

Биоинформатика для школьника на примере сравнительного анализа внутриклеточных протеолитических ферментов человека и лёгочных пресноводных моллюсков

А. А. Чиркин, профессор кафедры химии Витебского государственного университета имени П. М. Машерова, доктор биологических наук,

В. В. Долматова, аспирант кафедры химии Витебского государственного университета имени П. М. Машерова

Начало XXI века характеризуется катастрофически быстрым ростом объёмов биологической информации — «информационным взрывом», следствием которого явился «информационный кризис», когда человек испытывает огромные трудности при получении доступа к этой информации, а его базовое образование не позволяет её эффективно использовать. Это порождает профессиональную несостоятельность человека, поскольку сейчас «период полураспада компетентности» биологов, медиков, учителей и других специалистов составляет 4 года, а период обновления знаний — 6–8 лет [1]. Для поддержания компетентности в научных исследованиях и преподавании биологических дисциплин целесообразно использовать международные открытые компьютерные базы биологических знаний, в том числе крупнейшие хранилища первичных структур ДНК и аминокислотных последовательностей (такие как EMBL, GenBank, DDBJ, Uni-Prot, Ensembl и др.). Работа с такими базами данных требует владения технологиями биоинформатики. Эта наука применяет машинные алгоритмы и статистические методы к наборам биологических данных, состоящих, как правило, из большого числа ДНК, РНК и аминокислотных последовательностей. К ним относят: 1) математические методы компьютерного анализа в сравнительной геномике (геномная биоинформатика); 2) разработку алгоритмов и программ для предсказания пространственной структуры биополимеров (структурная биоинформатика); 3) описание молекул и механизмов внеклеточного и внутриклеточного сигналинга (метаболическая биоинформатика). Главная цель биоинформатики — способствовать пониманию биологических процессов. В прикладном смысле под биоинформатикой

понимают любое использование компьютеров для обработки биологической информации.

Целью данной статьи является описание алгоритма решения вопроса о степени сходства протеолитических ферментов человека и лёгочных пресноводных моллюсков. Такая информация будет полезна школьникам, владеющим компьютером на бытовом уровне, для понимания эволюционной консервативности ключевых процессов жизни клеточных организмов, в частности протеолиза. Практическая значимость решения этого вопроса связана с обоснованием целесообразности введения лёгочных пресноводных моллюсков в аквакультуру для получения из их тканей важных для фармацевтики и питания человека ферментов, как это осуществляется в настоящее время из тканей морских гидробионтов.

В большинстве сообщений, описывающих практическое использование лёгочных пресноводных моллюсков (прудовики — *Lymnaea stagnalis* L. and катушки — *Planorbarius corneus* L.), указывается на их участие в пищевых цепях экосистем, их роль как промежуточных переносчиков возбудителей некоторых заболеваний, использование в качестве модельных организмов для изучения физиологических процессов (размножение, нервная регуляция, клеточный метаболизм, генетика и др.). Однако нам не удалось найти систематизированных исследований, в которых бы обсуждался вопрос о применении пресноводных лёгочных моллюсков для получения ферментов, в частности ферментов протеолиза. Следует отметить, что на их кодирование приходится 5 % генома живых организмов. Раньше считали, что протеазы разрушают отработанные белки для поддержания общего гомеостаза. Однако последние исследования

продемонстрировали, что эти ферменты намного сложнее, поскольку играют ключевые роли в качестве сигнальных молекул и регуляторов важных клеточных процессов, таких как экспрессия генов, деление и гибель клеток. Ранее было показано, что в гемолимфе и гепатопанкреасе *Lymnaea stagnalis* L. и *Planorbarius corneus* L. имеются трипсиноподобные протеиназы, а также антитрипсиновые ингибиторы протеолиза и 2-макроглобулин [2; 4].

