

УДК 577.322

ВСТРЕЧАЕМОСТЬ МОТИВОВ ИЗ ШЕСТИ АМИНОКИСЛОТНЫХ ОСТАТКОВ В ТРЕХ ЭУКАРИОТИЧЕСКИХ ПРОТЕОМАХ

© 2012 г. М. Ю. Лобанов, Н. С. Богатырева, О. В. Галзитская*

Институт белка Российской академии наук, Пушкино, Московская обл., 142290

Поступила в редакцию 18.04.2011 г.

Принята к печати 14.07.2011 г.

В настоящее время известно восемнадцать наследственных неврологических заболеваний, связанных с мутациями многократной вставки одного аминокислотного остатка в последовательность белка. В связи с этим изучение функционального назначения простых аминокислотных мотивов в последовательности белка представляет важную биологическую задачу. В данной работе исследовано, как часто встречаются участки из шести одинаковых аминокислотных остатков, а также простые мотивы длиной в шесть остатков, составленные из двух аминокислот, расположенных в любом порядке. Исследование проводили на трех эукариотических протеомах хорошо исследованных организмов: *Homo sapiens*, *Drosophila melanogaster*, *Caenorhabditis elegans*. Оказалось, что многие простые мотивы встречаются очень часто: данные по встречаемости представлены на сайте: http://antares.protres.ru/motifs_six_residues.html. Можно предположить, что такие мотивы ответственны за наличие общих функций у негомологичных, неродственных белков из разных организмов.

Ключевые слова: эукариотический протеом, неструктурированные остатки, мотивы с низкой сложностью, гомоповторы.

OCCURRENCE OF MOTIFS WITH SIX AMINO ACID RESIDUES IN THREE EUKARYOTIC PROTEOMES, by M. Yu. Lobanov, N. S. Bogatyreva, O. V. Galzitiskaya* (Institute of Protein Research, Russian Academy of Sciences, Moscow Region, Pushchino, 142290; *e-mail: ogalzit@vega.protres.ru). Now it is known that 18 neurological inherited diseases connected with mutations of multiple insertion of one amino acid residue in protein sequence. Therefore, studying the functional role of such simple motifs is an important task in biology. In this work we have investigated how often homorepeats, i.e. runs of a single amino acid residue, of 6 amino acid residues long as well as simple motifs consisting from two amino acid residues of 6 residues long situated in any position occur in three eukaryotic well studied proteomes: *Homo sapiens*, *Drosophila melanogaster*, *Caenorhabditis elegans*. It turns out that many simple motifs occur very often. The occurrence for each motif can be found at our site: http://antares.protres.ru/motifs_six_residues.html. One can suggest that such short similar motifs are responsible for the common functions for nongomologous, unrelated proteins from different organisms.

Keywords: eukaryotic proteome, disordered residues, motifs with low complexity, homorepeats.

Предсказание структуры и функции белков – одно из главных направлений в структурной геномике. Особый интерес представляет поиск специальных участков и мотивов белковой цепи, которые выполняют важную функциональную роль в белковой молекуле. На сегодня для этих целей разработаны специализированные базы данных, такие как PROSITE, InterPro и Pfam [1–3].

Участки белков, не имеющие фиксированной пространственной структуры, называют неструктурированными. Особый интерес представляет предсказание таких неструктурированных участков белковой цепи (не имеющих в нативном со-

стоянии белка фиксированной пространственной структуры, т.е. не разрешенные методом рентгеноструктурного анализа) [4–7]. Неструктурированные участки часто играют важную функциональную роль, например их присутствие в последовательности белка приводят к сложностям при экспрессии, очищении и кристаллизации белка. Как правило, протяженные участки белков, состоящие из одной многократно повторяющейся аминокислоты, не имеют фиксированной пространственной структуры, т.е. они неструктурированы. В протеоме человека встречаются белки, содержащие простые аминокислотные мотивы, в том числе и состоящие из одной многократно повторенной аминокислоты.

