

## **РЕЗУЛЬТАТЫ АПРОБАЦИИ АЛГОРИТМА ДЛЯ ПРЕДСКАЗАНИЯ ВТОРИЧНОЙ СТРУКТУРЫ БЕЛКОВ И ПЕПТИДОВ**

*Научный руководитель: канд. биол. наук, доц. В. В. Хрусталёв*

*Кафедра общей химии*

*Белорусский государственный медицинский университет, г. Минск*

***Резюме.** Протестирован оригинальный алгоритм PentaFOLD 2.0 для предсказания вторичной структуры белков и пептидов по их аминокислотной последовательности. Работа алгоритма основана на двух вероятностных шкалах. По сравнению с существующими методами алгоритм PentaFOLD 2.0 отличается более высокой чувствительностью к выявлению бета-тяжей. PentaFOLD 2.0. использовался для конструирования вакцинного прионного пептида CC36.*

***Ключевые слова:** оригинальный алгоритм, вторичная структура, вероятностная шкала, чувствительность, специфичность.*

***Resume.** The original algorithm PentaFOLD 2.0 for secondary structure of proteins and peptides prediction using their amino acid sequence as an input has been tested. The work of the algorithm is based on two probability scales. Algorithm PentaFOLD 2.0 has a higher sensitivity for beta-strands than currently existing methods. PentaFOLD 2.0 has been used for construction of vaccine prion peptide CC36.*

***Keywords:** original algorithm, secondary structure, probability scale, sensitivity, specificity.*

**Актуальность.** Предсказание вторичной структуры белка имеет большое значение для медицины при создании вакцин от различных заболеваний на основе коротких пептидов, несущих антигенные свойства полноразмерного белка. Наибольший интерес для исследователей представляют алгоритмы, чья работа не основана на гомологии, которые в своей работе используют различные вероятностные шкалы. Данный тип алгоритмов позволяет предсказать участки прионного белка, в которых возможны структурные переходы, что особенно важно при выборе участка белка для

конструирования пептидов, которые можно использовать в качестве кандидатов для разработки вакцины против прионных заболеваний.

Согласно Chou и Fasman аминокислотные остатки делятся на три группы: те, которые достоверно чаще встречаются в альфа-спиралях (аланин, глутамин, глутаминовая кислота, аргинин, лейцин, метионин, лизин); те, которые чаще встречаются в бета-тяжах (валин, изолейцин, цистеин, фенилаланин, тирозин, триптофан, треонин), и те, которые чаще встречаются в неструктурированных участках (в «койлах») белков и пептидов (пролин, глицин, аспарагиновая кислота, аспарагин, серин, гистидин) [1]. Однако для большинства из них не существует резко выраженных предпочтений для формирования конкретного элемента вторичной структуры, вероятность существования которого возрастает, если соответствующие аминокислоты образуют кластеры. Влияние на фолдинг данного кластера оказывают фланкирующие элементы вторичной структуры, что также учитывается в алгоритме PentaFOLD 2.0.

**Цель** настоящего исследования: протестировать алгоритм для предсказания вторичной структуры белков и пептидов PentaFOLD 2.0.

**Задачи** исследования включали:

1. Создать вероятностный алгоритм PentaFOLD 2.0 для предсказания вторичной структуры белков и пептидов.
2. Сформировать выборки негомологичных белков четырёх структурных классов.
3. Определить чувствительность и специфичность алгоритма по альфа- и бета-паттернам для каждого класса белков.
4. Сравнить результаты предсказаний NPS@Consensus и алгоритма PentaFOLD 2.0 по бета-структурной выборке белков.

**Материалы и методы.** Материалом для создания алгоритма PentaFOLD 2.0 послужила выборка, состоящая из 542 трёхмерных структур бактериальных белков из Protein Data Bank ([www.rcsb.org](http://www.rcsb.org)). Чтобы исключить использование гомологичных последовательностей максимальный процент сходства первичных структур используемых белков друг с другом не превышал 25% по алгоритму Decrease Redundancy ([web.expasy.org/decrease\\_redundancy](http://web.expasy.org/decrease_redundancy)). На основании анализа аминокислотного состава альфа-спиралей, бета-тяжей и участков полипептидной цепи, находящихся в неструктурированном состоянии, была получена аминокислотная вероятностная шкала. В алгоритме PentaFOLD 2.0 вероятность включения аминокислотного остатка в тот или иной элемент вторичной структуры рассчитывается как среднее значение для пентапептида, в центре которого находится рассматриваемый остаток. Чередования гидрофобных и гидрофильных аминокислотных остатков в альфа-спиралях, бета-тяжах и неструктурированных фрагментах изучали во фрагментах длиной в пять аминокислотных остатков (в пентапептидах). Так как в каждом пентапептиде может присутствовать только два вида аминокислотных остатков (гидрофобный или гидрофильный), то максимальное количество пентапептидов равно 32 ( $2^5$ ). Разделение элементов вторичной структуры на пентапептиды проводилось методом скольжения с шагом в одну аминокислоту [2]. Так была получена пентапептидная вероятностная шкала алгоритма PentaFOLD 2.0. Результаты предсказаний вторичной структуры алгоритм выдаёт по двум паттернам – альфа-спиральному и бета-структурному, каждый

