БИООРГАНИЧЕСКАЯ ХИМИЯ
том 16 • № 2 • 1990

УДК 577.214.662

П. В. Костецкий, И. В. Артемьев

АВТОМАТИЧЕСКАЯ ИДЕНТИФИКАЦИЯ ФРАГМЕНТОВ СТРУКТУРНОГО ГЕНА С ПОМОЩЬЮ НАБОРА ПЕПТИДОВ

Институт биоорганической химии им. М. М. Шемякина АН СССР, Москва

Для автоматизации трудоемкой работы по идентификации фрагментов структурного гена с помощью набора пептидов нами разработана программа для ЭВМ «Искра-226». Программа написана на алгоритмическом языке Бейсик. С целью ускорения работы отдельных блоков программы использован автокод микроЭВМ. Для опознания 50 фрагментов реконструируемой ДНК с помощью 10 пептидов требуется не более 1 ч машинного времени. Результат работы программы – распечатываемые фрагменты структурного гена совместно с опознавшими их пептидами.

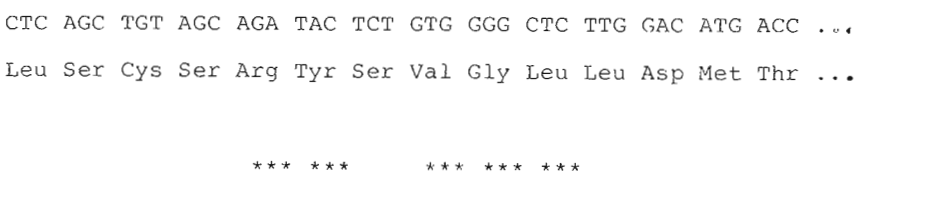
Дано математическое обоснование предложенного метода. Показано, что число ошибочных опознаний фрагментов структурного гена может быть предсказано на основании распределения Пуассона и сведено к минимуму при правильном выборе критериев. Рассматриваемый подход позволяет оценивать достоверность правильного опознания при наличии ошибок в нуклеотидных последовательностях фрагментов гена или аминокислотных последовательностях пептидов.

В настоящее время все чаще белковая последовательность становится известной одновременно с прочтением нуклеотидной последовательности соответствующего структурного гена. При этом на начальной стадии исследования из гидролизата микроколичеств изучаемого белка выделяют ряд пептидов, аминокислотную последовательность которых необходимо знать для синтеза наиболее удачных зондов. В результате последующего скрининга банка генов получают ряд клонов кДНК, причем некоторые клоны по причине недостаточной специфичности олигонуклеотидного зонда являются ложными. В ходе дальнейших структурных исследований образуется набор большого количества разнообразных фрагментов ДНК. Весьма важны среди этих фрагментов те, в которых содержится информация об аминокислотных последовательностях пептидов, полученных при деградации исходного белка.

Идентификация фрагментов нужного структурного гена с помощью пептидов облегчает нахождение перекрытий полинуклеотидных цепей, необходимых для реконструкции ДНК, кодирующей исследуемый белок. Естественно, что идентификация возможна для тех фрагментов, в которых транслируемая аминокислотная последовательность содержит хотя бы один пептид из выделенного набора. Такая комбинированная стратегия определения первичной структуры белка и нуклеотидной последовательности соответствующего структурного гена неоднократно применялась ранее [1–6].

Очевидно, что опознание фрагментов структурного гена с помощью набора пептидов без применения ЭВМ – длительный и трудоемкий процесс. Настоящая статья посвящена автоматизации процесса идентификации с помощью персональной ЭВМ «Искра-226», которая широко используется в биохимических лабораториях [7–13, 15]. При этом большое внимание уделено установлению критериев правильного применения разработанной программы автоматического поиска фрагментов структурного гена.

