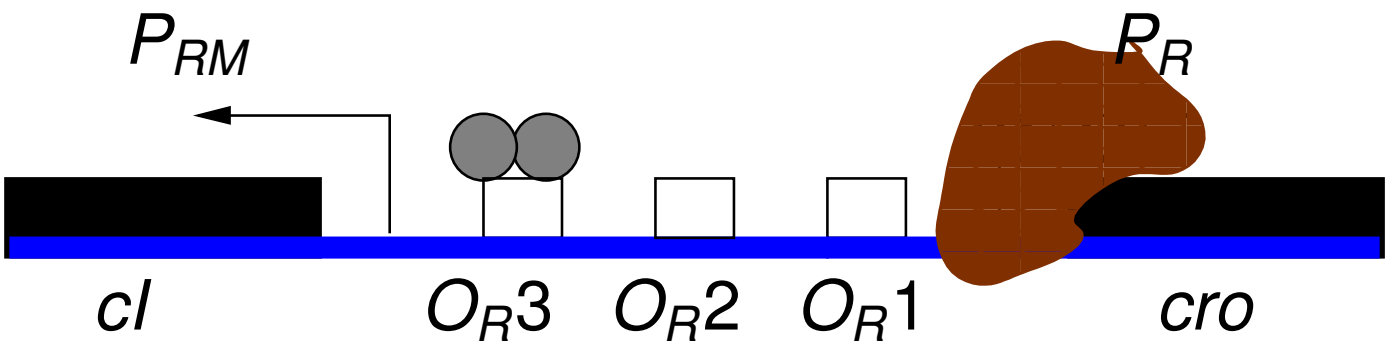
CI_2 Bound to O_R3 Turns Off P_{RM}



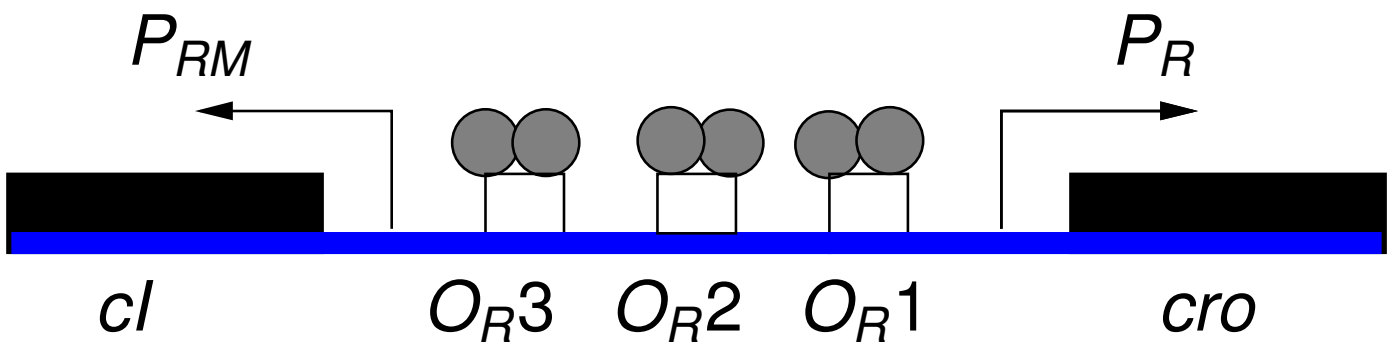
P_R Active When O_R Sites Are Empty



