Model-guided combinatorial optimization of complex synthetic gene networks

Joerg Schreiber, Ms. Meret Arter, Mr. Nicolas Lapique, Benjamin Haefliger and Yaakov Benenson

Corresponding author: Yaakov Benenson, ETH Zurich

Review timeline:

Submission date: 16 September 2016
Editorial Decision: 19 October 2016
Revision received: 18 November 2016
Accepted: 25 November 2016

Editor: Maria Polychronidou

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 19 October 2016

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the two referees who agreed to evaluate your manuscript. As you will see below, the reviewers raise a number of concerns, which should be carefully addressed in a revision of the manuscript.

The reviewers' recommendations are rather clear so I think that there is no need to repeat the points listed below. In line with comment #3 of reviewer #2 we would ask you to make sure that the miRNA biosensor, its design, intended function etc. as well as the related modeling analyses are described in sufficient detail. The study should be readable/understandable as a standalone paper, without the need to refer to the Mohammadi et al study.

REFREEE REPORTS

Reviewer #1:

This paper shows that it is possible to use an approach that combines predictive modeling with diverse construction and screening to obtain synthetic gene networks that work in the way you like. This is an important goal for synthetic biology and the present paper does a solid job in addressing this goal. However the authors need to address the following issues before the work is suitable for publication:

1. There have been previous papers that combine predictive modeling with diverse construction and screening to obtain synthetic gene networks that work in the way you like; notably a 2009 Nature Biotechnology paper by Tom Ellis (Imperial College). The authors need to better acknowledge these earlier efforts and explain how their work is novel compared to these studies. Moreover the authors
should better acknowledge other, earlier computational-experimental studies in synthetic biology, ones that go back to the early days of the field.

2. It would be helpful in the abstract and other places in the paper to better highlight how the good and bad matches between model and experiments were used to derive meaningful mechanistic insights.

3. The Introduction contains many general statements and platitudes about modeling. It would be good if the authors could provide specific examples illustrating specific points and principles.

4. The paper quickly goes from generics and general statements to specifics, dealing with the authors' sensors. A better setup of the specific work that is to be presented would help the reader appreciate what is to come.

5. The authors should discuss how their approach can be extended to other synthetic gene networks. What challenges remain? What are the open questions still out there?

6. The authors should include a figure with a schematic workflow of their approach. It is quite difficult to appreciate how the modeling was coupled with the experiments from the present paper.

7. The schematics in Figure 1 should be expanded and enhanced to give a better sense of the functionality of the synthetic gene networks.

8. In the figures, the authors should more clearly indicate which results are from the modeling and which results are from the experiments.

Reviewer #2:

The authors apply mechanistic modeling to design a genetically encoded biosensor for specific microRNAs, and experimentally construct variants of the designed biosensor to validate the designs’ performances. They vary the number and positions of miRNA binding sites, protein-coding exons, and transcriptionally regulated promoters. From their characterization, they find that the developed model was able to explain some aspects of how the miRNA biosensor design affected its function.

Specific Comments

1. In the introduction, the authors mention that metabolic pathway engineering attempts to maximize only a single parameter (yield), but that is not true. For example, pathways are engineered to maximize the product titer (e.g. grams/liter) and organism productivity (grams/liter/hour) as well. These are distinct objectives as high yield pathways may not have high productivity (and vice versa).

2. In the introduction, the authors mention that the prediction of parameter values from DNA sequence is still poorly understood. This is an overly broad statement; it’s generally true when engineering genetic systems inside mammalian cells, but much less true when bacterial hosts are used. There are several models that enable one to predict the sequence-expression relationship inside bacteria. For example, the RBS Calculator model is able to predict a bacterial mRNA’s translation rate. The authors can further distinguish their approach by highlighting that there are much fewer examples of rational, model-based design when engineering genetic systems in mammalian cells.

3. On pages 3-4, there needs to be a much more complete description of the miRNA biosensor that what is currently presented. The authors should not assume that the reader will have read Mohammadi et. al prior to reading this manuscript, and therefore the authors need to describe and explain the composition of the miRNA biosensor circuit and its intended function. Currently, a typical reader will have no real understanding of how the biosensor works and the importance of the simulations & design.

4. The authors present a highly detailed comparison between their simulation results and measurements, particularly on the choices of where to position miRNA binding sites. The resulting analysis is very, very specific to the miRNA biosensor characterized here, and it remains uncertain if
there are general design principles that could be "take home" messages to the reader. The authors should highlight the general design principles that could be re-used when engineering a different miRNA biosensor or another entirely different genetic circuit in mammalian cells. The Discussion section does not convey any of this information as the authors have tried to off-load this essential component into a separate publication. Unfortunately, this significantly detracts from the quality of this manuscript.

