

в то время как пары оснований АТ имеют только 2. ДНК с более высоким содержанием GC из-за ее источника или из-за SNP будет иметь более высокую температуру плавления, чем ДНК с более высоким содержанием АТ.

Эта информация также дает важные сведения о способе взаимодействия молекулы с ДНК. Молекулы, такие как интеркаляторы, располагаются между парами оснований и взаимодействуют посредством укладки. Это оказывает стабилизирующее действие на структуру ДНК, что приводит к повышению температуры ее плавления. Точно так же увеличение концентрации солей помогает рассеять отрицательное отталкивание между фосфатами в остове ДНК. Это также приводит к повышению температуры ее плавления. И наоборот, рН может оказывать негативное влияние на стабильность ДНК, что может привести к снижению ее температуры плавления.

Анализ кривой плавления имеет несколько ограничений. Абсолютное положение и ширина кривых плавления зависят от концентрации красителя, температуры и скорости перехода. Добавление интеркаляторов, таких, как бромид этидия, повышает температуру плавления и уширяет переход плавления.

**Заключение.** Таким образом, высокочувствительный анализ кривых плавления – это относительно новый метод анализа, позволяющий выявлять новые варианты генов, делать скрининг образцов ДНК на однонуклеотидные полиморфизмы, выявлять вставки и делеции и другие неизвестные мутации и определять процент метилированной ДНК в образцах. Однако при анализе гетерозиготных вариантов кривые плавления могут быть трудно различимы, поэтому в таких случаях для анализа используют другие графические зависимости, такие, как графики пиков плавления и сравнения кривых плавления.

Точность алгоритмов зависит от многих факторов, поэтому при анализе результатов кривых плавления не стоит полагаться лишь на один вариант детектирования. В рамках научных исследований в МГЭИ им. А. Д. Сахарова БГУ ведется разработка программных модулей, реализующих некоторые алгоритмы кластеризации, и позволяющих с их помощью провести соответствующий анализ [1–3].

#### ЛИТЕРАТУРА

1. *Jain, A. Data Clustering: A Review / A. Jain, M. Murty, P. Flynn // ACM Computing Surveys, 1999.31(3). P. 264–323.*
2. *Ririe, K. Product Differentiation by Analysis of DNA Melting Curves during the Polymerase Chain Reaction / K. Ririe, R. Rasmussen, C. Witter // Anal. Biochem. 1997.252(2). P. 857–859.*
3. *Wittwer, C. T. High-resolution DNA melting analysis: advancements and limitations / C. T. Witter // Human Mutation, 2009.30(6). P. 857–859.*

## ПРОЕКТ ПРОГРАММНОЙ СИСТЕМЫ ДЛЯ АНАЛИЗА ВЗАИМОДЕЙСТВИЯ БЕЛКОВ В УСЛОВИЯХ СЛАБОГО СТРУКТУРНОГО ПОДОБИЯ SOFTWARE SYSTEM PROJECT FOR PROTEINS' INTERACTION ANALYSIS UNDER CONDITIONS OF WEAK STRUCTURAL SIMILARITY

***А. Д. Казмерчук<sup>1,2</sup>, Е. В. Снытков<sup>1,2</sup>, Б. А. Тонконогов<sup>1,2</sup>***

***A. D. Kazmerchuk<sup>1,2</sup>, E. V. Snytkov<sup>1,2</sup>, B. A. Tonkonogov<sup>1,2</sup>***

*<sup>1</sup>Белорусский государственный университет, БГУ*

*г. Минск, Республика Беларусь*

*<sup>2</sup>Учреждение образования «Международный государственный экологический институт имени А. Д. Сахарова» Белорусского государственного университета, МГЭИ им. А. Д. Сахарова БГУ*

*г. Минск, Республика Беларусь*

*antkaz566@gmail.com, evgeni.snytkov@iseu.by*

*<sup>1</sup>Belarusian State University, BSU*

*Minsk, Republic of Belarus*

*<sup>2</sup>International Sakharov Environmental Institute of Belarusian State University, ISEI BSU*

*Minsk, Republic of Belarus*

Рассмотрены практическое назначение, алгоритм функционирования, архитектура, технологии и средства реализации, функциональность и тестирование проекта программной системы для анализа взаимодействия белков в условиях слабого структурного подобия, базирующейся на биоинформационных технологиях, реализующей расчетные методы и математические модели на различных вычислительных уровнях и позволяющей принимать исследовательские решения в различных областях биологии и медицины.

Practical purpose, functioning algorithm, architecture, technologies and means of implementation, functionality and testing of software system project for proteins' interaction analysis under conditions of weak structural similarity, based on bioinformatics technologies, implementing computational methods and mathematical models at

various computational levels and allowing making research decisions in various fields of biology and medicine are considered.

