

**Московский Государственный Университет им. М.В.Ломоносова**

**Факультет биоинженерии и биоинформатики**

# **FoldIt**

**Работа студента 4-ого курса**

**Трофимова Дениса**

Одними из основных объектов изучения биоинформатики являются белки – макромолекулы, в основе которых лежат пептидные цепочки аминокислотных остатков. Функции белков предопределяет пространственное положение цепочек; знание о структуре позволяет понять принципы работы и потенциальные механизмы управления или дизайна.

Существуют два принципа определения структуры белка: экспериментальный и вычислительный. Большинство известных структур получено экспериментальными методами, в основном, рентгеноструктурного анализа (РСА) и ядерно-магнитного резонанса (ЯМР). Несмотря на развитие технологий и автоматизацию процессов, в некоторых случаях не удается получить приемлемые результаты (нельзя закристаллизовать белок или получение неточных структур).

Вычислительные методы получения структуры можно разделить на построение по гомологии с другим белком и прогнозирование на основе знания лишь первичной последовательности. И если решение первым методом неотделимо от априори известных экспериментальных данных, что довольно сильно ограничивает в применении задач, то построением *de novo* теоретически можно получить пространственные структуры подавляющего большинства белков. Проблема заключается в сложности разработки алгоритмов методов *de novo*.

В настоящее время существуют несколько программ, которые могут с хорошей вероятностью предсказать укладку небольших, однодоменных белков в пределах 1,5 Å по всей структуре. Одно из таких успешных решений представлено в программе Rosetta: поиск путем конформационного перебора каждого сегмента структур, имеющих минимум по энтропии и свободной энергии белка. Методы *de novo*, как правило, требуют больших огромных вычислительных ресурсов и именно по этой причине используются только для относительно небольших белковых молекул. К сожалению, мощность имеющихся на сегодняшний день вычислительных ресурсов не может удовлетворять решению всех задач подобного рода, поэтому создатели выше упомянутой Rosetta решили взять курс на человеческое пространственное мышление и интуицию, предложив эту задачу для решения людям и создав программу FoldIt.

### **Foldit**

Foldit - это игра-головоломка, цель которой получить точную модель белковой структуры, сложенной наилучшим образом (рис.1). Сначала перед игроками возникают «тренировочные структуры», немного неправильно сложенные структуры, которые нужно в интерактивном режиме корректировать так, чтобы набрать наивысший балл (минимум энергии). При этом текущее состояние отображается, а также результаты других игроков и групп игроков (рис.1, 8-9). Чтобы игра была понятна людям, не имеющим дело с научной терминологией, многие биологические термины заменены на более общие понятия. Также создателями были спрятаны некоторые элементы, которые препятствовали бы более явному решению проблемы и загромождали собой пространство, и сделали возможным показывать области, которые получаются в ходе не самого удачного фолдинга, с тем, чтобы игрок обратил на них внимание и попытался улучшить положение дела (рис.1, 1-5). Есть специальные визуальные подсказки, отображающие гидрофобность (“незащищенная гидрофобность”), межатомные отталкивания (“столкновение”) и образующиеся полости (“пустота”). Игроки интуитивно манипулируют представленными инструментами. Наиболее быстрый способ манипуляции – непосредственно потянуть за любую часть белка. Это также возможно для поворота спиралей и стягивания бета-листов (“tweak”). Игроки также могут наложить небольшие ограничения на подвижность молекулы (“rubber bands”) и зафиксировать ее (“freezing”) (Рис.1, 6-7).

