

БИОЛОГИЧЕСКИ АКТИВНЫЕ ФРАГМЕНТЫ БЕЛКОВЫХ СТРУКТУР ГЛОБАЛЬНОГО ПРОТЕОМА

Замятнин А.А., Белозерская Т.А.

Институт биохимии им. А.Н. Баха, ФИЦ «Фундаментальные основы биотехнологии» РАН

Ленинский просп., 33, г. Москва, 119071, РФ; e-mail: aaz@inbi.ras.ru

Поступила в редакцию: 02.07.2020

Аннотация. В основе гигантского биоразнообразия лежит также гигантское, молекулярное разнообразие линейных комбинаций 20 аминокислотных остатков в пептидных структурах. Их совокупность представляет собой глобальный протеом Земли. Расшифрованы аминокислотные последовательности ~200 миллионов природных пептидных структур. Полагается, что каждая из них предназначена для выполнения определенной роли в живом организме. Несмотря на столь значительное число известных пептидных молекул, большая часть глобального протеома еще не расшифрована. Тем не менее, известная информация уже используется для различных анализов и обобщений. Огромная часть подобных анализов посвящена изучению фрагментов представителей глобального протеома. Теоретически возможное число разных пептидных структур, составленных всего лишь из 50 аминокислотных остатков, характеризуется величиной $\sim 10^{34}$. Огромное число фрагментов таких структур послужило основанием для возникновения понятия фрагментоники – направления, в рамках которого исследуется структура и функции совокупности белковых фрагментов. Эти данные в течение 30 лет собираются и изучаются в базе данных EROP-Moscow, и их число составляет более 26000. Среди них большое количество регуляторов нервной, эндокринной и иммунной систем, а также разнообразные антимикробные олигопептиды, ингибиторы ферментов и многие другие, обладающие специфическими физико-химическими особенностями. Анализ информации базы данных EROP-Moscow позволил сделать вывод о том, что регуляторные олигопептиды возникают из трех источников. Это – олигопептидные регуляторы, высвобождаемые из эндогенных специализированных предшественников с помощью особых ферментов, обычные эндогенные белки-полипептиды (например, гемоглобин, альбумин и т.д.), расщепляемые протеолитическими ферментами до небольших фрагментов, а также экзогенные белки или их фрагменты, поступающие в организм извне в результате приема пищи или укуса (например, насекомых). Они в организме формируют постоянно меняющийся пул эндогенных и экзогенных веществ пептидной природы (континуум), включенный в функционирование всех регуляторных систем.

Ключевые слова: протеом, белковые структуры.

Введение.

К настоящему времени изучены и описаны первичные структуры (аминокислотные последовательности) и функции более 560 тысяч природных пептидных молекул (<https://www.uniprot.org/uniprot/?query=reviewed:yes>, [1]), обычно называемых белками. Еще больше первичных структур (около 200 миллионов) расшифровано в результате использования генетического кода при трансляции генетической информации с нуклеотидных последовательностей на аминокислотные (<https://www.uniprot.org/uniprot/?query=reviewed:no>) [2]. Работы по получению новых данных в этом направлении ведутся непрерывно, но, вероятно, что на сегодняшний день выявлено еще менее половины первичных структур белков всех разнообразных живых организмов (архей, бактерий, животных, растений и грибов) и вирусов, населяющих Землю. Показано, что пептидные структуры глобального протеома содержат до почти ~36 000 аминокислотных остатков [3, 4]. Достаточно отметить, что и в наши дни постоянно открываются не только новые микроорганизмы, но и ранее неизвестные представители крупных животных (в том числе и млекопитающих, как ископаемых [5], так и ныне живущих [6]), пополняющие наши знания о гигантском природном биоразнообразии. Следовательно, следует ожидать выявления все новых и новых аминокислотных последовательностей вновь открываемых живых организмов.

В процессе исследований множества новых белков появилась и новая терминология. Так, Марк Уилкинс ввел термин «протеом» [7] в 1994 году на симпозиуме «2D-электрофорез: от карт белков до геномов», проходившем в Сиене, Италия. Автор использовал этот термин для описания всего набора белков, экспрессируемых геном, клеткой, тканью или организмом. Таким образом, если протеом представляет собой совокупность всех белков одного живого организма, то совокупность белков всех живых организмов следует называть глобальным протеомом.

Несколько позже (в 2001 г. [8]) наравне с протеомом стал использоваться термин «пептидом», который предназначен для характеристики полного набора пептидов, кодируемых определенным геномом или присутствующих в конкретном типе клеток или в целом организме. А для характеристики полного набора фрагментов одного белка нами был предложен термин «фрагментом» [9-11].

Разнообразие функций белков следует признать гигантским. Однако в основе этого разнообразия и биоразнообразия, заключающемся во всевозможном разнообразии форм жизни на Земле, лежат уникальные последовательности всего 4 нуклеотидов в нуклеиновых кислотах и 20 аминокислотных остатков в белках. Это

разнообразии началось выявляться в 1953 г. в пионерских работах по расшифровке аминокислотной последовательности окситоцина [12, 13], вазопрессина [14] и инсулина [15, 16]. В этих и последующих работах белки расщеплялись с помощью ферментов до небольших фрагментов, определялась химическая структура этих фрагментов и затем из них собиралась полная аминокислотная последовательность исходной расщепленной белковой молекулы. Полученные фрагменты использовались лишь для выявления полной первичной структуры изучаемого вещества, а их функциональные свойства, как правило, не исследовались. Однако в последнее время появляется все больше и больше данных о том, что фрагменты могут выполнять многочисленные функции несвойственные материнской молекуле. В связи с этим в наши дни белки с уже известной аминокислотной последовательностью исследователи расщепляют вновь, но уже не для определения их роли в первичной структуре белка, а для выявления функциональных свойств их фрагментов. Таким образом, быстро растет число данных о функциональной роли фрагментов глобального протеома, которые могут служить основой для целого ряда анализов и обобщений.

Биогенез белковых фрагментов.

Абсолютное большинство известных пептидных структур образовались в результате матричного синтеза. Как известно, полная структура белка обычно формируется в результате считывания нуклеотидной последовательности, начинающейся с единственного триплета, кодирующего аминокислоту метионин, и заканчивается перед одним из трех триплетов, останавливающих считывание – так называемых стоп-кодонах. Однако образовавшаяся молекула, как правило, еще не является сформированным белком, а представляет собой его предшественник, в котором могут быть такие участки, как пре-пептид (сигнальный пептид), про-пептид и собственно сам белок. Далее с помощью специальных ферментов осуществляется фрагментация исходной молекулы. Эти фрагменты и выявляются в экспериментальных исследованиях белков. Рассмотрим три источника обнаруживаемых в живых организмах фрагментов.

Специализированные предшественники. В случае больших белков простым примером такого предшественника является сывороточный альбумин быка (рис. 1А), состоящий из 583 аминокислотных остатков [17]. Его предшественник, содержащий 607 остатков, распадается на фрагменты: сигнальный пептид (остатки 1–18), короткий про-пептид (19–24) и собственно сывороточный альбумин (25–607) [18]. Именно этот большой белок среди трех фрагментов, полученный в эксперименте, рассматривается как функционально значимая структура.

