



УДК 612.82+577.15/17

СПЕЦИАЛИЗИРОВАННЫЙ БАНК ДАННЫХ ПО ПРИРОДНЫМ РЕГУЛЯТОРНЫМ ОЛИГОПЕПТИДАМ EROR—MOSCOW

ЗАМЯТНИН А. А.

Институт нормальной физиологии им. П. К. Анохина АМН СССР, Москва

К настоящему времени известно несколько сотен природных олигопептидов, участвующих в функционировании нервной, иммунной и эндокринной систем. Сведения об их структуре и функциях занесены в специально созданный банк данных. В нем собрана информация о 619 регуляторных олигопептидах с суммарным числом 11158 аминокислотных остатков. Большинство молекул содержит около 10 остатков при выбранном интервале от 2 до 50. На основании полученных результатов дано определение различия между олиго- и полипептидами.

Известно, что относительно небольшие природные соединения олигопептиды в сравнении с белками обладают не только существенно меньшим числом аминокислотных остатков, но и заметно иным спектром биологической активности [1]. Эти вещества привлекают к себе все большее внимание ввиду их важной роли в функционировании различных регуляторных систем организма. Короткие пептиды обладают свойством быть регуляторами различных отделов нервной (нейропептиды), иммунной (пептидные иммуномодуляторы), эндокринной (пептидные гормоны) систем. При этом отмечено, что один и тот же олигопептид способен участвовать в различных типах регуляторных процессов [2, 3], и это позволяет считать их полифункциональными.

Наибольшая информация о структуре и функциях природных молекул пептидной природы к настоящему времени заключена в банке данных, основой которых являются аминокислотные и нуклеотидные последовательности [4—6]. При этом сформированы банки данных как для любых природных пептидных структур [4], так и банки, специализированные по различным структурным и функциональным категориям, например, для сигнальных пептидов [5] или для характеристики функциональных свойств α -химотрипсина [6].

Для того, чтобы иметь возможность рассматривать класс регуляторных олигопептидов с единых позиций, нами сформирован банк данных об их структуре и функциях, ограниченных определенным спектром функциональной активности. Этот банк построен на основе доступных данных, с использованием специальных программ типа DBASE для персонального компьютера системы IBM—PC/AT.

Ему присвоено название EROP – Moscow (Endogenous Regulatory Oligopeptides).

Методические приемы

Основой банка является специализированный файл SEQUENCE, предназначенный для характеристики однозначной величины—первичной структуры природных одноцепочечных пептидных молекул, а также для внесения сопутствующих сведений. Таким образом, каждая запись содержит уникальную природную аминокислотную последовательность, не встречающуюся в других записях.

Таблица:

• Display structure Structure for file: A: SEQUEN31
 DBF Number of records: 00619
 Date of last update: 01/04/89
 Primary use database

Fid	Name	Type	Width	Dec
001	FAM:NMB::K	C	096	
002	FAM:NAME:K	C	004	
003	PEP:NMB::K	C	003	
004	PEP:NAME::K	C	041	
005	ABB:NAME:K	C	013	
006	FST:B:C::K	C	003	
007	FST:SPEC:K	C	003	
008	FST:ORG::K	C	003	
009	FST:SUB::K	C	003	
010	EDG:ORIG:K	L	001	
011	EXG:SPEC:K	C	003	
012	MPTRINBAFK	C	010	
013	PRECLRS::K	L	001	
014	AAR:SUM::K	N	002	
015	TSR:SUM::K	N	002	
016	M:W:::::K	N	006	001
017	SEQ:1::::K	C	053	
018	SEQ:3::::V	C	209	
019	SYN:NAME:V	C	015	
020	TAXON::::V	C	027	
021	ANIX:SEQ:V	C	004	
022	ALL:B:CL:V	C	015	
023	ALL:SPEC:V	C	015	
024	ALL:ORG::V	C	015	
025	ALL:SUB:V	C	015	
026	SS:BONDS:V	C	006	
027	N:STND:R:V	C	012	
028	SEQ:REF:V	C	240	
029	COMMENTS:V	C	240	
030	YEAR:SEQ:V	C	035	
031	COUNTRY::V	C	006	
032	ADD:INFM:V	C	019	