из которых основан на результатах предсказаний вторичной структуры как по аминокислотной, так и по пентапептидной шкалам.

На первом этапе работы алгоритм вне зависимости от паттерна предсказывает койл. Поскольку наиболее характерными особенностями обладает койл между двумя бета-тяжами, алгоритм и сравнивает вероятность его образования с таковой для бета-тяжей, расположенных между двумя другими бета-тяжами. Если образование неструктурированного участка для данной аминокислоты имеет высокую вероятность (более 0,6), то алгоритм предсказывает койл. Так алгоритм определяет вероятность образования койла на протяжении всей аминокислотной последовательности белка. Далее, для бета-паттерна алгоритм выбирает наиболее характерные (вероятность выше 0,75) альфа-спиральные пентапептиды (по данным сравнения состава альфа-спиралей, расположенных между двумя бета-тяжами, с таковым для бета-структуры между двух бета-тяжей, и состава альфа-спиралей, расположенных между альфа-спиралью и бета-тяжем в сравнении с аналогичным случаем для бета-структуры), а затем предсказывает наиболее характерные бета-тяжи (вероятность выше 0,6) по результатам сравнения состава бета-тяжей между двумя спиральями с таковым для альфа-спиралей между двумя спиральями. Для альфа-паттерна, наоборот, сначала выбираются наиболее характерные бета-структурные пентапептиды, а потом – с более низким порогом – характерные альфа-спиральные фрагменты. На заключительном этапе по альфа-паттерну как альфа-спиральные предсказываются все остатки, которые являются таковыми хотя бы по одной из двух вероятностных шкал, по бета-паттерну – наоборот [3].

Для тестирования алгоритма PentaFOLD 2.0 были использованы выборки 3D структур белков человека и животных, относящиеся к четырём структурным классам: альфа-спиральному, бета-структурному, «альфа + бета» и «альфа/бета». Каждая выборка включает по 100 негомологичных аминокислотных последовательностей, полученных из международной базы данных PDB. Для исключения гомологии максимальный процент сходства аминокислотных последовательностей белков в каждой выборке друг с другом не превышал 25% по алгоритму Decrease Redundancy. Информация по границам альфа-спиралей и бета-тяжей была получена из PDB файлов по результатам работы алгоритма DSSP [4]. Регионы белков, которые не содержали ни альфа-спиралей, ни бета-тяжей были отнесены к койлу.

Алгоритм для предсказания консенсусной вторичной структуры включает одиннадцать различных вероятностных методов [5]. После обработки первичной последовательности белков каждым из алгоритмов формируется консенсусная вторичная структура. Результаты работы алгоритмов NPS@Consensus по бета-структурной выборке сравнивали с предсказанием вторичной структуры белков данной выборки алгоритмом PentaFOLD 2.0.

