ADAPTABLE

Tutorial

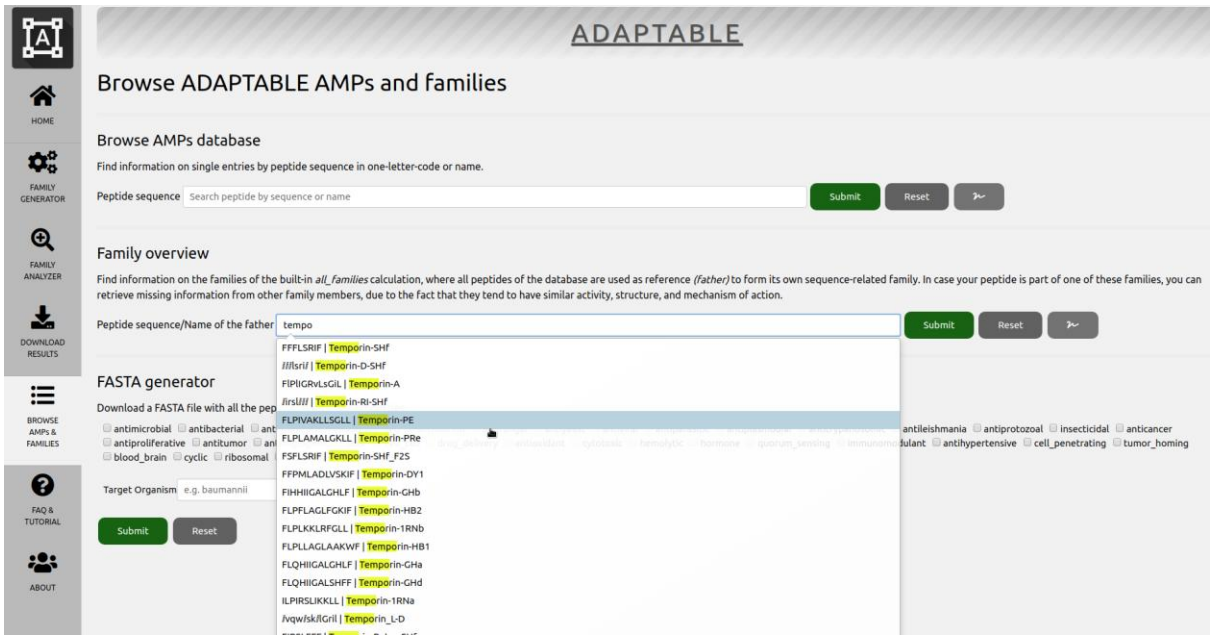
Content

ADAPTABLE TUTORIAL – BROWSING INFORMATION ON PEPTIDE SEQUENCES

How to easily retrieve combined information from external databases and from the ADAPTABLE families database

Suppose you want to retrieve information on a Temporin peptide, whose sequence is FLPIVAKLLSGLL. Go to the "BROWSE AMPs & FAMILIES" page and start to type the name (or part of the sequence) in the dedicated box. While typing, all available entries containing the sequence will be displayed.

Screenshot 1.



Just click on the desired peptide (the fifth in *screenshot 1*) and click "Submit". You will see a table as shown in *screenshot 2*.

In the first lines of the table, we learn that your peptide is able to generate a family similar to it (family 2044) but is also a member of other families of the built-in ADAPTABLE families database (f653, f7918, f1651…) that can be also accessed clicking the provided links.

In the field "External database ID" you can see that it has entries in the DBASSP and APD databases (just click on the links to view the related pages of each database). The complete set of information is gathered in the remaining part of the table. We can see if the family generated by our peptide (family 5092) can provide more information related with properties that are still not known for our peptide.

Screenshot 2

AMP properties

Properties of AMP **FLPIVAKLLSGLL**

Property	Value
Best representing member (father) of the following ADAPTABLE family:	f2044
Member of the following ADAPTABLE families (all_families DB):	f1 f2 f3 f4 f7 f8 f9 f10 f11 f15 f16 f17 f18 f19 f26 f27 f28 f29 f30 f33 f34 f35 f36 f39 f40 f49 f54 f59 f75 f79 f80 f87 f88 f95 f99 f100 f102 f110 f112 f117 f124 f129 f149 f159 f160 f178 f181 f185 f186 f188 f192 f203 f205 f250 f289 f329 f342 f369 f372 f379 f397 f399 f469 f496 f565 f569 f573 f612 f634 f653 f671 f672 f676 f678 f691 f717 f749 f758 f793 f805 f811 f882 f1145 f1186 f1651 f1706 f1919 f1954 f1962 f1999 f2028 f2044 f2047 f2132 f2196 f2197 f2293 f2322 f2340 f2370 f2377 f2423 f2446 f2487 f2489 f2508 f2522 f2534 f2550 f2598 f2609 f2614 f2634 f2639 f2650 f2678 f2694 f2728 f2730 f2739 f2749 f2768 f2821 f2865 f2877 f2905 f2925 f2938 f2951 f2953 f2954 f2964 f2973 f2994 f2997 f3008 f3035 f3055 f3071 f3083 f3120 f3124 f3125 f3149 f3170 f3171 f3172 f3188 f3242 f3243 f3249 f3250 f3275 f3280 f3290 f3330 f3358 f3360 f3368 f3386 f3450 f3451 f3452 f3475 f3482 f3547 f3579 f3581 f3613 f3615 f3619 f3636 f3688 f3701 f3718 f3758 f3769 f3775 f3801 f3810 f3853 f3880 f3881 f3917 f3918 f3969 f4014 f4015 f4041 f4058 f4060 f4078 f4089 f4092 f4134 f4165 f4166 f4181 f4188 f4221 f4234 f4251 f4303 f4373 f4425 f4426 f4437 f4492 f4493 f4494 f4523 f4552 f4553 f4591 f4601 f4607 f4633 f4634 f4658 f4672 f4743 f4762 f4810 f4892 f4898 f4899 f4910 f4951 f5052 f5065 f5098 f5199 f5323 f5350 f5445 f5479 f5492 f5531 f5581 f5582 f5648 f5732 f5734 f5853 f5880 f5910 f5927 f6121 f6318 f6378 f6390 f6757 f6796 f6908 f7076 f7145 f7918 f14649
External database ID	DBAASP10906 APD_AP02977
Sequence	FLPIVAKLLSGLL
Name	Temporin-PE
Source	Pelophylax esculentus; skin secretions, Pelophylax kl. esculentus, Europe
Stereo	L
C Terminus	amidation
PTM	amidation
Target	Lipid Bilayer
Antimicrobial	Yes
Antibacterial	Yes
Antigram Positive	Yes
Antigram Negative	Yes
Antifungal	Yes
Anticancer	Yes
PMID	29191658
Taxonomy	animalia; eumetazoa; edible frog, amphibians, animals; Leu-rich; ; UCLL1c
Targeted Organisms	Candida albicans NCPF 1467 (MIC=4 µM); Staphylococcus aureus NCTC 10788 (MIC=2 µM); Escherichia coli NCTC 10418 (MIC=16 µM); Staphylococcus aureus NCTC 12493 (MIC=4 µM); Enterococcus faaecalis NCTC 12697 (MIC=8 µM); Pseudomonas aeruginosa ATCC 27853 (MIC=128 µM); Human squamous lung carcinoma NCI-H157 (IC50=34.56 µM); Human glioblastoma U251-MG (IC50=25.13 µM); Human prostate adenocarcinoma PC-3 (IC50=38.56 µM); Human breast adenocarcinoma MDA-MB-435S (IC50=33.23 µM); Staphylococcus aureus (MIC=2 µM); MRSA; E. faaecalis; Candida albicans (MIC=4 µM)
Ribosomal	Yes
Experimental	Yes
Antibiofilm	Yes, against: Candida albicans NCPF; Staphylococcus aureus NCTC; Pseudomonas aeruginosa ATCC

Clicking on f2044 (first line of the table), or typing the sequence of the peptide used as reference in "BROWSE AMPs & FAMILIES" section ("Family overview" subsection), we are redirected to a page (see *screenshot 3*) describing the characteristics of the family constituted by similar peptides (peptides that can be considered as generated by our sequence, with few amino acid modifications).

The page also shows the list of members of the families (*screenshot 4*) with links to the entries of external databases. From the table in *screenshot 3* we discover that antibacterial data are available for over 78% of the members of the family and that indeed our peptide might display antimicrobial activity. We also learn that most of the similar peptides have a wide range of activities including antifungal but they also can be toxic and hemolytic. The button "Show frequency of aminoacids" allows you to review the members of the family highlighting the frequencies of their aminoacids to make motifs more easily viewable.

In conclusion, the ADAPTABLE family database can suggest properties that have never been tested for some specific entries.

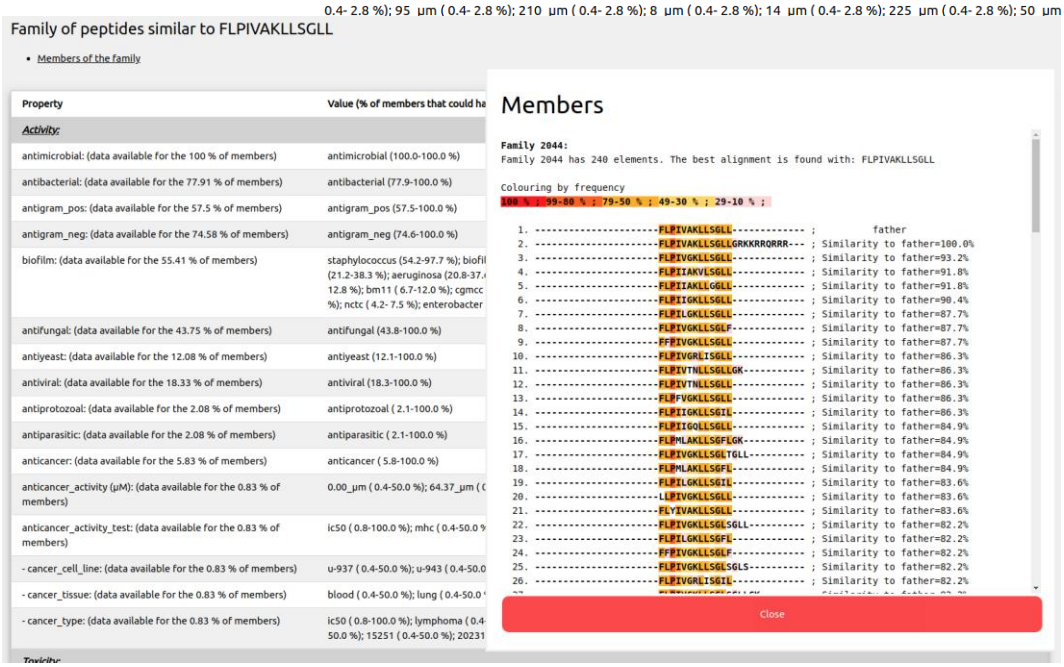
Screenshot3.

Family 2044 properties

Family of peptides similar to FLPIVAKLLSGLL

- Members of the family

Property	Value (% of members that could have it depending on available data per member)
Activity:	
antimicrobial: (data available for the 100 % of members)	antimicrobial (100.0-100.0 %)
antibacterial: (data available for the 77.91 % of members)	antibacterial (77.9-100.0 %)
antigram_pos: (data available for the 57.5 % of members)	antigram_pos (57.5-100.0 %)
antigram_neg: (data available for the 74.58 % of members)	antigram_neg (74.6-100.0 %)
biofilm: (data available for the 55.41 % of members)	staphylococcus (54.2-97.7 %); biofilm (55.4-100.0 %); aureus (53.8-97.0 %); atcc (27.9-50.4 %); candida (37.5-67.7 %); albicans (37.5-67.7 %); pseudomonas (21.2-38.3 %); aeruginosa (20.8-37.6 %); klebsiella (12.1-21.8 %); pneumoniae (11.7-21.1 %); bacillus (9.6-17.3 %); megaterium (9.6-17.3 %); epidermidis (7.1-12.8 %); bm11 (6.7-12.0 %); cgmcc (4.2- 7.5 %); cowan (5.4- 9.8 %); 08040724 (4.6- 8.3 %); atcc2002 (4.6- 8.3 %); enterococcus (4.6- 8.3 %); faecium (4.6- 8.3 %); nctc (4.2- 7.5 %); enterobacter (4.2- 7.5 %); cloacae (4.2- 7.5 %); atcc2592 (4.2- 7.5 %); streptococcus (2.9- 5.3 %); pyogenes (2.9- 5.3 %); a170 (2.9- 5.3 %)
antifungal: (data available for the 43.75 % of members)	antifungal (43.8-100.0 %)
antiyeast: (data available for the 12.08 % of members)	antiyeast (12.1-100.0 %)
antiviral: (data available for the 18.33 % of members)	antiviral (18.3-100.0 %)
antiprotozoal: (data available for the 2.08 % of members)	antiprotozoal (2.1-100.0 %)
antiparasitic: (data available for the 2.08 % of members)	antiparasitic (2.1-100.0 %)
anticancer: (data available for the 5.83 % of members)	anticancer (5.8-100.0 %)
anticancer_activity (µM): (data available for the 0.83 % of members)	0.00_µm (0.4-50.0 %); 64.37_µm (0.4-50.0 %); 2.48_µm (0.4-50.0 %); 2.15_µm (0.4-50.0 %); 2.20_µm (0.4-50.0 %); 1.88_µm (0.4-50.0 %); 1.86_µm (0.4-50.0 %)
anticancer_activity_test: (data available for the 0.83 % of members)	ic50 (0.8-100.0 %); mhc (0.4-50.0 %)
- cancer_cell_line: (data available for the 0.83 % of members)	u-937 (0.4-50.0 %); u-943 (0.4-50.0 %); a-549 (0.4-50.0 %); 293 (0.4-50.0 %); hep3b (0.4-50.0 %); mcf-7 (0.4-50.0 %)
- cancer_tissue: (data available for the 0.83 % of members)	blood (0.4-50.0 %); lung (0.4-50.0 %); renal (0.4-50.0 %); liver (0.4-50.0 %); breast (0.4-50.0 %)
- cancer_type: (data available for the 0.83 % of members)	ic50 (0.8-100.0 %); lymphoma (0.4-50.0 %); renal (0.4-50.0 %); mhc (0.4-50.0 %); 000357628 (0.4-50.0 %); 3731 (0.4-50.0 %); lung (0.4-50.0 %); 48573 (0.4-50.0 %); 15251 (0.4-50.0 %); 20231 (0.4-50.0 %); liver (0.4-50.0 %); 88058 (0.4-50.0 %); breast (0.4-50.0 %); 86909 (0.4-50.0 %)
Toxicity:	
toxic: (data available for the 21.25 % of members)	toxic (21.2-100.0 %)
cytotoxic: (data available for the 7.5 % of members)	cytotoxic (7.5-100.0 %)
hemolytic: (data available for the 15.83 % of members)	hemolytic (15.8-100.0 %)
hemolytic_activity (µM): (data available for the 15 % of members)	40_µm (5.8-38.9 %); 1.25_µm (5.8-38.9 %); 300_µm (1.7-11.1 %); 5_µm (1.2- 8.3 %); 120_µm (0.8- 5.6 %); 0.14_µm (0.4- 2.8 %); 30_µm (0.8- 5.6 %); 75_µm (0.4- 2.8 %); 95 µm (0.4- 2.8 %); 210 µm (0.4- 2.8 %); 8 µm (0.4- 2.8 %); 14 µm (0.4- 2.8 %); 225 µm (0.4- 2.8 %); 50 µm (0.4- 2.8 %); 0.05 µm (0.4- 2.8 %);



Screenshot 4.

