Занятие №13. Обработка протеомных данных (масс-спектрометрия)

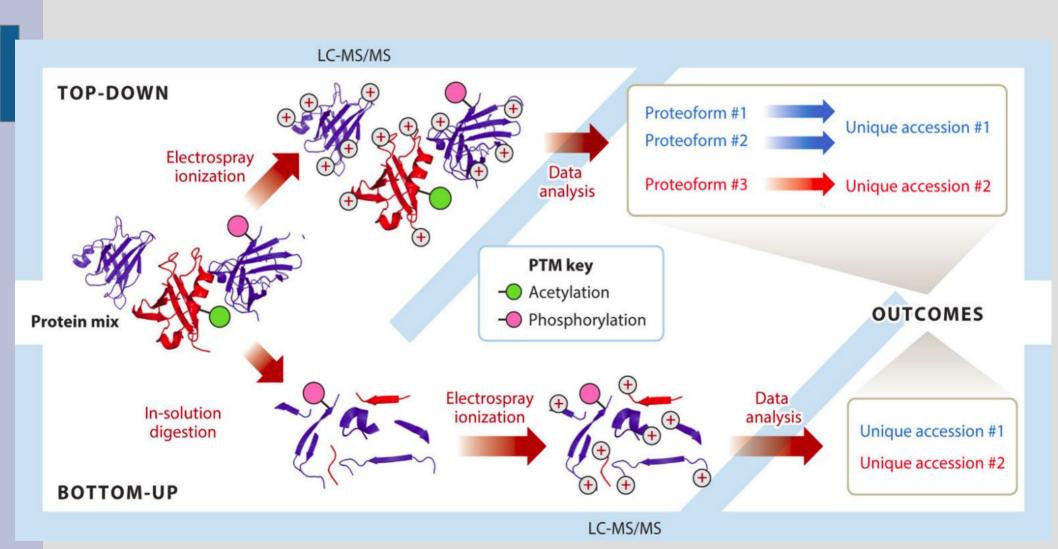
Подходы к анализу в протеомной масс-спектрометрии

- I. "Восходящий" анализ Bottom-up
- гидролиз белка/белков, масс-спектрометрия малых пептидных фрагментов

- II. "Нисходящий" анализ Top-down
- анализ без гидролиза (интактные протеины)

- III. Middle-down
- гидролизат состоит из более крупных пептидов

Два главных подхода к протеомному анализу



Top-down: изучение интактных белков

- 1). Позволяет изучать индивидуальные протеины.
- 2). Является низкопроизводительным подходом.

 Исключение: изучение белкового спектра для идентификации бактериальных культур. Данная методика предусматривает высокопроизводительное сравнительное исследование целых протеомов (однако, имеет ограниченную информативность о природе белков).

Bottom-up

- Данный подход преобладает в протеомных исследованиях.
- Существует в двух вариантах:
- 1. Выделение одного/нескольких белков (PAGE), их гидролиз —> масс-спектрометрия.
- 2. Гидролиз сложной смеси белков —> хроматомассспектрометрия.

Второй вариант известен также как протеомное исследование по методу дробовика – shotgun proteomics.

