

posed as the first step to decrease the number of comparisons. The both methods have been tested on the model sequences as well as on the prokaryotic origins of replication.

#### СПИСОК ЛИТЕРАТУРЫ

1. Kornberg A. Mechanisms of replication of the *Escherichia coli* genome // Eur. J. Biochem.—1983.—197, N 3.—P. 377—382.
2. Уотсон Дж., Туз Дж., Куртс Д. Рекомбинантная ДНК.— М.: Мир, 1986.— 285 с.
3. Pribnow D. Biological regulation and development // Gene expression.— New York: Plenum press, 1979.— Vol. 1, N 4.—P. 219—277.
4. The *ccs* element, a 16 base pair palindrome essential for activity of the otopine synthase enhancer / J. G. Ellis, D. J. Llewellyn, J. C. Walker et al. // EMBO J.—1987.— 6, N 11.—P. 3203—3208.
5. Введение в теорию генетических текстов / В. В. Соловьев, А. Э. Кель, И. В. Рогозин, Н. А. Колчанов.— Новосибирск: Изд-во Новосиб. ун-та, 1987.— 92 с.
6. Вайнштейн Н. К. Симметрия кристаллов. Современная кристаллография.— М.: Наука, 1979.— Т. 1.— 383 с.
7. Заморзаев А. М., Галярский Э. И., Палистрант А. Ф. Цветная симметрия, ее обобщения и приложения.— Кишинев: Штиинца, 1978.— 277 с.
8. Pattern recognition of sequence similarities in globular proteins by Fourier analysis. A novel approach to molecular evolution / A. M. Liquori, A. Ripamonti, C. Sadun et al. // J. Mol. Evol.—1986.— 23, N 1.—P. 80—87.
9. Cosic I., Nestic D. Prediction of not spots in SV40 enhancer and relation with experimental data // Eur. J. Biochem.—1987.— 170, N 1/2.—P. 247—252.
10. Миронов А. А., Александров Н. Н. Быстрый метод поиска гомологий нуклеотидных последовательностей // Биофизика.— 1988.— 33, № 2.— С. 229—232.

Казан. гос. ун-т

Получено 10.05.90

УДК 577.112.5.0.87

**П. В. Костецкий, И. В. Артемьев, О. И. Пожилыцова, В. В. Ульяшин**

#### **ПАКЕТ ПРИКЛАДНЫХ ПРОГРАММ ДЛЯ АНАЛИЗА АМИНОКИСЛОТНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ НА ПЕРСОНАЛЬНОЙ МИКРО-ЭВМ «ИСКРА-226»**

*Приведено описание пакета программ для обработки аминокислотных последовательностей белков и пептидов. Входящие в пакет программы дают возможность: 1) ввода и редакции аминокислотных последовательностей; 2) распечатки выравненных аминокислотных последовательностей семейств гомологичных белков; 3) попарного сравнения гомологичных последовательностей; 4) расчета филогенетических деревьев гомологичных белковых последовательностей; 5) идентификации местоположения переменных и консервативных участков в аминокислотных последовательностях гомологичных белков; 6) поиска сходства двух сравниваемых белков с распечаткой точечной матрицы сравнения; 7) предсказания локализации В- и Т-клеточных антигенных детерминант; 8) перенесения аминокислотных последовательностей банка белковых последовательностей PIR, хранящихся на магнитных лентах, на гибкий магнитный диск персональной ЭВМ «Искра-226» и обратно; 9) идентификации фрагментов структурного гена с помощью набора пептидов; 10) трансляции структурных генов, содержащих интроны. Пакет программ написан на алгоритмическом языке Бейсик. С целью ускорения работы отдельных наиболее трудоемких блоков программ использован автокод микро-ЭВМ. Программы реализованы в диалоговом режиме.*