Members of family 2044

- Properties of the family

Num	Sequence (Click each for more information)	ID
1	----------FLPIVAKLLSGLL----------	DBAASP10906 APD_AP02977
2	----------FLPIVAKLLSGLLGRKKRRQRRR---	DBAASP10908
3	----------FLPIVGKLLSGLL----------	satpdb21344 satpdb21479 hemo1130 DRAMP01342 DRAMP01830 CAMPSQ23 CAMPSQ444 CAMPSQ998 CAMPSQ4071 DBAASP1127 APD_AP00112 ANTISTAPHYStaph._P_13 DADP_2659 DADP_P82848 LAMPL01A000025 ADAM_1230 YADAMP317 uniprotP82848
4	----------FLPIIAKVLSGLL----------	satpdb15096 satpdb16614 DRAMP01810 CAMPSQ505 DBAASP3535 APD_AP00866 DADP_P84116 LAMPL01A000541 ADAM_1204 YADAMP307 uniprotP84116
5	----------FLPIIAKLLGGLL----------	satpdb27817 DRAMP03682 CAMPSQ9 DBAASP1565 InverPep577 LAMPL01A000011 ADAM_1203 YADAMP306 uniprotP0C1M1
6	----------FLPIIGKLLSGLL----------	satpdb22159 satpdb28122 hemo1131 DRAMP01816 CAMPSQ999 CAMPSQ1057 DBAASP1129 DBAASP5609 APD_AP00595 APD_AP02368 InverPep510 InverPep581 ANTISTAPHYStaph._P_14 DADP_2660 LAMPL01A003167 ADAM_1210 YADAMP308
7	----------FLPILGKLLSGLL----------	satpdb16682 satpdb28993 DRAMP01724 DBAASP6006 LAMPL10D5MTH50 ADAM_1221 YADAMP313
8	----------FLPIVGKLLSGLF----------	satpdb11723 satpdb24155 DRAMP01727 CAMPSQ2413 DBAASP1883 APD_AP02056 ADAM_1229 YADAMP316
9	----------FFPIVGKLLSGLL----------	satpdb16610 satpdb23945 DRAMP01736 CAMPSQ3637 CAMPSQ4081 DBAASP1864 APD_AP01936 ANTISTAPHYStaph._P_303 LAMPL02A001936 ADAM_0968 YADAMP282
10	----------FLPIVGRLISGLL----------	satpdb10939 satpdb13725 hemo2483 DRAMP01749 CAMPSQ2921 DBAASP5287 APD_AP00864 DADP_2703 LAMPL02A000864 ADAM_1235 YADAMP318
11	----------FLPIVTNLLSGLLGK----------	DRAMP01776
12	----------FLPIVTNLLSGLL----------	satpdb13503 satpdb22050 CAMPSQ1004 APD_AP00899 DADP_2690 LAMPL01A003172 ADAM_1238 YADAMP319
13	----------FLPFVGKLLSGLL----------	satpdb11702 satpdb17794 DBAASP8038 APD_AP02467 ADAM_1178
14	----------FLPIIGKLLSGIL----------	satpdb29086 LAMPL10Q2PGA80 ADAM_1209
15	----------FLPIIGQLLSGLL----------	satpdb10701 satpdb21465 hemo1618 DRAMP01760 CAMPSQ2922 DBAASP4222 APD_AP00865 LAMPL02A000865 ADAM_1211 YADAMP309
16	----------FLPMLAKLLSGFLGK----------	DRAMP01166 CAMPSQ1319 LAMPL01A002468
17	----------FLPIVGKLLSGLTGLL----------	CAMPSQ4078
18	----------FLPMLAKLLSGFL----------	satpdb16624 ADAM_1304

Show frequency of aminoacids

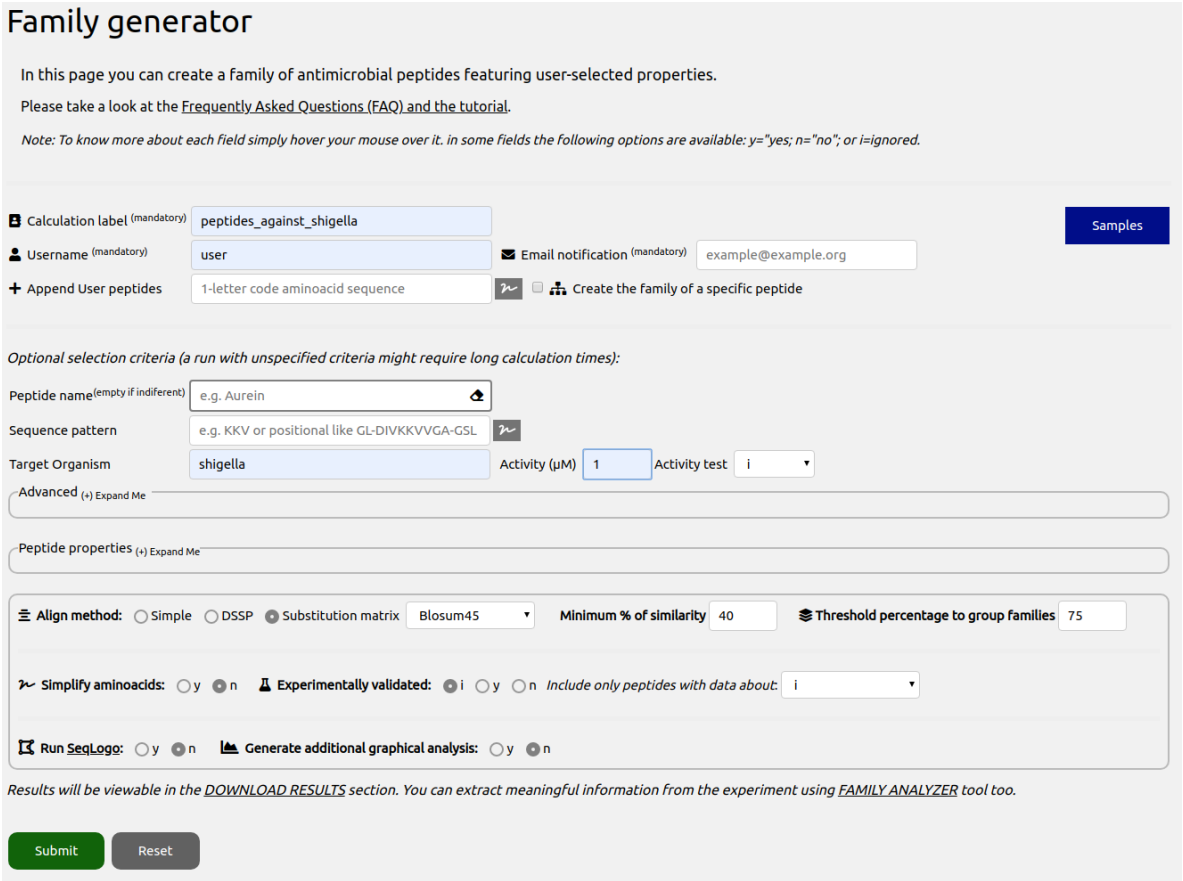
ADAPTABLE TUTORIAL - CASE EXAMPLE n.1

Designing new peptides active towards a specific organism and highlighting motifs

The case of *Shigella*

Suppose you want to design a new antimicrobial peptide active against *Shigella* sp. You want to be inspired by the most active entries in the database, so you look for peptides which are active at concentrations lower than 1 μM. Go to the "FAMILY GENERATOR" section on the left bar. For the Calculation label, you can choose a name such as "peptides_against_shigella". As you do not want to restrict the generation of families only taking into account "Aurein" peptides (provided as an example), remove all from "Peptide name" field. Go to "Target Organism" and write the name of the organism (or part of it) and the maximum μM activity value just like in *screenshot 1.1.* You will also need to choose the username ("user" in the example) to be able to access your experiments:

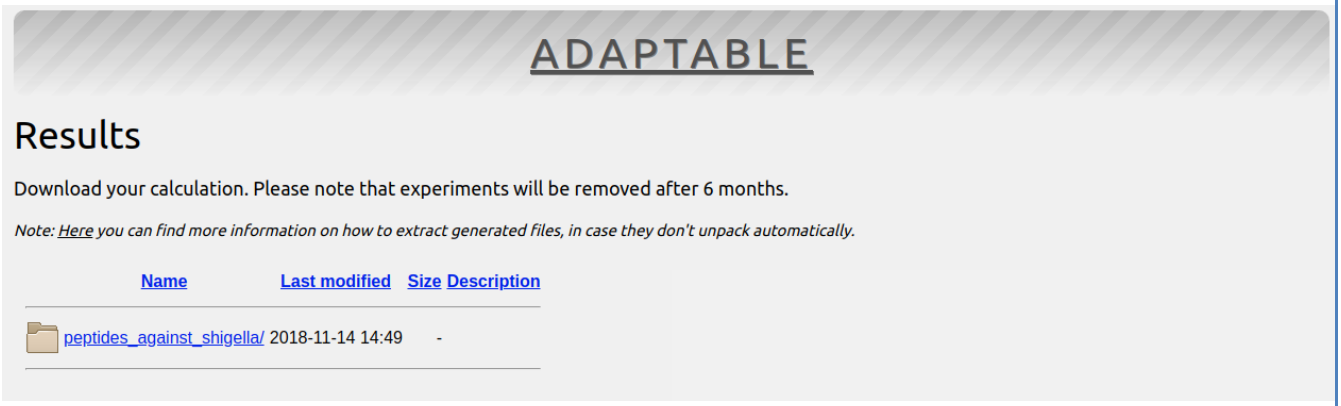
Screenshot 1.1.



You might want to see a detailed analysis so you can also set the options "Generate additional graphical analysis" and "Run SeqLogo" at the end of the page to yes ("y").

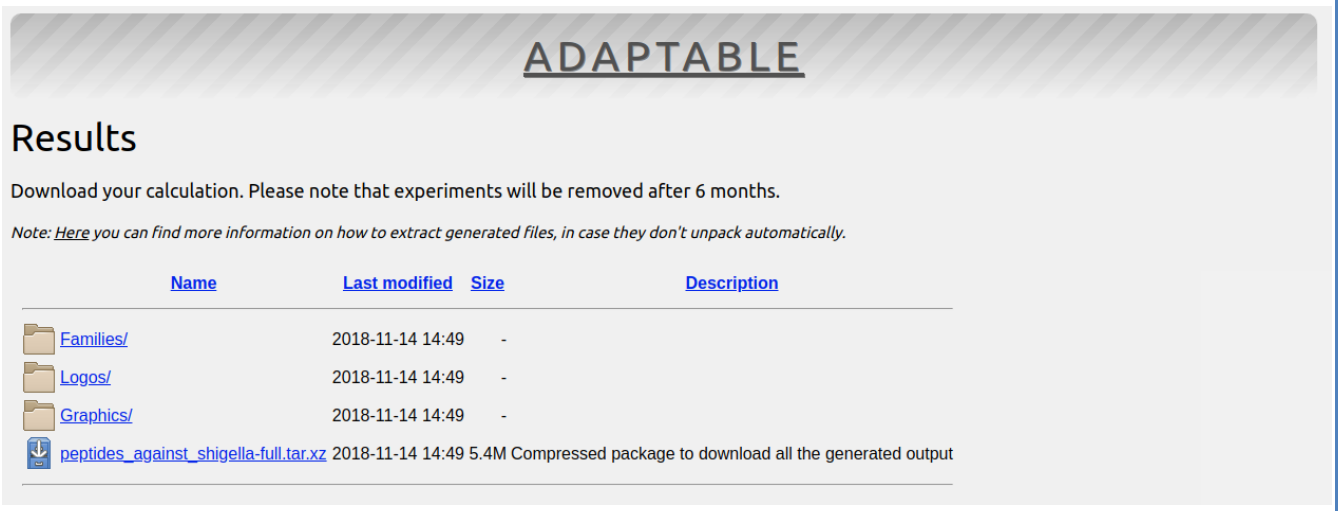
Click on "Submit" at the end of the page. After the calculation is finished, a page summarizing the properties of each generated family will be shown (always accessible from "FAMILY ANALYZER" section). You can also go to the "DOWNLOAD RESULTS" section on the left bar to download all the generated files. In both case you will be asked to type your username:

Screenshot 1.2.



Open the directory of your calculation “peptides_against_shigella”. Here is what you see:

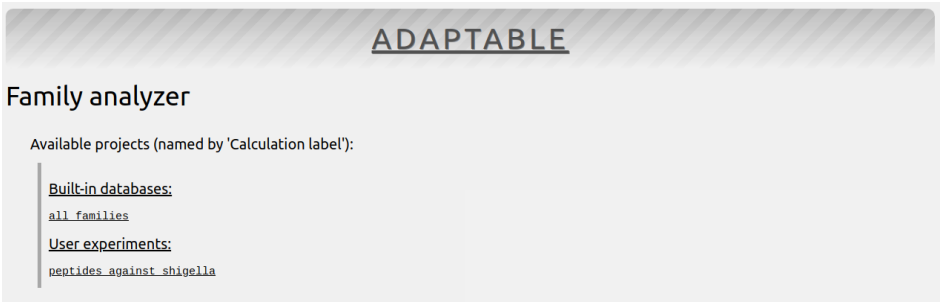
Screenshot 1.3.



You will see a compressed package to download all the generated files, additionally, in the subfolder “Families” you will find the full information for each family (as a whole html file showing different properties and statistics), in the subfolder “Logos” you will find the logos generated by Weblogo and in “Graphics” you will see the generated additional graphical analysis files.

