

Запросы к базам данных для предсказания и анализа пространственной структуры фермента протеин-дисульфидизомеразы семейства A4 человека и животных

Сушков Р.В.^{1*}, Мочалов А.В.¹, Орлов Ю.Л.^{2, 3**}, Савина Е.А.²

¹ ФГБОУ ВО ВолГМУ Минздрава России, Волгоград, Россия

² Первый МГМУ им. И.М. Сеченова Минздрава России (Сеченовский Университет), Москва, Россия

³ Институт цитологии и генетики СО РАН, Новосибирск, Россия

* roman.sushkov.01@mail.ru; ** orlov@bionet.nsc.ru

Ключевые слова: медицинская информатика; базы данных; протеин-дисульфидизомераза; статистическая обработка; корреляция белковых структур; филогенез

Мотивация и цель: Запросы к биоинформационным базам данных активно используются для предсказания и анализа пространственной структуры белков и ее изменчивости в медицине и генетике. Инструменты биоинформатики помогают анализировать генетические данные в симуляции и моделировании ДНК, РНК и белковых структур, в оценке молекулярных взаимодействий, разработке лекарственных веществ [1]. Протеин-дисульфидизомераза семейства A4 (PDIA4) – изомераза, катализирует образование дисульфидных связей и контролирует перестройки молекул в ЭПР, что обеспечивает их «гибкость», формирует шаперон-опосредованную укладку белка [2]. Нарушение вышеописанных функций приводит к различным дегенеративным изменениям нервной системы, например, болезни Альцгеймера и Паркинсона [3]. Практическая цель данной работы – поиск наиболее схожих по пространственной структуре и функциональной активности белков PDIA4 человека и животных, что может использоваться в медицине для создания лекарственных препаратов на базе его аналогов и для дальнейшего исследования биохимических свойств биомолекулы в лаборатории.

Методы и алгоритмы: Составление списков белков выполнялось на основе запросов к базам данных UniProtKB (<https://uniprot.org/>) и BioGraph (<http://andromeda.matf.bg.ac.rs:54321/>) [4]. В результате был получен список из 28 животных белков, анализ структуры которых мог дать позитивный результат в исследовании. Обработка аминокислотных последовательностей выполнялась в формате FASTA. Проводили попарное выравнивание аминокислотных последовательностей PDIA4 человека и животных в программе Clustal Omega (ebi.ac.uk/jdispatcher/msa/clustalo). Полученные вероятности различия двух последовательностей заносились в таблицу Excel «Анализ сходства первичных последовательностей фермента PDIA4 у человека к этому же ферменту других организмов». Параллельно производилось попарное сравнение аминокислотных последовательностей в системе BLAST. Полученными значениями (Total score; Query cover; E-value; Indent; Positives) дополнили таблицу Excel. Данные были обработаны с помощью ранговых оценок, на основании которых получена сводная таблица, включающая в себя три группы белков: наиболее сходные с белком человека, минимально сходные и медиана. Далее с использованием сайта RCSB

(<https://www.rcsb.org/>) построили 3D модели белков PDIA4. Сравнивались аминокислоты, подверженные модификациям у человеческого и животного белка.

Результаты: В качестве объектов для сравнения использовали белок человека и 28 животных. В результате выравнивания аминокислотных последовательностей в системе Clustal Omega и BLAST было выявлено следующее: наиболее сходен с человеческим вариантом PDIA4 белок хомяка; для проведения исследований *in vivo* следует использовать хомяка или крысу, поскольку они являются стандартным лабораторным животным; для получения чистого фермента и проведения исследований *in vitro* рациональнее рассматривать свинью, ввиду доступности ее тканей. Сравнение 3D структур белка PDIA4 человека и крысы на сайте RCSB дало результат в 0,88 TM-store, высокий процент корреляции.

Выводы: Путем использования запросов к базам данных проведено исследование пространственной структуры белка PDIA4. Произведен поиск первичных аминокислотных последовательностей. Выполнено попарное сравнение белков человека и животных. Получены данные ранговой обработки результатов исследования и построены доверительные интервалы. Найдены и коррелированы 3D модели наиболее близких по структуре белков. Можно сделать вывод о перспективности области исследований по поиску и разработке лекарственных препаратов на матрице белка животных PDIA4 [5], в частности крысы и хомяка, которые показали наибольшую связь с человеческим.

