Department of Medical Protein Research, VIB, B-9000 Ghent, Belgium
Department of Biochemistry, Ghent University, B-9000 Ghent, Belgium
http://www.proteomics.be
http://www.compomics.com

TOPPR MANUAL

Niklaas Colaert

http://iomics.ugent.be/toppr/

Contents

Introduction

The Online Protein Processing Resource (TOPPR)

Over two percent of all human and mouse genes encode for proteases. This large group of enzymes controls many biological processes and they are crucially important in relatively simple processes such as food digestion as well as in highly regulated processes such as controlled cell death. In addition, proteases add to several pathologies including cancer, cardiovascular and inflammatory diseases. It is commonly recognized that a more detailed understanding of protease-controlled processes can be achieved by extending our overall knowledge on proteases, their (preferred) substrates and specificities.

The N-terminal COFRADIC (COmbinded FRActional DIagonal Chromatography) technique developed in our lab enables isolation and identification of protein N-terminal peptides using peptide chromatography and mass spectrometry. Given that protein processing induces new N-terminal protein ends, the resulting neo-N-terminal peptides are also isolated and identified and represent clear markers for the actual cleavage position in the protease substrate . Recently, a similar C-terminal COFRADIC technique was developed to select and identify neo C-terminal peptides, also identifying the processing event. These COFRADIC techniques made it possible to identify large numbers of processing events for several individual proteases or in cellular setups.

Managing, analyzing, comparing and integrating processed events from different proteases have always been very cumbersome. Here we present The Online Protein Processing Resource (TOPPR) that stores high quality processed events and are made available in and easy and intuitive analysis platform.

This TOPPR user manual is divided in three parts. The first part (Searching TOPPR, chapter 1) describes different search strategies that can be followed in the process of retrieving relevant data from TOPPR. The second part (Viewing TOPPR, chapter 2) illustrates and explains how

TOPPR visualizes the search results and which type of analysis can be performed on the results. The last part (Additional terms, chapter 3) explains some common and useful terms.

Chapter 1

Searching TOPPR

TOPPR is a complete protease substrate database. Three different search methods (parameter search (1.1), UniProtKB/Swiss-Prot search (1.2) and motif search(1.3)) were created and are briefly discussed in the following sections.

1.1 Parameter search

The parameter search is present to find large lists of substrates rather than presenting different processed sites in one protein. The simplest way of searching with the parameter search is selecting a treatment from the treatment window (see figure 1.1). Additional information concerning these treatments can be obtained by clicking the treatment(s) link. All the substrates of this treatment will be given in the result section of the search. Only processed sites generated by the selected treatment(s) will be shown. processed sites found in the substrates that were not generated by the selected treatment(s) will not be shown.

Select (one or) more treatment(s):
48h paclitaxel
human µ-calpain
mouse caspase-1
24h paclitaxel
hepacivirin
HIV-1 retropepsin
mouse granzyme C
cathepsin B pH 5.5

Figure 1.1: One or more treatments can be selected to perform a simple parameter search.

Multiple treatments can be selected by holding down the control key and clicking the different treatments. Then, only proteins cleaved by all selected treatments will be given in the search result section.

Advanced search queries can be created with the parameter search. By clicking “advanced search?” in the right bottom of the parameter search field, the advanced search options will be visualized.

The first parameter that can be changed is the search type. The search type is the method TOPPR uses to merge different results. The default setting is “And”. Only the substrate overlap of the selected parameters (e.g. multiple treatments) will be shown as a result. If the search method is set to “Or”, all the substrates that are found for any search parameter will be shown. When the search method is set to “And not” a positive selection drop-down menu appears. Only the substrates that are found for the selected treatment and not the substrates found using the other search parameters will be shown. The different search types are indicated in figure 1.2.

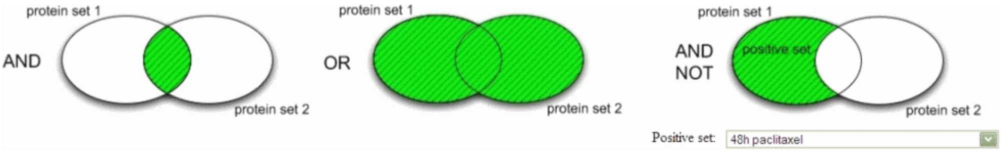


Figure 1.2: The different parameter search types that can be selected.

The second parameter that can be changed is “same site search?”. The default setting is “No” indicating that the different selected treatments are not required to cleave the substrate on the same position. If same site search is set to “Yes”, only proteins that are cleaved by different selected treatments on the same position will be shown (1.3).

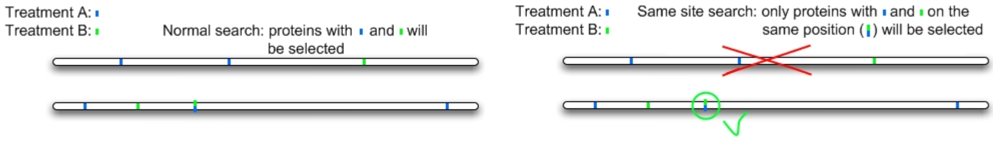

Treatment A:
Treatment B:
Normal search: proteins with and will be selected
Treatment A:
Treatment B:
Same site search: only proteins with and on the same position ( ) will be selected

Figure 1.3: The effect of using the “Same site search” parameter.

If “same site search” is set to “Yes”, changing the search type from “And” to “Or” will have no effect. Changing the search type from “And” to “And not” on the other hand will result in an empty search result.

The third parameter that can be set is the “pre” and “post” motif. Only the processed sites that match the motif will be included in the search result. More information about the motif search can be found in section 1.3.

