**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 и анализ экспериментальных данных по функционированию генетических схем в бактериях *E.сoli*.

## Выполнил студент

3 курса 302 группы

## Винников Ренат Сергеевич

(подпись студента)

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

Шайтан Алексей Константинович

(подпись руководителя)

# Москва 2024

**Оглавление**

[**Введение** 2](#_Toc135600093)

[**Материалы и методы** 6](#_Toc135600094)

[**Результаты и обсуждение** 8](#_Toc135600095)

[**Заключение** 16](#_Toc135600096)

[**Сопроводительные материалы** 16](#_Toc135600097)

[**Список литературы** 17](#_Toc135600098)

# **Введение**

В литературном обзоре прошлой курсовой работы мы рассмотрели синтетическую биологию как раздел генетической инженерии; некоторые задачи, поставленные в данной области науки, в том числе создание искусственных генетических схем – аналогов электронных логических схем; подчеркнули использование белка dCas9 в качестве регулятора транскрипции и его роли в дизайне генетических сетей, а также обсудили кинетическое моделирование данных систем.

Повторимся, что dCas9 обрел популярность благодаря возможности запрограммировать их для регуляции любого участка в геноме. Механизм связывания данного белка с ДНК, опосредованный гидовой РНК (гРНК) позволяет создать любое количество гРНК, а значит, сделать возможным регуляцию нескольких промоторов одновременно, без привлечения дополнительных белков-репрессоров [1]. К сожалению, использование dCas9 ограничивается следующими факторами: токсичностью белка, связанной с его неспецифичным связыванием короткого мотива PAM (2-6 п.н.), отсутствием кооперативного характера связывания комплекса белок:РНК с ДНК и крайне медленным отсоединением этого комплекса от ДНК [1,2].

Проблема кооперативности связывания особенно остро стоит при создании сложных систем, в том числе, включающих в себя несколько уровней генетических логических элементов. Несмотря на данные, указывающие на возможность создания сложных генетических схем, обусловленную неспецифичным связыванием dCas-белков [1], работы, моделирующие кинетику данных генетических сетей, используют уравнение Хилла – кинетическое уравнение, описывающее кооперативное связывание фермента с субстратом или белка с ДНК [2].

Для решения данных проблем создаются различные варианты dCas9 с внесенными в них мутациями, в том числе, ослабляющими связывание с PAM, используются гибридные белки, где с помощью линкерного пептида к dCas9 присоединяются транскрипционные факторы – активаторы (CRISPRa) или репрессоры (CRISPRi), изменяется также время деградации мРНК и белков dCas9 [2,3,4,5]. Также можно настраивать и другие компоненты, составляющие генетическую сеть – изменять силу промоторов и сайтов связывания рибосом – таким образом будет изменена скорость транскрипции и трансляции различных молекул [6].

При моделировании схемы, состоящей из нескольких уровней логических элементов, следует также учитывать конкуренцию различных гРНК за пул dCas-белка; пороги входных и выходных сигналов [3, 6, 7] (в случае CRISPRi/a входным сигналом может служить молекула, индуцирующая транскрипцию, а выходным – флуоресцентный белок или, в случае промежуточного «слоя» логических элементов – гРНК) – их необходимо согласовывать для корректной работы сети – общий принцип здесь гласит, что концентрации используемой в качестве индуктора, должны отличаться от концентрации получаемой в результате работы генетической схемы гРНК2 не более, чем на два порядка [8]. По этой же причине следует снижать текучесть промоторов (например, увеличивая силу репрессоров) для избавления от шума, который может повлиять на работу сети [6]; Шум также можно снизить, используя активаторные и репрессорные комплексы в одной генетической схеме [9].

В настоящий момент существует довольно мало работ по кинетическому моделированию генетических схем, в особенности сложных, состоящих из нескольких уровней. По этой причине, научившись моделировать базовые генетические схемы, мы переходим к моделированию логического элемента «ИЛИ-НЕ», состоящего из двух «слоев» логических элементов, и совмещающего в себе CRISPRi и CRISPRa, а также белки dCas9 и dCas12.

Еще одной важной задачей как в повторении экспериментов других авторов, так и в проведении своих для создания генетических схем, является анализ роста бактерий. Рост бактерий описывается кривой роста, которая включает в себя индукционный период (lag-фазу), во время которого клетки адаптируются к субстрату, например, синтезируя ферменты, необходимые для его метаболизма, экспоненциальную (log-фазу), во время которой достигается максимальная скорость роста, фазу линейного роста, характеризующуюся замедлением роста (в связи с уменьшением концентрации питательного субстрата), стационарную фазу, на протяжении которой число клеток остается практически постоянным (т.е. клетки либо почти не делятся и не отмирают, либо число делящихся клеток равно числу умирающих клеток) и фазу отмирания культуры, где умирающих клеток больше, чем делящихся [10].

