

УДК: 577.122

©Коллектив авторов

ОПРЕДЕЛЕНИЕ “АМИНОКИСЛОТНЫХ КОНФЛИКТОВ” И АМИНОКИСЛОТНЫХ ЗАМЕН В ПЕРВИЧНЫХ СТРУКТУРАХ 41 БЕЛКА ЧЕЛОВЕКА ПРОТЕОМНЫМИ ТЕХНОЛОГИЯМИ

М.А. Ковалева¹, Л.И. Ковалев^{1}, Л.С. Еремина¹, А.А. Макаров¹, М.В. Буракова¹,
И.Ю. Торопыгин², М.В. Серебрякова³, С.С. Шишкин¹, А.И. Арчаков²*

¹Институт биохимии им. А.Н. Баха Российской академии наук, 119071, Москва,
Ленинский проспект, 33, стр. 2; тел.: (095) 9525886; факс: (095) 9542732;
эл. почта: kovalyov@inbi.ras.ru

²ГУ НИИ биомедицинской химии РАМН им. В.Н.Ореховича, 119121, Москва,
Погодинская ул., д. 10

³НИИ физико-химической медицины МЗ РФ, 119992, Москва

При протеомных исследованиях некоторых тканей и органов человека (скелетные мышцы, миомеры, кора мозга, простата), а также культивируемых миобластов среди идентифицированных белков удалось выявить 41 белок, в аминокислотных последовательностях которых показано присутствие определенных (вариантов) аминокислотных остатков в зарегистрированных “конфликтных” участках. Среди 93 охарактеризованных “аминокислотных конфликтов” семь случаев оказались проявлениями белкового полиморфизма, обусловленного соответствующей единичной аминокислотной заменой. Кроме того, при протеомном анализе белков простаты были идентифицированы две изоформы простат-специфического антигена, которые образуются за счет альтернативного сплайсинга. Таким образом, показано, что с помощью протеомного анализа удастся характеризовать разные виды биохимического полиморфизма у многих белков человека.

Ключевые слова: протеомика, белки человека, аминокислотные конфликты.

ВВЕДЕНИЕ. Исследования белков человека в начавшейся постгеномной эре развития биохимии характеризуются выраженной структурной направленностью и тесно ассоциированы с изучением кодирующих их генов [1, 2]. Как результат, все большее внимание привлекает анализ феномена белкового полиморфизма и его значения для установления молекулярных основ онтогенеза в норме, а также для различных видов патологии. Важные роли при этом отводят сведениям о единичных аминокислотных заменах (single amino acid substitution, SAAS), которые часто получают по результатам обнаружения несинонимичных одонуклеотидных полиморфизмов (single nucleotide polymorphisms, SNP's) в соответствующих генах или кДНК [2, 3]. Кроме того, в международных компьютерных базах данных (в частности, в одной из наиболее детальных белковых баз данных - Swiss-Prot/TrEMBL, Интернет адрес: <http://ca.expasy.org>) в описаниях свойств белков приводятся многочисленные указания на различные

* - адресат для переписки

“единичные конфликты при определении аминокислотной последовательности” (single amino acid conflict, SAAC). Обычно SAAC в первичной структуре белка представляют как альтернативное присутствие той или иной аминокислоты (т.е. в виде двух вариантов) в какой-либо позиции. В подавляющем большинстве таких случаев сведения о первичных структурах белков являлись результатами перерасчетов данных секвенирования соответствующих нуклеиновых кислот, полученными двумя группами авторов (см. например [4, 5]). По-видимому, разрешению указанных “конфликтных определений” может способствовать широкое привлечение к исследованиям SAAS протеомных технологий, включающих прецизионное фракционирование белков и их детальную идентификацию методами масс-спектрометрии [6, 7].

В данной работе представлены результаты, подтверждающие присутствие одной из аминокислот (а иногда даже существование обоих вариантов, т.е. SAAS) в 93 случаях “конфликтных определений в аминокислотной последовательности” у 41 белка человека, которые были получены благодаря сочетанному применению двумерного электрофореза по О’Фарреллу, MALDI-TOF масс-спектрометрии и компьютерному анализу.

МЕТОДИКА. Материалами для изучения белков скелетных мышц, миометрия и коры мозга являлись аутопсийные образцы соответствующих тканей, полученных в Бюро судебно-медицинской экспертизы Департамента здравоохранения г. Москвы от лиц, погибших в результате несчастного случая (со сроками аутолиза 3-6 часов). Исследования белков простаты проводились на образцах тканей этого органа, взятых при операциях по поводу гиперплазии или рака простаты в НИИ урологии МЗ РФ. Белки скелетномышечных миобластов человека получали из клеточных культур, образцы которых были предоставлены к.б.н. Т.Б. Крохиной; для выращивания необходимой клеточной массы использовали методику, описанную ранее [8].

Экстракцию белков из исследованных тканей и клеток человека, последующее фракционирование белков модифицированным методом двумерного электрофореза в ПААГ по О’Фарреллу и окрашивание белковых фракций кумасси голубым R-250, а также азотнокислым серебром проводили, как описано ранее [7, 9]. Далее отдельные белковые фракции вырезали из гелевых пластин (из двумерных электрофореграмм) и полученные фрагменты гелей после измельчения обрабатывали трипсином. Гидролиз анализируемых белков трипсином и экстракцию полученных пептидов выполняли согласно протоколам [10] с некоторыми модификациями [11].

Выбор и использование методик масс-спектрометрического анализа белков (MALDI-TOF MS и MS/ MS) были проведены ранее [7]. Обычно 0,5 мкл образца, содержащего триптические пептиды, смешивали на масс-спектрометрической подложке с таким же объемом раствора 10 мг/мл 2,5-дигидроксibenзойной кислоты (“Sigma”, США) в 20%-ном ацетонитриле, содержавшем 0,1% трифторуксусной кислоты, и высушивали на воздухе. Масс-спектры получали на MALDI-времяпролетном-масс-спектрометре Ultraflex (“Bruker Daltonics”, Германия) с УФ-лазером (336 нм) в режиме положительных ионов в диапазоне масс 500-8000 Да и калибровали их, используя известные данные о распределении масс-спектров продуктов аутолиза трипсина.