* Эл. почта: ogalzit@vega.protres.ru

Проанализировав пространственные структуры в банке белковых структур PDB [<http://www.rcsb.org>], мы построили первую и пока единственную библиотеку неструктурированных мотивов. На настоящий момент библиотека содержит 109 шаблонов [8]. Интересно отметить, что туда попали такие простые мотивы, как QQQQQQ, PRRRRR, NNNNNN. При этом оказалось, что мотивы с низкой сложностью встречаются в эукариотических протеомах (включая протеом человека) чаще, чем другие мотивы белковых последовательностей. Например, такие мотивы как NEGHHN, QPELAP, MDSPDL, GSHMSK, MASMTG, NIEGRHM ни разу не встретились в трех тестовых эукариотических (*Homo sapiens*, *Drosophila melanogaster*, *Caenorhabditis elegans*) и трех бактериальных протеомах (*Escherichia coli*, *Lactococcus lactis*, *Mycobacterium tuberculosis*) [8]. Следует отметить, что полигистидиновые фрагменты часто попадают в банк белковых структур (часто на N- и на C-концах белковой цепи), но это искусственно введенные фрагменты в аминокислотную последовательность. Шести подряд идущих гистидиновых остатков достаточно для эффективной очистки рекомбинантного белка на колонке, содержащей никель [9]. Установлено, что повторы из гистидиновых остатков можно часто встретить в протеоме человека, но там назначение данных фрагментов имеет функциональный характер [8]. Предполагают, что если один и тот же аминокислотный мотив не имеет определенной пространственной структуры в большинстве белковых структур, то такой мотив с большой вероятностью будет неструктурирован и в другом белке, с неизвестной пока пространственной структурой. При этом предсказано, что большинство гомоповторов неструктурированы [10].

Следует отметить, что чем короче аминокислотный гомоповтор, тем более вероятна случайность его появления в последовательности белка, следовательно меньше его значимость и влияние его на структуру и функцию белка. Минимальная длина повтора, которая принимается во внимание в геномном анализе — это 5–7 аминокислотных остатков [11–13]. Это как раз та начальная длина, при которой гомоповтор может оказывать свое влияние как на функцию, так и на структуру белковой молекулы. Показано, что белки, содержащие аланиновый повтор, состоящий из 10 и более остатков, способны к агрегации [14].

При изучении амилоидообразования методами структурной биологии, применяемыми при исследовании сворачивания белков, установлено, что увеличение числа гомоповторов приводит к увеличению агрегации полимера и, как следствие, к образованию амилоидных фибрилл [15]. Восемнадцать известных неврологических заболеваний связаны с генетическими нарушениями,

удлиняющими простые однобуквенные мотивы [16]. Так наличие длинных, превышающих порог допустимого значения, полиглутаминовых и полиаланиновых повторов в белках связывают с такими болезнями, как болезнь Гентингтона [17] и дистрофия мышц глазного яблока, соответственно [14, 18]. Болезнь Гентингтона вызывается многократной вставкой кодона CAG, кодирующего глутамин, в гене *IT-15*. В гене дикога типа у разных людей присутствует разное количество CAG-повторов, однако, если число повторов превышает 36, развивается болезнь. Как правило, участки с простыми повторами неструктурированы и способны образовывать амилоидные фибриллярные структуры [15]. Участки поли-Q или поли-P могут активировать транскрипцию, когда связываются с ДНК-связывающим доменом фактора GAL4 [20].

Все вышеуказанные данные подчеркивают важность изучения функциональной роли аминокислотных гомоповторов. Можно предположить, что похожие мотивы белков могут иметь общую функцию у негомологичных, неродственных белков. В данной работе мы исследовали, как часто встречаются гомоповторы длиной в 6 аминокислотных остатков для всех 20 аминокислотных остатков, а также простые мотивы длиной в 6 остатков, составленные из двух аминокислот, расположенных в любом порядке в трех эукариотических протеомах.

Оказалось, что в рассматриваемых в этой работе трех эукариотических протеомах (*Homo sapiens*, *Drosophila melanogaster*, *Caenorhabditis elegans*), наиболее часто представленные гомоповторы специфичны для каждого протеома. Во всех трех протеомах часто (более ста раз) встречаются гомоповторы из шести E, P, A, S, G и Q аминокислот.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Создание базы белков. В данной работе мы рассмотрели три эукариотических протеома: протеом человека (*Homo sapiens*), включающий 51 778 белковых последовательностей, протеом плодовой мухи (*Drosophila melanogaster*), включающей 15101 белковых последовательностей, и протеом круглого червя (*Caenorhabditis elegans*), включающий 23 817 белковых последовательностей. Данные протеомы доступны на сайте <ftp://ftp.ebi.ac.uk/pub/databases/SPproteomes/uniprot/proteomes/>.