* Total *

01001

Распечатка структуры файла SEQUENCE (31-й релиз). Каждому полю соответствует сокращенное название вводимой в банк характеристики (1—5: нумерация и названия олигопептидов и их семейств; 6—9: биологическая локализация олигопептида, выделенного впервые; 10—11: эндогенность/экзогенность; 12: известные типы функций; 13: наличие данных о предшественнике; 14—15: число аминокислотных остатков, концевых радикалов; 16: величина Mr; 17—18: аминокислотные последовательности; 19: синонимы названия олигопептида; 20: таксон; 21: способ определения последовательности; 22—25: прочие источники выделения; 26—27: до-

Файл SEQUENCE, как и все другие файлы банка, содержит константную и переменную части (таблица). Первая составлена из полей 1—17 и содержит сведения об общих характеристиках олигопептида, включающих терминологию, элементы классификации, биологический источник, из которого данный олигопептид был выделен впервые, аминокислотную последовательность, величины числа аминокислотных остатков и величины M , перечень основных типов активности, наличие сведений о предшественнике. Эти данные целесообразно повторять и во всех других файлах, предназначенных для подробной характеристики функциональных или иных свойств олигопептида. Переменная часть, различная для разных файлов, занимает поля с 18 по 32 и в случае файла SEQUENCE содержит более подробные сведения о структуре, источниках выделения, литературные ссылки и другие сопутствующие сведения.

Для определенности нами были выбраны четыре произвольные широко распространенные функциональные характеристики, которыми часто описывают активность нейропептидов: влияние на сокращение гладкой мускулатуры, регуляция кровяного давления, терморегулирующие свойства и рилзинговая активность. Олигопептид считался активным в том случае, если он обладал одним (или более) типом указанной активности, и это служило основанием для включения его в банк данных [10].

Банк данных EROP—Moscow сформирован с помощью стандартных программ типа dBase. В то же время создан целый ряд специализированных программ, которые позволяют решать задачи, связанные с особенностями структуры, функций, локализации и т. д. В целом сформированный банк представляет собой также самостоятельный объект исследований, на котором можно проводить анализ сходства и различий важнейших характеристик регуляторных олигопептидов.

Результаты исследования

По публикациям, содержащим данные до 1988 г. включительно, в банк данных введены сведения о 525 природных олигопептидах, обладающих заданным спектром активности. Как выяснилось, указанное выше ограничение спектра функциональной активности не привело к отсечению большинства олигопептидов, широко известных не по выбранному нами функциям. Среди введенных в банк структур оказались натрийуретические пептиды млекопитающих, адипокинетические гормоны насекомых и другие, одни названия которых уже свидетельствуют об иных типах функциональной активности. Реальный функциональный спектр активности указанных 525 олигопептидов много богаче, чем выбранный вначале и включает в себя действие многих из них на поведенческую, сон, обучаемость, память, диурез, иммунитет и т. д. (подобная особенность, по нашему мнению, заслуживает отдельной публикации). Нами были обнаружены также еще 94 олигопептида, для которых в настоящее время отсутствуют указания о принадлежности к заданному вначале функциональному

полнотелые характеристики химической структуры: 28, 30—31: ссылки на публикации по определению нервной структуры; 29, 32: комментарии и дополнительная информация). Константные и переменные поля помечены соответственно буквами K и V. Типы полей обозначены тремя буквами: C—текстовая, N—числовая и L—логическая информация, Width характеризует длину поля. Для числовых данных можно указать количество знаков после занятой (Dec).

спектру активности. Сведения об этих олигопептидах также включены в банк данных.

Таким образом, на 1 января 1989 г. банком данных описано 619 природных олигопептидов с общим числом аминокислотных остатков, равным 11158. Если учитывать не только боковые, но и концевые радикалы олигопептидов, то суммарное число всех выступающих из пептидного остова радикалов составит 12310.

Спектр биологических источников регуляторных олигопептидов оказался довольно широким. Включенные в банк молекулы выделялись из 6 биологических типов: Protozoa, Coelenterata, Echinodermata, Amphirota, Mollusca и Chordata. Среди последних наибольшее число получено из представителей млекопитающих (от мыши до слона) и многочисленных видов лягушек.

