**Федеральное государственное бюджетное образовательное учреждение высшего образования**

**«Московский государственный университет имени М.В. Ломоносова»**

**Факультет фундаментальной физико-химической инженерии**

**Биологический факультет, кафедра биоинженерии**

# КУРСОВАЯ РАБОТА

Математическое моделирование кинетики реакций в генетических сетях, основанных на репрессии транскрипции с помощью белков.

## Выполнил(а) студент(ка)

2 курса 202 группы

## Винников Ренат Сергеевич

(подпись студента)

## Научный руководитель д.ф-м.н., чл.-корр. РАН, доцент

Шайтан Алексей Константинович

(подпись руководителя)

# Москва 2023

**Оглавление**

[**Введение** 2](#_Toc135600093)

[**Материалы и методы** 5](#_Toc135600094)

[**Результаты и обсуждение** 6](#_Toc135600095)

[**Заключение** 14](#_Toc135600096)

[**Сопроводительные материалы** 14](#_Toc135600097)

[**Список литературы** 15](#_Toc135600098)

# **Введение**

В основе существования любой клетки лежит экспрессия её генетического материала. Многие гены связаны между собой в сложные системы, где одни могут управлять экспрессией других. Такие сети управляют множеством процессов, таких как рост, размножение, дифференциация и адаптация к окружающей среде.

Помимо важности понимания принципов функционирования естественных генетических сетей (схем) для изучения живых организмов, в последние 20 лет получила активное развитие синтетическая биология - область науки, которая занимается созданием новых живых систем или изменением существующих с помощью инженерных методов. Она включает в себя создание искусственных генетических схем, клеток и организмов, а также изменение существующих генетических элементов для улучшения их функциональности. Синтетическая биология использует инженерный подход к биологическим системам, технологии и методы из различных областей, таких как генетика, биохимия, физика и математика, для создания новых биологических систем с определенными свойствами и функциями. Она имеет потенциал для решения многих проблем, таких как создание биосенсоров, борьба с болезнями, синтез новых источников энергии, и улучшение производства пищевых продуктов [1,2]. Ещё одним важным направлением синтетической биологии является создание аналогов электронных логических схем – это направление открывает перспективы создания биокомпьютеров, которые также могут использоваться в различных областях, от медицины до промышленности [3].

В 2000 году в Nature были опубликованы две статьи, заложившие основу в принципы дизайна искуственных генетических схем. В статье М. Эловитца и С. Ляйблера описано создание репрессилятора – биологического осцилятора, биологических клеточных часов, созданных человеком и основанных на трех белках-репрессорах [4]. Статья Т. Гарднера, Ч. Кантора и Д. Коллинза описывает создание переключателя – системы, в которой с помощью индукции ИПТГ активируется экспрессия GFP – флуоресцентного репортерного белка (состояние «ВКЛ»). В свою очередь удаление ИПТГ из системы (вымыванием) приводит к снижению экспрессии репортерного белка. Эта схема основана на двух белках-репрессорах [5]. В этих работах были продемонстрированы принципы создания колебательных и мультистабильных (имеющих несколько равновесных состояний) систем – в том числе, необходимость кооперативного связывания белками-репрессорами операторов (коэффициент Хилла n > 1) и снижения текучести промоторов, сопоставимость скоростей распада белков и мРНК, использование сильных промоторов и эффективных сайтов связывания рибосом.

В последнее десятилетие в дизайне генетических сетей начали активно использовать белки Cas (Cas9, Cas12a), измененные так, чтобы убрать их нуклеазную активность – такие варианты Cas белков называются dCas. Они получили широкое распространение благодаря их специфичности – в комплексе с гидовой РНК dCas можно направить на любой участок генома (можно синтезировать РНК с любой последовательностью), рядом с которым находится короткий мотив из 2-4 нуклеотидов, называемый PAM (Protospacer Adjacent Motif) [6]. dCas-белки имеют определенные недостатки, затрудняющие их использование для дизайна генетических сетей – они токсичны для клетки, что связано с их неспецифичным связыванием мотивов PAM; их связывание с оператором некооперативно (n ≈ 1.0); Комплекс dCas с ДНК медленно диссоциирует в отсутствие репликации в клетке, что замедляет динамику систем, основанных на CRISPR; Соответственно, вносятся различные изменения, направленные на улучшение свойств регуляторного белка – к dCas присоединяют дополнительные репрессорные или активаторные домены, что позволяет увеличить кооперативность их связывания с ДНК (например, dCas9-PhlF имеет n = 1.6 для этой реакции), вносят мутации, направленные на потерю офф-таргетной активности для снижения токсичности, ускоряют деградацию белков dCas и т.д. [7,8]. В то же время, стоит заметить, что некоторые сложные системы, основанные на Cas-белках, в том числе репрессилятор (CRISPRlator) и переключатель были успешно созданы в клетках без избавления от неспецифичного связывания dCas9 с ДНК, напротив, математическое моделирование системы показало, что именно неспецифическое связывание в отсутствии кооперативности позволяет создавать динамические или мультистабильные системы [9].