Внеклеточные белки и некоторые белки поверхности клетки поглощаются эндоцитозом и деградируют в лизосомах. Некоторые цитозольные белки и митохондрии деградируют в аутофагических везикулах после слияния с лизосомами. Разрушение мембраны лизосом ведёт к автолизу клетки. Во всех тканях живых организмов большинство внутриклеточных белков деградируют с помощью убиквитин — протеасомного пути. Деградация белков по АТФ-зависимому убиквитин-протеасомному пути включает в себя два этапа: ковалентное присоединение к субстрату полиубиквитиновой цепочки и деградацию помеченного белка 26S-протеасомой. Реакция убиквитирования осуществляется каскадом ферментов: E1 (Ubiquitin activating enzyme) — E2 (Ubiquitin conjugating enzyme) — E3 (Ubiquitin ligase). В протеасомах млекопитающих каталитически активными являются только β 1-, β 2- и β 5-субъединицы, причём все эти субъединицы обладают разными ферментативными активностями (каспазоподобной, трипсиноподобной и химотрипсиноподобной соответственно). Протеасомное разрушение белков представляет собой быстрый процесс, который обеспечивает переключение важнейших процессов экспрессии генов, клеточного цикла и апоптоза путём разрушения регуляторных белков p19, p21, p27, p53 и других, ряда транскрипционных факторов. Протеасомы и каспазы взаимодействуют в борьбе за жизнь и смерть клетки [3].

Считают, что протеолитические ферменты весьма консервативные, и эта система регуляции сформировалась на уровне первичных клеток. Эволюционным источником многих пептидаз высших эукариот являются пептидазы прокариот. В процессе эволюции совершенствовались эндосомно-лизосомальная система и действующие в цитоплазме и ядре клетки убиквитин- и АТФ-зависимая (протеасомная), а также кальций-зависимая (кальпаиновая) системы [2].

Материал и методы

Общий алгоритм работы.

Поиск и отбор нуклеотидных последовательностей, кодирующих белки человека, осуществлялись на сервере <https://www.ensembl.org>; поиск гомологичных последовательностей для моллюсков осуществлялся на сервере <https://www.ncbi.nlm.nih.gov> при помощи ресурса BLAST; описание белков для человека было взято с ресурса <https://www.uniprot.org>; парное выравнивание и сравнение последовательностей человека и моллюсков выполнены в программе MEGA5.2; построение 3D-структур ферментов для моллюсков выполнялось на сервере <https://swissmodel.expasy.org> по шаблону 3D-структуры ферментов человека, найденных в банке данных трёхмерных структур белков и нуклеиновых кислот <http://www.rcsb.org>. В работе использован следующий алгоритм: поиск нуклеотидной последовательности → построение аминокислотных последовательностей сравниваемых белков → их парное выравнивание и оценка степени гомологии первичных структур → построение 3D-структур по шаблону структуры сравниваемого белка человека → оценка третичной структуры по архитектуре молекул и их доменной организации. Исследование мотивов и строения активных центров ферментов не входило в задачи данной работы.

Конкретные этапы работы.

1. Поиск и отбор нуклеотидных последовательностей, кодирующих белки человека, осуществлялись на сервере <https://www.ensembl.org/>. На главной странице сайта имеется выпадающее меню, в котором можно выбрать геном того организма, для которого мы ищем нуклеотидные последовательности. Также на следующей строчке можно написать интересующий нас белок. Например, выбираем в строке *Search Human*, на следующей — пишем название белка *caspase* и нажимаем кнопку **Go**. На следующей странице открываются результаты поиска. На меню слева можно выбрать необходимые параметры (текущий выбор генома — *Current selection*, ограничение категорий — *Restrict category to*, количество результатов на странице — *Per page* и макет их отображения — *Layout*). В ограничении категорий (*Restrict category to*) выбираем *Transcript* и необходимый транскрипт белка, например *CASP14-201 (Human Transcript)*. Нажимаем на его название и по ссылке переходим на следующую страницу с описанием данного белка.

Для получения нуклеотидной последовательности необходимо нажать на кнопку «Показать таблицу транскрипта» (*Show transcript table*), в которой будут содержаться транскрипты. В данном случае выбираем CASP14-201 (в колонке **Transcript ID** нажимаем на ссылку ENST00000427043.3). Страница загружается ещё раз. Теперь необходимо слева нажать на кнопку «Экспортировать данные» — **Export data**. Затем открывается новое окно, в котором в графе **Output** должен стоять формат последовательности *FASTA sequence*, в графе **Strand** — *Feature strand*. И нажимаем кнопку **Next**.

