**BioGuide**

**Biological System Design Tool, Application Paper**

Muhammed Akif Ağca, Cihan Taştan, Hassan Salehe Matar, Burak Yılmaz, Yener Tunçel, Sibel Ataol, Tolga Can, Yeşim Aydin Son

**Abstract**

**Introduction**

As the field of Synthetic Biology is on the rise, iGEM is growing up very fast and the number of parts in the parts registry is increasing with the addition of more complex parts each day. Finding a part according to desired purposes is becoming more and more difficult. After facing some difficulty while running our algorithms on the parts registry and looking the parts for desired purposes, the need for more effective standardization of parts become more apparent. We have decided to investigate the information on parts in iGEM’s 2010 distribution and reorganize the information on the parts registry forms according to the needs of our algorithm and the needs of the other parts registry users. Furthermore, to provide alternative pathways to construct the most reliable and functional Biobrick devices we have used Graph Theoretic modeling to be able to refine and traverse the parts according to their input and output information by using BioGuide application.

**Results**

To overcome all of the difficulties and to enable faster BioBrick construction we have developed BioGuide application. The new standard is developed to apply graph theory on the parts according to needs of our algorithm and part registry users. The needs have been analyzed via a survey. The most reliable parts in parts registry are imported to our database. We have extracted the information from different sources and integrated in our database. Input output information of each part is extracted manually from parts registry web site. Then, a graph data structure is constructed to find possible input output relations between each part. Finally, an algorithm is developed to find possible pathways according to desired input and output information.

**Conclusion**


**Introduction**

As the field of Synthetic Biology is on the rise, iGEM is growing up very fast and the number of parts in the parts registry is increasing with the addition of more complex parts each day. After facing some difficulty while running our algorithms on the parts registry, the need for more effective standardization of parts entry was apparent. We have investigated the information on parts in iGEM’s 2010 distribution and reorganized the information on the parts registry forms according to the **needs of our algorithm**. Then we have used **graph theoretic modeling** to visualize the relations between iGEM Parts and to **standardize the representation** of the parts as much as possible by graph theoretical methods. This helped us to **find input output relations** between the parts. Furthermore, our program BioGuide is now able to **provide alternative pathways** to construct the **most reliable and functional Biobrick devices** with respect to given inputs and expected outputs as a guide to Biobricks parts registry.

The algorithm is working according to input output information of the parts. Each part is recorded to our database with the information specified in database section. The important information about the parts, which are revoked manually are; Promoter, Activity, Inducer, Activator, Repressor, Inhibitor for input. For output
information Reporter, Regulator, Inducer, Activator, Repressor, Inhibitor information is recorded. The details are explained in database part.

There are two graphs in BioGuide. The first one is possible combination of devices. The second one is possible combination of parts, which construct BioBrick devices. The graph constructed according to basics of graph theory each part is a node has an edge if a part’s output is input to another part. Graph data structure of devices constructed by the method. The combination rules are specified by image ID of each part. Each part type has an image ID possible combinations are specified by using the IDs.

The method has enabled us to find relations between all parts. By the method, we can search the parts to find pathways according to desired input and output information. The user can select inputs and outputs. The software is looking for all possible combinations, and highlighting shortest one to be constructed as a BioBrick device.

### Material

Our main data source for BIO Guide Software Program was the available background information of parts distributed in 2010 iGEM plates (Total of three 384-well plates of dried DNA) to the wetlab teams. This data was available through both the parts registry main website (http://partsregistry.org/Main_Page) in XML format and parts registry libraries (http://partsregistry.org/assembly/libraries.cgi?id=31) in Excel format. Data from parts with specific part IDs have been parsed with a custom code developed to modify SAX Parser. Then, the rest of the data which needs to be standardized according to biological importance have been extracted from the Registry of Parts Page manually. The chemical (IPTG, galactose etc.) or physical (UV irradiation, temperature etc) external inputs and proteins synthesized from a biobrick coding sequence can affect promoters on the parts. These effectors are identified under the title “Input”. And the “Output” s of these effectors are classified as inducers (a molecule that starts gene expression), repressors (blocker of attachment of RNA polymerase to promoter), activators (increasing the rate of transcription) and inhibitors (decreasing the rate of transcription). These standardizations on the database helped us to build the algorithm based on input/output relationships.