1st Revision - authors' response 18 November 2016

We would like to thank the Reviewers for their positive assessment of our work, their insightful comments and useful advice. In the point-by-point response below we detail specific changes that we hope will address the concerns and clarify the impact of this study on model-guided circuit engineering and optimization.

Point-by-point response:

Reviewer #1:

This paper shows that it is possible to use an approach that combines predictive modeling with diverse construction and screening to obtain synthetic gene networks that work in the way you like. This is an important goal for synthetic biology and the present paper does a solid job in addressing this goal. However the authors need to address the following issues before the work is suitable for publication:

We thank the Reviewer for their positive assessment of our work and the constructive comments.

1. There have been previous papers that combine predictive modeling with diverse construction and screening to obtain synthetic gene networks that work in the way you like; notably a 2009 Nature Biotechnology paper by Tom Ellis (Imperial College). The authors need to better acknowledge these earlier efforts and explain how their work is novel compared to these studies. Moreover the authors should better acknowledge other, earlier computational-experimental studies in synthetic biology, ones that go back to the early days of the field.

Response 1.1 We thank the Reviewer for the comment and apologize for any omissions in covering the prior art. We have extensively rewritten the introduction to place our work in the context of prior work on modeling and optimization of synthetic circuits. We also acknowledge early works in synthetic biology that used modeling to guide their design efforts.

2. It would be helpful in the abstract and other places in the paper to better highlight how the good and bad matches between model and experiments were used to derive meaningful mechanistic insights.

Response 1.2. In this study, we found no substantial discrepancies between the model and the experiments, and thus our initial mechanistic understanding proved to be correct. Thus, good matches between the predictions and the data, e.g., the fact that improved miRNA knockdown increases the dynamic range, and many other matches referring to the expected trends in On and Off sensor values, are discussed in the relevant places in the manuscript. However we do emphasize in a few places in the manuscript that the data from high-throughput screens and detailed characterization can and should be used to modify the initial model in case of discrepancy.
3. The Introduction contains many general statements and platitudes about modeling. It would be good if the authors could provide specific examples illustrating specific points and principles.

Response 1.3 See also Response 1.1, we have extensively reworked the introduction to make it more concrete, with specific examples showcasing prior work in the field and highlighting the differences between published approaches and this study. We also introduced a Rationale section to explain our reasoning in developing the workflow discussed in the manuscript.

4. The paper quickly goes from generics and general statements to specifics, dealing with the authors' sensors. A better setup of the specific work that is to be presented would help the reader appreciate what is to come.

Response 1.4 We have attempted to smoothen this transition by better introducing the rationale behind our workflow (see above), as well as explaining how the sensors work and what their potential usage could be.

5. The authors should discuss how their approach can be extended to other synthetic gene networks. What challenges remain? What are the open questions still out there?

Response 1.5 Both the new “rationale” section and the discussion, as well as the new schematics (new Figure EV1, see point 6 below) are phrased to highlight the general applicability of our optimization strategy.

6. The authors should include a figure with a schematic workflow of their approach. It is quite difficult to appreciate how the modeling was coupled with the experiments from the present paper.

Response 1.6 Thank you for the suggestion, we have created such a scheme as a new Figure EV1.

7. The schematics in Figure 1 should be expanded and enhanced to give a better sense of the functionality of the synthetic gene networks.

Response 1.7 The schematic were extended to clearly show the two logical sensor states (on and off) in terms of component activity levels etc.

8. In the figures, the authors should more clearly indicate which results are from the modeling and which results are from the experiments.

Response 1.8 We have clarified this in the individual legends.

Reviewer #2:

The authors apply mechanistic modeling to design a genetically encoded biosensor for specific microRNAs, and experimentally construct variants of the designed biosensor to validate the designs’ performances. They vary the number and positions of miRNA binding sites, protein-coding exons, and transcriptionally regulated promoters. From their characterization, they find that the developed model was able to explain some aspects of how the miRNA biosensor design affected its function.

We thank the Reviewer for their positive assessment of our work and for the constructive comments.
Specific Comments

Comment 2.1. In the introduction, the authors mention that metabolic pathway engineering attempts to maximize only a single parameter (yield), but that is not true. For example, pathways are engineered to maximize the product titer (e.g. grams/liter) and organism productivity (grams/liter/hour) as well. These are distinct objectives as high yield pathways may not have high productivity (and vice versa).

Response 2.1: we thank the Reviewer for this comment; indeed this has been an omission on our side. The statement has been rectified and a reference was added pointing to those different optimization factors.