**Введение.** Расшифровка аминокислотных последовательностей белков и пептидов в настоящее время стала рутинной работой, и уже накоплено много данных о первичных структурах белков и пептидов. В организованном в США банке PIR (Protein Identification Resource) сейчас находится более 1,5 млн аминокислотных остатков (а. к. о) [1]. С накоплением знаний о первичной структуре белков и их функциональных особенностях закономерно встает вопрос о взаимосвязи биологической активности белков и их строения. Для исследования свойств белков и пептидов с использованием ЭВМ написан ряд пакетов про-

© П. В. КОСТЕЦКИЙ, И. В. АРТЕМЬЕВ, О. И. ПОЖИЛЬЦОВА, В. В. УЛЬЯШИН, 1990

грамм, среди которых наибольшей популярностью пользуются программы для персонального компьютера IBM PC [2, 3] и Apple II [4, 5]. Программы различных пакетов выполняют примерно схожие функции: расчет аминокислотного состава и молекулярной массы белка, предсказание локализации антигенных детерминант, построение гидропатичных профилей, построение диаграмм сходства двух аминокислотных последовательностей и пр.

Наш пакет, имеющий аналогичные и ряд дополнительных возможностей (таблица), предназначен для отечественной ЭВМ «Искра-226»,

*Функции набора программ для обработки аминокислотных последовательностей белков и пептидов*

*Names and assignment of programs for analysis of protein amino acid sequences*

Название программы	Выполняемые функции
INPUT	Ввод и редакция аминокислотных последовательностей
FAMILY	Распечатка группы аминокислотных последовательностей гомологичных белков
MATRIX	Анализ попарного сходства аминокислотных последовательностей
TREE	Построение филогенетических деревьев
CONST	Поиск варибельных и консервативных участков аминокислотных последовательностей гомологичных белков
DOTMAT	Поиск неявной гомологии белков и пептидов
PROTSCAN	Предсказание местоположения В-клеточных антигенных детерминант
HELAMPHI	Предсказание местоположения амфипатических спиралей—Т-клеточных антигенных детерминант
TRANSFER	Обмен информацией с банком аминокислотных последовательностей PIR, установленным на ЭВМ «ЕС-1055»
GENIDENT	Идентификация фрагментов структурного гена с помощью набора пептидов
TRANSL	Трансляция структурных генов, содержащих интроны

которая широко применяется в биохимических исследованиях [5—13]. Ряд программ нашего пакета носит оригинальный характер. Пять из них предназначены для сравнения свойств аминокислотных последовательностей гомологичных белков (FAMILY, MATRIX, DOTMAT, CONST, TREE). Программы PROTSCAN и HELAMPHI позволяют предсказывать местоположение В- и Т-клеточных антигенных детерминант соответственно. Программа GENIDENT используется в процессе установления строения структурных генов и предоставляет возможность автоматической идентификации фрагментов структурного гена с помощью имеющегося набора пептидов [7]. Программа TRANSL применяется для трансляции структурных генов, в последовательностях которых допустимо наличие нетранслируемых вставок — интронов.

**Материалы и методы.** Все программы пакета написаны на алгоритмическом языке Бейсик в приложении к отечественной персональной микро-ЭВМ «Искра-226». Характерное время исполнения наиболее употребляемых программ — от нескольких секунд до нескольких минут. Для ускорения работы отдельных программ пакета использован автокод микро-ЭВМ.

Для большинства программ пакета достаточно минимального комплекта стандартных периферийных устройств «Искры-226»: накопителя на гибких дисках НГМД «ЕС-5074» и печатающего устройства «ROBOTRON-1154». Программы попарного сравнения белковых последовательностей DOTMAT и PROTSCAN требуют незначительных изменений печатных плат принтера, обеспечивающих вывод графической информации. При работе программы TRANSFER для переноса аминокислотных последовательностей банка PIR [1] с ЭВМ «ЕС-1055» в комплектацию «Искры-226» необходим накопитель на магнитной ленте «СМ-5300.01», позволяющий производить запись с плотностью 800 байт/дюйм. Аминокислотные последовательности белков, приведенные в настоящей работе, были взяты из 19-й версии банка PIR.

