



УДК 547.962 : 543.51

ИССЛЕДОВАНИЕ ПЕПТИДНЫХ СОЕДИНЕНИЙ МЕТОДОМ МАСС-СПЕКТРОМЕТРИИ ВЫСОКОГО РАЗРЕШЕНИЯ

III. КОМПЬЮТЕРНАЯ ПРОГРАММА ДЛЯ ОПРЕДЕЛЕНИЯ АМИНОКИСЛОТНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ В ЛИНЕЙНЫХ ПЕПТИДАХ

*Меримсон В. Г., Розынов Б. В., Садовская В. Л.,
Овчинников Ю. А.*

*Институт биоорганической химии им. М. М. Шемелкина
Академии наук СССР, Москва*

Разработана компьютерная программа для определения аминокислотной последовательности в пептидах на основании данных аминокислотного анализа и масс-спектров высокого разрешения. В основу работы программы положен метод субмолекулярного группового анализа, позволяющий отбирать из масс-спектра с помощью ЭВМ ионы, несущие информацию об аминокислотной последовательности. Работа программы проверялась при установлении аминокислотной последовательности по масс-спектрам восьми модельных пептидов, содержащих остатки цистеина, триптофана, гистидина, орнитина и других аминокислот.

В течение долгого времени определение аминокислотной последовательности в пептидах осуществлялось по масс-спектрам низкого разрешения [1]. Идентификацию начинали с нахождения в масс-спектре пика молекулярного иона или пиков, отвечающих элиминированию из молекулярного иона простейших элементов (H_2O , алкильный радикал и т. д.). Далее рассматривались ионы, соответствующие последовательной потере аминокислотных остатков из молекулярного иона. Такой подход требовал сведений об аминокислотном составе изучаемого пептида.

В дальнейшем для установления аминокислотной последовательности стали использовать масс-спектры высокого разрешения, при этом процесс идентификации пептидов осуществлялся с помощью ЭВМ. Был опубликован алгоритм программы, использующей для установления строения пептида принцип последовательного вычитания точных масс аминокислотных остатков из массы молекулярного иона [2]. Кроме того, были созданы программы для ЭВМ, устанавливающие аминокислотную последовательность в пептидах путем последовательного прибавления точных масс аминокислотных остатков [3] или групп $NHCHCO +$ боковая цепь [4] к точной массе N-ацильной группировки. В этих случаях не требовалось предварительной информации об аминокислотном составе пептида. Размер пептида также не являлся препятствием для работы предложенных программ. Основным ограничением применимости этих программ являются жесткие требования, предъявляемые к масс-спектру: необходимо наличие интенсивного пика молекулярного иона и ионов, образующихся в результате последовательного разрыва связей $CO-NH$ и $C-CO$.

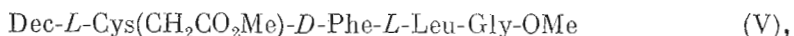
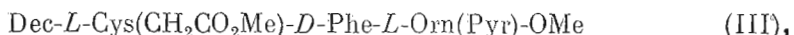
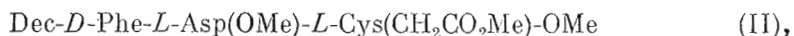
Метод субмолекулярного группового анализа (СГА) [5] расширяет возможности определения аминокислотной последовательности в пептидах с помощью ЭВМ. Согласно этому методу, остатки аминокислот и защитные группировки в пептиде разбиваются на субмолекулярные единицы, из которых с помощью ЭВМ собираются структуры ионов, содержащихся в масс-спектре высокого разрешения изучаемого пептида. При этом из общего числа ионов программой отбираются лишь ионы, несущие информацию об аминокислотной последовательности (ионы типа АП). В число ионов типа АП обязательно входят ионы, образующиеся в результате разрыва связей C—NH, NH—CO и C—CO, а также ряд других фрагментов: ионы, образующиеся в результате перегруппировочных процессов; ионы, возникающие при разрыве боковых цепей; ионы, построенные из двух или более аминокислотных остатков и не содержащие ни одной из концевых защитных группировок. При таком подходе значительно возрастает количество ионов, используемых для установления аминокислотной последовательности, причем в их число попадает много неучитываемых прежними методами ионов высокой интенсивности.