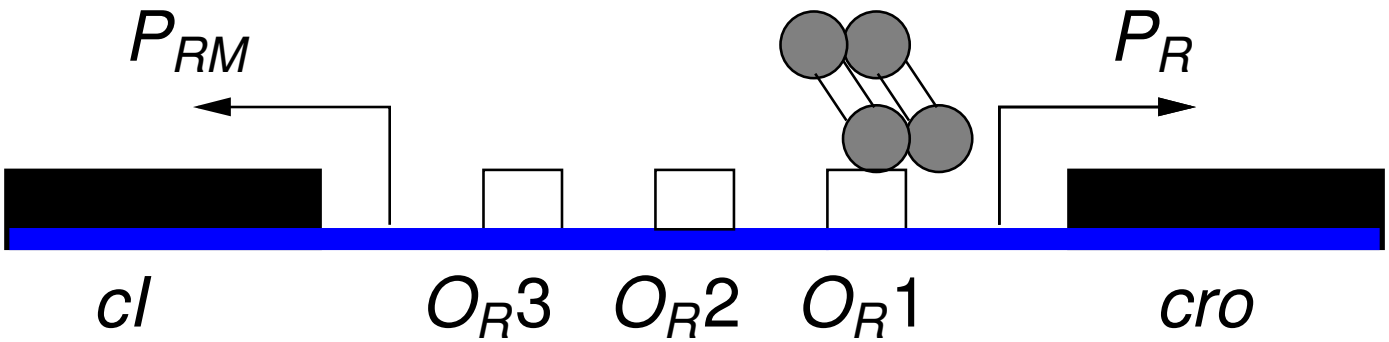
Cro_2 Bound to O_R3 Turns off P_{RM}



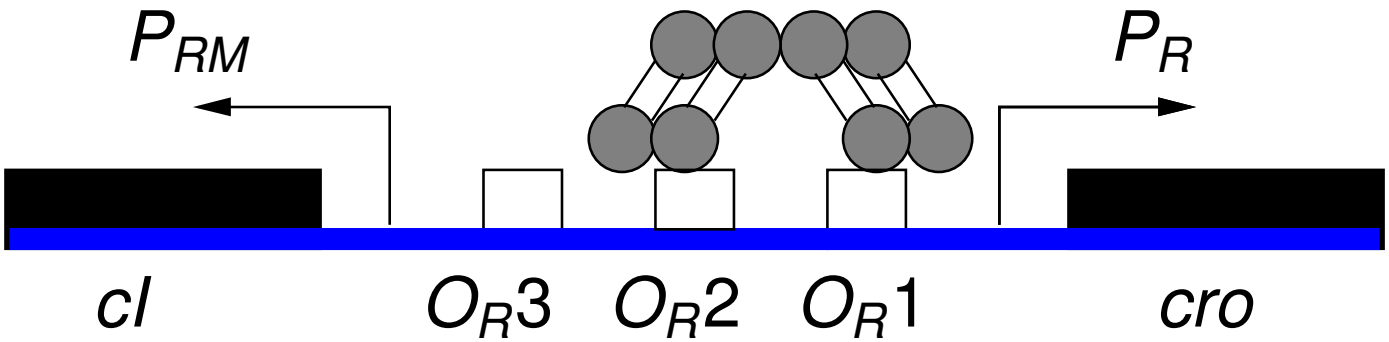
Cro_2 Bound to O_R1 or O_R2 Turns off P_R