Существующие белковые банки данных достаточно полно отражают аминокислотные последовательности, содержащие многие десятки, сотни и даже тысячи аминокислотных остатков. В описываемый же нами банк данных включены олигопептиды с числом остатков от 2 до 50 (рис. 1). Как оказалось, именно в этом интервале раз-

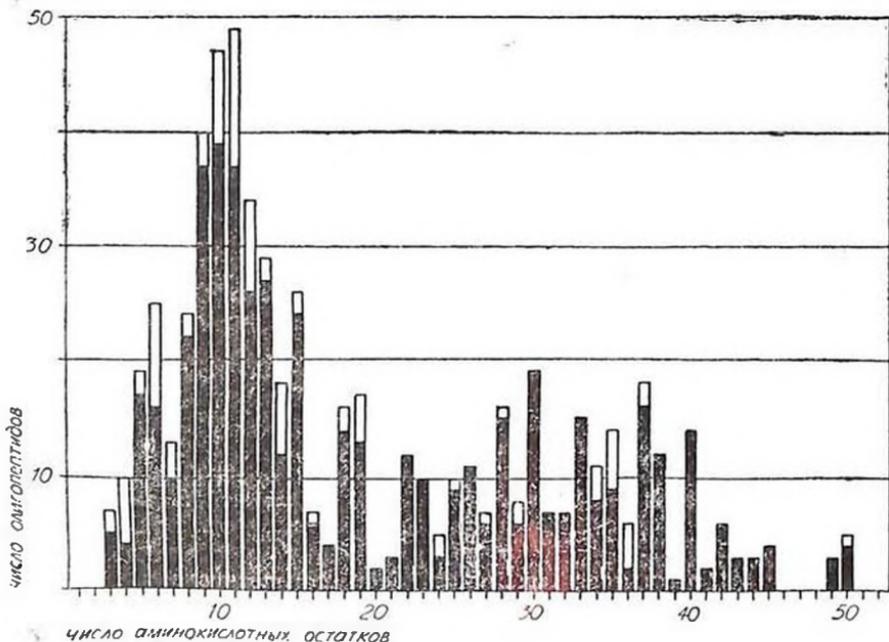


Рис. 1. Распределение числа регуляторных олигопептидов по числу аминокислотных остатков N. Заштрихованная часть представляет собой число олигопептидов с заданным (см. текст) спектром активности.

местилось наибольшее число олигопептидов, обладающих наиболее характерными регуляторными свойствами, которые описываются заданным нами спектром функциональной активности. Следует также

отметить, что ввиду значительной конформационной подвижности олигопептидов [11] их выступающие наружу радикалы (боковые и концевые) хорошо доступны окружению (например, рецепторам). В то же время, из-за усадки внутримолекулярных взаимодействий конформационная подвижность олигопептидов с увеличением длины ладаст и при $N > 50$, где N —число аминокислотных остатков; из них могут образовываться достаточно устойчивые глобулы с иным спектром функциональной активности [12]. Этим, по-видимому, можно объяснить то, что на рис. 2 число олигопептидов с заданной длиной активности при больших N заметно снижается. В этом распределении достаточно ясно виден максимум вблизи $N=10$, а также, возможно, и в районе 30 аминокислотных остатков. Заметим при этом, что среднее число аминокислотных остатков в одном олигопептиде из 619 должно было бы равняться примерно 18. Выявленные на рис. 2 преимущественные области содержания числа аминокислотных остатков могут оказаться оптимальными с физико-химической точки зрения для взаимодействия лиганда с рецептором.

Обсуждение результатов

Сформированный банк данных вместе с имеющимися для него стандартными и специально созданными программами может быть основой для структурно-функционального классифицирования любого нового природного или искусственно синтезированного (но составленного из стандартных аминокислотных остатков) олигопептида. После введения в банк сведений о его аминокислотной последовательности в результате соответствующей обработки можно ответить, например, на следующие вопросы:

обладает ли исследуемый олигопептид аминокислотной последовательностью, характерной для уже ранее известной молекулы; является ли этот олигопептид фрагментом уже изученного;

имеется ли структурное сходство у исследуемого олигопептида с каким-либо уже известным, а также к какому структурно-функциональному семейству он может быть отнесен;

имеет ли исследуемый олигопептид уникальную (новую) аминокислотную последовательность, не выявленную ранее ни для одного из изучавшихся олигопептидов.

Таким образом, выявление достаточного сходства структуры (на уровне аминокислотной последовательности) изучаемого олигопептида и представителя хорошо изученного семейства может открыть новые возможности прогнозирования его функциональных свойств до проведения соответствующего биологического тестирования.

На основании полученных результатов можно также сделать попытку ответить на вопрос о том, какие пептиды следует считать малыми (олигопептидами), а какие большими (полипептидами). Для этого воспользуемся соображениями о малых и больших числах, упомянутыми автором в 1958 г. на семинаре А. Н. Колмогорова, которые не удалось обнаружить в опубликованном виде [13].