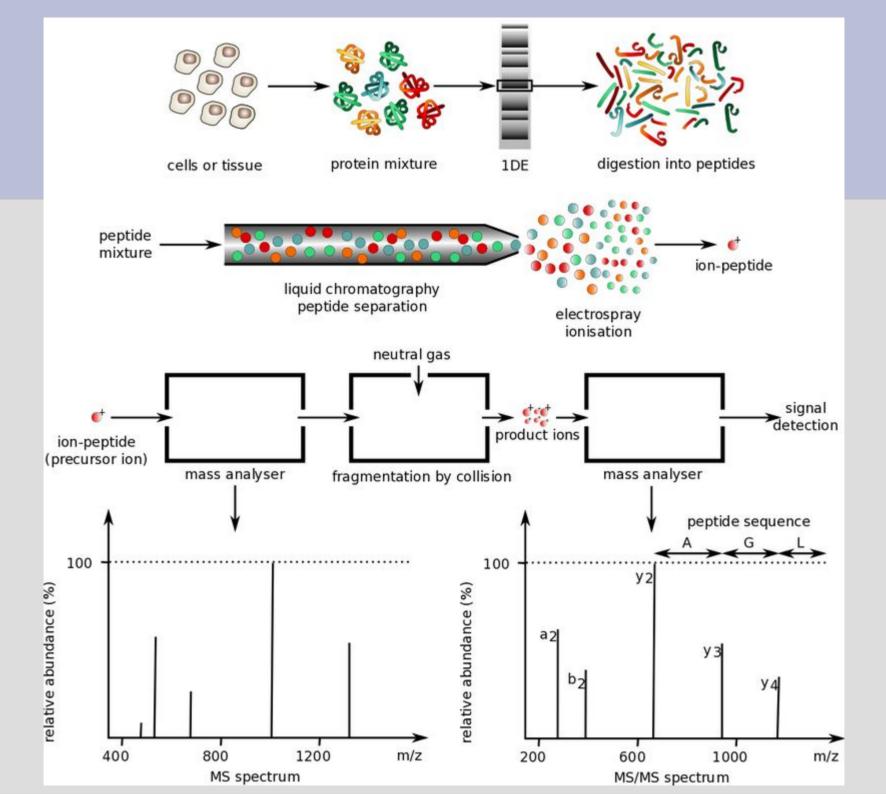
Типы протеомного исследования

1. Определение наличия в образце белков с известной последовательностью ("ресеквенирование" белков).

2. Определение последовательности белков de novo.

Определение белков de novo

Применяется на основе bottom-up подхода (анализ пептидных спектров). В данном случае требуется (особенно для сложных смесей белков) использование тандемного хроматомасс-спектрометра (LC-MS/MS). MS/MS обеспечивает большую разрешающую способность за счёт образования большего числа ионов.



Основа анализа масс-спектров – определение типов образующихся ионов

- В подавляющем большинстве методов МС в результате ионизации образуются положительно заряженные ионы (протонирование, [M+H]⁺). При этом из одной изначальной молекулы (или иона в случае второго этапа MS/MS) может образовываться несколько типов ионов:
- 1). Протонированная исходная молекула простое присоединение иона H+.
- 2). Множество ионизированных кусочков исходной молекулы в результате миграции иона H+ по молекуле и его "оседания" около какой-либо связи происходит разрыв по этой связи.

Главные типы ионов, образующихся в протеомной МС

- Существует три вида связей в пептидах, по которым происходит ионная фрагментация:
- а). Алкил-карбонильная: HC(R) CO (ионы типа а или х)
- б). Пептидная: C(O) NH (ионы типа b или у)
- в). Амино-алкильная: HN CHR (ионы типа с или z)
- 6 типов ионов возникают из-за того, что заряженным может быть фрагмент как слева от связи (ближе к N-концу; типы а, b, c), так и справа (ближе к С-концу; типы х, у, z). В качестве нижнего индекса к типу иона приписывают количество аминокислотных остатков, его образующих (напр., у₂, b₁).
- Наиболее распространёнными являются ионы типов b и у (комплементарные ионы). Часто встречаются а-ионы, образующиеся из b-ионов в результате потери карбонильной группы (СО).

Дополнительные пептидные ионы

- При более жёсткой ионизации наряду с главными ионами происходит образование саттелитных ионов (d-, v-, и w-типы).
- В этих ионах по сравнению с главными ионами происходит дополнительное присоединение Н+ к аминокислотному радикалу.

Главные и саттелитные пептидные ионы

$$R_1 O CHR' I^{H+}$$
 $H_2N-C-C-N-C$
 $H H H H$
 $H_2N-C-C-N-C$
 $H H H H$
 H_1
 $H_2N-C-C-N-C$
 H_1
 $H_2N-C-C-N-C-COOH$
 H_1
 $H_2N-C-C-C-N-C-COOH$
 H_1
 $H_2N-C-C-C-N-C-COOH$
 H_1
 $H_2N-C-C-C-N-C-COOH$
 H_1
 $H_2N-C-C-C-COOH$
 H_1
 $H_2N-C-C-C-COOH$
 H_1
 $H_2N-C-C-C-C-COOH$
 H_1
 $H_2N-C-C-C-C-C-COOH$
 H_1
 H_2
 H_1
 H_1
 H_2
 H_1
 H_2
 H_1
 H_2
 H_1
 H_1
 H_2
 H_1
 H_2
 H_1
 H_2
 H_1
 H_1
 H_1
 H_2
 H_1
 H_2
 H_1
 H_1
 H_1
 H_2
 H_1
 H_2
 H_1
 H_2
 H_1
 H_2
 H_1
 H_1
 H_1
 H_2
 H_1
 H_2
 H_1
 H_1
 H_1
 H_2
 H_1
 H_2
 H_1
 H_2
 H_1
 H_1
 H_2
 H_1
 H_2
 H_1
 H_2
 H_1
 H_1
 H_1
 H_2
 H_1
 H_1
 H_1
 H_2
 H_1
 H_1
 H_1
 H_2
 H_1
 H_2
 H_1
 H_1
 H_1
 H_2
 H_1
 H_2
 H_1
 H_1
 H_1
 H_2
 H_1
 H_1
 H_1
 H_2
 H_1
 H_1
 H_1
 H_1
 H_2
 H_1
 H_1
 H_1
 H_2
 H_1
 H_1
 H_2
 H_1
 H_1
 H_1
 H_1
 H_1
 H_1
 H_1
 H_1
 H_1