Данные о росте бактерий во времени получают фотометрией, после чего необходимо перейти от оптической плотности (OD) при заданной длине волны (обычно 600 нм) к числу клеток. Для этого необходимо построить калибровочную кривую, используя растворы с известным числом клеток. После этого можно приступать к приближению экспериментальной кривой различными моделями роста клеток [11]. Модели делятся на структурированные и неструктурированные: структурированные модели рассматривают рост клеток на молекулярном уровне – включают в себя дифференциальные уравнения, составленные на основе закона действующих масс для отдельных компонентов внутри клетки и также описывающие влияние этих компонентов на рост клеток. Неструктурированные модели описывают рост клеток основываясь только на начальной концентрации биомассы и концентрациях субстратов, то есть рассматривают рост на клеточном уровне. В целом, неструктурированные модели лучше описывают сбалансированный рост, то есть такой рост, при котором происходит пропорциональное увеличение количества всех клеточных компонентов (и в итоге роста получаются одинаковые в размерах и составе клетки), в то время как структурированные хорошо справляются с описаниями несбалансированного роста, однако из-за большого числа уравнений в структурированных моделях, они имеют довольно большую вычислительную сложность и используются реже, чем неструктурированные [10,11].

В августе 2023 наши коллеги повторяли эксперименты Santos-Moreno по созданию логического элемента «НЕ» в клетках *E. coli* (штамм BL21), и после обработки данных и получения кривой зависимости OD600 от времени, перед нами встала проблема анализа данной кривой. Успешный анализ этой кривой позволит в будущем анализировать и другие экспериментальные данные.

По этим причинам, мы сформулировали следующие цели работы:

1. Изучение работы двуслойной генетической схемы «ИЛИ-НЕ» на основе CRISPRa/i с различной кооперативностью с помощью детерминистического кинетического моделирования.
2. Получение кривой зависимости числа бактерий от времени на основе экспериментальных данных и её анализ.

Для достижения целей были поставлены следующие задачи:

1. Задать математическую модель генетической сети, составить систему уравнений, описывающую её динамику.
2. Используя различные методы численного моделирования, создать программу, позволяющие в интерактивном режиме осуществлять численное решение систем дифференциальных уравнений, описывающих функционирование выбранной генетической схемы.
3. Изучить режимы работы и особенности функционирования данной генетической схемы на основе разработанной модели.
4. Построить калибровочную кривую для спектрофотометра (планшетного ридера), связывающую OD600 с количеством клеток.
5. На основании калибровочной кривой построить кривую роста бактерий и проанализировать их рост.

# 

# **Материалы и методы**

Мы исследовали модель логического элемента «ИЛИ-НЕ», где в качестве как репрессора, так и активатора выступает комплекс белка dCas9 c различными гидовыми РНК.

Уравнения динамики логического элемента «ИЛИ-НЕ» были выведены из основных уравнений химической кинетики и экспрессии генов [13].

Численные значения параметров для соответствующих систем были взяты из баз данных BioModels [14] и BioNumbers [15].

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 визуализации использовалась библиотеки bqplot (0.12.33), matplotlib (3.9.0), seaborn (0.13.2). Интерактивные виджеты создавались с помощью ipywidgets (7.6.5).

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

Для построения калибровочной кривой использовались следующие материалы:

* 10 мл ночной культуры *E. coli* (штамм BL21(DE3)), выросшей в среде LB (1% триптон, 0.5% экстракт дрожжей, 1% NaCl)
* 4 стерильных бутыли с 99 мл фосфатного буфера
* Микропипетки
* 8 чашек Петри
* Твердая среда LB (LB + агар (1,5%))
* 4 пробирки на 10 мл
* Планшетный ридер POLARstar Omega

Мы следовали следующему протоколу:

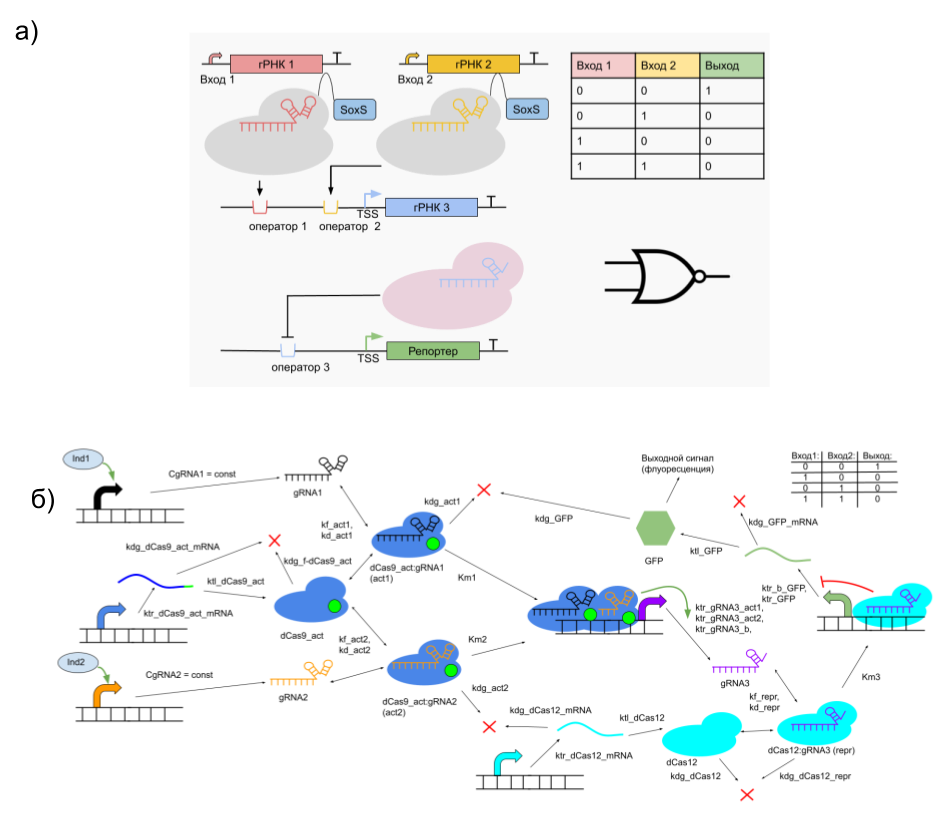
1. Получили ночную культуру: замороженную культуру клеток (хранящуюся при температуре -80ºС) разогрели и перенесли на чашку Петри с твердой средой LB, распределив их равномерно методом штриха с помощью предварительно стерилизованной бактериальной петли, оставили инкубироваться при 37ºC на ночь, на следующий день выкололи в среду LB клетки кончиком носика микропипетки.
2. Пронумеровали чашки Петри номерами 1-8, а бутыли с фосфатным буфером – 10-2, 10-4, 10-6 и 10-8.
3. Соблюдая стерильность, встряхнули колбу с ночной культурой и микропипеткой перенесли 1 мл культуры в бутыль с фосфатным буфером, получив разведение в 100 (10-2) раз.
4. Встряхнув бутыль, открутили пробку и перенесли 1 мл раствора культуры в фосфатном буфере в бутыль 10-4. Повторили эти действия с другими бутылями для получения разведения в 10-6 и 10-8 раз.
5. Залили чашки Петри расплавленным агаром, подождали, пока агар остынет.
6. Встряхнув бутыль 10-2, перенесли в чашку Петри №1 1 мл раствора, равномерно распределив его по чашке Петри с помощью стеклянного шпателя (предварительно стерилизованного) и в чашку №2 – 0.1 мл раствора. Аналогично повторили для других бутылей и чашек.
7. Перевернули чашки и оставили их инкубироваться на 24 часа при 37ºC.
8. После окончания инкубационного периода, мы посчитали количество колоний в чашках – по статистическим соображениям, число колоний не должно быть значительно меньше 30 и значительно больше 300. В результате две чашки Петри (6 и 7) подошли для дальнейших расчетов (21 и 144 колонии).
9. Мы рассчитали количество бактерий на мл. Для этого мы сначала пересчитали число колоний на мл раствора, а затем построили линейную регрессию числа колоний на мл от разведения (10-6 и 10-8).
10. Подписали 4 пробирки на 10 мл: 1/2, 1/4, 1/8 и 1/16. Внесли в каждую пробирку 5 мл среды LB.
11. Взяли колбу с исходной ночной культурой *E. coli*, встряхнули ее и перенесли 5 мл культуры в первую пробирку, перемешали – таким образом получили разведение 1/2.
12. Встряхнув, перенесли 5 мл раствора из первой пробирки во вторую. Повторили еще два раза, получив ряд разведений от 1/2 до 1/16.
13. В планшете в 5 лунок 1-го ряда внесли по 200 мкл чистой среды LB. В 5 лунок 2-го ряда внесли по 200 мкл оригинальной культуры клеток. В 5 лунок следующих 4 рядов вносили растворы из пробирок: 1/2, 1/4, 1/8 и 1/16.
14. Загрузили планшет в планшетный ридер и произвели измерение OD600.