Создание банка мотивов длиной в 6 аминокислотных остатков. Нами составлена библиотека всех возможных мотивов из двух аминокислот, при условии, что обе аминокислоты могут встречаться в любой позиции, в любом соотношении и длина такого мотива составляет 6 аминокислот. Таких мотивов получилось $11\,780 = (2^6 - 2)C_{20}^2$ (исключаем два гомоповтора для каждой пары

Таблица 1. Число белков хотя бы с одним гомоповтором длиной в 6 аминокислотных остатков для 20 аминокислот в трех эукариотических протеомах и частоты встречаемости для 20 аминокислот в этих протеомах

Гомоповтор	<i>H. sapiens</i> (51778 белков)	<i>D. melanogaster</i> (15101 белков)	<i>C. elegans</i> (23817 белков)	АК	<i>H. sapiens</i> (20894630 а.о.)	<i>D. melanogaster</i> (8251519 а.о.)	<i>C. elegans</i> (10278458 а.о.)
EEEEEE	840	103	128	Е	0.0702	0.0644	0.0659
PPPPPP	692	320	251	Р	0.0639	0.0553	0.0496
AAAAAA	653	588	124	А	0.0700	0.0734	0.0637
LLLLLL	560	48	43	L	0.0989	0.0908	0.0857
SSSSSS	488	418	258	S	0.0833	0.0837	0.0814
GGGGGG	435	440	220	G	0.0664	0.0609	0.0537
QQQQQQ	325	939	254	Q	0.0476	0.0521	0.0420
KKKKKK	272	38	59	K	0.0567	0.0558	0.0628
NNNNNN	110	149	57	N	0.0264	0.0264	0.0232
DDDDDD	98	97	92	D	0.0472	0.0517	0.0534
RRRRRR	61	40	9	R	0.0573	0.0555	0.0522
TTTTTT	59	241	243	T	0.0538	0.0573	0.0591
FFFFFF	26	1	2	F	0.0362	0.0352	0.0467
CCCCCC	13	4	1	C	0.0230	0.0188	0.0202
VVVVVV	6	3	3	V	0.0599	0.0592	0.0621
MMMMMM	3	2	0	M	0.0218	0.0233	0.0264
IIIIII	3	0	4	I	0.0430	0.0493	0.0608
NNNNNN	3	208	66	N	0.0353	0.0471	0.0488
YYYYYY	2	0	0	Y	0.0265	0.0297	0.0314
WWWWWW	0	0	0	W	0.0128	0.0101	0.0109

Примечание. АК – аминокислота.

аминокислот). Полученные мотивы можно разбить на три группы. В первую группу мы поместили мотивы, в которых одна из двух аминокислот встречается всего один раз и находится либо в первом, либо в шестом положениях, т.е. с краю. Во вторую группу вошли мотивы, где вторая аминокислота встречается также один раз, только находится уже внутри мотива. И в третью группу вошли все остальные мотивы, где каждая из двух аминокислот встречается как минимум два раза и в любом порядке.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Гомоповторы в протеомах человека, мухи и червя

Нами подсчитано число белков, в которых хотя бы один раз встретился данный гомоповтор из шести аминокислотных остатков в трех эукариотических протеомах: человека, мухи и червя (*Homo sapiens*, *Drosophila melanogaster*, *Caenorhabditis elegans*) (см. табл. 1). Ранее показано, что наиболее часто встречающиеся гомоповторы в эукариотических протеомах (в порядке возрастания) – это поли-Q, поли-N, поли-A, поли-S, поли-G и поли-P [19].

В случае протеома человека часто можно обнаружить следующие гомоповторы из 6 аминокис-

лотных остатков (в порядке убывания и более 100 раз) EEEEEEE, PPPPPP, AAAAAA, LLLLLL, SSSSSS, GGGGGG, QQQQQQ, KKKKKK, НННННН. Поли-Е встречается чаще, чем поли-D, а поли-K – чем поли-R. Следует отметить, что очень редко встречается гомоповтор NNNNNN – всего три раза.

У плодовой мушки набор таких гомоповторов из 6 аминокислотных остатков уже немного другой. Наиболее часто можно встретить следующие гомоповторы (в порядке убывания и более 100 раз): QQQQQQ, AAAAAA, GGGGGG, SSSSSS, PPPPPP, TTTTTT, NNNNNN, НННННН, EEEEEEE.

