



УДК 577.112.5.087

© 1992 г. И. И. Парилис, Е. Ю. Казанов,  
Р. С. Салихов, Д. Х. ХамидовКОМПЬЮТЕРНЫЙ АНАЛИЗ И СРАВНЕНИЕ СТРУКТУР  
НЕЙРОТОКСИНОВ ЗМЕЙ И АКТИНИИ  
*RADIANTHUS MACRODACTYLUS*

Институт биохимии АН Республики Узбекистан, Ташкент

Результаты компьютерных методов сопоставления исследуемых белков позволили восстановить их эволюционные деревья и выявить уровень структурно-функциональной зависимости.

Введенное расстояние между процентными аминокислотными составами было использовано для классификации группы токсинов из актинии *Radianthus macrodactylus*. Они оказались длинными нейротоксинами. Классификационные баллы были рассчитаны на основании теории распознавания образов.

Было произведено разделение выборки из 10 фосфолипаз A2 на две группы по их принадлежности к различным организмам с использованием этого нового расстояния.

Сопоставление гомологичных белковых молекул с целью построения филогенетического древа производится обычно путем сравнения выравненных по длине аминокислотных последовательностей. При этом для каждой пары сравниваемых белков оценивают «расстояние» с использованием простейшей матрицы 0-1 или матрицы 0-3. В первом случае подсчитывают число различающихся аминокислотных остатков, а во втором — минимально возможное количество замещений нуклеотидных остатков в кодонах соответствующих аминокислотных остатков в сравниваемых позициях [1, 2].

При определении этих «расстояний» для белков различной длины более короткий приходится искусственно «вытягивать» за счет введения в определенные позиции дефисов, т. е. «разрывать» исходную структуру. Нами была введена новая оценка «расстояний», при которой не приходится «перекраивать» нативную первичную структуру, так как используется не аминокислотная последовательность, а аминокислотный состав сравниваемых белков. Такое «расстояние», называемое «эвклидовым», вычисляется как квадратный корень из суммы квадратов разностей процентного содержания одноименных аминокислот в составах белков. В этом случае каждому белку соответствует в 20-мерном пространстве точка, координатами которой являются процентные содержания каждой из 20 аминокислот. Это унифицированное числовое признаковое пространство было использовано нами для классификации большой выборки (порядка 100) токсических полипептидов и для идентификации новых (не вошедших в обучение) белков с использованием алгоритмов теории распознавания образов [3, 4].

Нами были впервые в работе [4] модифицированы для сопоставления белковых структур описанные ранее [3] алгоритмы теории распознавания образов. В основу теории положен расчет информативности всех признаков исследуемых объектов, отбор наиболее информативных при рассматриваемой классификации и вычисление соответствующих баллов, позволяющих идентифицировать объект, не вошедший в выборку обучения. Этот метод

получил широкое распространение при математическом прогнозировании в геологической разведке нефти, в медицинской диагностике, при компьютерном конструировании лекарственных препаратов. Эффективность его применения определяется прежде всего удачным выбором признаков, выбранных для описания исследуемых объектов. Для токсических полипептидов были использованы процентные составы лишь четырех наиболее информативных аминокислот, что позволило по вычисленным баллам произвести разделение токсинов на нейро- и цитотоксины [4].

Интерес к определению расстояния между белками с нерасшифрованной первичной структурой проявился во многих работах, однако подходы и цели были различны. Так, например, в работе [5] определяется эволюционная близость таких белков и строится филогенетическое древо на основе данных электрофореза и «иммунологических расстояний».

В работе [6] предложен статистический метод определения скорости эволюции белков, использующий среднее от стандартных отклонений по каждой входящей в рассматриваемые белки аминокислоте. Корректность метода доказывается сопоставлением полученных результатов с расчетами, основанными на сравнении первичных структур. Однако его применение связано с рядом требований, ограничивающих его возможности, а результаты характеризуют лишь эволюционную близость группы гомологичных белков, а не «расстояние» между отдельными белками.