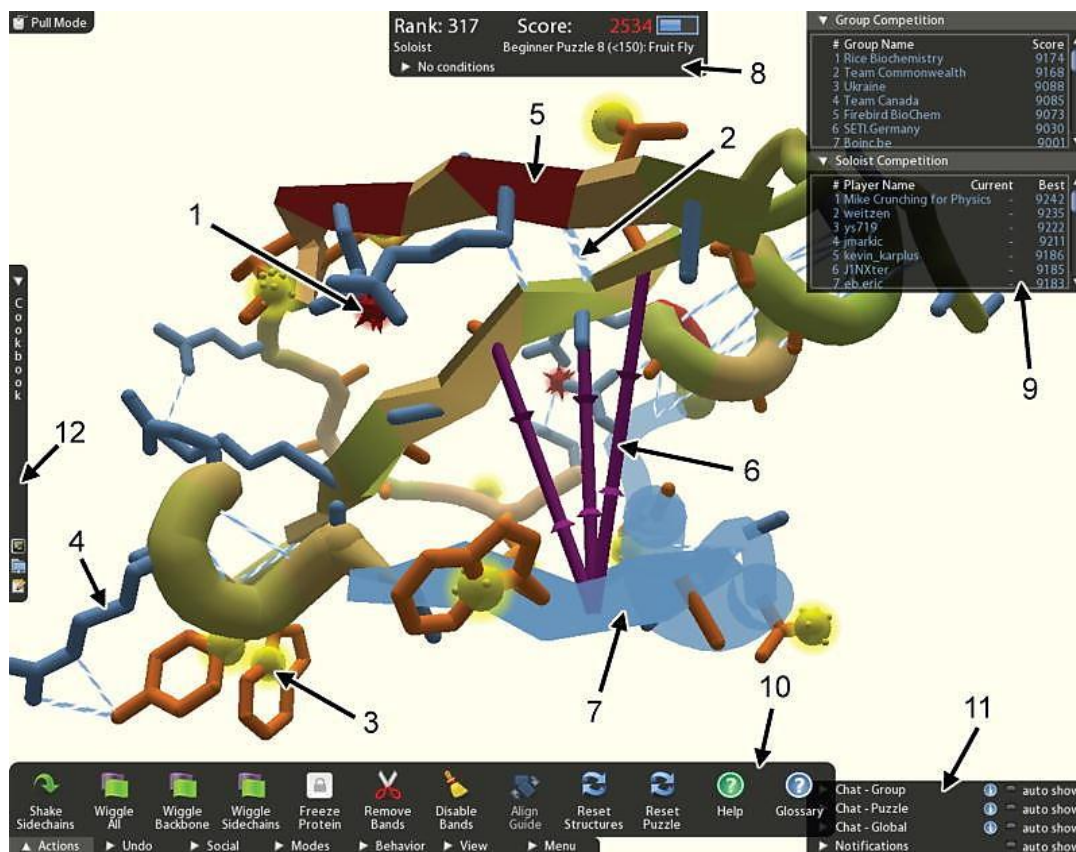


Рис.1. Скриншот FoldIt, показывающий инструменты и отображение. 1 – визуализация атомов, которые слишком близко располагаются, 2 – водородные связи, 3 – гидрофобная боковая цепь с желтыми вставками, где есть незащищенные группы, 4 – гидрофильная боковая цепь, 5 – сегмент остова, окрашенного в красный цвет из-за высокой энергии остатков, 6 – ограничение подвижности молекулы, 7 – «замораживание» молекулы, 8 – набранные очки, 9 – лидеры, 10 – панель инструментов, 11 – чат для общения с другими игроками, 12 – автоматизированные инструменты/ «рецепты».

Доступные также и автоматические инструменты: изменение положения ротамеров в боковой цепи путем комбинаторных переборов ("shake"), локальная градиентная минимизация энергии ("wiggle"), вставка фрагмента ("rebuild"). Более того, разработаны скрипты на языке Lua - «рецепты» (recipes) - как создателями игры, так и собственно игроками, которые значительно упрощают жизнь (рис. 2). Этими алгоритмами можно делиться с другими игроками или подгрузить себе в FoldIt в «кулинарную книгу» ("cookbook"). Некоторые рецепты фактически повторяют уже известные решения: так например Blue Fuse, один из наиболее часто используемых скриптов, очень похож на на протокол работы скрипта Fast Relax для Rosetta (рис. 3)

