

8. Pajot P. Fluorescence of protein in 6M guanidine hydrochloride. A method for the quantitative determination of tryptophan // Eur. J. Biochem.—1976.—63, N 1.—P. 263—270.
9. Burstein E. A., Vedenkina N. S., Ivkova M. N. Fluorescence and the location of tryptophan residues in protein molecules // Photochemistry and Photobiology.—1973.—18, N 2.—P. 263—279.
10. Демченко А. П. Люминесценция и динамика структуры белков.— Киев: Наук. думка, 1988.— 278 с.

Ин-т молекуляр. биологии и генетики АН УССР, Киев

Получено 10.06.91

УДК 577.112.5.087

И. И. Парилис, Е. Ю. Казанов, Р. С. Салихов,  
Л. Я. Юкельсон, Д. Х. Хамидов

### ПАКЕТ ПРИКЛАДНЫХ ПРОГРАММ ДЛЯ КЛАССИФИКАЦИИ БЕЛКОВ ПО ИХ АМИНОКИСЛОТНОМУ СОСТАВУ И ПЕРВИЧНОЙ СТРУКТУРЕ

*В работе предложен метод классификации белков по аминокислотным составам и первичным структурам на основе нового попарного расстояния. Описан пакет прикладных программ для реализации этого метода на IBM-совместимой персональной ЭВМ. Входящие в пакет программы дают возможность сопоставлять полученные для нового расстояния результаты с традиционными. Приводится достаточный критерий для классификации белков.*

В настоящее время накоплено большое количество данных о первичных структурах белков и пептидов в различных атласах и банках [1]. Например, в банке PIR (США) более 1,5 миллиона аминокислотных остатков. Сопоставление этих структур для ориентировочного определения функциональной принадлежности белка и восстановления филогенетического древа производится обычно по двум типам расстояний:

1) попозиционное сравнение аминокислот (расчет числа различий в парах гомологично расположенных аминокислотных последовательностей) — расстояние по матрице 0—1;

2) мутационное расстояние (минимальное количество замещений нуклеотидных остатков в цепи ДНК для замены каждой аминокислоты на другую в соответствующей позиции белковой молекулы) — расстояние по матрице 0—3 [1, 2].

При определении этих расстояний для белков различной длины более короткий приходится искусственно «вытягивать» за счет введения в определенные позиции делеций, т. е. «разрывать» исходную структуру. Нами было введено новое расстояние, при расчете которого не приходится «перекраивать» нативную структуру, так как используется не аминокислотная последовательность, а аминокислотный состав сравниваемых белков. Это расстояние, названное «эвклидовым», рассчитывается как квадратный корень из суммы квадратов разности процентного содержания одноименных аминокислот в составах белков. В этом случае каждый белок можно определить в двадцатимерном пространстве точкой, координатами которой являются процентные содержания каждой из 20 аминокислот. Такое унифицированное признаковое пространство было использовано нами для классификации большой выборки (порядка 100) токсических полипептидов и для идентификации новых белков с применением алгоритмов теории распознавания образцов [3—5].

Другим важным преимуществом такого подхода является возможность сопоставления белков с нерасшифрованной первичной структурой

© И. И. ПАРИЛИС, Е. Ю. КАЗАНОВ, Р. С. САЛИХОВ, Л. Я. ЮКЕЛЬСОН, Д. Х. ХАМИДОВ, 1991.

даже в случае малой выборки. До сих пор попытки классификации полипептидов с неизвестной первичной структурой осуществлялись на основе данных электрофореза и иммунологии [6].

Вычисленные попарные расстояния могут использоваться для восстановления эволюционной истории рассматриваемых белков (в виде филогенетических деревьев) по любому из существующих алгоритмов [5].