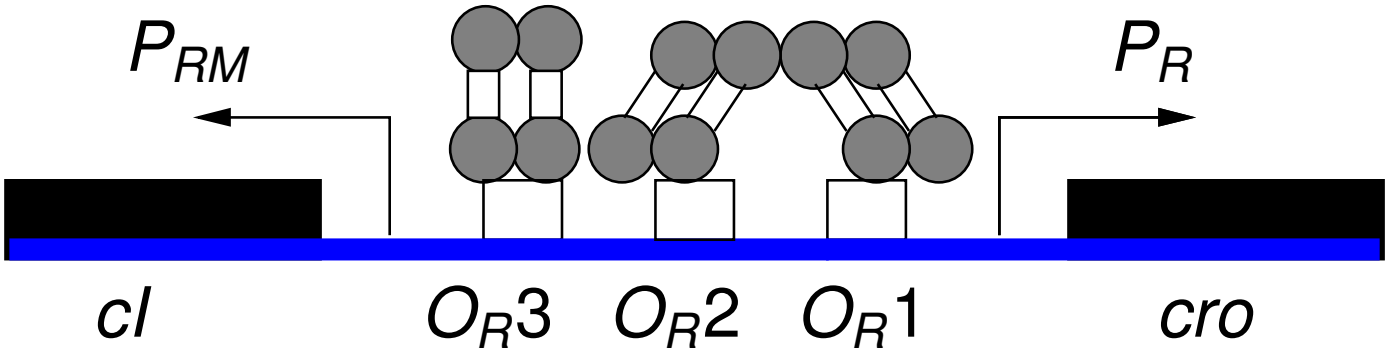
Low Concentrations of CI_2



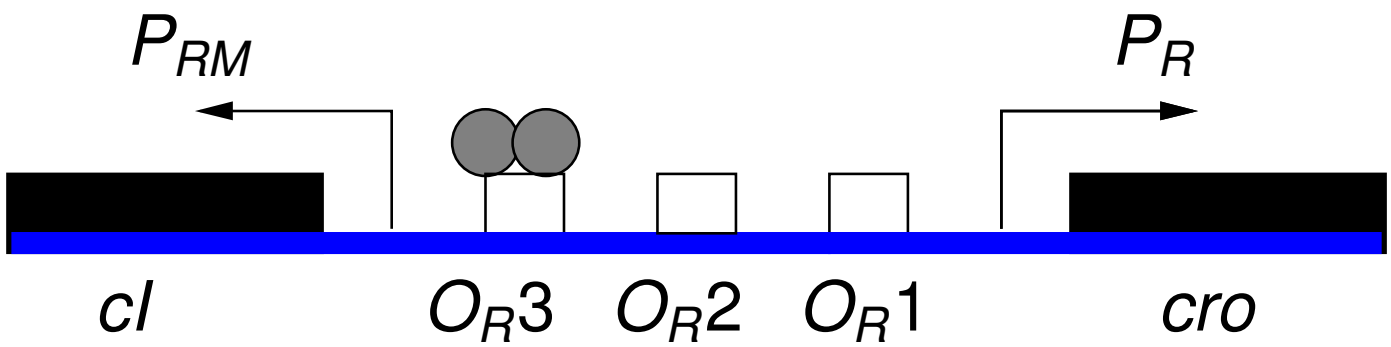
Cooperativity Aids CI_2 Binding to O_R2



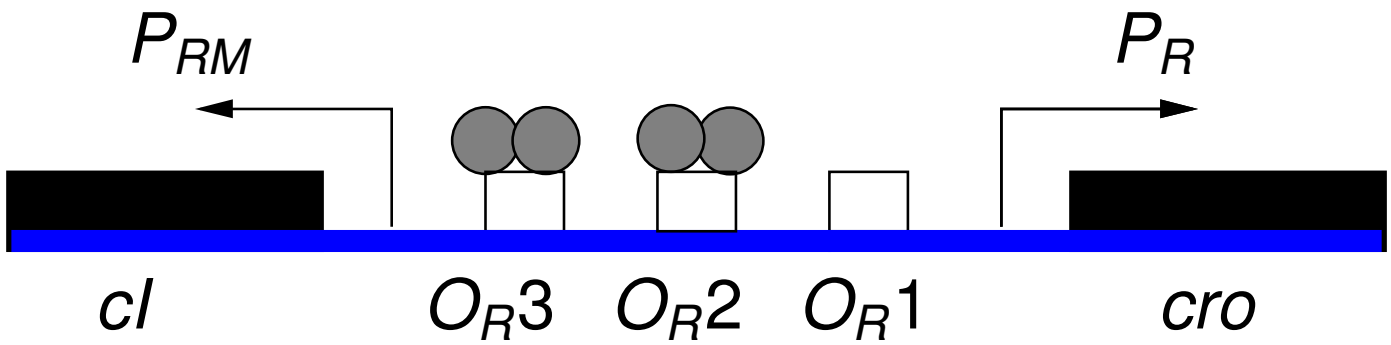
High Concentrations of CI_2



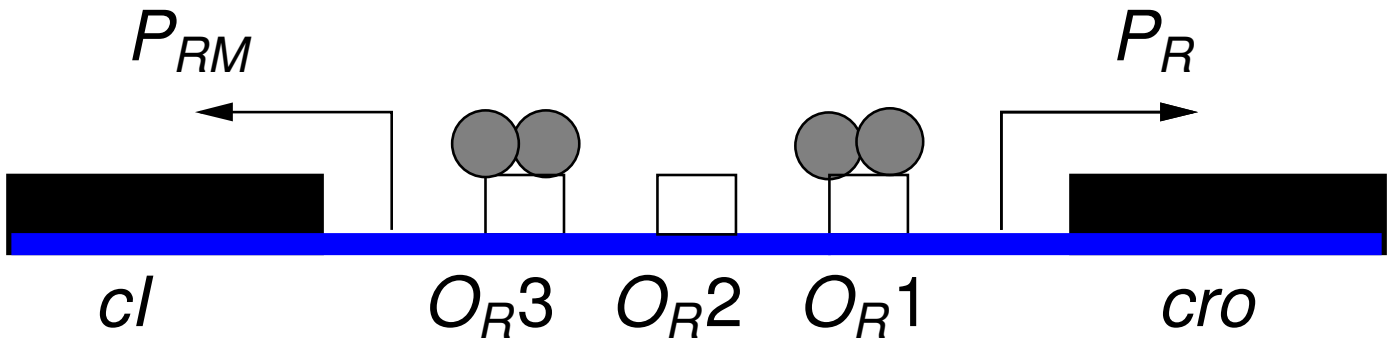
Low Concentrations of Cro_2



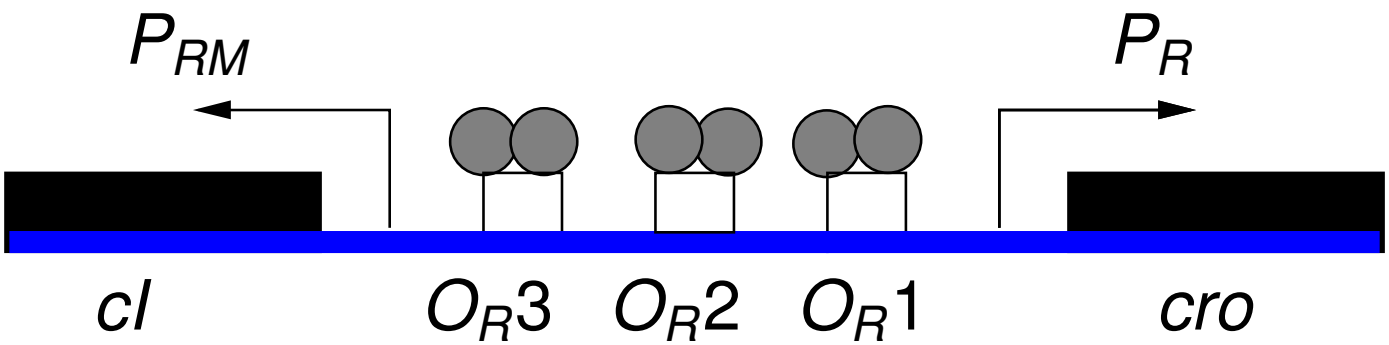
Moderate Concentrations of Cro_2



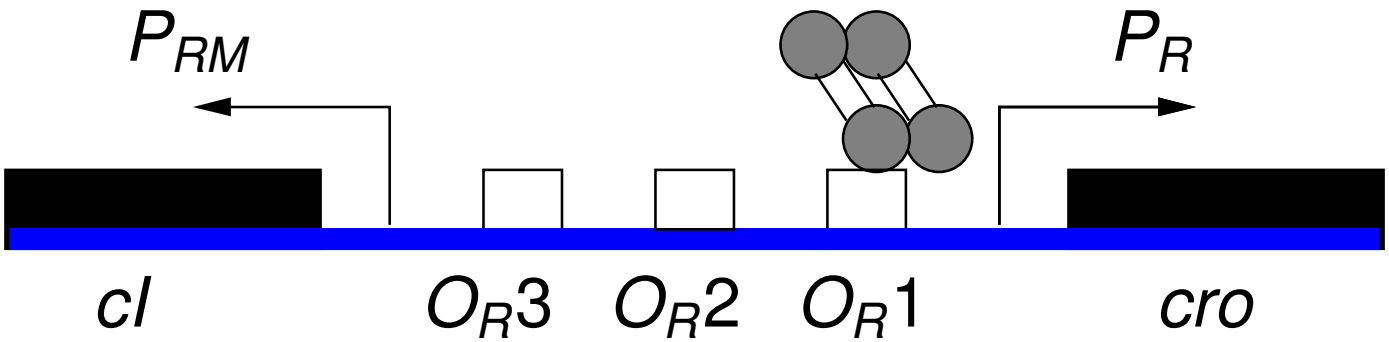
OR



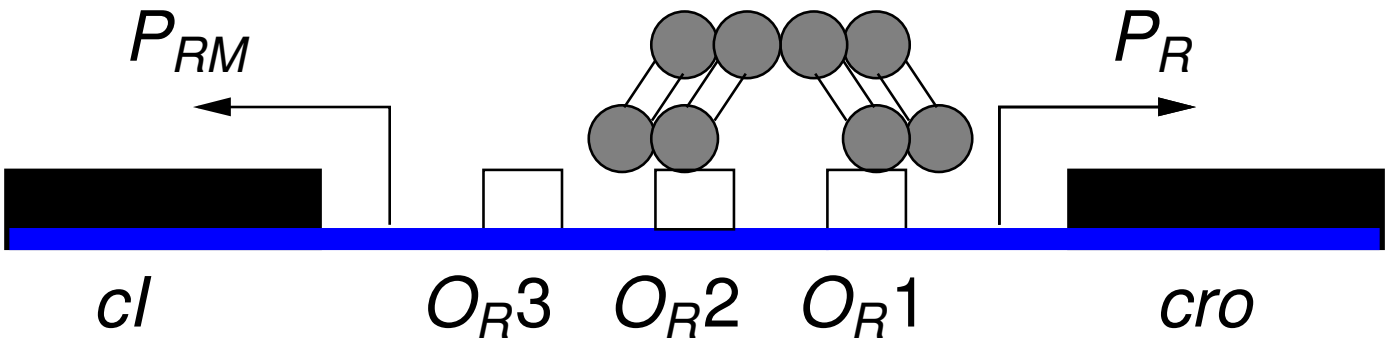
High Concentrations of Cro_2



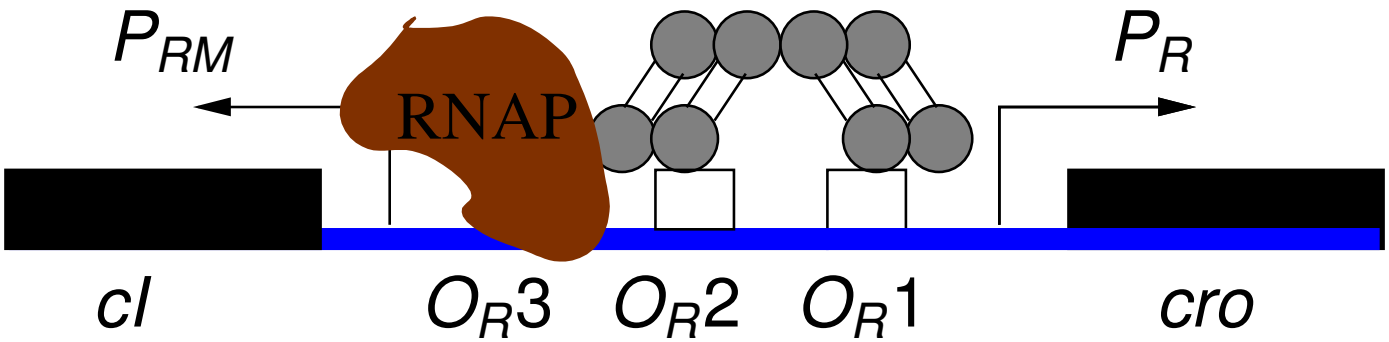
Cooperativity of CI_2 Binding



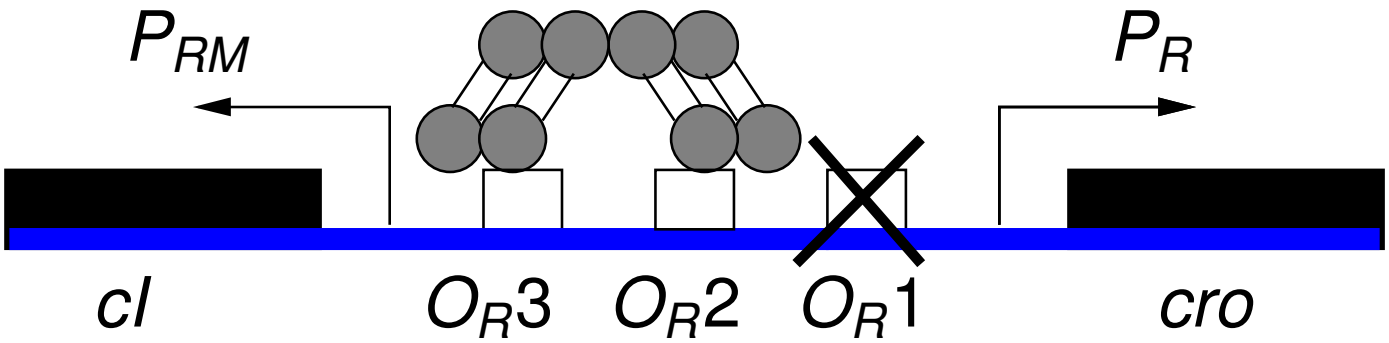
Cooperativity of CI_2 Binding