*Ключевые слова:* программная система, взаимодействие белков, структурное подобие.

*Keywords:* software system, proteins' interaction, structural similarity.

<https://doi.org/10.46646/SAKH-2022-2-419-422>

**Введение.** Из-за большого количества видов белков (протеинов) в биологических системах и дорогостоящих экспериментальных решений, эффективное решение многокритериальной задачи докинга (метода молекулярного моделирования, позволяющего предсказать наиболее выгодную для образования устойчивого комплекса ориентацию и конформацию одной молекулы в сайте связывания другой) возможно только с применением информационных технологий и средств автоматизации, реализующих специальные расчетные методы и математические модели взаимодействия протеинов.

Белково-пептидные взаимодействия играют важную роль в различных биологических процессах, таких как передача сигналов, клеточная локализация и межклеточное взаимодействие, иммунная система, апоптоз и так далее. Фармакологические препараты, основанные на пептидах, могут быть использованы для предотвращения заболеваний, связанных со сбоями в работе белков из-за нежелательных межбелковых взаимодействий. Многие базы данных и алгоритмы были разработаны специально для использования в области терапии на основе пептидов. Существует более 200 терапевтических пептидов, одобренных Управлением по санитарному надзору за качеством пищевых продуктов и медикаментов (США), для лечения различных заболеваний.

Докинг позволяет идентифицировать новые соединения, представляющие терапевтический интерес, предсказывать взаимодействие лиганд-мишень на молекулярном уровне или определять взаимосвязь «структура-активность», не владея информацией априори о химической структуре других модуляторов-мишеней. Хотя изначально он был разработан, чтобы помочь понять механизмы молекулярного распознавания между малыми и большими молекулами, использование и применение докинга при разработке, например, лекарств сильно изменились за последние годы. Возможности молекулярного докинга при разработке лекарств хорошо известны. Тем не менее, стыковка имеет внутренние ограничения, которые снижают его производительность прогнозирования. Хотя докинг в основном использовался как самостоятельный метод для разработки лекарств, теперь он часто интегрируется в рабочие процессы, которые включают другие вычислительные методы, такие как подходы на основе лигандов и структур и искусственный интеллект (рис. 1).

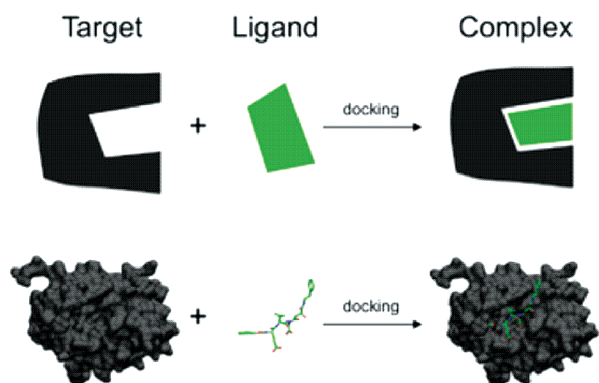


Рисунок 1 – Визуальная интерпретация процесса докинга

Подходы, основанные на структуре, такие как молекулярная динамика и оценка свободной энергии связывания, широко используются в сочетании с докингом для улучшения результатов виртуального скрининга. В частности, молекулярная динамика позволяет оценить гибкость остатков в целевом сайте связывания, а также исследовать более крупные конформационные изменения, потенциально доступные для данного белка. Следовательно, такой подход представляет собой эффективный инструмент для идентификации конформаций рецепторов для докинга и для оценки стабильности предсказанных комплексов. Возможности, предлагаемые молекулярной динамикой в перспективном скрининге *in silico*, особенно привлекательны для гибких мишеней с ограниченным числом зарегистрированных кристаллографических конформаций.

Более продвинутые и дорогостоящие в вычислительном отношении методы прогнозирования возмущения свободной энергии, термодинамического интегрирования и воронкообразной метадинамики также могут использоваться для молекулярного докинга и постобработки его результатов. Некоторые методы при прогнозировании средства «лиганд-белок» требуют больших вычислительных ресурсов, поэтому потенциально менее подходят для скрининга больших библиотек соединений.

Совсем недавно подходы статистики и машинного обучения также закрепились в изучении лекарственных веществ. Фактически, эти методы позволяют легко использовать постоянно растущий источник информации,

содержащийся в общедоступных базах данных по структурам, химическим составам и биологическим активностям, что приводит к более точным прогнозам аффинности связывания. В частности, подходы машинного обучения, включая «случайный лес» и машины опорных векторов, применялись для улучшения прогнозов аффинности связывания на основе стыковки. Подходы машинного обучения также использовались для улучшения функций оценки докинга.