Метод СГА был применен [5] при масс-спектрометрическом установлении строения единственного пептида Boc-Ala-Pro-Ser(Bzl)-Gly-OEt. В его масс-спектре присутствовал интенсивный пик молекулярного иона, значение массового числа которого позволило авторам определить аминокислотный состав пептида, а наличие характеристических ионов дало возможность установить последовательность аминокислотных остатков в нем, причем эта трудоемкая процедура проводилась без применения ЭВМ.

Опыт использования ЭВМ для установления строения органических молекул на основании масс-спектрометрических данных показал, что наиболее доступными для этих целей являются малые лабораторные компьютеры, непосредственно соединенные с масс-спектрометрами [6, 7]. Использование малых компьютеров требует создания специальных программ, позволяющих обрабатывать большой объем информации при минимальном объеме памяти ЭВМ [8].

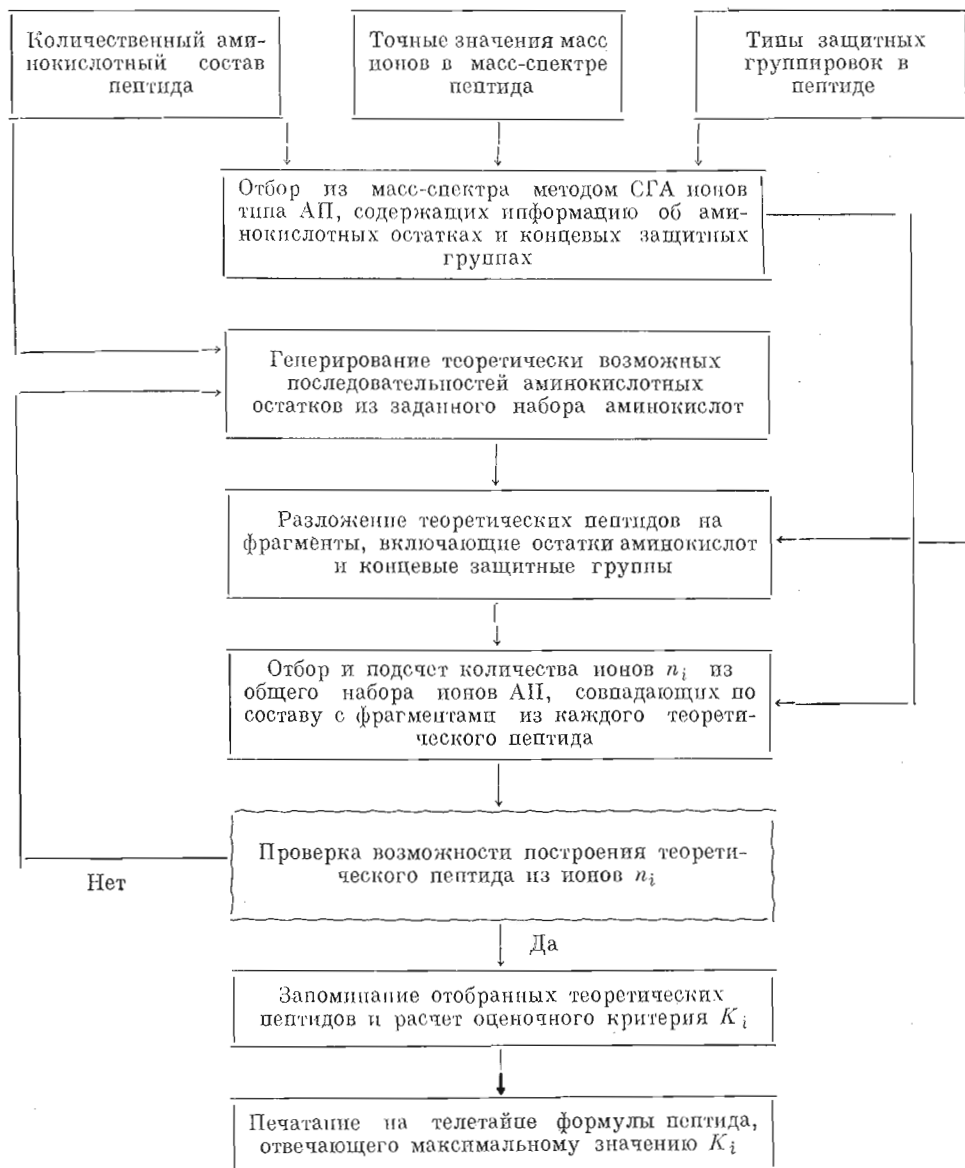
Мы попытались разработать на основе метода СГА программу для мини-компьютера PDP-8/I для определения аминокислотной последовательности в линейных пептидах, содержащих функциональные группы в боковых цепях (см. схему).