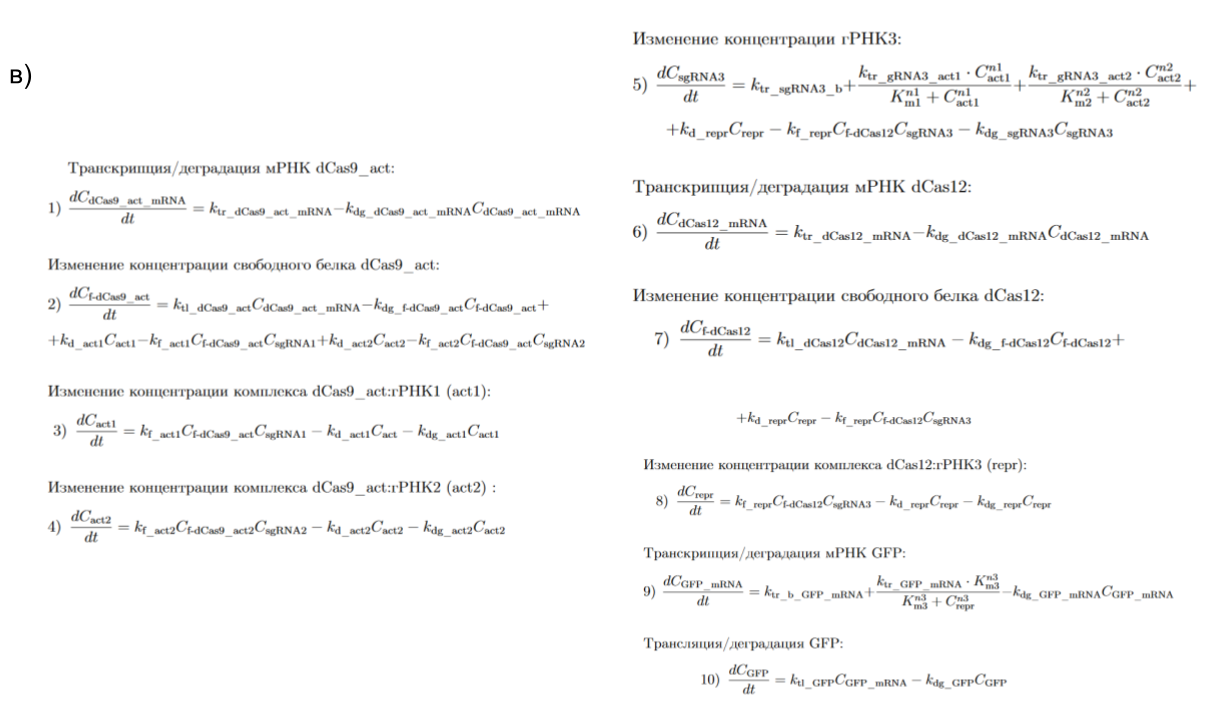
Данные (зависимость OD600 от времени) были получены в формате .ruc (внутренний формат программного обеспечения планшетного ридера), откуда они были экспортированы в формате .tsv и загружены в Google colab для обработки данных библиотеками pandas (2.2.2) и scipy (1.13.1). На основании коэффициентов регрессии, полученных в п. 8 протокола, было посчитано количество бактерий в пробирках (п. 9-11) и построена еще одна линейная регрессия, для OD600 и числа клеток.

# **Результаты и обсуждение**

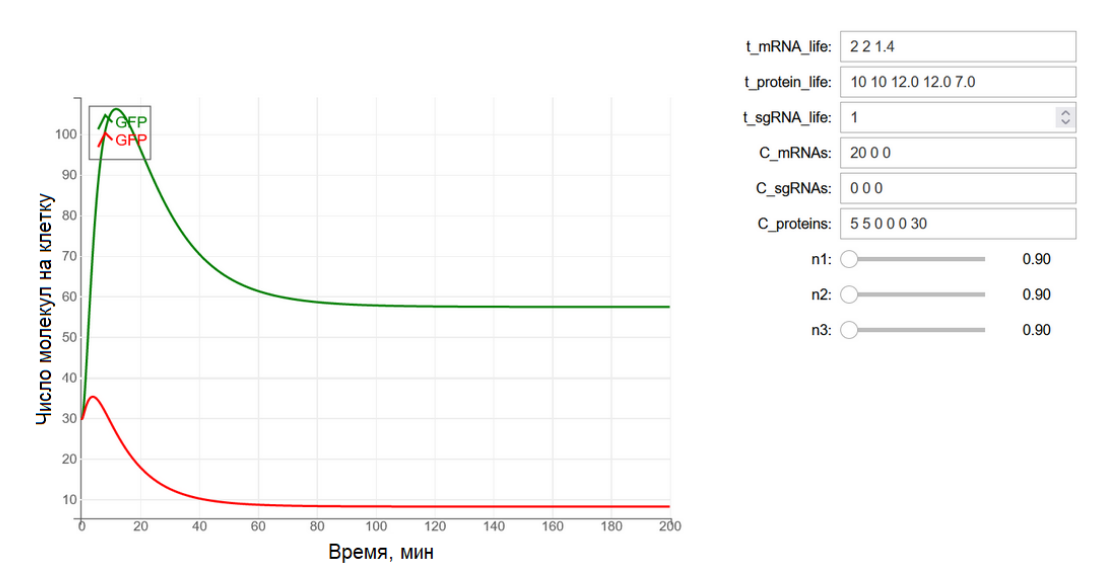
Модель логического элемента «ИЛИ-НЕ» выполнена в виде веб-приложения на основе файла Jupyter Notebook с кодом, чтобы пользователи могли работать с нашей моделью непосредственно в браузере. Для этого мы использовали JupyterLite – среду Jupyter, написанную на WebAssembly.