Идентификацию исследуемых белков проводили по полученным масс-спектрам с помощью программы Mascot (“Matrixscience”, США) и по базам данных Национального центра биотехнологической информации США (NCBI), принимая точность определения массы ионов равной 0,01% и допуская возможность модификации цистеинов акриламидом и окисления метионинов.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ. Одним из наглядных примеров, демонстрирующих широкие возможности и фактически универсальный характер протеомных технологий, включающих двумерный электрофорез по О’Фарреллу, масс-спектрометрию, компьютерные методы анализа изображений и др., можно

считать результаты системных исследований белков мышечных и других тканей человека [7, 10, 12]. Обычно в таких исследованиях параллельно анализируются сотни белковых продуктов генной экспрессии, и проведение их идентификации масс-спектрометрией позволяет получить информацию о протяженных участках аминокислотных последовательностей изучаемых белков. В нашей работе к настоящему времени удалось охарактеризовать структурные свойства более 300 белков, которые были получены после фракционирования двумерным электрофорезом по О'Фарреллу экстрактов поперечно-полосатых мышц, а также некоторых других тканей и культивируемых клеток человека. Соответствующие результаты фракционирования были опубликованы ранее [7, 10, 12].

Полученные таким образом отдельные белковые фракции вырезали из гелевых пластин, обрабатывали трипсином и триптические пептиды анализировали MALDI-TOF масс-спектрометрией. Далее выполняли поиск по компьютерным базам данных с помощью программы Mascot, что обеспечило идентификацию определенных пептидов по соответствующим масс-спектрам (в ходе поиска учитывались только значимые результаты, при $p < 0,05$). В целом, среди всех идентифицированных белков у ряда белков ($n=41$), перечисленных в таблице 1, оказалось 93 участка, в которых по материалам компьютерной базы данных Swiss-Prot/TrEMBL присутствовали так называемые “конфликты” по позициям отдельных аминокислотных остатков. При этом подавляющее большинство “конфликтов” затрагивали единственный аминокислотный остаток (single amino acid conflict, SAAC), но иногда они представляли собой и некоторые другие разночтения в первичных структурах. Обобщенные результаты проведенного анализа представлены в таблице 2.

Как видно из приведенных данных, полученные результаты в подавляющем большинстве случаев свидетельствовали о присутствии пептидов, содержащих аминокислотный остаток, соответствующий одному из “конфликтных” значений. При этом в нескольких случаях один и тот же SAAC удавалось зарегистрировать при анализе структуры белка, выявленного из разных источников, но идентифицированного как один и тот же продукт генной экспрессии. Например, в препаратах белка из скелетных мышц, идентифицированного как енолаза 1 (ENO1), был обнаружен пептид с m/z 1143.60 и с SAAC 187 E / G (здесь и далее подчеркнут выявленный вариант). Тот же самый SAAC 187 E / G присутствовал в пептиде с близким значением m/z (1143.64), который был получен при анализе енолазы 1 из коры мозга. (В таблице 2 все ENO1 пептиды из коры мозга выделены серым цветом). Таким же способом удалось подтвердить идентификацию одного из двух вариантов для другого известного SAAC (252 E / S) в структуре енолазы 1, благодаря определению общего пептида с экспериментальными значениями m/z 1556.67 и 1556.81 в препаратах енолазы 1, полученных из мышц и из мозга.

Сходная ситуация оказалась при анализе белка, идентифицированного как триозофосфат изомераза 1 (ТФИ 1). В частности, один из вариантов SAAC 167 R / N присутствовал в пептидах с экспериментальными значениями m/z 1602.71 и 1602.89, которые были обнаружены в образцах скелетных мышц и в простате (в таблице 2 все ТФИ 1 пептиды из простаты выделены курсивом). Аналогичные данные были получены еще для четырех известных SAAC в структуре ТФИ 1: 20-21 QS / KN; 27 G / S; 30-31 NA / QG и 43-44 AP / IG.

В качестве других доводов, подтверждающих идентификацию SAAC, рассматривали случаи, когда один и тот же SAAC удавалось определить в двух или более пептидах, выявленных при анализе одного и того же белка из одного и того же источника. В частности, SAAC 346 D / N присутствовал в пептидах с m/z 2407.21 и с m/z 2535.31, полученных при анализе ацетоацетил КоА тиолазы (АКАТ), а также SAAC 129 Q / E регистрировался в пептидах с m/z 1515.70 и с m/z 1531.71 миоглобина (МГ) (табл. 2). Многократное определение одного и того же SAAC, очевидно, позволяет разрешить “конфликт” в пользу того варианта, который удалось идентифицировать протеомными технологиями.

Таблица 1. Общие характеристики белков человека, выбранных для протеомного анализа конфликтов, зарегистрированных при определении аминокислотных последовательностей (SAAC и др.) в Swiss-Prot/TrEMBL (по версиям с обновлениями ноябрь-декабрь 2007). Пояснения в тексте.