Фрагмент ДНК является частью структурного гена, если в транслируемой аминокислотной последовательности присутствует хотя бы один пептид из экспериментально определенного набора. Поиск пептида на транс-



Опознание пептидом Tyr-Ser-Ala-Gly-Leu-Leu фрагмента ДНК фосфодиэстеразы циклического GMP. Звездочками помечены совпадающие позиции пептида и транслированной аминокислотной последовательности части структурного гена

лированной аминокислотной последовательности осуществляется по всем шести возможным рамкам трансляции прямой и комплементарной цепи. Фрагмент ДНК фосфодиэстеразы циклического GMP из сетчатки быка (ФДЭ) [5] опознан гексапептидом как часть структурного гена, причем в транслированной аминокислотной цепи имеется участок, который совпадает по 5 аминокислотным остаткам с первичной структурой пептида (рисунок). Наличие одного несовпадающего остатка вызвано ошибкой в определении нуклеотидной последовательности ДНК. По причине распространенности таких ошибок идентификация многих фрагментов структурного гена возможна только при допущении несовпадения одного или более аминокислотных остатков. Для набора из нескольких пептидов и малого числа фрагментов ДНК совпадение 5 из 6 аминокислотных остатков свидетельствует о высоком уровне достоверности. Однако при большом числе пептидов и фрагментов ДНК требуется математическое моделирование и численная оценка достоверности правильного опознания структурного гена.

Действительно, гексапептид на рисунке может узнавать и случайные участки ДНК, если число совпадающих остатков понизить до 4. Очевидно, что математическое моделирование и должно дать возможность оценивать вероятность правильной идентификации в случаях наличия ошибок в нуклеотидных последовательностях фрагментов гена.

Для обоснования критериев идентификации фрагментов структурного гена — минимально необходимого числа совпадающих позиций пептида с транслированным участком гена — нами использовались данные о генах, кодирующих следующие белки: α- и γ-субъединицы фосфодиэстеразы циклического GMP из сетчатки быка [5, 14], АТФ-зависимой La-протеиназы *E. coli* [15], α-субъединицы G-белка быка [16], α-субъединицы Na, K-АТРазы из почек свиньи [4]. Нуклеотидные последовательности указанных генов, включая фланкирующие участки, взяты из оригинальных работ и 57-й версии GENBANK'a [17].

Наиболее подробные статистические исследования выполнены на примере α-субъединицы фосфодиэстеразы циклического GMP [5], называемого далее для краткости ФДЭ. При этом в качестве опознающих пептидов брали произвольные участки ФДЭ, а фрагментами ДНК служили случайные нуклеотидные последовательности с соотношением оснований A : C : : G : T = 28 : 24 : 26 : 22, отвечающим составу прямой цепи гена ФДЭ быка. Нуклеотидный состав обратной цепи роли не играет, так как фрагменты обратной цепи в процессе компьютерной обработки конвертируются в прямую цепь.

Для проводимых численных экспериментов использовали программный датчик псевдослучайных чисел. С его помощью случайным образом из аминокислотной последовательности белка ФДЭ брали серии по 100 пептидов фиксированной длины. С использованием датчика получали также искусственные фрагменты ДНК постоянной длины и состава, тождественного составу гена, кодирующего этот белок. Алгоритм получения таких ДНК основан на произвольной случайной перестановке 250 нуклеотидных

оснований, взятых в соответствующем ФДЭ соотношении. Серии пептидов постоянной длины, случайно взятые из аминокислотной последовательности белка, использовали в опознании 50 полученных случайных нуклеотидных последовательностей, которые заведомо не являются какими-либо структурными генами.

В качестве примера рассмотрим численный эксперимент по опознанию случайных ДНК серией пептидов длиной 9 аминокислотных остатков. В результате моделирования оказалось, что высокий уровень сходства — 7 и более совпадений из 9 остатков — гарантирует от «случайных» опознаний. При этом не наблюдалось ни одного опознания таких искусственных ДНК. При числе совпадений, равном 6 ($n=6$), имелось четыре фрагмента, принятых за структурный ген ($M=4$). При уменьшении уровня совпадений до 5 ($n=5$) число случайно опознанных фрагментов резко возрастает ($M=86$).