**Результаты и обсуждение.** Программа ввода и редакции аминокислотных последовательностей INPUT работает в диалоговом режиме с использованием главного меню. Программа позволяет вводить аминокислотные последовательности как с гибкого магнитного диска, так и непосредственно с клавиатуры компьютера.

При вводе данных с клавиатуры предоставлена возможность применения как однобуквенного кода аминокислот, так и трехбуквенного представления аминокислотной последовательности, используя при этом специализированную накладку на панель клавиатуры. Введенные

	10	20	30	40	50	60
1	NLYQFKKMLGCTGYNRNSWWDFAFYGGYCGNGGSSGTFVDDLRCQGVHINQCYDEAEKLSRCWPFYFK					
2	.....S.....N.....G.....					
3	.....K...S...L...N.....I...N...G...G.....					
4	.....H...P...H...N.....K.....I...K.....G...I...					
5	.....E.....H...N.....I...G...G...I...					
6	.....H...S...P...H.....E...A.....G...L...G...L...F					
7	.....H...S...P...H.....K.....EK...G...M...G...T					

  

	70	80	90	100	110	120
1	TYSYECSSQCTI TCKNGHNCACAAAVCKCKDRLAAICFAGAPYNNNNYNIILKARCQ					
2	.....G...CA.....D...D...N.....E					
3	.....GDD...N...S.....Y.....D...N.....					
4	...T...SC.....D...G...K...S.....V...N...R...T...DK...FN...					
5	...T...DSC.....S...GAA...N...S.....V...N...R...TK...FN...					
6	L...K.....K...S...K...E...N...LV...N.....IDA...VN...E...					
7	L...K...K...K...S...G...K...G...N...LV...N.....B...IDA...NF...K...					

Рис. 1. Сравнение аминокислотных последовательностей фосфолипаз A2 рода *Naja*: *N. n. kaouthia* (1), *N. n. atra* (2), *N. melanoleuca I* (3), *N. melanoleuca III* (4), *N. m. mossambica* (5), *N. m. pallida* (6), *N. n. oxiana* (7). Для достижения максимума сходства в гомологичные последовательности введены делеции, обозначенные дефисом. Одинаковые с последовательностью (1) остатки в последовательностях (2)—(7) обозначены точками

Fig. 1. A comparison of snake phospholipase A2 amino acid sequences: *N. n. kaouthia* (1), *N. n. atra* (2), *N. melanoleuca I* (3), *N. melanoleuca III* (4), *N. m. mossambica* (5), *N. m. pallida* (6), *N. n. oxiana* (7). Gaps (—) have been inserted to achieve maximum homology. Dots represent residues identical to those of *N. n. kaouthia* sequence.

аминокислотные последовательности вместе с дополнительной информацией о ней могут быть сохранены на ГМД в виде файла данных со своим именем.

При чтении данных с ГМД возможна стыковка нескольких последовательностей в заданном порядке. Редакция аминокислотной последовательности производится стандартными клавишами редактирования клавиатуры компьютера и допускает вставку, удаление, замену любого аминокислотного остатка последовательности и введение пробелов для обозначения делеций.

Программным меню предоставляется возможность вывода на печать аминокислотной последовательности, а также вычисления молекулярной массы и аминокислотного состава рассматриваемого объекта.

Подпрограмма FAMILY является дополнением программы INPUT и позволяет распечатывать в виде двумерного символьного массива группу аминокислотных последовательностей гомологичных белков. Предварительно гомологичные последовательности выравнивают по длине с помощью введения делеций для достижения максимума сходства. Эта операция выполняется вручную с помощью редактирующей части программы INPUT. При распечатке выбранных последовательностей в однобуквенном виде позиции, где при сравнении с первой последовательностью остальные последовательности имеют одинаковые аминокислотные остатки, помечаются символом «.» (рис. 1).