№	Название белка (Условное обозначение)	Источники (биомагистраты)*	Номер в Swiss-Prot/TrEMBL	Символ гена	% совпадения **	Электрофоретические свойства, Мм/рI»***
1	2	3	4	5	6	7
1	Δ3,5-Δ2,4-дигеноил-коэнзим А изомеразы (ДКАИ)	CM	Q13011	ECH1	82	32,2/6,57
2	Аконитаза 2, (АКО2)	CM, П	Q99798	ACO2	48, 46	89/7,02
3	Пируваткиназа, мышечная изоформа (ПКМ)	CM	P14618	PKM2	31	59,6/7,10
4	Колаген 1, изоформа А 2 (Кол1A2)	CM, М6, П	P08123	COL1A2	14, 18, 13	116/7,10
5	Легкая цепь ферритина, (ЛЦФ)	CM, П	P02792	FTL	41,78	450.0****/5.07
6	Енолаза 1, (ЕНО1)	CM, М6, П КМ	P06733	ENO1	22, 59, 86, 56	47,3/6,58
7	Креатинфосфокиназа, мышечная изоформа, (КФК)	CM	P06732	CKM	26	41,2/7,09
8	Десмин, (Дес)	CM	P17661	DES	74	52,0/5,47
9	Миозенин 1, (МЗН)	CM	Q9NP98	MYOZ1	62	34,0/7,31
10	Легкая цепь миозина 1, изоформа (ЛЦМ1)	CM	P05976	MYL1	66	23,9/5,12
11	Легкая цепь миозина 2, изоформа скелетномышечная, (ЛЦМ2)	CM	Q96A32	MYLPF	94	20,4/5,06
12	Белок теплового шока 27 кДа, (БТПШ 27)	CM	P04792	HSBP1	64	27,0/6,20
13	Белок сходный с белком теплового шока 20 кДа, (sim БТПШ 20)	CM	O14558	HSPB6	42	17,5/6,26
14	Белок теплового шока 70 кДа изоформа 5, (БТПШ 70)	CM, М6	P11021	HSPA5	37, 50	72,0/5,10
15	Альфа В кристаллин, (Кр αВ)	CM	P02511	CRYAB	66	20,8/7,09

Таблица 1. Продолжение.

1	2	3	4	5	6	7
16	Шаперонин 60, (Ш60)	СМ	P10809	CH60	49	62,5/5,51
17	Триозофосфат изомеразы 1, (ТФИ 1)	СМ, М6, П	P60174	TRP1	58, 31, 76	25,8/6,68
18	Фосфоглицеромутаза мышечная изоформа, (ФГМ)	СМ	P15259	PGAM2	38	28,8/7,48
19	3-гидрокси-изобутирил-коэнзим А гидролаза, (3-ИКИ)	СМ	Q6NVY1	HIBCH	25	40,0/6,95
20	Миоглобин, (МГ)	СМ	P02144	MB	48	17,5/7,26
21	Ацетоацетил КоА тиаза, (АКАТ)	СМ	P24752	ACAT1	55	40,0/6,72
22	Тропомозин 4, (ТМ4)	М6	P67936	TRM4	28	30,0/4,62
23	Ядерный фосфопротеин В 23, (ЯФП В23)	М6	P06748	NPM1	52	17,5/8,37
24	Аннексин V, (Ан V)	М6	P08758	ANXA5	70	31,0/4,75
25	Актин бета, (Ак-β)	М6	P60709	ACTB	54	40,0/4,75
26	Белок теплового шока 47, колаген связывающий белок 2, (БТПШ47)	М6	P50454	SERPINH1	37, 32	39,5/8,72 45,5/8,35
27	Виментин, (Вим)	М6	P08670	VIM	69	56,0/5,30
28	Дисульфид изомеразы (ДСИ А3)	М6, П	P30101	PDIA3	62, 46	52,0/6,15
29	Транскеллин, (Тр)	М6, П	Q01995	TAGLN	72	22,0/7,15
30	Альдолаза В, (Ал-В)	М	P05062	ALDOB	27	39,4/8,30
31	Белок, содержащий SH3 домен и подобный богатому глутаминовой кислотой белку (Б-SH3D)	М	O75368	SH3BGRL	72	12,8/5,22
32	Фосфатидилэтаноламин связывающий белок 1, (ФЭСБ1)	П	P30086	PEBP1	77	21,7/7,10
33	Пептидил-пролил цис-транс изомеразы А, (ППИ-А)	П	P62937	PRPA	50	18,5/7,30

Таблица 1. Продолжение.

1	2	3	4	5	6	7
34	Пероксиредоксин 1, (ПР)	П	Q06830	PRDX1	47	22,2/6,95
35	Супероксиддисмутаза (Mn), митохондриальная, (СОД)	П	P04179	SOD2	42	22,5/7,10
36	Простаг-специфический антиген, основная изоформа, после процессинга - 237 а.о. (ПСА-ой);	П	P07288	KLK3	62, 21	35,5/7,30*****
37	Миозин, регуляторная легкая цепь 2 (ПЦМ глад)	П	P24844	MYL9	41	19,8/4,80
38	Белок S100 - A9, (S100 - A9)	П	P06702	S100A9	39	14,2/6,25
39	Глицеральдегид 3 фосфат дегидрогеназа, (Г3ФД)	КМ	P04406	GAPDH	67	36,0/7,50
40	Синапсин 1, (Син 1)	КМ	P17600	SYN1	40	70,0/9,50
41	Глиальный фибриллярный кислый белок, (ГФКБ)	КМ	P14136	GFAP	60	50,0/5,40

Примечание: * - Аббревиатуры, обозначающие использованные биоматериалы: СМ - скелетные мышцы; Мб - миоциты; М - миоциты; П - ткани простаты; КМ - кора мозга; ** - совпадения масс выявленных триптических пептидов с последовательностью белка; *** - "pI" - экспериментально определенные значения электрофоретической подвижности, полученные при неравновесном изоэлектрофокусировании, которое выполнялось на первом этапе двумерного электрофоретического анализа; эти значения, как правило, не совпадают с точными значениями pI; **** - Экспериментально определенная молекулярная масса ферритинового комплекса, состоящего из легких и тяжелых цепей; ***** - Экспериментально определенные показатели для особого комплекса из ПСА и специального пептида - ингибитора

Таблица 2. Результаты протеомного анализа 93 "аминокислотных конфликтов", зарегистрированных при определении аминокислотных последовательностей у 41 белка человека и представленных в базе данных Swiss-Prot/TrEMBL. Пояснения в тексте.