Основной задачей наших исследований [4, 7, 8] является сопоставление различных белков на основе введенного попарного «расстояния» между их аминокислотными составами. Результаты компьютерных расчетов этих «расстояний» используются для классификации белков по их основной функции, для идентификации новых белков, а также для восстановления эволюционной истории рассматриваемых белков (в виде филогенетических деревьев) по любому из существующих алгоритмов [1, 2, 5]. В качестве примера в настоящей работе проведено сравнение анемонотоксинов из работ [9–12] с различными токсинами из ядов змей [13] и проинсулинами лошади, утки, крысы и человека [1].

В табл. 1 приведены попарные евклидовы расстояния для 22 белков из четырех групп: анемонотоксинов (RTX) из актинии *Radianthus macrodactylus* [9–12], длинных и коротких нейротоксинов из ядов змей [13] и проинсулинов [1].

Из таблицы видно, во-первых, что данные белки разделяются на четыре группы, так как попарные расстояния внутри каждой из них меньше, чем для представителей разных групп. Во-вторых, построчное сканирование строк 1–5 таблицы показывает, что анемонотоксины ближе к длинным нейротоксинам (с номерами 6–11), чем к коротким нейротоксинам (12–17) и проинсулинам (18–22). Наконец, показано, что самыми близкими из анемонотоксинов являются RTX-III и RTX-V, значит, в процессе эволюции они дивергировали последними. Этот результат совпадает с результатом, полученным при сопоставлении аминокислотных последовательностей данных белков по матрицам 0–1 и 0–3 [1, 2], что показывает применимость нашего метода для построения филогенетических деревьев.

Аминокислотные последовательности анемонотоксинов содержат 47–48 остатков, в то время как нейротоксины имеют длину 60–75, что исключает возможность их попозиционного сравнения, так как в структуру анемонотоксинов пришлось бы при выравнивании вводить до 50% делеций.

Таким образом, предложенная методика эффективна при анализе белков, не сопоставимых другими методами.

В качестве другого примера была рассмотрена группа из 7 фосфолипаз A2 змей из рода *Naja* и 3 – из рода *Bitis*, аминокислотные последовательности которых сопоставлялись в работе [14]. Там же с применением матрицы 0–1 была подтверждена принадлежность фосфолипаз к одной из

Попарные эвклидовы расстояния между аминокислотными составами белков различных групп \*

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	12,2	7,8	15,0	15,0	16,7	14,3	14,4	14,6	13,7	13,8	21,2	20,6	20,7	20,5	20,0	19,2	22,0	21,7	26,4	25,9	25,0
2		8,3	8,3	6,9	14,3	11,6	9,3	9,8	8,4	11,8	14,8	14,7	15,1	14,3	13,9	12,6	18,3	18,8	21,4	21,2	21,0
3			13,2	12,4	14,2	11,7	10,3	10,5	8,9	10,6	15,9	14,8	15,3	14,9	14,3	14,1	20,0	20,2	24,7	23,3	22,8
4				3,5	15,8	14,3	12,7	12,7	11,9	14,2	18,5	18,3	18,3	17,7	17,3	17,0	17,6	18,2	20,3	19,3	20,3
5					14,5	13,2	11,4	11,4	11,0	13,5	17,0	17,3	17,0	16,3	16,0	15,8	16,9	17,3	19,3	18,1	19,4
6						7,4	8,4	7,9	9,6	9,7	17,5	17,5	17,0	15,8	14,7	16,3	19,0	20,1	23,9	21,2	22,7
7							7,4	7,3	7,6	6,6	15,3	15,1	14,8	13,6	12,8	13,6	17,8	18,6	22,8	21,1	21,4
8								2,3	3,6	7,1	13,4	13,2	13,4	12,0	11,3	12,3	18,7	19,6	22,3	21,1	21,9
9									3,9	6,9	13,6	13,6	13,6	12,2	11,6	12,5	18,7	19,6	22,7	20,9	22,1
10										6,4	13,8	12,8	13,2	11,8	11,2	11,7	18,9	20,0	23,2	21,5	22,5
11											14,7	14,0	13,7	12,8	13,0	14,2	18,6	19,4	23,3	21,0	21,5
12												4,7	4,1	4,1	5,8	7,1	22,0	21,9	25,1	22,9	22,8
13													3,3	3,3	4,7	6,7	21,6	21,8	26,0	23,2	22,8
14														2,4	4,7	7,5	21,5	21,5	25,8	22,7	22,4
15															3,3	6,2	20,9	21,2	25,1	22,1	22,3
16																5,8	20,4	20,8	24,8	22,2	22,3
17																	20,2	20,8	24,7	22,6	22,7
18																		3,7	40,4	9,0	7,5
19																			10,1	10,2	6,2
20																				10,7	10,0
21																					10,5

