

УДК 577.322.53

## ВЫБОР ОПТИМАЛЬНОГО МЕТОДА КОНСТРУИРОВАНИЯ ТРЕТИЧНОЙ СТРУКТУРЫ БЕЛКОВ НА ПРИМЕРЕ ИЗОФОРМЫ P00749-2 УРОКИНАЗЫ

*Ринейская О. Н., Байроченко Д. С.*

*Учреждение образования «Белорусский государственный медицинский университет»,  
г. Минск, Республика Беларусь*

**Реферат.** Урокиназа существует в виде нескольких изоформ, но только для одной из них установлена кристаллографическая структура, информацию о которой можно получить в Protein Data Bank. Поэтому актуальным является моделирование пространственной структуры белков *in silico*, для которых пока не определена третичная структура при помощи физических методов исследования. В настоящей работе проведено сравнение качества 3D-моделей изоформы урокиназы P00749-2, полученных с использованием MODELLER, Phyre2, Robetta и SWISS-MODEL. Установлено, что использование сервиса SWISS-MODEL является наиболее оптимальным для прогнозирования третичной структуры изоформ урокиназы.

**Ключевые слова:** гомологичное моделирование, урокиназа, SWISS-MODEL, Robetta, Phyre2, MODELLER.

**Введение.** Все более важную роль в науке играют исследования *in silico*, представляющие собой изучение биологически важных процессов и структур с использованием специализированных программ. Применение методов *in silico* позволяет делать прогнозы и при этом экономить человеческие и денежные ресурсы при проведении начальных этапов научных исследований. Но, безусловно, выдвинутые гипотезы требуют валидации в экспериментах *in vitro* и *in vivo*. Последние четыре десятилетия внимание исследователей в области биологии и медицины сфокусировано на получении структур молекул методами рентгеноструктурного анализа и электронной микроскопии. Каждый год международные библиотеки и базы данных пополняются новыми кристаллографическими структурами. Так, только в 2023 году в базу данных Protein Data Bank (PDB) было добавлено 12 057 новых структур. Эти биологические молекулы могут быть потенциальными мишенями для разработки таргетной терапии заболеваний растений и животных или же ключом к пониманию фундаментальных процессов в биологии и медицине. Однако, несмотря на столь стремительное пополнение баз данных информацией о третичных структурах биомолекул, традиционных методов установления структур, очевидно, недостаточно для удовлетворения имеющегося спроса на эти данные у

научного сообщества. Развитие новых методов позволяет снабдить исследователей более полной информацией о структурах белков и их изомеров, столь необходимой для целостного понимания функционирования человеческого организма и дальнейшего проведения исследований. Из сказанного выше следует, что вопрос о количестве и качестве 3D-структур биомолекул стоит достаточно остро. Решением этой задачи явилась разработка ряда методов, позволяющих предсказывать третичную структуру белковых молекул *in silico*. Данные методы могут использоваться не только для предсказания пространственного строения отдельных белков, но и для конструирования их изоформ, что может быть полезно при разработке высокоаффинных ингибиторов белков-мишеней [1], а также в контексте комплексного изучения *in silico* эффективности действия потенциальных лекарственных средств.

Урокиназный активатор плазминогена (урокиназа) — это внеклеточная сериновая протеаза, состоящая из 411 аминокислотных остатков, кодируемая геном PLAU (находится в длинном плече 10-й хромосомы человека). Данный фермент участвует в фибринолизе посредством активации плазминогена, способен подвергать гидролизу некоторые компоненты внеклеточного матрикса. Он также активирует внеклеточные цинксодержащие эндопептидазы,

представителей семейства матриксных металлопротеиназ, ответственных за ремоделирование и деградацию внеклеточного матрикса, регуляцию пролиферации, миграции и дифференциации клеток, могут играть важную роль в ангиогенезе, регуляции апоптоза и восстановления тканей [2].

Экспрессия гена PLAU может быть индуцирована различными типами факторов роста, гормонами, цитокинами, а также морфологическими изменениями тканей. В опухолевых клетках экспрессия усиливается в несколько раз, что ведет к повышению активности урокиназы. Это вызывает усиление процессов разрушения внеклеточного матрикса и ведет к нарушению межклеточных адгезионных контактов, что, в свою очередь, может способствовать прогрессированию опухолей и метастазированию [3].