На примере этих статей видна тенденция авторов к использованию кинетического моделирования. Это лишь один из подходов к математическому моделированию генетических сетей, однако он позволяет отслеживать динамику системы, в том числе, в явном виде наблюдать изменение концентраций различных компонентов [10]. Во многом этот подход основан на численном решении дифференциальных уравнений, описывающих изменения концентраций компонентов системы. Существуют различные программы для кинетического моделирования, поддерживающие, в том числе, импорт моделей, написанных на SBML – языке разметки системной биологии, который позволяет не только описать систему с точки зрения компартментов, параметров, веществ и реакций, но и задать математические выражения (уравнения), устанавливающие связь между компонентами, что делает модели, описанные на SBML, удобными для моделирования в разных программах [11]. Существует база данных BioModels, в которой хранятся различные кинетические модели биохимических и клеточных систем, в том числе в формате SBML, что позволяет импортировать эти модели для симуляции в различные среды [12]. К числу таких сред для моделирования относятся CellDesigner, DBSolve Optimum, COPASI, MATLAB SimBiology и др. [13-15]. Однако, следует заметить, что часть из этих программ получала последнее обновление несколько лет назад и не поддерживает новый уровень и версию формата SBML (level 3 version 2). Упомянутые программы, кроме последней, не позволяют импортировать код на каком-либо языке программирования, что делает их менее гибкими (COPASI позволяет экспортировать уравнения, описывающие динамику системы в виде кода на C). MATLAB SimBiology является коммерческим продуктом, что лимитирует его использование.

По этим причинам, мы сформулировали следующие цели работы:

1. Разработка вычислительных подходов к исследованию кинетики функционирования генетических сетей на основе решения дифференциальных уравнений с использованием языка программирования Python.
2. Изучение режимов работы элементарных генетических сетей на основе белков-репрессоров с различной кооперативностью с помощью детерминистического кинетического моделирования.
3. Представление полученных методов в удобном для использования виде.

Для достижения целей были поставлены следующие задачи:

1. Задать математическую модель генетической сети, составить систему уравнений, описывающую её динамику, определить диапазоны изменения параметров, соответствующих реальным системам.
2. Исследовать различные методы численного моделирования, создать программу, позволяющие в интерактивном режиме осуществлять численное решение систем дифференциальных уравнений, описывающих функционирование выбранных генетических схем.
3. Изучить режимы работы и особенности функционирования генетических схем на основе разработанных моделей.

# **Материалы и методы**

В качестве исследуемых моделей были выбраны репрессилятор [4], интерактивный график изменения концентраций белков-репрессоров в котором был получен в ходе предыдущей курсовой работы, и логический элемемент «НЕ», где в качестве репрессора выступает комплекс белка dCas9 c гидовой РНК.

Уравнения, описывающие динамику системы репрессилятора были взяты из соответствующей статьи. Уравнения динамики логического элемента «НЕ» были выведены из основных уравнений химической кинетики и экспрессии генов [16].

Численные значения параметров для соответствующих систем были взяты из баз данных BioModels [12] и BioNumbers [17].

Для численного моделирования использовался язык программирования Python 3.8.5 с библиотеками numpy (1.24.2), scipy (1.5.2). Численное интегрирование было проведено методом Рунге-Кутта порядка 4(5), c помощью команды scipy.integrate.solve\_ivp(method='RK45'). Для нахождения равновесного состояния система ДУ была решена гибридным методом Пауэлла, командой scipy.optimize.root().

Для визуализации использовалась библиотека bqplot (0.12.33). Интерактивные виджеты создавались с помощью ipywidgets (7.6.5).

Для представления готовых графиков использовался сайт на базе JupyterLite (GitHub Pages).

# **Результаты и обсуждение**

В прошлом году мы создали модель репрессилятора, представленную в виде интерактивного графика – пользователь мог ввести новое значение параметра в соответствующее ему текстовое поле, что приводило к перестройке графика. Эта модель запускалась локально, из файла. В наши планы на будущее входило создание веб-приложения на основе файла Jupyter Notebook с кодом, чтобы пользователи могли работать с нашей моделью непосредственно в браузере. Для этого мы использовали JupyterLite – среду Jupyter, написанную на WebAssembly, что позволяет программе запускаться и производить вычисления в браузере. Для этого нам пришлось отказаться от визуализации с помощью библиотеки Bokeh (из-за несовместимости ее с JupyterLite) в пользу bqblot и ipywidgets. Таким образом мы смогли перенести интерактивную модель репрессилятора в веб-приложение (рис.1).