На основе анализа известных разработок в исследуемой предметной области, связанной с автоматизированной информационно-аналитической оценкой взаимодействия двух протеинов, предложен обобщенный подход к разработке методов анализа взаимодействия двух протеинов, включающий 3 стадии:

- 1) математическое описание анализируемых физико-химических процессов при взаимодействии двух протеинов;
- 2) проведение комплексного расчета начальной позиции взаимодействия двух протеинов, согласно некоторому гипотетическому допущению, при котором гипотеза может выходить за рамки доступных знаний о биохимической структуре протеинов;
- 3) уточнение результатов расчета значения энергетической функции взаимодействующего протеин-протеинового комплекса на основе некоторого метаэвристического метода, основанного на высокоуровневых проблемно-независимых алгоритмических структурах и избираемого некоторым осознанным или игровым способом с возможностью полного перебора всех доступных метаэвристик в пределах метаэвристического фреймворка.

Проект интегрированной информационной системы, предназначенной для проведения компьютерного анализа возможного взаимодействия двух протеинов в условиях слабого структурного подобия, создан на основании разработанных новых методов и специальных алгоритмов, программная реализация которых позволит обеспечить платформенно-независимый доступ к системе из различных вычислительных сред, включая как серверные вычислительные средства, так и встраиваемые системы и мобильные устройства.

В процессе создания указанной системы выполняются следующие виды работ:

- поиск и оценивание доступных гипотез о взаимодействии протеинов, включая метафорические гипотезы близкие по значению к понятию метаэвристик, то есть гипотез, которые научно не подтверждены, но, по крайней мере, не хуже, чем случайное начальное расположение протеинов для докинга согласно их информационной характеристике;
- создание расширяемого модуля с объединенным интерфейсом для аккомодации всех возможных типов гипотез, в том числе и неучтенных на текущий момент;
- поиск доступных для программной реализации метаэвристик в опубликованных информационных источниках, а также оценка их состоятельности;
- программная реализация метаэвристик как отдельно оформленных программных библиотек;
- разработка метаэвристической программной оболочки для аккомодации созданных метаэвристик;
- разработка архитектуры и алгоритмов работы программных модулей для компьютерной реализации информационной системы, включая интеграцию метаэвристической программной оболочки в финальную библиотеку, с учетом стандартизированных интерфейсов с целью бесперебойного функционирования программных модулей;
- выбор инструментальных средств разработки компонентов информационной системы, включающих в себя средства тестирования программного кода;
- реализация программных модулей информационной системы и создание консольного приложения или программной библиотеки, представляющей из себя ядро системы;
- в выполнение программного тестирования и апробации информационной системы.

**Практическое назначение и алгоритм функционирования интегрированной информационной системы.** Разрабатываемая интегрированная информационная система может быть апробирована и практически применена в интересах учреждений Министерства здравоохранения Республики Беларусь и Национальной академии наук Беларуси для поддержки принятия исследовательских решений в различных областях биологии и медицины, к примеру, дизайна лекарств, исследования протеинопатий и летальных губчатых энцефалопатий.

Алгоритм разрабатываемой системы базируется на двух понятиях: спекуляции и метаэвристическом фреймворке. Под спекуляцией в системе понимается расширяемая совокупность спекулятивных гипотез о природе взаимодействия протеинов (эволюционные спекуляции, гипотеза Анфинсена, гипотеза схожести вторичных структур и иные), которая генерирует за детерминированное время первоначальное взаимное расположение протеинов, которое может быть как неточным, так и сразу удовлетворить условию останки, из-за чего алгоритм завершит работу.

Следует заметить, что спекулятивность гипотез подразумевает тот факт, что, несмотря на то, что в биологии существует достаточно недоказанных гипотез о взаимодействии протеинов, рассматриваемые гипотезы не обязаны быть строго биологическими, потому что проблема докинга протеинов по своей сути является комбинаторной проблемой, а значит, не имеет стандартного детерминированного решения. Это подтверждается, к примеру, описанными нарушениями гипотезы Анфинсена и гипотезами передачи прионных взаимодействий. Основным критерием алгоритма спекулятивной гипотезы, таким образом, является детерминированное время работы отдельно взятой спекуляции.

Если условие останки не достигнуто, то алгоритм запускает метаэвристический фреймворк, представляющий из себя расширяемое множество метаэвристик (высокоуровневых проблемно-независимых алгоритмических структур, предоставляющих набор рекомендаций или стратегий для разработки алгоритмов эвристической оптимизации), и случайным образом последовательно начинает применять метаэвристики из данного множе-

ства. Далее алгоритм остановится либо в случае достижения условия остановки, либо в случае перебора всех доступных метаэвристик, и в данном случае ответом является минимум энергии по множеству всех доступных метаэвристик.