Урокиназа является потенциальной мишенью для противоопухолевой терапии. В настоящее время на разных этапах клинических испытаний находятся несколько лекарственных препаратов [4], которые способны ингибировать ферментативную активность урокиназы путем связывания с ее активным центром, таким образом значительно сдерживая метастазирование и дальнейшее прогрессирование опухолей.

Урокиназа существует в виде нескольких изоформ, но только для одной из них установлена кристаллографическая структура, информацию о которой можно получить в PDB [<https://www.rcsb.org>]. При разработке ингибиторов фермента *in silico* желательно проводить анализ эффективности действия предложенных лигандов в отношении всех изоформ. Решению подобной задачи служит предсказание пространственных структур изоферментов урокиназы, необходимых для последующего анализа эффективности связывания в области активного центра ингибиторов, например Нафамостата и ряда его производных.

**Цель работы** – сравнение качества третичных структур изоформы урокиназы P00749-2, полученных путем моделирования *in silico* с использованием ряда онлайн-сервисов и программ.

**Материалы и методы.** Поиск аминокислотных последовательностей изоформ урокиназы осуществлялся в ресурсе UniProt [<https://www.uniprot.org>], содержащем данные о последовательностях белков и аннотации к ним. Были найдены два варианта фермента, обозначенные как P00749-1 и P00749-2. Вариант P00749-1 имеет длину 431 аминокислотного остатка и массу около 48,5 кДа, является каноническим, для него ранее была предсказана *in silico* трехмерная структура, которую можно найти в

базе данных AlphaFold [<https://alphafold.ebi.ac.uk>]; пространственное строение его также установлено методами рентгеноструктурного анализа (PDB ID 7vm4). Вариант P00749-2 (изоформа 2) образуется в результате альтернативного сплайсинга, имеет длину 414 аминокислотных остатков и массу около 46,9 кДа. Именно изоформа 2 была выбрана для предсказания третичной структуры, так как информация о ее третичной структуре не представлена в PDB. Для построения пространственных моделей протеина использовался ряд специализированных программ. SWISS-MODEL [5] представляет собой онлайн-сервис, предназначенный для гомологичного моделирования белковых молекул. Построение моделей происходит в несколько этапов: определение структурных референсов, выравнивание интересующей последовательности и референсных последовательностей, построение модели и оценка ее качества. SWISS-MODEL обеспечивает интегрированную работу программного обеспечения и актуальных баз данных, содержащих последовательности и третичные структуры белков, что необходимо для корректной работы сервиса. Показатель GMQE (Global Model Quality Estimate — глобальная оценка качества модели) характеризует основные геометрические аспекты белковой структуры.

Robetta — сервер прогнозирования конфигураций белковых структур, в основе которого лежит программный пакет Rosetta для моделирования на макромолекулярном уровне. Основным функционал Robetta сводится к предсказанию трехмерной структуры белка на основе его аминокислотной последовательности. Предсказание структуры осуществляется методом гомологического моделирования или моделирования *ab initio* — в зависимости от выбранных опций. В данном случае построение образцов осуществляется с использованием протокола RoseTTAFold, представляющего собой нейронную сеть, которая одновременно оценивает паттерны белковых последовательностей, взаимодействие между аминокислотами и оценивает потенциальную 3D-структуру белка [6].

Phyre2 представляет собой комбинацию нескольких программных компонентов [7], которые обеспечивают гомологичное моделирование на основе одной или нескольких референсных последовательностей и при их отсутствии моделирование *ab initio*. Это позволяет воссоздать 3D-структуру с использованием полной аминокислотной последовательности, учитывая стереохимические взаимодействия между остатками моделируемого участка, включая длину связей,

торсионные углы, вандерваальсовы и электростатические взаимодействия. Сервис автоматически подбирает референсные структуры, обеспечивая максимальную достоверность, идентичность, охват интересующей последовательности и создает только одну модель.