Подпрограмма MATRIX дает возможность вывода матрицы попарного сходства гомологичных белковых последовательностей на экран или АЦПУ. Верхний угол матрицы показывает сумму совпадающих

аминокислотных остатков, а нижний угол матрицы — процент совпадений в сравниваемой паре последовательностей. При этом общая длина не считается совпадением и длина пары уменьшается на единицу. На главной диагонали матрицы даны длины каждой последовательности (рис. 2).

Программа CONST производит сравнение двух групп гомологичных выравненных аминокислотных последовательностей одного семейства. Каждую группу составляют последовательности, принадлежащие наиболее близким организмам. При «межгрупповом» сравнении сопо-

		1	2	3	4	5	6	7
1	<i>N.n.kaouthia</i>	119	110	103	94	92	89	95
2	<i>N.n.atra</i>	92	120	105	91	91	92	97
3	<i>N.melanoleuca I</i>	87	88	119	92	94	86	110
4	<i>N.melanoleuca III</i>	79	76	77	118	104	86	86
5	<i>N.m.mossambica</i>	77	76	79	88	119	85	88
6	<i>N.m.pallida</i>	75	77	72	73	71	118	79
7	<i>N.n.oxiana</i>	80	81	92	73	74	67	119

Рис. 2. Матрица попарного сходства аминокислотных последовательностей семейства фосфолипаз A2. Элемент верхней части матрицы — число совпадающих а. к. о. между гомологичными белками; элемент нижнего угла матрицы — процент совпадений в сравниваемой паре последовательностей

Fig. 2. Homology matrix from pairwise comparison of 7 amino acid sequences. Upper half of matrix represent the number of identical amino acid residues; lower half — the percentage of identical amino acid residues.

ставляются только последовательности разных групп. Такой способ сравнения позволяет достаточно надежно идентифицировать консервативные и варьируемые участки в гомологичных белках, так как результат сравнения мало зависит от количественного состава групп [8]. Результатом сравнения является символично-цифровая строка (рис. 3), в каждой позиции которой представлен процент аминокислотных замен при всех попарных сравнениях аминокислотных остатков каждого из столбцов.

Разбивка исходного семейства выравненных гомологичных последовательностей на две группы может быть осуществлена с помощью программы TREE. На основании матрицы попарного сходства аминокислотных последовательностей (рис. 2) программа формирует дихотомичное филогенетическое дерево, каждая вершина которого делит семейство нисходящих ветвей на две группы [14].

Программа DOTMAT предназначена для сравнения аминокислотных последовательностей двух белковых структур. Результатом работы программы является распечатка точечной матрицы гомологии выбранных объектов [15]. Исследователь задает длину участков сравнения и минимальное количество совпадающих аминокислотных остатков в нем. Каждая точка распечатываемой матрицы соответствует гомологичным фрагментам сравниваемых последовательностей, а ее координаты — соответственно положениям гомологичных фрагментов в анализируемых последовательностях.

Приводимая программа может сравнивать объекты, длина меньшего из которых не превосходит 780 а. к. о. (ограничения ширины бумаги печатающего устройства). Для ускорения работы программы использован автокод «Искры-226», что позволило значительно увеличить быстродействие программы. Время получения точечной матрицы гомологии практически не зависит от выбираемой длины сравниваемого сегмента и фактически сводится к времени ее печати. Полученная мат-

рица дает возможность проведения качественных оценок степени сходства исследуемых объектов.

