

# ОТЧЕТ О ПРОДЕЛАННОЙ РАБОТЕ С ИСПОЛЬЗОВАНИЕМ ОБОРУДОВАНИЯ ИВЦ НГУ

## **1. Аннотация**

Эпистаз – феномен, при котором эффект мутации гена определяется генетическим контекстом. В процессе эволюции белков также наблюдается сходный феномен: известно, что влияние одиночных замен аминокислот на стабильность и функцию белка (и, следовательно, их вклад в приспособленность организма) не всегда является аддитивным.

Одна из математических моделей, которая позволяет описать влияние контекста на эффект от замены аминокислоты в белке основана на понятии гиперкуба в многомерном пространстве, в котором вершинами являются последовательности белка, а ребра соответствуют одиночным аминокислотным заменам. Такая модель реализована в программе HypercubeME [1]. Она позволяет на основе модели гиперкуба и набора последовательностей мутантных белков оценить число последовательностей, в которых может наблюдаться неаддитивный эффект мутаций.

Настоящая работа посвящена оценке влияния контекста на изменение стабильности белков в результате аминокислотных замен для последовательностей семейств гемоглобинов и миоглобинов позвоночных, а также белков вируса гриппа А – гемагглютининов, нейраминидаз и нуклеопротеинов на основе использования алгоритма HypercubeME.

## **2. Тема работы**

Оценка эволюционной значимости множественных эпистатических взаимодействий сайтов белков

## **3. Состав коллектива**

1. Афонников Дмитрий Аркадьевич, к.б.н., ведущий научный сотрудник, Институт цитологии и генетики СО РАН, заведующий лабораторией эволюционной биоинформатики и теоретической генетики
2. Вишневский Олег Владимирович, научный сотрудник, Институт цитологии и генетики СО РАН
3. Генаев Михаил Александрович, старший научный сотрудник, Институт цитологии и генетики СО РАН

4. Дорошков Алексей Владимирович, научный сотрудник, Институт цитологии и генетики СО РАН
5. Зубаирова Ульяна Станиславовна, научный сотрудник, Институт цитологии и генетики СО РАН
6. Игнатьева Елена Васильевна, старший научный сотрудник, Институт цитологии и генетики СО РАН
7. Кривенко Ольга Валериевна, старший научный сотрудник, Институт цитологии и генетики СО РАН
8. Левицкий Виктор Георгиевич, старший научный сотрудник, Институт цитологии и генетики СО РАН
9. Ощепков Дмитрий Юрьевич, старший научный сотрудник, Институт цитологии и генетики СО РАН
10. Подколотная Ольга Александровна, старший научный сотрудник, Институт цитологии и генетики СО РАН
11. Пономаренко Михаил Павлович, ведущий научный сотрудник, Институт цитологии и генетики СО РАН
12. Суслов Валентин Владимирович, научный сотрудник, Институт цитологии и генетики СО РАН
13. Турнаев Игорь Иванович, младший научный сотрудник, Институт цитологии и генетики СО РАН
14. Фомин Эдуард Станиславович, старший научный сотрудник, Институт цитологии и генетики СО РАН
15. Чадаева Ирина Витальевна, младший научный сотрудник, Институт цитологии и генетики СО РАН

#### **4. Научное содержание работы**

##### **4.1. Постановка задачи**

Данный проект является дипломной работой, который выполнялся в лаборатории эволюционной биоинформатики и теоретической генетики под руководством Дмитрия Аркадьевича Афонникова. Работа основывается на моделировании процесса эволюции белковых последовательностей с разным режимом эволюции с последующей оценкой эволюционной значимости белков с наблюдаемыми заменами. Оценка значимости рассчитывалась при помощи дополнительных программ.

#### 4.2. Подробное описание работы, включая используемые алгоритмы

Для моделирования эволюции белковых последовательностей глобиновых и вирусных белков необходимо иметь полный набор мутантных последовательностей. Этот набор можно было получить в программе HypercubeME [1], куда на вход подавались множественные выравнивания аминокислотных последовательностей. Выравнивания проводились в программе MUSCLE. Предварительно последовательности белков были взяты из базы данных BLAST nr и Influenza Research Database, отфильтрованы от повторов, коротких последовательностей, последовательностей с низким процентом покрытия (coverage). После получения наборов мутаций, составляющих многомерные гиперкубы, в программе HypercubeME, была проведена оценка стабильности белковых структур через изменение свободной энергии фолдинга в программе FoldX. На вход в программу FoldX подавались трехмерные структуры белков дикого типа, сгенерированных с помощью скрипта на языке Python, а также списки наблюдаемых мутаций. После оценки стабильности структур рассчитывалась средняя величина свободной энергии на одну замену, и полученные значения сравнивались для множественных мутаций и сумм единичных мутаций, которые составляли наблюдаемые множественные мутации.