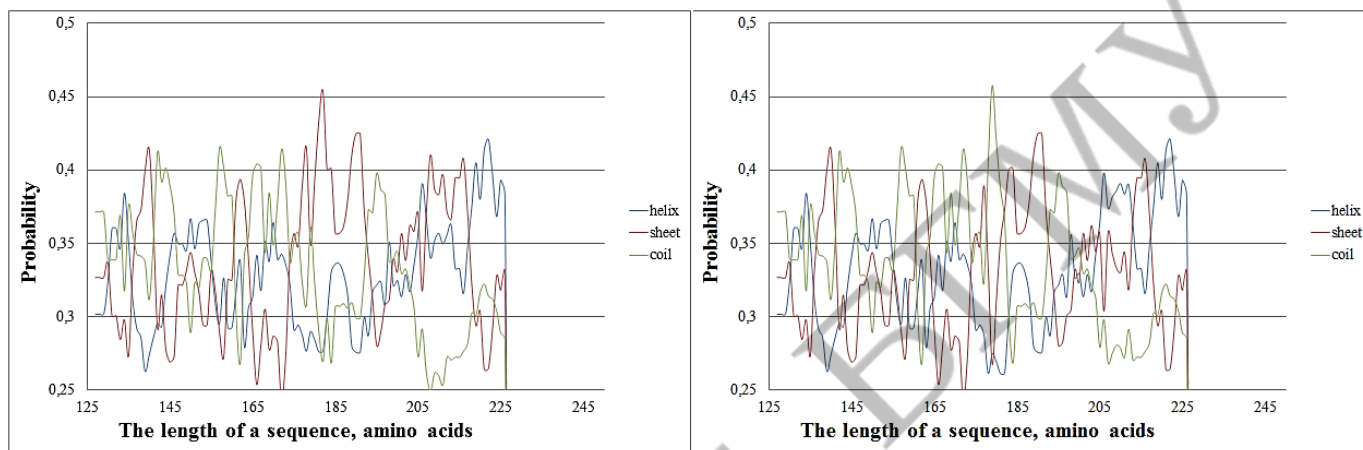
**Результаты и их обсуждение.** Результаты работы алгоритма PentaFOLD 2.0 следующие: чувствительность к альфа-спиралям по альфа-спиральному (66,90%; 65,36%; 65,04% и 67,15%) и бета-структурному паттерну (31,48%; 32,02%; 31,27%; 30,12%) алгоритма не зависит от структурного класса белка (структурные классы белков расположены в порядке – альфа + бета, альфа/бета, бета, альфа). Специфичность к альфа-спиралям демонстрирует ярко выраженную зависимость от структурного

класса. По альфа-спиральному паттерну наибольшего значения (74,50%) специфичность достигла в выборке альфа-спиральных белков, наименьшего (17,06%) – в выборке бета-структурных белков, специфичность в выборке белков альфа/бета класса (52,89%) превысила таковую в выборке белков альфа + бета (48,43%). Похожая ситуация наблюдается и с предсказанием бета-тяжей. Чувствительность к ним как по бета-паттерну, так и по альфа-паттерну находится примерно на одном и том же уровне для белков альфа + бета, бета и альфа классов (61,33% и 29,54%; 59,68% и 27,91%; 55,89 и 26,18%). Несколько более высокие значения характерны для чувствительности к бета-тяжам среди альфа/бета белков: 66,31% и 35,94%. Что касается специфичности к бета-тяжам, то самые высокие её значения достигаются в выборке бета-структурных белков (60,89% и 69,21%), наименьшие (5,48% и 9,25%) – в выборке альфа-спиральных белков. Для белков класса альфа + бета и альфа/бета специфичность к бета-тяжам практически идентична (33,80% и 45,95%; 32,31% и 47,51%). Что касается неструктурированных участков белков, то результаты предсказаний алгоритма PentaFOLD 2.0 по альфа- и бета-паттернам отличаются незначительно. При обработке четырёх выборок чувствительность к койлу как по альфа-паттерну, так и по бета-паттерну находится примерно на одном и том же уровне для белков альфа, альфа+бета, альфа/бета классов (63,30% и 64,07%; 63,47% и 63,95%; 63,70% и 64,01%). Несколько более высокие значения характерны для чувствительности к койлу по альфа- и бета-паттерну бета-структурных белков – 66,92% и 67,10% соответственно. Что касается специфичности к койлу, то самые высокие её значения достигаются также в бета-структурной выборке белков (70,61% и 70,68%), наименьшие (65,29% и 64,50%) – в альфа-спиральной выборке белков. Для белков класса альфа + бета и альфа/бета специфичность к неструктурированным фрагментам практически идентична (69,82% и 69,60%; 68,87% и 68,32%).

Анализируя бета-структурную выборку с помощью алгоритмов NPS@Consensus, выяснено, что чувствительность для бета-тяжей составила 52,87%, а специфичность – 74,83%. Анализируя же смешанную выборку, состоящую из 50 первичных структур белков и пептидов различных структурных классов, алгоритмы из NPS@Consensus предсказывают бета-тяжи с чувствительностью в 50,80% и со специфичностью в 56,41%, в то время как по алгоритму PentaFOLD 2.0 получены результаты – 67,75% и 35,96% соответственно. Полученные результаты говорят о том, что оригинальный алгоритм PentaFOLD 2.0 способствует выявлению тех участков белков, которые склонны к структурным переходам, что способствует совершенствованию критериев отбора фрагментов белков возбудителей инфекционных болезней, на месте которых, при определённых условиях могут сформироваться бета-тяжи, для включения их в состав коротких вакцинных пептидов.