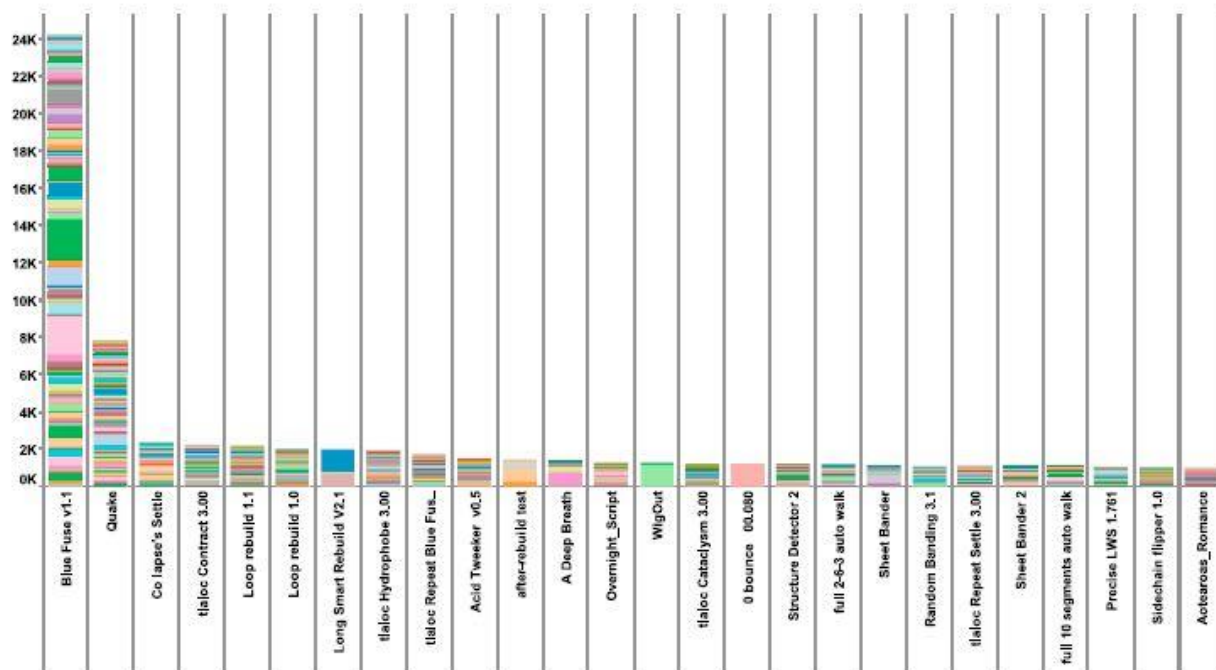


Рис.2. Наиболее часто используемые рецепты FoldIt. Каждый столбец представляет отдельный рецепт. По вертикали показана частота использования. Цветами обозначены разные игроки. Blue Fuse является самым популярным рецептом