We will go to “FAMILY ANALYZER” section, that includes tools that will allow to parse the generated information in a more detailed way. It will ask for the username and, after that, the following section will be shown to allow choosing the desired user experiment:

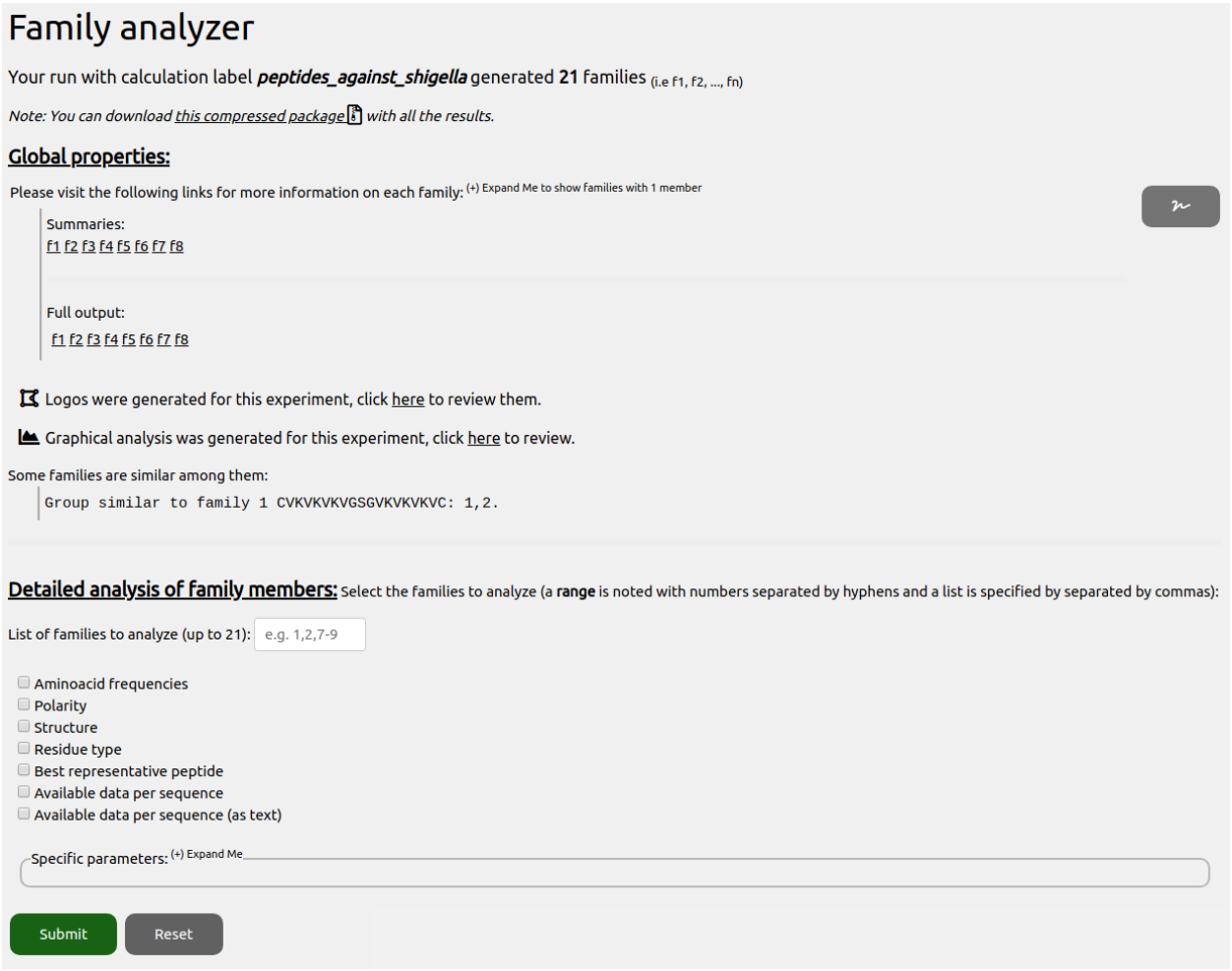
Screenshot 1.4.



After going into our experiment, we can see a page that will provide us with some information (Screenshot 1.5).
The run has generated a total of 8 families with more than one member (you can see the rest clicking on “Expand Me to show families with 1 member”. Regarding the global properties of the families, you can review a summary and the full html output for each.
It is also shown that this run has generated Logos and Graphical plots that can be accessed directly from here too.
Finally, it’s shown that 2 of the 21 generated families are similar among them for the case the user wants to take this into account when reviewing the features of each family.

Going into the “Detailed analysis of family members” we will be able to select the families to analyze for getting more detailed information.

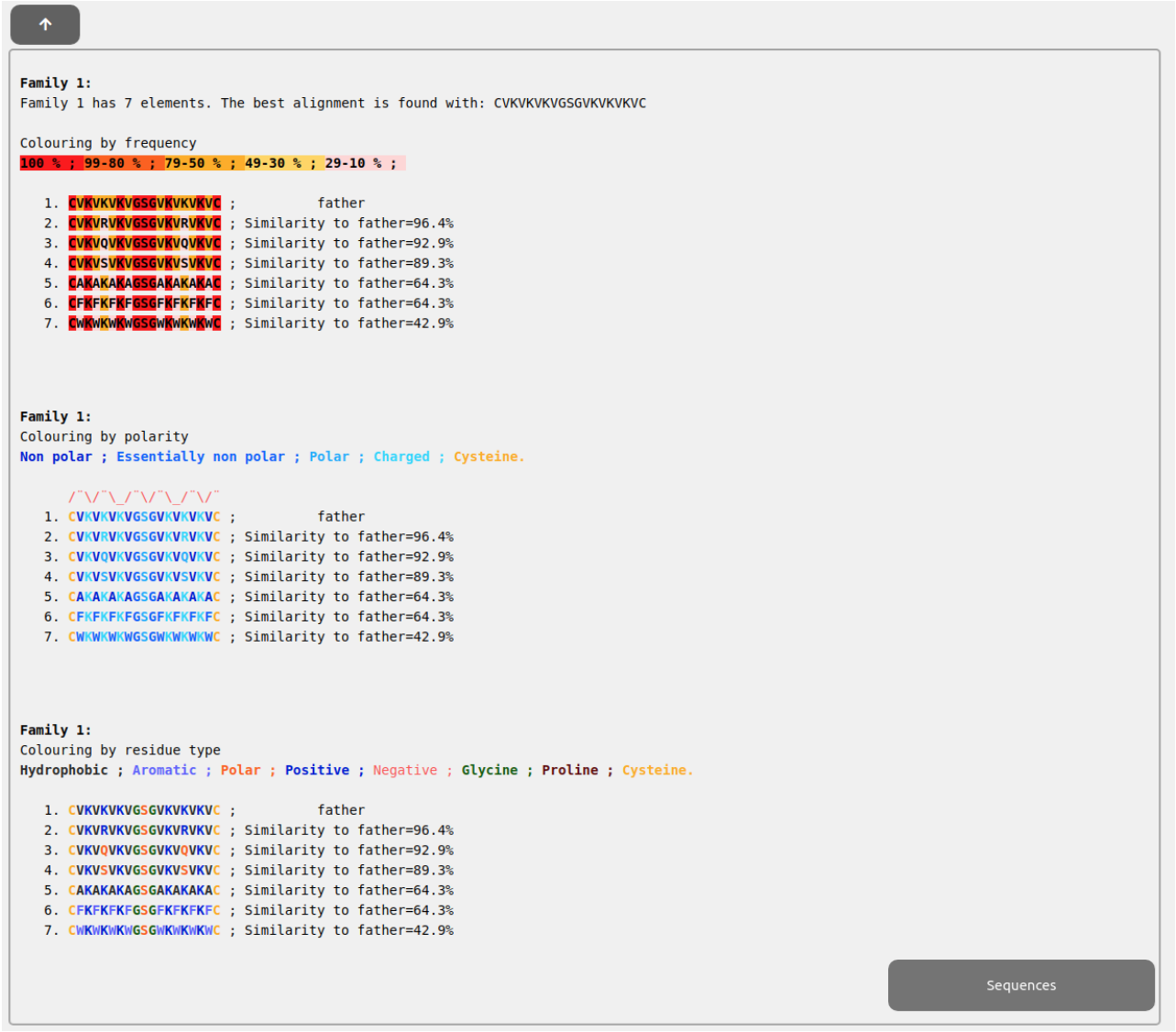
Screenshot 1.5.



Based on sequence alignment, ADAPTABLE has classified all peptides of its database in groups of sequence-related peptides (families), each represented by a single sequence chosen among the members of the family (called the father). The father has the highest resemblance to all other peptides. In this terminology, each family is composed by a father and multiple sons (the other members of the family). This classification is purely based on sequence; it is therefore independent on the activity or other properties.

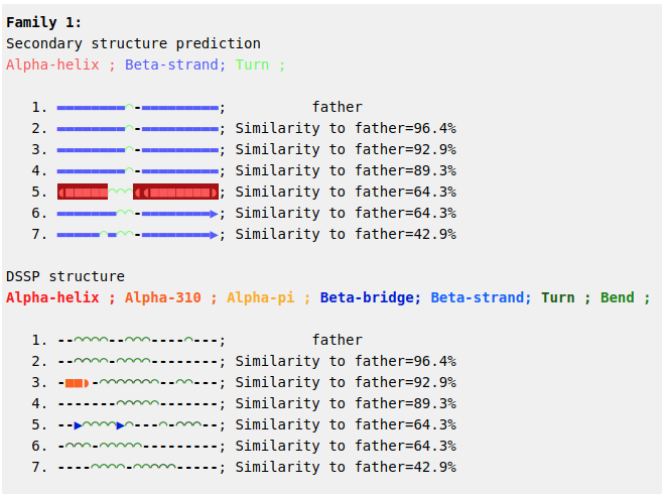
We will choose to review the members of family 1 and get them shown according to the “Aminoacid frequencies”, “Polarity” and type of residue.
Among the 7 members, the first sequence is the one which has been used as reference to build the family, and is therefore called the “father”.
Judging by the frequency of occurrence, it seems that some amino acids (in red) are highly repeated in the same position and therefore potentially important for the antimicrobial mechanism of action. The statistical analysis performed at the end of the calculation will help us to better understand this information (see below).
Coloring the sequence alignment by residue type or polarity can sometime reveal other features (as the presence of amphipathic helix).

Screenshot 1.6.



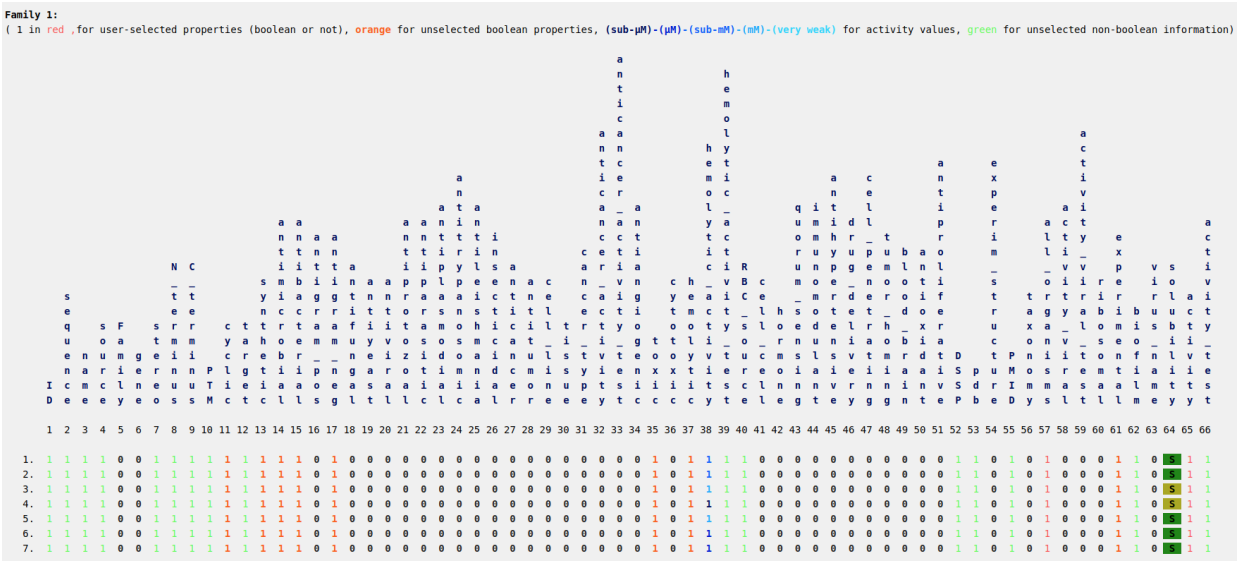
In this case, the interpretation is not straightforward. Polar and nonpolar residues do not seem to alternate with the frequency of a helix (3.5 residues per turn). The secondary structure prediction, along with the DSSP structure (“Structure” option) confirms that a helical structure, commonly found in antimicrobial peptides, is not expected in this case:

Screenshot 1.7.



We can also see a summary of all the information available for the family clicking on “Available data per sequence”:

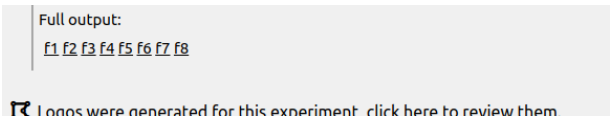
Screenshot 1.8.



In this table, each line corresponds to one member of the family and each column corresponds to a property (“1” indicates presence of the property).

But a more readable output showing several statistics and tables can be obtained in the generated HTMLs with the full output as accessed in “Global properties” section:

Screenshot 1.8.



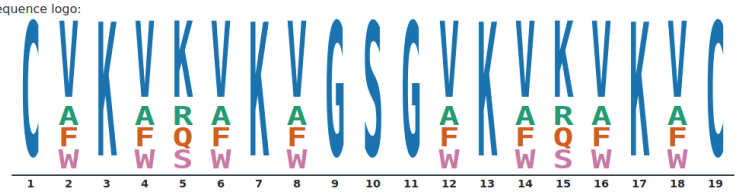
In this case, we will click into f1 output and we will see the whole output divided in multiple sections that can be navigated using its left bar to jump from one section to other.

Screenshot 1.9.

Results from family similar to CVKVKVKVGSGVKVKVKVC (#1 out of 21)

Summary

- This family has **7** elements
- Peptide used to generate this family (father): **CVKVKVKVGSGVKVKVKVC**
- Sequence logo:



Best representing peptide

Optimal peptide sequences as calculated from the most abundant residue at each position (1-19):

Non polar **Essentially non polar** **Polar** **Charged** **Cysteine**

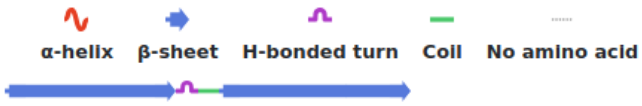
1. CVKVKVKVGSGVKVKVKVC
2. CVKVRVKVGSGVKVRVKVC
3. CVKVKVKVGSGVKVKVKVC
4. CVKVRVKVGSGVKVRVKVC
5. CAKAKAKAGSGAKAKAKAC
6. CVKVRVKVGSGVKVRVKVC
7. CVKVKVKVGSGVKVKVKVC
8. CVKVRVKVGSGVKVRVKVC
9. CVKVKVKVGSGVKVKVKVC
10. CVKVKVKVGSGVKVKVKVC
11. CVKVKVKVGSGVKVKVKVC
12. CVKVRVKVGSGVKVRVKVC
13. CVKVKVKVGSGVKVKVKVC
14. CVKVRVKVGSGVKVRVKVC
15. CAKAKAKAGSGAKAKAKAC
16. CVKVRVKVGSGVKVRVKVC
17. CVKVKVKVGSGVKVKVKVC
18. CVKVRVKVGSGVKVRVKVC
19. CVKVKVKVGSGVKVKVKVC

Best representing sequence:

CVKVKVKVGSGVKVKVKVC (IDs : satpdb28089, hemo2373, DBAASP347)

Consensus secondary structure:

Legend:



Consensus secondary structure (from DSSP):

Legend:

Family members

Family members coloured by frequency:

Legend: 100 % | 99-80 % | 79-50 % | 49-30 % | 29-10 %

1. *Father*
2. *96.4% similarity to father*
3. *92.9% similarity to father*
4. *89.3% similarity to father*
5. *64.3% similarity to father*
6. *64.3% similarity to father*
7. *42.9% similarity to father*

Family members coloured by residue type:

Legend: Hydrophobic | Aromatic | Polar | Positive | Negative | Glycine | Proline | Cysteine

1. *Father*
2. *96.4% similarity to father*
3. *92.9% similarity to father*
4. *89.3% similarity to father*
5. *64.3% similarity to father*
6. *64.3% similarity to father*
7. *42.9% similarity to father*

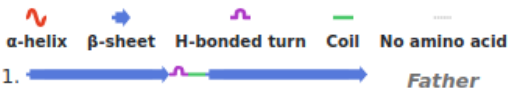
Family members coloured by polarity:

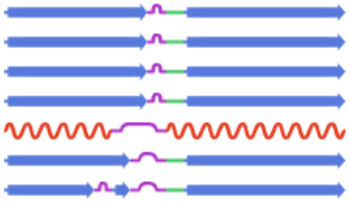
Legend: Non polar | Essentially non polar | Polar | Charged | Cysteine

1. *Father*
2. *96.4% similarity to father*
3. *92.9% similarity to father*
4. *89.3% similarity to father*
5. *64.3% similarity to father*
6. *64.3% similarity to father*
7. *42.9% similarity to father*

Secondary structure prediction for family members:

Legend:

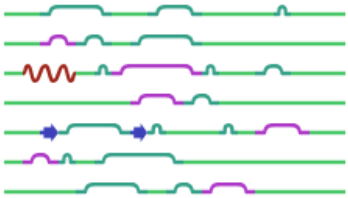




1. *Father*
2. *96.4% similarity to father*
3. *92.9% similarity to father*
4. *89.3% similarity to father*
5. *64.3% similarity to father*
6. *64.3% similarity to father*
7. *42.9% similarity to father*

Secondary structure from DSSP for family members:

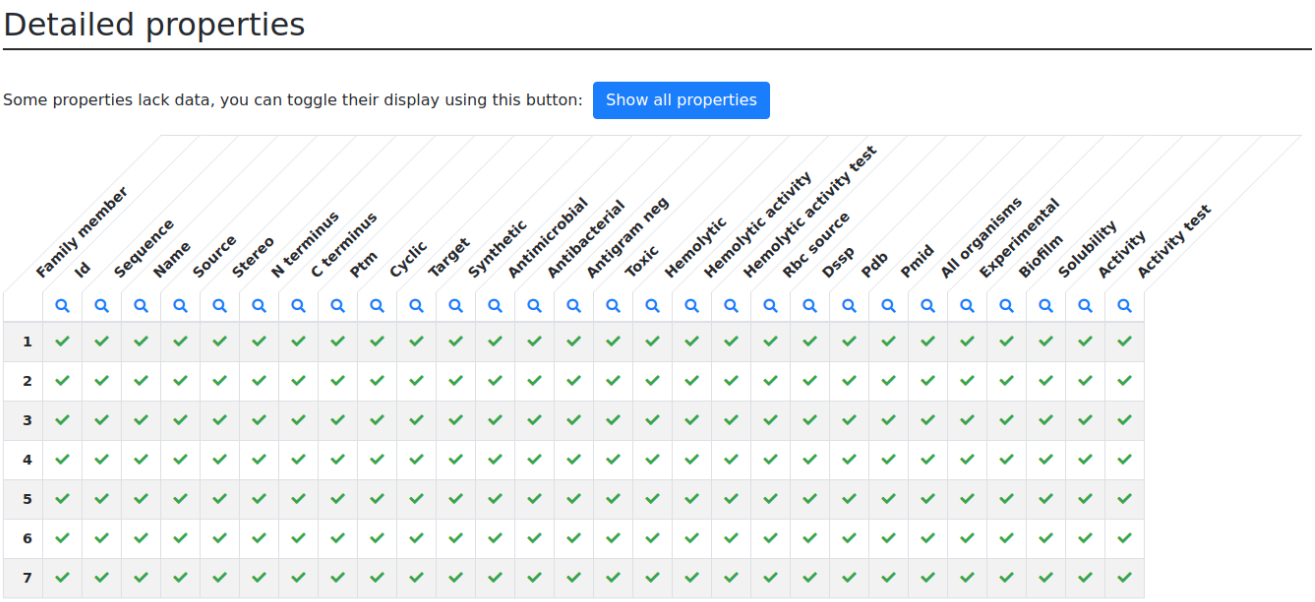
Legend:



1. *Father*
2. *96.4% similarity to father*
3. *92.9% similarity to father*
4. *89.3% similarity to father*
5. *64.3% similarity to father*
6. *64.3% similarity to father*
7. *42.9% similarity to father*

A table is also plotted here showing a summary of all the relevant properties:

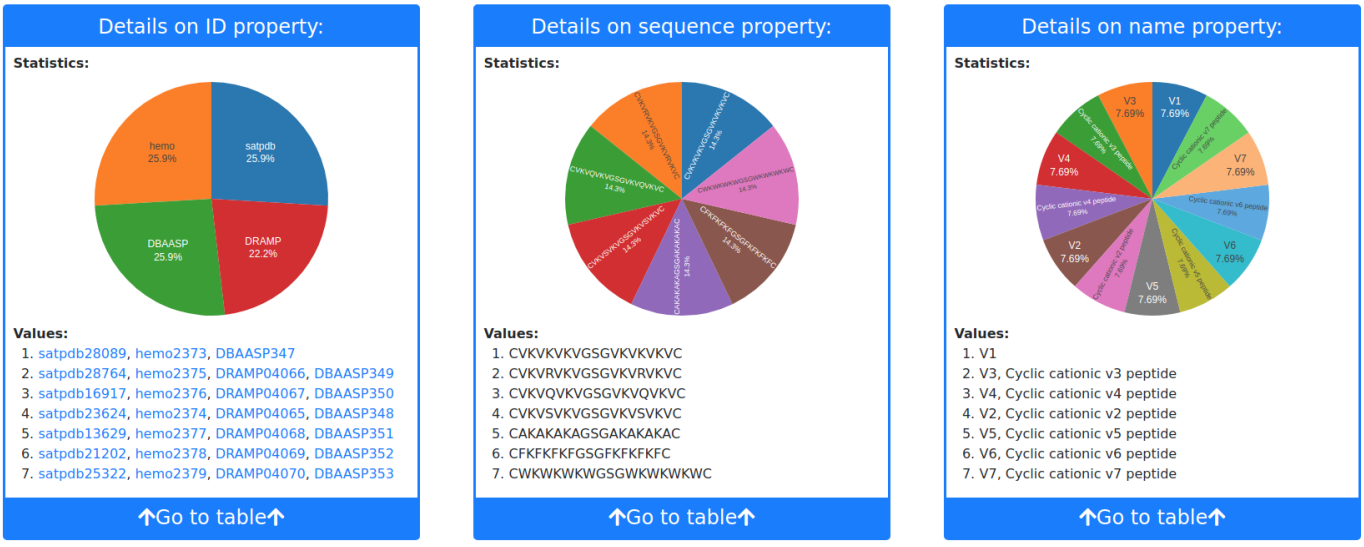
Screenshot 1.10.



We immediately see that these peptides are mostly cyclic, synthetic, actives against gram-negative organisms. We also see that this family tends to be hemolytic and toxic (column 38). Therefore, chemical modifications should be performed to reduce their toxicity.

After that table, some more details are listed:

Screenshot 1.11.



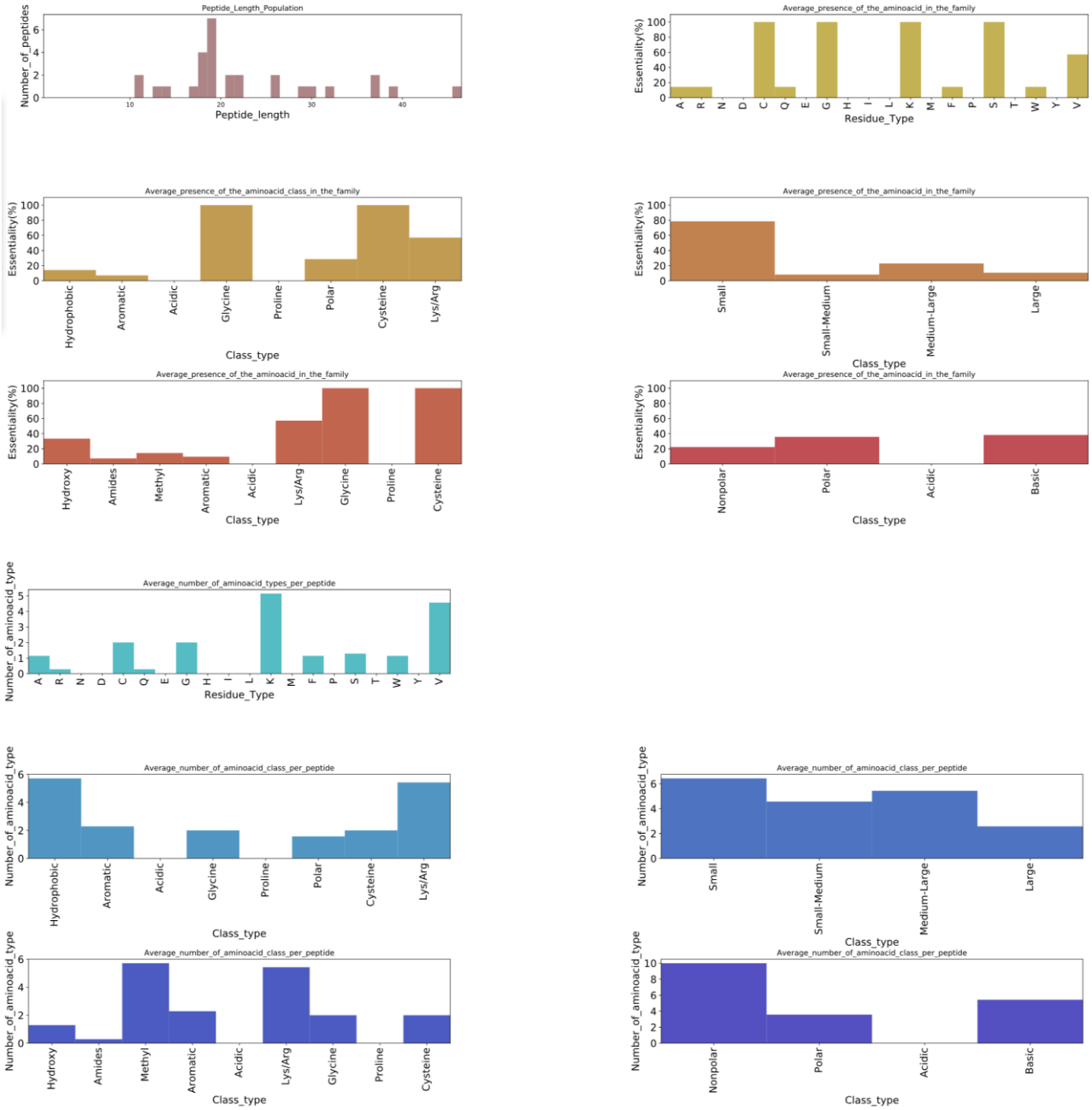
An additional graphical statistical analysis on the amino acid composition and location helps clarifying the properties of this family. The analysis can be found in the subdirectory Graphics in the "DOWNLOAD RESULTS" section on the left bar or simply through the link in "Global properties" section of "Family Analyzer".

Screenshot 1.12.

Name	Last modified	Size	Description
Graph1_1_fam1.svg	2019-02-06 15:30	291K	Graph plot files
Graph2_1_fam1.svg	2019-02-06 15:30	500K	Graph plot files
Graph3_1_fam1.svg	2019-02-06 15:30	474K	Graph plot files
Graph4_1_fam1.svg	2019-02-06 15:30	1.4M	Graph plot files
Graph5_1_fam1.svg	2019-02-06 15:30	771K	Graph plot files
Graph5_2_fam1.svg	2019-02-06 15:30	97K	Graph plot files
Graph6_1_fam1.svg	2019-02-06 15:30	474K	Graph plot files
Graph6_2_fam1.svg	2019-02-06 15:30	107K	Graph plot files

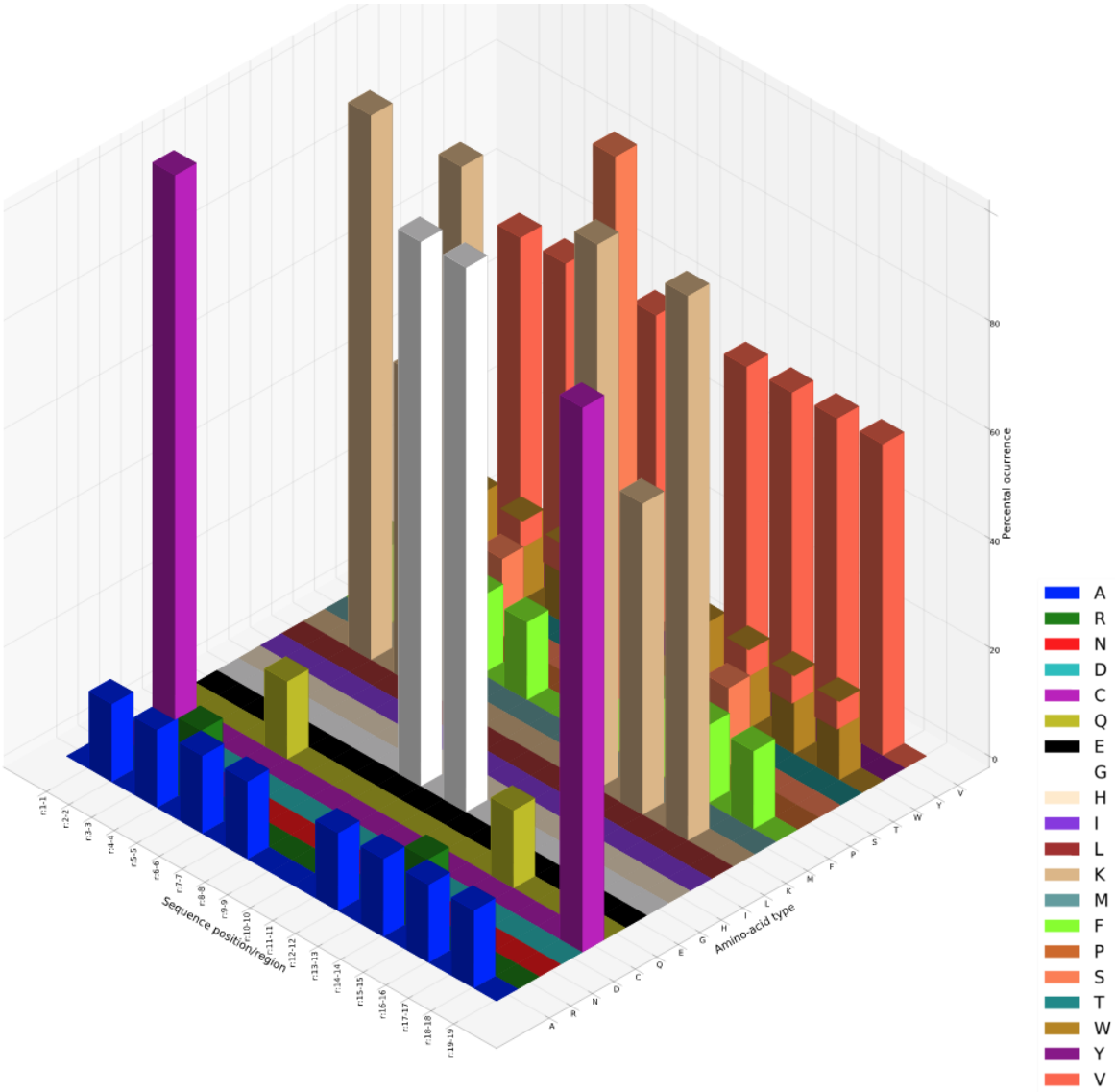
We find 6 different types of graphs. In graph 1, we can see that the members of family 1 contain on average 5 lysine residues (K) and 4.5 valines (V) but also 2 cysteines (C) per peptide. Hydrophobic and positively charged amino acids dominate the sequence and amino acid side chains tend to be of reduced size.