Однако (рис. 1Б) в предшественнике пентапептидов экефалинов (проэнкефалин-А) помимо сигнального пептида содержится несколько про-пептидов, а также одна аминокислотная последовательность лей-энкефалина (**YGGFL**) и пять копий мет-энкефалинов (**YGGFM**) [19].

В природе в одной аминокислотной последовательности предшественника также часто встречается множество копий олигопептидов, разделенных дипептидными парами (фланкирование), которые содержат остатки лизина **K** или аргинина **R**. Так, например, предшественник тетрапептида моллюска *Aplysia californica* FMRFамид содержит 28 одинаковых копий этого олигопептида и один гомологичный тетрапептид FLRFамид (рис. 1В) [20].



Рисунок 1. Предшественники белковых структур. А – сывороточный альбумин быка; Б – энкефалины человека; В – нейропептиды моллюска *Aplysia californica*. Цвета фрагментов: розовый – пре-пептид (сигнальный пептид), черный – про-пептид, зеленый – основная функционально-значимая структура, красный – лизил/аргиниловые участки, узнаваемые расщепляющим предшественник ферментом трипсином.

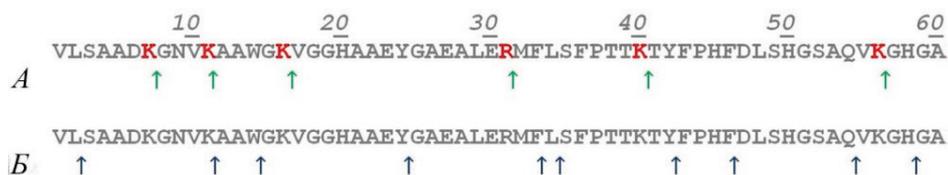


Рисунок 2. Первые 60 аминокислотных остатков α -цепи гемоглобина быка. Стрелками показаны места расщепления А – трипсином и В – химотрипсином



Рисунок 3. Первые 60 аминокислотных остатков α -цепи гемоглобина быка. Стрелками показаны места расщепления пепсином при трех величинах pH

Подобным образом из специализированных предшественников формируются многочисленные фрагменты, которые хорошо экстрагируются и известны как регуляторные олигопептиды.

Эндогенные белки. В последнее время появляется все больше и больше свидетельств тому, что в разных органах и тканях живых организмов присутствуют еще и пептидные структуры, которые образуются не из специализированных предшественников, а являются природными фрагментами хорошо известных белков, образованных из специализированных предшественников. Одна из первых таких работ была посвящена обнаружению в мозге быка почти 200 фрагментов α - и β -цепей гемоглобина и ряда других белков [21]. На сегодняшний день с бурным развитием методов протеомики число таких работ быстро растет, пополняя базы данных о природных фрагментах белков, обнаруженных в живом организме. Однако функции большинства этих фрагментов до сих пор неизвестны. Так, в одной из последних работ описано около 300 фрагментов различных известных белков крысы [22]. Все они авторами названы нейропептидами лишь потому, что выделены из мозга, т.е. из нервной системы. Однако экспериментальное подтверждение их участия в нервной регуляции не было приведено.

Экзогенные белки. В любом живом организме могут быть обнаружены не только фрагменты собственных белков, но белков, сформированных в другом организме. Самым простым путем они попадают в чужой организм в пищу. Их переваривание ферментами желудочно-кишечного тракта приводит к образованию фрагментов, обладающих регуляторными функциями.

На рисунке 2 показано образование фрагментов α -цепи гемоглобина быка под действием специфических ферментов трипсина и химотрипсина, а на рисунке 3 – образование фрагментов у того же белка под действием неспецифического фермента пепсина. Пепсин обладает способностью разрывать большинство из 400 возможных разных пептидных связей белка и доводить процесс до образования отдельных аминокислот. Однако вероятность разрыва ферментами разных связей разная [23], и поэтому в определенный момент времени могут существовать фрагменты с самыми разными комбинациями соседствующих в целом белке аминокислотных остатков. Ранее нами было показано, что время жизни ряда фрагментов может быть достаточным для достижения ими потенциальных рецепторов и осуществления свойственных им регуляторных функций [24].

Кроме этого пути существует и другой путь поступления в живой организм белков и фрагментов извне. Множество животных существуют благодаря способности добывать себе пищу ввода фекально-пептидные токсины в животное-жертву. В этом случае жертва погибает. Однако этого не происходит, например, после укуса кровососущего животного, когда в него из укушенного животного попадают белки крови, фрагменты которых могут быть идентифицированы. Так в гемолимфе клеща *Boophilus microplus* был обнаружен фрагмент α -цепи гемоглобина быка, причем этот фрагмент обладал антибактериальной и антигрибной активностью [25].

Таким образом, фрагменты, выявляемые в живых организмах, отличаются по биогенезу, но все они могут участвовать в процессах регуляции живого организма.

Коллекция данных об известных природных физиологически активных пептидных фрагментах.

Природные физиологически активные фрагменты часто принято называть регуляторными олигопептидами, поскольку в них содержится сравнительно небольшое число аминокислотных остатков – от 2 до 50 [26]. Являющиеся, по-видимому, регуляторами всех жизненно важных процессов всех живых организмов они привлекают внимание исследователей уже более века. Первый олигопептид – дипептид карнозин был открыт в России в 1900 году [27,28], а через 18 лет в США была установлена его химическая структура [29]. С тех пор идентифицировано более 26 тысяч природных олигопептидных регуляторов. Каждый год всей совокупности

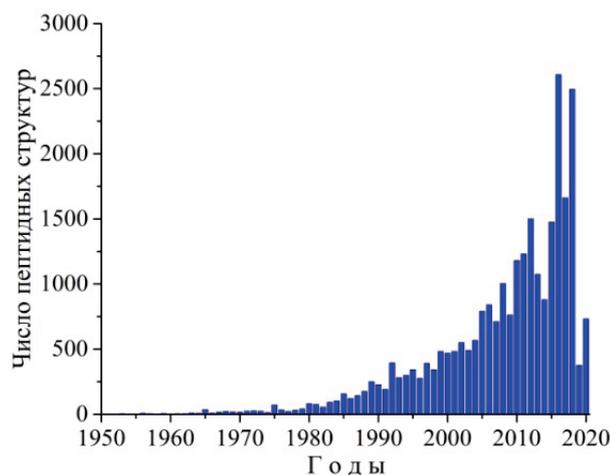


Рисунок 4. Выявление числа функционально значимых фрагментов белков (за 2019-2020 гг. данные еще неполные)

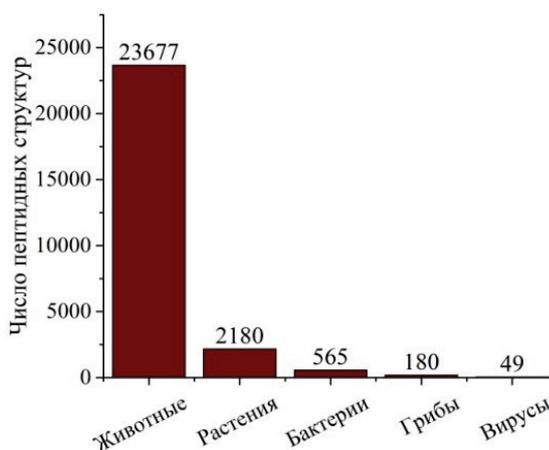


Рисунок 5. Число функционально значимых фрагментов белков у представителей разных биологических доменов, царств и вирусов

исследований в области олигопептидной регуляции посвящается около 10000 публикаций в год, в которых, в частности, впервые описывается более 2000 новых природных структур веществ этого химического класса (рис. 4).