В случаях, когда M мало, как, например, для пептидов длиной 9 аминокислотных остатков и числе совпадающих позиций $n\geqslant 6$, можно ожидать, что количество ошибочных опознаний будет случайной величиной, распределение которой будет подчиняться закону Пуассона. Очевидно [18], что для случайной величины с пуассоновским распределением вероятность «случайного» опознания одним пептидом K фрагментов ДНК выражается формулой

$$P_K = (M_0)^K \cdot \exp(-M_0)/K!, \quad K = 0, 1, 2, 3, \ldots \tag{1}$$

где M_0 — параметр распределения.

Аппроксимируем приведенные в табл. 1 данные численного эксперимента пуассоновским распределением с параметром M_0, равным математическому ожиданию числа «случайных» опознаний. Среднее число случайных находок, вычисляемое по формуле $\bar{M} = \left(\sum_{i=1}^{100} M_i\right)\Big/100$, может быть принято за параметр пуассоновского распределения [19]: $\bar{M} \to M_0$.

Для критериев поиска $n=5$ и $n=6$ из 9 позиций (табл. 1), исходя из среднего числа «случайных» находок ($\bar{M}$), легко вычислить по формуле (1) теоретические частоты K опознаний фрагментов ДНК, не являющихся структурными генами. Для критерия $n=6$ из 9 ($\bar{M}=0{,}04$) аналитически полученные из (1) данные хорошо совпадают с экспериментальной частотой идентификации искусственных ДНК (табл. 1). В то же время для критерия $n=5$ из 9 ожидаемые и наблюдаемые экспериментально частоты идентификаций различаются существенно, и это свидетельствует об отклонении от закона распределения Пуассона для $n\leqslant 5$. О качестве согласования эмпирического и теоретического распределения можно судить по величине $\chi^2 = \sum (B_\text{р} - B_\text{э})^2/B_\text{р}$. В этой формуле $B_\text{р}=P_K \cdot 100$ и $B_\text{э}$ — соответственно ожидаемые и наблюдаемые числа K опознаний искусственных ДНК, не являющихся какими-либо структурными генами. Если величина χ^2 не превышает заданного значения χ_0^2, то с некоторой доверительной вероятностью справедливо приближение экспериментального распределения теоретическим. Численное значение берется из таблиц вероятностного распределения Пирсона при выбранном уровне доверительной вероятности [19].

Величина критерия χ^2 свидетельствует о том, что число искусственных фрагментов ДНК (M), случайно опознаваемых с помощью пептидов длиной 9 аминокислотных остатков как часть структурного гена для критерия поиска $n=6$ (табл. 1), имеет с вероятностью не менее 0,95 пуассоновский закон распределения. Это дает возможность оценивать вероятность (P_0) правильной идентификации реальным нонапептидом одного или более фрагментов структурного гена из 50 фрагментов ДНК. Очевидно, что $P_0 = (\bar{M})^0 \cdot \exp(-\bar{M})/0! = \exp(-0{,}04) = 0{,}961$.

При возрастании числа фрагментов, ошибочно принятых за структурный ген, случайная величина M не всегда подчиняется с выбранной доверительной вероятностью закону Пуассона и, как следствие этого, формула

(1) становится непригодной для оценки достоверности идентификации. Поэтому важен подбор таких критериев опознания, для которых поведение случайной величины M подчиняется пуассоновскому закону распределения.