Связь между массой главного иона и массами аминокислотных остатков иона

 Масса образующихся главных ионов определяется суммой масс аминокислот, входящих в их состав с учётом особенностей образования (с N- или С-конца; первичный или вторичный ион):

```
    m(b)=∑m(aa) + 1 ; 1 – масса протона
```

 m(y)=∑m(aa) + 19 ; 19 – масса протона+молекула воды

• m(a)=m(b) - 28 ; 28 – масса карбонила (CO)

Массы аминокислотных остатков, главных и соответствующих главных (а₁-ионы; immonium ion) и сателлитных ионов (related ions)

Name	3-letter code	1-letter code	Residue Mass	Immonium ion	Related ions	Composition	
Alanine	Ala	A	71.03711	44		C ₃ H ₅ NO	
Arginine	Arg	R	156.10111	129	59,70,73,87,100,112	C ₆ H ₁₂ N ₄ O	
Asparagine	Asn	N	114.04293	87	70	C ₄ H ₆ N ₂ O ₂	
Aspartic Acid	Asp	D	115.02694	88	70	C ₄ H ₅ NO ₃	
Cysteine	Cys	C	103.00919	76		C ₃ H ₅ NOS	
Glutamic Acid	Glu	E	129.04259	102		C ₅ H ₇ NO ₃	
Glutamine	Gln	Q	128.05858	101	56,84,129	C ₅ H ₈ N ₂ O ₂	
Glycine	Gly	G	57.02146	30		C ₂ H ₃ NO	
Histidine	His	Н	137.05891	110	82,121,123,138,166	C ₆ H ₇ N ₃ O	
Isoleucine	Ile	I	113.08406	86	44,72	C ₆ H ₁₁ NO	
Leucine	Leu	L	113.08406	86	44,72	C ₆ H ₁₁ NO	
Lysine	Lys	K	128.09496	101	70,84,112,129	C ₆ H ₁₂ N ₂ O	
Methionine	Met	M	131.04049	104	61	C ₅ H ₉ NOS	
Phenyalanine	Phe	F	147.06841	120	91	C ₉ H ₉ NO	
Proline	Pro	P	97.05276	70		C ₅ H ₇ NO	
Serine	Ser	S	87.03203	60		C ₃ H ₅ NO ₂	
Threonine	Thr	T	101.04768	74		C ₄ H ₇ NO ₂	
Tryptophan	Trp	W	186.07931	159	11,117,130,132,170,100	$C_{11}H_{10}N_2O$	
Tyrosine	Tyr	Y	163.06333	136	91,107	C ₉ H ₉ NO ₂	
Valine	Val	V	99.06841	72	44,55,69	C ₅ H ₉ NO	

Массы b2-ионов

								-										
	G	Α	S	Р	٧	T	С	I/L	N	D	K/Q	E	M	Н	F	R	Υ	W
G	115									-	2 11 7			8 8				
Α	129	143					0 0	-									23 12	
S	145	159	175				Î											
Р	155	169	185	195									i i					
٧	157	171	187	197	199		9 10 10 30							30				
Т	159	173	189	199	201	203	, u									ž.	V. 130	
С	161	175	191	201	203	205	207											
I/L	171	185	201	211	213	215	217	227						5.				
N	172	186	202	212	214	216	218	228	229					9 8				
D	173	187	203	213	215	217	219	229	230	231			20	- 5			2: :3	
K/Q	186	200	216	226	228	230	232	242	243	244	257							
Е	187	201	217	227	229	231	233	243	244	245	258	259			,			
M	189	203	219	229	231	233	235	245	246	247	260	261	263	18				
Н	195	209	225	235	237	239	241	251	252	253	266	267	269	275			2.	
Fb	205	219	235	245	247	249	251	261	262	263	276	277	279	285	295			
R	214	228	244	254	256	258	260	270	271	272	285	286	288	294	304	313		
Υ	221	235	251	261	263	265	267	277	278	279	292	293	295	301	311	320	327	
W	244	258	274	284	286	288	290	300	301	302	315	316	318	324	334	343	350	373

GG=N=114; GA=K/Q=128; GV=R=156; GE=AD=SV=W=186.