Для автоматизации анализа информации с целью классификации, идентификации и выявления функциональной близости белков и полипептидов, а также получения данных для построения филогенетического

2 3 4 5 6 7 8 9 10 11 12 13 14	2 3 4 5 6 7 8 9 10 11 12 13 14
1 45 45 43 84 189 184 248 225 201 215 214 145 161	1 13 18 21 43 115 114 114 116 113 114 115 117 116
2 20 40 90 183 178 242 222 198 209 211 146 150	2 8 16 41 115 115 114 116 113 113 114 117 116
3 33 82 178 172 233 217 192 203 204 146 147	3 17 42 116 116 115 117 114 114 115 117 116
4 76 168 165 224 209 186 192 193 135 141	4 39 113 113 113 115 112 112 113 115 114
5 187 189 235 203 201 212 208 170 178	5 112 112 113 114 111 111 114 115 114
6 36 103 90 75 82 75 116 116	6 4 51 58 16 40 47 95 97
7 101 101 61 78 67 113 112	7 50 59 15 39 46 96 97
8 106 100 65 69 176 165	8 48 50 30 39 94 96
9 104 96 93 164 169	9 56 50 54 95 97
10 83 70 136 137	10 39 41 95 97
11 46 137 129	11 38 96 97
12 134 128	12 95 97
13 65	13 19

Рис. 1. Попарные евклидовы расстояния, рассчитанные по аминокислотным составам для факторов роста нервов (NGF) (1 — мыши; 2 — человека; 3 — быка; 4 — цыпленка; 5 — змеи), проинсулинов (PI) (6, 7 — крысы; 8 — лошади; 9 — утки; 10 — человека; 11 — быка; 12 — свиньи) и инсулиноподобных факторов роста человека (ILGF) (13, 14)

Рис. 2. Количество попарных различий в аминокислотных последовательностях (здесь и на рис. 3 обозначения, как на рис. 1)

древа нами написана программа для IBM-совместимого персонального компьютера.

Программа выполнена в виде системы иерархического меню.

Главное меню содержит четыре альтернативных субменю: а) работа с записями; б) расчет расстояний; в) печать отчетов; г) выход в DOS.

Субменю «Работа с записями» позволяет вводить, редактировать, просматривать, удалять записи, рассчитывать аминокислотный состав белков и пептидов.

Субменю «Расчет расстояний» позволяет: 1) определять общие участки двух белков (при этом альтернативные аминокислоты в переменных позициях замещаются звездочками) с вычислением всех трех расстояний; 2) определять консенсус для белков одного семейства; 3) вычислять все три расстояния для белков нескольких семейств, при этом результат работы может выводиться как на принтер, так и в файл для последующего редактирования, печати и хранения. Расстояния выводятся в виде треугольной матрицы. Расстояние по матрице 0—3 во всех случаях рассчитывается со «штрафом за делецию», равным 3; 4) определять позиционное процентное содержание аминокислот (для одного семейства), что может быть использовано для прогнозирования аминокислотных последовательностей изофункциональных полипептидов.

Субменю «Печать отчетов» позволяет выводить на принтер первичные структуры, количественные и процентные составы белков.

Программа написана для СУБД «FoxBase+» и реализована на персональном компьютере «Amstrad 1640» с жестким диском на 32 мегабайта и матричным принтером «Amstrad DMP4000».

В качестве примера работы приводятся значения всех трех расстояний, вычисленные с помощью данной программы для факторов

роста нервной ткани (NGF) [7], проинсулинов (PI) и инсулиноподобных факторов роста человека (ILGF) [1] (рис. 1—3).

Для определения функциональной принадлежности того или иного белка вычисляются максимумы попарных расстояний внутри каждого класса и минимум для белков из различных классов. Из их сопоставления делается вывод о разделении белков на непересекающиеся классы.

Достаточным условием для кластеризации является выполнение неравенства:

$$\min D(x, y) > \{\max D(x, x), \max D(y, y)\},$$

где  $\max D(x, x)$  — максимальное значение расстояния для белков одного класса;  $\max D(y, y)$  — для белков другого класса;  $\min D(x, y)$  — минимальное расстояние между белками разных классов.