MySQL Server is used for Database development and organization. All of our illustrations for ER and algorithm is created in SmartDraw (trial version). Java Programming Language, and NetBeans Development environment is used or for software development. The graphical visualization of the software is done with Cytoscape and yfiles libraries (trial version) are used for the presentation of graphical events. We have utilized css Javascripts for our webdesign. Autodesk Maya 2011 with Academic Licence, Adobe Creative Suite 5 Master Collection (Trial Version) have been used for animations and illustrations. Video tutorial for the BioGUIDE has been created by camstudio and trial versions of Flash and After Effects are also used for the videos.

### Database Standardization

Two main focuses of our project was the organization of the available information about Biobricks on iGEM’s website and development of a software application to help synthetic biologists at the experimental set-up level
by providing all available construct combinations for any given input and output relations, which they can utilize for their own project.

Normalization and re-organization of the part information at iGEM’s website was needed in order to develop our application, which will automatically search the possible construct combinations. For the organization and analysis of the Biobricks, we used part info for Spring 2010 distribution. The information on all three 384 well plates distributed by iGEM scrutinized and checked individually to specify the standards available and needed. iGEM is providing so many parts within a hierarchical way, but there is no order in the information flow and no common standards. Furthermore, the information bulk is being used in an ineffective manner. Some of the parts distributed are known to be nonfunctional. Web pages for parts contain lots of information, but majority of them, are again not ordered. Moreover, some additional information had to be removed or replaced in such a way that the information for parts can be used effectively. And removal of the redundant bulk information related with parts at iGEM’s website had been recommended for future.

Although, the final standardization, which we have suggested is not for general public use and it was urgently needed in order to satisfy the needs of our algorithm. But, still it will be a valuable resource, since it summarizes the basic information about the parts.

As the first step to build the proposed standardization template, the headings selected related to parts are listed on Table 1. Submission of part IDs for individual parts is an accepted and quite valuable way of tracking information. Although, every part has unique partID, for every part there is a need to assign unique part names as official iGEM names. Part names will have an important role as they will be providing the short description about the part, which synthetic biologists can immediately recognize and utilize during the construction of unique Biobricks. Additionally unique part names will be helpful to identify the devices with more than one Biobrick in their constructs. Assignment of unique and distinct names for parts describing their nature and content will be helpful to researchers for the recognition of and search for the parts.

**Headings Selected From Previous Entry Forms for Indication of Standardized Information**

-----------------------------------------------

PartID:

PartName:

Bricks:

BrickIDs:

ImageIDs:

RFC10:

RFC21:
Table 1: The table above basically describes and designates qualities of parts which identifies their compositions and demonstrates the status of previously assigned standards. PartID refers to the unique ID number for parts including atomic parts and assemblies. PartName refers to the given unique names to parts. Bricks, refers to the shortcut names which specifies atomic parts. ImageIDs, refers to individual or combination of numbers that are assigned by us. RFCs refers to the states of parts based on RFC standards.

iGEM both provides individual, atomic parts and pre-combined constructs such as devices and systems. Availability of combined constructs is important to the researchers as combining individual bio-bricks one at a time will be very time consuming. These previously merged constructs, serve as the repository for puzzle and they can be used for different purposes. Up to date the largest and most trustworthy source, for synthetic biology and its components, is iGEM’s parts registry. In 2010, iGEM provided over 1000 parts that have initiated many projects. Having more atomic parts available in the iGEM’s repository, will lead to the design of more complex and robust constructs, and we would have a better chance to design different constructs for unique purposes. Also, for the parts that are already available, extra steps needs to be taken for the quality control and surveillance of these products. The quality control of the information for the parts is essential for the future of iGEM and synthetic biology. Even though we have found pre-determined RFC standards useful and included those to our standardized template, some individual parts still requires re-organization of the information as RFC standards alone for the functionality of parts, does not satisfy the needs for wet lab biologists.

Without a question there is an urgent need to build a distinct and specific database well organized with its own standards for synthetic biology; however, development of such a database is not an easy task.