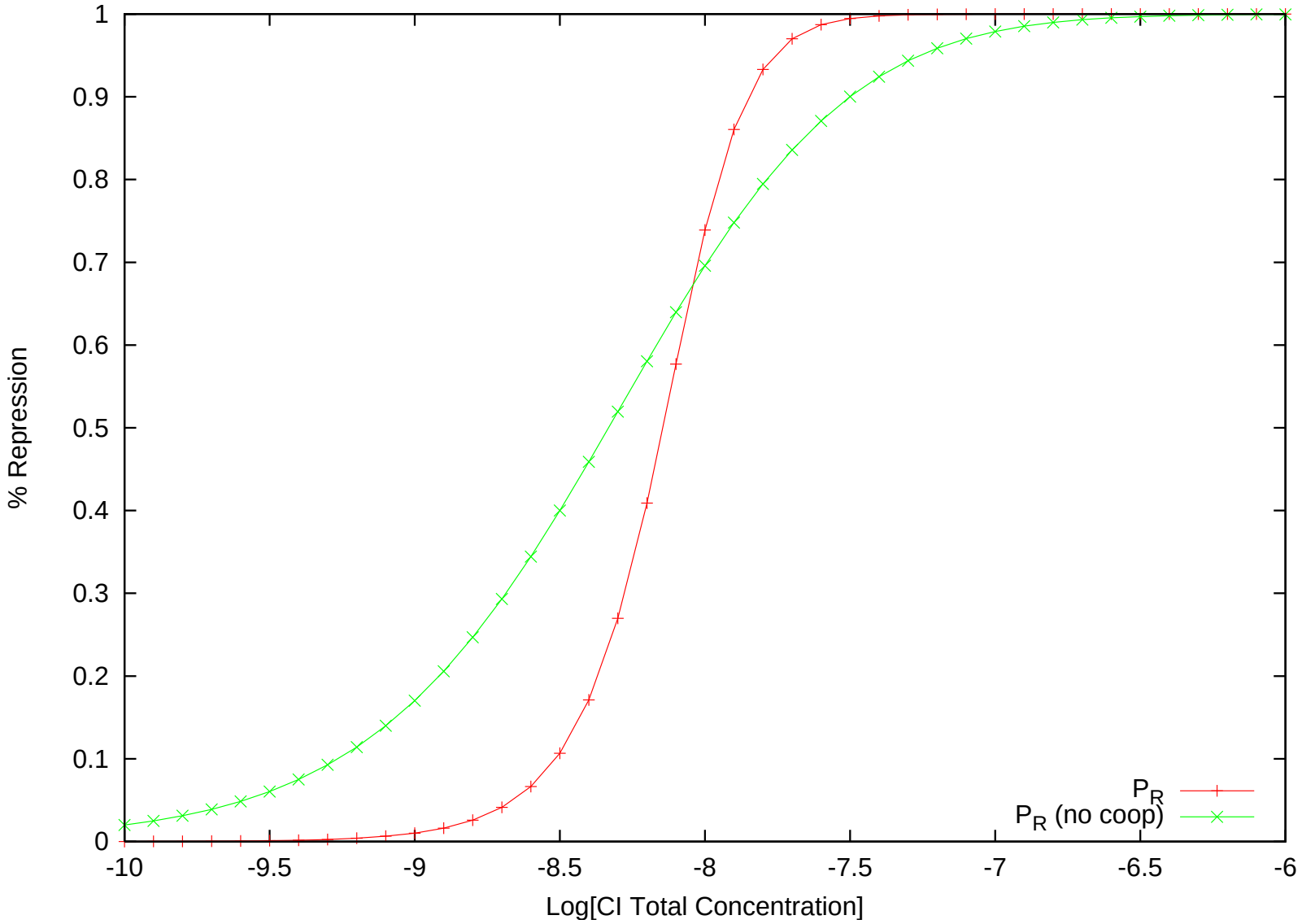
Cooperativity of CI_2 Binding



Cooperativity of CI_2 Binding



Effect of Cooperativity



Configurations of O_R

- At low to moderate concentrations of CI and Cro, there are three common configurations:
  - No molecules bound to O_R, Cro produced at full rate and CI produced at low basal rate.
  - CI_2 bound to O_R1 and O_R2, Cro production repressed, and CI activated.
  - Cro_2 bound to O_R3, CI cannot be produced, Cro is produced.
- Feedback of the products as transcription factors coupled with affinities makes O_R behave as a bistable switch.
- In *lysis* state, Cro produced locking out production of CI.