Screenshot 1.13.



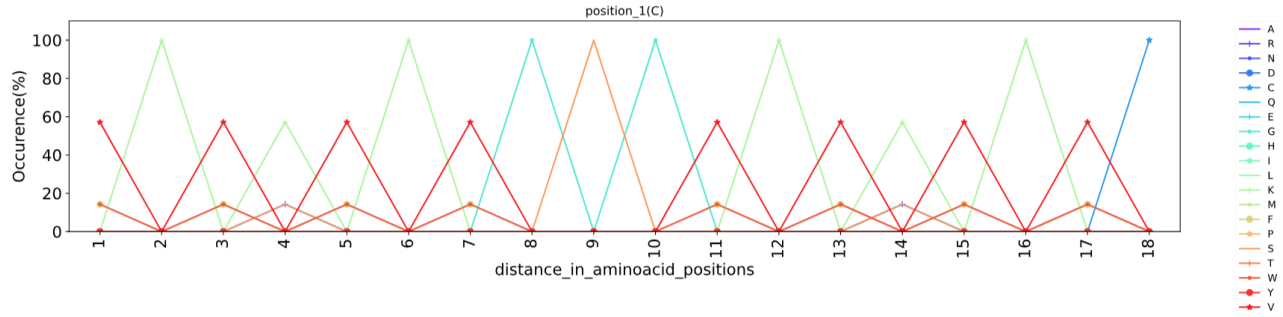
In graph 2, a 3D scheme (its 2D version is in graph 3) shows that the position of valines (green) and of lysines (cyan) along the sequence does not seem to matter. On the contrary, Cysteine (magenta) residues tend to be at the termini. This might suggest the presence of a head to tail disulphide bond (in the “data table” previously examined the peptides are actually reported to be cyclic).

Screenshot 1.14.



To confirm this hypothesis, we can check the probability to find a cysteine in last position when a Cysteine is present in the first position. This information can be retrieved from graph5 of family 1:

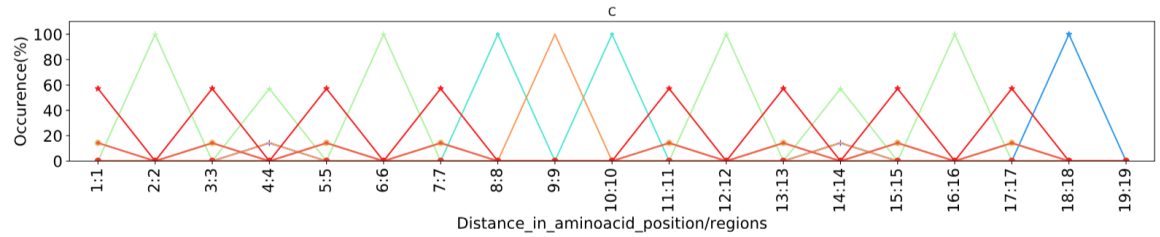
Screenshot 1.15.



In this graph, different amino acid types have different color code.
When a Cysteine is present in position 1 the probability to find another Cysteine (blue line) 18 amino acids (aa) apart, is 100%. This might sound as trivial information in the present case (these are synthetic peptides) but it might be very useful for peptides whose structure have not been studied yet.

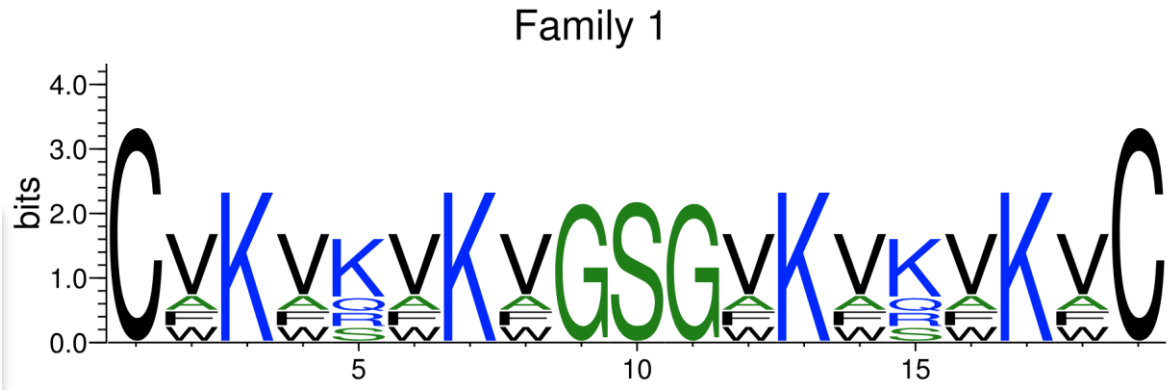
Graph 4 analyzes motifs (groups of amino acids in well-defined relative positions within the sequence), an important information often shedding light on the mechanism of action. Motifs are found by calculating the position-independent probability to find a residue type at a certain distance from a reference residue type (e.g. if you have an Alanine in a peptide of the family, you can calculate which is the probability to find each of the 20 amino acid types n positions apart).
In our case, when C is present, graph 4 states that we have a 100 % probability to find K (green) 2, 6, 12 and 16 positions apart, G (cyan) 8, 10 position apart, S 9 position apart and C 18 positions apart. This means that C-K---K-GSG-K---K-C is a common motif for family 1.

Screenshot 1.16.



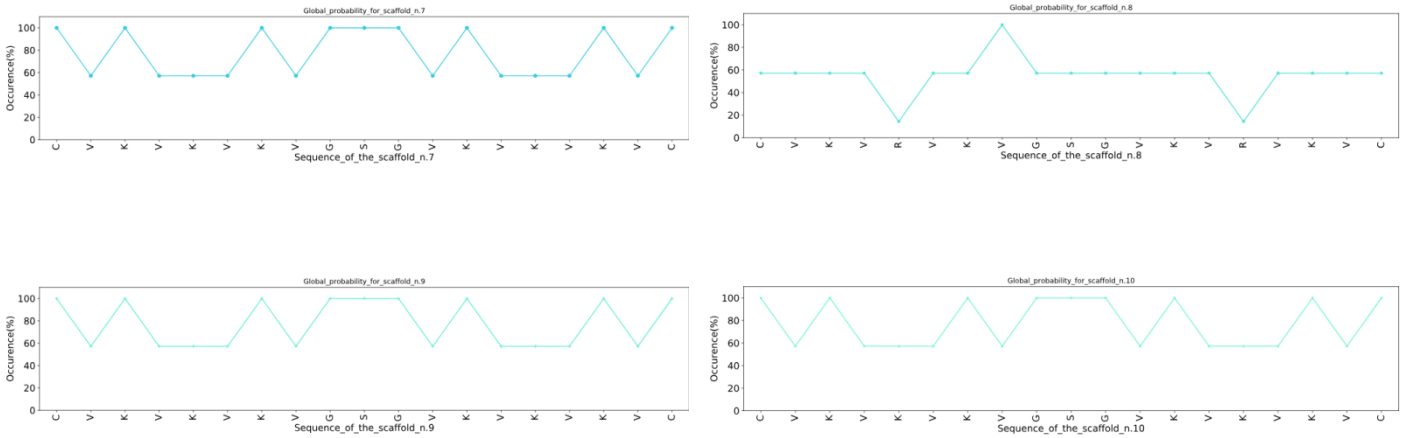
Among this family of peptides highly active against *Shigella* sp., we have highlighted the "father", which is the peptide which better represents the sequences of all other peptides of the family. Can we design a new sequence better resuming the properties of family 1, including the father?
Creating a peptide by concatenating the most frequent residue per position is a tempting but inefficient way to solve this problem. The resulting peptide would be a chimera where the synergic action of specific amino acid types in optimal relative positions would be likely lost. However, starting from the most abundant amino acid in each position, we can fill all other positions with the most frequent amino acid found in related peptides. As shown by the logo image, the most abundant amino acid in position 8 is Valine (you can find the logo image in the logos directory of "DOWNLOAD RESULTS" section). This feature of ADAPTABLE is powered by WebLogo/SeqLogo software; http://weblogo.threeplusone.com/ , Department of Plant and Microbial Biology, University of California, Berkeley):

Screenshot 1.17.



Analyzing all peptides with Valine in position 8 (see graph 6), we can find that the most frequent amino acid in position 1 is C, in position 2 is V and so on. As shown in the figure, we can do the same starting from position 7 or 9 or 10.

Screenshot 1.18.



In this way, we can create many different optimal peptide sequences, one per each position. The best representative peptide is chosen selecting the curve maximizing the sum of probabilities over all positions (the area of the graphs).

The optimal peptide can be found searching for “Best representative peptide”.

Screenshot 1.19.

Family 1:
Best representing sequence is sequence: CVKVKVKVGSGVKVKVKVC (>satpdb28089;hemo2373;DBAASP347;)

Secondary structure prediction
Alpha-helix ; Beta-strand; Turn ;

Helix
HHHHHHHH---HHHHHHHH
Strand
BBBBBBBBB-BBBBBBBBB
Turn
--------T----------
Consensus
=========^-=========;

DSSP prediction
Alpha-helix ; Alpha-310 ; Alpha-pi ; Beta-bridge; Beta-strand; Turn ; Bend ;

--^^^--^^^----^---;

You can also specify more parameters simply clicking on the box to expand it and show them.

Screenshot 1.20.

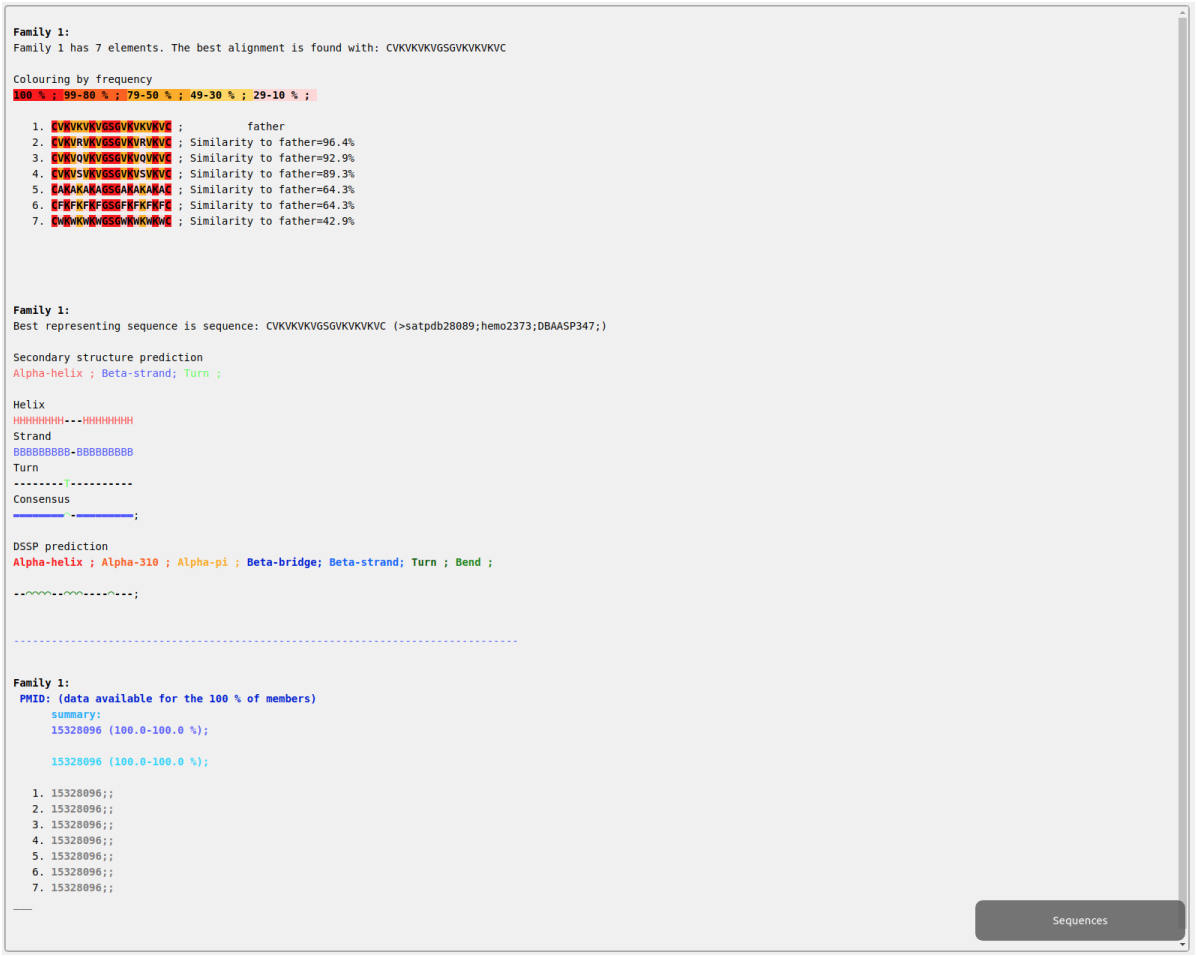
Specific parameters: (-) Collapse Me

ID sequence name source taxonomy Family gene stereo N_terminus C_terminus PTM cyclic ribosomal target all_organisms activity activity_test experimental
antimicrobial antiprotozoal insecticidal anticancer toxic cell_cell drug_delivery antioxidant antiproliferative antibacterial antigram_pos antigram_neg antifungal antiyeast
antiviral virus_name activity_viral activity_viral_test antiparasitic antiplasmodial antitrypanosomic antileishmania antitumor cell_line tissue cancer_type anticancer_activity
anticancer_activity_test antiangiogenic cytotoxic hemolytic hemolytic_activity hemolytic_activity_test RBC_source hormone quorum_sensing immunomodulant antihypertensive
cell_penetrating tumor_homing blood_brain DSSP pdb experim_structure PMID synthetic antibiofilm solubility

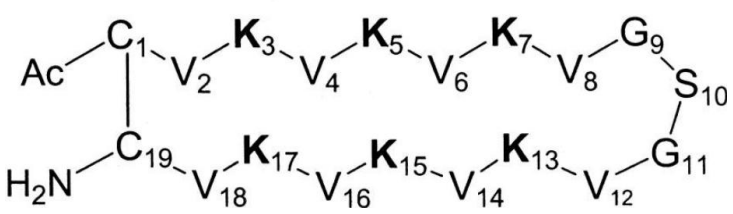
Submit Reset

For example, selecting “Aminoacid frequencies”, “Best representing peptide” and “PMID” (the reference to the publication on the selected peptides) you can visualize all this information in one page:

Screenshot 1.21.



Using the PMID 15328096 we can check the publication related to the family of peptides and have a confirmation of what we have “discovered”. You can also visit the page for getting the details of each family. The template peptide for this family of synthesized peptides, which was not present in the database, coincides with our optimal sequence. The family do has a disulphide bridge connecting the N and C terminus. No helical structure is present.



This image was taken from Frecer, V., Ho, B. & Ding, J. L. De novo design of potent antimicrobial peptides. Antimicrob. Agents Chemother. 48, 3349–3357 (2004).