2. In the introduction, the authors mention that the prediction of parameter values from DNA sequence is still poorly understood. This is an overly broad statement; it's generally true when engineering genetic systems inside mammalian cells, but much less true when bacterial hosts are used. There are several models that enable one to predict the sequence-expression relationship inside bacteria. For example, the RBS Calculator model is able to predict a bacterial mRNA's translation rate. The authors can further distinguish their approach by highlighting that there are much fewer examples of rational, model-based design when engineering genetic systems in mammalian cells.

Response 2.2 We thank the Reviewer for the suggestion; we have modified the introduction to better reflect the state of the art in predicting function from sequence. We distinguish between de novo predictions and those based on large dataset combined with machine learning.

3. On pages 3-4, there needs to be a much more complete description of the miRNA biosensor that what is currently presented. The authors should not assume that the reader will have read Mohammadi et. al. prior to reading this manuscript, and therefore the authors need to describe and explain the composition of the miRNA biosensor circuit and its intended function. Currently, a typical reader will have no real understanding of how the biosensor works and the importance of the simulations & design.

Response 2.3 We have added an explanation about the potential utilization of miRNA classifier networks, and added a description of sensor mechanism of operation. Figure 1A was also expanded to illustrate the configuration of different sensor states.

4. The authors present a highly detailed comparison between their simulation results and measurements, particularly on the choices of where to position miRNA binding sites. The resulting analysis is very, very specific to the miRNA biosensor characterized here, and it remains uncertain if there are general design principles that could be "take home" messages to the reader. The authors should highlight the general design principles that could be re-used when engineering a different miRNA biosensor or another entirely different genetic circuit in mammalian cells. The Discussion section does not convey any of this information as the authors have tried to off-load this essential component into a separate publication. Unfortunately, this significantly detracts from the quality of this manuscript.

Response 2.4 We thank the Reviewer for raising this point. This "take home message" is an important component of the study. First, we note that every circuit family has its own unique features and therefore a model of a given circuit, and the conclusions thereof, might be only applicable to the specific family under investigation or closely-related families. In our case, the circuit family is a particular double inversion topology (See Fig. 1A top panel, on the right). Regardless of what the inputs and the intermediate components are, the repression steps must be very efficient and triggered by low amounts of either the input or the repressor; at the same
time the induction should operate far from saturation. This is the immediate take home message of this study. The broad importance of the work is to demonstrate the workflow that can be applicable to diverse circuit families: introducing a performance metric and evaluating multiple parameter combinations on a grid, and uncovering trends, correlations and cross-correlations to reveal optimal parameter regimes. Next, in the experimental part, designing a library of diverse components that span the parameter phase space and combinatorial screening of multiple circuit variants in an effort to reach the favorable parameter regime and either validate or adjust the original model. The manuscript has been modified to convey these ideas in the Introduction, the Rationale and the Discussion sections, and we hope the general applicability of our study comes across much better now.

2nd Editorial Decision

25 November 2016

Thank you for sending us your revised manuscript. We are satisfied with the modifications made and I am pleased to inform you that your paper has been accepted for publication in Molecular Systems Biology.
EMBO PRESS

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND

Corresponding Author Name: Yakov Benenson
Journal Submitted to: NMR
Manuscript Number: NMR-16-725SR

Reporting Checklist for Life Sciences Articles (Rev. July 2015)

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. These guidelines are consistent with the Principles and Guidelines for Reporting Preclinical Research issued by the NIH in 2004. Please follow the journal’s authorship guidelines in preparing your manuscript.

A- Figures

1. Data

The data shown in figures should satisfy the following conditions:

- the data were obtained and processed according to the field’s best practice and are presented to reflect the results of the experiments in an accurate and unbiased manner.
- figure panels include only data points, measurements or observations that can be compared to each other in a scientifically meaningful way.
- graphs include clearly labeled error bars for independent experiments and sample sizes. Unless justified, error bars should not be shown for technical replicates.
- if n < 5, the individual data points from each experiment should be plotted and any statistical test employed should be justified.
- Source Data should be included to report the data underlying graphs. Please follow the guidelines set out in the authorship guidelines on Data Presentation.