По сравнению с алгоритмами оптимизации и итерационными методами, метаэвристики не гарантируют, что глобально оптимальное решение может быть найдено для некоторого класса проблем. Многие метаэвристики реализуют ту или иную форму стохастической оптимизации, так что найденное решение зависит от сгенерированного набора случайных величин. В комбинаторной оптимизации путем поиска по большому набору возможных решений метаэвристики часто могут найти хорошие решения с меньшими вычислительными затратами, чем алгоритмы оптимизации, итерационные методы или простые эвристики. По сути, они полезны для решения задач оптимизации и отвечают ряду требований:

- способны рассчитывать решения задачи, которые являются очень близкими к оптимальным;
- достаточно просты для понимания, дальнейшей поддержки программного кода, его отладки и доработки;
- могут достаточно быстро вычислять решения задач.

Условием остановки алгоритма является заранее заданный уровень энергии протеин-протеинового комплекса. Применимость данного алгоритма допускается следствием из «теоремы об отсутствии бесплатных обедов», утверждающее, что при наличии конкретной оптимизационной задачи существует, по крайней мере, одна метаэвристика, которая даст результат лучше, чем остальные.

**Архитектура, технологии и средства реализации интегрированной информационной системы.** Архитектура системы представляет собой совокупность следующих расширяемых программных слоев:

- 1) слой парсеров – набор адаптеров для определенных типов файлов, содержащих информацию о трехмерной структуре протеина;
- 2) слой парсеров экспериментального докинга – набор адаптеров для определенных типов экспериментальных баз данных докинга, формирующий на выходе объект обобщенной структуры;
- 3) слой обобщенной структуры – методы работы с объектом как абстракцией взаимодействия двух протеинов, инициализация которой производится либо из двух объектов, представляющих из себя протеины, либо из результата парсинга записи об экспериментальном докинге, и которая содержит в себе метод подобия, позволяющий оценить схожесть двух результатов докинга;
- 4) слой спекуляций – спекулятивные детерминированные методы, реализующие недоказанные предположения о природе докинга и оперирующие на обобщенных структурах;
- 5) слой метаэвристического фреймворка – метаэвристики, уточняющие результат докинга на выходе слоя спекуляций и оперирующие на обобщенных структурах.

Система целиком реализуется на языке JavaScript без применения сторонних библиотек, фреймворков и платформ. Система планируется быть совместимой с платформой Node.js, однако реализуется с учетом независимости от нее.

**Заключение.** Таким образом, задача протеин-протеинового докинга имеет важное значение для молекулярной биологии, физиологии и фармацевтики (разработки лекарств с определенными свойствами). На текущий момент все существующие серверы и алгоритмы докинга используют некоторую определенную широко известную метаэвристику и упрощение поверхности взаимодействующих протеинов, так как докинг является комбинаторной, а значит неполиномиальной задачей. В Республике Беларусь не имеется подобных собственных серверов. Авторами предлагается осуществить автоматизированный поиск поверхностей взаимодействующих протеинов, исходя из совокупности недоказанных научных гипотез о природе таких взаимодействий. Структурный докинг предлагается расширить до полноценного метаэвристического фреймворка вместо использования разрозненных метаэвристик [1–5].

#### ЛИТЕРАТУРА

1. *Siyavoush, D*, et al. *Methods and Algorithms for Molecular Docking-based Drug Discovery* / Tebriz University of Iran. Chapter 1. *Molecular Docking at a Glance*. – Penn.: IGI-Global, 2016.
2. *H3ABioNet\_Docking.pdf* [Электронный ресурс]. – Электрон. дан. – URL: [https://training.h3abionet.org/IBT\\_2016/wp-content/uploads/2016/09/H3ABioNet\\_Docking.pdf](https://training.h3abionet.org/IBT_2016/wp-content/uploads/2016/09/H3ABioNet_Docking.pdf), свободный. – Загл. с экрана.
3. *Wang, Ch*. Improving Scoring-Docking-Screening Powers of Protein-Ligand Scoring Functions using Random Forest / *Ch. Wang, Y. Zhang* // *J Comput Chem*. – 2017. – Vol. 38 (3). – P. 169–177. doi: 10.1002/jcc.24667.
4. *Protein-protein interaction databases - OpenWetWare* [Электронный ресурс]. – Электрон. дан. – URL: [https://openwetware.org/wiki/Protein-protein\\_interaction\\_databases](https://openwetware.org/wiki/Protein-protein_interaction_databases), свободный. – Загл. с экрана.
5. *Atomic Coordinate Entry Format Version 3.3* [Электронный ресурс]. – Электрон. дан. – URL: <http://www.wwpdb.org/documentation/file-format-content/format33/v3.3.html>, свободный. – Загл. с экрана.