MODELLER – программа для гомологичного и сравнительного моделирования трехмерной структуры белка с учетом пространственных ограничений [8]. Метод основан на выравнивании входной последовательности, подлежащей моделированию, и референсной аминокислотной последовательности («шаблонным белком», структура которого ранее была установлена). С целью повышения точности моделирования задается алгоритм, позволяющий использовать несколько последовательностей в качестве референсных (multi-template homology modelling).

Первоначальная оценка качества трехмерных моделей урокиназы P00749-2, предложенных каждой из программ с целью выбора наиболее успешного варианта для последующего сравнительного анализа, производилась посредством предлагаемых каждой программой инструментов и параметров. Для выявления наиболее подходящего сервиса трехмерного моделирования протеина было отобрано по одной модели, полученной при помощи каждого из описанных сервисов. Для последующего сравнительного анализа качества 3D-структур использовались программы SAVES v6.0 и Инструмент Structure Assessment.

Сервис SAVES v6.0 при помощи алгоритма *ERRAT* позволяет анализировать белковые структуры на предмет наличия несвязанных остатков в пределах 3,5 Å между различными атомами, рассчитывая общий коэффициент качества (для высококачественной модели значение должно превышать 50). Программа *VERIFY3D* сопоставляет трехмерную структуру с аминокислотной последовательностью, проводя сравнение качества исследуемой молекулы с белковыми структурами с высоким разрешением из базы данных. *PROCHECK* обеспечивает анализ общей геометрии и оценку стереохимического качества прогнозируемой модели, строит карту Рамачандрана с разрешением 2,0 Å.

Инструмент Structure Assessment рассчитывает ряд показателей, позволяющих оценить качество подготовленной структуры. При расчете значения «Сβ» анализируются только попарные взаимодействия между позициями β-углеродных атомов аминокислот. Показатель «All Atom» учитывает потенциал взаимодействия между всеми атомами, при котором оцениваются попарные

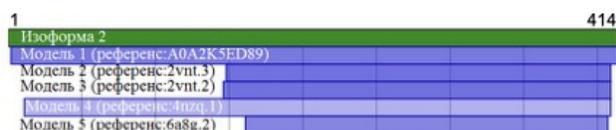
взаимодействия между ними. «Solvation» показывает, насколько соответствует доступность аминокислотного остатка построенной модели растворителю, его ожидаемому значению. «Torsion» оценивает соответствие каждых трех последовательно расположенных аминокислот модели пространственным ограничениям, задаваемым двугранными углами ψ и φ. QMEAN и QMEAN-DisCo — это оценочные функции, которые позволяют получать как глобальные (для всей структуры), так и локальные (для каждого остатка) абсолютные оценки качества на основе одной модели. Основным отличием QMEANDisCo от QMEAN является возможность прогнозировать согласованность парных расстояний α-углеродных атомов на основе модели с пространственными ограничениями (DisCo — distance constraints), полученной из гомологичных структур. Все оценки объединяются с использованием нейронной сети. Перечисленные показатели формируются на основании сравнения анализируемой модели и экспериментально определенных структур аналогичного размера. Структуры, получившие значения близкие к нулю по каждому из вышеназванных параметров, считаются нативными. Неудовлетворительными считаются значения, соответствующие диапазону от -4 и ниже. Сервис также позволяет построить карту Рамачандрана с разрешением 2,5 Å.

**Результаты и их обсуждение.** Моделирование изоформы 2 урокиназы при помощи SWISS-MODEL проводилось в автоматическом режиме. Из представленного отчета на основании значения параметра GMQE были отобраны 5 структур, характеристики которых представлены в таблице 1.

Таблица 1 – Характеристики выбранных референсных структур в SWISS-MODEL

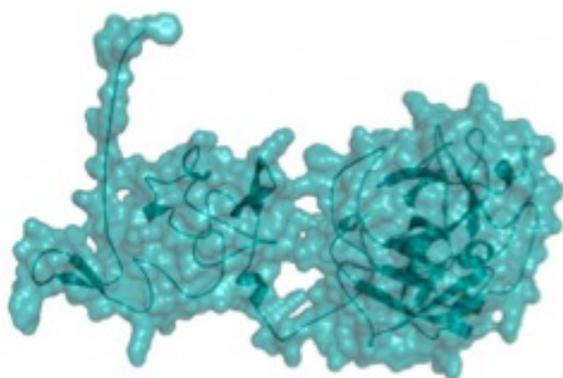
№	Условное обозначение	GMQE	Идентичность последовательности, %	Метод
1	A0A2K5ED89	0,84	89,08	AlphaFold v2
2	2vnt.3	0,61	99,64	X-ray 2,20Å
3	2vnt.2	0,59	99,63	X-ray 2,20Å
4	4nzq.1	0,58	28,35	X-ray 2,81Å
5	6a8g.2	0,50	69,92	X-ray 2,53Å