В следующем открывшемся окне предлагается выбрать выходной формат для экспорта. При открытии ссылки **HTML** последовательность открывается как страница браузера с дополнительными меню и ссылками; при открытии ссылки **Text** — как простой текст на странице; при открытии ссылки **Compressed text (.gz)** — скачивается как архив.

2. Поиск гомологичных последовательностей для моллюсков осуществлялся на сервере <https://www.ncbi.nlm.nih.gov> при помощи ресурса **BLAST**. Заходим на главную страницу сервера. В правом меню выбираем ресурс **BLAST**. Далее на странице предлагается несколько вариантов этого ресурса (например, *Nucleotide BLAST* — ищет гомологичные нуклеотидные последовательности, *Protein BLAST* — ищет гомологичные аминокислотные последовательности, *blastx* — переводит нуклеотидные последовательности в аминокислотные, *tblastn* — переводит аминокислотные последовательности в нуклеотидные).

Мы будем использовать *Nucleotide BLAST*, для этого нажимаем на соответствующую кнопку. Перед нами открывается окно **Standard Nucleotide BLAST**. Далее слева вверху (Enter accession number(s), gi(s), or FASTA sequence(s)) нам нужно ввести либо номер последовательности, либо номер gi, либо вставить последовательность в *FASTA*-формате. *FASTA*-формат — это формат, когда последовательность начинается со знака >. Копируем необходимую последовательность, найденную на сервере <https://www.ensembl.org/>, и вставляем её в это окно. Ниже на одну строчку можно прикрепить файл с последовательностью, если она сохранена отдельным файлом. Ещё ниже на одну строчку можно задать имя для поиска, например *caspase*. Далее в выпадающем меню **Database** должен стоять параметр *Nucleotide collection (nr/nt)*.

В строке **Organism** можно указать интересные нас организмы, т. е. название самого организма, семейства или отряда на латинском языке (например мы ищем последовательности белка для лёгочных пресноводных моллюсков семейства *Planorbidae*).

В меню **Program Selection** можно оптимизировать поиск. Предлагается три варианта: *Highly similar sequences (megablast)*, *More dissimilar sequences (discontiguous megablast)* и *Somewhat similar sequences (blastn)*: первый вариант предназначен для сравнения последовательностей с большим процентом гомологии и лучше всего работает, если целевой процент идентичности составляет 95 % или более (проводит сравнение очень быстро); второй вариант использует первоначальное семя, которое игнорирует некоторые основания (позволяя несоответствия), и предназначен для межвидовых сравнений; третий вариант проводит сравнение медленно, но позволяет размер слова довести до семи баз. Мы выбираем последний вариант, после чего ниже нажимаем кнопку **BLAST**. Если наша последовательность имеет большой размер, то после нажатия кнопки нужно подождать какое-то время до получения результата (страница при этом периодически перезагружается).

На странице с результатами есть графическое резюме (*Graphic Summary*), в котором найденные последовательности отображаются в виде точек или полос разного цвета (это зависит от размера найденной последовательности и количества гомологичных нуклеотидов). Все последовательности здесь кликабельные. После графического резюме идёт описание (*Descriptions*) каждой найденной последовательности в виде таблицы. В этой таблице есть название последовательности, максимальный и общий вес (*Max Score, Total Score*), процент длины исходной последовательности, выровненный с найденной последовательностью (*Query Coverage*), величина *E-value*, процент идентичности (*Ident*) и ссылка на последовательность (*Accession*).

После описания в таблице идут сами выравнивания последовательностей с графическим отображением. Здесь можно по отдельности сохранить необходимые последовательности (кнопка **Download** перед каждой последовательностью). Сохранить последовательности можно и в таблице описания (например, выделить несколько последовательностей галочкой слева от названия или выбрать в самом

верху таблицы *Select:All* и нажать кнопку *Download*). Сохраняем необходимые последовательности. Теперь у нас есть последовательности, по которым можно построить 3D-структуры белков для моллюсков.