Для понимания сложных регуляторных процессов, осуществляющихся с участием огромного числа олигопептидов, необходима их классификация, первым шагом которой является систематизация химических и биологических данных об этих веществах. Около 30 лет назад нами была предпринята первая попытка провести такую систематизацию [27, 28] и классификацию на ее основе [32, 33]. Одним из результатов этой работы стало создание MS DOS версии базы данных EROP-Moscow (Endogenous Regulatory OligoPeptides) о структуре, функциях и локализации известных к тому времени олигопептидов. Однако долгое время такая база не была доступна широкому кругу пользователей. В дальнейшем ее содержание (при сохранении основной структуры базы) было преобразовано в соответствующий формат и введено в Интернет для всеобщего пользования [34, 35].

Следует отметить, что к настоящему времени создан целый ряд общедоступных белковых баз данных, содержащих огромную информацию. Однако данные собственно о природных олигопептидах в этих базах занимают менее одного процента от общей информации, и их поиск подчас представляет собой весьма трудоемкий процесс. Кроме того, имеющаяся там информация не отличается полнотой.

В то же время база EROP-Moscow предназначена для максимально полного описания известных на сегодняшний день эндогенных и экзогенных функционально значимых фрагментов белков, среди которых нейрорегуляторы (нейропептиды) представляют собой самый большой функциональный класс. Она предоставляет возможность пользователю легко осуществлять поиск олигопептидов по многим характеристикам, быстро проводить сравнительный анализ их свойств, а также получать суммарные статистические данные из всей совокупности информации, введенной в базу. До появления этой базы таких возможностей у пользователей не было.

На рисунке 5 показано, сколько на сегодняшний день изучено разных фрагментов белков глобального протеома, полученных из разных биологических источников.

Источники информации. Целью создания EROP-Moscow являлся сбор воедино основных структурных, функциональных и других характеристик всех известных к настоящему времени природных олигопептидов с максимальной достоверностью. Чтобы избежать ошибок, которые могут встречаться во вторичных материалах,

эта информация экстрагируется только из первичных источников. Абсолютным большинством источников являются публикации в научных журналах. Чтобы ее получить, постоянно проводится скрининг содержания всех выпусков более 100 журналов биохимического, биофизического, физиологического, нейрохимического и общепрофильного профиля, в которых выявляются публикации с описанием структур новых природных олигопептидов. Наряду с тем, что самое большое количество новых пептидных структур опубликовано в специализированном журнале *Peptides* (почти 3000), более 150 разных журналов опубликовали данные лишь об одном физиологически активном фрагменте. Так в журнале *Geology* были опубликованы данные об остеокальцине (олигопептиде, активно связывающим кальций в костях) представителя вымершего вида степного бизона *Bison priscus* [36]. Общее число информационных источников, содержащих первичные структуры функционально значимых фрагментов белков, составляет более 500, а число ученых из 89 стран, принимающих участие в обнаружении новых пептидных структур, превышает 17000.

Отбор природных фрагментов. В EROP-Moscow вводятся данные только о тех олигопептидах, химическая структура которых прямо или в результате трансляции с нуклеотидной последовательности установлена полностью и может быть описана в виде стандартного однобуквенного аминокислотного кода (исключение составляют дипептид карнозин и его природные аналоги). Если в публикации приведена первичная структура, в которой какие-либо аминокислотные остатки (например, аспарагина/аспарагиновой кислоты или глутамин/глутаминовой кислоты, а также лейцина/изолейцина) точно не идентифицированы, то данные об этом олигопептиде в базу данных не вводятся. Не вводятся также данные и об искусственно синтезированных молекулах, которые в живой природе не обнаружены.

Большинство данных, содержащихся в EROP-Moscow, посвящено олигопептидам, сформировавшимся в результате матричного синтеза. Однако существует большое разнообразие олигопептидов (преимущественно у бактерий и грибов), которые образуются в результате нерибосомального тРНК-зависимого [37] или ферментативного синтеза. Если их структуру составляют аминокислотные остатки, которые могут быть описаны стандартным аминокислотным кодом, то данные о них также вводятся в EROP-Moscow.

Сравнение EROP-Moscow с другими базами. В Интернете открыт свободный доступ к большому числу как глобальных, так и специализированных белково-пептидных баз данных. Среди глобальных, наиболее известными являются PIR (**Protein Identification Resource**, <https://proteininformationresource.org/pirwww/>) и UniProt (<https://www.uniprot.org/>).

Существуют также и сравнительно небольшие базы, которые содержат сведения об олигопептидах с особыми структурными или функциональными свойствами. Однако, на сегодняшний день около половины природных олигопептидных структур, имеющихся в EROP-Moscow, в них отсутствует. Очень многие олигопептиды, являющиеся фрагментами более крупных предшественников, в этих базах данных не упоминаются и их там можно найти только по аминокислотной последовательности. Например, в SwissProt имеется запись P01019 о предшественнике ангиотензинов человека, в которой указаны его участки, представляющие аминокислотные последовательности ангиотензинов I и II [38]. Однако о существовании, например, ангиотензинов V и VI [39] там не упоминается, в то время как в EROP-Moscow соответствующие записи имеются (E00165 и E00166).

Значительное число олигопептидов с уникальной аминокислотной последовательностью не включено во многие базы данных еще и потому, что, по-видимому, в поле зрения создателей этих баз не попадает целый ряд изданий, в которых публикуется немалое число новых олигопептидных структур (например, в *Biological Bulletin*). Кроме того, по-видимому, не принимаются во внимание также и сведения, содержащиеся во вторичных публикациях, на которые имеются ссылки в первичных источниках информации при сравнении гомологичных аминокислотных последовательностей.

Следствием указанных выше причин служит то, что, например, в SwissProt имеется существенно меньше сведений о самых коротких олигопептидах, чем в EROP-Moscow (рис. 6).