№	Ссылка белка	Заголовок m/z идентифицирован- ных пептидов	Аминокислотные последовательности пептидов, соответствующих идентифицированным m/z (возможны). Жирным шрифтом выделены обнаруженные аконформативные аминокислотные остатки	Некоторые зарегистрированные SAAS в др. конформаты. Подчеркнуты а.о., обнаруженные при масс-спектрометрии
1	2	3	4	5
1	ДКАИ	1542.84	EVDVGLAADVGTLPQR (197 – 211)	210 Q / E
		1564.80 1580.81	KMMAD EALGSG LVSR (231 – 245) с учетом окисления одного (M) и KMMAD EALGSG LVSR (231 – 245) с учетом окисления двух (M)	230* R / H и 239 G / D
		1731.91	EVM DAALALLAAE ISSK (251 – 267)	259 A / P
		916.55	<u>SPVAVQ</u> STIK (268 – 276)	268 / 271 <u>SPVA</u> / TTVL
		1347.77	SVQATPENKELK (310 – 321)	313 A / P
		2648.1 2590.2	LTGSSAQEEASGV ALGEAPDH SYESLR (33-59) LTGSSAQEEASGV ALGEAPDH SYESLR (33-59)	41 E / A
2	AKO2	2351.10	VAMSHFERPNEYHNDLLEK (32 – 50) с учетом окисления (M)	35 S / T
		1953.01	AKDINQEVYNFLATAGAK (143 – 160)	159 A / D
		1219.64	LTGSLSGWSSPK (234 – 245)	242 S / T
		1685.82	FNPEIDYLTIGIDGKK (507 – 521)	517 T / K
3	ПКМ	2036.01	QKGA D FLVTEVENGSGSLGSK (187 – 206)	204 G / V
		2407.27	KGDVWVLTGWKPGSGFTINMR (505 / 526)	507 D / H
4	Кол1A2	1562.84	GETGPSGPGVPAGAVGPKR (979 – 996)	980 E / V
		1251.69	GIPGPGVGAAGATGAR (328 – 342)	333 V / P и 338 A / T
5	ЛЦФ	1874.87	ALFQDIKKPAE DE WVK (77 – 92)	87 E / Q и 89 E / W
		1209.63	MGDHLTNLH R (145 – 154) с учетом окисления (M)	154 R / A
		2315.18	LGGPEAGLGEYLFERLTKH D (155 – 175)	175 D / N
6	ЕНО1	1143.60	IGAENVYHNLK (184 – 193)	187 E / G
		1556.67	VWIGMDVAASEF R (240 – 253)	252 E / S
		1143.64	IGAENVYHNLK (184 – 193)	187 E / G
		1556.81	VWIGMDVAASEF R (240 – 253) с учетом окисления (M)	252 F / S
		2033.07	FTASAGIQVVGDDLTVTNPK (307 – 326)	310 S / I

Таблица 2. Продолжение.

1	2	3	4	5
7	КОК	2571.95	TDLNHNKGGDDLDPNVLSRR (108 – 130)	130 R / P
		759.35	DWPDAR (210 – 215)	210 D – H и 215 R – P
		1994.79	GTGGVDTAAGVSMFVSNADR (321 – 341)	324 G / A
8	Дес	867.44	SPDYMEK (119 – 125)	119 – 123 SPDYM → FANYI
9	МЭН	3468.50	GGSQAGGSGSAGQYGSDQQHHLGSGAGGTGGPAGQAGR (114 – 153)	153 R / K
10	ЛЦМ1	1527.66	IEFSKEQQDEFK (45 – 56)	45 I / M
11	ЛЦМ2	1776.83	NICYVITHGDAKDQE (155 – 169) с учетом модификации цистеина (C) акриламидом	169 E / D
12	БПШ 27	987.58	RVPFSLLR (5 – 12)	10 L / I
13	sim БПШ 20	1349.73	RASAPLPGLSAPGR (14 – 27)	20 P / L
14	БПШ 70	1964.99	KSQIFSTASDNQPTVTK (447 – 464)	447 K / N
15	Кр αВ	1822.07	TIPTREEKPAVTAAPK (158 – 174)	165 E / K
16	Ш	1344.68	TVIEQSWGSPK (61 – 72)	67 S / G
17	ТФМ 1	1414.68	QSLGELIGTLNAAK (20 – 33)	20 – 21 QS / KN; 27 G / S и 30 – 31 NA / QG
		2205.74	VPADTEWVCAPPTAYIDFAR (34 – 53)	43 – 44 AP / IG
		1602.71	VMLAYEPVWAIGTGK (161 – 175)	167 P / N
		1542.86	KQSLGELIGTLNAAK (19 – 33)	20-21 QS / KN и 27 G / S
		1414.79	QSLGELIGTLNAAK (20 – 33)	30-31 NA / QG
		2206.10	VPADTEWVCAPPTAYIDFAR (34 – 53)	43-44 AP / IG
		1602.89	VMLAYEPVWAIGTGK (161 – 175)	167 P / N
		3029.71	ELASQPDVDGFLVGGASLKPEFVDINAK (220 – 248)	244 I / L
18	ФТМ	1357.61	HGESTWNQENR (11 – 21)	14 S / T
		1059.56	HYGGLTGINK (91 – 100)	98 L / F
		826.41 и 630.41	HGEQVVK (107 – 113) и IWRR (114 – 117)	113 – 114 KI / RS

Таблица 2. Продолжение.

