

Министерство образования и науки Российской Федерации  
Федеральное государственное автономное образовательное учреждение  
высшего образования  
«Санкт-Петербургский политехнический университет Петра Великого»

УДК

№

Инв. №

У Т В Е Р Ж Д А Ю

Зав. НИЛ «Математическая биология  
и биоинформатика», ИПММ  
ФГАОУ ВО «СПбПУ»,  
д.б.н.

\_\_\_\_\_ М. Г. Самсонова

«\_\_\_\_\_» \_\_\_\_\_ 2016 г.

ОТЧЕТ  
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

по теме:

«Стохастическое моделирование экспрессии генов»

Выполнил студент гр. №53601/4

\_\_\_\_\_ Д. В. Яковлев

«\_\_\_\_\_» \_\_\_\_\_ 2016 г.

Руководитель НИР, к.ф.-м.н.

\_\_\_\_\_ В.В. Гурский

«\_\_\_\_\_» \_\_\_\_\_ 2016 г.

Санкт-Петербург 2016

## РЕФЕРАТ

Отчёт 14 стр., 1 часть, 5 рис., 5 источников

**СТОХАСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ЭКСПРЕССИИ ГЕНОВ** В данной работе описываются реакции моделирования экспрессии генов. Приводятся результаты стохастического моделирования экспрессии генов.

# СОДЕРЖАНИЕ

Введение . . . . .	4
1 Модель . . . . .	5
1.1 Транскрипционные факторы . . . . .	5
1.2 Транскрипция и трансляция . . . . .	5
1.3 Диффузия и деградация . . . . .	6
2 Инструменты для моделирования . . . . .	9
2.1 StochPy, GillesPy . . . . .	9
2.2 fern . . . . .	9
2.3 StochKit . . . . .	9
2.4 DEEP . . . . .	9
3 Результаты . . . . .	10
3.1 Описание реакций . . . . .	10
3.2 Разработка программного обеспечения . . . . .	11
3.3 Анализ результатов . . . . .	11
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ . . . . .	15

## Введение

Экспрессия генов - это процесс преобразования последовательности нуклеотидов ДНК в функциональный продукт РНК и белок.

Существует два основных направления моделирования экспрессии генов - аналитическое и стохастическое.

Цель научно-исследовательской работы:

- Описать реакции экспрессии генов
- Разработка программного обеспечения для стохастического моделирования экспрессии генов.
- Анализ экспериментальных расчётов
- Выводы

# 1 Модель

В данной работе будут рассмотрены 5 основных видов реакций в процессе экспрессии генов:

- Присоединение и отсоединение транскрипционных факторов (Рис. 1)
- Начало транскрипции (Рис. 2)
- Трансляция (Рис. 2)
- Деградация белка и тф (Рис. 3)
- Диффузия между ядрами (Рис. 3)

## 1.1 Транскрипционные факторы

Наиболее важную роль в экспрессии генов играет присоединение транскрипционных факторов. Связывание происходит к специфичным участкам ДНК - сайтам. Основным вкладом транскрипционных факторов в экспрессию генов является запуск процесса транскрипции. В данной работе будет предполагаться, что транскрипционные факторы самостоятельно взаимодействуют, а не в комплексе. Стоит отметить, что для каждого белка свои сайты присоединения на ДНК, которые известны из экспериментальных расчётов.