"Правила" ионизации пептидов

- 1). Подавляющее большинство ионов приходится на b-, y- и а- ионы. а-Ионы образуются из b-ионов в результате отщепления СО.
- 2). Аминокислоты, содержащие гидроксильную группу в радикале (Ser, Thr, Asp, Glu), при ионизации теряют её в виде воды (-18).
- 3). Аминокислоты, содержащие аминогруппу в радикале (Lys, Arg, Asn, Gln), при ионизации теряют её в виде аммиака (-17).
- 4). Наличие на С-конце основной аминокислоты (Arg, His, Lys) приводит к образованию вместо иона b_n иона (b_{n-1}+18).
- 5). Комплементарные ионы (b_n и у_{N-n}), образующиеся из одного пептида, в сумме дают одно и то же значение массы пептида из N аминокислот (точнее, массу пептида + 2 из-за двух ионов водорода).

Количественный протеомный анализ

- Позволяет оценивать относительные количества одинаковых белков их разных источников в результате анализа смешанного образца.
- Выделяют несколько методов:
- I. С введением метки исследуемые белки метят молекулами, содержащими разное количество стабильных изотопов 13С и 15N.
- 1). Метаболическая метка
- 2). Химическая метка
- II. Без введения метки анализ хроматограмм или пиков специфических пептидных ионов, отвечающих отдельным протеинам.

Виды количественного протеомного МС-анализа

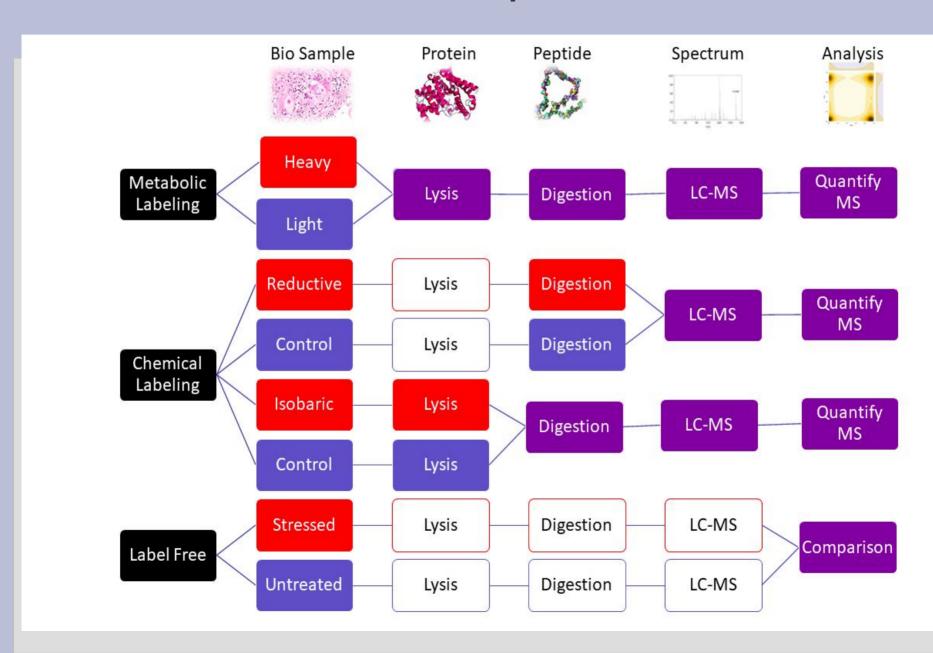
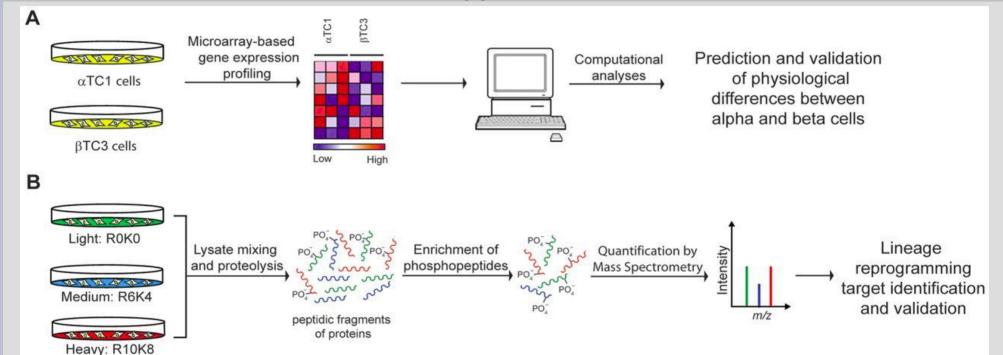