В качестве моделей для отладки программы были выбраны пептиды, которые имели в своем составе остатки цистеина, аспарагиновой кислоты, триптофана и других аминокислот:



где Dec — деканоил, Pyr — пиримидил, Orn — орнитин. Масс-спектры высокого разрешения пептидов (I) — (VII) в большинстве своем не содержали пиков молекулярных ионов [9], а также отдельных характеристических фрагментов, обычно используемых для установления строения пептидов, поэтому при составлении программы не представлялось возмож-

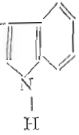
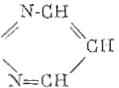


Блок-схема программы определения аминокислотной последовательности в линейных пептидах



ным применить алгоритмы, приведенные в работах Бимана [4] и Мак-Лафферти [3].

Используемые при составлении программы субмолекулярные группы представляли собой структурные элементы пептидной молекулы, составленные с учетом известных закономерностей распада этих соединений при электронном ударе (табл. 1). Точные значения масс субмолекулярных групп, приведенные в табл. 1, были заложены в программу и использовались ЭВМ для составления комбинаций субмолекулярных групп, соответствующих массам ионов, измеренным с точностью 5—10 м. д. массы. При составлении комбинаций программа ЭВМ учитывала взаимное соотношение числа и типа субмолекулярных групп, вытекающее из вероятной структуры ионов.

Массы субмолекулярных групп, используемых в программе для анализа пептидов

Субмолекулярные группы	Обозначение в программе	Массы, а. е. м.
NH-CH ₂	Gl	29,026547
NH-CH(CH ₃)	Al	43,042196
NH-CH(C ₂ H ₅)	Va	71,073495
NH-CH(C ₄ H ₉)	Le	85,089144
NH-CH-CH ₂	As	42,034372
NH-CH-CH ₂	Cy	42,034372
NH-CH-CH ₂ O	Se	58,029286
NH-CH-CH ₂ C ₆ H ₅	Ph	119,073495
NH-CH-CH ₂ C ₆ H ₄ OH	Ty	135,068409
NH-CH-CH ₂ - 	Tr	158,084393
NH-CH-(CH ₂) ₃ NH-C 	Or(Pyr)	164,106190
NH-CH-CH ₂ - 	Hi	109,063993
NH-CH-CH ₂ - 	MeHi	123,079642
C ₃ H ₁₀	De	127,148667
OCH ₃	OMe	31,018387
CO	CO	27,994914
NH	NH	15,010898
H	H	1,007824
SCH ₂	SCH ₂	45,987721
O	O	15,994914
C ₄ H ₉	Bu	57,070421

В тех случаях, когда точные значения массовых чисел субмолекулярных групп, характеризующих остатки аминокислот, совпадают (например, для цистеина и аспарагиновой кислоты), программа предусматривает несколько комбинаций субмолекулярных групп для одного и того же иона (см. в табл. 2 ионы с *m/e* 233 и 102). В таких случаях все комбинации субмолекулярных групп, даваемые ЭВМ, рассматриваются как самостоятельные ионы типа АП.

Из-за отсутствия в масс-спектрах высокого разрешения пептидов (I) — (VII) пиков молекулярных ионов при составлении программы в качестве входных данных нами использована независимо полученная информация об аминокислотном составе пептида. Наличие такой информации не требует включения в набор ионов типа АП фрагментов, содержащих все аминокислотные остатки, а также ионов, состоящих из одного аминокислотного остатка без каких-либо концевых защитных группировок.