Для червя наиболее часто обнаруживали следующие гомоповторы (в порядке убывания и более 100 раз): SSSSSS, QQQQQQ, PPPPPP, TTTTTT, GGGGGG, EEEEEEE, AAAAAA. Во всех трех протеомах часто (более ста раз) встречали гомоповторы из шести Е, Р, А, S, G и Q аминокислот.

Интересно, что гомоповтор из шести лейциновых остатков (LLLLLL) можно обнаружить в протеоме человека в 11 и 13 раз чаще, чем в протеомах мухи и червя соответственно. При этом гомоповтор из шести треониновых остатков (TTTTTT) находили в 4 раза чаще в протеомах мухи и червя, чем в протеоме человека. А самое большое различие наблюдали для гомоповтора из шести аспара-

Рис. 1. Попарное сравнение протеомов человека, мухи и червя. Отображено число белков, в которых встречаются гомоповторы для 20 аминокислотных остатков. По осям отложены число белков с гомоповторами в первом и втором из пары сравниваемых протеомов: *a* – человек–муха, *б* – человек–червь, *в* – муха–червь. Количества белков с мотивом из шести одинаковых остатков представлены буквой, обозначающей соответствующий остаток, если таких белков более 100 хотя бы в одном из протеомов, и точкой – в противном случае.

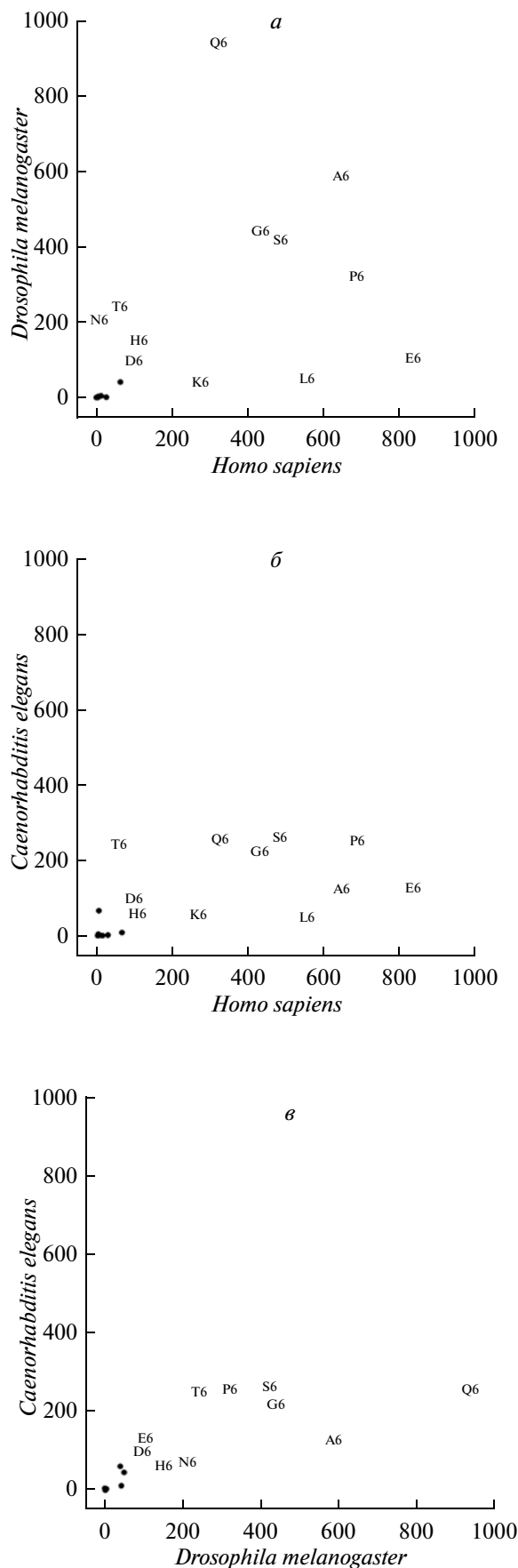
гинов (NNNNNN). Он встречается чаще в 69 раз в протеоме мухи, чем в протеоме человека. В целом, повторы из гидрофобных аминокислот встречаются реже, чем из полярных аминокислотных остатков, при том, что частоты встречаемости самих аминокислот в протеомах незначительно отличаются от ожидаемых (см. табл. 1).