Проведенные численные эксперименты для серий пептидов фиксированной длины L ($4 \leqslant L \leqslant 12$ аминокислотных остатков) позволили установить критерии опознания, при которых вероятность верной идентификации фрагмента структурного гена превышает 0,95 (табл. 2). Ослабление этих критериев (уменьшение числа совпадающих остатков пептида с транслированной аминокислотной последовательностью) приводит к резкому увеличению числа фрагментов ДНК (M), ошибочно идентифицированных как фрагмент структурного гена. При длине пептида менее 4 остатков число случайных опознаний достаточно велико и сильно зависит от аминокислотного состава пептида. Тем не менее пептиды такой длины могут быть использованы для опознания фрагментов гена. В этом случае требуется привлечение дополнительных экспериментальных данных, например учета специфичности при ферментативном расщеплении (см. ниже).

Естественно, что в работе по определению строения структурных генов могут встречаться пептиды длиной более 12 остатков. В таких случаях для идентификации фрагментов ДНК можно брать участки пептидов меньшей длины. Это вполне оправданно, так как одной из распространенных ошибок при определении нуклеотидных последовательностей является пропуск или вставка основания в какой-либо позиции, что приводит к резкому искажению аминокислотной последовательности. Легко показать, что влияние такого искажения будет в значительной степени устранено при использовании двух концевых пептидов половинной длины. Действительно, по крайней мере один из укороченных пептидов окажется пригодным для правильной идентификации фрагментов ДНК.

При использовании больших наборов пептидов фиксированной длины число ошибочных идентификаций фрагментов ДНК, полученных с использованием приведенных в табл. 2 критериев и подчиняющихся закону распределения Пуассона, в реальной работе может оказаться недостаточно малым. В связи с этим необходимо уметь оценивать общее число «случайных» находок, встречающихся при опознании всех фрагментов ДНК с помощью полного набора пептидов.

Поскольку для пептидов фиксированной длины число ошибочно принятых за структурный ген фрагментов ДНК — независимая случайная величина, подчиняющаяся распределению Пуассона, приведенные в табл. 2 статистические данные позволяют предсказывать математическое ожидание ошибочных прилипаний (E) при опознании фрагментов структурного гена полным набором пептидов по формуле

$$E = N \sum_{L} R_L m_L, \qquad (2)$$

где N — число фрагментов ДНК; R_L — число пептидов фиксированной длины L; m_L — коэффициент, вычисляемый из табл. 2 по формуле $m_L = M_L/50$; M_L — среднее число находок одного пептида на 50 случайных ДНК; L изменяется от минимальной до максимальной длины встречающихся пептидов.

С помощью соотношения (1), полагая $M_0 = E$, можно легко показать, что если $E \leqslant 0{,}05$, то вероятность хотя бы одного случайного опознания для набора пептидов не превышает $1 - P_0 = 0{,}05$. В случае большой группы пептидов наличие одного случайного опознания вполне допустимо. В таком случае, если значение $E \leqslant 0{,}37$, вероятность более одного случайного опознания $P = 1 - (P_0 + P_1)$ не превосходит 0,05. Разумеется, ужесточая критерий сходства, можно потребовать для любой группы пептидов практически безошибочной идентификации фрагментов структурного гена. В табл. 3 приведены размеры групп пептидов равной длины, для которых с вероятностью не менее 0,95 можно предсказать отсутствие или наличие не более одной случайной идентификации какого-либо фрагмента ДНК. Подобные оценки легко выполнимы и для групп пептидов различной дли-

ны. Однако, как видно из табл. 3, в любом наборе пептидов не должно быть более 1—2 тетрапептидов и для коротких пептидов (до гексапептидов) желательно полное совпадение с транслируемым участком ДНК.

Для проверки соотношения (2) нами был проведен численный эксперимент, в котором 10 пептидов различной длины (1 пента-, 2 гекса-, 3 гепта-, 2 окта- и по одному нона- и декапептиду) участвовали в опознании 30 серий по 50 случайных ДНК длиной 250 нуклеиновых оснований, заведомо не являющихся частями структурного гена. Как и следовало ожидать, экспериментальное число случайно «опознанных» искусственных ДНК и предварительно вычисленное по формуле (2) оказались достаточно близкими по величине: 4 и 5,65 соответственно.