*Рис. 1. (ниже) а) Принципиальная схема разработанной модели логического элемента “ИЛИ-НЕ” и соответствующая ему таблица истинности. б) Более детальная схема логического элемента “ИЛИ-НЕ”. в) Математическая модель элемента “ИЛИ-НЕ”* – *дифференциальные уравнения. Здесь Km обозначает равновесные константы диссоциации комплекса dCas:гРНК от ДНК, ktr* – *скорости транскрипции,  ktl* – *константы трансляции,  kdg* – *константы деградации,  kf* – *константы формирования комплексов,  kd* –*константы распада комплексов, n - коэффициент Хилла.*





Модель отражает различные режимы работы системы (см. Рис. 2). В качестве входных сигналов служат гРНК. В качестве выходного – концентрация зеленого флуоресцентного белка (что дает сигнал флуоресценции).  Для простоты интенсивность входного сигнала (концентрации молекул гРНК) в рамках модели считается постоянной на протяжении всего периода времени.



*Рисунок 2. Результаты кинетического моделирования с помощью интерактивной модели логического элемента “ИЛИ-НЕ”, показанного на рис. 1. Зеленая кривая обозначает концентрацию GFP (число молекул на клетку) в отсутствии гРНК1 и гРНК2 (Вход: 0 0, Выход: 1), красная - концентрацию GFP в присутствии гРНК1 и/или гРНК2 (Вход: 0 1/1 0/1 1, Выход: 0).*

Равновесная константа диссоциации комплекса dCas9:гРНК от ДНК получена в экспериментальной части работы, остальные параметры взяты из источников литературы и базы данных Biomodels.

Моделирование показывает, что повышение коэффициента Хилла (мера кооперативности связывания) улучшает работу сети – происходит более сильное подавление/активация транскрипции элементов системы, что делает отличие разных состояний системы более явным.

Как видно из рис.1, «ИЛИ-НЕ» состоит из элемента первого уровня «ИЛИ» и элемента второго уровня «НЕ». Связь уровней обеспечивает гРНК3 – ее транскрипция активируется только в присутствии комплекса dCas9\_act:гРНК1 или dCas9\_act:гРНК2. dCas9\_act – гибридный белок, в котором dCas9 соединен линкерным пептидом с прокариотическим активаторным доменом (например, SoxS, TetD). Для связывания уровней и корректной работы схемы нужно, чтобы получаемые в результате работы генетической схемы концентрация гРНК2 не отличалась от концентрации гРНК1, используемой в качестве индуктора более, чем на два порядка – это правило распространяется на более многоуровневые сети. Такое правило выполняется для нашей модели в рамках выбранных параметров (рис. 3).

*Рисунок 3 (ниже). Зависимость количества молекул гидовых РНК в расчете на клетку для модели логического элемента «ИЛИ-НЕ» от времени в ходе кинетического моделирования работы элемента.  На 1 молекулу гРНК2 приходится ~5.5 молекул гРНК3, большая часть из которых находится в форме, связанной белком dCas12.*