Схема метаболического мечения белков – культивирование клеток на средах с разным содержанием "лёгких" и "тяжёлых" аминокислот – метод SILAC



C

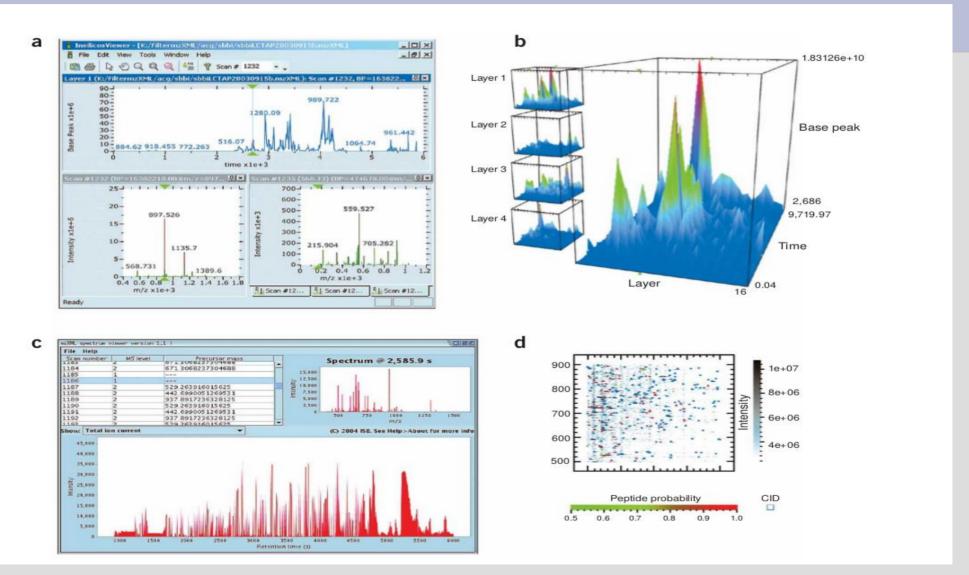
Experiment	Light	Medium	Heavy αTC1 cells		
1a	βTC3 cells	[αTC1+ βTC3] cells			
1b	[αTC1+ βTC3] cells	αTC1 cells	βTC3 cells		
2a*	No treatment	DMSO, 1h	BRD7389, 1h		
2b*	No treatment	DMSO, 120h	BRD7389, 120h		
2c*	No treatment	DMSO, 120h	GW8510, 120h		

* all alpha cells

Форматы представления МС-данных

- 1. "Родные" форматы приборов
- TDF (Bruker), t2d (ABI), lcd (Shimadzu), tdc (Physical Electronics) и мн. др.
- 2. Универсальные форматы
- mzXML, mzXL

Примеры представления данных из mzXML файлов в разных программах (a,b – InsilicosViewer; c – mzXML Spectrum Viewer; d - Pep3D)