Следует отметить, что наблюдаемое среднее число случайных опознаний для группы пептидов оказалось несколько ниже ожидаемого значения. Это естественно, так как в формуле (2) подставляются из табл. 2 значения частот случайных опознаний одиночных пептидов, округленные в сторону увеличения. Данные табл. 3 позволяют также предсказывать результаты опознания фрагментов ДНК и для больших наборов из 100 и более пептидов, как это имело место при определении строения гена α-субъединицы ФДЭ [5], кодирующего белок длиной 886 аминокислотных остатков.

При получении и обосновании критериев идентификации табл. 3 использовали соотношение оснований прямой цепи гена ФДЭ из сетчатки быка A : C : G : T = 28 : 24 : 26 : 22. Нуклеотидный состав четырех остальных объектов заметно отличается от модельного. Например, для гена α-субъединицы G белка быка длиной 3099 пар оснований нуклеотидный состав A : C : G : T=30 : 16 : 21 : 33. Нами проведены статистические расчеты по опознанию тремя сериями из 50 пептидов α-субъединицы G белка длиной 4, 7 и 11 остатков случайных фрагментов ДНК длиной 250 п. о. и нуклеотидным составом, тождественным составу природного гена. Полученные результаты сопоставляли с данными численного эксперимента на примере α-субъединицы ФДЭ (табл. 2). Оказалось, что вероятность случайного опознания фрагментов ДНК практически постоянна для экспериментов на обоих объектах. Таким образом, основываясь на проведенных исследованиях, можно предположить, что формулы (1) и (2) применимы для оценки вероятности случайных опознаний и при соотношениях оснований A : C : G : T, заметно отличающихся от 25 : 25 : 25 : 25.

В экспериментальной работе исследователи для увеличения достоверности опознания фрагментов ДНК учитывают также специфичность гидролиза белка при получении набора пептидов. Например, в случае трипсинового гидролиза С-концевым остатком опознанного участка транслированного полипептида могут быть только остатки Arg или Lys.

С учетом дополнительного критерия специфичности гидролиза получаем оценку числа ожидаемых случайных находок — E:

$$E = gN\Sigma R_L m_L, \tag{3}$$

где g — коэффициент учета специфичности гидролиза, равный $T/64$ (T — суммарное число триплетов, кодирующих специфичные при данном типе гидролиза аминокислоты). Остальные величины ранее определены в соотношении (2).

Так, например, при трипсиновом гидролизе белка образуются пептиды, имеющие на С-конце остатки Arg и Lys, которые кодируются соответственно 6 и 2 триплетами. Тогда T=8 и g=8/64=0,125.

Естественно, что при учете специфичности гидролиза критерии надежной идентификации фрагментов с помощью пептидов могут быть смягчены по сравнению с данными табл. 2 и 3. Это особенно важно при использовании коротких пептидов из 4—5 остатков, число которых в применяемых наборах может быть увеличено. В случае более длинных пептидов можно понизить на единицу минимальное число совпадающих позиций, необходимое для опознания фрагмента структурного гена.

Для проверки соотношения (3) нами был проведен численный эксперимент, в котором 10 пептидов трипсинового гидролиза α-субъединицы

Таблица 1

Проверка на соответствие распределению Пуассона частот опознаний 50 искусственных ДНК, не являющихся структурными генами, с помощью 100 нонапептидов

	Число совпадений							
	$n=6$			$n=5$				
K	0	1	$\geqslant 2$	0	1	2	3	$\geqslant 4$
B_p	96,079	3,843	0,078	43,316	36,392	15,649	4,486	1,157
$B_э$	96	4	0	53	23	14	6	4
	$\overline{M}=0{,}040$; $\chi^2=0{,}084$ *			$\overline{M}=0{,}860$; $\chi^2=15{,}296$ **				

* С вероятностью более 0,95 случайная величина M имеет закон распределения Пуассона.
** Поведение случайной величины M не подчиняется закону Пуассона.