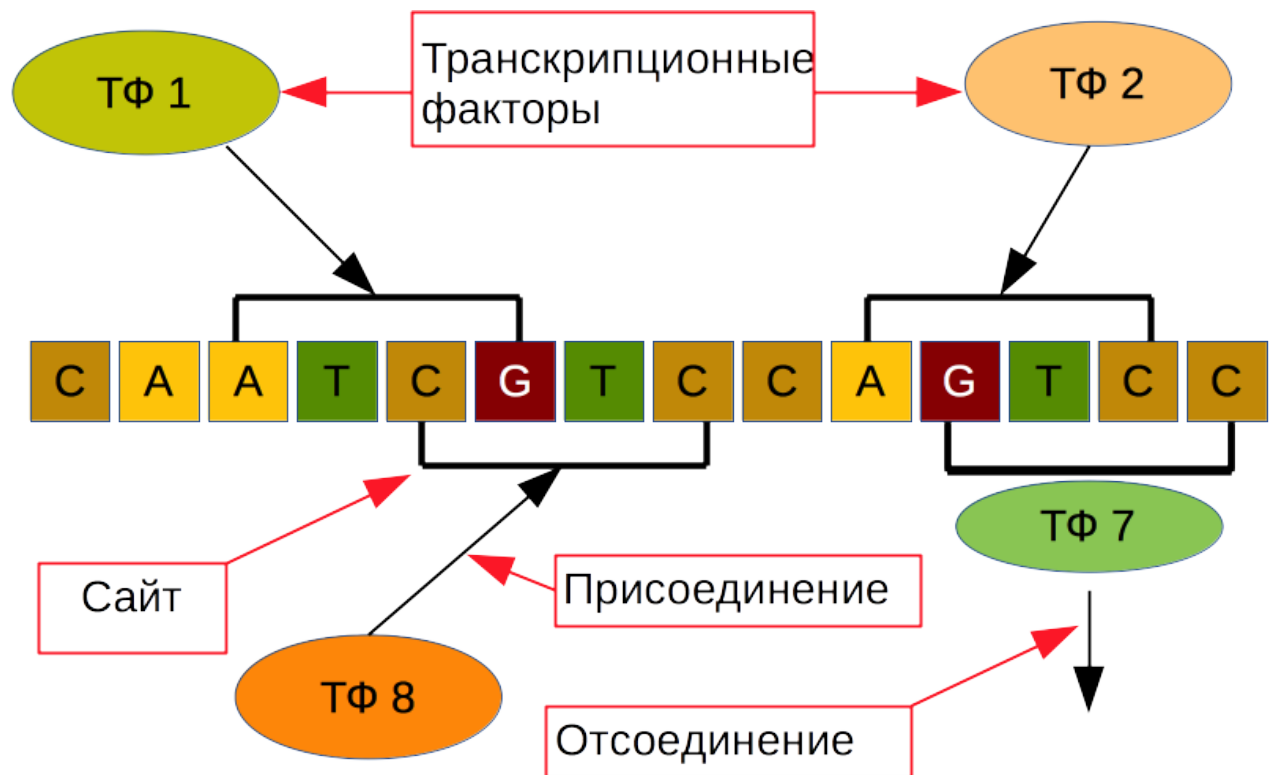
Транскрипционные факторы делятся на два типа:

- **Активаторы** - белки, которые активируют процесс транскрипции. В нашей работе это будут - **cad** и **bcd**.
- **Репрессоры** - белки, которые останавливают процесс транскрипции. В нашей работе это будут - **tll**, **hkb**, **hb**, **Kr**, **gt**

Таким образом в экспрессии генов будут задействованы 7 транскрипционных факторов - **cad**, **bcd**, **tll**, **hkb**, **hb**, **Kr**, **gt**

## 1.2 Транскрипция и трансляция

Транскрипция - процесс синтеза РНК с использованием ДНК в качестве матрицы, происходящий во всех живых клетках. Как упоминалось ранее активаторами для старта транскрипции являются белки **cad** и **bcd**. В результате транскрипции образуются молекулы РНК, которые будут участвовать в реакции трансляции, а именно синтезировать молекулы белка.