На основании этих референсных структур были построены модели изоформы 2 урокиназы (рисунок 1). Очевидным недостатком данного сервиса является тот факт, что построение 3D-структур зависит от идентичности последовательностей и длины, которой обладают выбранные ранее референсные структуры, в связи с чем возможна ситуация визуализации с использованием неполной последовательности белка, что и произошло в нашем случае.



**Рисунок 1 – Результаты гомологичного моделирования изоформы 2 урокиназы при помощи SWISS-MODEL. Номера 1–414 соответствуют аминокислотным остаткам**

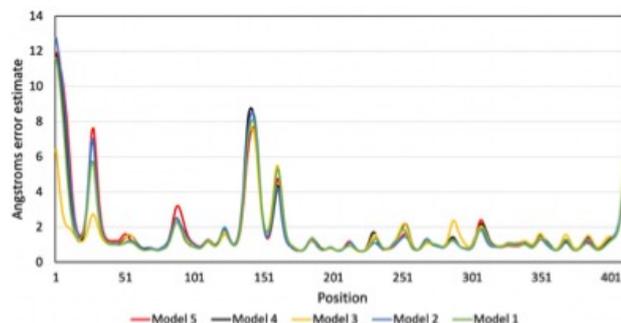
Таким образом, модель 1 покрывает аминокислотные остатки 3–414, являясь лидером среди полученных моделей по данному показателю. Значение GMQE распределяется по моделям 1–5 следующим образом: 0,84; 0,61; 0,59; 0,58; 0,50 соответственно. По сравнению с остальными вариантами модель 1 обладает наиболее высоким значением GMQE, что позволяет предполагать большую правдоподобность построенной модели по сравнению с остальными. В связи с этим именно модель 1 (изображена на рисунке 2) была отобрана для дальнейшего сравнительного анализа.



**Рисунок 2 – Модель 1 изоформы 2 урокиназы, полученная с помощью SWISS-MODEL**

Алгоритмами сервиса Robetta было предложено 5 потенциальных моделей. Сервер предоставляет данные о возможной погрешности (отклонении от координат данной аминокислоты по

сравнению с гомологичной последовательностью) в ангстремах для каждого остатка всех 5 моделей, на основании значения которой можно судить о качестве сконструированных структур (рисунок 3).



**Рисунок 3 – Оценка погрешности (Å) для каждого аминокислотного остатка (1–414) моделей изоформы 2 урокиназы**

Из приведенных данных видно, что модель 3 имеет наименьшие показатели погрешности, что свидетельствует о более высокой точности построения модели. Среднее значение погрешности для нее (1,58 Å) можно считать приемлемым, в связи с чем именно данная модель была отобрана для последующего анализа. В программе Phyre2 нами использовался интенсивный режим (Intensive mode), поскольку он позволяет одновременно использовать несколько последовательностей (в данном случае 4 аминокислотные последовательности сравнения). Онлайн-сервис предоставил карту достоверности, описывающую качество различных участков сконструированной модели (рисунок 4): 17 аминокислотных остатков были смоделированы ab initio (обладают низкой степенью достоверности, на рисунке 4 обозначены синим); 96 % Остатков смоделированы с достоверностью более 90 %, значит, качество модели можно считать достаточным для принятия варианта для дальнейшего сравнения.