Для каждого рассматриваемого в статье протеома мы подсчитали число белков, в которых встречаются гомоповторы из шести идущих подряд остатков данной аминокислоты для 20 аминокислотных остатков (см. табл. 1). Коэффициент корреляции между числами белков для человека и мухи составляет 0.47, человека и червя 0.59, а мухи и червя 0.79 (см. рис. 1). Коэффициент корреляции выше для пары мухи и червя, чем для пар, в которую вошел протеом человека. Хотя сами коэффициенты корреляции достаточно велики, они все же существенно меньше корреляций между частотами аминокислот. Так коэффициенты корреляций между встречаемостью для 20 аминокислот для рассматриваемых пар составили: 0.97, 0.92 и 0.96 соответственно.

Встречаемость мотивов длиной шесть аминокислотных остатков, составленных из двух аминокислот

Нами составлена библиотека всех возможных мотивов из двух аминокислот, при условии, что две аминокислоты могут встречаться в любой позиции, в любом соотношении, и длина такого мотива составляет 6 аминокислот. Число таких мотивов получилось равным 11780. Полученные мотивы разбили на три группы (см. раздел “Экспериментальная часть”). Число мотивов первой, второй и третьей группы равно 760 (6), 1520 (13) и 9500 (81%) соответственно.

Мы оценили встречаемость этих мотивов в трех эукариотических протеомах (*Homo sapiens*, *Drosophila melanogaster*, *Caenorhabditis elegans*). На сайте (http://antares.protres.ru/motifs_six_residues.html) размещены эти данные. На рис. 2*a–в* представлена статистика простых мотивов по трем группам для трех протеомов. Среди мотивов из первой группы в лидерах находятся: мотив EEEEEED встречается 422 раза в человеческом протеоме, DEEEEE – 370, затем идет LPPPPP – 327, APPPPP – 264 и