Полноту данных, содержащихся в EROP-Moscow можно оценить, сравнив ее другими узко-специализированными коллекциями регуляторных пептидов, например с числом олигопептидов, обладающих

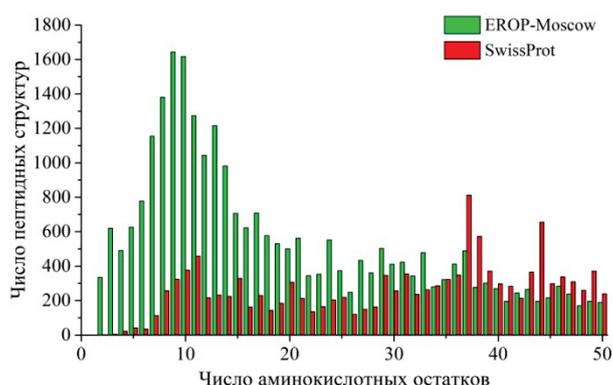


Рисунок 6. Сравнение содержания олигопептидных структур в олигопептидной базе EROP-Moscow и белковой базе SwissProt

Таблица 1. Природные модификации 26571 олигопептидов, занесенных в базу EROP-Moscow

Химическая особенность	Количество пептидов
Образование S–S-связей	7316 (27,3%)
Циклизация	528 (2,0%)
Амидирование С-конца	9574 (35,8%)
Образование пироглутамила на N-конце	1336 (5,0%)
Другие модификации N-конца	132 (0,5%)
Модификации аминокислотных остатков	1328 (5,0%)

определенной функцией. Среди наиболее представленных различных регуляторных олигопептидов база EROP-Moscow содержит сведения о более чем 8000 нейропептидах, около 6000 антимикробных олигопептидах и около 3300 ингибиторах ферментов. Специализированной базы данных, посвященной нейропептидам не существует. Однако антимикробным олигопептидам посвящена база APD (Antimicrobial Peptide Database, <http://aps.unmc.edu/AP/>). В ней содержатся данные об олигопептидах, содержащих и более 50 аминокислотных остатков. Однако общее число разных структур составляет лишь около 3200, т.е. почти в 2 раза меньше, чем в EROP-Moscow.

Другая база данных BioPep (<http://bis.zju.edu.cn/biopepdb/>) посвящена пептидам, обладающим, как и олигопептиды EROP-Moscow, разными функциями и содержащим как менее, так и более чем 50 аминокислотных остатков. Однако, в ней число антимикробных пептидов не превышает 2500, а число ингибиторов ангиотензин-I превращающего фермента (ACEI – angiotensin I-converting enzyme inhibitor) составляет менее 1700, в то время как в EROP-Moscow их содержится более 1700.

Характерные физико-химические особенности фрагментов.

Прямое определение точной химической структуры олигопептидного фрагмента во многих случаях сопряжено с большими трудностями, поскольку их природное содержание часто составляет всего лишь 10^{-15} – 10^{-12} М [40]. Большие проблемы возникают также при определении локализации олигопептида в клетке. Все это требует использования высокочувствительной биохимической и иммуногистохимической техники анализа.

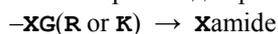
В то же время определение аминокислотной последовательности, осуществляемое трансляцией с соответствующей нуклеотидной последовательности, не требует привлечения высокоточных аналитических биохимических методов.

Посттрансляционные модификации. Однако при таком подходе остаются без внимания многие типы возможных посттрансляционных модификаций, которым подвергается примерно половина известных олигопептидов (табл. 1).

Самая простая из них (но не самая распространенная) – это соединение с новой пептидной связью N- и С-концевых аминокислотных остатков олигопептида. В результате образуется циклическая структура, представляющая собой макроцикл, максимально возможный для данной молекулы. В этом случае понятия N- и С-концов теряют смысл и (если неизвестна первичная структура предшественника) возникает проблема начала считывания аминокислотной последовательности.

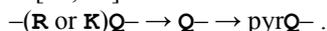
Примерно четверть известных олигопептидов содержит цистеиниловые пары, замыкающиеся в дисульфидные связи (S–S), число которых в одном олигопептиде может достигать шести. При большом числе S–S-связей структура образует несколько макроциклов и как бы завязана в сложный узел, замораживающий большинство степеней свободы молекулы. Следует отметить, что S–S-связи встречаются также и у некоторых циклических олигопептидов.

Наиболее часто встречающейся посттрансляционной модификацией у олигопептидов является амидирование С-конца. Эта модификация, осуществляющаяся при участии различных ферментов [41] или неферментативным путем [42], может быть выявлена и без привлечения аналитических биохимических методов, поскольку амидирование обычно касается того аминокислотного остатка X, за которым следуют остатки глицина (G) и аргинина (R) или лизина (K), в результате чего происходит реакция



(как на рис. 1B в случае FMRFамида).

Также часто (хотя и менее) встречается модификация N-конца, при которой в оказавшемся на этом конце остатке глутамина (Q) боковой радикал неферментативным путем замыкается на N-концевую группу, в результате чего образуется пироглутаминил [43, 44]:



Эта модификация касается большинства N-концевых остатков глутамина и потому после трансляции нуклеотидной последовательности, как правило, может быть предсказана без дополнительного биохимического анализа.

Всего же у олигопептидов известно ~100 различных природных химических модификаций аминокислотных остатков.

Пространственная структура. Олигопептиды и особенно те, которые не содержат ковалентных связей между удаленными аминокислотными остатками, обладают большой конформационной подвижностью. Одна и та же молекула в свободном состоянии с течением времени может изменять свою конфигурацию. Если же рассматривать совокупность одинаковых молекул, то в один и тот же момент времени они могут находиться в разных конформациях. Этим, в частности, объясняются большие трудности при попытках кристаллизовать олигопептиды [45] для дальнейшего проведения рентгеноструктурного анализа.

Возможное конформационное многообразие одной молекулы олигопептида демонстрируется и компьютерным моделированием с помощью программ, позволяющих минимизировать энергии и оптимизировать конфигурацию молекулы [46].

Примером экспериментального исследования пространственной структуры олигопептидов является рентгеноструктурный анализ энкефалинов. В результате такого исследования, например, пентапептида лей-энкефалина **YGGFL** в его кристаллах были обнаружены четыре пространственные формы – от вытянутой до сложной [47]. Этот результат послужил основанием для предположения о том, что данная молекула, находясь в одной конформации, может селективно связываться с одним подтипом опиоидного рецептора, а в другой – с рецептором другого подтипа.

Разнообразие форм одной молекулы олигопептида было также обнаружено с помощью ЯМР-спектроскопии другого опиоидного пентапептида – мет-энкефалина **YGGFM** [48]. Более того, этим методом у данного соединения, находящегося в растворе, было показано существование димеров. Однако следует отметить, что, как в случае рентгеноструктурного анализа, так и ЯМР-спектроскопии, исследовавшиеся молекулы олигопептидов находились в условиях, весьма далеких от тех, которые характерны для живого организма. Поэтому трудно полагать, что к настоящему времени нам известны все возможные конформационные состояния этих и других олигопептидных молекул, которые значимы для осуществления ими регуляторных процессов.