Программа PROTSCAN предназначена для предсказания местоположения В-клеточных антигенных детерминант белка, исходя из шкалы антигенности аминокислотных остатков Веллинга [16]. Результа-

	10	20	30	40	50	60
1	NLYQFKNMTQCTVPNRSWDFADYGCYCGRGSGTVPDDLDRCQVHDNDCYDEAEKISRCPWPK					
2	.....S.....G.....N.....G.....					
3	.....K...S...L...N.....I...N...G...G.....					
4	.....H.....P...H...N.....K.....I...K.....G...R...I...					
5	.....H.....H...N.....I.....G.....G...I...I...					
6	.....H...S.P.H.....K..A.....C...L-G...LT					
7	.....H...S.P.H.....K.....EK.G.M-G...T					
8	..I..G...SAMTCK-.SLAY.S.....W..K.Q.R..T...F...C...GK.D.C.PRM---I					
9	D.T..G...NKMCG--VF.YIY.....W..K.K.I.AT...F...C...GKMGTYDTK...-T					
10	D.T..G...NKMCG--VF.YIY.....W..Q.K.R.AT...F...C...GKMGTYDTK...-T					
	70*00*000*****5**8*7*000000*0080*2*07*0000*500*0089797*8**4***0					
	70	80	90	100	110	120
1	TYSYECSSGCTLTCKNGNNAACAAAVCBCDRLAAICFAGAPYNNN-NYNIDLKARCQ					
2	.....G...CA.....D.-D...N.....E					
3	.....GDD.N...S.....Y.....D...N.....					
4	..T..SC.....D.G-K...S.....V..N...R.T..DK-....FN....					
5	..T..DSC...SCGAA.N...S.....V..N...R...IDK-....FN....					
6	L.K...K...SG...K.E...N..LV..N...IDA...VN..E...					
7	L.K.K...K...SG...SK.G...N..LV..N...R.IDA-...NF.K...					
8	L...KPHN.NIV.-GDK...KKK...E...V.....ASKHSY.K.LWRYPSSK.TTGTAEK					
9	S.N..IQN.GID.--DEDPQKKEL.E...V.....NNRNTY.S..FGHSSSK.TGTEQC--					
10	S.N..FQD.DIT.-GDKDPQKKEL.E...V.....NSRNTY.SK.FGYSSSK.TETEQC--					
	*0805***0***2*89*8*8***70*0635008000*****8*54*****0*					

Рис. 3. Сравнение аминокислотных последовательностей фосфолипаз A2 рода *Naja* и рода *Bitis*: *N. n. kaouthia* (1), *N. n. atra* (2), *N. melanoleuca* I (3), *N. melanoleuca* III (4), *N. m. mossambica* (5), *N. m. pallida* (6), *N. n. oxiana* (7), *B. gabonica* (8), *B. nasicornis* (9), *B. caudalis* (10). Для достижения максимума сходства в гомологичные последовательности введены дефисы, обозначенные дефисом. Одинаковые с последовательностью (1) остатки в последовательностях (2) — (10) обозначены точками. Каждый элемент нижней строки — процент аминокислотных замен (*d*) в каждой позиции при 21 возможном межгрупповом парном сравнении фосфолипаз родов *Naja* и *Bitis*. Использована символично-цифровая шкала: «0» —  $d=0\%$ ; «1» —  $0 < d < 10\%$ ; «2» —  $10 < d < 20\%$ ; ...«\*» —  $90 < d < 100\%$ .

Fig. 3. A comparison of snake phospholipase A2 amino acid sequences of *Naja* and *Bitis* genera: *N. n. kaouthia* (1), *N. n. atra* (2), *N. melanoleuca* I (3), *N. melanoleuca* III (4), *N. m. mossambica* (5), *N. m. pallida* (6), *N. n. oxiana* (7), *B. gabonica* (8), *B. nasicornis* (9), *B. caudalis* (10). Gaps (—) have been inserted to achieve maximum homology. Dots represent residues identical to those of *N. n. kaouthia* sequence. Each element of lower row is the percentage of amino acid residues replacement in given column by 21 possible intergroup *Naja/Bitis* comparison. Alphanumerical symbols are used «0» —  $d=0\%$ ; «1» —  $0 < d < 10\%$ ; «2» —  $10 < d < 20\%$ ; ...«\*» —  $90 < d < 100\%$ .

