

## Thesis work

Table peptides 2-50 AAR - SwissProt (555426 proteins)

Felipe Araya Barrera

201173501-3

October 7th, 2017

## Introduction

I attach the table with the results obtained based on the file **uniprot\_sprot.fasta** (updated in September - 555426 proteins). This table is similar to Table 1 in the file *Searching the protein sequence database*, with the updated results:

Length of peptide ( $N$ )	Total possible N-peptides ( $20^N$ )	Number found	Percentage found (%)
1	20	20	100
2	400	400	100
3	8000	8000	100
4	160000	159999	99.99938
5	3200000	3113509	97.29716
6	64000000	32921109	51.43923
7	1280000000	84118859	6.571786
8	25600000000	100896814	0.3941282
9	512000000000	105834330	0.02067077
10	10240000000000	108976567	0.001064224
11	204800000000000	111551595	0.00005446855
12	$4.096000 \times 10^{15}$	113751287	$2.777131 \times 10^{-6}$
13	$8.192000 \times 10^{16}$	115660117	$1.411867 \times 10^{-7}$
14	$1.638400 \times 10^{18}$	117330023	$7.161256 \times 10^{-9}$
15	$3.276800 \times 10^{19}$	118797220	$3.625403 \times 10^{-10}$
16	$6.553600 \times 10^{20}$	120090802	$1.832440 \times 10^{-11}$
17	$1.310720 \times 10^{22}$	121236766	$9.249631 \times 10^{-13}$
18	$2.621440 \times 10^{23}$	122255013	$4.663659 \times 10^{-14}$
19	$5.242880 \times 10^{24}$	123160391	$2.349098 \times 10^{-15}$
20	$1.048576 \times 10^{26}$	123966286	$1.182235 \times 10^{-16}$
21	$2.097152 \times 10^{27}$	124685091	$5.945448 \times 10^{-18}$
22	$4.194304 \times 10^{28}$	125325735	$2.987998 \times 10^{-19}$
23	$8.388608 \times 10^{29}$	125896961	$1.500809 \times 10^{-20}$
24	$1.677722 \times 10^{31}$	126406287	$7.534402 \times 10^{-22}$
25	$3.355443 \times 10^{32}$	126859712	$3.780714 \times 10^{-23}$
26	$6.710886 \times 10^{33}$	127262721	$1.896362 \times 10^{-24}$
27	$1.342177 \times 10^{35}$	127619577	$9.508399 \times 10^{-26}$
28	$2.684355 \times 10^{36}$	127934582	$4.765935 \times 10^{-27}$
29	$5.368709 \times 10^{37}$	128211241	$2.388120 \times 10^{-28}$
30	$1.073742 \times 10^{39}$	128452495	$1.196307 \times 10^{-29}$
31	$2.147484 \times 10^{40}$	128661754	$5.991280 \times 10^{-31}$
32	$4.294967 \times 10^{41}$	128841734	$2.999830 \times 10^{-32}$
33	$8.589935 \times 10^{42}$	128994975	$1.501699 \times 10^{-33}$
34	$1.717987 \times 10^{44}$	129123214	$7.515960 \times 10^{-35}$
35	$3.435974 \times 10^{45}$	129227999	$3.761030 \times 10^{-36}$

Length of peptide (N)	Total possible N-peptides ( $20^N$ )	Number found	Percentage found (%)
36	$6.871948 \times 10^{46}$	129311151	$1.881725 \times 10^{-37}$
37	$1.374390 \times 10^{48}$	129374023	$9.413199 \times 10^{-39}$
38	$2.748779 \times 10^{49}$	129418106	$4.708203 \times 10^{-40}$
39	$5.497558 \times 10^{50}$	129444962	$2.354590 \times 10^{-40}$
40	$1.099512 \times 10^{52}$	129455435	$1.177390 \times 10^{-42}$
41	$2.199023 \times 10^{53}$	129450779	$5.886740 \times 10^{-44}$
42	$4.398047 \times 10^{54}$	129431855	$2.942940 \times 10^{-45}$
43	$8.796093 \times 10^{55}$	129399352	$1.471100 \times 10^{-46}$
44	$1.759219 \times 10^{57}$	129354043	$7.352926 \times 10^{-48}$
45	$3.518437 \times 10^{58}$	129296497	$3.674827 \times 10^{-49}$
46	$7.036874 \times 10^{59}$	129227572	$1.836434 \times 10^{-50}$
47	$1.407375 \times 10^{61}$	129147725	$9.176498 \times 10^{-52}$
48	$2.814750 \times 10^{62}$	129057687	$4.585050 \times 10^{-53}$
49	$5.629500 \times 10^{63}$	128957792	$2.290751 \times 10^{-54}$
50	$1.125900 \times 10^{65}$	128848422	$1.144404 \times 10^{-55}$

I also attach the Table 1 of the file mentioned in the beginning, so that you can see the notorious differences of the results obtained.

TABLE I

Length of peptide	Total possible N-peptides	Number found	Percentage found
$N$	$20^{*N}$		
2	400	400	100.0
3	8,000	7,995	99.9
4	160,000	115,817	72.4
5	3,200,000	333,965	10.4
6	64,000,000	387,925	0.61
7	1,280,000,000	399,330	0.031
8	25,600,000,000	405,682	0.0016