- In *lysogeny* state, CI produced locking out production of Cro.

Recognition of the Operators

- CI_2 and Cro_2 bind to operator sites that are 17 base pairs long.
- How do these proteins locate these sequences from amongst the millions within the bacteria?
- Observing from midpoint, a strand on one side is nearly symmetric with complimentary strand on other side.

Near Symmetry in the Operator Sequences

Op	Operator sequences	Operator half sequences
O_R1	*T A T C A C C G c C A G A G G T A* *A T A G T G G C g G T C T C C A T*	*T A T C A C C G c* *T A C C T C T G*
O_R2	*T A A C A C C G t G C G T G T T G* *A T T G T G G C a C G C A C A A C*	*T A A C A C C G t* *C A A C A C G C*
O_R3	*T A T C A C C G c A A G G G A T A* *A T A G T G G C g T T C C C T A T*	*T A T C A C C G c* *T A T C C C T T*
O_L1	*T A T C A C C G c C A G T G G T A* *A T A G T G G C g G T C A C C A T*	*T A T C A C C G c* *T A C C A C T G*
O_L2	*T A T C T C T G g C G G T G T T G* *A T A G A G A C c G C C A C A A C*	*T A T C T C T G* *C A A C A C C G c*
O_L3	*T A T C A C C G c A G A T G G T T* *A T A G T G G C g T C T A C C A A*	*T A T C A C C G c* *A A C C A T C T*
Con.		T_9 A_{12} T_6 C_{12} A_9 C_{11} C_7 G_9 C_5 C_2 C_3 T_2 T_1 T_4 T_2 T_1 A_1 A_3 C_1 G_1 C_1