**Рисунок 4 – Оценка качества участков сконструированной модели изоформы 2 урокиназы (Phyre 2). Смещение в красную область означает высокое качество, в синюю – низкое**

Несмотря на то, что программа MODELLER предоставляет алгоритмы выравнивания

последовательностей, для поиска потенциальных шаблонных белков была использована программа BLAST (алгоритм blastp), которая осуществляет поиск среди большого количества последовательностей (было установлено значение 5000), давая более точные результаты. Из предоставленного списка было отобрано 6 шаблонных последовательностей на основании следующих параметров: значение E (отражает вероятность случайного совпадения последовательностей при выравнивании, не обусловленного истинной биологической взаимосвязью между последовательностями, в том числе общим эволюционным происхождением) для всех структур равно 0, что говорит о высокой схожести последовательностей по отношению к изоформе 2; процент покрытия (для всех выбранных равен 97) и идентичности последовательностей (характеристика последовательностей представлена в таблице 2).

Таблица 2 – Характеристика референсных последовательностей, использованных для моделирования в программе MODELLER

№	Условное обозначение	Длина, аминокислот	Идентичность последовательности, %	Биологический вид
1	P00749.3	431	99,75	Homo sapiens
2	QRF29.1	431	97,52	Pongo abelii
3	P16227.1	433	92,61	Papio cynocephalus
4	Q8MNY7.1	433	83,70	Oryctolagus cuniculus
5	P04185.1	442	79,81	Sus scrofa
6	Q05589.1	433	75,50	Bos taurus

Было сконструировано 5 моделей. Для первичной оценки их качества использовался параметр DOPE — discrete optimized protein energy (дискретная оптимизированная энергия белка). DOPE — потенциал, зависящий от межатомного расстояния. Значения DOPE (округлены до целых) всей белковой молекулы предсказанных структур были следующими: -29323; -29440; -28987; -28585; -29077 для моделей 1–5 соответственно. Модель 2 была выбрана в связи с тем, что она имела лучший (наименьший) показатель DOPE. Значение DOPE для каждого аминокислотного остатка предпочтительной модели продемонстрировано на рисунке 5.

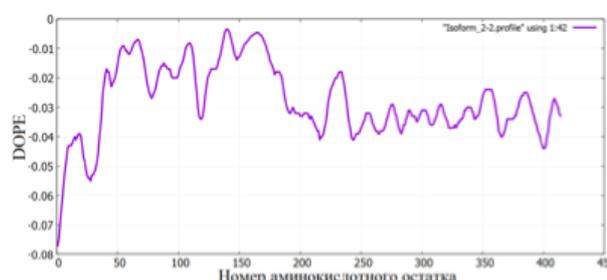


Рисунок 5 – Значение DOPE каждого аминокислотного остатка для модели 2 изоформы 2 урокиназы, полученной с помощью MODELLER

Для сравнительного анализа лучших моделей, предложенных каждым из 4 сервисов, был использован онлайн-сервис Structure Assessment.

При построении карты Рамачандрана лучшие результаты (наибольшее количество остатков в разрешенных регионах и наименьшее — в запрещенных) соответствовали модели, сконструированной в SWISS-MODEL. По значениям QMEANDisCo и QMEAN данная модель также занимает лидирующую позицию. Эта модель уступает лишь по показателям C $\beta$ , Solvation, Torsion модели Robetta (таблица 3).

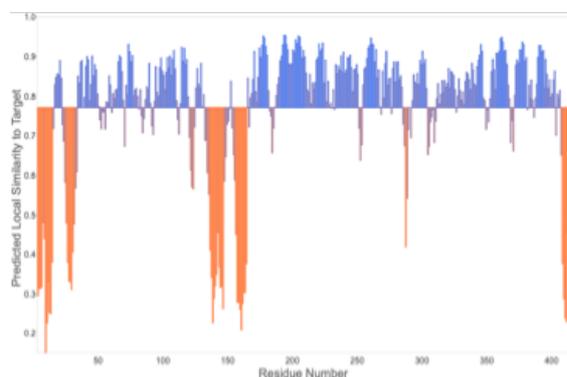
Таблица 3 – Количественные характеристики моделей изоформы 2 урокиназы, полученные при помощи различных сервисов (Structure Assessment)

Показатели	SWISS-MODEL	Robetta	Phyre2	MODELLER
Карта Рамачандрана, основной регион, %	95,85	92,96	89,32	87,38
Карта Рамачандрана, запрещенный регион, %	0,49	1,46	4,13	5,10
Карта Рамачандрана, другие регионы, %	3,66	5,58	6,55	7,52
QMEANDisCo	0,77	0,70	0,64	0,48
QMEAN	-0,02	-0,49	-4,94	-7,00