Для эвклидова расстояния (см. рис. 1):

$\max(\text{NGF}, \text{NGF})=9,0$ ;  $\max(\text{PI}, \text{PI})=10,6$ ;  $\max(\text{ILGF}, \text{ILGF})=6,5$ ;  
 $\min(\text{NGF}, \text{PI})=16,5$ ;  $\min(\text{NGF}, \text{ILGF})=13,5$ ;  $\min(\text{PI}, \text{ILGF})=11,2$ .

Для количества различий (см. рис. 2):

$\max(\text{NGF}, \text{NGF})=43$ ;  $\max(\text{PI}, \text{PI})=59$ ;  $\max(\text{ILGF}, \text{ILGF})=19$ ;  
 $\min(\text{NGF}, \text{PI})=111$ ;  $\min(\text{NGF}, \text{ILGF})=114$ ;  $\min(\text{PI}, \text{ILGF})=94$ .

Для мутационных расстояний (см. рис. 3):

$\max(\text{NGF}, \text{NGF})=64$ ;  $\max(\text{PI}, \text{PI})=105$ ;  $\max(\text{ILGF}, \text{ILGF})=37$ ;  
 $\min(\text{NGF}, \text{PI})=224$ ;  $\min(\text{NGF}, \text{ILGF})=250$ ;  $\min(\text{PI}, \text{ILGF})=220$ .

Из приведенных рисунков видно, что введенное нами новое расстояние между аминокислотными составами является адекватным при раз-

	2	3	4	5	6	7	8	9	10	11	12	13	14
1	14	19	29	64	231	229	228	247	228	227	226	264	256
2		8	22	62	233	231	229	245	226	228	226	266	258
3			23	63	234	232	231	245	227	230	228	267	259
4				58	231	229	234	245	228	231	228	259	251
5					227	226	228	245	224	226	227	258	250
6						4	87	104	18	66	72	221	220
7							85	105	16	64	70	221	220
8								83	84	43	68	222	225
9									103	91	98	231	230
10										63	66	220	220
11											67	224	227
12												221	222
13													37

делении белков по функциям, т. е. для их классификации, а также для идентификации новых белков. К этому заключению приводит аналогичный вычислительный эксперимент, проведенный для ряда групп изофункциональных белков: длинных и коротких пейротоксинов, цитотоксинов, сарафотоксинов и эндотелиинов, анемотоксинов [8—10].

Рис. 3. Попарные мутационные расстояния для аминокислотных последовательностей

Данный метод позволяет идентифицировать функциональную принадлежность белков со значительно отличающейся длиной. Так, нами была установлена функциональная близость токсинов RTX-I—RTX-V из актинии *Radianthus macrodactylus* к длинным нейротоксинам. Аминокислотные последовательности анемотоксинов содержат 47—48 остатков, в то время как нейротоксины имеют длину от 60 до 75, что исключает возможность их попозиционного сравнения.

#### Резюме

В роботі запропоновано метод класифікації білків по амінокислотному складу та первинній структурі на основі нової попарної відстані. Описано пакет прикладних програм для реалізації цього методу на IBM-сумісній персональній ЕОМ. Програми, що входять в пакет, дають можливість співставляти одержані для нової відстані результати з традиційними. Приведено достатній критерій для класифікації білків.

#### Summary

A method of protein classification by their amino acid composition and primary structures using a new distance is proposed. A program for realization of this method for an

IBM-compatible personal computer is described. Presented software package allows to compare the results obtained by the new method with the traditional ones. A sufficient criteria for protein classification have been developed.