В случае чисел малыми предлагалось называть такие, величина которых составляет значение порядка величины системы счисления (то есть в десятичной системе величину порядка 10). Большими же числами тогда будут те, которые много больше величины системы счисления.

Аналогом порядка системы счисления в пептидах естественно выбрать число составляющих их аминокислотных остатков, природное разнообразие которых описывается величиной 20. Продолжая

аналогно, олигопептидами будем называть такие молекулы, для которых N порядка 20, то есть $N > 20$ или $N \leq 20$ при $N \neq 1$. В то же время для полипептидов (обычно белков) $N \gg 20$. Однако исследование физико-химического сходства ряда радикалов аминокислотных остатков (остатки треонина и серина, аргинина и лизина и др.) [1] реальная система счисления в случае олигопептидов составит величину, меньшую, чем 20.

На основании проведенных рассуждений рассмотрим общую схему распределения природных пептидов в зависимости от числа N . На рис. 3 вся занимаемая ими область находится в пределах от 2 (дипептиды) до примерно 10 аминокислотных остатков. Левое и правое подраспределения относятся к олиго- и полипептидам соответственно. Их частичное перекрывание обусловлено многообразием возможных вариантов аминокислотных составов и последовательностей, среди которых в области перекрывания могут обнаружиться молекулы с явной тенденцией к глобулярной структуре при достаточно малых N (этому, например, способствует наличие дисульфидных мост-

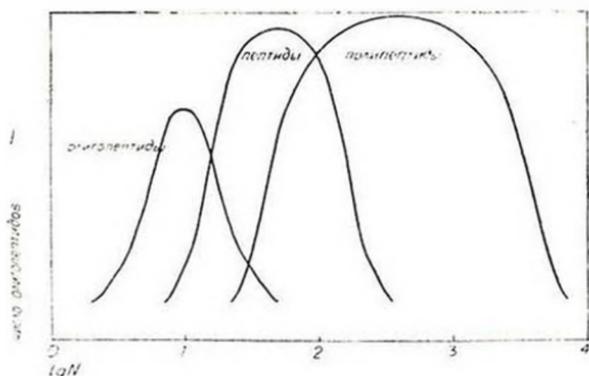


Рис. 2. Классификация общепринятых определений пептидов. По оси ординат—число веществ пептидной природы в произвольных единицах

тиков у остатков цистеинов). На рис. 2 показана также и промежуточная область с не слишком большим N («средние пептиды»). Эту область в значительной мере занимают ферменты [14], одна из функций которых заключается в том, чтобы переводить полипептиды в олигопептиды. Очевидно, что выбор конкретных спектров функций может привести к другим типам распределений.

Число расшифрованных аминокислотных последовательностей регуляторных олигопептидов продолжает быстро расти. Из рис. 3 видно, что к настоящему времени в год расшифровывается уже более 60 таких последовательностей. Судя по оценке [15], общее число природных регуляторных олигопептидов для одного организма составляет, по-видимому, величину порядка 10000. Поэтому следует ожидать, что параллельное развитие компьютерной техники и увеличение возможности записывания большей информации на од-

ном носителе позволит и далее использовать персональные компьютеры для хранения данных и исследования природных олигопептидов без использования крупных вычислительных комплексов.