Окончание табл. 3

СВ	0,69	-0,38	-2,94	-5,68
All Atom	-1,87	-2,73	-5,72	-6,89
Solvation	-1,31	-0,99	-3,83	-7,96
Torsion	0,33	-0,04	-3,37	-3,77

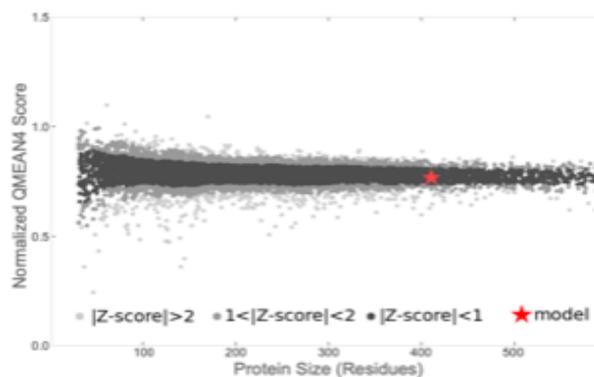
На данном этапе сравнения моделей на основании показателей Structure Assessment лучшей следует считать ту, которая получена с помощью SWISS-MODEL. Оценка локального качества, то есть ожидаемое сходство с «шаблонной» структурой (ось ординат) каждого аминокислотного остатка (ось абсцисс) данной модели представлена на рисунке 6. Принято считать, что остаткам, имеющим значение ниже 0,6, присуще низкое качество структуры. На рисунке 6 горизонтальной линией отмечено глобальное значение QMEANDisCo; отклонения для каждого остатка от данного значения в лучшую (обозначено синим цветом) или худшую (обозначено оранжевым цветом) сторону продемонстрированы вертикальными линиями.



**Рисунок 6 – Оценка локального качества аминокислотных остатков модели изоформы 2 урикиназы SWISS-MODEL**

Оценка глобального качества SWISS-MODEL модели 3D-структуры изоформы 2 урикиназы была произведена при помощи Инструмента Structure Assessment (рисунок 7).

Совокупность точек — это экспериментальные структуры белка с оценкой QMEAN в пределах 1 стандартного отклонения от среднего значения (Z-балл меньше 1). К данной области относится и анализируемая модель (представлена в виде красного астериска), что говорит о высоком качестве ее структуры. С целью уточнения значений в картах Рамачандрана для каждой из четырех моделей был использован сервис SAVES v6.0. Результаты исследования представлены в таблице 4.



**Рисунок 7 – Оценка глобального качества модели SWISS-MODEL изоформы 2 урикиназы**

Таблица 4 – Количественные характеристики моделей изоформы 2 урикиназы, полученных при помощи различных сервисов (SAVES v6.0)

Показатели	SWISS-MODEL	Robetta	Phyre2	MODELLER
ERRAT	89,26	90,5	42,47	38,48
VERIFY3D, %	70,63	73,43	76,09	59,66
Карта Рамачандрана, основной регион, %	89,3	82,6	78,9	79,8
Карта Рамачандрана, разрешенный регион, %	10,5	15,2	16,6	16,6
Карта Рамачандрана, в целом разрешенный регион, %	0,3	1,7	2,8	2,2
Карта Рамачандрана, запрещенный регион, %	0,0	0,6	1,7	1,4

При этом лучшим значением по показателю наличия несвязанных остатков в пределах 3,5 Å между различными атомами (ERRAT) обладает модель Robetta; по значению VERIFY3D (на основании сравнения с белковой структурой с высоким разрешением) — образец Phyre 2. В то же время по вышеперечисленным показателям модель из SWISS-MODEL занимает только второе и третье места соответственно. Однако по

всем параметрам карты Рамачандрана образец из SWISS-MODEL обладает лучшими показателями. Более того, данная модель не имеет остатков, расположенных в запрещенном регионе карты, что является весомым показателем, говорящим в пользу высокой степени качества структуры. Карта Рамачандрана для модели SWISS-MODEL представлена на рисунке 8.