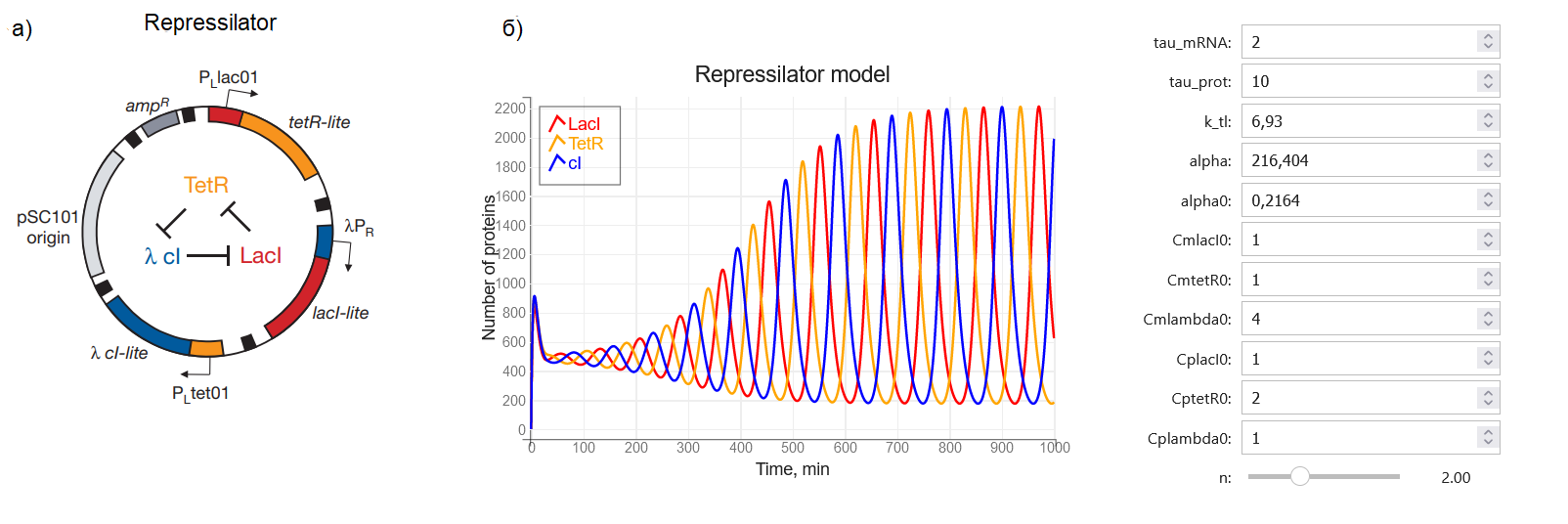
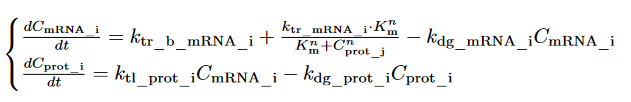
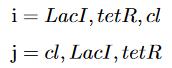


Рис.1. а) Схема репрессилятора, взята из соответствующей статьи [4]. б) Скриншот интерактивного графика зависимости количества молекул белков-репрессоров от времени, представленного на сайте на базе Github Pages (см. раздел «Сопроводительные материалы»).

Мы расширили функционал этого приложения, добавив возможность проанализировать, как изменение конкретных параметров (коэффициента Хилла (n) и времен полураспада белка и мРНК (tau\_mRNA, tau\_prot), упомянутых в статье [4] как особенно важных для возникновения осцилляций, изменяет динамику системы. Для этого мы добавили возможность расчета равновесного состояния и построения фазовых портретов системы (зависимость концентрации (в молекулах на объем клетки) белка i от концентрации белка j, i **≠** j). Приведем уравнения, описывающие динамику репрессилятора:





Заметим, что запись выше обозначает три системы уравнений, каждая из которых описывает изменение концентрации мРНК и молекул репрессора. ktr – константы скорости транскрипции (ktr\_b – базальная скорость транскрипции), kdg – константы скорости деградации, ktl – константа скорости трансляции, Km – константа Михаэлиса, n – коэффициент Хилла. Индексы i и j, как было описано выше, указывают на конкретный белок-репрессор.

Фазовый портрет данной системы при начальном значении параметров представляет из себя предельный цикл – неустойчивое стационарное состояние. Можно заметить, что изменение коэффициента Хилла в меньшую сторону (снижение кооперативности связывания репрессора с ДНК), при прочих фиксированных параметрах, приводит к стабилизации равновесия и, следовательно, затуханию осцилляций. Также к затуханию осцилляций приводит увеличение полужизни мРНК, а уменьшение полужизни белка или увеличение коэффициента Хилла увеличивает амплитуду осцилляций (рис.2).

Вторая система представляет собой логический элемент «НЕ» на основе комплекса dCas9 и гидовой РНК (sgRNA). Схема этой системы и ее интерактивный график представлены на рис.3. Приведем уравнения, описывающие динамику этой системы:

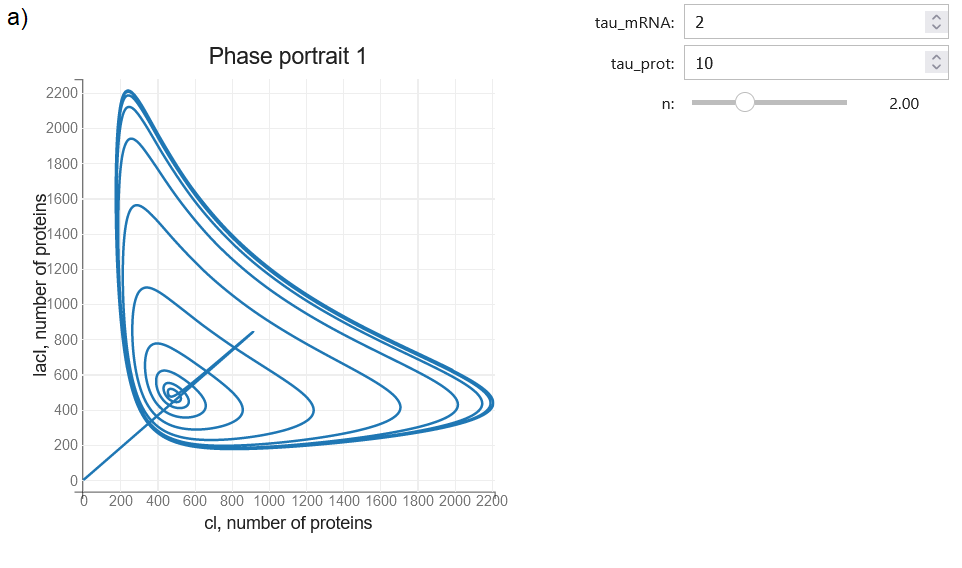


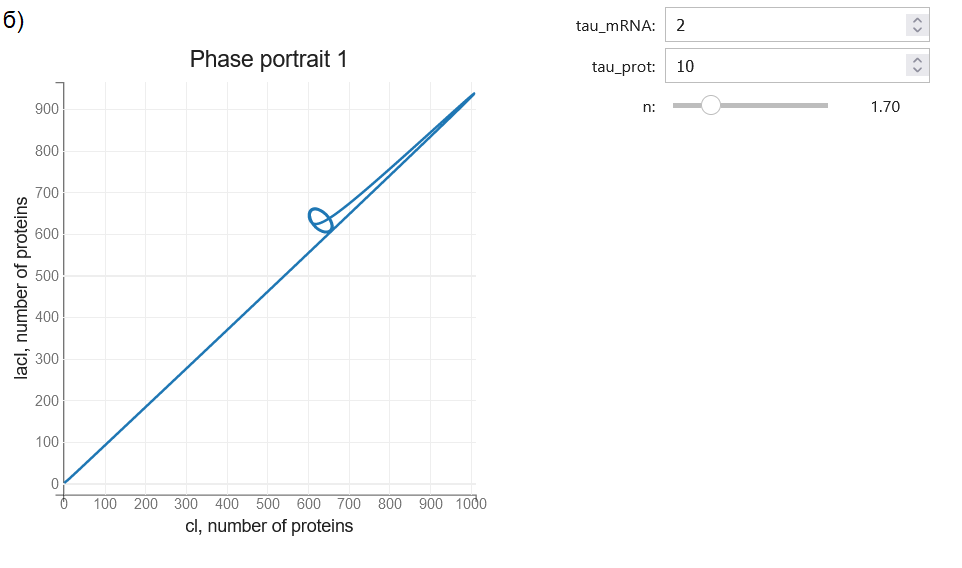






Уравнение 1) описывает изменение концентрации (в молекулах на объем клетки) мРНК dCas9. Уравнение 2) – изменение концентрации свободного белка dCas9. Уравнение 3) отвечает за формирование, распад и деградацию комплекса dCas9:гРНК. Уравнение 4) описывает изменение концентрации мРНК GFP. Уравнение 5) – изменение концентрации белка GFP. Поговорим о параметрах:



****

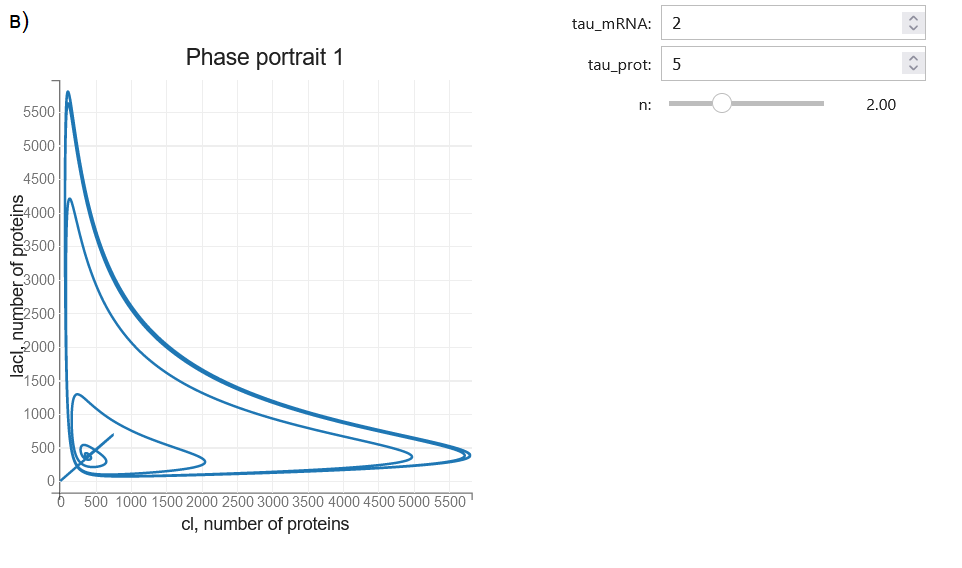
****

Рис. 2. а) Фазовый портрет системы (Ось x – количество молекул белка cl, ось y – количество молекул белка LacI) при начальном значении всех параметров. б) Фазовый портрет системы при понижении значения коэффициента Хилла (n) до 1.7. в) Фазовый портрет системы при понижении значения полужизни белка (tau\_prot) до 5 минут.

ktr, kdg, ktl,Km,n– обозначения аналогичны указанным выше для репрессилятора. kd – константа скорости распада комплекса, kf – константа скорости образования комплекса. В качестве упрощения, в рамках модели мы принимаем концентрацию гРНК постоянной.