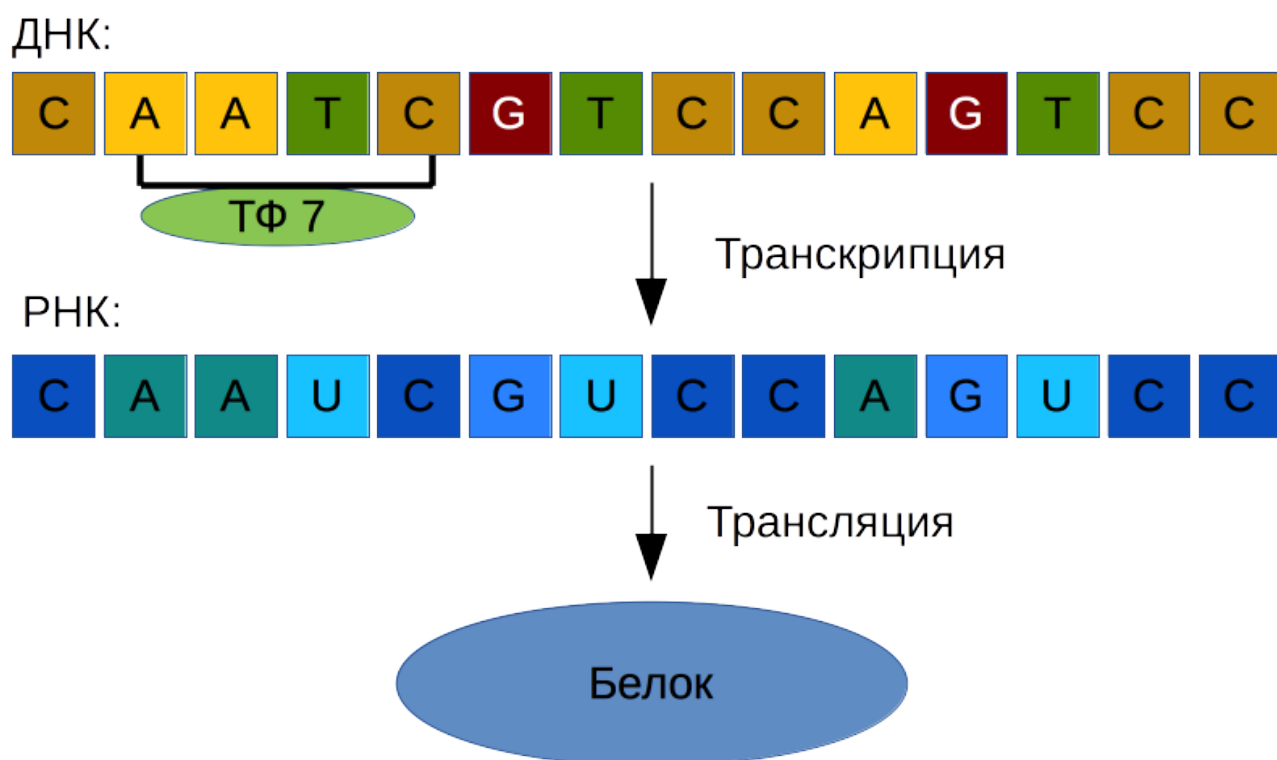


Для каждого сайта определено какой ТФ может присоединиться. Сайтов – 611 и ТФ – 8.

Рисунок 1 – Процесс присоединения и отсоединения ТФ

### 1.3 Диффузия и деградация

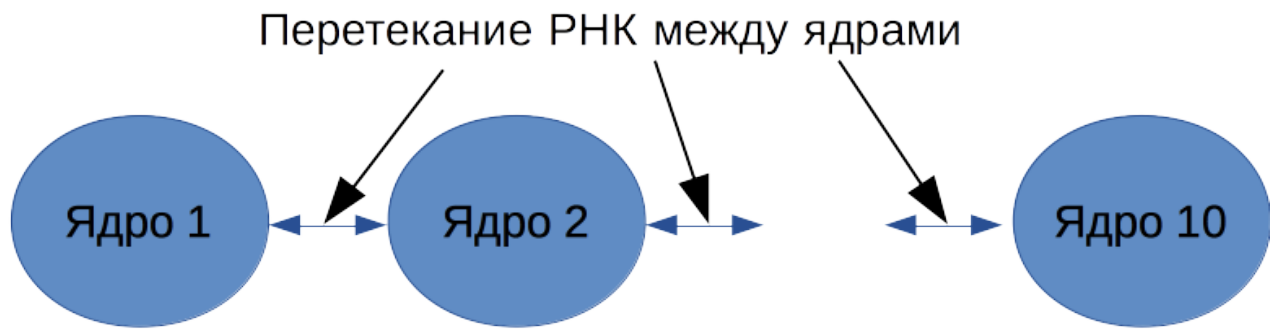
Будем предполагать, что все молекулы у нас линейно расположены. Тогда между ядрами может происходить процесс диффузии - переход РНК в соседнее ядро. Так же под действием внешней среды может происходить распад РНК и белка.



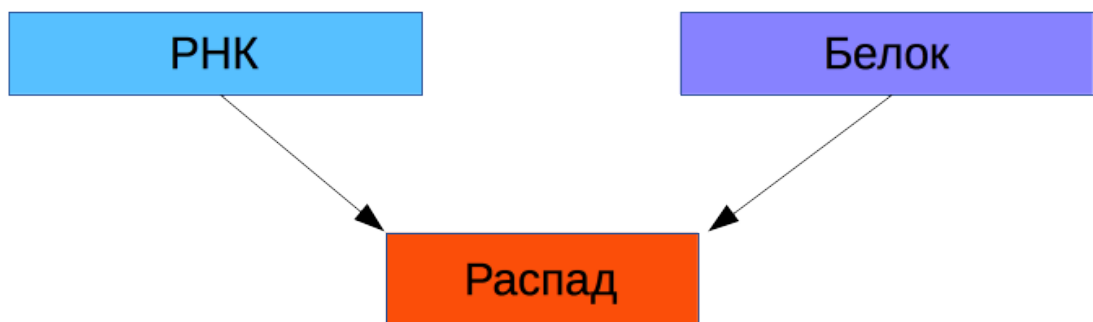
Вероятность транскрипции зависит от количества присоединённых ТФ, трансляции – от количества РНК.