Далее программа оперирует комбинациями субгрупп, оставляя лишь остатки аминокислот и защитные группировки. Причем если в конкретной субгрупповой композиции содержатся два или более аминокислотных остатка, имеющих в боковой цепи защитную группу, эквивалентную концевым защитным группам (Dec и OMe), то программа приписывает этой композиции несколько различных структур, часть которых содержат

Отбор методом СГА ионов типа АП в масс-спектре пептида (II)

Измеренная масса иона, а.е.м.	Субмолекулярный состав ионов типа АП						Преобразованный набор ионов типа АП, содержащий информацию об аминокислотных остатках и концевых защитных группах		
	De	Ph, As	(CO) ₃	OMe	+H		De	Ph, As	
404,2696	De	Ph, As	(CO) ₃	OMe	+H		De	Ph, As	
302,2076	De	Ph	(CO) ₂				De	Ph	
274,2167	De	Ph	CO				De	Ph	
260,0908		Ph, As	(CO) ₃	OMe	-H	-NH		Ph, As	OMe
233,1051		Ph, Cy	(CO) ₂	OMe		-NH		Ph, Cy	OMe
		Ph, As	(CO) ₂			-NH		Ph, As	OMe
								Ph, As	OMe
102,0552		Cy	CO	OMe	+H			Cy	OMe
		As	CO	OMe	+H			As	OMe

концевые защитные группы, а в остальных эта группа принадлежит боковой цепи остатка аминокислоты. Очевидно информация об аминокислотной последовательности в обоих случаях различна и, следовательно, первоначальный набор ионов типа АП возрастает. В качестве примера в табл. 2 приведен результат компьютерной обработки масс-спектра высокого разрешения трипептида (II).

Полученный таким образом набор ионов типа АП используется далее программой для установления структуры изучаемого пептида. Этот процесс начинается с генерирования всех теоретически возможных последовательностей из заданного набора аминокислот. Обозначим символом i порядковый номер каждой последовательности. Теоретические пептиды далее разлагаются программой на все возможные фрагменты, содержащие остатки аминокислот и концевые защитные группировки. Фрагменты, полученные из i -го теоретического пептида, сравниваются по составу последовательно со всеми ионами из общего набора ионов АП. Программа отбирает и подсчитывает часть ионов n_i из набора АП, которые совпадают по составу с фрагментами из каждого теоретического пептида, и запоминает все теоретические последовательности, которые могут быть построены из отобранных ионов n_i .

Для оценки реальности полученных последовательностей аминокислотных остатков нами вводится критерий K_i . Этот критерий представляет собой выраженное в процентах отношение числа ионов n_i к общему числу ионов типа АП в наборе

$$K_i = \frac{n_i}{\sum_{АП}} 100\%.$$

Можно ожидать, что истинной аминокислотной последовательности должно соответствовать максимальное значение величины K_i . В табл. 3 приведен результат окончательной машинной обработки масс-спектров высокого разрешения по установлению аминокислотной последовательности в пептидах (I) — (VII).

Программа построена таким образом, что после расчета K_i для каждой из отобранных последовательностей она печатает на телегайле истинную аминокислотную последовательность, которой соответствует максимальное значение K_i (табл. 3).

С использованием разработанной программы была установлена аминокислотная последовательность в тетрапептиде $C_{32}H_{56}N_6O_6$, Dec-His-Leu-Ala-Leu-OMe (VIII) (см. табл. 3). Особенности масс-спектрометрического поведения этого пептида хорошо изучены [10]. В масс-спектре низкого

Т а б л и ц а 3

Результат компьютерной обработки масс-спектров-высокого разрешения пептидов (I) — (VIII)