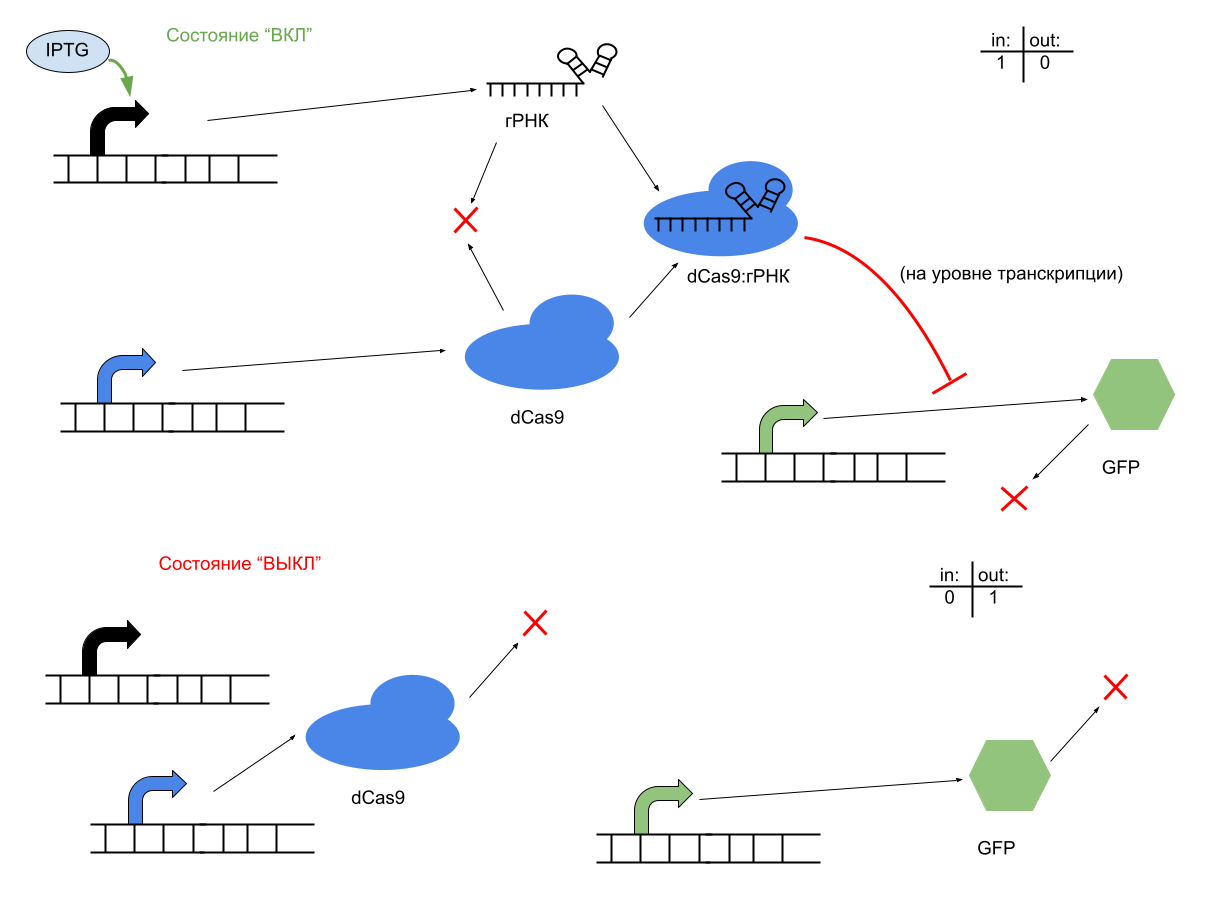
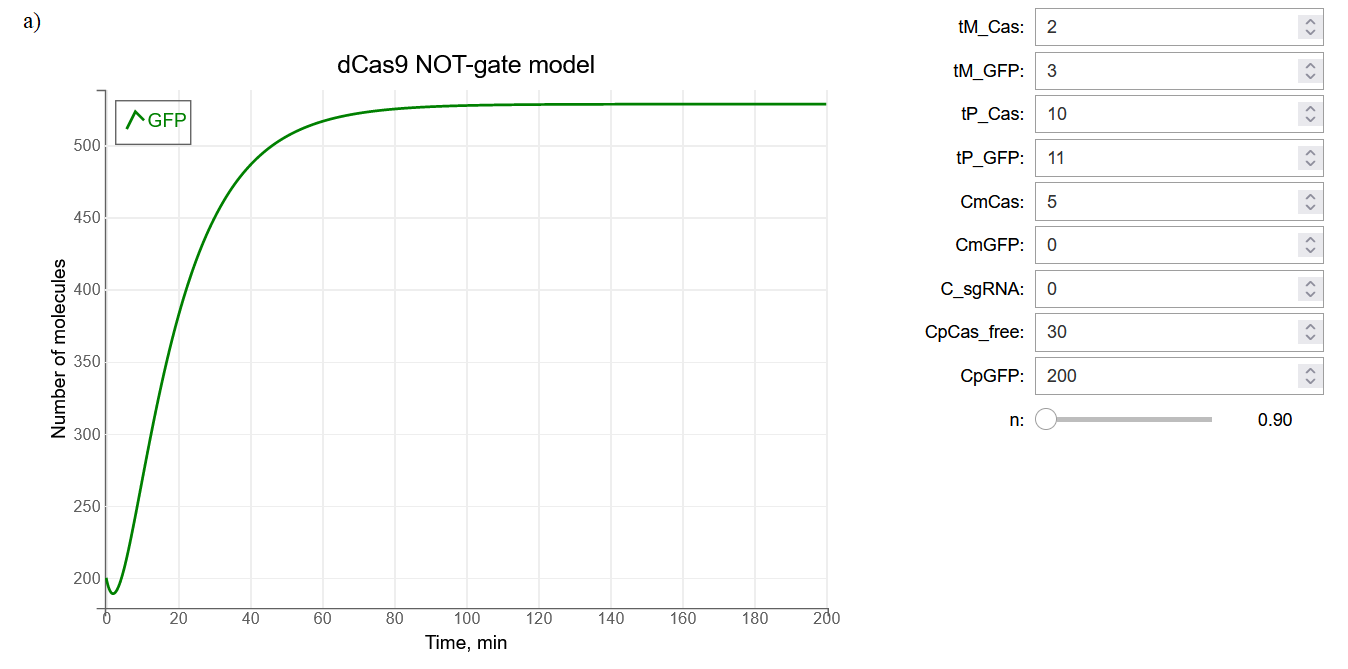
****