**Contact Information of Part Owners and Qualitative Group Comments about Parts**

Designers: Mail:
GroupFavorite:
StarRating:
Parameters:

Table 2: The above table simply depicts information about possessors of parts and their contact information and the popularity of the parts for groups. Parameters heading, refers distinctive experimental details unique to the usage of parts which should be decided by groups.
Second step for building the standardized template was to get the phylogenetic information about the parts development process which includes the name of the group, designer and contact information, along with the comments from the group on the parts they have submitted. Contact information is especially important for iGEM as other groups who need extra information about the available part can reach to the required information. Even though contacting with the designers of the individual parts which are available is highly encouraged by iGEM, unavailability of contact information points out the fact that iGEM's parts registry needs strong re-organization in order to serve to the synthetic biology community properly.

Additionally, the “group favorite” and “starRating” fields are also important for individual evaluation of the parts, which doesn’t get the deserved attention from the iGEM groups. “Group Favorite” defines the confidence on the part by the designer group. “StarRating” defines the related part in terms of popularity and usage efficiency among the groups. According to our observations, most groups are not aware of either of the fields or they are used incorrectly or ineffectively. For example for a part with a full reporter which is known to be functional and gives precise and expected results the StarRating should be at least 2 stars, but for most of the parts in 2010 distribution, it is very difficult to observe a part whose “StarRating” is above one. For quick determination of functionality of the parts these two evaluations are important so they have been included in the proposed standardization template. But, as they were not properly used up to now for the re-organization of the parts information during the development of our software application we had to include all parts to our queries regardless of their evaluations based on “Group Favorites” and “StarRatings” 

Second step for building the standardized template was to get the phylogenetic information about the parts development process which includes the name of the group, designer and contact information, along with the comments from the group on the parts they have submitted. Contact information is especially important for iGEM as other groups who need extra information about the available part can reach to the required information. Even though contacting with the designers of the individual parts which are available is highly encouraged by iGEM, unavailability of contact information points out the fact that iGEM’s parts registry needs strong re-organization in order to serve to the synthetic biology community properly.

Additionally, the “group favorite” and “starRating” fields are also important for individual evaluation of the parts, which doesn’t get the deserved attention from the iGEM groups. “Group Favorite” defines the confidence on the part by the designer group. “StarRating” defines the related part in terms of popularity and usage efficiency among the groups. According to our observations, most groups are not aware of either of the fields or they are used incorrectly or ineffectively. For example for a part with a full reporter which is known to be functional and gives precise and expected results the StarRating should be at least 2 stars, but for most of the parts in 2010 distribution, it is very difficult to observe a part whose “StarRating” is above one. For quick determination of functionality of the parts these two evaluations are important so they have been included in the proposed standardization template. But, as they were not properly used up to now for the re-organization of the parts information during the development of our software application we had to include all parts to our queries regardless of their evaluations based on “Group Favorites” and “StarRatings”
Input and Output Characteristics of Parts

Parameters:
-Input:
  • Promoter:
  • Activity:
  • Inducer:
  • Activator:
  • Repressor:
  • Inhibitor:
-Output:
  • Reporter:
  • Reporter2:
  • Regulator:
  • Inducer:
  • Activator:
  • Repressor:
  • Inhibitor:
Table 3: The table above elaborately describes the input relations based on promoters and the output products based on the functional genes and RNAs which are included within the parts. Working condition simply describes any influencing factor or circumstance which is directly related with the functional properties of parts.

Third part of our standardization template includes parameters of contingent input and output elements. These parameters are classified into two groups for simplicity as presented on Table 3. This final part of the standardization template includes the upmost important information about the Biobricks that are required for the BioGuide Software to run its searching algorithm.

Briefly, BioGuide application is designed to catch the input and output relations of individual parts to examine possible Biobricks pathways for specific input and output queries. In other words, at pre-experimental stage, it helps wet lab biologists to design their unique constructs by revealing possible alternative options for pre-determined purposes, along with the primary paths. Our ultimate goal is to improve the algorithm designed for iGEM 2010 and present a new version of the BioGuide in iGEM 2011, which will provide optimum design of constructs for predetermined parameters.