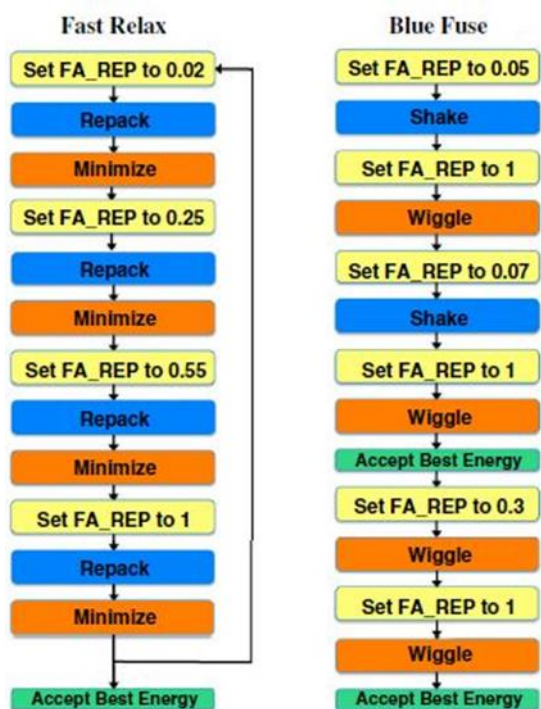


Рис.3. Сходство протокола Fast Relax для Rosetta и Blue Fuse для FoldIt

Одна из наиболее важных задач, решенных игроками FoldIt, - разрешение структуры ретровирусной протеазы обезьяньего вируса Мэйсона-Пфайзера (M-PMV). M-PMV вызывает обезьяний СПИД (SAIDS) у обезьян рода *Macaca*. Ретровирусные протеазы (PR) играют ключевую роль в созревании вируса и пролиферации, и находятся в центре внимания разработок антиретровирусных лекарственных препаратов. Биофизические исследования и ЯМР показывали, что при отсутствии субстратов или ингибиторов M-PMV протеазы должны складываться в стабильный мономер, но кристаллическая структура не могла быть разрешена с помощью

молекулярного замещения в течение нескольких десятилетий, даже при наличии гомологов. С целью определить, может ли интуиция человека обойти автоматизированные системы, игрокам FoldIt был брошен вызов: построить точную модель M-PMV PR, основываясь на координатах, полученных методом ЯМР. По окончании трехнедельного соревнования, были проанализированы модели, набравшие наибольшее количество очков. Примечательно то, что несколько решений FoldIt

оказались весьма хорошими, чтобы в дальнейшем позволить разрешить кристаллическую структуру

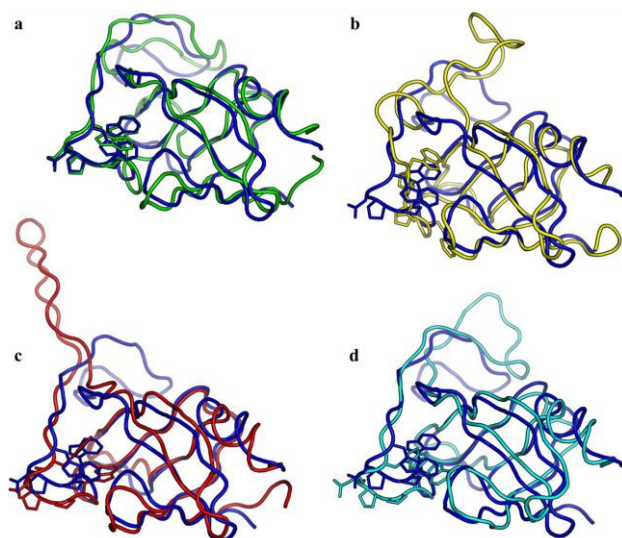


Рис.4. Лучшие попытки разрешения структуры M-PMV PR. Позже определенная кристаллическая структура представлена синим на всех 4 картинках. (a) в зеленом цвете: предсказание FoldIt, следует отметить точное совпадение боковой цепи (b) в желтом цвете: предсказание Rosetta с учетом перестройки-уточнения структуры (c) в красном: предсказание Rosetta с учетом релаксации (d) в голубом: предсказание Rosetta используя CS-Rosetta (система химического сдвига на основе Rosetta).

Далее игрокам были предложены 10 разных структур от ЯМР, которые были плохо описаны при использовании функцией энергии Rosetta. Улучшения в точности моделей представлены на рисунке 5. Процесс уточнения структуры в ходе игры занял 16 дней. Решение при определенном молекулярном замещении считалось правильным, если модель имела функцию правдоподобия больше, чем в случае случайных моделей.