3. Построение 3D-структур ферментов для моллюсков выполнялось на сервере <https://swissmodel.expasy.org> по шаблону 3D-структуры ферментов человека, найденных в банке данных трёхмерных структур белков и нуклеиновых кислот <http://www.rcsb.org>. Для этого в банке данных трёхмерных структур белков и нуклеиновых кислот <http://www.rcsb.org> ищем 3D-структуры для человека. На главной странице сайта <http://www.rcsb.org> в поисковую строку вводим название интересующего нас белка (например, *caspase*) и дописываем слово *human* (чтобы в результате поиска отобразились только белки человека). После нажимаем кнопку *Go*. Далее мы попадаем на страницу с результатами поиска. Находим интересующий нас белок, открываем его страницу по ссылке в названии (например, *Crystal Structure of Ligand Free Human Caspase-6*). Чтобы скачать его 3D-структуру, нажимаем кнопку *Download File*, в выпадающем меню выбираем *PDB Format*. После этого открывается окно с выбором двух вариантов: *Открыть* или *Сохранить*. Выбираем пункт *Сохранить* и нажимаем *Ok*. В папке загрузок у нас появляется файл с 3D-структурой этого белка (расширение файла должно быть *.pdb*).

4. Теперь у нас есть последовательность белка для моллюска и 3D-структура для человека — можно строить 3D-структуру белка для моллюска. Открываем сервер <https://swissmodel.expasy.org/> и нажимаем кнопку *Start Modelling*. В открывшемся окне справа в меню «Поддерживаемые входы» (*Supported Inputs*) выбираем *User Template* (пользовательский шаблон). Слева в поле *Target Sequence(s)*: вставляем последовательность, которая будет преобразована в 3D-структуру (последовательность для моллюска). Либо нажимаем кнопку *Upload Target Sequence File ...* и загружаем файл с этой последовательностью. После этого в поле *Template File*: загружаем файл с 3D-структурой для человека (это будет наш шаблон), скачанный с сервера <http://www.rcsb.org>. В поле *Project Title*: пишем название проекта. И нажимаем кнопку *Build Model* — начинается построение модели. Это может занимать какое-то количество времени (зависит от размера последовательности, которую мы преобразовываем, и от размера шаблона).

После завершения построения в окне откроются построенные модели, которые можно скачать и использовать для дальнейшей работы.

Полученные результаты и обсуждение.

Сравнительный биоинформатический анализ протеолитических ферментов человека и моллюска *Biomphalaria glabrata*, являющегося родственным организмом с *Planorbarius corneus*, представлен в таблице 1.

Таблица 1 — Оценка гомологии первичных и третичных структур молекул внутриклеточных протеолитических ферментов человека и моллюска

Фермент	Гомология	
	Последовательности аминокислот	Третичные структуры, однотипные участки
1	2	3
Prolyl oligopeptidase (КФ 3.4.21.26)	66 %	Альфа-спирализованный и бета-складчатый домены
ATP-dependent Clp protease proteolytic subunit (КФ 3.4.21.92)	68 %	Признаки структуры типа семилучевой звезды
Furin — КФ 3.4.21.75	69 %	Альфа-спирализованный и бета-складчатый домены
Signal Peptide Peptidase (КФ 3.4.11.2)	66 %	Повторяющиеся 4 типа альфа-спиральных структур
Leucyl aminopeptidases (cytosol aminopeptidase, КФ 3.4.11.1)	66 %	Аналогичная гексамерная структура
Thimet oligopeptidases (КФ 3.4.24.15)	66 %	Преобладание альфа-спирализованных участков