Структуры, полученные как в эксперименте, так и в случае теоретического моделирования, пока не отражают всех особенностей конфигурации олигопептидных молекул, находящихся в реальных условиях. При рентгеноструктурном анализе выявляются лишь те конформации, которые образовались в результате кристаллизации, а данные компьютерного моделирования сильно зависят от качества программ, в которых для расчетов используются приближенные величины, характеризующие разные типы внутримолекулярных взаимодействий в олигопептиде. Поэтому большинство таких данных не дает достоверного представления о молекулярных конфигурациях, в особенности тех, которые олигопептиды принимают при взаимодействии с рецепторными структурами.

Функционально значимые химические группы. Специфический спектр функциональной активности олигопептидов послужил основой для предположения о том, что аминокислотные составы олигопептидных и полипептидных белковых молекул различаются [49, 50]. При сравнении среднего аминокислотного состава регуляторных олигопептидов с суммарным составом белков это различие было обнаружено [51-53] и оказалось, что для некоторых функциональных классов оно довольно заметно. Было показано, что некоторые из этих классов отличаются значительным количеством аминокислотных остатков, содержащих положительно заряженные (**K**, **R**) и циклические (**F**, **H**, **P**, **W**, **Y**) боковые радикалы. Превалирование содержания ряда этих остатков, особенно несущих положительный заряд (при пониженном содержании отрицательно заряженных остатков **D** и **E**), наиболее ярко проявляется у многочисленных представителей функционального класса антимикробных олигопептидов [54] и олигопептидных токсинов [55]. Подобные особенности, но в меньшей степени выявлены также у нейропептидов [56] и олигопептидных гормонов – либеринов и статинов [57].

Существование определенных рецепторных структур (рецепторы, ионные каналы и др.) позволило высказать предположение о том, что выявленные особенности отражают специфику взаимодействия олигопептида (например, гормона или нейропептида) с соответствующей мишенью. Процесс взаимодействия может осуществляться в две стадии [56]. Сначала в результате электростатического взаимодействия положительных зарядов олигопептидного лиганда с отрицательными зарядами рецепторной структуры происходит их неспецифическое сближение. После этого осуществляется необходимая конформационная перестройка олигопептидного лиганда (процесс узнавания), включающая подгонку элементов, участвующих в узнавании функционально значимых радикалов на основе различных слабых взаимодействий, например, с помощью образования межмолекулярных водородных связей и осуществления стэкинг и гидрофобных взаимодействий.

Однако в случае олигопептидных иммунорегуляторов подобное отличие обнаружено не было [58]. По-видимому, это объясняется тем, что в осуществлении иммунного ответа участвует множество органов и специализированных клеток и между ними существует большое разнообразие функциональных связей, которые изучены еще недостаточно. Эти связи осуществляются как специфическими олигопептидными иммунорегуляторами, так и представителями иных функциональных классов олигопептидов (например, нейропептидами). Это обстоятельство не позволяет выявить характерные особенности олигопептидных регуляторов иммунной системы всех вместе взятых. Возможно, что такие закономерности существуют, но не на уровне иммунной регуляции в целом, а на уровне конкретных процессов связывания олигопептидов с определенными структурами, участвующими в этой регуляции. Решение этой задачи станет возможным, когда будет сформировано ясное представление о структурной организации как всей иммунной системы, так и отдельных ее компонентов, включая олигопептиды.

Таким образом, разнообразие аминокислотного состава и аминокислотной последовательности фрагментов, разнообразие посттрансляционных модификаций и разнообразие пространственных форм лежит в основе огромного разнообразия функций, которые могут осуществлять пептидные фрагменты. Очевидно, что 250 функциональных классов, которые можно обнаружить в статистических данных базы EROP-Moscow, не являются пределом, и это число будет увеличиваться в процессе дальнейших исследований.

Структурно-функциональное разнообразие фрагментов глобального протеома.

Несмотря на многообразие структур и функций белков и их фрагментов все они сводятся к уникальным последовательностям аминокислотным последовательностям. А функциональный спектр олигопептидов, по-видимому, практически безграничен. С одной стороны, это объясняется огромным числом возможных разных аминокислотных последовательностей олигопептидов N . Так, в соответствии с известной формулой

$$N_p = n^p, \quad (1)$$

(где p – число аминокислотных остатков в молекуле, а n – число разных аминокислотных остатков, равное 20) в случае дипептидов, когда $p = 2$, $N_2 = 400$, а на другом конце интервала $2 \leq p \leq 50$, т.е. при $p = 50$, $N_{50} = \sim 10^{34}$. Правда, в реальности это число меньше: из-за ограниченных размеров геномов в природе реализуются не все комбинации как нуклеотидов, так и аминокислот. Так, в пептидных структурах встречаются все дипептидные (400) и трипептидные (8000) комбинации, а среди тетрапептидных (160000 комбинаций) уже одна (CQWW) не выявлена. Однако разнообразие возможных первичных структур следует все же признать гигантским.

С другой стороны, как следует из рассмотрения физических особенностей олигопептидов, одна молекула, элементы которой обладают большим числом степеней свободы, может принимать множество пространственных конфигураций (конформаций), позволяющих ей взаимодействовать с самыми разными рецепторными структурами и, таким образом, участвовать в разнообразных функциональных реакциях.

Структурно-функциональное разнообразие фрагментов одного белка может быть продемонстрировано с помощью соответствующих фрагментов, полученных из разных источников (живых организмов, белков) или с помощью синтеза. Такие данные взяты нами из базы EROP-Moscow и сведены вместе для части α -цепи гемоглобина быка (рис. 7). Даже на таком сравнительно небольшом участке только одного белка оказалось несколько десятков фрагментов с тремя типами известной функциональной активности. Подобные иллюстрации могут быть сделаны и для других белков (β -цепи гемоглобина, разных белков казеина, сывороточного и яичного альбуминов и т.д.).

Следует отметить, что функциональные свойства целого ряда выделенных и идентифицированных фрагментов белков (в том числе и гемоглобина), обнаруженных в разных органах и тканях, пока не изучены. Кроме того, очевидно, что идентифицированная и изученная часть фрагментов представляет собой малую часть всех возможных фрагментов. Для оценки этого произведем следующий расчет.

Теоретически возможное максимальное число всех (включая одинаковые) перекрывающихся фрагментов пептидной структуры N_k с заданным числом аминокислотных остатков k описывается выражением [10]:

$$N_k = p - (k - 1), \quad (2)$$

а максимальное суммарное число всех возможных перекрывающихся пептидных фрагментов белка (также включая одинаковые) для всех k , т.е. от $k = 2$ (дипептиды) до $k = p - 1$, может быть получено, используя следующее выражение:

$$N_k = \sum_2^{k=p-1} = \frac{p(p+1)}{2} - 1. \quad (3)$$

Приведенные формулы характеризуют верхний предел рассчитываемых величин. В частности, как следует из формулы 2, белковая последовательность, состоящая из 70 ($p = 70$) аминокислотных остатков (как N-концевая часть α -цепи гемоглобина быка на рисунке 7), содержит 69 перекрывающихся дипептидных фрагментов ($k = 2$), 68 трипептидных ($k = 3$) и т.д., а согласно формуле (3), в этой последовательности суммарное число всех возможных перекрывающихся фрагментов ($k = 2, 3, \dots, p-1$) составит величину 2484. Следовательно, число возможных фрагментов у всего этого белка (фрагментом), состоящего из 142 аминокислотных остатков, равно 10152. Таким образом, даже с учетом небольшого числа повторяющихся фрагментов (как дипептидный фрагмент AA, рисунок 7) число еще неисследованных фрагментов только этого белка существенно больше, чем исследованных.