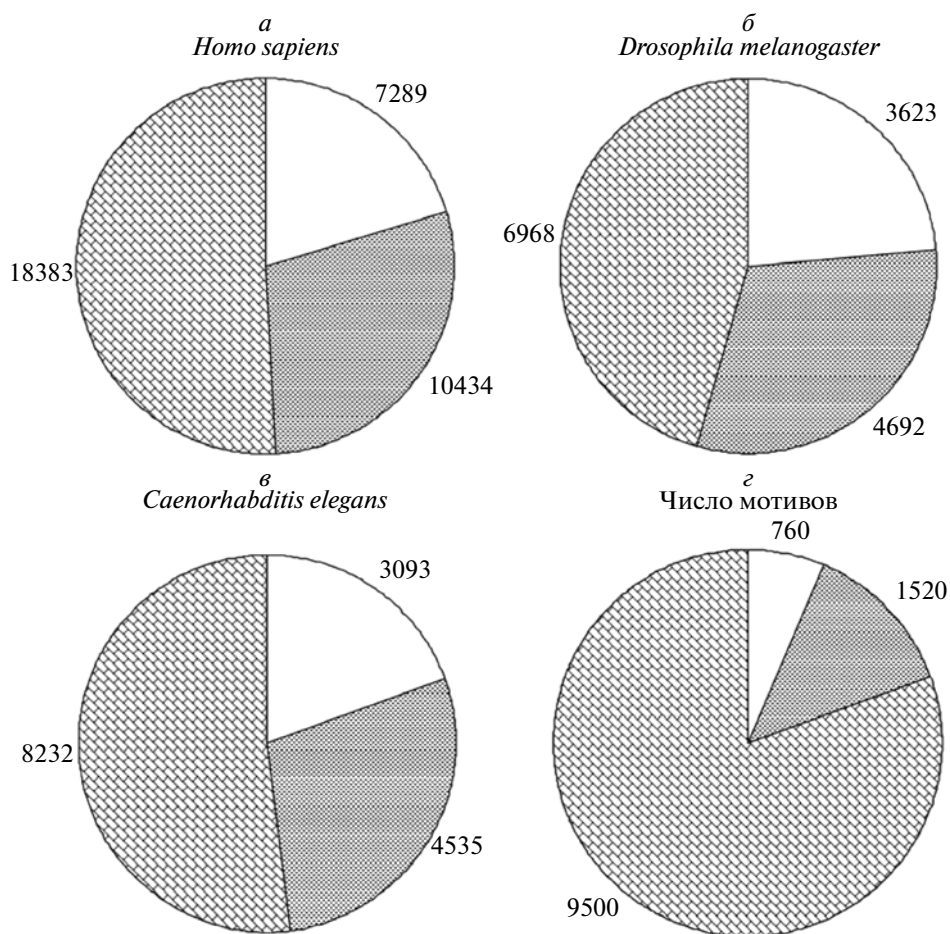


Рис. 2. Статистика встречаемости мотивов из шести остатков двух аминокислот по трем группам для трех протеомов: а – *H. sapiens*; б – *D. melanogaster*; в – *C. elegans*; г – возможное число мотивов из двух аминокислот для каждой группы. Первая группа – белый цвет, вторая – мелкая клетка, третья – диагональный кирпич (см. Создание банка мотивов длиной в 6 аминокислотных остатков).

PLLLLL – 251, а PPPPPL – 216 раз. Среди мотивов, где одна аминокислота обнаружена один раз и только внутри мотива, на первое место выходит мотив EEEEEDE – 288, EDEEEEE – 279, EEDEEEE – 248, EEEDEEE – 250, PLPPPP – 239, PPPPLP – 207 раз. Среди лидеров, где появляются уже две аминокислоты на первое место по встречаемости выходят такие мотивы как SGSGSG – 135, EEEEDD – 157, GPPGPP – 162, RSRRSRS – 153 раза. Среди мотивов, которые встречаются более 100 раз в протеоме человека, часто встречаются такие аминокислоты, как E, D, P, A, L, G, S, Q, V, T, K, R. Стоит отметить следующие редкие мотивы, встречающиеся всего лишь в двух белках: из первой группы – FFFFFN, FFFFFP, CННННН, MVVVVV, IННННН, WKKKKK, IIIIF; из второй группы – ПМПП, RRFRRR, YLYYYY, NNCNNN, ННТННН, DDQDDD; из третьей – CCCRRR, MMMGGG, TTTDDD, FFSFFS, FPFPPF, VVRVVR, QKQKQK, DDHDDH. На рис. 3 показаны часто встречающиеся мотивы.

В случае протеома червя список мотивов, которые встречаются чаще всего, – уже немного другой. Среди мотивов из первой группы лидируют: STTTTT – 166, APPPPP – 139, TSSSSS – 130, PTTTTT – 120, DEEEEE – 107, PSSSSS – 95, TTTTTS – 91, SSSSST – 89, QQQQQP – 96 раз. Среди второй группы: STSSSS – 126, TSTTTT – 112, SSTSSS – 96, TTTSTT – 90, EDEEEE – 85 раз. Среди третьей группы: GPPGPP – 135, PPGPPG – 136, PGPPGP – 92, RSRRSRS – 80, SRSRSR – 62, DEEEEE – 61 раз. На рис. 3 показаны часто встречающиеся шаблоны.

В часто встречающихся мотивах протеома мухи лидируют такие аминокислоты, как глутамин, аланин, глицин. Среди мотивов из первой группы лидируют: QQQQQH – 470, HQQQQQ – 410, LQQQQQ – 359, QQQQQL – 359, QQQQQP – 276, PQQQQQ – 260, QQQQQA – 221, AQQQQQ – 219, SAAAAA – 224 раза. Среди второй группы: QQQQHQ – 319, QQQHQQ – 297, QQHQQQ – 290, QHQQQQ – 284, QQQQLQ – 243, QLQQQQ –