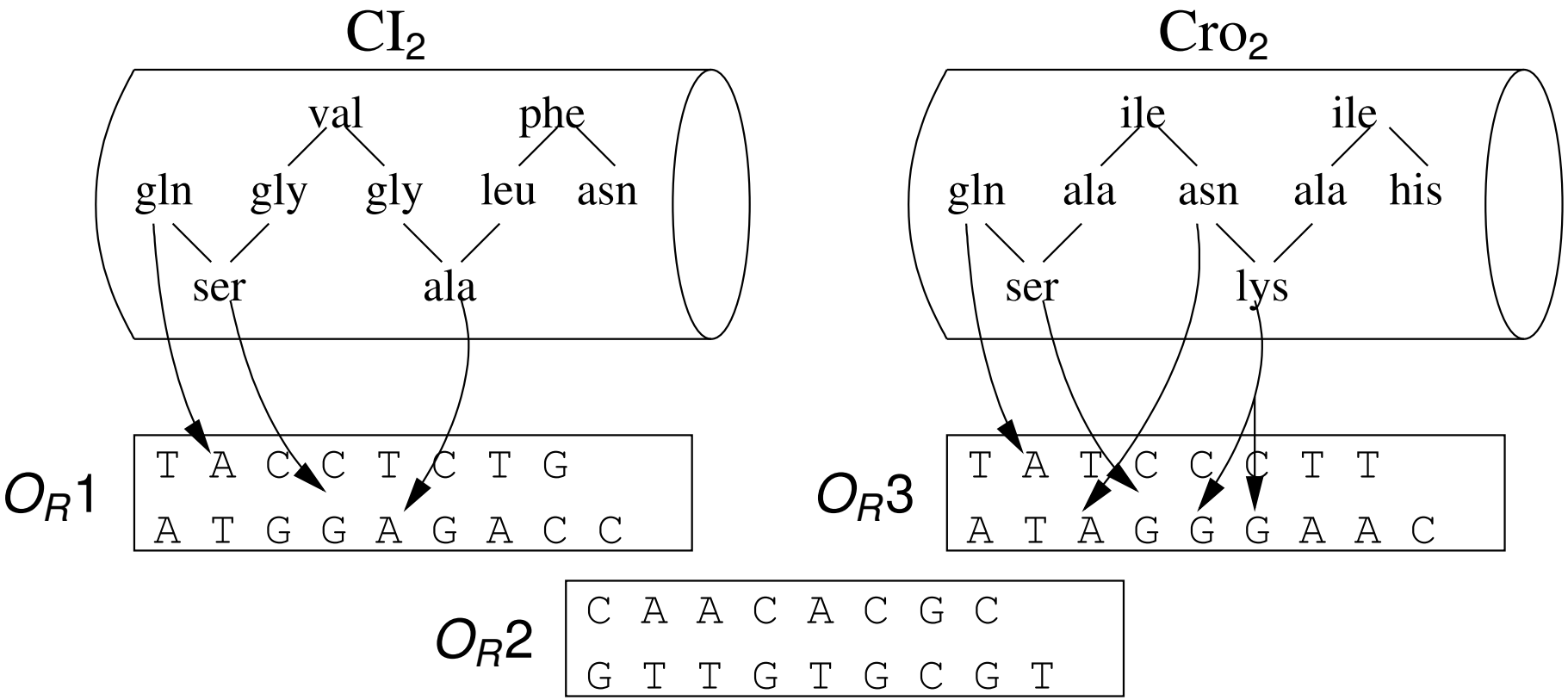
Consensus Sequence

- The *consensus sequence* is as follows:

```
TATCACCGcCGGTGATA
ATAGTGGCgGCCACTAT
```

- Many entries are highly preserved.
- Differences exist that cause the differences in affinity for CI_2 and Cro_2 for the different operators.
- Notice that the first half of the operator sites O_R1 and O_R3 agree perfectly with the consensus sequence while second half has several differences.

Amino Acid-Base Pair Interactions



Ribosome Binding Sites

- A sequence in mRNA to which ribosomes bind and initiate translation.
- Bacteria usually require RBS and start codon.
- Also called *Shine-Dalgarno sequence*.
- See `http://parts.igem.org/Ribosome_Binding_Sites`

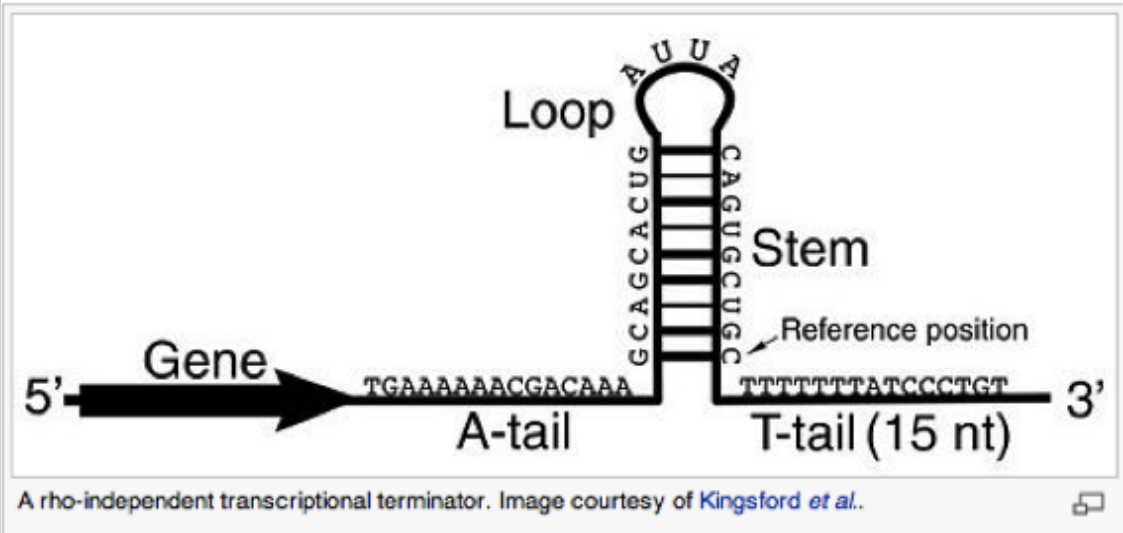
A typical RBS sequence is located about 6 nucleotides upstream of a start codon in an mRNA. The ribosomal holoenzyme binds to both the RBS and the start codon. The start codon and everything downstream are translated by the ribosome.