Рис. 3. Графическая схема логического элемента «НЕ» на основе комплекса dCas9 с гРНК. Для упрощения рисунка опущен этап транскрипции, но он учитывается в модели. Красный крест обозначает деградацию белков/мРНК.

Для изучения режимов работы данного комплекса были построены график зависимости концентрации GFP от времени и фазовый портрет (зависимость концентрации GFP от концентрации комплекса dCas9:гРНК) (рис.4). Наша модель показывает, что для успешной работы схемы в режиме «ВЫКЛ» следует увеличивать концентрацию гидовой РНК, уменьшать время жизни мРНК GFP и самих молекул белка, увеличивать время жизни dCas9 и его мРНК, уменьшать Km (данная константа равна количеству молекул репрессора, нужных для снижения скорости транскрипции с промотора в 2 раза [4]) и увеличивать коэффициент Хилла – то есть делать GFP менее «долгоживущим», репрессор – напротив, более стабильным и усиливать связывание промотора репрессором.



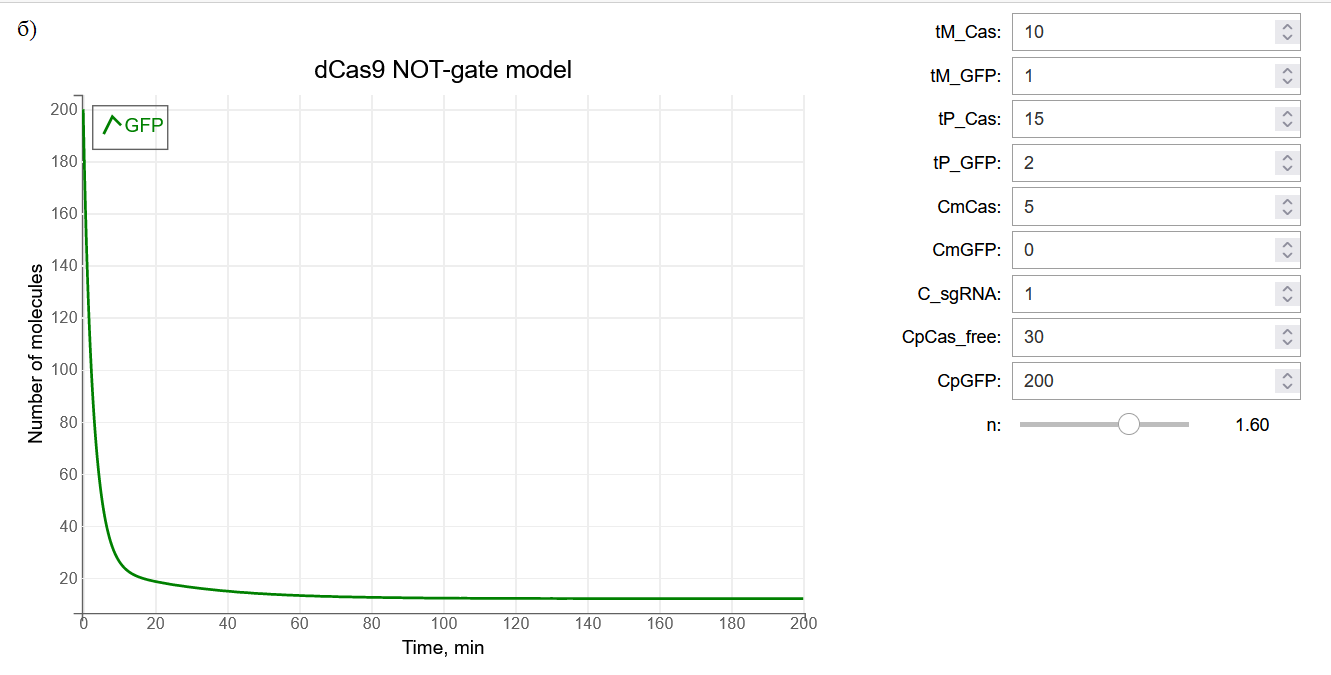
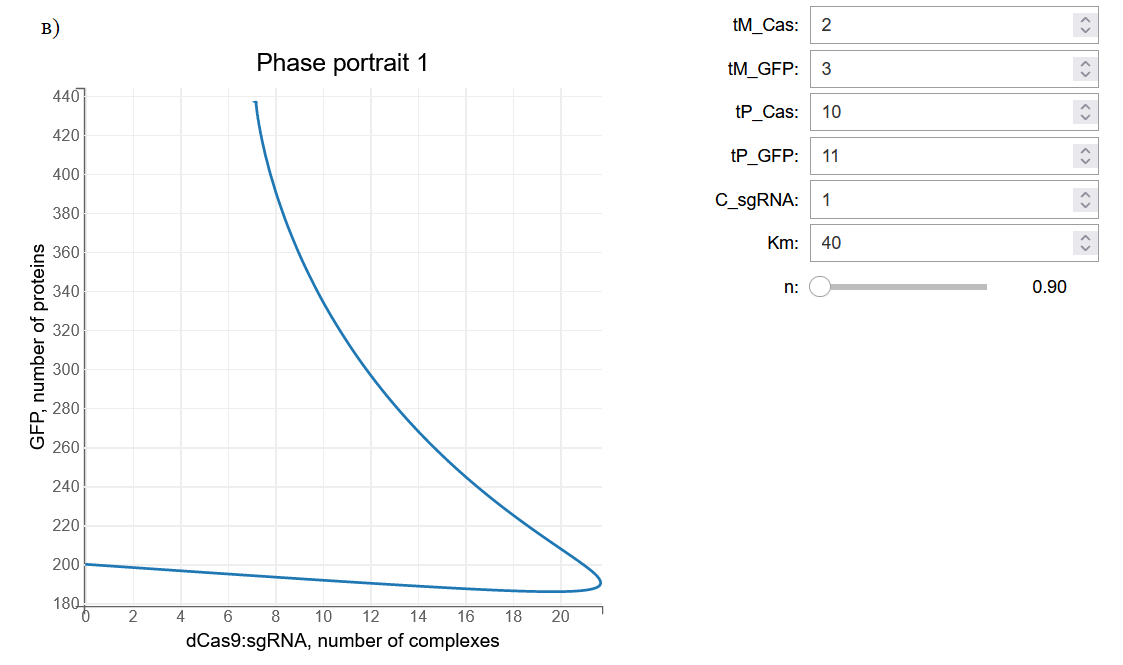
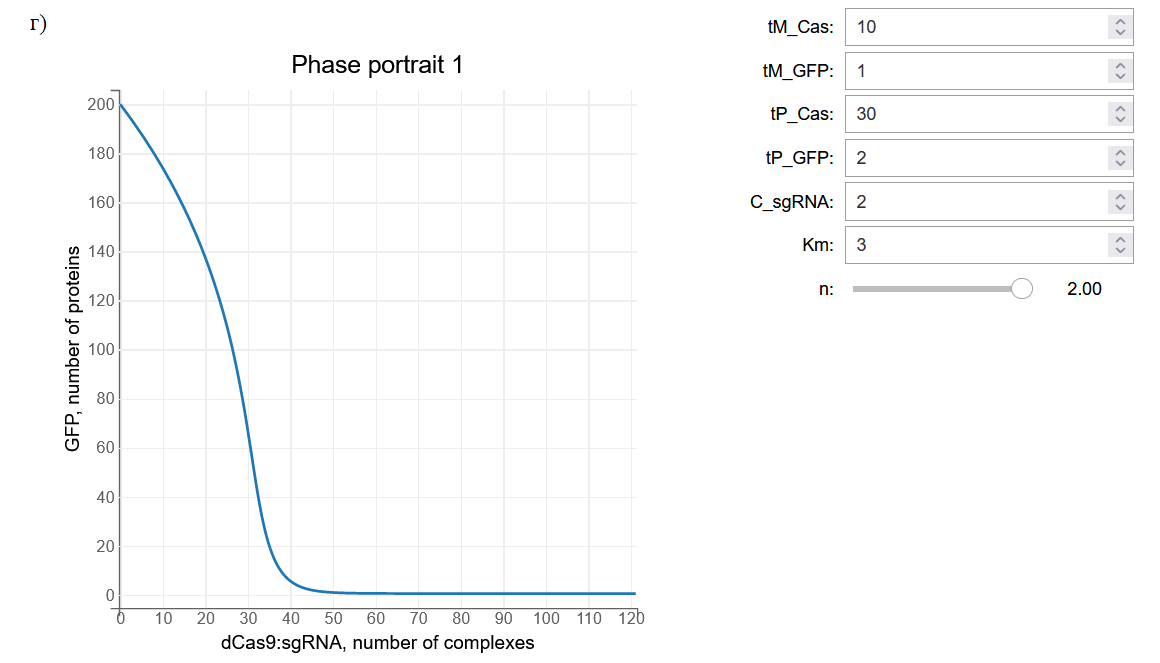


Рис. 4. а) Режим работы «ВЫКЛ». б) Режим работы «ВКЛ», параметры подобраны для приближения к максимальному снижению концентрации GFP. в) (ниже) Фазовый портрет, соответствующий неработающему режиму «ВКЛ» - сигнал дан (концентрация гидовой РНК не равна 0), однако снижения концентрации GFP по мере увеличения концентрации репрессора не наблюдается. г) (ниже) Фазовый портрет системы в режиме работы «ВКЛ» - концентрация GFP по мере увеличения концентрации репрессора стремится к нулю.