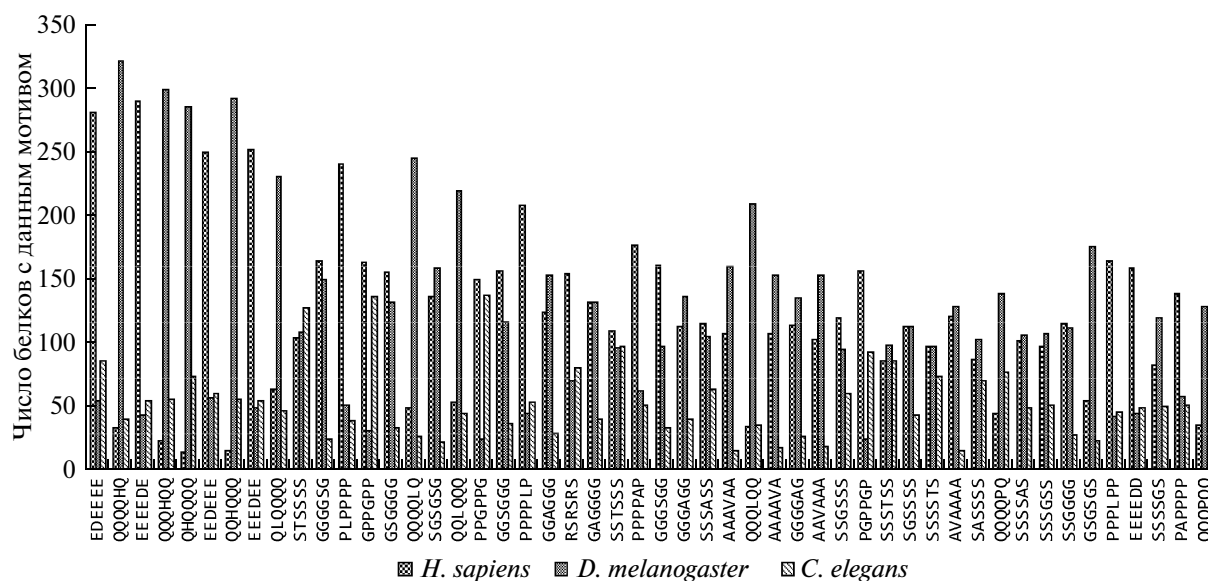


Рис. 3. 50 Мотивов (50) из второй и третьей группы, часто встречающиеся в трех протеомах.

229, QQLQQQ – 218 раз. Среди третьей группы: GSGSGS – 174, SSGSGS – 157, HHQQQQ – 163, QQQQHH – 166, SSGGGG – 110, GSGSGS – 105 раз. На рис. 3 показаны часто встречающиеся шаблоны. Из 11780 мотивов 3601 мотивов не встретились ни разу ни в одном из трех протеомов (см. http://antares.protres.ru/motifs_six_residues.html).

Интересно, что для всех трех протеомов число белков, в которых обнаружен хотя бы один раз мотив из первой группы, составляет около 20%. В случае второй и третьей группы – 30 и 50% соответственно (см. рис. 2). При этом следует отметить, что среднее число белков в протеоме человека, в которых встретился мотив из первой группы хотя бы один раз равно 27, для второй и третьей группы – 16 и 6 соответственно. Данное соотношение постоянно и для двух других рассмотренных протеомов. То есть белки, содержащие мотивы из третьей группы, встречаются чаще белков, содержащих мотивы из двух других групп, только потому, что в третьей группе намного больше мотивов (в 12.5 раз больше по сравнению с первой).

Можно отметить, что мотивы первой группы самые простые. Это гомоповтор с примыкающей аминокислотой. Мотивы второй группы – это гомоповтор с вкраплением другой аминокислоты. В то же время, третью группу трудно получить из гомоповтора. Чаще всего встречаются мотивы максимально похожие на гомоповторы, т.е. мотивы первой группы, немного реже мотивы из второй группы и совсем редко мотивы, непохожие на гомоповторы.

Лидирующие мотивы в каждом протеоме свои, и видно, что часто встречаются те аминокислоты, которые лидируют среди повторов из 6 аминокислот.

В табл. 2 представлены общие мотивы, которые часто встречаются в трех протеомах, как минимум в 50 белках. Можно предположить, что такие короткие похожие мотивы ответственны за наличие общих функций у негомологичных, неродственных белков из разных организмов.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных наук (11-04-00763), Российской академии наук (программы “Молекулярная и клеточная биология”)

Таблица 2. Общие простые мотивы, составленные из двух аминокислот, часто встречающиеся в трех протеомах

Мотив	<i>H. sapiens</i>	<i>D. melanogaster</i>	<i>C. elegans</i>
EDEEEE	279	53	85
EEDEEE	248	56	59
STSSSS	103	107	126
RSRSRS	153	69	80
SSTSSS	108	95	96
PPPPAP	175	62	51
SSSASS	114	104	63
SSGSSS	118	94	59
SSSTSS	85	97	85
SSSSTS	96	96	72
SASSSS	86	101	69
SSSGSS	96	106	50
PAPPPP	138	57	50
SSASSS	82	86	64
SSSSDS	100	62	55

(01200959110), “Фундаментальные науки – медицине”), Федерального агентства по науке и инновациям (№ 16.512.11.2204).