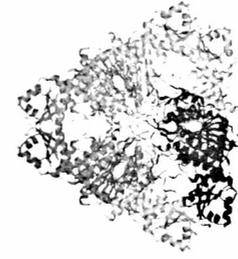
1	2	3
Ubiquitin conjugating factor E4 B-like (КФ 6.3.2.19)	72 %	Однотипные альфа-спирализованные фрагменты
Ubiquitin conjugating factor E2 W-like	75 %	Одинаково чередуются альфа-спиральные, бета-структурные и неупорядоченные фрагменты
Ubiquitin carboxyl-terminal hydrolase L5	72 %	Сходные каталитические домены
Ubiquitin-like modifier-activating enzyme 5	76 %	Построение домена, состоящего из альфа-спиральных и бета-структурных участков
Amidophosphoribosyl-transferase (phosphoribosyl pyrophosphate amidotransferase) (КФ 2.4.2.14)	64 %	Четыре однотипных домена
Adenylosuccinate lyase (adenylosuccinase, КФ 4.3.2.2)	68 %	Мономеры имеют по три идентичных домена

Первые шесть ферментов, представленных в таблице 1, являются ферментами, которые можно отнести к группе строго не регулируемых протеолитических ферментов: 1) Prolyl oligopeptidase представляет собой цитозольную сериновую пептидазу, которая расщепляет пептидную связь С-концевого пролина; 2) ATP-dependent Clp protease proteolytic subunit входит в состав высокоактивной сериновой эндопептидазы Clp; 3) Furin является сериновой протеазой клеток животных, расположенной в аппарате Гольджи, и напоминает бактериальный протеолитический фермент субтилизин; 4) Signal Peptide Peptidase — это внутримембранная аспартил-протеаза, расщепляющая остаточные сигнальные пептиды, оставшиеся в мембране после действия сигнальной пептидазы; 5) Leucyl aminopeptidase (cytosol aminopeptidase) относится к ферментам, которые преимущественно катализируют гидролиз лейциновых остатков на N-конце пептидов и белков; 6) Thimet oligopeptidases, известные как TOPs, являются металлопептидазами, и у животных они участвуют в деградации пептидов — брадикинина, нейро-тензина, ангиотензина I и пептида Аβ. Первичные структуры этих ферментов человека и моллюска *Biomphalaria glabrata* после парного выравнивания демонстрируют гомологию в интервале 66–69 %. 3D-структуры этих ферментов у человека и моллюска близки по архитектуре и наличию доменных структур. Следующие четыре фермента относятся к управляемому убиквитин-протеасомному пути протеолиза; 7) Ubiquitin conjugating factor E4 B-like, конъюгирующие убиквитин

ферменты, также известные как ферменты E2; 8) Ubiquitin conjugating factor E2 W-like; 9) Ubiquitin carboxyl-terminal hydrolase L5; 10) Ubiquitin-like modifier-activating enzyme 5. Нуклеотидные последовательности этих ферментов для моллюсков оказались неполными, но тем не менее проведенный анализ фрагментов первичных структур показал более высокую степень гомологии при парном выравнивании 72–76 %, а фрагменты 3D-структур позволили выявить сходство третичных структур на уровне общей архитектуры молекул и их доменного строения. Для сравнения биоинформатическому анализу были подвергнуты два фермента пуринового обмена, важного для синтеза нуклеотидов; 11) Amidophosphoribosyl-transferase (phosphoribosyl pyrophosphate amidotransferase), который содержит аналог каталитической триады цистеиновых протеаз и 12) Adenylosuccinate lyase (adenylosuccinase), который превращает аденилосукцинат в АМР и фумарат. Гомология этих ферментов у человека и моллюска такая же, как и у ферментов нерегулируемого протеолиза.

В таблице 2 в качестве примера приведены 3D-структуры нерегулируемого и регулируемого ферментов протеолиза человека и моллюска *Biomphalaria glabrata*. **Третичные структуры определяют функции белков.** Анализ гомологии третичных структур был ограничен наличием сходных черт в архитектуре ферментов и их доменного строения. Анализ поиска сходных мотивов полипептидных цепей и структур активных центров ферментов не входил в задачи данной статьи.