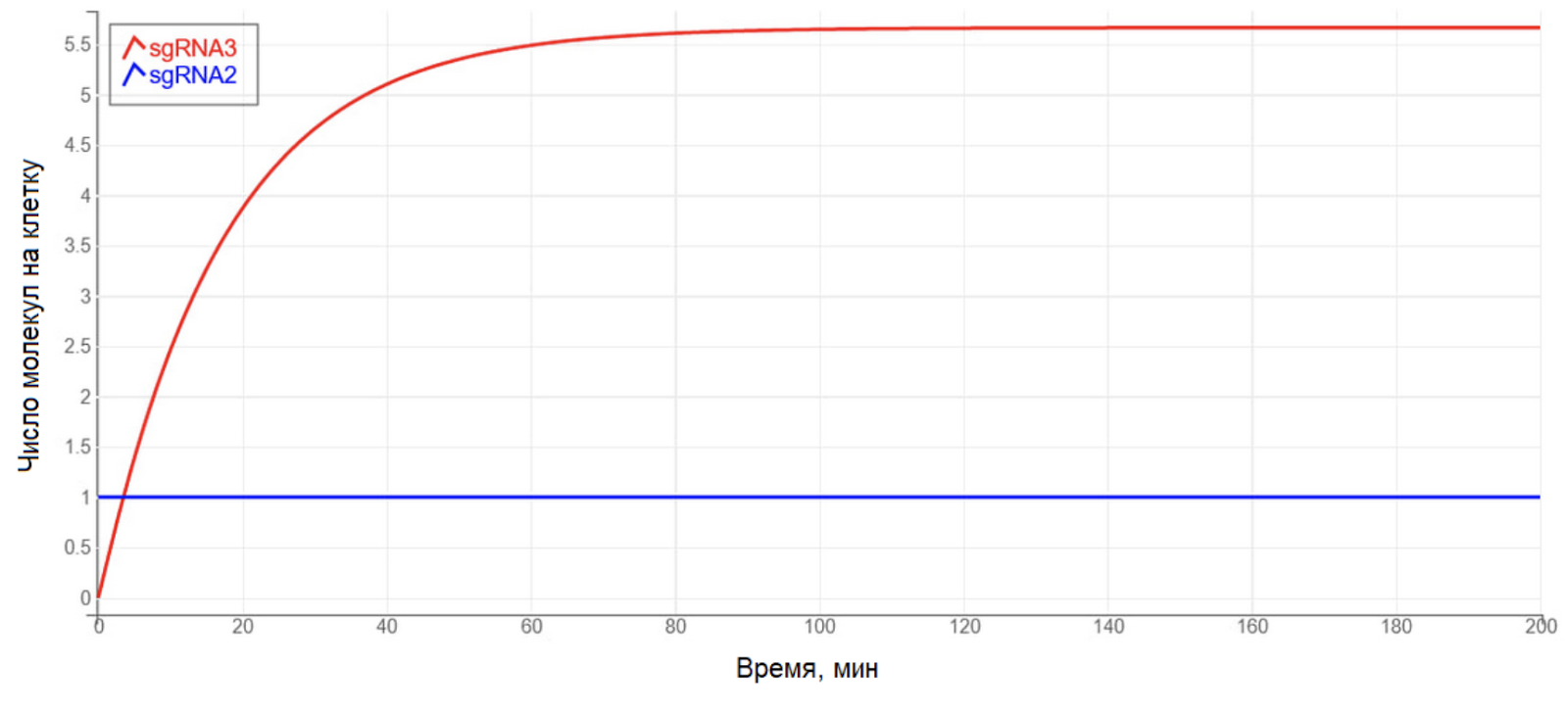
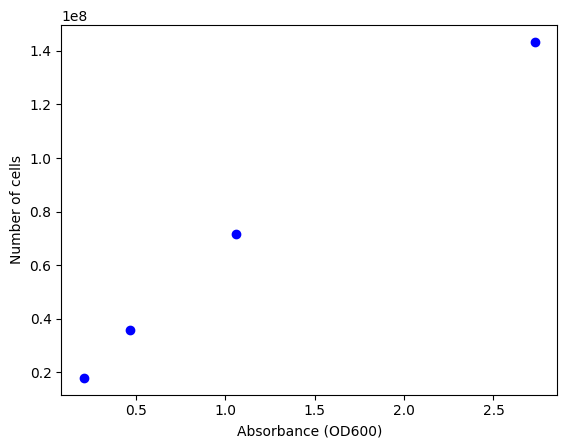
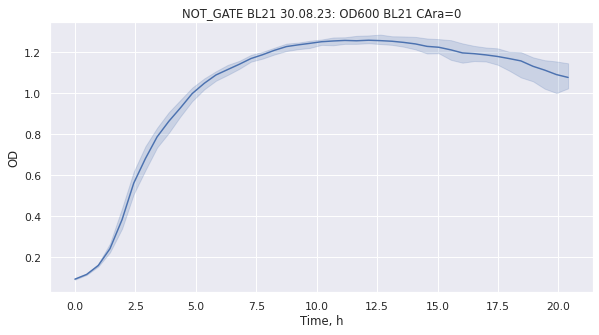
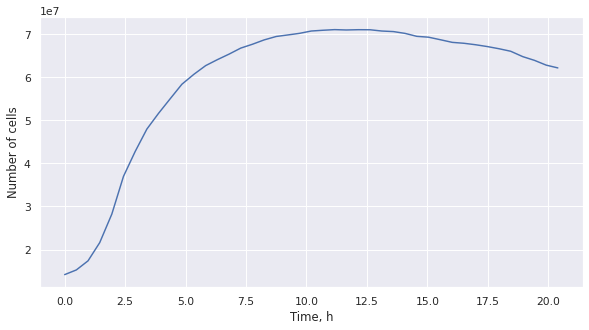


Рис. 4 показывает, как связана OD600 с числом клеток. Коэффициенты этой регрессии были использованы для перевода экспериментальной кривой (зависимость OD600 от времени) в кривую роста бактерий (рис. 5).