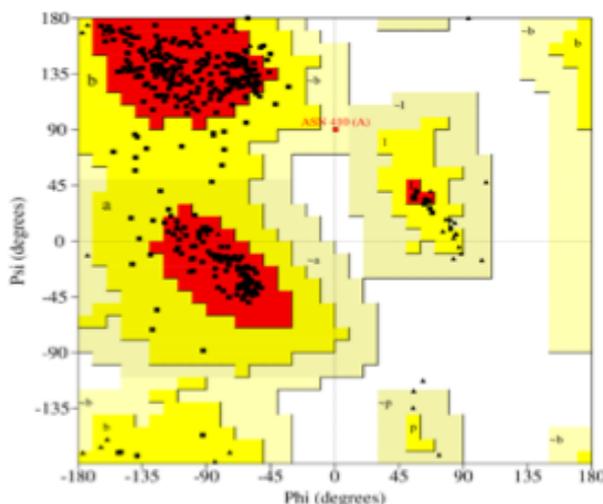


Рисунок 8 – Карта Рамачандрана для модели SWISS-MODEL изоформы 2 урокиназы

**Заключение.** Таким образом, был построен ряд трехмерных моделей изоформы 2 урокиназы с использованием четырех сервисов. Из результатов, предоставленных каждой программой, было отобрано по одной лучшей модели для сравнительного анализа. На основании 15 проанализированных характеристик, полученных с использованием Structure Assessment и SAVES v6.0,

наилучшими результатами обладает модель SWISS-MODEL, лидируя по 10 величинам и значительно опережая по показателям иные третичные структуры.

В результате проделанной работы для конструирования третичной структуры данной и иных изоформ урокиназы рекомендуется использование сервиса SWISS-MODEL.

#### Список цитированных источников

1. Modelling three-dimensional protein structures for applications in drug design / T. Schmidt [et al.] // *Drug discovery today*. – 2014. – Vol. 19, № 7. – P. 890–897.
2. Wang, X. Matrix Metalloproteinases, Vascular Remodeling, and Vascular Disease / X. Wang, R. A. Khalil // *Advances in pharmacology* (San Diego, Calif.). – 2018. – Vol. 81. – P. 241–330.
3. Madunić, J. The Urokinase Plasminogen Activator System in Human Cancers: An Overview of Its Prognostic and Predictive Role / J. Madunić // *Thromb Haemost.* – 2018. – Vol. 118, № 12. – P. 2020–2036.
4. Therapeutic Strategies Targeting Urokinase and Its Receptor in Cancer / M. T. Masucci [et al.] // *Cancers*. – 2022. – Vol. 14, № 3. – P. 498.
5. SWISS-MODEL: homology modelling of protein structures and complexes / A. Waterhouse [et al.] // *Nucleic Acids Research*. – 2018. – Vol. 46, № W1. – P. W296–W303.
6. Accurate prediction of protein structures and interactions using a three-track neural network / M. Baek [et al.] // *Science*. – 2021. – Vol. 373, № 6557. – P. 871–876.
7. The Phyre2 web portal for protein modeling, prediction and analysis / L. A. Kelley [et al.] // *Nat Protoc.* – 2015. – Vol. 10, № 6. – P. 845–858.
8. Webb, B. Protein Structure Modeling with MODELLER / B. Webb, A. Sali // *Methods Mol Biol.* – 2021. – Vol. 2199. – P. 239–255.

## Selection of an optimal method for construction of the tertiary structure of proteins using the example of urokinase isoform P00749-2

*Ryneiskaya O. N., Bairachenka D. S.*

*Belarusian State Medical University, Minsk, Republic of Belarus*

Urokinase exists in the form of several isoforms, but only one of them has a crystallographic structure, information about which can be obtained in the Protein data bank. Therefore, it is relevant to model the spatial structure of proteins in silico, for which the tertiary structure has not yet been determined using physical research methods. This work compares the quality of 3D models of urokinase isoform P00749-2, obtained using MODELLER, Phyre2, Robetta and SWISS-MODEL. It was revealed that the use of the SWISS-MODEL service is the most optimal for predicting the tertiary structure of urokinase isoforms.

**Keywords:** homologous modelling, urokinase, SWISS-MODEL, Robetta, Phyre2, MODELLER.

*Поступила 05.06.2024*