СПИСОК ЛИТЕРАТУРЫ

1. Finn R.D., Mistry J., Tate J., Coggill P., Heger A., Pollington J.E., Gavin O.L., Gunasekaran P., Ceric G., Forslund K., Holm L., Sonnhammer E.L., Eddy S.R., Bateman A. 2010. The Pfam protein families database. *Nucl. Acids Res.* **38** (Database issue): D211–222.
2. Hunter S., Arweiler R., Attwood T.K., Bairoch A., Bateman A., et al. 2009. InterPro: the integrative protein signature database. *Nucl. Acids Res.* **37** (Database issue): D211–215.
3. Sigrist C.J., Cerutti L., de Castro E., Langendijk-Genevaux P.S., Bulliard V., et al. 2010. PROSITE, a protein domain database for functional characterization and annotation. *Nucl. Acids Res.* **38** (Database issue): D161–166.
4. Галзитская О.В., Гарбузинский С.А., Лобанов М.Ю. 2006. Предсказание нативно-развернутых участков белковой цепи. *Молекуляр. биология.* **40**(2), 341–348.
5. Dosztanyi Z., Csizmok V., Tompa P., Simon I. 2005. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, **21**, 3433–3434.
6. Lobanov M.Y., Galzitskaya O.V. 2011. The Ising model for prediction of disordered residues from protein sequence alone. *Phys. Biol.* **8**, 035004.
7. Сердюк И.Н., Галзитская О.В. 2007. Неструктурированные области в элонгационных факторах EF1A трех доменов живого мира. *Молекуляр. биология.* **41**(6), 1042–1055.
8. Lobanov M.Y., Furletova E.I., Bogatyreva N.S., Roytberg M.A., Galzitskaya O.V. 2010. Library of disordered patterns in 3D protein structures. *PLoS Comput. Biol.* **6**(10), e1000958.
9. Hengen P. 1995. Purification of His-Tag fusion proteins from *Escherichia coli*. *Trends Biochem. Sci.* **20**(7), 285–286.
10. Huntley M.A., Golding G.B. 2002. Simple sequences are rare in the Protein Data Bank. *Proteins.* **48**(1), 134–140.
11. Karlin S. 1995. Statistical significance of sequence patterns in proteins. *Curr. Opin. Struct. Biol.* **5**(3), 360–371.
12. Katti M.V., Ranjekar P.K., Gupta V.S. 2001. Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Mol. Biol. Evol.* **18** (7), 1161–1167.
13. Karlin S., Brocchieri L., Bergman A., Mrazek J. 2002. Amino acid runs in eukaryotic proteomes and disease associations. *Proc. Natl. Acad. Sci. USA.* **99** (1), 333–338.
14. Fan X., Dion P., Laganieri J., Brais B., Rouleau G.A. 2001. Oligomerization of polyalanine expanded PABPN1 facilitates nuclear protein aggregation that is associated with cell death. *Hum. Mol. Genet.* **10** (21), 2341–2351.
15. Fraser R.D.B., MacRae T.P. 1973. Conformation in Fibrous Proteins. N.-Y. and London: Acad. Press.
16. Jorda J., Xue B., Uversky V.N., Kajava A.V. 2010. Protein tandem repeats – the more perfect, the less structured. *FEBS J.* **277**, 2673–2682.
17. Lagier-Tourenne C., Cleveland D.W. 2010. Neurodegeneration: An expansion in ALS genetics. *Nature.* **466** (7310), 1052–1053.
18. Sackewitz M., Scheidt H.A., Lodderstedt G., Schierhorn A., Schwarz E., Huster D. 2008. Structural and dynamical characterization of fibrils from a disease-associated alanine expansion domain using proteolysis and solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **130** (23), 7172–7173.
19. Jorda J., Kajava A.V. 2010. Protein homorepeats sequences, structures, evolution, and functions. *Adv. Protein Chem. Struct. Biol.* **79**, 59–88.
20. Gerber H.P., Seipel K., Georgiev O., Höffler M., Hug M., Rusconi S., Schaffner W. 1994. Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science.* **263** (5148), 808–811.