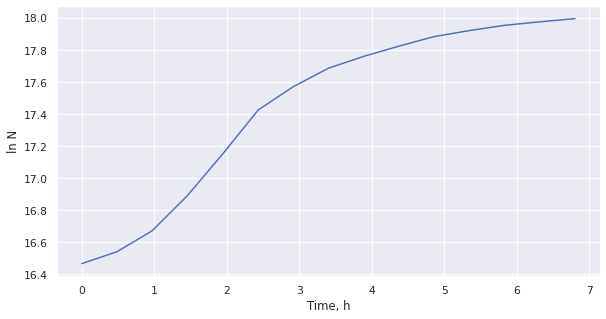
*Рис. 4. График, сопоставляющий OD600 и количество клеток.*





*Рис. 5. Верхняя кривая – зависимость OD600 от времени. Нижняя кривая – кривая роста бактерий.*

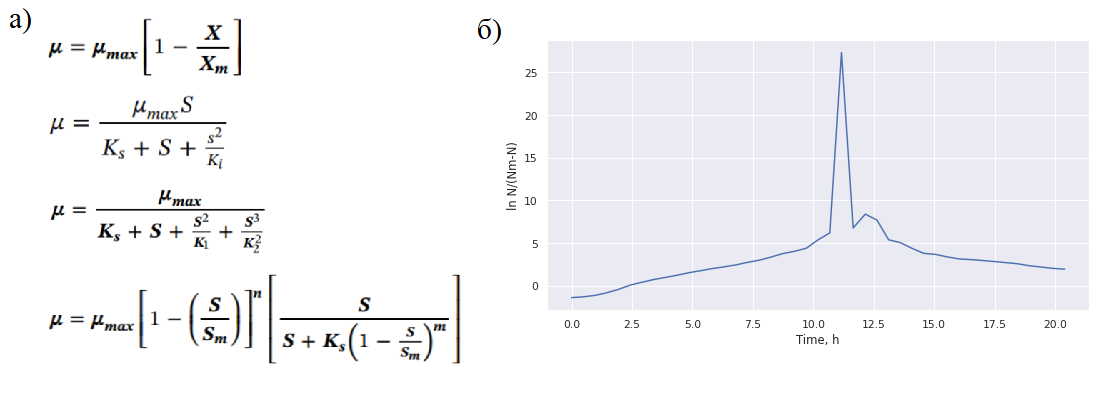
Дальнейший анализ требовал переноса кривой в логарифмические координаты (рис. 6). Нужно было понять, какие фазы включает в себя данная кривая роста. Особенно интересен начальный участок кривой. Скорость роста бактерий, как видно из рисунка, неоднородна во времени, значит, в нашем случае наблюдается lag-фаза и фаза замедления роста, что накладывает ограничения на приближение этой кривой моделями роста бактерий. Значения удельной скорости роста (т.е. изменение количества клеток в пересчете на одну клетку в промежуток времени) в разные промежутки времени были определены из формулы µ = ln(NT/N0)/T, где NT – количество клеток в конце временного промежутка, N0 – количество клеток в начале промежутка, T – длина временного промежутка. В lag-фазу удельная скорость роста составила 0.2105 клетки в час, в фазу экспоненциального роста – 0.5149, затем, в фазу замедления роста скорость снизилась до 0.2297 кл/ч и, наконец, до 0.0809 кл/ч, что соответствует переходу в стационарную фазу.



*Рис 6. Кривая из рис. 5 в логарифмических координатах. Показан начальный участок кривой. Как видно, рост не является однородным. Наблюдается лаг-фаза (в течение первого часа) и фаза замедления роста (начинается между вторым и третьим часом).*

Неструктурированные модели, приведенные в литературе [10–12], являются в основном расширениями модели Моно. Подобные модели хорошо описывают лишь фазу экспоненциального роста. Впрочем, ряд других моделей способен описывать прочие фазы – к таким моделям относятся логистическая модель, модели Холдейна, Йана-Кога и Хана-Левенспиля (рис. 7а). Довольно просто проверить соответствие кривой логистической модели – для этого стоит построить кривую в координатах ln N/(Nm-N), t (рис.7б). В данном случае, линеаризации не произошло, следовательно, данная модель не подходит для описания нашей экспериментальной кривой.