Пептид	Набор вероятных последовательностей аминокислотных остатков	K_i
(I)	Dec-Cys(CH ₂ CO ₂ Me)-Tyr-OMe	60,0
(II)	Dec-Phe-Asp(OMe)-Cys(CH ₂ CO ₂ Me)-OMe	60,0
	Dec-Asp(OMe)-Phe-Cys(CH ₂ CO ₂ Me)-OMe	50,0
(III)	Dec-Cys(CH ₂ CO ₂ Me)-Phe-Orn(Pyr)-OMe	100,0
	Dec-Phe-Cys(CH ₂ CO ₂ Me)-Orn(Pyr)-OMe	54,6
	Dec-Cys(CH ₂ CO ₂ Me)-Orn(Pyr)-Phe-OMe	45,5
(IV)	Dec-Gly-Leu-Val-OMe	100,0
	Dec-Gly-Val-Leu-OMe	57,2
	Dec-Leu-Gly-Val-OMe	42,9
(V)	Dec-Cys(CH ₂ CO ₂ Me)-Phe-Leu-Gly-OMe	82,4
	Dec-Cys(CH ₂ CO ₂ Me)-Phe-Gly-Leu-OMe	64,7
	Dec-Cys(CH ₂ CO ₂ Me)-Leu-Phe-Gly-OMe	53,0
	Dec-Cys(CH ₂ CO ₂ Me)-Gly-Phe-Leu-OMe	41,2
	Dec-Cys(CH ₂ CO ₂ Me)-Gly-Leu-Phe-OMe	41,2
	Dec-Phe-Cys(CH ₂ CO ₂ Me)-Leu-Gly-OMe	41,2
	Dec-Cys(CH ₂ CO ₂ Me)-Leu-Gly-Phe-OMe	35,3
	Dec-Phe-Cys(CH ₂ CO ₂ Me)-Gly-Leu-OMe	29,4
	Dec-Leu-Phe-Cys(CH ₂ CO ₂ Me)-Gly-OMe	29,4
	Dec-Leu-Cys(CH ₂ CO ₂ Me)-Phe-Gly-OMe	23,5
	Dec-Phe-Leu-Cys(CH ₂ CO ₂ Me)-Gly-OMe	17,7
(VI)	Dec-Gly-Trp-Ala-Ala-OMe	69,2
	Dec-Gly-Ala-Trp-Ala-OMe	61,5
	Dec-Trp-Gly-Ala-Ala-OMe	38,5
	Dec-Gly-Ala-Ala-Trp-OMe	30,8
	Dec-Trp-Ala-Gly-Ala-OMe	30,8
	Dec-Ala-Gly-Trp-Ala-OMe	30,8
	Dec-Ala-Trp-Gly-Ala-OMe	30,8
(VII)	Dec-Cys-(CH ₂ CO ₂ Me)-Ser(Bu)-Gly-Leu-OMe	52,5
	Dec-Cys(CH ₂ CO ₂ Me)-Ser(Bu)-Leu-Gly-OMe	47,6
	Dec-Cys(CH ₂ CO ₂ Me)Gly-Ser(Bu)-Leu-OMe	38,1
	Dec-Ser-(Bu)-Cys(CH ₂ CO ₂ Me)-Gly-Leu-OMe	38,1
	Dec-Cys(CH ₂ CO ₂ Me)-Leu-Ser(Bu)-Gly-OMe	33,3
	Dec-Ser(Bu)-Cys(CH ₂ CO ₂ Me)-Leu-Gly-OMe	33,3
	Dec-Gly-Ser(Bu)-Cys(CH ₂ CO ₂ Me)-Leu-OMe	33,3
	Dec-Gly-Cys(CH ₂ CO ₂ Me)-Ser(Bu)-Leu-OMe	31,3
	Dec-Ser(Bu)-Gly-Cys(CH ₂ CO ₂ Me)-Leu-OMe	23,8
	Dec-Cys(CH ₂ CO ₂ Me)-Gly-Leu-Ser(Bu)-OMe	23,8
	Dec-Cys(CH ₂ CO ₂ Me)-Leu-Gly-Ser(Bu)-OMe	23,8
(VIII) *	Dec-His-Leu-Ala-Leu-OMe	78,3
	Dec-His-Leu-Leu-Ala-OMe	56,5
	Dec-Leu-His-Leu-Ala-OMe	43,5
	Dec-His-Ala-Leu-Leu-OMe	43,5
	Dec-Leu-His-Ala-Leu-OMe	39,1
(VIII) **	Dec-His-Leu-Ala-Leu-OMe	76,0
	Dec-His-Leu-Leu-Ala-OMe	52,0
	Dec-His-Ala-Leu-Leu-OMe	52,0
	Dec-Leu-His-Ala-Leu-OMe	40,0
	Dec-Leu-His-Leu-Ala-OMe	36,0
	Dec-Ala-His-Leu-Leu-OMe	16,0

* Без учета межмолекулярного метилирования.