Таблица 2

Данные численного эксперимента по опознанию 50 случайных фрагментов ДНК для серий из 100 пептидов длиной от 4 до 12 остатков

Длина пептидов	Число совпадающих позиций	Число находок на 50 ДНК	Среднее число случайных находок на 50 фрагментах ДНК: $\overline{M}=M/100$	Вероятность случайного опознания $1-P_0$
4	4	19	0,19	0,173
5	5	1	0,01	0,010
6	5	5	0,05	0,049
7	5	17	0,17	0,156
	6	1	0,005 (*)	0,005
8	6	1	0,01	0,010
9	6	4	0,04	0,039
10	6	5	0,05	0,049
11	6	10	0,10	0,095
	7	1	0,003 (**)	0,003
12	7	1	0,01	0,010

Примечание. Для получения среднего числа случайных находок было взято 100 (*) и 150 (**) искусственных фрагментов ДНК.

ФДЭ циклического GMP постоянной длиной L (L=6, 9, 12 остатков) участвовали в опознании 30 серий по 50 случайных ДНК длиной 250 нуклеиновых оснований, заведомо не являющихся частями структурного гена.

Как и ожидалось, экспериментальное число «опознанных» ДНК и предварительно вычисленное по формуле (3) достаточно близки между собой. Например, для пептидов длиной 6 остатков среднее ожидаемое число искусственных ДНК, идентифицированных как часть структурного гена, совпадает с полученным в численном эксперименте (0,065). Этими расчетами показана возможность применения аналитической формулы (3) для практической оценки числа возможных случайных опознаний.

Полученные с помощью статистических испытаний критерии опознания структурных генов были использованы для идентификации модельных фрагментов длиной 250 пар оснований из пяти реальных ДНК: *Lon*-гена, кодирующего АТР-зависимую La-протеиназу *E. coli* [15] (2812 пар оснований, A : C : G : T = 26 : 24 : 28 : 22), генов α- и γ-субъединиц ФДЭ циклического GMP быка [5, 14] (2215, A : C : G : T = 28 : 24 : 26 : 22, 833 пары оснований, A : C : G : T = 21 : 33 : 28 : 18); гена α-субъединицы G белка быка [16] (3099 пар оснований, A : C : G : T = 30 : 16 : 21 : 33); гена α-субъединицы Na, K-АТР-азы из почек свиньи [4] (3420 пар оснований, A : C : G : T = 24 : 28 : 27 : 21). Как отмечалось ранее, нуклеотидный состав обратной цепи роли не играет. Естественно, что выбранные последовательности ДНК включали в себя вместе с кодирующими областями и прилегающие фланкирующие участки.

Таблица 3

Рассчитанное максимальное число пептидов в наборе для достоверной идентификации фрагментов структурного гена *

Длина пептида	Минимальное число совпадений с транслируемым участком гена	Максимально допустимое число пептидов в наборе	
		при вероятности 0,95 правильной идентификации фрагментов	при допущении не более 1 «лишнего» опознания с вероятностью 0,95
4	4	–	2
5	5	5	35
6	5	1	7
7	6	10	>50
8	6	5	35
9	6	1	9
10	6	1	7
11	7	17	>50
12	7	5	35

* Приводимые данные справедливы, если общее число фрагментов ДНК не превосходит 50.