Таблица 2 — 3D-структуры ферментов протеолиза человека и моллюска

Нерегулируемый протеолиз. Гомология 64–69 %		Регулируемый протеолиз. Гомология 69–76 %			
Фермент	3D-структура фермента человека	3D-структура фермента моллюска	Фермент	3D-структура фермента человека	3D-структура фермента моллюска
<p>Prolyl oligopep-tidase Два похожих домена — α-спиральный, β-структурный</p>			<p>Ubiquitin conjugation factor E2 E1 Аналогичная структура олигомерного белка</p>		
<p>Leucyl aminopeptidases (cytosol aminopeptidase) Одинаковая олигомерная структура</p>			<p>Ubiquitin carboxyl-terminal hydrolase L5 Аналогичный каталитический домен</p>		
<p>Thimet oligopep-tidases Близкие по структуре глобулярные белки</p>			<p>E3 ubiquitin ligase (WD40 domain) Фрагмент фермента у моллюска имеет признаки сходства с ферментом человека</p>		

При анализе данных таблицы 2 следует учитывать, что 3D-структуры белков, как правило, являются более консервативными, чем белковые последовательности, поэтому сходство структур может указывать на сходство и функций белков. Следовательно, полученные материалы позволяют положительно решить вопрос об использовании тканей лёгочных пресноводных моллюсков для моделирования патологических процессов человека, связанных с нарушениями системы протеолиза. Кроме того, из тканей моллюсков могут быть выделены протеолитические ферменты, которые затем могут найти применение в фармакодинамических исследованиях регуляторов протеолиза.

Заключение

Представленные данные показывают, что высокая степень гомологии ферментов протеолиза у человека и моллюска сопряжена с формированием близких третичных структур белков. Это свидетельствует в пользу предположения, что сравниваемые протеолитические ферменты человека и моллюсков выполняют однотипные функции. Можно также предположить, что более высокая степень гомологии первичных структур у протеолитических ферментов убиквитин-протеасомного пути может быть связана с более высокой степенью консервативности этого регулируемого пути внутриклеточного протеолиза.

Практическая значимость полученных данных о высокой степени гомологии протеоли-

тических ферментов у человека и лёгочных пресноводных моллюсков обосновывает формирование аквакультуры моллюсков с целью получения из их тканей белковых ферментативных препаратов протеолитического действия в рамках задач биофармации, для совершенствования косметических средств и применения в пищевой промышленности.

Выводы

Описан алгоритм построения третичных структур сравниваемых белков с помощью открытых международных баз данных и доступных серверов.

Гомология ферментов нерегулируемого протеолиза у человека и лёгочных пресноводных моллюсков находится в пределах 66–69 %, а убиквитин-протеасомного пути — 72–76 %.

Практическое значение высокой степени гомологии протеолитических ферментов у людей и пресноводных лёгочных моллюсков обосновывает возможность получения из тканей последних белковых ферментативных препаратов протеолитического действия для практического применения.

Благодарность

Авторы благодарны доценту В. В. Хрусталёву (кафедра общей химии Белорусского государственного медицинского университета) за консультации по методам биоинформатики.

Список использованных источников

1. Основы безопасности жизнедеятельности: 10–11 классы: методическое пособие / С. В. Алексеев, С. П. Данченко, Г. А. Костецкая. — М. : Вентана-Граф, 2015. — 120 с.
2. Чиркин, А. А. Сравнительный биохимический анализ тканей лёгочных пресноводных моллюсков, обитающих в озёрах Витебской и Гомельской областей Республики Беларусь / А. А. Чиркин, О. М. Балаева-Тихомирова, Е. О. Данченко [и др.] // SWorld. Научный мир. — 2018. — № 51. — Т. 1. — С. 90–95.
3. Цимоха, А. С. Протеасомы: участие в клеточных процессах / А. С. Цимоха // Цитология. — 2010. — Т. 52. — № 4. — С. 277–300.
4. Chirkin, A. Inhibition of proteolysis in plant and animal tissues / A. Chirkin, V. Dolmatova // Agrobiodiversity for improving nutrition, health and life quality. — 2017. — № 1. — P. 60–65.
5. Материалы Википедии.