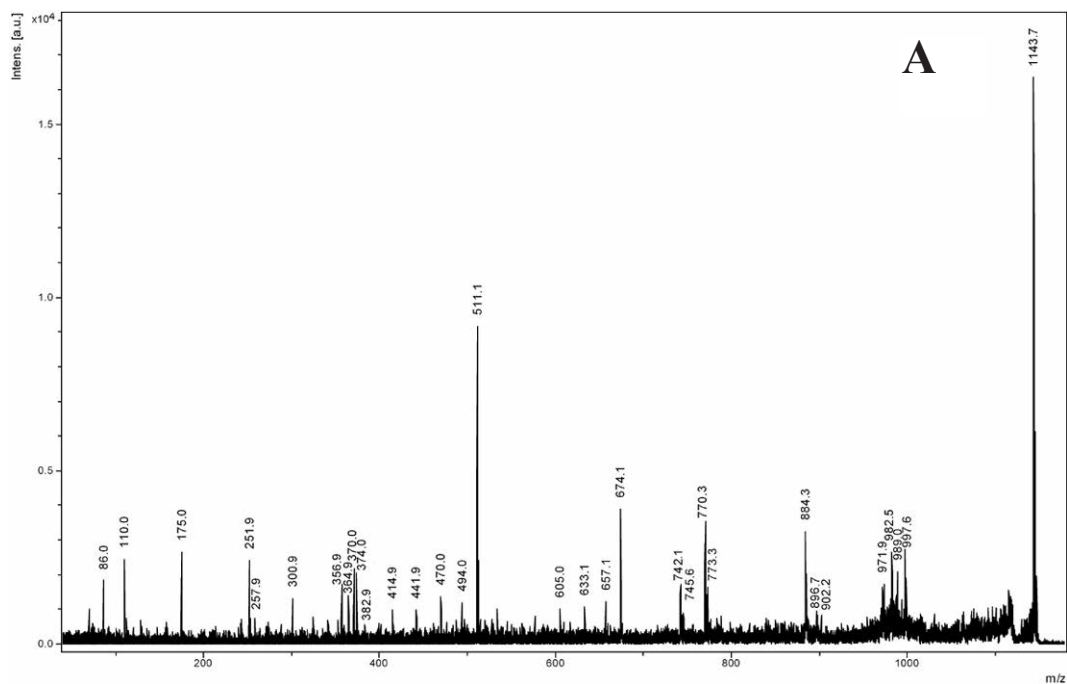
1	2	3	4	5
19	3-ГКГ	2312.92 1257.68	WKPADLKEVTEEDNMMFK (359 - 377) LGYFLALIGR (191 - 201)	369 E / K 184 - 198 IPRIQKLGYFLALT / FATPRKTVLLPCIN
20	МГ	1515.70 1531.71	HPGDFGADAQ GANNK (120 - 134) HPGDFGADAQ GANNK с учетом окисления (M)	129 Q / E
21	АКАТ	2407.21 2535.31 2225.12	EDIAMWVNEAFSLVLANIK с учетом окисления (M) (345 - 365) KEDIAMWVNEAFSLVLANIK с учетом окисления (M) (344 - 365) QGEYGLASICNGGGGASAMLIQK (404 - 426)	346 D / N 412 I / F
22	ТМ4	1243.70	IQLVHEHLDK (56 - 65)	59 V / F
23	ЯФП ВЗ3	745.46 910.48 783.42	VPLATLK (74 - 80) SKGQESFK (222 - 229) FINVYK (268 - 273)	80 K / E 219 - 221 TPR / SSS 271 Y / C
24	АН V	1835.94	MIDQEAIQDLWQWNR (278 - 291) с учетом окисления (M)	288 - 294 WQWRKSL - CLAVEEVSRLK - CMAVEEVSRLK - CVAVEEVSRLK 290 - 294 WRKSL - SLAQVSLRK - SLERVSRLK 135 S / L
25	Ак-β	2888.31 1106.62 1234.72 1954.15	QWTEEEYGSSLEDWWDTSGYQR (126 - 150) SEIDLFMIR (276 - 284) SEIDLFMIRK (276 - 285) VAREEPVLLTEAPLMPK (91 - 98)	279 I / T 97 P / A
26	БПШ47	1473.76 1359.64 1294.71 2000.08	SAGLAFSLYQAMAK (47 - 60) с учетом окисления (M) SYTVGVNPMHR (229 - 239) с учетом окисления 3 (M) LQIVEMPLANK (253 - 263) с учетом окисления (M) LSSLPLMPHMEPLER (264 - 280) с учетом окисления (M)	46 - 48 RSA / PST 235 M / T 255 I / L 270 L / F
27	Вмн	870.42 1323.63 1023.51 905.45 2186.96	FANYDK (114 - 120) BEAENPLQSEK (197 - 207) QOYESVANK (274 - 282) GTINESLER (335 - 342) EMEENFAVEAANYQDTIGR (346 - 364)	113 R / P* 197 E / G и 201 N / S 278 S / I 339 S / C 350 N / K

Таблица 2. Продолжение.

1	2	3	4	5
28	ДСМ АЗ	939.48	TVAYTEQK (219 – 226)	225 Q / P
		2877.37	FQENIFGICRNIPEDNKLIQK (235 – 258) с учетом окисления (M) и модификации цистеина (C) акриламидом	238 E / G
		1515.77	FLQDYFDGNLKR (352 – 363)	355 D / G и 358 D / G
		1368.67	SEPIRESNDGPK (367 – 379)	368 E / D
29	Тр	990.59	RLGFQVMLK (50 – 57)	49 R / P
		1516.84	LVNELYPDGSKPK (65 – 78)	72 D / E
		965.51	AAEDYGVK (100 – 108)	104 Y / C или 104 Y / S
		1227.63	HNVLQMGSMR (173 – 183) с учетом окисления (M)	179 M / V
30	Ал-В	1804.08	YPIREQAMATVTLHR (243 – 258)	250 A / D
31	Б-ШЕД	1734.91	IGFEKDIANNEER (32 – 46)	35 E / K
32	ФЭСБ1	3471.65	WSSGPLSLQEVDEQFPLNVTYAGAAVDLQK (7 – 38)	8 W / K
33	ПМ-А	1831.91	SYGKGFEDENLTK (76 – 90)	89 I / T
		1154.57	FEDEMILK (82 – 90)	
34	ПР 1	2420.21	HGEVCSAGWKPGSDTPKRVQK (169 – 190) с учетом модификации цистеина (C) акриламидом	189 Q / P
35	СОД	1738.82	HNAAUYNNLVTEK (54 – 68)	65 I / N
		2542.29	HNAAUYNNLVTEKYQALAK (54 – 75)	66 E / Q
		2637.22	FNAGGGHNNISIFWTNLSPIGGGPK (90 – 114)	112 E / Q
		1028.61	GELLEAKR (115 – 123)	123 R / L
36	ПСА-ом	3538.63	VMDLPTQEPALGTPICVAGSWGSPERPEELPK (114 – 145) с учетом модификации цистеина (C) акриламидом	165-168 FLP / HLLYDQM
		3554.52	VMDLPTQEPALGTPICVAGSWGSPERPEELPK (114 – 145) с учетом окисления (M) и с учетом модификации цистеина (C) акриламидом	
37	ЛМ-глад	1051.55	ELLTPMGDR (125 – 133) с учетом окисления (M)	125 E / K
38	5100-А9	1807.04	NIEPMPHNYGVK (11-25)	25 K / F
39	ГЭФД	870.50	VIRELDGK (220 – 227)	225 D / N
40	СМН 1	924.52	INGEDIK (135 – 142)	138 E / G
41	ГФКБ	1177.66	LADYQAEIR (112 – 121)	121 R / P
		2233.13	LQDETMLRLEAENMLAAYR (155 – 173)	158 E / G
		2348.14	LEAENMLAAYRQDEADLAK (163 – 183)	174 Q / QQ
		1208.62	EAASTQAEALR (320 – 330)	326 E / V