В опознании участвовали по 20 пептидов соответствующего белка различной длины (2 тетра-, 2 пента-, 5 гекса-, 5 гепта-, 1 окта-, 3 нона- и 2 декапептида). И фрагменты ДНК и пептиды брали произвольно. В модельных экспериментах использовали критерии идентификации, приведенные в первой и второй колонках табл. 3. Оценив по формуле (2) число случайных опознаний, можно рассчитать вероятность ошибочных опознаний пептидами фрагментов ДНК. Например, для 14 фрагментов гена α-субъединицы Na, K-АТР-азы и 20 опознающих пептидов по данным табл. 2 и формуле (2) математическое ожидание ошибочных опознаний $E=0{,}25$. И по формуле (1) имеем вероятность одного или более случайных опознаний $P=1-P_0=0{,}22$. Для каждого из остальных четырех приводимых объектов вероятность одного или более случайных опознаний еще меньше. Поэтому неудивительно, что при идентификации набором пептидов фрагментов ДНК всех пяти структурных генов с использованием критериев табл. 3 было всего лишь одно случайное опознание.

Таким образом, при расшифровке структуры белка с использованием программы идентификации фрагментов структурного гена с помощью набора пептидов можно по данным табл. 2 и формулам (1) и (2), а в случае ферментативного специфического гидролиза по формуле (3) произвести оценку числа случайных опознаний. При этом, задавшись необходимым уровнем достоверности, возможно выбрать критерии надежной идентификации фрагментов структурного гена.

С целью автоматизации трудоемкой работы идентификации фрагментов структурного гена с помощью набора пептидов нами разработана программа «GENIDENT» для ЭВМ «Искра-226». Программа состоит из двух основных блоков. Первый блок осуществляет ввод набора пептидов для идентификации фрагментов ДНК, ввод самих фрагментов а при необходимости и ввод аминокислотной последовательности гомологичного белка. Второй блок позволяет производить поиск местоположения пептидов на транслированных фрагментах структурного гена.

Программа позволяет вводить до 150 фрагментов ДНК длиной до 4000 пар оснований с файлов гибкого магнитного диска и до 150 пептидов, что вполне достаточно в практических случаях. Ввод полученных гидролизом пептидов исследуемого белка предусматривается как с гибкого магнитного диска, так и непосредственно в диалоговом режиме с клавиатуры ЭВМ. При поиске фрагментов структурного гена используются критерии, приведенные в двух левых столбцах табл. 3. Критерии поиска возможно модифицировать применительно к каждому конкретному пептиду. Для поиска местоположения пептидов в аминокислотной по-

следовательности гомологичного белка в первом и втором блоках программы имеются соответствующие возможности.

Программа «GENIDENT» написана на алгоритмическом языке Бейсик с использованием языка Ассемблер для ускорения работы отдельных блоков. Время работы программы практически не зависит от средней длины фрагмента ДНК; просмотр одного фрагмента ДНК по одной рамке трансляции с помощью одного пептида осуществляется за 1 с. Для опознания 50 фрагментов реконструируемой ДНК с помощью 10 пептидов требуется не более 1 ч машинного времени. Программа «GENIDENT» рассчитана на использование специалистами, обладающими минимальными навыками применения ЭВМ в лабораторных исследованиях, и доступна от авторов в виде объектного модуля, записанного на дискете размером 203 мм. Предлагаемая программа самодокументирована и снабжена основным и вспомогательными меню.

Авторы признательны В. В. Губанову и А. Н. Обухову за участие в постановке задачи и ценные рекомендации, высказанные в процессе обсуждения и реализации отдельных этапов.