Таким образом, в результате осуществления собственных процессов метаболизма и внешнего воздействия в организме формируется постоянно меняющийся пул эндогенных и экзогенных веществ пептидной природы – регуляторных олигопептидных молекул (континуум [59]), включенный в работу всех регуляторных систем. Существенным в этом процессе является то, что фрагменты могут обладать иными функциями по сравнению с молекулами, из которых они образовались.

Заключение.

Фрагментом одного белка – это только часть совокупности фрагментов всех белков клетки, живого организма, совокупности всех живых организмов. Все они являются составными частями глобального фрагментома, который усилиями многочисленных исследователей формируется в белково-пептидных базах данных. В свою очередь, этот белково-пептидный фрагментом представляется составной частью глобального фрагментома всех химических веществ всех живых организмов, который и является объектом биохимических исследований.

4. Zamyatnin A.A., Belozerskaya T.A. Multiple diversity of mitochondrial cytochrome b amino acid sequences of the same length in animals. *Front. Mol. Biosci.*, 2020, p. e00102, DOI: 10.3389/fmolb.2020.00102.
5. Davis S.N., Torres C.R., Musser G.M., Proffitt J.V., Crouch N.M.A., Lundelius E.L., Lamanna M.C., Clarke J.A. New mammalian and avian records from the late eocene La Meseta and Submeseta formations of Seymour Island, Antarctica. *PeerJ*, 2020, vol. 8, p. e8268, DOI: 10.7717/peerj.8268.
6. Helgen K.M., Pinto C.M., Kays R., Helgen L.E., Tsuchiya M.T., Quinn A., Wilson D.E., Maldonado J.E. Taxonomic revision of the olingos (Bassaricyon), with description of a new species, the olinguito. *ZooKeys*, 2013, vol. 324, pp. 1-83, DOI: 10.3897/zookeys.324.5827.
7. Wilkins M. Proteomics data mining. *Expert Review of Proteomics*, 2009, vol. 6, no. 6, pp. 599-603. DOI: 10.1586/epr.09.81.
8. Clynen E., Baggerman G., Veelaert D., Cerstiaens A., Van der Horst D., Harthoorn L., Derua R., Waelkens E., De Loof A., Schoofs L. Peptidomics of the pars intercerebralis-corpora cardiaca complex of the migratory locust, *Locusta migratoria*. *Eur. J. Biochem.* 2001, vol. 268, no. 7, pp. 1929-1939, DOI: 10.1046/j.1432-1327.2001.02067.x.
9. Zamyatnin A.A. Fragmentomics of oligopeptides and proteins. *4th International Peptide Symposium in conjunction with 7th Australian Peptide Symposium and 2nd Asia-Pacific International Peptide Symposium*. Cairns, Australia, Conference Program and Abstracts, 2007, p. 130.
10. Замятнин А.А. Фрагментомика белков и природных олигопептидов. *Биофизика*, 2008, т. 53, № 5, с. 725-733. [Zamyatnin A.A. Fragmentomics of proteins and natural oligopeptides. *Biophysics*, 2008, vol. 53, no. 5, pp. 329-335. (In Russ.)].
11. Zamyatnin A.A. Fragmentomics of natural peptide structures. *Biochemistry*, 2009, vol. 74, no. 13, pp. 1575-1585, DOI: 10.1134/s0006297909130100.
12. Tuppy H. The amino acid sequence in oxytocin. *Biochim. Biophys. Acta*, 1953, vol. 11, no. 3, pp. 449-450.
13. du Vigneaud V., Ressler C., Trippett S. The sequence of amino acids in oxytocin, with a proposal for the structure of oxytocin. *J. Biol. Chem.*, 1953, vol. 205, no. 2, pp. 949-957.
14. du Vigneaud V., Lowler H.C., Popenoe E.A. Enzymatic cleavage of glycylamide from vasopressin and a proposed structure for this pressor-antidiuretic hormone of the posterior pituitary. *J. Am. Chem. Soc.*, 1953, vol. 75, no. 19, pp. 4880-4881.
15. Sanger F., Thompson E.O.P. The amino-acid sequence in the glycyl chain of insulin. I. The identification of lower peptides from partial hydrolysates. *Biochem J.*, 1953, vol. 53, no. 3, pp. 353-366, DOI: 10.1042/bj0530353.
16. Sanger F., Thompson E.O.P. The amino-acid sequence in the glycyl chain of insulin. II. The identification of lower peptides from partial hydrolysates. *Biochem J.*, 1953, vol. 53, no. 3, pp. 366-374, DOI: 10.1042/bj0530366.
17. Holowachuk E.W., Stoltenberg J.K., Reed R.G., Peters T.Jr. Submitted (AUG-1991) to the EMBL/GenBank/ DDBJ databases.
18. Brown J.R. Structure of bovine serum albumin. *Fed. Proc.*, 1975, vol. 34, pp. 591-591.
19. Comb M., Seeburg P.H., Adelman J., Eiden L., Herbert E. Primary structure of the human Met- and Leu-enkephalin precursor and its mRNA. *Nature*, 1982, vol. 295, no. 5851, pp. 663-666, DOI: 10.1038/295663a0.
20. Taussig R., Scheller R.H. The *Aplysia* FMRFamide gene encodes sequences related to mammalian brain peptides. *DNA*, 1986, vol. 5, no. 6, pp. 453-461. DOI: 10.1089/dna.1.1986.5.453.
21. Иванов В.Т., Карелин А.А., Михалева, И.И., Васильковский Б.В., Свиряев В.И., Назимов И.В. Выделение, структура и свойства новых эндогенных пептидов. *Биоорг. Химия*, т. 18, № 10-11, с. 1271-1311. [Ivanov V.T., Karelin A.A., Mikhaleva I.I., Vaskovsky B.V., Sviryaev V.I., Nazimov I.V. Isolation, structure and properties of novel endogenous peptides. *Bioorganicheskaya Khimiya*, vol. 18, no. 10-11, pp. 1271-311. (In Russ.)]
22. Yang N., Anapindi K.D.B., Rubakhin S.S., Wei P., Yu Q., Li L., Kenny P.J., Sweedler J.V. Neuropeptidomics of the rat habenular nuclei. *J. Proteome Res.*, 2018, vol. 17, no. 4, pp. 1463-1473, DOI: 10.1021/acs.jproteome.7b00811.
23. Palashoff M.H. Determining the specificity of pepsin for proteolytic digestion, Chemistry Master's Theses, 2008, Paper 1, URL: <http://hdl.handle.net/2047/d10016636>.
24. Замятнин А.А., Воронина О.Л. Фрагменты пищевых белков – регуляторные олигопептиды. *Биохимия*, 2012, т. 77, № 5, с. 622-632. [Zamyatnin A.A., Voronina O.L. Food protein fragments are regulatory oligopeptides. *Biochemistry (Moscow)*, 2012, vol. 77, no. 5, pp. 502-510. (In Russ.)]
25. Fogaca A.C., da Silva P.I.Jr, Miranda M.T., Bianchi A.G., Miranda A., Ribolla P.E., Daffre S. Antimicrobial activity of a bovine hemoglobin fragment in the tick *Boophilus microplus*. *J. Biol. Chem.*, 1999, vol. 274, no. 36, pp. 25330-25334, DOI: 10.1074/jbc.274.36.25330.
26. Zamyatnin A. Structural-functional diversity of the natural oligopeptides. *Progr. Biophys. Mol. Biol.*, 2018, vol. 133, pp. 1-8, DOI: 10.1016/j.pbiomolbio.2017.09.024.
27. Gulevitch V.S., Amiradzhibi S. Über das Carnosin, eine neue organische Base des Fleischextrakt. *Deutsch. Chem. Ges.*, 1900, no. 33, pp. 1902-1903.
28. Замятнин А.А. Место карнозина среди физиологически активных веществ пептидной природы. *Биохимия*, 1992, т. 57, № 9, с. 1296-1301. [Zamyatnin A.A. Place of carnosine among natural active peptide substances. *Biokhimiya*, 1992, vol. 57, no 9, pp. 1296-1301. (In Russ.)]
29. Baumann L., Ingwaldsen T. Concerning histidine and carnosine. The synthesis of carnosine. *J. Biol. Chem.* 1918, vol. 35, no. 2, pp. 263-276.