Примечание: * - Образование этих пептидов из-за субстратной специфичности трипсина возможно в результате трипсинолиза следующих белков: ДКАИ - только для аллеля 230R., а не для аллеля 230N; ЯФП В23 - только для аллеля 221R, а не для аллеля 221S

Кроме того, прямое подтверждение результата определения SAAC в некоторых белках было получено с помощью метода MALDI-TOF MS/MS. В качестве примера на рисунке 1 приведено расшифровка аминокислотной последовательности и выявление SAAC 187 E / G в пептиде с m/z 1143.60, полученном при анализе белка, идентифицированного как енолаза 1 (ENO1).



Mascot Search Results

Peptide View

MS/MS Fragmentation of IGAEVYHNLK

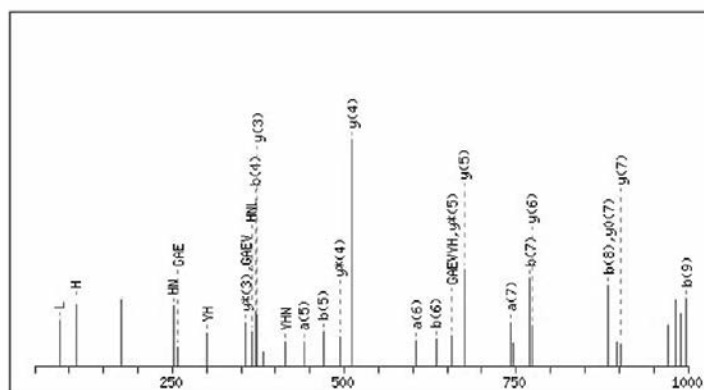
Found in [gi|693933](#), 2-phosphopyruvate-hydratase alpha-enolase, carbonate dehydratase [Homo sapiens]

Match to Query 1: 1142.692724 from(1143.700000,1+)

Data file DATA.TXT

Click mouse within plot area to zoom in by factor of two about that point

Or, Plot from to Da



Monoisotopic mass of neutral peptide Mr(calc): 1142.61

Fixed modifications: Propionamide (C)

Ions Score: 49 **Expect:** 0.0019

Matches (Bold Red): 29/112 fragment ions using 30 most intense peaks

B

#	Immon.	a	a*	a ⁰	b	b*	b ⁰	Seq.	y	y*	y ⁰	#
1	86.10	86.10			114.09			I				10
2	30.03	143.12			171.11			G	1030.53	1013.51	1012.52	9
3	44.05	214.16			242.15			A	973.51	956.48	955.50	8
4	102.05	343.20		325.19	371.19		353.18	E	902.47	885.45	884.46	7
5	72.08	442.27		424.26	470.26		452.25	V	773.43	756.40		6
6	136.08	605.33		587.32	633.32		615.31	Y	674.36	657.34		5
7	110.07	742.39		724.38	770.38		752.37	H	511.30	494.27		4
8	87.06	856.43	839.40	838.42	884.43	867.40	866.42	N	374.24	357.21		3
9	86.10	969.52	952.49	951.50	997.51	980.48	979.50	L	260.20	243.17		2
10	101.11							K	147.11	130.09		1

Рисунок 1.

Результаты MALDI-TOF MS/MS анализа пептида с m/z 1143.60, полученного при анализе белка, идентифицированного как енолаза 1 (ENO1).

A – Масс-спектры полученных фрагментов, на графике показаны регистрируемые ионы с соответствующими молекулярными массами (Mz);

Б – Результаты анализа масс-спектров фрагментов пептида с Mz 1143.60

В – Расчет аминокислотной последовательности пептида с Mz 1143.60

Как видно из таблицы 2, одним из наиболее полно охарактеризованных оказался продукт гена *ECN1* - Δ3,5-Δ2,4-диеноил-коэнзим А изомераза (ДКАИ), который ранее был идентифицирован среди белков поперечнополосатых мышц человека [7]. Определение по данным масс-спектрометрии строения шести участков в этом белке (табл. 2) позволило подтвердить наличие в его аминокислотной последовательности по одному из двух вариантов для четырех известных SAAS, а также пятого конфликтного участка, состоящего из четырех аминокислотных остатков. Относительно еще одного единичного “конфликта”, который был выявлен при протемном анализе [7] и еще ранее был отмечен при определении аминокислотной последовательности ДКАИ (по записям 11433007 и 16924265 в базе данных Protein/ NCBI), стало известно, что его причина - это одонуклеотидный полиморфизм. Таким образом, указанный SAAS в действительности является не конфликтным определением, а аминокислотной заменой, т.е. SAAS, которая обуславливает соответствующий белковый полиморфизм (41 E / Δ).

Далее, проведенный протеомный анализ и идентификация на двумерных электрофореграммах аконитазы 2 (АКО2) привели к обнаружению аллельных вариантов (ниже во всех случаях они подчеркнуты), относящихся к еще двум видам известных SAAS - 159 Δ / D и 242 S / T (табл. 2). Отдельные аллельные варианты известных SAAS были выявлены еще в двух типичных ферментах – в мышечной изоформе пируваткиназы (ПКМ - 204 G / V) и в альдолазе В (Ал-В - 250 Δ / D), а также в трех структурных белках - тропомиозине 4 (ТМ4 - 59 V / F), β-актине (Ак-β - 97 P / A), ядерном фосфопротеине, изоформе В 23 (ЯФП В23 - 80 K / E) и в одном из известных шапероноподобных белков - пептидил-пролил цис-транс изомеразе А (ППИ-А - 89 I / T).