том работы программы является график, каждая точка которого отвечает среднему значению антигенности сегмента фиксированной длины, содержащего не более 30 а. к. о. Значение антигенности каждой точки отвечает начальному номеру сегмента. При необходимости возможно распечатывать числовые значения антигенного профиля или его любого фрагмента.

Программа PROTSCAN в одном из режимов работы строит график амфипатичности в соответствии с методом Кайта и Дулитла [17], или график гидрофильности с использованием шкалы Хоппа — Вудса [18]. Максимальная длина белковой молекулы при работе с программой PROTSCAN — не более 1 000 а. к. о.

Программа HELAMPNI предсказывает местоположение амфипатических спиралей как возможных Т-клеточных антигенных детерминант [19]. Численные расчеты для белка длиной 30 а. к. о. и распечатка результатов требуют не более 1 ч (рис. 4). Помимо координат амфипатических блоков и соответствующих им значений амфипатичности на печать выводятся значения углов, при которых достигается наи-

лучшая периодичность гидрофобных свойств каждого сегмента из 11 последовательных аминокислотных остатков. Программа HELAMPRI может обрабатывать белки длиной до 1 000 а. к. о.

Программа TRANSFER предназначена для перенесения аминокислотных последовательностей из белкового банка PIR (США) с ЭВМ

HUMAN IL 3					
10	20	30	40	50	60
MSRPLRVLLLELQELVRFGLQAFMTQTTSLKTSWVNCNMIDEITTHLKQPPFLLDFNNLN					
CFEDQDILMENNLRFPNLEAFNRAVKSLQNASATESLLKNLLPQPLATAAFTTRPRLHKDQ					
IWNQEPNRKILSTFYLKTLQNAQACQPTISLALF					
Координаты амфипатических блоков		Углы		Амфипатический показатель (AS)	
30	51	85 - 130		35,0	
72	107	85 - 135		69,9	
127	143	95 - 135		12,2	
Длина белка		151 а.к.о.		Средн AS	
Число предсказанных блоков		- 3		- 8,0	
Длина сегмента сканирования		- 11			

Рис. 4. Предсказание возможных Т-клеточных антигенных детерминант на аминокислотной последовательности белка интерлейкина-3 человека

Fig. 4. T-cell antigenic sites prediction for human interleukin-3 amino acid sequence.

«ЕС-1055» на гибкий магнитный диск для дальнейшего их анализа на персональной ЭВМ «Искра-226». Предусмотрена также возможность передачи данных с «Искры-226» на «ЕС-1055» для решения задач, тре-

120	130	140	150	160
!	!	!	!	!
CTC AGC TGT AGC AGA TAC TCT GTC GGG CTC TTG GAC ATG ACC....				
Leu Ser Cys Ser Arg Tyr Ser Val Gly Leu Leu Asp Met Thr....				
*** **				
Tyr-Ser-Ala-Gly-Leu-Leu				

Рис. 5. Оpozнание пептидом Tyr-Ser-Ala-Gly-Leu-Leu фрагмента ДНК фосфодиэстеразы циклического ГМФ. Звездочками помечены совпадающие позиции пептида с транскрипованной аминокислотной последовательностью части структурного гена

Fig. 5. The fragment of gene coding for phosphodiesterase cyclic GMP (bovine) identified by pentapeptide Tyr-Ser-Ala-Gly-Leu-Leu. The equivalent positions of the peptides and translated amino acid sequence are marked by asterisks.

бующих значительных вычислений. Передача данных осуществляется с использованием магнитной ленты в качестве носителя информации и требует наличия в составе периферийных устройств к ЭВМ «Искра-226» накопителя на магнитной ленте «Искра-005-61».