Most of the parts are composed of functional and nonfunctional constructs which are formed by atomic parts. And every part should carry the information for all of its atomic parts within itself. The “input” heading actually stands for promoters. Parts with one or more promoters can be found at iGEM’s Parts Registry. Along with the information on which and how many promoters a part might have, the activity level of promoters are also important to distinguish between a constitutively active promoter or a promoter activated by specific physiological processes or states etc. This information was crucial for us to dissect in order to run our algorithm as it directly affects which inputs can activate the devices or the systems.

Throughout our investigations on the Parts Registry, we found out that much of the terminology was being used ambiguously. Although this might not be vital for synthetic biologists, it is still endeavoring to understand the function of certain regulatory elements which also becomes a time consuming task for the researcher. Thus, we recommend that the explanations of certain regulatory elements should be redefined and fixed especially for synthetic biology for easy communication, sharing and searching of information.

Common misuses of the terminology can guide us to figure out how to construct a standard nomenclature for synthetic biology. We claim that a standard nomenclature is urgently needed for synthetic biology for the following reasons. First of all, synthetic biology is an emerging research discipline and an industrial application area which is highly promising. Secondly, redefinition of the terminology to build a standard nomenclature is needed as some of the terms are prone to be used instead of another causing problems related to misuse for the global communication about synthetic biology. Lastly, the nomenclature has major importance for the construction of a persistent and trustworthy database for synthetic biology which serves for the information exhibition and exchange globally. For instance, there are obvious misunderstandings about the words which are
predominantly used for regulation process. We have noticed that, the terms “inhibitor” and “repressor” are being used as equivocally in the part information pages. Like the lactose inhibitor protein, a widely used DNA-binding transcriptional repressor, that have been labeled both as “inhibitor” and “repressor” at iGEM’s Parts Registry. Similar problems resulting from ambiguous use of terminology also observed with regulatory elements. To sum up, we investigated all input elements for promoters and classify these elements in terms of their function, affect and required input element for them. So, we suggest that terminology used for regulation of transcription should be defined clearly on iGEM’s website and correct use of terminology should be enforced.

The second group of parameters was collected under the title “Output”, which refers to products of functional genes. In contradiction, the term “reporter” has also been described within the same list. Reporters are also genes whose products, can be used for screening as an output. According to our group, the usage of the term “reporter” for genes is unnecessary and cause extra complexity for information distribution and gives rise to discrepancies. Instead of using the term “reporter”, predefined “gene” description should be used for genes, which can function as reporters. The special information which is related with the characteristic of that gene should also be presented on part info web page.

Furthermore, the same terminology “reporter” was used for both atomic parts and composite bio-bricks. Also the overall image descriptions for these were defined as “reporters”. We want to point out that using same nomenclature for both atomic genes and for whole functional constructs contributes to the complexity and makes specific explorations difficult through the Parts Registry. So, assigning “reporter” for both atomic parts and for whole constructs is not a good practice. Instead, we are suggesting the usage of other available terminology for the parts listed as reporters, which most of the constructs, now known as reporters, can be grouped into, such as “protein generators”, “composite parts” or “inverters”.

**Methods**

**Part Extraction Standards**

All information about the parts that are essential in experimental setup of iGEM projects has been utilized. The information for the parts available provided with all three 384 well plates in Spring 2010 distribution have been standardized. Our standardization criteria have been discussed in detail under Database Standardization. ER diagram has been generated which simply describes the organization of the data. Around 70% of the parts information has been fetched by the custom parsing code from XML and Excel files provided by iGEM. Rest of the data had to be collected and organized manually as the organization of these data cannot be standardized to generate an algorithm. This step was one of the most time consuming steps in our project. For each construct and Biobrick the information collected was; Activity, Inducer, Activator, Repressor and Inhibitor for promoters and Inducer, Activator, Repressor and Inhibitor information valid for synthesized molecules (mostly proteins and RNA fragments etc.)
**Combination**