Рисунок 2 – Процесс транскрипции и трансляции

### Диффузия:



### Деградация:



Вероятности деградации и диффузии линейно зависят от концентраций.

Рисунок 3 – Процесс диффузии и деградации



## 2 Инструменты для моделирования

Были рассмотрены следующие пакеты для стохастического моделирования - fern[1], StochKit[2], StochPy[3], GillesPy[2]. Каждый из них рассчитан на языки программирования - java, c++, python соответственно. Далее будет каждый пакет рассмотрен более подробно, но наш выбор остановился на StochKit, как на самом быстром.

### 2.1 *StochPy, GillesPy*

Является обвёрткой над пакетом StochKit с некоторыми модификациями. В комплекте с моделированием есть возможность строить графики. Содержит две модификации алгоритма гиллеспи. На вход не принимает файлы в формате sbml (нужно конвертировать в xml).

### 2.2 *fern*

Является самым медленным из предложенных пакетов стохастического моделирования. Содержит более семи модификаций алгоритма гиллеспи. На вход принимает файлы в формате sbml, что является большим плюсом. Данный пакет подходит для моделирования небольшого количества реакций.

### 2.3 *StochKit*

Является самым быстрым, так как написан на C и все остальные пакеты либо опираются на него, либо являются обвёрткой. В комплекте идёт программа, написанная на Matlab, для построения графиков, но к сожалению данные графики неинтересны. В пакете реализовано несколько модификаций алгоритма гиллеспи. Выходом программы StochKit является текстовый файл с концентрациями веществ в любой момент времени (частота задаётся параметром).

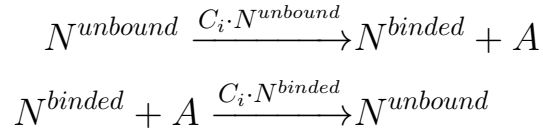
### 2.4 *DEEP*

Наша модель будет содержать параметры и для моделирования мы воспользуемся генетическими алгоритмами, реализованными в DEEP [4].

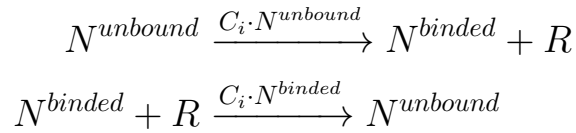
## 3 Результаты

### 3.1 Описание реакций

Опишем реакции для одного ядра (для остальных ядер аналогично). Для каждого сайта присоединения реакции присоединения и отсоединения будут иметь одинаковый вид. Основным отличием будет какой именно ТФ присоединяется/отсоединяется. Для активаторов (в нашей работе это **cad** или **bcd**) реакции примут следующий вид:

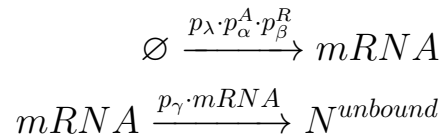


Для репрессоров реакции будут выглядеть так:



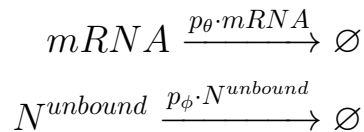
где  $C_i$  - коэффициент присоединения к  $i$ -ому сайту,  $R$  - количество репрессоров и  $A$  - количество активаторов.

Теперь запишем реакции транскрипции и трансляции:



где  $p_\lambda$ ,  $p_\alpha$ ,  $p_\beta$ ,  $p_\gamma$  - коэффициенты транскрипции, активации, регрессии и трансляции соответственно.

Реакции деградации:



где  $p_\theta$ ,  $p_\phi$  - коэффициенты деградации мРНК и белка соответственно.