После разработки и тестирования алгоритма PentaFOLD 2.0 он активно использовался для конструирования пептида СС36, содержащего С-конец второй альфа-спирали и N-конец третьей альфа-спирали прионного белка человека. В первоначальной версии пептида должна была присутствовать только одна аминокислотная замена – Phe20 на Trp20. Однако в процессе твёрдофазного синтеза пептида начала образовываться межмолекулярная бета-структура. С помощью алго-

ритма PentaFOLD 2.0 выяснено, что существуют два района бета-структурного потенциала в третьей альфа-спирали прионного белка человека. Так, участок самого сильного бета-структурного потенциала, находящегося в середине третьей альфа-спирали (206-MERVVEQMC-214), образовывал бета-шпильку с участком более слабого бета-структурного потенциала (199-TETDVKM-205), что и останавливало синтез необходимого пептида, вероятно, после формирования межмолекулярной бета-структуры (рисунок 1).



**Рисунок 1** – Предсказанная вторичная структура большого прионного белка человека по данным алгоритма PentaFOLD 2.0. Слева – вторичная структура белка до аминокислотных замен, справа – после аминокислотных замен

Поэтому в новой версии пептида СС36 Met28 был вынужденно заменён на Arg28, а Val32 на Arg32 с целью снизить бета-структурный потенциал С-конца пептида и не допустить образования межмолекулярной бета-структуры во время нового синтеза пептида СС36. В результате произведённых замен бета-шпилька в ходе синтеза С-конца пептида уже не образовывалась. На N-конце пептида Val2 заменён на Pro2 с целью повышения вероятности формирования дисульфидной связи между Cys1 и Cys36.

#### **Выводы:**

1 Создан алгоритм для предсказания вторичной структуры белков и пептидов, работа которого не основана на гомологии, т. е. данный алгоритм работает исключительно с первичной структурой белков и пептидов.

2 С помощью выборок белков, относящихся к четырём структурным классам: альфа-спиральному, бета-структурному, «альфа + бета» и «альфа/бета» протестирован оригинальный алгоритм PentaFOLD 2.0. Определены чувствительность и специфичность данного алгоритма для различных элементов вторичной структуры белков по двум паттернам.

3 Выяснено, что наилучшие результаты алгоритм PentaFOLD 2.0 достигает при использовании альфа-спирального паттерна для альфа-спиральных белков и бета-структурного паттерна – для бета-структурных.

4 Чувствительность алгоритма PentaFOLD 2.0 для бета-тяжей выше, чем у методов из NPS@Consensus, т.е. разработанный алгоритм лучше предсказывает участки первичной структуры белков и пептидов, где может сформироваться бета-тяж.

*V. V. Poboinev*

**RESULTS OF ALGORITHM APPROBATION FOR PREDICTION  
SECONDARY STRUCTURE OF PROTEINS AND PEPTIDES**

*Tutor: associate professor V. V. Khrustalev*

*Department of general chemistry*

*Belarusian State Medical University, Minsk*

**Литература**

1. Chou, P. Y. Prediction of the secondary structure of proteins from their amino acid sequence / P. Y. Chou, G. D. Fasman // *Adv. Enzymol. Relat. Areas. Mol. Biol.* – 1978. – Vol. 47. – P. 45-48.
2. Khrustalev, V. V. Stabilization of secondary structure elements by specific combinations of hydrophilic and hydrophobic amino acid residues is more important for proteins encoded by GC-poor genes / V. V. Khrustalev, E. V. Barkovsky // *Biochimie.* – 2012. – Vol. 94. – P. 2706-2715.
3. Побойнев, В. В. PENTAFOLD 2.0 – алгоритм для предсказаний вторичной структуры белков и пептидов / В. В. Побойнев, В. В. Хрусталёв, Т. А. Хрусталёва // *Инновации в медицине и фармации – 2016: материалы дистанционной научно-практической конференции студентов и молодых учёных.* – Минск: БГМУ. – 2016. – С. 549-554.
4. Kabsch, W. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features / W. Kabsch, C. Sander // *Biopolymers.* – 1983. – Vol. 22. – P. 2577-2637.
5. Combet, C. NPS@: Network Protein Sequence Analysis / C. Combet [et al.] // *TIBS.* – 2000. – Vol. 25. – P.147-150.