#### СПИСОК ЛИТЕРАТУРЫ

1. Dayhoff M. O. Atlas of protein sequence and structure.— Washington, 1973.— Suppl. 1.— P. 112.
2. Fitch W. M., Margoliash E. Construction of phylogenetic trees // Science.— 1967.— 155, N 2.— P. 279—284.
3. Вапник В. Н. Восстановление зависимостей по эмпирическим данным.— М.: Наука, 1979.— 340 с.
4. Классификация и идентификация токсических полипептидов методом распознавания образов / И. И. Парилис, Г. Л. Буссель, Л. Я. Юкельсон, Д. Х. Хамидов // Молекуляр. биология.— 1988.— 22, № 6.— С. 1697—1701.
5. Сравнительный анализ факторов роста нервов (ФРН) с применением компьютерных программ / И. И. Парилис, Р. С. Салихов, Д. Б. Мирходжаев и др. // Докл. АН УзССР.— 1990.— 11, № 1.— С. 53—55.
6. Айла Ф. Дж. Эволюция: Молекулярная биология предлагает эффективные методы реконструкции филогении // Журн. общ. биологии.— 1986.— 47, № 4.— С. 479—493.
7. Selby M. J., Edwards R. H., Rutter W. J. Cobra NGF: structure and evolutionary comparison // Neurosci. Res.— 1987.— 18, N 2.— С. 293—298.
8. Homology between snake venom sarafotoxins and mammalian endothelins / D. Graur, A. Bdelah, Z. Wollberg, E. Kochva // Isr. J. Zool.— 1988/1989.— 35.— P. 171—175.
9. Taniya N., Maeda N., Cogger H. G. Neurotoxins from the venoms of the sea snakes *Hydrophis ornatus* and *Hydrophis lapemoides* // Biochem. J.— 1983.— 213, N 1.— P. 36—38.
10. Зыкова Т. А., Козловская Э. П. Аминокислотная последовательность нейротоксина 1 из актинии *Radianthus macrodactylus* // Биоорг. химия.— 1989.— 15, № 10.— С. 1301—1306.

Ин-т биохимии АН УзССР, Ташкент

Получено 26.03.91

УДК 577.15

Е. О. Кальчева, М. М. Файзишев, С. С. Малюта

#### УРИДИНДИФОСФАТ-N- АЦЕТИЛМУРАМИЛ-L-АЛАНИЛ-D-ГЛЮТАМАТ: МЕЗО-2, 6-ДИАМИНОПИМЕЛАТЛИГАЗНАЯ АКТИВНОСТЬ В ЭКСТРАКТАХ КЛЕТОК *BACILLUS SUBTILIS*, *STREPTOCOCCUS BOVIS* И *ENTEROCOCCUS FAECIUM*

В экстрактах клеток *B. subtilis*, *S. bovis*, *E. faecium* обнаружена УДФ-N-ацетилмурамил-L-аланил-D-глутамат: мезо-2,6-ДАП-лигазная активность. УДФ-N-ацетилмурамил-L-аланил-D-глутамат: мезо-2,6-ДАП-лигаза *B. subtilis* выделена и очищена гель-фильтрацией и ионообменной хроматографией. Молекулярная масса фермента, определенная методом гель-фильтрации, составляет 120 000.

**Введение.** Мезо-диаминопимелиновая кислота (мезо-ДАП) является непосредственным предшественником лизина в его биосинтетическом пути и в то же время — важным компонентом пептидогликана бактериальной клеточной стенки (рис. 1). Декарбоксилирование мезо-ДАП в лизин осуществляется ферментом мезо-ДАП-декарбоксилазой (ЕС 4.1.1.20), а включение мезо-ДАП в состав УДФ-N-ацетилмурамил-трипептида катализируется УДФ-N-ацетилмурамил-L-аланил-D-глутамат: мезо-2,6-ДАП-лигазой (ЕС 6. 3. 2. 13). ДАП-лигазная активность обнаруживается в клетках бактерий, клеточная стенка которых содержит мезо-ДАП [1, 2]. В настоящее время указанный фермент выделен из *Corynebacterium xerosis* [3], *B. cereus* [4], *Escherichia coli* [5]; показана его способность функционировать как в прямом, так и в обратном

© Е. О. КАЛЬЧЕВА, М. М. ФАЙЗИШЕВ, С. С. МАЛЮТА, 1991