** С учетом межмолекулярного метилирования.

Таблица 4

Ионы типа АП без учета межмолекулярного метилирования, отобранные ЭВМ из масс-спектра пептида (VIII)

Измеренная масса иона, а.е.м.	Относительная интенсивность, % от максимальной	Субмолекулярный состав ионов
405,2854	2,2	De Le, Hi (CO) ₃
293,2028	1,8	De Hi (CO) ₂ +H
292,2025	16,4	De Hi (CO) ₂
291,1943	2,0	De Hi (CO) ₂ -H
265,2147	7,8	De Hi CO +H
264,2082	28,5	De Hi CO
249,1314	0,7	Le, Hi CO -H
217,1566	0,7	Le, Al (CO) ₂ +2H OMe
206,1309	4,5	Le, Hi CO -NH -H
183,1112	2,1	Le, Al (CO) ₂ -H
167,0953	0,9	Le, Al (CO) ₂ -NH -2H
146,1169	3,7	Le CO +2H OMe
142,1237	0,9	Le, Al CO -NH +H
139,1035	1,7	Le, Al CO -NH -2H

Таблица 5

Ионы типа АП с учетом межмолекулярного метилирования, отобранные ЭВМ из масс-спектра пептида (VIII)

Измеренная масса иона, а.е.м.	Относительная интенсивность, % от максимальной	Субмолекулярный состав ионов
419,3019	5,2	De Le, MeHi (CO) ₃
349,2667	1,3	De Al, MeHi (CO) ₂
307,2185	3,4	De MeHi (CO) ₂ +H
306,2189	19,3	De MeHi (CO) ₂
305,2092	1,5	De MeHi (CO) ₂ -H
293,2028	1,8	Le, Al, MeHi (CO) ₂ +H -HN
279,2272	16,0	De MeHi CO +H
278,2230	60,3	De MeHi CO
277,2149	1,3	De MeHi CO -H
220,1450	5,8	Le, MeHi CO -H -NH
217,1566	0,7	Le, Al (CO) ₂ +2H OMe
183,1112	2,1	Le, Al (CO) ₂ -H
167,0953	4,4	Le, Al (CO) ₂ -2H -NH
146,1169	3,7	Le CO +2H OMe
142,1237	0,8	Le, Al CO +H -NH
139,1035	1,7	Le, Al CO -2H -NH

разрешения наряду с пиком молекулярного иона (M^+) присутствует пик иона $(M + CH_2)^+$. Появление в масс-спектре пика иона $(M + CH_2)^+$ является характерной особенностью гистидинсодержащих пептидов, которые в условиях масс-спектрометрирования легко подвергаются межмолекулярным реакциям, приводящим к метилированию атома азота в боковой цепи гистидинового остатка [3, 11]. Часто пик молекулярного иона отсутствует в спектрах гистидинсодержащих пептидов, а наблюдается только

пик иона ($M + \text{CH}_2$)⁺. Ионы, образующиеся в результате разрыва в молекулярном ионе связей CO—NH и C—CO, сопровождаются ионами, отличающимися на 14 а.е.м., пики которых в масс-спектре имеют большую интенсивность. По этой причине в программе, разработанной Мак-Лафферти с соавт. [3], учитывается процесс метилирования гистидинового остатка и из масс-спектра выбираются пики ионов, включающих остаток метилгистидина.

Наша программа также позволяет учесть процесс молекулярного метилирования гистидинсодержащих пептидов в условиях масс-спектрометрирования, однако в ней предусматривается и возможность анализа гистидинсодержащих пептидов без учета этого процесса.