При протеомном изучении белков простаты особенно интересные результаты были получены при идентификации и выявлении белкового полиморфизма простат-специфического антигена (ПСА). Так, на типичных двумерных электрофореграммах небольшая по количеству белковая фракция, которая по электрофоретическим свойствам соответствовала значениями Мм 21,8 кДа и рI 6,72 (рис. 2А), по масс-спектрометрическому анализу была определена как одна из изоформ ПСА, соответствующая записи 24370944 в базе данных NCBI. Эта изоформа отсутствует в базе данных Swiss-Prot/TrEMBL и по имеющимся сведениям является не основным продуктом экспрессии гена *KLK3*, а образуется в результате альтернативного сплайсинга (ПСА-ас).

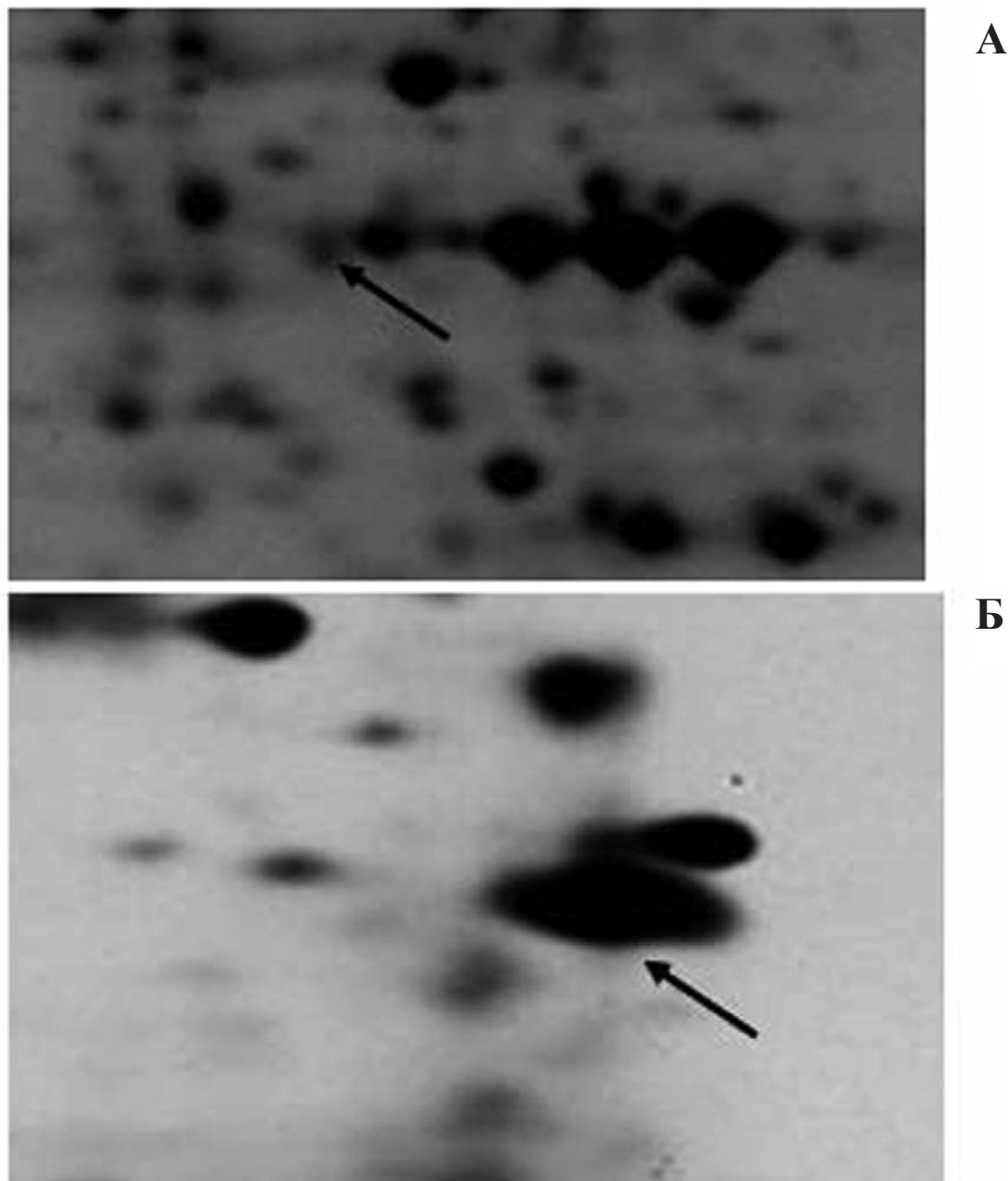


Рисунок 2.

Фрагменты двумерных электрофореграмм белковых экстрактов тканей простаты, полученных с использованием мочевины (А) и без неё (Б). Стрелками показаны: А – изоформа ПСА-ас с Мм 21,8 кДа и рI 6,72; Б – основная изоформа ПСА-ои, находящаяся, по-видимому, в комплексе с пептидным ингибитором SPAAT, благодаря чему обнаруженная белковая фракция, характеризовалась значениями Мм 35,5 и рI 7,30.

Основную изоформу ПСА удалось идентифицировать только на двумерных электрофореграммах белковых экстрактов тканей простаты, полученных без использования мочевины. В показанной на рисунке 2Б, значительной по количеству фракции (Мм 35,5 и pI 7,30;) были выявлены пептиды, соответствовавшие как основной изоформе ПСА (ПСА-ои, записи P07288 в SWISS-PROT), так и другому белковому продукту, обозначаемому аббревиатурой - SPAAT (N-terminal of short peptide alpha 1-antitrypsin, запись 257063 NCBI). Известно, что SPAAT является пептидом из 40 аминокислотных остатков, образующимся посттрансляционно из N-концевого участка аминокислотной последовательности α -1-антитрипсина. Этот пептид обладает свойствами ингибитора многих сериновых протеаз [13].