\* Токсены актинии *R. mastochoctylus* (9—12); анемонотоксены I (4), IV (2), II (3), III (4), V (5); длинные нейротоксены змей [7]: 6 — токсин *Acanthocheilus* b, 7 — *Notechis scutatus* III-4; 8 — *Dendroaspis viridis* токсин I, 9 — *D. viridis* токсин V, 10 — *D. viridis* токсин 4, 9, 3; 11 — *Bungarus multicinctus* — *Bungarus* токсин; короткие нейротоксены змей [7]: 12 — *Hydrophis cyanocinctus* токсин с, 13 — *H. cyanocinctus* токсин а, 14 — *H. cyanocinctus* токсин b, 15 — *Larabis kardecshii* Lh, нейротоксин, 16 — *H. kardecshii* s, 17 — *Aripasarus laevis* s, прописулины позвоночных: 18 — *Rattus norvegicus* прописулин-1, 19 — *R. norvegicus* прописулин-11, 20 — *Egus sabatius*, 21 — *Anas platyrhynchos*, 22 — *Homo sapiens*.

Содержание информативных аминокислотных остатков и классификационные баллы для отнесения токсинов змей

Содержание, %	Классификационные баллы			Содержание, %	Классификационные баллы		
	Нейротоксины		Цитоток- сины		Нейротоксины		Цитоток- сины
	длинные	короткие			длинные	короткие	
Gly 0-5,4	12	0	26	Trp 0-1	2	0	23
5,4-8,8	23	23	1	1-1,5	22	3	0
>8,8	0	12	0	1,5-1,7	0	24	16
Cys 0-12,7	6	3	0	>1,7	25	8	0
12,7-13,3	1	22	25	Gln 0-3,1	26	0	27
>13,3	29	20	0	>3,1	15	30	0

Таблица 3

Сравнение анемонотоксинов актинии RTX-I - RTX-V с нейротоксинами и цитотоксинами змей

Содержание информа- тивных аминокислот- ных остатков, %	Классификационные баллы		
	Нейротоксины		Цитотоксины
	длинные	короткие	
RTX-I и RTX-II			
Gly 6,25	23	23	1
Cys 12,5	6	3	0
Trp 2,2	25	8	0
Gln 0	26	0	27
	Всего 80	34	28
RTX-III - RTX-V			
Gly 10,4	0	12	0
Cys 12,5	6	3	0
Trp 2,1	25	8	0
Gln 0	26	0	27
	Всего 57	23	27

двух родов и построено филогенетическое древо. Нами воспроизведены эти же результаты с использованием лишь аминокислотных составов рассматриваемых белков.

Разработанная нами методика требует в каждом случае применения компьютерных расчетов, использующих процентное содержание всех 20 аминокислот, без учета их информативности. Применение алгоритмов теории распознавания образов обладает двумя преимуществами: во-первых, производится оценка признаков по их информативности и отбор необходимого числа для решения каждой классификационной задачи, во-вторых, после расчета баллов пользователь определяет принадлежность каждого нового белка одному из рассматриваемых классов суммированием соответствующих баллов [4].