30. Замятнин А.А. Специализированный банк данных EROP-Moscow о свойствах природных регуляторных олигопептидов. *Нейрохимия*, 1990, т. 9, № 1, с. 71-82. [Zamyatnin A.A. Specialized database EROP-Moscow for properties of natural regulatory oligopeptides. *Neirokhikia*, 1990, vol. 9, no. 1, pp. 71-82. (In Russ.)]
31. Zamyatnin, A.A. EROP-Moscow: specialized data bank for endogenous regulatory oligopeptides. *Protein Seq. Data Anal*, 1991, vol. 4, no. 1, pp. 49-52.
32. Замятнин А.А. Классификация эндогенных регуляторных олигопептидов по первичной структуре. *Докл. АН СССР*, 1990, т. 311, № 5, с. 1259-1265. [Zamyatnin A.A. Structural classification of endogenous regulatory oligopeptides. *Doklady Akademii Nauk SSSR*, 1990, vol. 311, no. 5, pp. 1259-1265. (In Russ.)]
33. Zamyatnin, A.A. Structural classification of endogenous regulatory oligopeptides. *Protein Seq. Data Anal*, 1991, vol. 4, no. 1, pp. 53-56.
34. Замятнин А.А., Борчиков А.С., Владимиров М.Г., Воронина О.Л. Клиент-серверная база данных EROP-Moscow о природных олигопептидах с Интернет-доступом. *Нейрохимия*, 2005, т. 22, № 1, с. 17-32. [Zamyatnin A.A., Borchikov A.S., Vladimirov M.G., Voronina O.L. Client-server database EROP-Moscow for endogenous oligopeptides with Internet access. *Neirokhikia*, 2005, no. 22, no. 1, pp. 17-32. (In Russ.)]
35. Zamyatnin A.A., Borchikov A.S., Vladimirov M.G., Voronina O.L. The EROP-Moscow oligopeptide database. *Nucleic Acids Res.*, 2006, vol. 34, Database Issue, pp. D261-D266. DOI: 10.1093/nar/gkj008.
36. Nielsen-Marsh C.M., Ostrom P.H., Gandhi H., Shapiro B., Cooper A., Hauschka P.V., Collins M.J. Sequence preservation of osteocalcin protein and mitochondrial DNA in bison bones older than 55 ka. *Geology*, 2002, vol. 30, no. 12, pp. 1099-1102. DOI: 10.1130/0091-7613(2002)030<1099:SPOOPA>2.0.CO;2.
37. Егоров Н.С., Силаев А.Б., Катруха Г.С., Орлова Т.А. *Антибиотики-полипептиды*. М.: МГУ, 1987, 264 с. [Egorov N.S., Silaev A.B., Katrukha G.S., Orlova T.A. *Antibiotics-polypeptides*. М.: MGU, 1987, 264 p. (In Russ.)]
38. Skeggs L.T.Jr., Lentz K.E., Kahn J.R., Shmway N.P., Woods K.R. The Amino Acid Sequence of Hypertensin II. *J. Exp. Med.*, vol. 104, no. 2, pp. 193-197. DOI: 10.1084/jem.104.2.193.
39. Tsai B.S., Peach M.J., Khoshla M.C., Bumpus F.M. Synthesis and Evaluation of (Des-Asp1)angiotensin I as a Precursor for (Des-Asp1)angiotensin II ("Angiotensin III"). *J. Med. Chem.*, 1975, vol. 18, no. 12, pp. 1180-1183, DOI: 10.1021/jm00246a002.
40. Ашмарин И.П., Обухова М.Ф. Содержание регуляторных пептидов в коре головного мозга и их центральная активность. *Журн. Высш. Нервн. Деят.*, 1985, т. 35, № 2, с. 211-221. [Ashmarin I.P., Obukhova M.F. Concentration of regulatory peptides in the cerebral cortex and their central activity. *Biokhimiya*, vol. 35, no. 2, pp. 211-221. (In Russ.)]
41. Kizer J.S., Busby W.H. Jr, Cottle C., Youngblood W.W. Glycine-directed peptide amidation: Presence in rat brain of two enzymes that convert p-Glu-His-Pro-Gly-OH Into p-Glu-His-Pro-NH₂ (Thyrotropin-Releasing Hormone). *Proc. Natl. Acad. Sci. USA*, 1984, vol. 81, no. 10, pp. 3228-3232, DOI: 10.1073/pnas.81.10.3228.
42. Bateman R.C., Youngblood W.W., Busby W.H., Kizer J.S. Nonenzymatic peptide alpha-amidation. implications for a novel enzyme mechanism. *J. Biol. Chem.*, 1985, vol. 260, no. 16, pp. 9088-9091.
43. Dimarchi R.D., Tam J.P., Kent S.B.H., Merrifield R.B. Weak acid-catalyzed pyrrolidone carboxylic acid formation from glutamine during solid phase peptide synthesis. Minimization by rapid coupling. *Int. J. Pept. Prot. Res.*, 1982, vol. 19, no. 1, pp. 88-93. DOI: 10.1111/j.1399-3011.1982.tb03027.x.
44. Abracham G.A., Podell D.N. Pyroglutamic acid. Non-metabolic formation, function in proteins and peptides, and characteristics of the enzymes effecting its removal. *Mol. Cell. Biochem.*, 1984, vol. 38, no. 1, pp. 181-190. DOI: 10.1007/BF00235695.
45. Karle I.L. In: *The Peptides: Analysis, Synthesis, Biology* (Gross, E., and Meienhofer, J., eds), Acad. Press, N. Y., pp. 1-54.
46. Schulz C.P. Illuminating folding intermediates. *Nature Struct. Biol.*, 2000, vol. 7, no. 1, pp. 7-10. DOI: 10.1038/71197.
47. Camerman A., Mastropaolo D., Karle I., Karle J., Camerman N. Crystal structure of leucine-enkephalin. *Nature*, 1983, vol. 306, no. 5942, pp. 447-450. DOI: 10.1038/306447a0.
48. Higashijima T., Kobayashi J., Nagai U., Miyazawa T. Nuclear-magnetic-resonance study on met-enkephalin and met-enkephalinamide. Molecular association and conformation. *Eur. J. Biochem.*, 1979, vol. 97, no. 1, pp. 43-57. DOI: 10.1111/j.1432-1033.1979.tb13084.x.
49. Sabesan M.N. Do peptide hormones have the structural features of epinephrine at their active sites? *Fed. Proc.*, 1980, vol. 39, pp. 1946-1946.
50. Sabesan M.N., Harper, E.T. Are aromatic residues essential at the "active sites" of peptide hormones? *J. Theor. Boil.*, 83, no. 3, pp. 457-467. DOI: 10.1016/0022-5193(80)90052-1.
51. Zamyatnin A.A. Amino acid, peptide, and protein volume in solution. *Ann. Rev. Biophys. Bioeng.*, 1984, vol. 13, pp. 145-165. DOI: 10.1146/annurev.bb.13.060184.001045.
52. Замятнин А.А. Физико-химические особенности эндогенных регуляторных олигопептидов. *Биофизика*, 1990, т. 35, № 4, с. 555-559. [Zamyatnin A.A. Physico-chemical features of endogenous regulatory oligopeptides. *Biofizika*, vol. 35, no 4, pp. 555-559. (In Russ.)]
53. Zamyatnin A.A. Specificity of the amino acid residue content in endogenous regulatory oligopeptides. *Protein Seq. Data Anal*, 1991, vol. 4, no. 1, pp. 57-60.
54. Замятнин А.А., Воронина О.Л. Общие физико-химические и физиологические особенности эндогенных антибактериальных олигопептидов. *Успехи Биол. Химии*, 1998, т. 38, с. 165-197. [Zamyatnin A.A., Voronina O.L.