Для получения масс-спектра пептид (VIII) брался в количестве ~ 0,1 мкмоль (60—70 мкг). Масс-спектр содержал пики 95 ионов с m/e от 591 до 73. Пик молекулярного иона в масс-спектре пептида (VIII) отсутствует, а пику с m/e 591 отвечает ион $[M + \text{CH}_2 - (\text{CH}_3)_2\text{CH}]^+$ (измеренная масса 591, 3858; вычисленная для $\text{C}_{30}\text{H}_{51}\text{O}_6\text{N}_8$ 591, 3865). Результаты установления последовательности аминокислотных остатков в пептиде (VIII) приведены в табл. 3. Из данных табл. 3 видно, что разработанная программа обеспечивает надежное установление аминокислотной последовательности в гистидинсодержащих пептидах как без учета процессов межмолекулярного метилирования, так и с учетом этих процессов.

В табл. 4 и 5 представлен субгрупповой состав ионов типа АII, которые выбираются ЭВМ из масс-спектра пептида (VIII) при анализе его аминокислотной последовательности. В первом случае программа оперирует с остатком гистидина, а во втором случае — с остатком N-метилгистидина (MeHis). Интенсивности ионов, содержащих остатки гистидина (табл. 4), значительно ниже интенсивностей ионов, содержащих остатки N-метилгистидина (табл. 5), однако это не сказывается на точности определения аминокислотной последовательности с помощью разработанной для ЭВМ программы.

Масс-спектры высокого разрешения пептидов (I) — (VIII) были получены на масс-спектрометре MS-902 с системой обработки данных DS-30 при температуре ионизационной камеры 250° и энергии ионизации 70 эВ. Программа была разработана для мини-компьютера PDP-8/I.

ЛИТЕРАТУРА

1. Shemyakin M. M., Ovchinnikov Yu. A., Kiryushkin A. A. (1971) «Mass Spectrometry: Techniques and Applications», New York, Ed. G. M. A., Milne, 289—325.
2. Barber M., Powers P., Wallington M. J., Wolstenholme W. A. (1966) Nature, 212, 784—787.
3. Senn M., Venkataraghavan R., McLafferty F. W. (1966) J. Amer. Chem. Soc., 88, 5593—5597.
4. Biemann K., Cone C., Webster B. R., Arsenault G. P. (1966) J. Amer. Chem. Soc., 88, 5598—5606.
5. Kunderd A., Spencer R. B., Budde W. L. (1971) Anal. Chem., 43, 1086—1090.
6. Gray N. A. B. (1974) Advances Mass Spectrometry, vol. 6, Barking — London 1073—1074.
7. Robertson D. H., Merrit G., Jr. (1974) Advances Mass Spectrometry, vol. 6, Barking — London, 1011—1017.
8. Koo D. H. K., Sedgwick R. D. (1974) Advances Mass Spectrometry, vol. 6, Barking — London, 1019—1025.
9. Розынов Б. В., Садовская В. Л., Меримсон В. Г., Овчинников Ю. А. (1975) Биоорганич. химия, 1, 494—505.
10. Овчинников Ю. А., Кирюшкин А. А., Виноградова Е. Н., Розынов Б. В., Шемякин М. М. (1967) Биохимия, 32, 427—438.
11. Milne G. W. A., Kiryushkin A. A., Alakhov Yu. A., Lipkin V. M., Ovchinnikov Yu. A. (1969) Tetrahedron, 26, 299—304.

Поступила в редакцию
8.IV.1975

INVESTIGATION OF PEPTIDE COMPOUNDS BY HIGH RESOLUTION MASS
SPECTROMETRY. III. THE COMPUTER PROGRAM FOR DETERMINATION
OF AMINO ACID SEQUENCE IN LINEAR PEPTIDES

MERIMSON V. G., ROSINOV B. V., SADOVSKAYA V. L.,
OVCHINNIKOV Yu. A.

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

The computer program for determination of the amino acid sequence in linear peptides based on amino acid analysis and high resolution mass spectra data is described. The sub-molecular group analysis was utilized for identification of sequential ions from the high resolution mass spectra. The program was applied to the mass spectra of cysteine, tryptophane, histidine, and ornithine containing peptides.