Rules (Image Combinations) In order to build our input/output relations graphs first we run our algorithm on the real combination dataset which contains all few thousand different possible combinations of the biobricks. But after performing all combinations for the first few hundred biobricks application’s rate slowed down tremendously, which also become very time consuming for displaying biobricks graphs. To overcome this bottleneck we have developed a new strategy, where we have only used the construct combinations of the biobricks distributed within the plates. Moreover, according to information gathered from the subparts of the constructs distributed, we also collected the subpart assembly order, such as 1st: promoter, 2nd: rbs, 3rd: coding seq, any internal parts and the Last: terminator. Each specific Biobrick type has been assigned a number as a unique image ID from 1 to 19. Gathering the information on subparts was not a direct forward process. ImageID assembly orders for each construct have been used to extract the type information for each subpart with that construct. This innovative approach helped us to reveal 400 possible brick combinations present within the 3x384 well plates distributed by iGEM in Spring 2010.

**Algorithms**

In this section, the step by step functioning of our application, along with the encapsulation of the algorithmic concepts of ‘standardization’ of functional iGEM devices are depicted in pictorial forms called flowcharts. Rectangular boxes represent the encapsulation of implementations of the computer programs to perform the particular tasks stated in that box on the flowcharts. These boxes are sometimes called subprograms, objects or packages in Object Oriented software Engineering context. The diamonds represent decision branching and they are found between two rectangular boxes. The arrows show the direction in which subprograms work and communicate. The subprogram at the head of the arrow starts executing after the termination of the subprogram at the tail of the arrow. Following flowcharts are the high level representations of our algorithms developed for the BioGuide software.
Information about the iGEM parts had to be collected in a standardized format for our application to function properly. Following data collection custom subprograms is needed to parse and forward the data the application’s database. In order to achieve this we have designed and implemented the algorithm shown in diagram 1. In this algorithm, the first stage was to find the list of part IDs of devices which were supplied by iGEM in Spring 2010 distribution. This information has been collected from two sources 1) plate files in excel format which was available online 2) device data provided in xml format, both provided by iGEM. The last step in the algorithm was to send the collected partID data to the application’s database.
Diagram 2 presents the main algorithm, which shows how BioGuide application works. In BioGuide the major components are device and Biobrick graphs. While the device graph represents input-output (promoter-regulator) compatibility combination of iGEM devices, the Biobrick graph represents combinations of atomic parts assembled in a device or system. The flowchart shows how these graphs are created and embedded into the
program, which displays both of the graphs to the user when launched. Application presents few interactive options to the user when started, which were shown on the flowchart under the horizontal, bolded line. As shown on the diagram 2, there are four interactive tasks BioGuide can do, where the device and Biobricks graphs are utilized. Upon clicking a node on a devices or Biobricks graph, that node changes in size and color and the various functions shown on the flowchart can be performed then after.

**Graph Model**

**Graphical Modeling for Bio-Guide**

Graphical Modeling Theory has been applied to construct four different graphs where relations of atomic parts, devices and systems and the functional combinations that can build new constructs are presented for the iGEMs parts registry database. Three graphs are composed of iGEM devices and one graph is based on Biobricks. Each graph comprises a set of vertices or nodes and a set of edges. In the set of nodes each node represents a device, while in the set of edges each edge represents the input-output combination of the nodes. These graphs are directed graphs as the edges are created according to input-output combination. All compatibilities between a regulator and a promoter of an edge is created, where the source of this edge is the device with the corresponding regulator and target of the edge is the device with the promoter in concern.

![Fig. 1: A node representing a device](image1)

![Fig. 2: Arrow representing an edge between two nodes](image2)

The atomic structures used in our graphical model have been represented in Figures 1 and 2. A node is represented with a solid circle where the label, the part/device ID according to iGEM standards, of the device is marked on the foreground. The blue arrows between nodes connect the related devices, representing the input-output connectivity. End style of the arrow helps us to determine the direction of the node, like in Figure 2 where the node labeled BBa_S03520 is the source and BBa_JO9250 is the target.
**Directivity**

All the four constructed graphs build for BioGuide are directed graphs. So that, for every edge there must be a single source and a target. There is no single edge which is bidirectional. In mathematical form this can be represented as:

If an edge $e$ has node $v$ as source and node $w$ as target then the edge can be expressed as

$$e := (v, w)$$

For a directed graph the combination $(v, w)$ is totally different from $(w, v)$. Therefore,

$$e \neq (w, v)$$

The direction of the edges has been represented with the arrows, as explained in Figure 2.