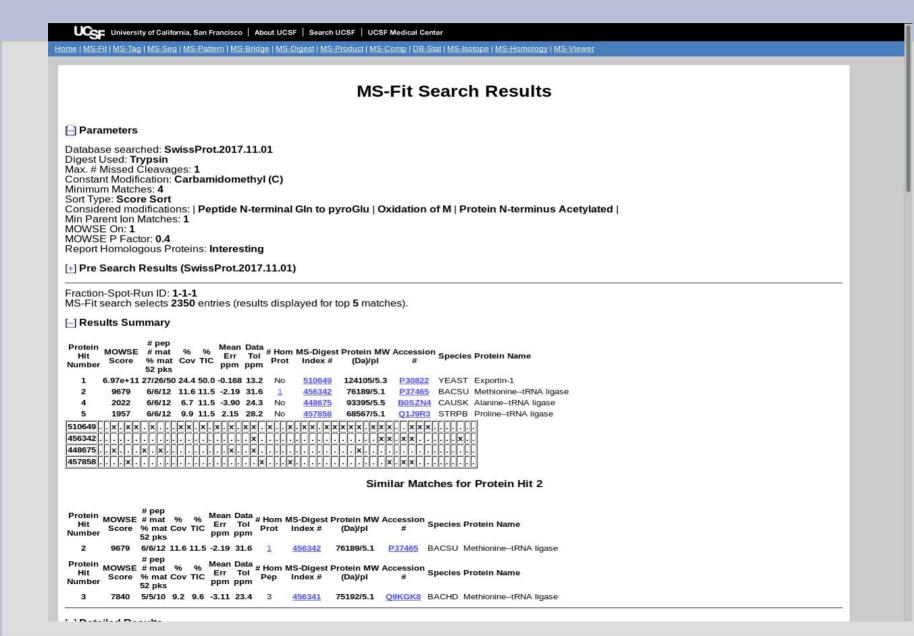


а,с – масс-спектры и хроматограммы; b – масс-спектры четырёх экспериментов; d – оценка результатов исследования после поиска в пептидной базе данных

Анализ результатов МС-эксперимента

- 1. Поиск по базам данных масс-спектров пептидных фрагментов записей, отвечающих (с учётом посттрансляционных модификаций) с полученными данными (напр., SEQUEST).
- 2. Расчёт вероятности правильного определения пептидов (напр., PeptideProphet).
- 3. Поиск по базам данных масс-спектров триптических гидролизатов белков соответствий с выявленными пептидами и оценка вероятности правильного определения белков (напр., ProteinProphet).
- 4. [Для количественных исследований] Определение количественного отношения белков из разных источников (напр., XPRESS).
- 5. Интерпретация качественных и количественных данных с помощью метаболических сетей (Cytoscape).

Пример результатов анализа масс-спектра с помощью инструмента MS-Fit онлайн-ресурса ProteinProspector



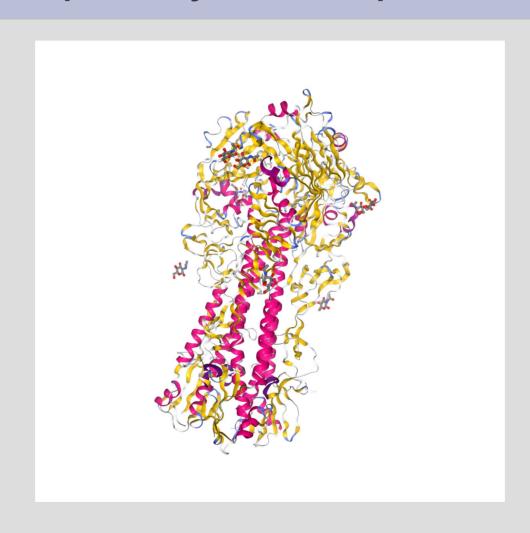
FASTA – универсальный формат представления первичной структуры белков и нуклеиновых кислот

 >BBB04705.1 M1 protein [Influenza A virus (A/WSN/1933(H1N1))]
 MSLLTEVETYVLSIVPSGPLKAEIAQRLEDVFAGKNTDLEVLMEWL KTRPILSPLTKGILGFVFTLTVPSERGLQRRRFVQNALNGNGDPNN MDKAVKLYRKLKREITFHGAKEIALSYSAGALASCMGLIYNRMGAV TTEVAFGLVCATCEQIADSQHRSHRQMVTTTNPLIRHENRMVLAS TTAKAMEQMAGSSEQAAEAMDIASQARQMVQAMRTVGTHPSSS AGLKDDLLENLQAYQKRMGVQMQRFK

Обозначение аминокислот в формате FASTA

Код	Значение	Код	Значение

Трёхмерная структура гемагглютинина Н1, полученная с помощью средств базы данных PDB (разным цветом обозначены разные участки вторичной структуры)



Протеомные базы данных и интернет-ресурсы

- Molbiol https://www.molbiol.ru
- операции с нуклеотидными и аминокислотными последовательностями и др.
- NCBI https://www.ncbi.nih.nlm.gov/Protein
- белковые последовательности, их сравнение, статьи и пр.
- Protein Data Bank (PDB) https://www.rcsb.org
- структура белковых молекул
- ExPASy https://www.expasy.org/proteomics
- портал биоинформатических ресурсов
- ProteinProspector

 https://www.prospector.ucsf.edu/prospector/mshome.htm
- инструменты для анализа масс-спектров макромолекул