ADAPTABLE TUTORIAL - CASE EXAMPLE n.2

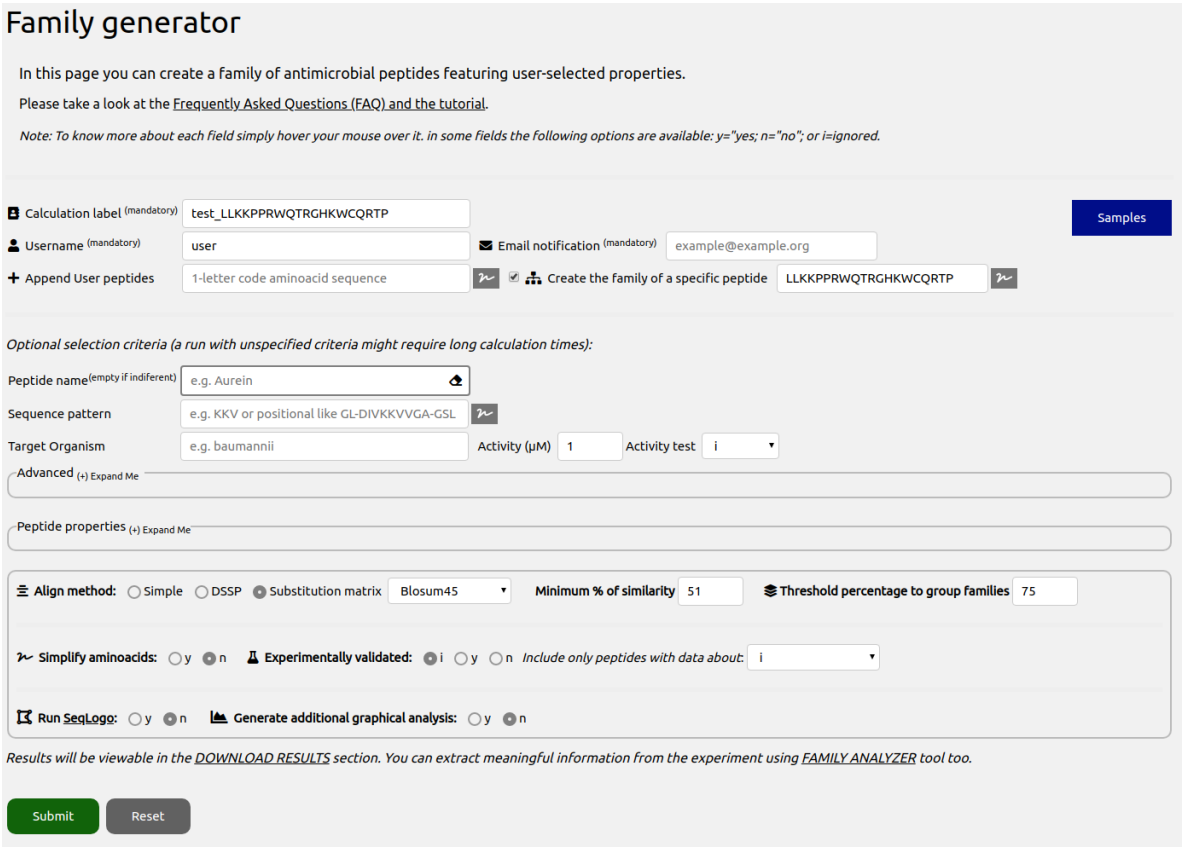
Predicting the antimicrobial or anticancer activity of a generic peptide sequence

The case of random new peptides

Suppose to have rationally designed an antimicrobial or anticancer peptide targeting microbial or cancerous cell membranes. Your sequence is entirely new and you want to predict its activity based on experimental data on similar peptides. Just for the sake of demonstrating the potentiality of ADAPTABLE in testing new sequences, we will use a completely random sequence: LLKKPPRWQTRGHKWCQRTP .

In order to predict the activity we will tell ADAPTABLE to "Create the family of a specific peptide" using this one as reference. We will give the name "test_LLKKPPRWQTRGHKWCQRTP" to calculation.

Screenshot 2.1.

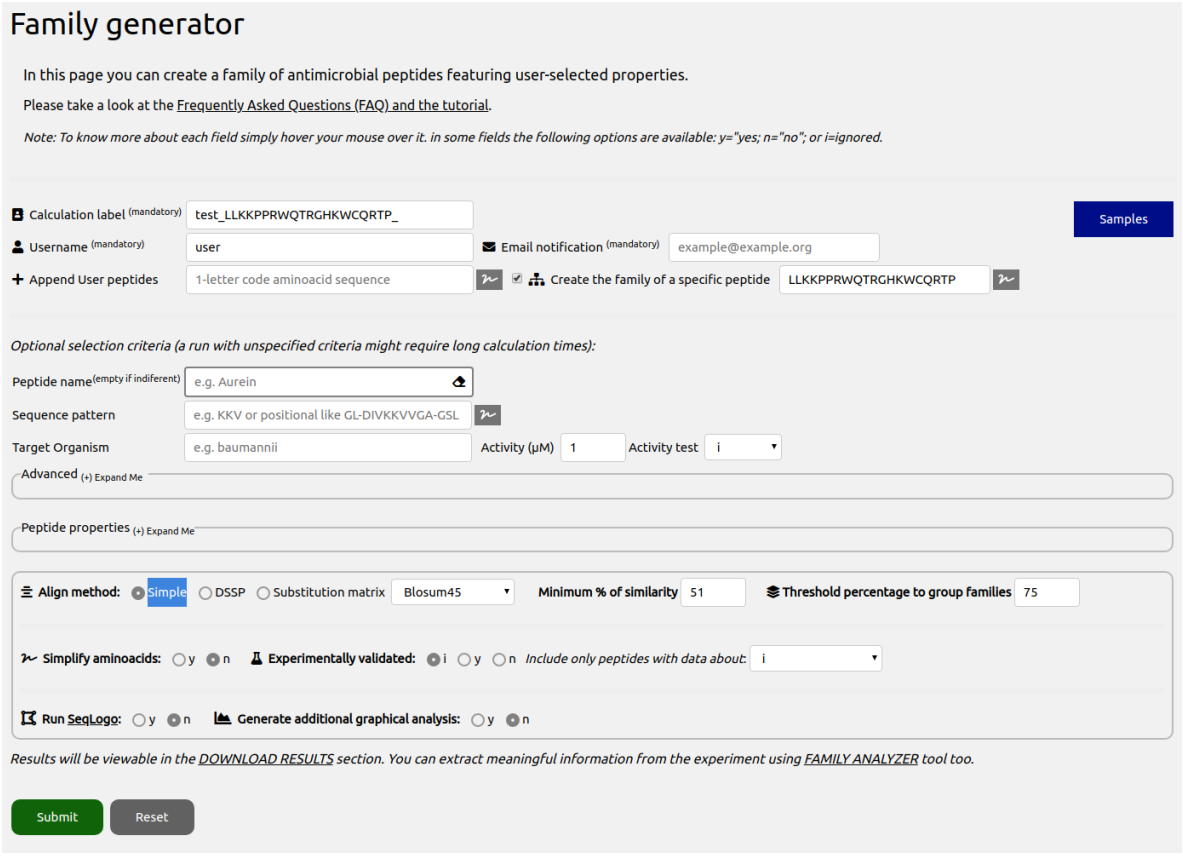


This tool allows to compare a test sequence (in our case LLKKPPRWQTRGHKWCQRTP), with the full database. In our case, running the prediction did not produce any interesting result. Despite the fact that our sequence contains many amino acids frequently found in antimicrobial peptides (K, A, V, R), the sequence did not resemble any of the peptides of our database. This, of course, does not mean that our random peptide does not exhibit antimicrobial or anticancer properties. However, if active, we can estimate that its mechanism of action is not common in nature.

In order to get more insight in the possible activity of our test peptide, the calculation can be repeated "in simplified space" (setting to "simple" the align method), where residue types are compared by their properties. In this way, hydrophobic residues (A, V, I, L, M) are represented by A,

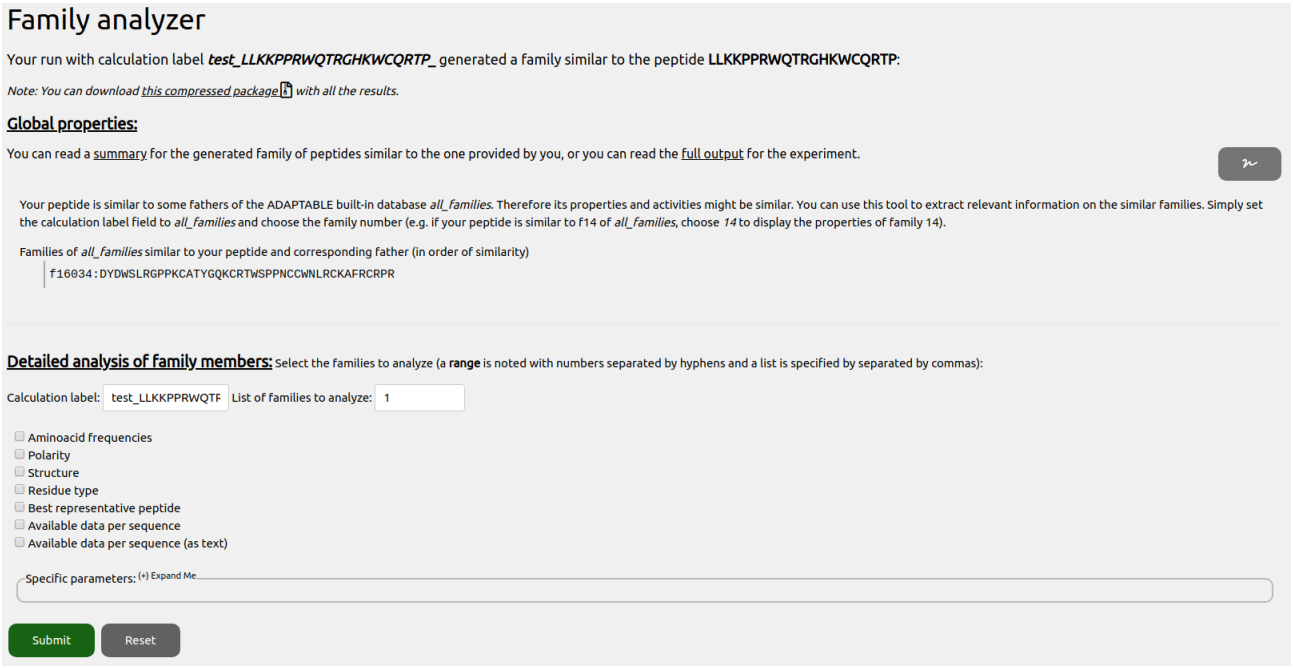
negative residues (D, E) by D, positive residues (K, R) by K, aromatic residues (W, Y, H, F) by F, polar residues (S, T, N, Q) by S, modified amino acids by (Ⓜ). Gly, Pro, Cys are treated individually.

Screenshot 2.2.



After the calculation, a page summarizing the features of the generated family will be shown. We can analyze the results more deeply by going to the "FAMILY ANALYZER" page and typing our user we can choose our run test_LLKKPPRWQTRGHKWCQRTP_:

Screenshot 2.3.



Our test peptide was able to generate a family of 3 similar peptides (the family can be visualized by clicking on the "Submit" button in the lower part of the page with the option "Aminoacid frequencies" checked). Remember that when running the experiment you can tune the minimum similarity to get peptides included in the family, and change the alignment matrix and method:

Screenshot 2.4.

In the family, the first sequence is our test peptide, followed by similar database sequences with decreasing similarity. We can now click on the "Up arrow" button in the upper part and re-run the analyzing tool to retrieve other information.

In the *Screenshot 2.3*, ADAPTABLE also reports that the test peptide is similar to the father of the family 16064 from "all_families" experiment (a built-in experiment that aims to group all the peptides into families with similar sequence). Therefore, it would be interesting to see its properties. In order to do so, we can use "all_families" as calculation label and we can type "16034" in the "List of families to analyze" field, as in *Screenshot 2.5*.

Screenshot 2.5.

Family 16034:
Family 16034 has 2 elements. The best alignment is found with: DYDWSLRGPPKCATYGQKCRTWSPPNCCWNLRCKAFRCRPR

Colouring by frequency
100 % ; 99-80 % ; 79-50 % ; 49-30 % ; 29-10 % ;

1. **DYDWSLRGPPKCATYGQKCRTWSPPNCCWNLRCKAFRCRPR** ; father
2. **DYDWSLRGPPKCATYGQKCRTWSPRNCCWNLRCKAFRCRPR** ; Similarity to father=96.6%

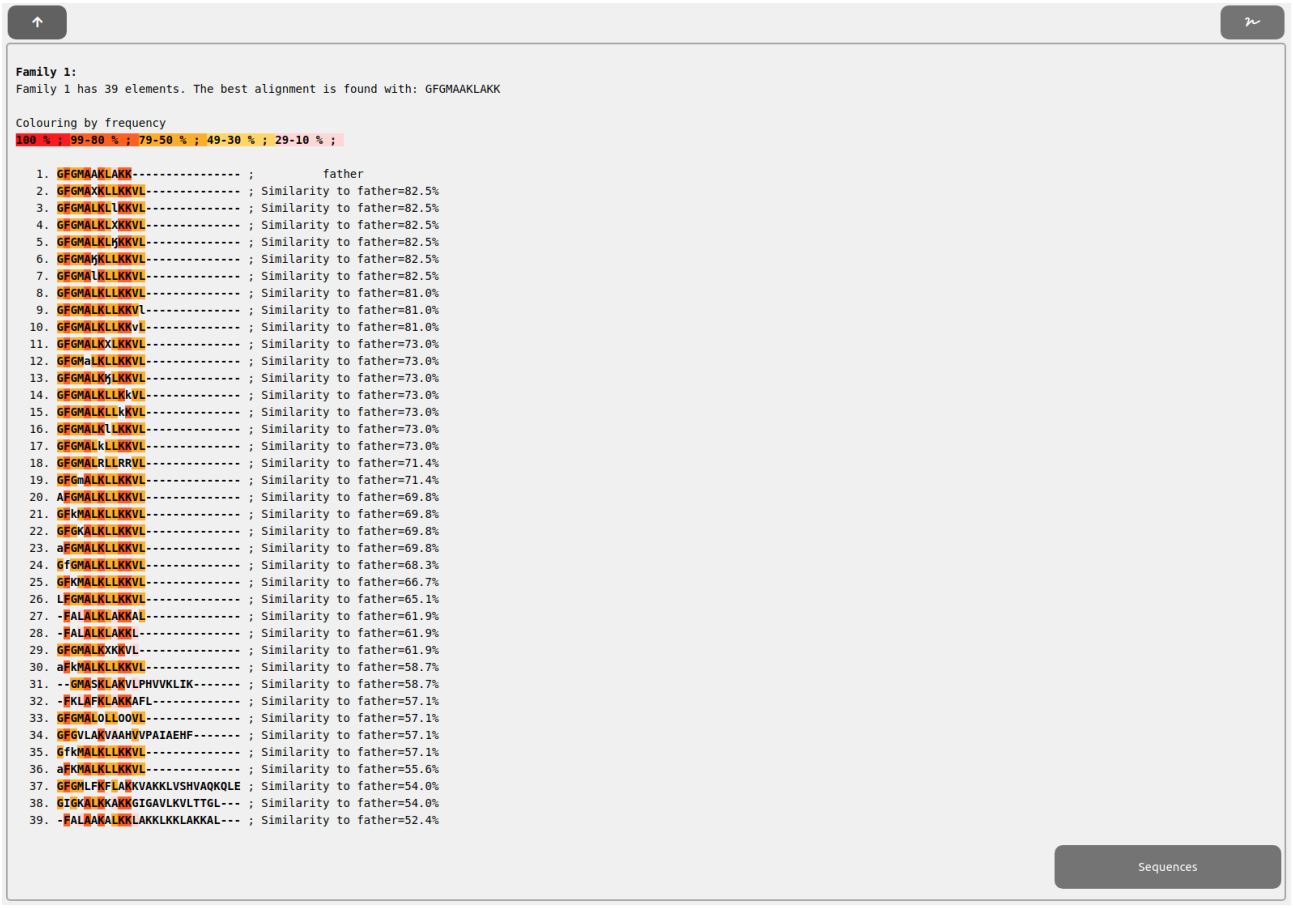
Selecting “Available data per sequence” and clicking on “Submit”, we can see that the similar family 16034 is composed of members with antibacterial activity.