**Connectivity**

The nodes forming their own sub-graphs disconnected from the rest of the nodes have been recognized, which showed us the presence of incompatibility between few regulators and promoters of the devices. We have observed this disconnection in all four of our graphs. The basis of the disconnection has been shown in Figure 3, where the two sub-graphs without any edge that connects them to the main graph has been presented on the right hand side of the diagram. These features classify our graphs as disconnected graphs [1].

![Fig. 3: A zoomed in screenshot showing two sub-graphs within the disconnected graph.](image)

"**Semi-Simplicity**"

A simple graph is a graph in which no more than one edge contains the same set of nodes. So, in a simple graph it is not possible to find more than one edge with the same source and the same target. Additionally, an edge with the same source and target, forming a loop is not allowed. But, in synthetic biology it is possible to construct a
device consisting of devices or bio bricks of the same species or type. Accordingly, our graphs are simple graphs with an exception of possible self-containing loops, where the edge starts from and ends on the same node. Our graphs have an exception of having loops and due to this permitted flexibility our graphs are "semi-simple".

For general information about graphs refer to:


Pathway Analysis

The user can select inputs and outputs from the given combo box or can type it manually. Pathway algorithm is searching the parts according to their input and output information. Dijkstra algorithm (http://en.wikipedia.org/wiki/Dijkstra%27s_algorithm) is being used to find shortest path. The algorithm is looking for all possible paths and the shortest one is highlighted to be constructed as a BioBrick device. Here are sample pathways;

First of the sample pathways could be constructed on reporter gene composites which consist of promoter, ribosome binding sites reporter coding sequence and double terminator. For this type composite searching, the promoter type is mostly chosen with inducer and the product is chosen from one of the reporter proteins. For example, arsenic arsenate or derivative of them induces specific promoter. As an input, the arsenic is chosen. Arsenic regulates the transcription of arsRp (arsenic promoter, BBa_K190015). This promoter is associated with dimer ArsR protein that is repressor of arsR promoter. Upon binding of arsenic, the dimer dissociates and allows RNA polymerase for transcription. For downstream of that promoter, from reporter subtitle of output combo box of program, lacZ protein (BBa_J33202) is chosen as reporter protein. This protein is classified as reporter for blue/white selection on Xgal plates. When the program is run, all possible pathways are subtracted. The option of the shortest pathway finds a device named as Bacillus subtilis ars promoter and arsR gene plus E. coli lacZ (BBa_J33206) for this example. This construct is designed with ribosome binding sites and designed to detect the transcription of arsenic dependent promoter by reporter protein obtained.

Second example on sample pathways could be chosen to show shortest pathway. IPTG is coded in input box and m RFP (BBa_E1010) is coded for output box, when right hand-sighted bar ‘show the parts’ are selected, any devices or systems which includes these parts were shown in device graph as dots. BBa_J04450 is one of the given results that consists of 3757bp and includes luxR gene and cl lam genes which is so long that hard to manipulate. However, in the bottom part, ‘shortest path’ option gives the simplest device which takes that input and gives the product as in output. BBa_J04450 is the simplest device for this example that includes lacI promter rbs, mRFP coding sequence and double terminator.

The BioGUIDE program provides us to choose any of options both for input and output such as inducer, activator repressor in toolbar. For chosen options, it gives not only parts and their brick codes but also combinations of devices and constructed devices. For example, input is chosen as an activator, temperature 42 C, and output is chosen as LucR protein. By the advances of this program, the part which has activation in 42 C is detected. RNA thermometer (ROSE) BBa_K115001 is used for temperature dependent post transcriptional regulation. The
translation initiates at 42C. For this example results, the shortest pathway with this input and output is found as Expression of luciferase with RNA-thermometer (BBa_K115020). This device consists of AraC promoter (BBa_R0080) with arabinose activator and RNA thermometer with 42C activator. As an output LucR protein expression is obtained. The expression of luciferase is measured by this device.