При идентификации анемонотоксинов по их аминокислотным составам к одной из трех групп токсинов — длинных и коротких нейротоксинов и цитотоксинов выявлены наиболее информативные признаки. Имми оказались процентные содержания глицина, цистеина, триптофана и глутамина. Соответствующие классификационные баллы приведены в табл. 2, а их

применение к рассматриваемым белкам — в табл. 3. По максимальной сумме баллов все анемонотоксины ближе всего к длинным нейротоксинам.

Нами создан оригинальный пакет прикладных программ для сопоставления белков по их аминокислотному составу, описанный в работе [7]. Исследователи, решающие подобные задачи, могут приобрести пакет у авторов.

#### СПИСОК ЛИТЕРАТУРЫ

1. Dayhoff M. O. Atlas of Protein Sequence and Structure. 1972. V. 5; 1973. Suppl. 1; 1976. Suppl. 2; 1979. Suppl. 3. Washington D. C.: Nat. Biomed. Res. Foundation.
2. Fitch W. M., Margoliash E. // Science. 1967. V. 155. № 2. P. 279–284.
3. Вапник В. Н., Глазкова Т. Г., Коцеев В. А., Михальский А. И., Червоенник А. Я. Алгоритмы и программы восстановления зависимостей. М.: Наука, 1984.
4. Парилис И. И., Буссель Г. Л., Юкельсон Л. Я., Хамидов Д. Х. // Молекулярн. биология. 1988. Т. 22. № 6. С. 1697–1701.
5. Айала Ф. Дж. // Журн. общей биологии. 1986. Т. XLVII. № 4. С. 479–493.
6. Озиевецкая М. М. // Молекуляр. биология. 1977. Т. 11. № 1. С. 74–81.
7. Парилис И. И., Казанов Е. Ю., Салихов Р. С., Юкельсон Л. Я., Хамидов Д. Х. // Биополимеры и клетка. 1991. Т. 7. № 6. С. 88–91.
8. Parilis I. I., Kazanov E. Yu. // Abst. International Conference on Modelling and Computer Methods in Molecular Biology and Genetics (Novosibirsk, 1990). P. 65–66.
9. Зыкова Т. А., Винокуров Л. М., Маркова Л. Ф., Козловская Э. П., Еляков Г. Б. // Биоорганич. химия. 1985. Т. 11. № 3. С. 293–301.
10. Зыкова Т. А., Козловская Э. П., Еляков Г. Б. // Биоорганич. химия. 1988. Т. 14. № 7. С. 878–882.
11. Зыкова Т. А., Козловская Э. П., Еляков Г. Б. // Биоорганич. химия. 1988. Т. 14. № 11. С. 1489–1494.
12. Зыкова Т. А., Козловская Э. П. // Биоорганич. химия. 1989. Т. 15. № 10. С. 1301–1306.
13. Endo T., Tamiya N. // Pharmac. Ther. 1987. V. 34. P. 403–451.
14. Костецкий П. В., Пожилыцова О. И., Ульяшин В. В. // Молекуляр. биология. 1990. Т. 24. № 6. С. 1590–1596.

Поступила в редакцию  
28.V.1991

После доработки  
18.XI.1991

I. I. PARILIS, E. Yu. KAZANOV, R. S. SALIKHOV, D. Kh. KHAMIDOV

#### COMPUTER ANALYSIS AND COMPARISON OF NEUROTOXINS FROM SNAKES AND THE SEA ANEMONE *RADIANTHUS MACRODACTYLUS*

*Institute of Biochemistry, Academy of Sciences of the Uzbek Republic, Tashkent*

Evolutionary trees of proteins analysed by means of computer comparative methods were reconstructed and the level of the structure and functional relationship was revealed.

The distance between the percent amin acid contents was used to classify a group of toxins from the sea anemone *Radianthus macrodactylus*, wich proved to belong to long neurotoxins. The scores were calculated by the pattern recognition method. By the use of the distance values, ten phospholipases A2 were divided into two groups, containing representatives of the most closely related organisms.