Для автоматизации трудоемкого процесса идентификации фрагментов структурного гена с помощью набора пептидов предлагается программа GENIDENT [20]. В качестве иллюстрации на рис. 5 фрагмент гена  $\alpha$ -субъединицы фосфодиэстеразы циклического ГМФ из сетчатки быка опознан гексапептидом Tyr-Ser-Ala-Gly-Leu-Leu. При опознании фрагментов гена программа использует критерии, обоснованные ме-

годами математической статистики [7]. Исследователь может оценить достоверность правильного опознания фрагментов структурного гена при возможности ошибок в первичных структурах пептидов и фрагментов ДНК. Программа включает в себя главное и вспомогательные меню. К дополнительным возможностям программы относится возможность поиска пептидов на известной аминокислотной структуре гомологичного объекта. Программа была использована при расшифровке структуры  $\alpha$ -субъединицы фосфодинэстеразы циклического ГМФ [20] и основного токсического компонента яда паука *Latrodectus mactants tredecimguttatus* — фермента латротоксина.

Программа TRANSL транслирует кодирующие участки структурного гена. По задаваемой информации об интронных участках нуклеотидной последовательности она распечатывает транслированную аминокислотную последовательность с учетом не кодирующих участков [21]. Программа требует указания границ соответствующих транслируемых подфрагментов ДНК.

В заключение следует отметить, что описанный в статье пакет программ для анализа аминокислотных последовательностей широко использовался в нашем институте [7, 8, 20]. Для исследователей, работающих на персональных компьютерах «Искра-226», пакет программ можно получить от авторов данной работы в виде набора объективных модулей, записанных на гибкий магнитный диск размером 203 мм.

Авторы признательны сотрудникам института Г. С. Монастырской, И. Е. Дулубовой за участие в постановке задач и ценные рекомендации, высказанные в процессе реализации отдельных программ пакета. Авторы благодарят также Н. А. Фонинову, Р. Р. Владимирову за помощь в составлении отдельных программ, В. Е. Костюшко — за изготовление накладной клавиатуры для ввода белковых последовательностей в трехбуквенном коде.

#### A SOFTWARE PACKAGE FOR ANALYSING AMINO ACID SEQUENCES BY PERSONAL MICROCOMPUTER «ISKRA-226»

*P. V. Kostetsky, I. V. Artemyev, O. I. Pozhiltsova, V. V. Ulyashin*

M. M. Shemyakin Institute of Bioorganic Chemistry,  
Academy of Sciences of the USSR, Moscow

#### Summary

Presented software package allows: 1) input and edition of amino acid sequences; 2) listing of aligned amino acid sequences of homologous protein; 3) pairwise comparison of homologous sequences; 4) calculation of phylogenetic trees of homologous protein sequences; 5) identification of variable and constant regions of amino acid sequences; 6) search for similarity between two amino acid sequences and list out comparison dot matrix; 7) prediction of B- and T-cell antigenic determinants; 8) trans forming Protein Identification Resource (PIR) from magnetic tape media to «Iskra-226» floppy disk media and vice versa; 9) identification of structure gene fragments using a set of peptides; 10) translation of structure genes containing introns.

Programs are written in BASIC for a personal computer «Iskra-226» (USSR). To accelerate some operations, computer code commands are used. Software package is realized in interactive mode and intended for specialists on peptide and protein chemistry.

#### СПИСОК ЛИТЕРАТУРЫ

1. *The protein identification resource (PIR)* / K. E. Sidman, D. G. George, W. C. Barker, L. T. Hunt // *Nucl. Acids Res.*— 1988.— 16, N 5.— P. 1869—1871.
2. *Cannon G. C.* Sequence analysis on microcomputer // *Science.*— 1987.— 238, N 1.— P. 97—103.
3. *Heijne G.* Getting sense out of sequence data // *Nature.*— 1988.— 333, N 6174.— P. 605—607.