Results

BioGuide is a search tool for the parts in Parts Registry according to their input and output information and a pathway analysis tool according to desired input and output information. BioGuide uses graph data structure; there are 400 nodes and 800 edges in database. The most reliable parts in 2010 plate are used as row data.

By using BioGuide software program a synthetic biologist mainly can construct simple devices with desired input and output derivatives. Since the library of this program depends on distributed kit plates, the designed devices on program could be constructed and handled easily and in time saving manner. For the new researchers on synthetic biology, this program represents comprehensive part design and all required information in one screen. In this program each tab in part’s page as hard information and experience is concluded in one page.

A synthetic biologist could use this program to check and find out the parts or devices for transcriptional regulation and polymerase per second (PoPS) measurements. The devices which were ligated with reporter proteins such as green fluorescence protein could be used in promoter efficiency experiments. The expressed protein amount directly represents the transcriptional regulation.

To find the shortest pathway of desired devices in library is one of the opportunities of this program for researchers. The undesired downstream genes in devices could be eliminated that simplifies the job of researchers.

Discussion

As a synthetic biologist in team, while choosing parts to characterize or ligating with main coding parts we used this software tool, BioGuide. For this year’s project in IGEM 2011, we designed a kill switch device. This module totally consists of parts in 2011 kit plate distribution and in 2010 kit plates. The library of BioGuide program consists of these distributed bricks. Therefore while modeling the kill switch device we could performed this software program. One of the advantages of BioGuide is that when the input and output parameters are given, whole parts in kit plates were demonstrated on graphical diagram. This program provides us with all possible options to decide on. Second advantage of this program is that within selected parts program suggests a system which starts with input and gives the output as product. Even in these systems the shortest path could be obtained. When the appeared parts on left hand sided graphical diagram is double clicked on, the details of part are represented in one page long and the whole sequence could also be obtained for sequence analysis. Meanwhile the right hand sided graphical diagram shows the parts which compose that device. The initial part is in pink color and the final part is in yellow color. This property of program enhances the understanding of the
synthetic biologist who is new in this field and parts registry. The only hardening about this program is that it requires setup in computers. The webpage based programs provides more opportunities when compared with BioGuide. However the previous years’ software projects in IGEM 2011 were based on sequence analysis and restriction site mapping for molecular biology protocols. They are more interpretenable for people who are experts in this field. The BioGuide project is more comprehensive for beginners in this field and IGEM competition rather than other software tools.

**Supplementary Materials**

All materials are published in project’s official page: http://2010.igem.org/Team:METU_Turkey_Software

**User Guide**

User guide is a how to material to show how to use the BioGuide.

**Survey**

Survey is applied to analyze the needs of all parts registry users.

**Scope**

The partregistry.org is a continuously growing collection of standard genetic parts that can be mixed and matched to build synthetic biology devices and systems. The Registry is based on the principle of "get some, give some". Registry users benefit from using the parts and information available in the Registry for designing their own genetically engineered biological systems. In exchange, the expectation is that Registry users will contribute back to the information and the data on existing parts and will submit new parts they have designed in order to improve this community resource.

As an expanding database partregistry.org needs to be more organized and the standardization template needs to be improved. Additionally, the potential of multiple ways of using each part in different construct combination brings out the necessity for an application to search through the database. BioGuide is the first designed software that organizes over 1000 parts in partregistry.org as possible atomics parts to build new biological device and systems for specific input and outputs based on graph theory. The requirement of similar applications and software tools are now inevitable in the emerging field of synthetic biology. The innovative approach that makes the partregistry.org easy to use for synthetic biology applications is the collection of standardized parts that can be used in any combination with minimal effort under one database. But while working on our algorithm to search for possible combinations of parts depending on the given input and output, we have realized that present standards are inadequate and parts registry form must be improved.

In very near future a new format for parts registry form is needed and few additional features should be implemented to have more control on the database. We are planning to suggest a new format and features for the parts registry based on the survey results we have received. And planning to build the next version of BioGuide based on the revised parts registry form. Along with using new parts registry standards we will be improving the
algorithm, so that the software can search through more complex relations and returns all possible functional constructs.