СПИСОК ЛИТЕРАТУРЫ

1. *Lipkin V. M., Chertov O. Y., Makarova I. A., Monastyrskaya G. S., Sverdlov E. D., Ovchinnikov Y. A.* // Chemistry of peptides and proteins. V. 2 / Eds Voelter W., Bayer E., Ovchinnikov Y. A., Wunsch E. B.; N. Y.: Walter de Gruyter and Co., 1984.
2. *Овчинников Ю. А., Свердлов Е. Д., Липкин В. М., Монастырская Г. С., Чертов О. Ю., Губанов В. В., Гурьев С. О., Модянов Н. Н., Гринкевич В. А., Макарова И. А., Марченко Т. В., Половникова И. Н.* // Биоорган. химия. 1980. Т. 6. № 5. С. 655–665.
3. *Ovchinnikov Yu. A., Lipkin V. M., Kumarev V. P., Gubanov V. V., Khramtsov N. V., Akhmedov N. V., Zagranichny V. E., Muradov K. G.* // FEBS Lett. 1986. V. 20. № 2. P. 288–292.
4. *Монастырская Г. С., Броуде Н. Е., Мелков А. М., Смирнов Ю. В., Малышев И. В., Арсенян С. Г., Соломатина И. С., Свердлов В. Е., Гришин А. В., Петрухин К. Е., Модянов Н. Н.* // Биорган. химия. 1987. Т. 13. № 1. С. 20–26.
5. *Овчинников Ю. А., Губанов В. В., Храмцов Н. В., Ахмедов Н. Б., Ищенко К. А., Заграничный В. Е., Василевская И. А., Ракитина Т. В., Атабекова Н. В., Быстров Н. С., Северцова И. В., Липкин В. М.* // Докл. АН СССР. 1987. Т. 296. № 2. С. 487–491.
6. *Овчинников Ю. А., Броуде Н. Е., Петрухин К. Е., Гришин А. В., Кияткин Н. И., Арзамазова Н. М., Гевондян Н. М., Чертова Е. Н., Мелков А. М., Смирнов Ю. В., Малышев И. В., Монастырская Г. С., Модянов Н. Н.* // Докл. АН СССР. 1986. Т. 287. № 6. С. 1491–1496.
7. *Александров А. А.* // Журн. Всесоюз. хим. о-ва. 1984. Т. 2. С. 64–69.
8. *Костецкий П. В., Доброва И. Е.* // Биоорган. химия. 1988. Т. 14. № 4. С. 515–521.
9. *Черепанов Д. А., Рапанович И. И.* // Молекулярн. биология. 1987. Т. 21. № 3. С. 820–830.
10. *Миронов А. А., Александров Н. Н., Люсиновская-Гурова Л. В., Кистер А. Э.* // Молекулярн. биология. 1986. Т. 20. № 4. С. 1014–1023.
11. *Бородовский М. Ю., Сприжицкий Ю. А., Голованов Е. И., Александров А. А.* // Молекулярн. биология. 1986. Т. 21. № 3. С. 672–677.
12. *Бородовский М. Ю., Сприжицкий Ю. А., Голованов Е. И., Александров А. А.* // Молекулярн. биология. 1986. Т. 20. № 4. С. 1024–1033.
13. *Бородовский М. Ю., Сприжицкий Ю. А., Голованов Е. И., Александров А. А.* // Молекулярн. биология. 1986. Т. 20. № 5. С. 1390–1398.
14. *Ovchinnikov Y. A., Lipkin V. M., Kumarev V. P., Gubanov V. V., Khramtsov N. V., Akhmedov N. B., Zagranichny V. E., Muradov K. G.* // FEBS Lett. 1986. V. 204. № 1. P. 288–292.
15. *Америк А. Ю., Чистякова Л. Г., Остроумова Н. И., Гуревич А. И., Антонов В. К.* // Биоорган. химия. 1988. Т. 14. № 3. С. 408–411.
16. *Mukada T., Tanabe T., Takahashi H., Noda M., Haga K., Haga T., Ichiyama A., Kangawa K., Hiranaga M., Matsuo H., Numa S.* // FEBS Lett. 1986. V. 197. № 1. P. 305–310.
17. *Bilofsky H. S., Burks C.* // Nucl. Acids Res. 1988. V. 16. № 5. P. 1861–1864.
18. *Феллер В.* Введение в теорию вероятностей и ее приложения. Т. 1. М.: Мир, 1984. С. 170–181.
19. *Закс Л.* Статистическое оценивание. М.: Финансы и статистика, 1986.

Поступила в редакцию
3.II.1989
После доработки
25.V.1989