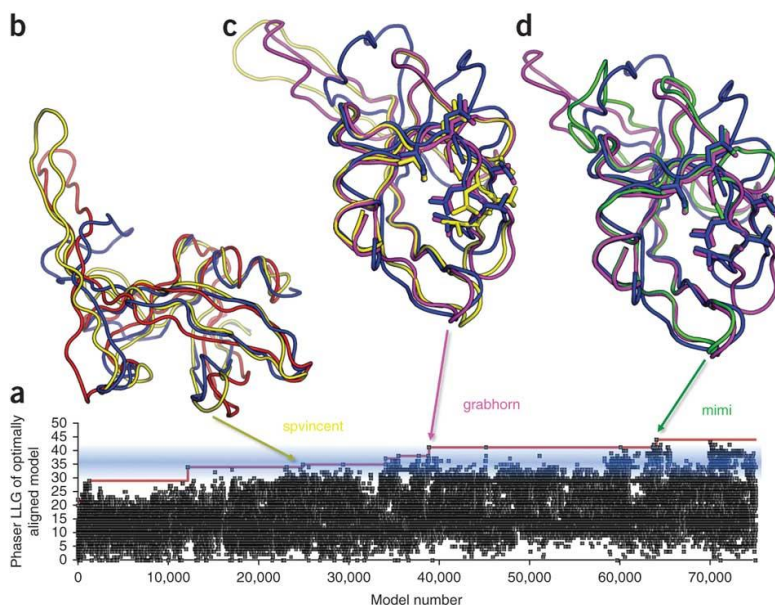


Рис.5. Улучшение структуры M-PMV PR при участии игроков FoldIt (a) процесс уточнения структуры в течение 16 дней игры. Ось X показывает прогресс во времени. Ось Y показывает логарифмическую функцию правдоподобия для каждой модели. Чтобы считать решение

*правильным при молекулярном замещении, модель должна иметь логарифмическую функцию правдоподобия больше, чем лучшие случайные модели. Распределения по интенсивности этих случайно выбранных моделей показаны бледно-синей полосой (b) начиная с весьма уточненной ЯМР модели (красный цвет), сгенерированная модель одним из игроков (желтый цвет) ближе к позже полученной кристаллической структуре (синий цвет) (c) начиная с уточненной модели одним из игроков, другой игрок получил свою модель (розовый цвет) (d) третий игрок вносит окончательные поправки и получает свою модель (зеленый цвет) со значительным уточнением для обеспечения однозначного решения молекулярного замещения.*

Функция правдоподобия полученной модели была достаточно высокой, чтобы говорить, что данное решение - лучшее из множества других: полученных из FoldIt, автоматически полученных, уточненных другими способами. Используя решение FoldIt, структура была окончательно определена в течение нескольких дней после этого.

Решающая роль игроков FoldIt в определении структуры M-PMV PR показывает всю мощь онлайн-игр и возможности человеческой интуиции в решении сложнейших научных задач.

## Список литературы

1. Good BM, Su AI. (2011); Games with a scientific purpose; *Genome Biology*; 12:135
2. Khatib F, Cooper S, Tyka MD, Xu K, Makedon I, Popovic Z, Baker D, Players F. (2011); Algorithm discovery by protein folding game players.; *Proc Natl Acad Sci U S A*. 108(47):18949-53
3. Cooper S, Khatib F, Treuille A, Barbero J, Lee J, Beenen M, Leaver-Fay A, Baker D, Popović Z, Players F. (2010); Predicting protein structures with a multiplayer online game. *Nature*; 466(7307):756-60.
4. Gilski M, Kazmierczyk M, Krzywda S, Ząbranska H, Cooper S, Popović Z, Khatib F, DiMaio F, Thompson J, Baker D, Pichová I, Jaskolski M.(2011); High-resolution structure of a retroviral protease folded as a monomer. *Acta Crystallogr. D. Biol. Crystallogr.* 67(Pt 11):907-14
5. Khatib F, DiMaio F; Foldit Contenders Group; Foldit Void Crushers Group, Cooper S, Kazmierczyk M, Gilski M, Krzywda S, Zabranska H, Pichova I, Thompson J, Popović Z, Jaskolski M, Baker D. (2011); Crystal structure of a monomeric retroviral protease solved by protein folding game players.; *Nat Struct Mol Biol*, 8(10):1175-7. doi: 10.1038/nsmb.2119