Заметим, что количественные параметры, используемые для второй модели по большей части были взяты из базы данных BioNumbers, но часть значений взята из модели репрессилятора – это вынужденная мера в связи с нехваткой экспериментальных данных. Далее, диапазоны параметров, в которых работает логический элемент «НЕ» и осуществляются осцилляции в системе репрессилятора приходится подбирать вручную, если пользоваться данными графическими методами, несмотря на их наглядность. Поэтому в будущем мы планируем расширить анализ наших моделей, используя подходы, которые позволят нам точнее изучить как упомянутые в этой работе системы, так и более сложные системы на основе dCas-белков, включая генетические схемы из статьи [9]. Также следует более подробно изучить функционал уже имеющихся приложений для анализа генетических схем с целью воспроизведения нужных функций в нашем веб-приложении и использования некоторых из них в нашей работе.

# **Заключение**

Нам удалось создать веб-приложение, позволяющее работать с интерактивными графиками – моделями репрессилятора и логического элемента «НЕ». Помимо графиков зависимости концентраций веществ от времени мы использовали фазовые портреты систем для изучения их режимов работы и изучения влияния различных параметров на динамику системы. В будущем мы планируем расширить анализ уже имеющихся и новых моделей на основе dCas-белков, в частности, определить диапазоны параметров, в которых определенная генетическая схема способна выполнять свои функции. Также планируется использование уже имеющихся сред для моделирования с целью автоматизации нашей работы.

# **Сопроводительные материалы**

1. Модель репрессилятора (веб-приложение) - <https://intbio.org/2022_synbio_webapp/lab?path=Repressilator_Runge-Kutta_bq.ipynb>
2. Модель логического элемента «НЕ» (веб-приложение) -

<https://intbio.org/2022_synbio_webapp/lab?path=CRISPR_NOT_Gate.ipynb>

# **Список литературы**

1. Cameron, D., Bashor, C. & Collins, J. A brief history of synthetic biology. *Nat Rev Microbiol* **12**, 381–390 (2014).
2. Khalil, A., Collins, J. Synthetic biology: applications come of age. *Nat Rev Genet* **11**, 367–379 (2010).
3. Kim, H., Bojar, D. & Fussenegger, M. A CRISPR/Cas9-based central processing unit to program complex logic computation in human cells. *PNAS* **116**(15), 7214-7219 (2019).
4. Elowitz, M., Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
5. Gardner, T., Cantor, C. & Collins, J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000).
6. Koonin, E., Makarova, K. Origins and evolution of CRISPR-Cas systems. *Phil. Trans. R. Soc.* *B* **374**: 20180087 (2019).
7. Zhang, S., Voigt, C. Engineered dCas9 with reduced toxicity in bacteria: implications for genetic circuit design. *Nucleic Acids Research* **46**(20), 11115-11125 (2018).
8. Shaytan, A., Novikov, R., Vinnikov, R. *et al.* From DNA-protein interactions to the genetic circuit design using CRISPR-dCas systems. *Front. Mol. Biosci., Sec. Biophysics* **9** (2022)
9. Santos-Moreno, J., Tasiudi, E., Stelling, J. *et al.* Multistable and dynamic CRISPRi-based synthetic circuits. *Nat Commun* **11**, 2746 (2020).
10. Bordbar, A., Monk, J., King, Z. *et al.* Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Genet* **15**, 107–120 (2014).
11. Hucka, M., Finney, A., Sauro, H. *et al.* The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**(4), 524-31 (2003).
12. Le Novère, N., Bornstein, B., Broicher, A. *et al.* BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems, Nucleic Acids Research **34**(1), D689–D691 (2006).
13. Funahashi A., Matsuoka Y., Jouraku A. *et al.* CellDesigner 3.5: A Versatile Modeling Tool for Biochemical Networks, Proceedings of the IEEE **96**(8), 1254-1265 (2008).
14. Gizzatkulov, N., Goryanin, I., Metelkin, E. *et al.* DBSolve Optimum: a software package for kinetic modeling which allows dynamic visualization of simulation results. *BMC Syst Biol* **4**, 109 (2010).
15. Hoops, S., Sahle, S., Gauges, R. *et al.* COPASI—a COmplex PAthway SImulator, Bioinformatics **22**(24), 3067–3074 (2006).
16. Klipp E., Liebermeister W., Wierling, C. *et al.* Systems Biology: A Textbook, Second Edition, *Wiley*,488 стр. (2016).
17. Milo, R., Jorgensen, P., Moran, U. *et al.* BioNumbers - the database of key numbers in molecular and cell biology, Nucleic Acids Research **38**(1), D750–D753 (2010).