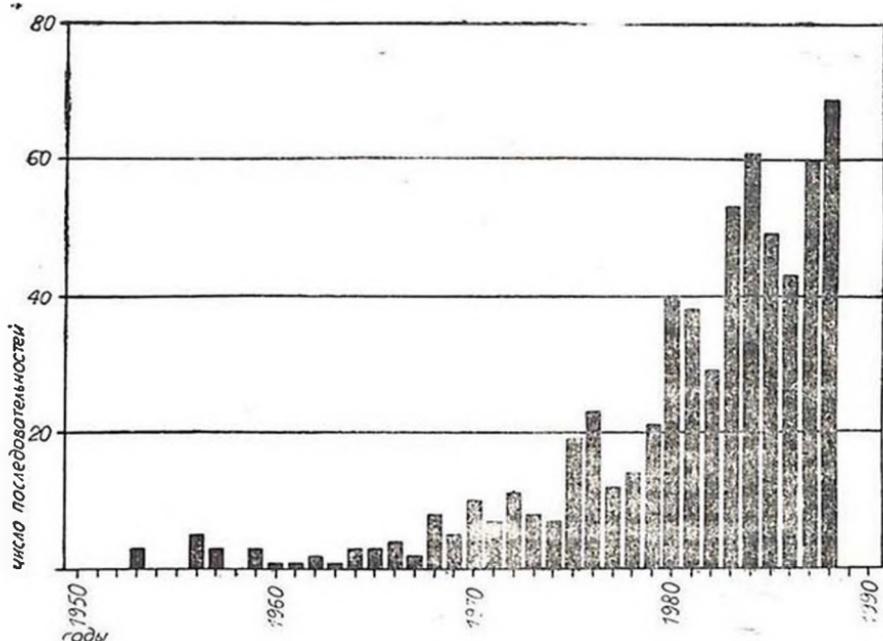


Рис. 3. Рост числа расшифрованных аминокислотных последовательностей природных регуляторных олигопептидов. Видно, что к настоящему времени расшифровывается более 60 первичных структур в год

Автор приносит благодарность К. В. Судакову и В. В. Шерстневу за постоянную поддержку этой работы, а также П. К. Климову, А. М. Молчанову и Ю. Л. Соколову за плодотворные дискуссии и советы.

EROP-MOSCOW SPECIALIZED DATA BANK FOR ENDOGENOUS REGULATORY OLIGOPEPTIDES

ZAMYATNIN A. A.

Anokhin Institute of Normal Physiology, USSR Academy of Medical Sciences, Moscow

At present we know several hundred endogenous oligopeptides participating in the functioning of the nervous, immune and endocrine systems. The information on their structure and functions was collected in a specially created data bank. Data on 619 regulatory oligopeptides with the total number of amino acid residues of 11158 have been collected.

Most molecules are about 10 amino acid residues in length, the range being from 2 to 50. The result obtained permitted us to define some differences between oligo- and polypeptides.

ЛИТЕРАТУРА

1. Ленинджер А. Основы биохимии, т. 1, М., Мир, 1985.
2. Morley J. E., Kay N. E., Solomon G. F., Plotnikoff N. P. *Life Sci.*, v. 41, № 5, p. 527—544, 1987.
3. Scharrer B. *Ann. Rev. Neurosci.*, v. 10, p. 1—17, 1987.
4. Orcutt B. C., George D. G., Dayhoff M. O. *Ann. Rev. Biophys. Bioeng.*, v. 12, p. 419—411, 1983.
5. Claverie J. -M., Sauvaget I. *Nature*, v. 318, № 6041, p. 19, 1985.
6. George D. G., Mewes H. W., Kihara H. *Protein Sequence and Data Analysis*, v. 1, № 1, p. 27—39, 1987.
7. Claverie J. -M., Bricault L. *Proteins: Structure, Function and Genetics*, v. 1, № 1, p. 60—65, 1986.
8. von Heijne G. *Protein Sequence and Data Analysis*, v. 1, № 1, p. 41—42, 1987.
9. Keil B. *Protein Sequence and Data Analysis*, v. 1, № 1, p. 13—20, 1987.
10. Замятнин А. А. Докл. АН СССР, т. 292, № 5, с. 1261—1264, 1987.
11. Karle I. L. -In: *The Peptides: Analysis, Synthesis, Biology* (eds. Gross F., Melnhofer J.), p. 1—54. New York, Acad. Press, 1981.
12. Привалов П. Л. *Биофизика*, т. 30, № 4, с. 722—733, 1985.
13. Zamyatnin A. A. *Ann. Rev. Biophys. Bioeng.*, v. 13, p. 145—165, 1984.
14. Nashikawa K., Kubota Y., Ooi T. J. *Biochem.*, v. 94, № 3, p. 981—995, 1983.
15. Ашмарин Н. П. Чтения им. А. Д. Сперанского, с. 5—16, 1984.

Поступила 18.VIII.1989

СПЕЦИАЛИЗИРОВАННЫЙ БАНК ДАННЫХ EROP—MOSCOW

Специализированный банк данных EROP—Moscow Endogenous Regulatory OligoPeptides содержит сведения об олигопептидах, для которых полностью расшифрована первичная структура (аминокислотная последовательность), охарактеризован спектр функциональной активности, известно место образования в живом организме. Эти и ряд других характеристик относятся к природным одноцепочечным олигопептидам, содержащим от 2 до 50 аминокислотных остатков. На 1 января 1990 года банк EROP—Moscow содержит записи о 819 регуляторных олигопептидах, включающих 16 811 аминокислотных остатков или 18 351 боковой аминокислотный радикал вместе с концевыми группами. Пополнение банка осуществляется 3 раза в год.