To conclude, out test peptide might have antibacterial activity although its sequence is not strikingly similar to any entry of the database.

Let's now try with a different peptide: GFGMAAKLAKK.

In this case, going to the "FAMILY ANALYZER" section and looking for "Aminoacid frequencies" information, we observe something more interesting:

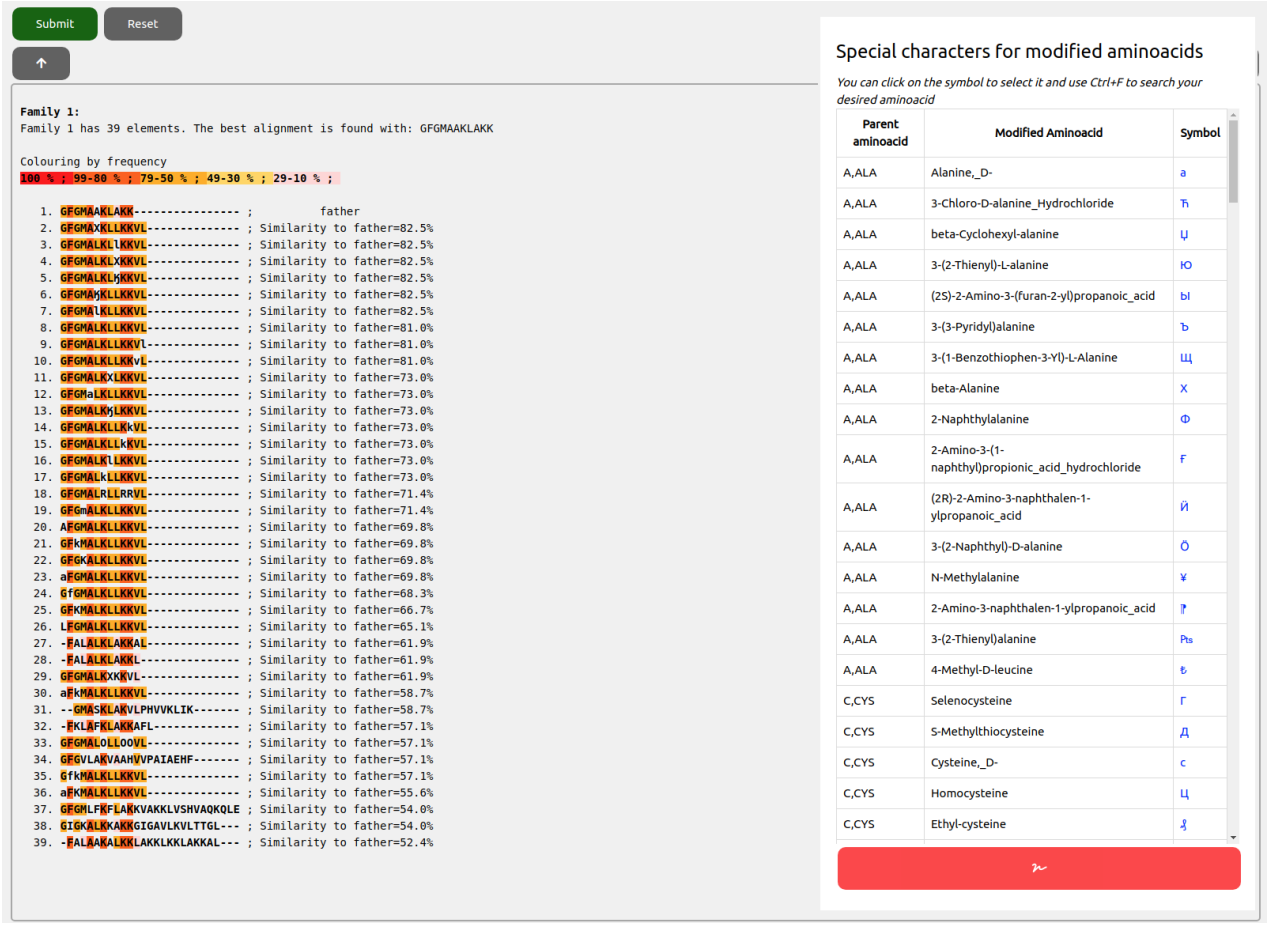
Screenshot 2.8.



Our peptide (peptide 1) is quite similar to other peptides in the databases. Some of them (like peptide number 5 or 13) have sequences with non-standard aminoacids. This aminoacids are handled in the same way as regular aminoacids and represented by different characters to recognize them. To try to handle them more easily, we provide a button when necessarily next to the fields/sections that would need it (Screenshot 2.9). The button will open a section (Screenshot 2.10) allowing you to search for characters or aminoacids (pressing Ctrl+F), and to paste those characters in relevant fields when needed (simply clicking on them will paste the characters to the field next to the button used to open the character picker).

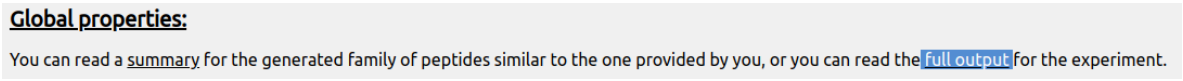
Screenshot 2.9.

Screenshot 2.10.



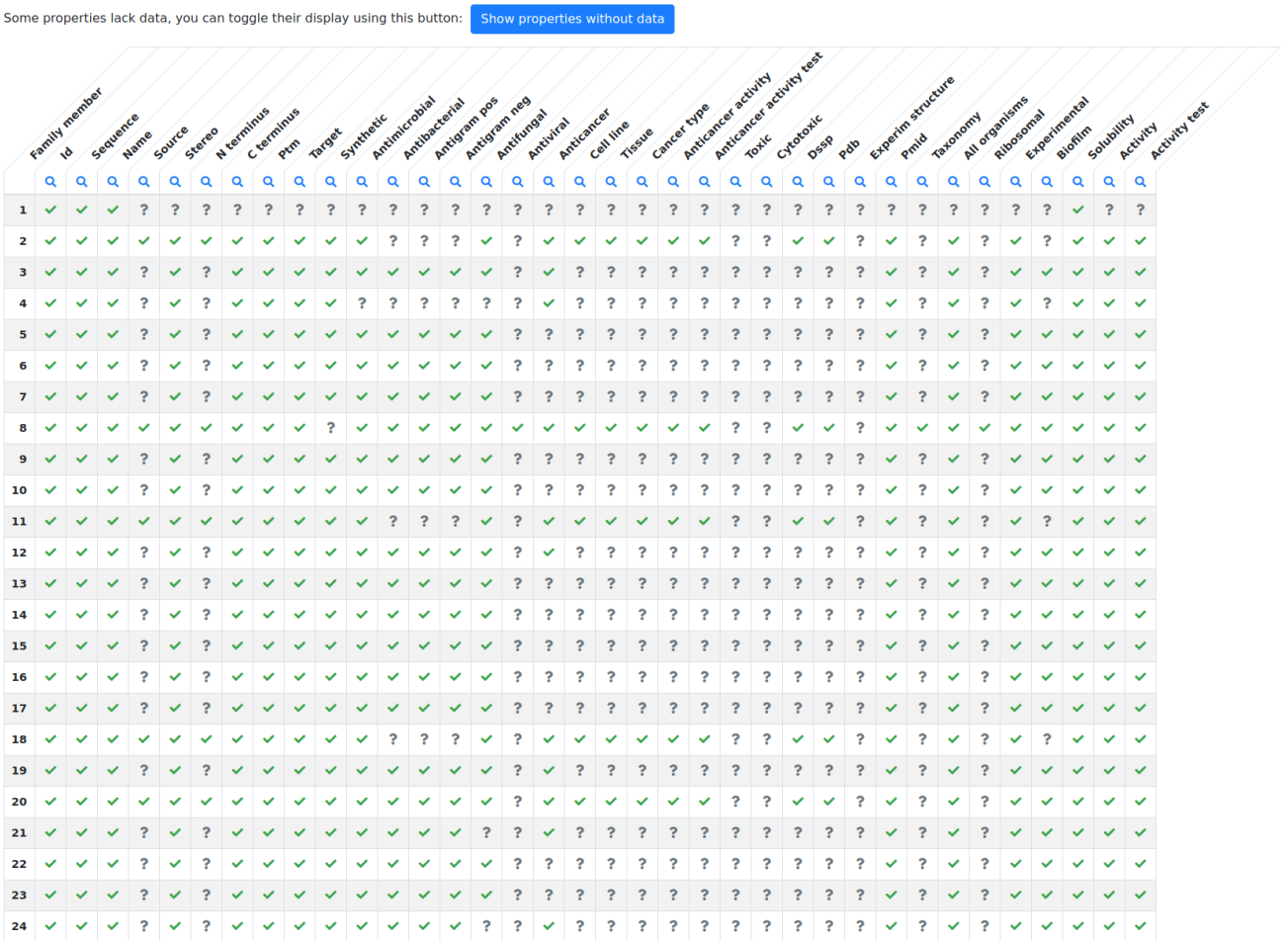
To find more information about the properties of the similar peptides we have several options: rely on FAMILY ANALYZER utilities to review each field we are interested in with full detail, click on "Available data per sequence" to show a table showing a summary or rely on the more readable table provided into the Full output html. We will opt for this last option and, to open it, we simply need to scroll up to "Global properties" section and click on "full output" link:

Screenshot 2.11.



Then, we click into "Property details" and we will see the table.

Screenshot 2.12.



We discover that the peptides similar to our test sequence are antifungal and anticancer. In the list of similar peptides, the order follows the degree of similarity. This means that the most similar molecule is peptide 2, which is also antimicrobial. Given the sequence similarity we can predict that our peptide is likely to be antifungal and anticancer, possibly also antimicrobial then.

Looking into “Specific parameters” section, we can find more detailed information on a possible “cell_line”, including “tissue” and “cancer_type”.

Screenshot 2.13.

```
Family 1:
 cell_line: (data available for the 33.3333 % of members)
      summary:
      hela (33.3-100.0 %); ccrf (23.1-69.2 %); cem (23.1-69.2 %); mcf (10.3-30.8 %); 480 (10.3-30.8 %); bmkc (10.3-30.8 %); 1299 (10.3-30.8 %);

      hela (33.3-100.0 %); ccrf-cem (23.1-69.2 %); mcf-7 (10.3-30.8 %); sw-480 (10.3-30.8 %); bmkc (10.3-30.8 %); h-1299 (10.3-30.8 %); pc-3 (10.3-30.8 %);

   1.
   2. HeLa;SW;CCRF-CEM;;
   3.
   4.
   5.
   6.
   7.
   8. HeLa;SW;CCRF-CEM;;
   9.
  10.
  11. HeLa;SW;CCRF-CEM;;
  12.
  13.
  14.
  15.
  16.
  17.
  18. HeLa;SW;CCRF-CEM;;
  19.
  20. HeLa;SW;CCRF-CEM;;
  21.
  22.
  23.
  24.
  25. HeLa;SW;CCRF-CEM;;
  26. HeLa;SW;CCRF-CEM;;
  27. MCF-7;SW-480;BMKC;H-1299;PC-3;HeLa;;
  28. MCF-7;SW-480;BMKC;H-1299;PC-3;HeLa;;
  29. HeLa;SW;CCRF-CEM;;
  30.
  31.
  32. MCF-7;SW-480;BMKC;H-1299;PC-3;HeLa;;
  33. HeLa;SW;CCRF-CEM;;
  34.
  35.
  36.
  37.
  38.
  39. MCF-7;SW-480;BMKC;H-1299;PC-3;HeLa;;
___


Family 1:
 tissue: (data available for the 33.3333 % of members)
      summary:
      cervix (33.3-100.0 %); colon (33.3-100.0 %); blood (23.1-69.2 %); breast (10.3-30.8 %); skin (10.3-30.8 %); lung (10.3-30.8 %); prostate (10.3-30.8 %);

      cervix (33.3-100.0 %); colon (33.3-100.0 %); blood (23.1-69.2 %); breast (10.3-30.8 %); skin (10.3-30.8 %); lung (10.3-30.8 %); prostate (10.3-30.8 %);
```

Following the list, we can predict the activity with decreasing accuracy. For example, our compound has relatively good chances to act on HeLa, SW and CCRF-CEM cell lines for Cervix, Colon and Lung cancers (properties of entry 2).

The output (section “Source”) also indicates that our compound can be considered part of the family of Macropins peptides:

Screenshot 2.11.

```
Family 1:
 source: (data available for the 35.8974 % of members)
      summary:
      macropins (23.1-64.3 %); their (23.1-64.3 %); analogs (23.1-64.3 %); flak (10.3-28.6 %); peptides (10.3-28.6 %); venom ( 2.6- 7.1 %); solitary ( 2.6- 7.1 %); fulvipes (
      2.6- 7.1 %); beemacropis ( 2.6- 7.1 %); bee ( 2.6- 7.1 %); macropis ( 2.6- 7.1 %); colletotrichum ( 2.6- 7.1 %); gloeosporioides ( 2.6- 7.1 %); nara ( 2.6- 7.1 %); gc5 (
      2.6- 7.1 %); cupiennius ( 2.6- 7.1 %); salei ( 2.6- 7.1 %); wandering ( 2.6- 7.1 %); spider ( 2.6- 7.1 %);

      macropins_and_their_analogs (23.1-64.3 %); flak_peptides (10.3-28.6 %); venom ( 2.6- 7.1 %); _the_solitary_beemacropis_fulvipes ( 2.6- 7.1 %); _the_solitary_bee_macropis
      _fulvipes ( 2.6- 7.1 %); colletotrichum_gloeosporioides_nara_gc5 ( 2.6- 7.1 %); cupiennius_salei__wandering_spider ( 2.6- 7.1 %);

   1.
   2. Macropins_and_their_analogs;;
   3.
   4.
   5.
   6.
   7.
   8. Macropins_and_their_analogs;Venom,_the_solitary_beeMacropis_fulvipes;venom,_the_solitary_bee_Macropis_fulvipes;Colletotrichum_gloeosporioides_Nara_gc5;;
```