Так же нужно учесть диффузию - перетекание белка в соседние ядра:

$$N_i \xrightarrow{p_\xi \cdot \max(N_i - N_{i+1}, 0)} N_{i+1}$$

$$N_i \xrightarrow{p_\xi \cdot \max(N_i - N_{i-1}, 0)} N_{i-1}$$

где  $p_\xi$  - параметр диффузии,  $i$  - номер ядра. В результате наша модель имеет 7 параметров -  $p_\lambda$ ,  $p_\alpha$ ,  $p_\beta$ ,  $p_\gamma$ ,  $p_\theta$ ,  $p_\phi$ ,  $p_\xi$ .

### 3.2 Разработка программного обеспечения

Пакет был разделён на четыре компоненты:

- Генерация системы реакций для пакета StochKit с помощью языка программирования Python
- Стохастическое моделирование с помощью пакета StochKit
- Оптимизация параметров модели с помощью DEEP
- Анализ результатов с помощью языка программирования R

Основной трудностью было наладить работу DEEP совместно с StochKit. Для этого нужно было изменить исходники StochKit, чтобы после каждого запуска он выводил ошибку моделирования при заданных параметрах.

### 3.3 Анализ результатов

В силу ограниченности возможности работы на кластере, было произведено несколько стартов. Было построены графики зависимости концентрации белка **kni** от ядра к концу моделирования (в нашем случае это 60 минута).

На графиках (4, 5) отчётливо видно, что поведение концентраций у белка **kni** и мРНК схожи с экспериментальными расчётами.

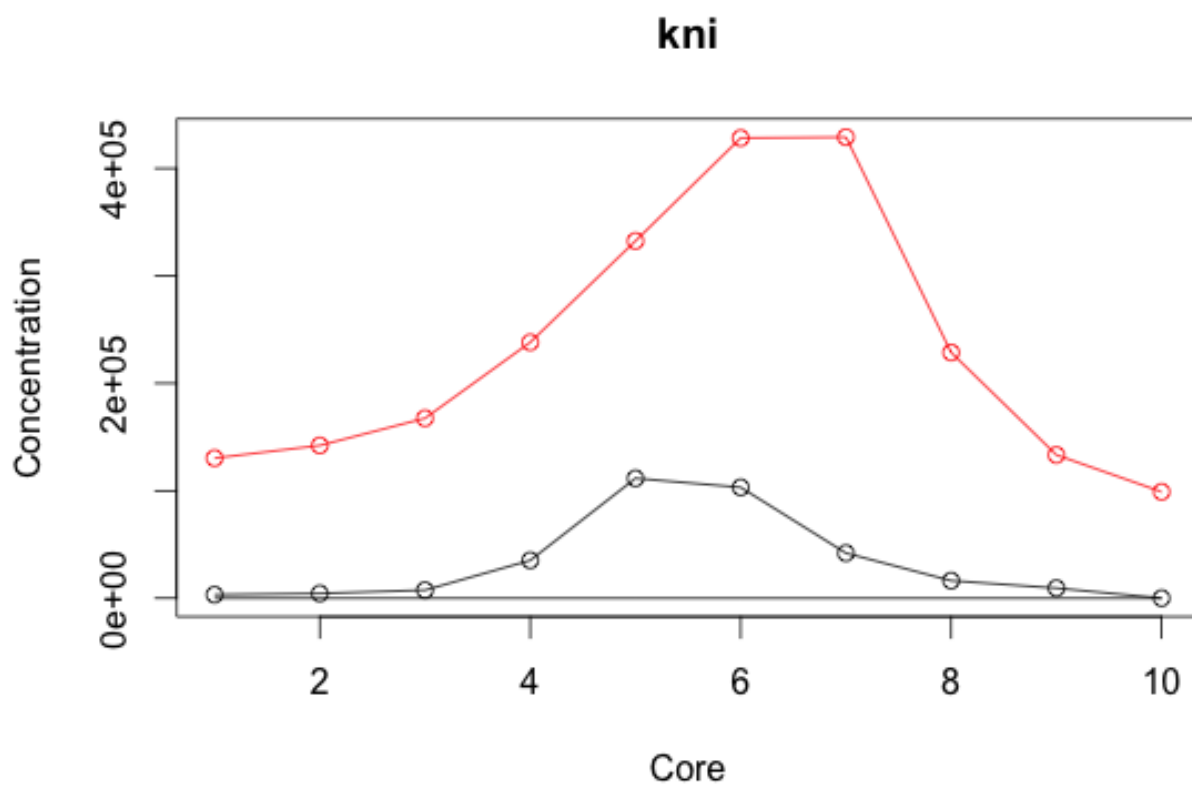


Рисунок 4 – Зависимость концентрации белка kni от ядра. Чёрным цветом - результат моделирования, красный цвет - экспериментальный расчёты