*Рис. 7. (ниже) а) Сверху-вниз: логистическая модель, модель Холдейна, модель Йана-Кога, модель Хана-Левенспиля. μ – удельная скорость роста, μmax – максимальная удельная скорость роста, S – концентрация субстрата, X − концентрация живой биомассы (в прочих уравнениях − N), Ks – константа полунасыщения (Km), Ki – константа ингибирования (Kd), Sm – критическая концентрация субстрата, действующего как ингибитор, Xm – предельная концентрация живой биомассы. б) кривая из рис. 5 в «логистических» координатах. Линеаризация не произошла, соответственно, данная модель не подходит для описания кривой.*



К сожалению, исходные экспериментальные данные не включали в себя скорости поглощения субстрата и концентрацию, при которой избыток (или недостаток) субстрата начинает ингибировать рост клеток, соответственно, на данный момент мы не можем приблизить кривую моделями, рассмотренными выше. Для дальнейшего анализа нам потребуется повторить эти эксперименты, измерив данные параметры.

# 

# **Заключение**

Нам удалось создать интерактивную модель двухуровневого логического элемента «ИЛИ-НЕ», включающего в себя как CRISPRi, так и CRISPRa. Нашу модель следует дополнить экспериментальными данными для лучшего прогнозирования поведения реальной системы. В будущем мы планируем моделировать еще более сложные системы. Кроме того, мы откалибровали планшетный ридер POLARstar Omega и построили кривую роста бактерий, соответствующую экспериментальным данным, измерили удельные скорости роста бактерий в разные промежутки времени. В дальнейшем нам следует провести дополнительные измерения, связанные с изменением концентрации субстрата и определить концентрацию субстрата, при которой скорость роста клеток изменяется. Это позволит описать кривую существующими моделями роста бактериальных клеток.

# **Сопроводительные материалы**

1. Модель логического элемента «ИЛИ-НЕ» (веб-приложение): <https://intbio.org/2022_synbio_webapp/lab?path=CRISPR_NOR_GATE.ipynb>
2. Та же модель в Google Colab: <https://colab.research.google.com/drive/194blU30HINWEQ8Yg1KHHlaWB78U-_pTE?usp=sharing>
3. Обработка кривой оптической плотности: <https://colab.research.google.com/drive/1PNeU_YQttQY0v_Z2MzIoM0rw8oo06d5M?usp=sharing>
4. Построение калибровочной кривой планшетного ридера: <https://colab.research.google.com/drive/1AcCk01m2SfKAPLBUdAUZ5KMiqzQBxGhE>

# **Список литературы**

1. Santos-Moreno, J., Tasiudi, E., Stelling, J. *et al.* Multistable and dynamic CRISPRi-based synthetic circuits. *Nat Commun* **11**, 2746 (2020).
2. Zhang, S., Voigt, C. Engineered dCas9 with reduced toxicity in bacteria: implications for genetic circuit design. *Nucleic Acids Research* **46**(20), 11115-11125 (2018).
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. 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).
5. Clamons, S., Murray, R. Modelling dynamic transcriptional circuits with CRISPRi. *Biorxiv (preprint)* (2022).
6. Brophy, J., Voigt, C. Principles of genetic circuit design. *Nature Methods,* **11**(5), 508–520 (2014).
7. Gander, M., Vrana, J., Voje, W. *et al.* Digital logic circuits in yeast with CRISPR-dCas9 NOR gates. *Nat Commun***8**,15459 (2017).
8. Tickman, B., Burbano, D., Chavali, V. *et al.* Multi-layer CRISPRa/i circuits for dynamic genetic programs in cell-free and bacterial systems. *Cell Systems* **13**, 215–229 (2022).
9. Elowitz, M., Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
10. Варфоломеев, С., Гуревич, К. Биокинетика. *ФАИР-ПРЕСС*, 715 стр. (1999).
11. Muloiwa, M., Nyende-Byakika, S., Dinka, M., Comparison of unstructured kinetic bacterial growth models. *South African Journal of Chemical Engineering* **33**, 141-150 (2020).
12. Panikov, N., Kinetics, microbial growth. *Encyclopedia of Industrial Biotechnology: bioprocess, Bioseparation, and Cell Technology* 1–34. (2009).
13. Klipp, E., Liebermeister, W., Wierling, C. *et al.* Systems Biology: A Textbook, Second Edition, *Wiley*,488 стр. (2016).
14. 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).
15. 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).