Таким образом, полученные данные позволили предположить, что белковая фракция с Мм 35,5 и pI 7,30 является комплексом ингибитора SPAAT (4 кДа) с ПСА-ои (27 кДа), представляющую собой известную сериновую протеазу из семейства калликреинов. При масс-спектрометрическом анализе в составе ПСА-ои, были выявлены два пептида (Mz 3538.63 и 3554.52), которые содержали один и тот же “конфликтный” участок из четырех аминокислотных остатков - FLTP (табл. 2). В целом, полученные результаты показывают, что с помощью протеомного анализа удастся характеризовать разные виды полиморфизма ПСА, включая идентификацию разных белковых продуктов экспрессии гена *KLK3*. Это открывает новые возможности для дальнейших исследований белкового полиморфизма ПСА, что особенно существенно в связи с недавно полученными данными об образовании, как минимум, 15 транскриптов за счет альтернативного сплайсинга, происходящего при функционировании гена *KLK3* [14].

ЗАКЛЮЧЕНИЕ. Проведенный протеомный анализ, охвативший несколько сотен белков из ряда тканей и органов человека, позволил уточнить первичные структуры 41 белка и параллельно выявить у некоторых из них полиморфные варианты. Соответственно, можно ожидать, что расширение подобных исследований будет способствовать развитию представлений о белковом полиморфизме человека и расширению возможностей его использования при решении различных биомедицинских задач.

Работа поддерживалась грантом “Изучение биохимического полиморфизма СН-домен содержащих и других актин-связывающих белков при раке простаты” Программы Президиума РАН “Фундаментальные науки – медицине”.

ЛИТЕРАТУРА

1. *Reijneveld J.C., Ginjaar I.B., Frankhuizen W.S., Notermans N.C.* (2006) *Muscle Nerve*, **34**, 656-658.
2. *Kovacs P., Stumvoll M., Bogardus C., Hanson R.L., Baier L.J.* (2006) *Diabetes*, **55**, 1497-1503.
3. *Mattarucchi E., Marsoni M., Binelli G., Passi A., Lo Curto F., Pasquali F., Porta G.* (2005) *J. Biochem. Mol. Biol.*, **38**, 555-562.
4. *Watt K.W.K., Lee P.J., M'Timkulu T., Chan W.P., Loo R.* (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3166-3170.
5. *Schulz P., Stucka R., Feldmann H., Combriato G., Klobeck H.-G., Fittler F.* (1988) *Nucleic Acids Res.*, **16**, 6226.
6. *Говорун В.М., Арчаков А.И.* (2002) *Биохимия*, **67**, 1341-1359.
7. *Ковалев Л.И., Ковалева М.А., Ковалев П.Л., Серебрякова М.В., Мошковский С.А., Шишкин С.С.* (2006) *Биохимия*, **71**, 554-560.
8. *Крохина Т.Б., Шишкин С.С., Раевская Г.Б., Ковалев Л.И., Еришова Е.С., Черников В.Г., Мирочник В.В., Бубнова Е.Н., Кухаренко В.И.* (1996) *Бюлл. эксп. биол. мед.*, **122**, 314-317.
9. *Kovalyov L.I., Shishkin S.S., Efimochkin A.S., Kovalyova M.A., Ershova E.S., Egorov T.A., Musalyamov A.K.* (1995) *Electrophoresis*, **16**, 1160-1169.

10. *Shevchenko A., Wilm M., Vorm O., Mann M.* (1996) *Anal. Chem.*, **68**, 850–858.
11. *Говорун В.М., Мошковский С.А., Тихонова О.В., Гоуфман Е.И., Серебрякова М.В., Момыналиев К.Т., Лохов П.Г., Хряпова Е.В., Кудрявцева Л.В., Смирнова О.В., Торопыгин И.Ю., Максимов Б.И., Арчаков А.И.* (2003) *Биохимия*, **68**, 52-60.
12. *Шишкин С.С., Ковалев Л.И., Крахмалева И.Н., Ковалева М.А., Еремина Л.С., Попов В.О.* (2007) В кн.: Молекулярный полиморфизм человека. Под. ред. С.Д. Варфоломеева. М. Из-во РУДН. т. 2. 483-539.
13. *Niemann M.A., Baggott J.E., Miller E.J.* (1997) *Biochim. Biophys. Acta*, **1340**, 123-130.
14. *Heuze-Vourc'h N., Leblond V., Courty Y.* (2003) *Eur. J. Biochem.*, **270**, 706-714.

Поступила: 17. 12. 2007.

**DETERMINATION OF “AMINO ACIDS CONFLICTS” AND AMINO ACIDS
SUBSTITUTIONS IN PRIMARY STRUCTURES OF 41 HUMAN PROTEINS
BY USE OF THE PROTEOMIC TECHNOLOGIES**

*M.A. Kovalyova¹, L.I. Kovalyov¹, L.S. Eryomina¹, A.A. Makarov¹, M.V. Burakova¹, I.Y. Toropygin²,
M.V. Serebryakova³, S.S. Shishkin¹, A.I. Archakov²*

¹Bakh Institute of Biochemistry, Russian Academy of Sciences, Leninsky pr., 33/2, Moscow,
119071 Russia; fax: (095)954-2733; e-mail: kovalyov@inbi.ras.ru

²Orekhovich Institute of Biomedical Chemistry, Pogodinskaya ul., 10, Moscow, 119121 Russia

³Institute of Physico-Chemical Medicine, Moscow, Russia

Proteomic studies of some human tissues and organs (skeletal muscles, myometrium, motor zone of the brain, prostate), and also cultivated myoblasts revealed 41 of 300 identified proteins, in which the present of certain variants of amino acids (“conflicts”) was recognized at several positions. Among the 93 registered amino acids “conflicts”, seven cases represented the results of the protein polymorphisms caused by corresponding substitution of individual amino acid. Moreover, among prostate proteins the proteomic analysis revealed two isoforms of prostate-specific antigen, formed due to alternative splicing. Thus, our results have shown, that proteomic technologies allow to specify effectively the features of primary structures and to characterize various kinds of polymorphism in many human proteins.

Key words: proteomics, human protein sequences, amino acid conflicts.