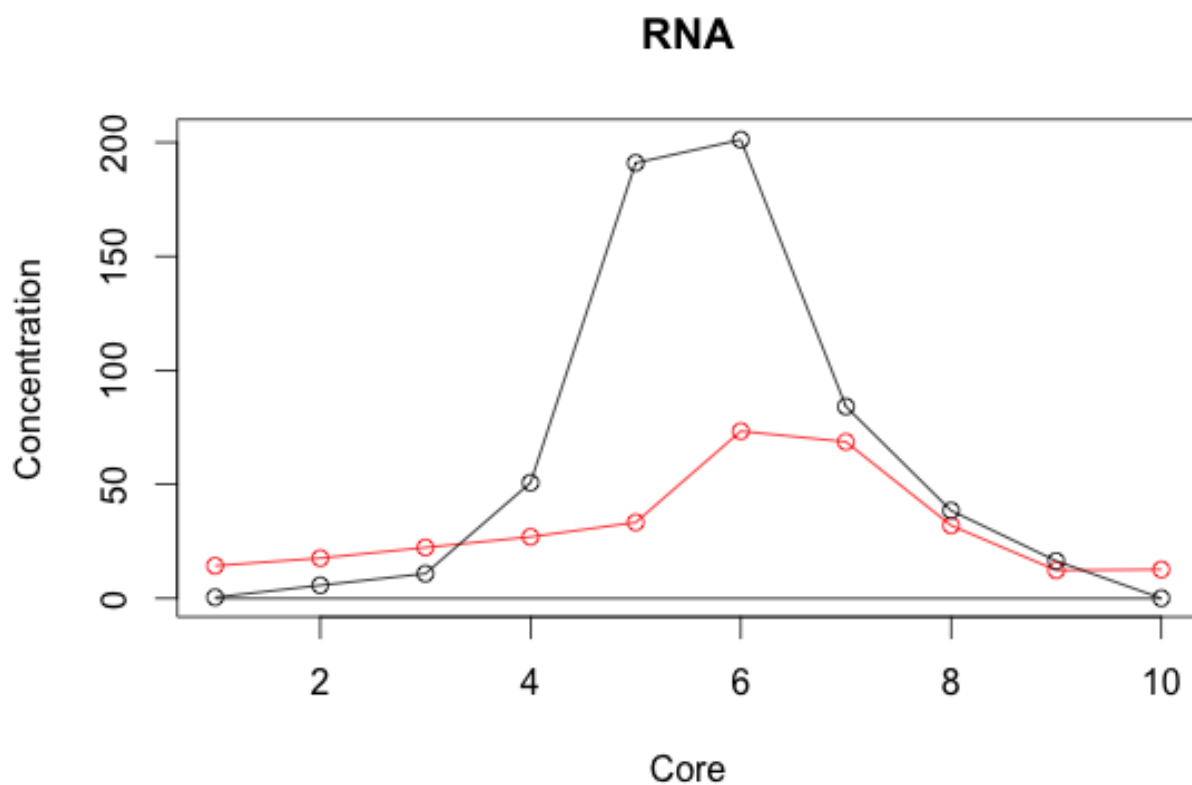


Рисунок 5 – Зависимость концентрации мРНК от ядра. Чёрным цветом - результат моделирования, красный цвет - экспериментальные расчёты

## ВЫВОДЫ

Основной сложностью моделирования является вычислительная сложность, так как в сумме получается нужно рассчитывать вероятности для 12404 реакций. Но даже при малом времени работы DEEP удалось подобрать параметры модели, чтобы концентрации белка **kni** и мРНК имели схожее поведение с экспериментальными расчётами.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 FERN – a Java framework for stochastic simulation and evaluation of reaction networks. — <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-356>. — Accessed: 2016-30-04.
- 2 StochKit. — <https://sourceforge.net/projects/stochkit/>. — Accessed: 2016-30-04.
- 3 StochPy. — <http://stochpy.sourceforge.net>. — Accessed: 2016-30-04.
- 4 *Kozlov K., Samsonov A.* DEEP—differential evolution entirely parallel method for gene regulatory networks. — <http://link.springer.com/article/10.1007/s11227-010-0390-6>. — Accessed: 2016-30-04.