Protein Coding Regions

- Specify an amino acid sequence for a protein using the genetic code.
- Begin with a start codon, typically ATG to initiate translation.
- End with a stop codon, TAA, TAG, TGA to end translation.
- See `http://parts.igem.org/Protein_coding_sequences`

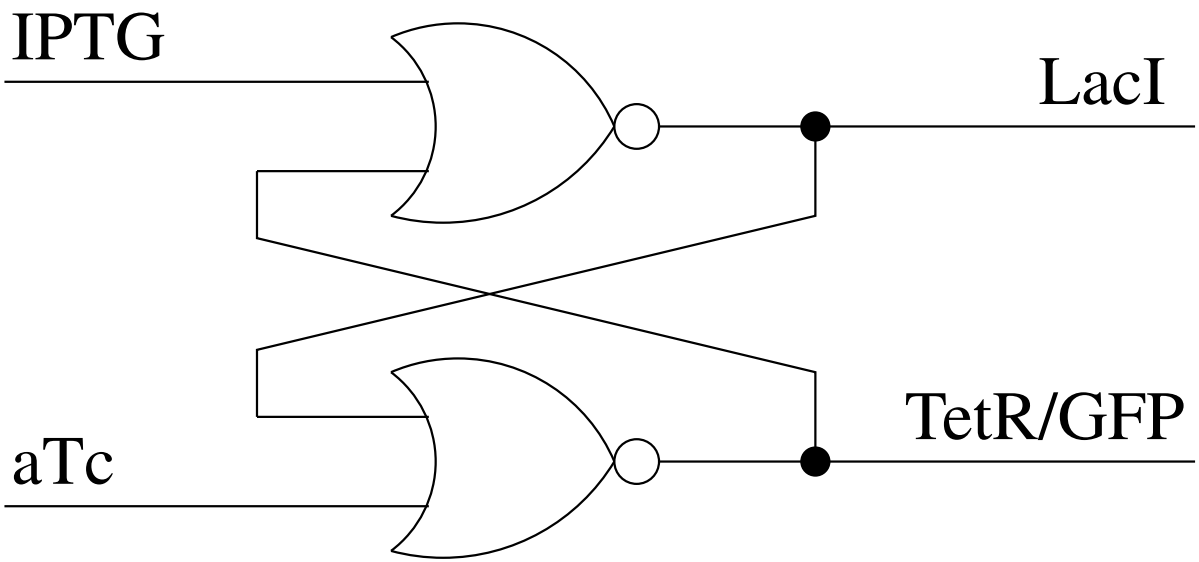
Terminators

- Sequence of DNA that signals the end of transcription.
- Placed past 3' end of the protein coding sequence.
- Often palindromic in sequence in prokaryotes.
- Weakens interaction between RNAP and the DNA strand being copied.
- Two categories in prokaryotes:
  - Rho dependent
  - Rho independent
- Rho independent most commonly used.
- See `http://parts.igem.org/Terminators`

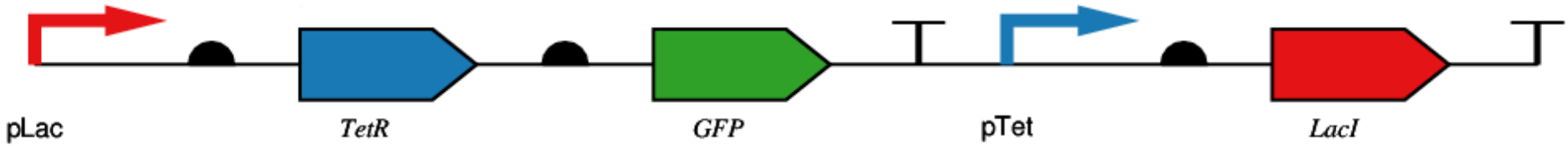


A rho-independent transcriptional terminator. Image courtesy of Kingsford *et al.*.

Genetic Toggle Switch (SR Latch) Logic Diagram



Genetic Toggle Switch (Gardner et al. 2000)



Assignment #1

1. Ensure that you understand genetic circuits to the level of detail presented in lecture 1.
2. Read the genetic toggle switch paper: `http://www.nature.com/nature/journal/v403/n6767/full/403339a0.html`
3. Read the paper that you selected for this course.
4. Download and examine the supplemental material for your paper.
5. Create an account on `https://synbiohub.utah.edu`.
6. Locate parts needed to construct a genetic toggle switch.
    1. Submit the parts into a collection in your private repository.
    2. Explain your choices in the details fields.
    3. Provide a share link to your private repository.
7. Locate parts needed to construct your paper's genetic circuit.
    1. Submit the parts into a collection in your private repository.
    2. Explain your choices in the details fields.
    3. Provide a share link to your private repository.