

Proposal for iGEM parameter database

Nicholas Smart^{1*}, Annika Wipprecht,¹ Stuart Campbell,¹ Liam Barry¹ and Rory MacDonald¹

¹University of Aberdeen

Editor: Nicholas Smart

ABSTRACT

Outline: Having spoken with Randy Rettberg at the iGEM meeting in Edinburgh several efforts have been made to produce a persuasive outline for a parameter database to simultaneously serve biologists and mathematical modellers within and without the iGEM competition

Motivation: Having been involved in the project for several weeks it quickly became apparent that the tools of mathematical modelling relied heavily on parameters such as promoter strengths and dissociation constants. These parameters are not currently available in a readily accessible database, nor are they consistently reported in standardised units.

Goal: It would be our aim to instigate the creation of an accessible database within the iGEM project where, for example, promoter strengths could be compared, dissociation constants collected and reviewed and standard modelling practises taught and evaluated. This database could be added to by iGEM teams every year- in much the same way as the standard registry of parts is.

1 INTRODUCTION

The purpose of this effort is not design the database ourselves; we understand there is a great deal of computer science knowledge within the iGEM project already. Rather we would suggest what we would deem the ideal content to be and the potential benefits of such a database.

Our experience joining the iGEM as mathematical modellers is that the methods used by modellers in teams around the world are very similar; at least in terms of the forms of equation used, the parameters needed and the software tools utilised. Despite this demand for parameters and the scarcity of many of them within the literature there is still not an adequate database within the iGEM project to store the collected values. This means that it is quite possible for much of the collected knowledge that has been unearthed to be reburied at the end of each iGEM year. A poignant example to introduce the need for a database is the case of the ETHZ wiki. To summarise, ETHZ listed a table of referenced parameters for common biobricks in 2007. Two years later it is hard to do a specific Google search for parameters or find an iGEM 2008 wiki that does not reference the 2007 ETHZ page in its modelling section as supplying one or more of their parameters. Essentially this small table of 20 parameters cleaved from the stacks of biological papers has proved invaluable to many iGEM teams. Imagine the use of an entire database of constantly improving parameters which is under easy peer review.

2 PROPOSED ASPECTS TO INCLUDE IN THE DATABASE

We have discussed several possibilities for the setup of the database very thoroughly and the final solution we came to is outlined here. We would have parameters in two database sections: Promoters and Proteins. Both databases would be primarily search based as both promoters and proteins have specific names (biobrick part numbers in the case of promoters). Browsing would also be an alternative but subsections would not be involved as this could lead to multiple instances of database entries in different subsections.

We imagine here clicking the "Promoter" tab and opening the promoter database. It would be initially empty and need populating. Assume we know something about a constitutive promoter J23113 and want to populate the database with this promoter and then our information. First we click a button called "Add Promoter" and are prompted for a biobrick number (BBa_J23113) and a brief description (which can be amended later by anyone viewing the page). We then add links to the registry of parts pages which use this promoter (more can be added later) and the registry page for this specific promoter. We then accept this and the promoter is added to the list. We click it and enter the specific J23113 page. It is empty save the list of links to biobricks using it and our brief description of it.

Realising we have information on the production strength of this promoter in POPS we click "Add Information". Ten columns of data input are then created for us to use. They could look something like this:

Parameter	Description	Value	Units	Error
N/A	N/A	N/A	N/A	N/A
Extra?	Ref1	Ref2	Ref3	Notes
N/A	N/A	N/A	N/A	N/A

These headings are suggestive / instructive but can be changed as required. We then populate this table with our information, which happens to be a promoter strength in POPS. We leave the boxes that we do not require with N/A in them.

Parameter	Description	Value	Units	Error
Promoter strength	N/A	3.52×10^{-4}	POPS	
Extra?	Ref1	Ref2	Ref3	Notes
N/A	"paper name"	"paper name"	N/A	cross calculated from ref1 & ref2 (POPS is molecules/s)

We then click submit and view the first piece of information for JS23113 in the following way (with all columns with value N/A omitted):

*Correspondence should be addressed to: n.smart.05@aberdeen.ac.uk

Added by Nick Smart (Aberdeen University) 2009					
Parameter	Value	Units	Ref1	Ref2	Notes
Promoter strength	3.52×10^{-4}	POPS	<i>paper</i>	<i>paper</i>	Cross calculated from ref1 & ref2 (POPS is molecules/s)
Comment	Comments(2)	Rate	Rating (2.3)		

Obviously "paper" would be the actual references for each reference. There should also be a button for any iGEM member to comment on the parameter and a way to view these comments. Also a button to rate the parameter and the average rating displayed. We hope that the process described would yield the following benefits:

1. All information would be accurately referenced and experimental errors would be included where possible.
2. iGEM teams should be able to add information or comments to the database easily.
3. Almost certainly more than one figure for a specific parameter will be found in the literature for addition to the database so peer review should be allowed to rate each entry separately on a scale of 1 to 5. Commenting could be used to ask specific questions about the promoters or proteins.
4. iGEM team members would be able to add information in a way that is useful to them, there will be a suggestion of which columns to fill in but essentially there will be no limitation to the type of information that can be included
5. Unlike the boxes in the tables described here each data entry section would be large enough to accommodate a lot of information. This would allow extensive notes to be written on the conditions in which the experiment was undertaken etc.
6. No limit to the type and number of information pieces available. All information tables are specific to the type of information being conveyed so are not constricted by the table they are being inserted into.
7. All information pieces are separate from each other so can be rated and reviewed / removed separately.

It is important that the author of each piece of information is recorded and that the information can be "flagged as inaccurate" by any iGEM member.

We would also recommend a "notes box" where iGEM team members can insert information simply by supplying a "title" and a "comment". These comments could also be commented on and rated.

The protein database would run in a similar fashion but would be used to collect degradation constants / half lives, molecule sizes and functions *etc.* for all proteins (enzymes, repressors, GFP *etc.*)

3 ADDITIONAL EXTRAS

We would also suggest that store and search / browse for user uploaded information would be useful for modellers. An example of what this could be used for would be collecting template c / Matlab programs for, for example, random number generation, Monte Carlo simulations and parameter sensitivity analysis. It could also be used to upload written articles about common modelling techniques such as those above. Articles and programs could be rated and those which score highly could be kept in the database for another year. We believe this would help to create a more cohesive structure for systems biology modelling within the iGEM project running in parallel with the extension of the Standard Registry of Parts.

4 CONCLUSION

The idea outlined in this paper should be not only an invaluable tool for the modelling sides of the iGEM teams but also the next logical step for the iGEM project to move in. Having already produced a large range of biobricks which are readily accessible to the biologists to test and review it would be an exciting leap forward to be able to collect all this information in the form of an evolving database of standardised parameters so that modellers can dive right in and confidently use their time to advance their models rather than having first to search for papers which have been found by iGEMers scores of times before.