2. Captions

Each figure caption should contain the following information, for each panel where they are relevant:

- a specification of the experimental system investigated (e.g., cell line, species name).
- a description of the methodology (e.g., transfections performed in both datasets. The criteria were not pre-established).
- an explicit mention of the biological and chemical entities that are being measured.
- an explicit mention of the biological and chemical entities that are altered/modified/ manipulated in a controlled manner.
- the exact sample size (n) for each experimental group/condition, given as a number, not a range;
- a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, littermates, cultures, etc.).
- a statement of how many times the experiment was independently replicated in the laboratory.
- definitions of statistical methods and measures:
  - common tests, such as z test (please specify whether paired vs. unpaired), simple χ2 tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - are tests one-sided or two-sided?
  - are there adjustments for multiple comparisons?
  - exact statistical test results, e.g., P-value ≥ a but not P value ≥ c;
  - definition of “center values” as median or average;
  - definition of error bars as d.s. or s.e.m.

Any descriptions too long for the figure legend should be included in the methods section and/or with the source data.

Please ensure that the answers to the following questions are reported in the manuscript itself. We encourage you to include a specific subsection in the methods section for statistics, magnets, animal models and human subjects.

B- Statistics and general methods

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?</td>
<td>We performed all experiments in biologic and independent replicates.</td>
</tr>
<tr>
<td>1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.</td>
<td>No, our experiments do not encompass animal studies.</td>
</tr>
<tr>
<td>2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?</td>
<td>We excluded the final data series of the robotic experiments for technical reasons. We repeated the experiment and normalized the data between the experimental runs using control transitions performed in both datasets. The criteria were not pre-established.</td>
</tr>
<tr>
<td>2. Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. randomization procedure)? If yes, please describe.</td>
<td>Randomization was not relevant in this project as we were not looking for statistical significance.</td>
</tr>
<tr>
<td>3. For animal studies, include a statement about randomization even if no randomization was used.</td>
<td>NA.</td>
</tr>
<tr>
<td>4.a. Were any steps taken to minimize the effects of subjective bias during group allocation or/and when pooling results (e.g., binding of the investigator)? If you please describe.</td>
<td>The data analysis of the high-throughput screening experiment was performed by a different scientist than the person in charge of the experiment itself.</td>
</tr>
<tr>
<td>4.b. For animal studies, include a statement about blinding even if no blinding was done</td>
<td>NA.</td>
</tr>
<tr>
<td>5. For every figure, are statistical tests justified as appropriate?</td>
<td>Yes, in all figures we report the mean and standard deviations.</td>
</tr>
<tr>
<td>6. Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to check it.</td>
<td>This is not directly applicable in the current study as we were not seeking statistical significance per se.</td>
</tr>
<tr>
<td>7. Is there an estimate of variation within each group of data?</td>
<td>NA.</td>
</tr>
<tr>
<td>8. Is there variance similar between the groups that are being statistically compared?</td>
<td>We did not perform statistical comparison between groups.</td>
</tr>
</tbody>
</table>

C- Reagents

Please fill out these boxes (Do not worry if you cannot see all your text once you press return)
### D. Animal Models

1. Report species, strain, gender, age of strain and genetic modification status where applicable. Please detail housing and husbandry conditions and the source of animals.

2. For experiments involving live organisms, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.

3. We recommend consulting the ARRIVE guidance (see link list at top right) to ensure that other relevant aspects of animal studies are adequately reported. See author guidelines, under ‘Reporting Guidelines’.

### E. Human Subjects

1. Identify the committee(s) approving the study protocol.

2. Include a statement confirming that informed consent was obtained from all subjects and that the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services. See author guidelines, under ‘Reporting Guidelines’. Please confirm you have submitted this list.

3. For tumor marker prognostic studies, we recommend that you follow the REMARK reporting guidelines (see link list at top right), see author guidelines, under ‘Reporting Guidelines’. Please confirm you have followed these guidelines.

### F. Data Accessibility

1. Provide accession codes for deposited data. See author guidelines, under ‘Data Deposition’.

2. Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Microscopic structures c. Crystallographic data for small molecules d. Functional genomics data e. Proteomics and molecular interactions

3. We provide numerical data corresponding to the experimental figures in the article.

4. As far as possible, primary and referenced data should be formally cited in a Data Accessibility section. Please state whether you have included this section.

Examples:

**Primary Data**

- Protein Data Bank (PDB)
- Protein DataBank (PDB) (see link list at top right)
- PROoki (see link list at top right)
- PDB (see link list at top right)
- Protein DataBank (PDB) (see link list at top right)
- PROoki (see link list at top right)

**Referenced Data**

- Protein DataBank (PDB) (see link list at top right)
- PROoki (see link list at top right)
- PDB (see link list at top right)

### G. Dual use research of concern

1. Could your study fall under dual use research? If yes, please provide appropriate reference(s). See author guidelines, under ‘Reporting Guidelines’.

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can your study fall under dual use research?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* For all hyperlinks, please see the table at the top right of the document.