Common physico-chemical and physiological features of endogenous antibacterial oligopeptides. *Uspekhi Biologicheskoi Khimii*, 1998, vol. 38, pp. 165-197. (In Russ.)]

55. Замятнин А.А. Физико-химические и биологические особенности эндогенных олигопептидных токсинов. *Нейрохимия*, 1996, т. 13, № 4, с. 243-259. [Zamyatnin A.A. Physico-chemical and biological features of endogenous oligopeptide toxins. *Neirokhimia*, 1996, vol. 13, no. 4, pp. 243-259. (In Russ.)]

56. Замятнин А.А., Воронина О.Л. Общие химические особенности эндогенных нейропептидов. *Нейрохимия*, 1997, т. 14, № 3, с. 263-272. [Zamyatnin A.A. Common chemical features of endogenous neuropeptides. *Neirokhimia*, 1997, vol. 14, no. 3, pp. 263-272. (In Russ.)]

57. Замятнин А.А., Воронина О.Л. Общие физико-химические особенности эндогенных олигопептидных гормонов – либерины и статины. *Биофизика*, 1998, т. 43, № 6, с. 438-446. [Zamyatnin A.A. Common physicochemical features of endogenous oligopeptide hormones: liberins and statins. *Biophysics*, 1998, vol. 43, no. 3, pp. 413-421. (In Russ.)]

58. Воронина О.Л., Замятнин А.А. Эндогенные олигопептиды и иммунная регуляция. *Нейрохимия*, 2001, т. 18, № 3, с. 163-181. [Voronina O.L., Zamyatnin A.A. Endogenous oligopeptides and immune regulation. *Neirokhimia*, 2001, vol. 18, no. 3, pp. 163-181. (In Russ.)]

59. Ашмарин И.П., Обухова М.Ф. Регуляторные олигопептиды. Функционально-непрерывная совокупность. *Биохимия*, 1986, т. 51, № 4, с. 531-544. [Ashmarin I.P., Obukhova M.F. Regulatory oligopeptides. A functional continuum. *Biokhimia*, 1986, vol. 51, no. 4, pp. 531-544. (In Russ.)]

BIOLOGICALLY ACTIVE FRAGMENTS OF PROTEIN STRUCTURES OF GLOBAL PROTEOME

Zamyatnin A.A., Belozerskaya T.A.

A.N. Bach Institute of Biochemistry,

Research Centre of Biotechnology Russian Academy of Sciences

Leninsky prosp., 33, Moscow 119071, Russia; e-mail: aaz@inbi.ras.ru

Abstract. Giant biodiversity is based on the gigantic molecular diversity of linear combinations of 20 amino acid residues in peptide structures. Their combination is a global proteome of the Earth. The amino acid sequences of ~200 million natural peptide structures have been decoded. It is believed that each of them is dedicated a specific role in a living organism. Despite such a significant number of known peptide molecules, a good part of the global proteome has not yet been deciphered. However, familiar information is already used for various analyses and generalizations. A large part of such analyses is devoted to the study of fragments representing the global proteome. Theoretically, the possible number of different peptide structures composed of only 50 amino acid residues is characterized by a value of $\sim 10^{34}$. A huge number of such structure fragments served as the basis for the emergence of the concept of fragmentomics – a direction which studied the structure and functions of a set of protein fragments. These data have been collected and studied in the EROP-Moscow database for 30 years, and their number is more than 26,000 now. A large number of regulators of the nervous, endocrine and immune systems, as well as a variety of antimicrobial oligopeptides, enzyme inhibitors, and many others possessing specific physical and chemical features were among them. The analysis of the EROP-Moscow database information allowed us to conclude that regulatory oligopeptides arise from three sources. These are oligopeptide regulators cleaved from specialized endogenous precursors using special enzymes, ordinary endogenous protein polypeptides (for example, hemoglobin, albumin, etc.), cleaved by proteolytic enzymes to small fragments, as well as exogenous proteins or their fragments entering the body from the outside as a result of a meal or bite (e.g., insects). They form in the organism a constantly changing pool of endogenous and exogenous substances of peptide nature (continuum), which is included in the functioning of all regulatory systems.

Key words: proteome, protein structures.