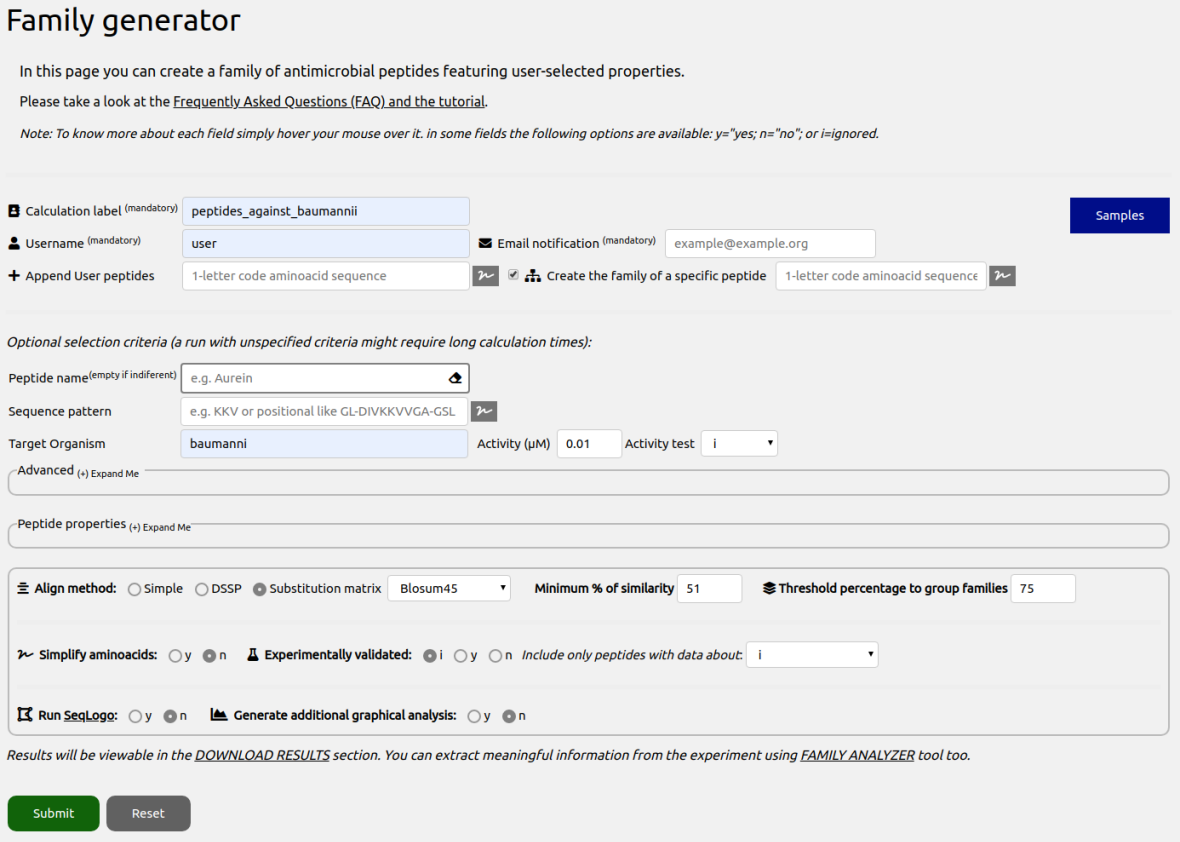
ADAPTABLE TUTORIAL - CASE EXAMPLE n.3

Discovering new activities of pre-existing sequences not tested experimentally

The case of *Acinetobacter baumannii*

Despite the huge amount of experimental data on antimicrobial peptides, databases contain only a limited number of active peptides against some specific organisms. For example, when searching for peptides active against *A. baumannii*, if we look into peptides with an activity of not more than 0.01µM to kill the target organisms we obtain only 2 families formed by only 1 aminoacid each.

Screenshot 3.1.



A. baumannii has been declared among the most dangerous pathogen in terms of antibiotic resistance by the World Health Organization (WHO). It is therefore very important to find new active compounds which can be used as drugs. To find new peptides that could be similar to this two ADAPTABLE can help in this task. For example, focalizing on peptide 1 (GWLKKIGKKIERVGQHTRDATIQTIGVAQQAANVAATLK), we can use the "Create the family of a specific peptide" feature (see details in the previous example) to find interesting peptides. It allows to compare a given sequence (in our case GWLKKIGKKIERVGQHTRDATIQTIGVAQQAANVAATLK), with our full AMPs database. When running the calculation we obtain 31 potentially active peptides which were never tested with *A. Baumannii* as can be seen in the summary page that is offered shown after the experiment run. That page summarizes the properties (*screenshot 3.2*) and lists the members (*screenshot 3.3*) of the family of peptides similar to GWLKKIGKKIERVGQHTRDATIQTIGVAQQAANVAATLK.

Screenshot 3.2.

Properties of the family of peptides similar to GWLKKIGKKIERVGQHTRDATIQTIGVAQQAANVAATLK

You can visit 'DOWNLOAD RESULTS' section to download all the data.

- Members of the family

Property	Value (% of members that could have it depending on available data per member)
Activity:	
antimicrobial: (data available for the 96.77 % of members)	antimicrobial (96.8-100.0 %)
antibacterial: (data available for the 87.09 % of members)	antibacterial (87.1-100.0 %)
antigram_pos: (data available for the 83.87 % of members)	antigram_pos (83.9-100.0 %)
antigram_neg: (data available for the 83.87 % of members)	antigram_neg (83.9-100.0 %)
biofilm: (data available for the 51.61 % of members)	biofilm (51.6-100.0 %); staphylococcus (38.7-75.0 %); micrococcus (35.5-68.8 %); luteus (35.5-68.8 %); aureus (35.5-68.8 %); pseudomonas (29.0-56.2 %); candida (25.8-50.0 %); aeruginosa (25.8-50.0 %); mycobacterium (25.8-50.0 %); smegmatis (25.8-50.0 %); albicans (25.8-50.0 %); enterobacter (16.1-31.2 %); dsm (19.4-37.5 %); 511 (19.4-37.5 %); 987 (19.4-37.5 %); cloacae (16.1-31.2 %); bacillus (12.9-25.0 %); megaterium (12.9-25.0 %); klebsiella (6.5-12.5 %); ot97 (9.7-18.8 %); serratia (9.7-18.8 %); marcescens (9.7-18.8 %); mi11 (9.7-18.8 %); ptcc (3.2- 6.2 %); pneumoniae (6.5-12.5 %); db11 (6.5-12.5 %); bm11 (6.5-12.5 %); b12 (6.5-12.5 %); shigella (6.5-12.5 %); sonnei (6.5-12.5 %); oxytoca (3.2- 6.2 %); aerogenes (3.2- 6.2 %); acinetobacter (3.2- 6.2 %); pittii (3.2- 6.2 %); nctc (3.2- 6.2 %); epidermidis (3.2- 6.2 %); β12 (3.2- 6.2 %); streptococcus (3.2- 6.2 %); pyogenes (3.2- 6.2 %); glabrata (3.2- 6.2 %); dbl (3.2- 6.2 %); fda (3.2- 6.2 %); atcc (3.2- 6.2 %); js11746 (3.2- 6.2 %); fda209p (3.2- 6.2 %); k799 (3.2- 6.2 %); fluorescens (3.2- 6.2 %); edwardsiella (3.2- 6.2 %); tarda (3.2- 6.2 %)
antifungal: (data available for the 35.48 % of members)	antifungal (35.5-100.0 %)
antiviral: (data available for the 9.67 % of members)	antiviral (9.7-100.0 %)
anticancer: (data available for the 3.22 % of members)	anticancer (3.2-100.0 %)
- cancer_type: (data available for the 3.22 % of members)	leukemia (3.2-100.0 %)
Toxicity:	
toxic: (data available for the 3.22 % of members)	toxic (3.2-100.0 %)
cytotoxic: (data available for the 3.22 % of members)	cytotoxic (3.2-100.0 %)
hemolytic: (data available for the 3.22 % of members)	hemolytic (3.2-100.0 %)
Source:	
name: (data available for the 100 % of members)	cecropin (71.0-71.0 %); sarcotoxin (22.6-22.6 %); insects (41.9-41.9 %); animals (41.9-41.9 %); precursor (29.0-29.0 %); lser (19.4-19.4 %); cec (9.7- 9.7 %); altname (9.7- 9.7 %); full (9.7- 9.7 %); drosophila (9.7- 9.7 %); a1a2 (6.5- 6.5 %); drovi (9.7- 9.7 %); mdc (6.5- 6.5 %); cec1 (6.5- 6.5 %); lucilin (6.5- 6.5 %); flags (6.5- 6.5 %); cec3 (6.5- 6.5 %); ceca1 (6.5- 6.5 %); drose (6.5- 6.5 %); cec2 (6.5- 6.5 %); gi23259 (6.5- 6.5 %); sarpe (6.5- 6.5 %); drosi (6.5- 6.5 %); cecb (6.5- 6.5 %)
source: (data available for the 96.77 % of members)	fly (77.4-80.0 %); drosophila (35.5-36.7 %); fruit (38.7-40.0 %); lucilia (29.0-30.0 %); sericata (22.6-23.3 %); peregrina (16.1-16.7 %); sarcophaga (16.1-16.7 %); flesh (12.9-13.3 %); melanogaster (19.4-20.0 %); ceratitis (6.5- 6.7 %); capitata (6.5- 6.7 %); musca (12.9-13.3 %); domestica (12.9-13.3 %); blow (19.4-20.0 %); morsitans (6.5- 6.7 %); sechellia (12.9-13.3 %); virilis (9.7-10.0 %); simulans (9.7-10.0 %); tsetse (6.5- 6.7 %); glossina (6.5- 6.7 %); medfly (6.5- 6.7 %); mediterranean (6.5- 6.7 %); boettcherisca (9.7-10.0 %); mauritiana (9.7-10.0 %); mojavensis (6.5- 6.7 %); tik (6.5- 6.7 %); flies (6.5- 6.7 %)
taxonomy: (data available for the 96.77 % of members)	animalia (93.5-96.7 %); insects (87.1-90.0 %); arthropoda (71.0-73.3 %); arthropods (54.8-56.7 %); invertebrates (54.8-56.7 %); animals (54.8-56.7 %); insecta (38.7-40.0 %); eumetazoa (41.9-43.3 %); diptera (35.5-36.7 %); sarcophaga (12.9-13.3 %); drosophila (12.9-13.3 %); cecropins (19.4-20.0 %); sarcophagidae (12.9-13.3 %); ceratitis (6.5- 6.7 %); peregrina (12.9-13.3 %); eukaryota (9.7-10.0 %); metazoa (9.7-10.0 %); ecdysozoa (9.7-10.0 %); hexapoda (9.7-10.0 %); pterygota (9.7-10.0 %); neoptera (9.7-10.0 %); holometabola (9.7-10.0 %); brachycera (9.7-10.0 %); muscomorpha (9.7-10.0 %); oestroidea (9.7-10.0 %); drosophilidae (9.7-10.0 %); melanogaster (9.7-10.0 %); cecropin (6.5- 6.7 %); tephritidae (6.5- 6.7 %); capitata (6.5- 6.7 %); boettcherisca (6.5- 6.7 %)
Family: (data available for the 54.83 % of members)	cecropin (54.8-100.0 %)

Show frequency of aminoacids

Screenshot 3.5.

Members of the family of peptides similar to GWLKKIGKKIERVGQHTRDATIQTIGVAQQAANVAATLK

- Properties of the family

Num	Sequence (Click each for more information)	ID
1	**-----GWLKKIGKKIERVGQHTRDATIQTIGVAQQAANVAATLK-**	**My Peptide**
2	-----GWLKKIGKKIERVGQHTRDATIQTIGVAQQAANVAATLK-	satpdb16356 satpdb18295 canPPD4420 DRAMP03077 CAMPSQ1244 DBAASP11810 APD_AP00135 InverPep97 LAMPL01A002386 ADAM_2948 YADAMP2322
3	GSPEFGWLKKIGKKIERVGQHTRDATIQTIGVAQQAANVAATLKG	satpdb21418 CAMPSQ625 InverPep618 LAMPL01A000665 ADAM_6656
4	-----GWLKKIGKKIERVGQHTRDATIQTIGVAQQAANVAATLKG	satpdb27249 DBAASP6281 APD_AP02498 InverPep534 LAMPL07APD0040 ADAM_2949
5	-----GWLKKLGKKIERVGQHTRDATIQTIGVAQQAVNVAATLKG	DBAASP11570 APD_AP03009
6	-----GWLKKFGKKIERVGQHTRDATIQAIGVAQQAANVAATLKG	satpdb17922 DBAASP8177 APD_AP02503 InverPep539
7	-----GWLKKIGKKIERVGQHTRDATIQVIGVAQQAANVAATAR-	satpdb21316 DRAMP03102 CAMPSQ126 APD_AP00231 InverPep280 LAMPL01A000138 ADAM_2950 YADAMP2323
8	-----GWLKKIGKKIERVGQHTRDATIQTIAVAQQAANVAATAR-	satpdb18232 DRAMP03076 CAMPSQ513 APD_AP00132 InverPep96 LAMPL01A000550 ADAM_2947 YADAMP2321
9	-----GWLKKIGKKIERVGQHTRDATIQVLGVAQQAANVAATARG	satpdb13980 DBAASP8174 APD_AP02500 InverPep536
10	-----GWLKKIGKKIERVGQHTRDATIQGLGVAQQAANVAATAR-	DRAMP03079 CAMPSQ1265 CAMPSQ2136 LAMPL01A002409
11	-----GWLKKIGKKIERVGQHTRDASIQAIGIAQQAANVAATARG	satpdb20523 DBAASP8175 APD_AP02501 InverPep537
12	-----GWLKKIGKKIERVGQHTRDATIQGLGIAQQAANVAATAR-	satpdb10156 satpdb18038 DRAMP03089 CAMPSQ125 CAMPSQ1269 DBAASP1855 APD_AP00230 InverPep279 LAMPL01A000137 ADAM_2945 YADAMP2320
13	-----GWLRKIGKKIERVGQHTRDATIQVLGIAQQAANVAATAR-	satpdb20196 DRAMP03101 CAMPSQ127 APD_AP00232 InverPep281 LAMPL01A000139 ADAM_2956 YADAMP2326 uniprotP08377
14	-----GWLKKIGKKIERVGQHTRDATIQGLGIAQQAANVAATARG	satpdb21100 CAMPSQ886 DBAASP5153 InverPep124 LAMPL01A002923 ADAM_2946 YADAMP2367
15	-----GWLKKIGKKIERIGQHTRDATIQGVGIAQQAANVAATAR-	InverPep121 LAMPL01A002414
16	-----GWLKKIGKKIERIGQHTRDATIQGLGIAQQAANVAATAR-	DRAMP03114 CAMPSQ1271 InverPep126 LAMPL01A002415
17	-----GWIRDFGKRIERVGQHTRDATIQTIAVAQQAANVAATLKG	satpdb12541 DRAMP03103 CAMPSQ128 APD_AP00233 InverPep282 LAMPL01A000140 ADAM_2932 YADAMP2365 uniprotP18312
18	-QSEAGWLKKIGKKIERVGQHTRDATIQGLGVAQQAPNVAATAR-	DRAMP03117 CAMPSQ1268 LAMPL01A002412

Show frequency of aminoacids

A quick look to the data table (“Available data per sequence”) reveals that some of the new potentially active peptides are not hemolytic. In the table, each line corresponds to one peptide and each column corresponds to a property (1 indicates presence of the property). A “1” in column 37 indicates that the peptide has been tested experimentally for its hemolytic activity, while column 38 expresses whether the peptide is hemolytic (value 1) or not (value 0). In the former case, the intensity of the blue color indicates the extent of the hemolytic activity (**(sub-µM)-(µM)-(sub-mM)-(mM)-(very weak)**).

This way we can select peptides with activities we want or activities we want to skip (like toxic or hemolytic activities).

Screenshot 3.7.