Database queries for prediction and analysis of the spatial structure of the protein disulfide isomerase enzyme of the A4 family of humans and animals

Sushkov R.V.^{1*}, Mochalov A.V.¹, Orlov Y.L.^{2, 3**}, Savina E.A.²

¹ Volgograd State Medical University, Volgograd, Russia

² Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia

³ Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

* roman.sushkov.01@mail.ru; ** orlov@bionet.nsc.ru

Key words: medical informatics; databases; protein disulfide isomerase; statistical processing; correlation of protein structures; phylogeny

Motivation and Aim: Database queries are actively used to predict and analyze the spatial structure of proteins and its variability in medicine and genetics. Bioinformatics tools help to analyze genetic data in the simulation and modeling of DNA, RNA and protein structures, in the assessment of molecular interactions, and in the development of drugs [1]. Protein disulfide isomerase of the A4 family (PDIA4) isomerase, catalyzes the formation of disulfide bonds and controls rearrangements of molecules in the EPR, which ensures their “flexibility”, forms chaperone-mediated protein folding [2]. Violation of the above-described functions leads to various degenerative changes in the nervous system, for example, Alzheimer's and Parkinson's diseases [3]. The practical purpose of this work is to search for the most similar in spatial structure and functional activity of human and animal PDIA4 proteins, which can be used in medicine to create

drugs based on its synthetic analogues, and for further research of the biochemical properties of biomolecules in laboratories.

Methods and Algorithms: Protein lists were compiled based on queries to UniProtKB databases (<https://uniprot.org/>) and BioGraph (<http://andromeda.matf.bg.ac.rs:54321/>) [4]. As a result, a list of 28 animal proteins was obtained, the analysis of the structure of which could give a positive result in the study. The processing of amino acid sequences was performed in the FASTA format. Pairwise alignment of human and animal PDIA4 amino acid sequences was performed in the Clustal Omega program (ebi.ac.uk/jdispatcher/msa/clustalo). The obtained probabilities of differences between the two sequences were entered into an Excel spreadsheet “Analysis of the similarity of the primary sequences of the PDIA4 enzyme in humans to the same enzyme of other organisms”. In parallel with the above, a pairwise comparison of amino acid sequences in the BLAST system was performed. The obtained values (Total score; Query cover; E-value; Indent; Positives) were added to the Excel table. The data were processed using rank estimates, on the basis of which a summary table was obtained, including three groups of proteins: the most similar to human protein, the least similar and the median. Further using the RCSB website (<https://www.rcsb.org/>) we built 3D models of PDIA4 proteins. Amino acids subject to modifications in human and animal proteins were compared.

Results: Human protein and 28 animals were used as objects for comparison. As a result of the alignment of amino acid sequences in the Clustal Omega and BLAST system, the following was revealed: the hamster protein is most similar to the human variant of PDIA4; for *in vivo* studies, a hamster or rat should be used, since they are a standard laboratory animal; for obtaining a pure enzyme and conducting *in vitro* studies, it is more rational to consider a pig, due to the accessibility of her tissues. A comparison of the 3D structures of the human and rat PDIA4 protein on the RCSB website yielded a result of 0.88 TM-store, a high percentage of correlation.

Conclusion: By using database queries, a study of the spatial structure of the PDIA4 protein was conducted. A search for primary amino acid sequences has been performed. A pairwise comparison of human and animal proteins was performed. The data of the rank processing of the research results were obtained, and confidence intervals were constructed. 3D models of the proteins closest in structure have been found and correlated. It can be concluded that the field of research on the search and development of drugs based on the animal protein matrix PDIA4 is promising [5]. In particular, rats and hamsters, which showed the greatest connection with the human.

Список литературы/References

1. Muldowney M.W., Duncan K.R., Elsayed S.S. et al. Artificial intelligence for natural product drug discovery. *Nat Rev Drug Discov.* 2023;22(11):895-916. doi 10.1038/s41573-023-00774-7
2. Fu J., Gao J., Liang Z., Yang D. PDI-Regulated Disulfide Bond Formation in Protein Folding and Biomolecular Assembly. *Molecules.* 2020;26(1):171. doi 10.3390/molecules26010171
3. Medinas D.B., Rozas P., Hetz C. Critical roles of protein disulfide isomerases in balancing proteostasis in the nervous system. *J Biol Chem.* 2022;298(7):102087. doi 10.1016/j.jbc.2022.102087
4. Veljković A.N., Orlov Y.L., Mitić N.S. BioGraph: Data Model for Linking and Querying Diverse Biological Metadata. *Int J Mol Sci.* 2023;24(8):6954. doi 10.3390/ijms24086954
5. Kuo T.F., Hsu S.W., Huang S.H. et al. Pdia4 regulates β -cell pathogenesis in diabetes: molecular mechanism and targeted therapy. *EMBO Mol Med.* 2021;13(10):e11668. doi 10.15252/emmm.201911668