#### 4.3. Полученные результаты

В результате работы были получены гиперкубы для глобинов с максимальной размерностью 2 и гиперкубы для белков вируса гриппа А с максимальной размерностью 3 (см. Таб 1).

Таб 1. Количество гиперкубов для разных белковых семейств

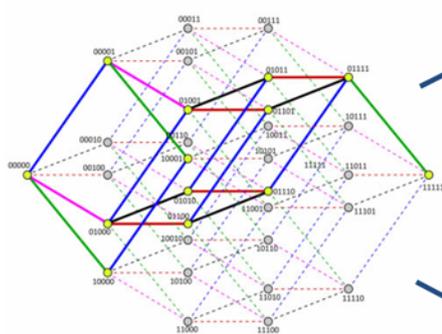
Названия белков	Количество гиперкубов порядка n		
	n = 1	n = 2	n = 3
Глобины	593	13	-
Гемагглютинины	23 157	1934	10

Нейраминидазы	19 462	1565	6
Нуклеопротеины	11 992	2499	83

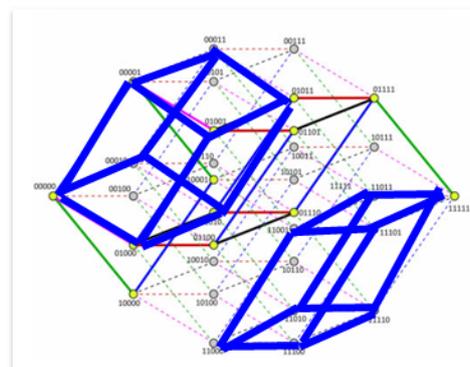
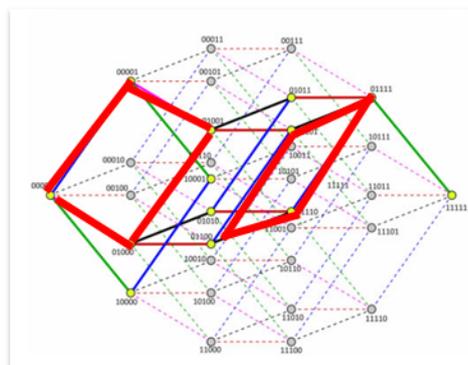
Визуально эти результаты можно представить следующим образом (см. Рис 1).

Для глобинов максимальный порядок составил 2

## МАКСИМАЛЬНЫЙ ПОРЯДОК ГИПЕРКУБОВ



Максимальный порядок гиперкуба для белков с искусственной селекцией: 12  
Количество гиперкубов **199 847 053** (Esteban et al., 2019)



Для вирусов максимальный порядок составил 3

Рис 1. Сравнение размеров гиперкубов для экспериментальных данных из статьи Esteban et al., 2019 и данных, полученных в нашем исследовании (выделены красными и синими линиями).

Примечательно то, что больше всего гиперкубов третьего порядка было найдено для нуклеопротеинов вируса гриппа А, что может быть связано с нейтральным отбором на них со стороны иммунной системы. Эту гипотезу мы планируем проверить в дальнейшей работе с помощью анализа других вирусных белков, в частности вируса гепатита С, ВИЧ и SARS-CoV-2.

Оценка стабильности белковых структур не выявила неаддитивного эффекта для множественных замен и соответствующих сумм единичных замен. Необходимы дополнительные исследования параметров, с помощью которых можно оценить эволюционную значимость замен или вклад контекста в приспособленность белка.

Исследование будет продолжаться в период после защиты дипломной работы, а также в магистратуре Сколковского института науки и технологий.

1. Esteban L. A. et al. HypercubeME: two hundred million combinatorially complete datasets from a single experiment //Bioinformatics. – 2020 – Т. 36. – №. 6. – С. 1960-1962.

## **5. Эффект от использования кластера в достижении целей работы**

Вся работа, за исключением статистического анализа, проводилась на кластере ИВЦ НГУ ввиду того, что для обработки сотен тысяч аминокислотных последовательностей, а также моделирование трехмерных структур белков необходимы большие вычислительные мощности. На персональном компьютере невозможно было бы провести вычисления с такой скоростью, с какой они производились на кластере. Таким образом, работа на кластере ИВЦ НГУ повысила эффективность работы на много порядков.