Банк EROP—Moscow сформирован с помощью компьютера системы IBM—PC/AT на основе стандартных программ типа dBASE, а также снабжен рядом специализированных программ.

С помощью банка EROP—Moscow на договорной основе можно выполнять следующие заказы:

1. Выдавать информацию о первичной структуре пептидов с указанием литературных источников (25 руб. за 1 запись);

Most molecules are about 10 amino acid residues in length, the range being from 2 to 50. The result obtained permitted us to define some differences between oligo- and polypeptides.

ЛИТЕРАТУРА

1. Ленинджер А. Основы биохимии, т. 1, М., Мир, 1985.
2. Morley J. E., Kay N. E., Solomon G. F., Plotnikoff N. P. *Life Sci.*, v. 41, № 5, p. 527—544, 1987.
3. Scharrer B. *Ann. Rev. Neurosci.*, v. 10, p. 1—17, 1987.
4. Orcutt B. C., George D. G., Dayhoff M. O. *Ann. Rev. Biophys. Bioeng.*, v. 12, p. 419—411, 1983.
5. Claverie J. -M., Sauvaget I. *Nature*, v. 318, № 6041, p. 19, 1985.
6. George D. G., Mewes H. W., Kihara H. *Protein Sequence and Data Analysis*, v. 1, № 1, p. 27—39, 1987.
7. Claverie J. -M., Bricault L. *Proteins: Structure, Function and Genetics*, v. 1, № 1, p. 60—65, 1986.
8. von Heijne G. *Protein Sequence and Data Analysis*, v. 1, № 1, p. 41—42, 1987.
9. Keil B. *Protein Sequence and Data Analysis*, v. 1, № 1, p. 13—20, 1987.
10. Замятнин А. А. Докл. АН СССР, т. 292, № 5, с. 1261—1264, 1987.
11. Karle I. L. -In: *The Peptides: Analysis, Synthesis, Biology* (eds. Gross F., Melnikhofer J.), p. 1—54. New York, Acad. Press, 1981.
12. Привалов П. Л. *Биофизика*, т. 30, № 4, с. 722—733, 1985.
13. Zamyatnin A. A. *Ann. Rev. Biophys. Bioeng.*, v. 13, p. 145—165, 1984.
14. Nashikawa K., Kubota Y., Ooi T. J. *Biochem.*, v. 94, № 3, p. 981—995, 1983.
15. Ашмарин Н. П. Чтения им. А. Д. Сперанского, с. 5—16, 1984.

Поступила 18.VIII.1989

СПЕЦИАЛИЗИРОВАННЫЙ БАНК ДАННЫХ EROP—MOSCOW

Специализированный банк данных EROP—Moscow Endogenous Regulatory OligoPeptides содержит сведения об олигопептидах, для которых полностью расшифрована первичная структура (аминокислотная последовательность), охарактеризован спектр функциональной активности, известно место образования в живом организме. Эти и ряд других характеристик относятся к природным одноцепочечным олигопептидам, содержащим от 2 до 50 аминокислотных остатков. На 1 января 1990 года банк EROP—Moscow содержит записи о 819 регуляторных олигопептидах, включающих 16 811 аминокислотных остатков или 18 351 боковой аминокислотный радикал вместе с концевыми группами. Пополнение банка осуществляется 3 раза в год.

Банк EROP—Moscow сформирован с помощью компьютера системы IBM—PC/AT на основе стандартных программ типа dBASE, а также снабжен рядом специализированных программ.

С помощью банка EROP—Moscow на договорной основе можно выполнять следующие заказы:

1. Выдавать информацию о первичной структуре пептидов с указанием литературных источников (25 руб. за 1 запись);

2. Выдавать дополнительную информацию о функциях пептидов с литературными источниками (15 руб. за 1 источник);

3. Проводить анализ первичной структуры любого пептида с целью выявления сходства со структурами, содержащимися в банке (250 руб. за 1 анализ);

4. Формировать новые специализированные базы данных (500 руб. за разработку структуры 1 банка);

5. Решать прочие задачи (по договорной цене).

В ряде случаев по согласованию с заказчиком указанная ориентировочная цена может быть увеличена или уменьшена.

Результаты выдаются в виде распечатки или в виде записи на стандартный (3- или 5-дюймовый) диск, предоставляемый заказчиком.

Продаваемая информация не может быть передана третьим организациям или лицам.

С заказами обращаться по адресу:

103009, Москва, ул. Герцена, 6,
Институт нормальной физиологии
им. П. К. Анохина АМН СССР
Тел. 155-47-30