4. *ANTHEPROT*: a package for protein sequence analysis using a microcomputer / G. Deleage, F. F. Clerc, B. Roux, D. C. Gautheron // *Computer Appl. Biosci.*—1988.—4, N 3.—P. 351—356.
5. *Bellon B.* Apple Macintosh programm for nucleic and protein sequence analysis // *Nucl. Acids Res.*—1988.—16, N 5.—P. 1837—1846.
6. Костецкий П. В., Доброва И. Е. Реконструкция длинных полинуклеотидных последовательностей из фрагментов с использованием персональной ЭВМ «Искра-226» // *Биоорг. химия.*—1988.—14, № 4.—С. 515—521.
7. Костецкий П. В., Артемьев И. В. Автоматическая идентификация фрагментов структурного гена с помощью набора пептидов // Там же.—1990.—16, № 2.—С. 202—210.
8. Костецкий П. В., Владимирова Р. Р. Графическая идентификация консервативных и вариабельных участков в аминокислотных последовательностях гомологичных белков // Там же.—1989.—15, № 11.—С. 1573—1576.
9. Головинов А. П., Суханов С. В., Барсуков Л. И. Влияние  $Mn^{2+}$  и  $Cd^{2+}$  на проводимость грамицидиновых каналов // *Биол. мембраны.*—1989.—7, № 2.—С. 149—153.
10. Черепинов Д. А., Рапапович И. И. Моделирование кинетики действия олигомерных ферментов на примере фосфофруктокиназы // *Молекуляр. биология.*—1987.—21, № 3.—С. 820—830.
11. *Статистические закономерности в первичных структурах функциональных областей генома Escherichia coli. I.* Частотные характеристики / М. Ю. Бородовский, Ю. А. Сприжичский, Е. И. Голованов, А. А. Александров // Там же.—1986.—20, № 4.—С. 1014—1023.
12. *Статистические закономерности в первичных структурах функциональных областей генома Escherichia coli. II.* Неоднородные марковские модели / М. Ю. Бородовский, Ю. А. Сприжичский, Е. И. Голованов, А. А. Александров // Там же.—С. 1024—1033.
13. *Статистические закономерности в первичных структурах функциональных областей генома Escherichia coli. III.* Компьютерное распознавание кодирующих областей // Там же.—№ 5.—С. 1390—1398.
14. *Fitch W. M., Margoliagh E.* Construction of phylogenetic trees // *Science.*—1967.—155, N 2.—P. 279—284.
15. *Transforming growth factor  $\beta 2$ : cDNA cloning and sequence analysis* / L. Madisen, N. Webb, T. M. Rose et al. // *DNA.*—1988.—7, N 1.—P. 1—8.
16. *Prediction of sequential antigenic regions in protein* / G. W. Welling, W. J. Weijer, R. Zee, S. Welling-Wester // *FEBS Lett.*—1985.—188, N 2.—P. 215—218.
17. *Kyte J., Doolittle R. F.* A simple method for displaying the hydropathic character of a protein // *J. Mol. Biol.*—1982.—157, N 1.—P. 105—132.
18. *Hopp T. P., Woods K. R.* Prediction of protein antigenic determinants from amino acid sequences // *Proc. Nat. Acad. Sci. USA.*—1981.—78, N 6.—P. 3824—3828.
19. *Prediction of immunodominant helper T cell antigenic sites from the primary sequence* / H. Margalit, J. L. Spouge, J. L. Cornette et al. // *J. Immunol.*—1987.—138, N 7.—P. 2213—2229.
20. Костецкий П. В., Артемьев И. В. Автоматизация идентификации фрагментов структурного гена с помощью набора пептидов // VII Всесоюз. симпоз. по химии белков и пептидов.—Таллин, 1987.—С. 27.
21. *Family of human  $Na^+$ ,  $K^+$ -ATPase genes* / Yu. A. Ovchinnikov, G. S. Monastyrskaya, N. E. Broude et al. // *FEBS Lett.*—1988.—233, N 1.—P. 87—94.

Ин-т биоорг. химии им. М. М. Шемякина АН СССР, Москва

Получено 10.05.90