By changing and selecting the last parameters (experimental type, cell source, taxonomy and inhibitor) specific experimental conditions can be selected and queried. These parameters are influenced by the search type. If the search type is set to “Or” and there is one treatment and one taxonomy selected, all the proteins stored in TOPPR for this taxonomy and all the proteins cleaved by this treatment will be shown.

1.2 UniProtKB/Swiss-Prot search

Searching for a specific substrate and its corresponding processed sites can be done with the UniProtKB/Swiss-Prot search. Three different search methods were developed to search the different UniProtKB/Swiss-Prot protein characteristics.

- **UniProtKB/Swiss-Prot accession search:** Search for a specific UniProtKB/Swiss-Prot accession number. A substrate and its corresponding processed sites will be found if there is a protein stored in TOPPR with this UniProtKB/Swiss-Prot accession number. If no substrate is found for this accession an orthologue search will be performed. A RefSeq NCBI accession will first be identified for this UniProtKB/Swiss-Prot accession number. This is done by PICR (Protein Identifier Cross-Reference Service) a tool developed at the EBI. The orthologues and paralogues for this RefSeq protein can be found by using the HomoloGene NCBI database. HomoloGene is a system for automated detection of homologues (orthologues and paralogues) among annotated genes of several completely sequenced eukaryotic genomes. The found RefSeq orthologues and paralogues will be translated back by PICR to a UniProtKB/Swiss-Prot accession. This accession will finally be searched in TOPPR. The diagram (see figure 1.4) summarizes this orthologue finding method.

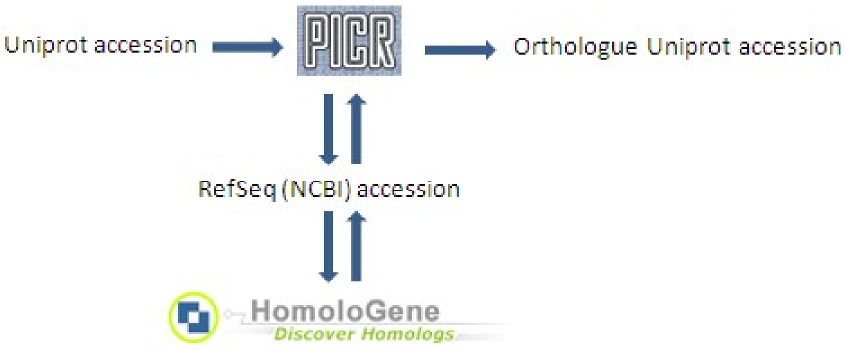
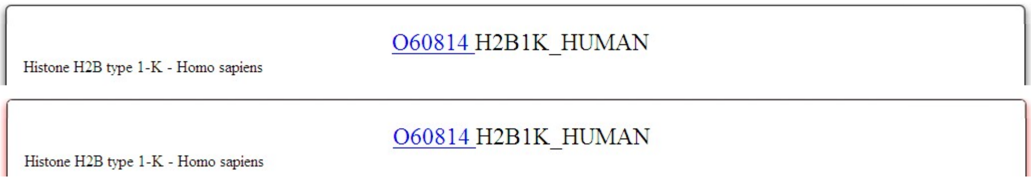


Figure 1.4: The orthologue finding method.

If an orthologue or paralogue is found in TOPPR the protein will be visualized in a red box, and not in a black box (see figure 1.5).



O60814 H2B1K_HUMAN
Histone H2B type 1-K - Homo sapiens

O60814 H2B1K_HUMAN
Histone H2B type 1-K - Homo sapiens

Figure 1.5: The box at the top is the normal representation of a found substrate. The box at the bottom with the red halo is shown when an orthologue or paralogue is found.

- UniProtKB/Swiss-Prot entry name search: Search for a specific UniProtKB/Swiss-Prot entry name. *Example: search for ACTB_HUMAN.*
- UniProtKB/Swiss-Prot description search: Searh for a term in the UniProtKB/Swiss-Prot name or description of a protein. *Example: search for **actin**. TOPPR will retrieve several substrates. Obviously it will find **Actin** , cytoplasmic 1 - Homo sapiens and **Actin** cytoplasmic 1 Mus musculus but also **Actin**-related protein 2 - Homo sapiens and C-jun-amino-terminal kinase-inter **actin**g protein 4 - Homo sapiens.*

1.3 Motif search

The amino acids surrounding a processed site are often very important for substrate recognition by a protease. Searching with a specific amino acid motif before ("pre-site motif") and after ("post-site motif") the processed site can be done in TOPPR. Different search strategies can be used. The easiest strategy is to search for a specific sequence of amino acids. *Example: Pre-site motif = DEVD. TOPPR will only report processed sites where the non-prime (P4-P1) amino acids are DEVD. Pre-site motif = DEVD, post-site motif = G. TOPPR will only find processed sites where the non-prime amino acids (P4-P1) are DEVD and the prime amino acid (P1') is G.*

Another method to find specific processed sites is to use the character *. Any amino acid can occur at the position of *. *Example: Pre-site motif = D**D. TOPPR will find processed site where the non-prime amino acids are DEVD, DSED, DGAD,... .*

Searching for two specific amino acids on one position can be done by searching for [AA 1/AA 2]. *Example: pre-site motif = DE[V/I]D. TOPPR will only find processed site where the non-prime amino acids are DEVD or DEID.*

Chapter 2

Viewing TOPPR

Four different sections (search summary (2.1), motif section (2.2), peptide section (2.3) and protein section (2.4)) appear in a search result. At the bottom of every search result page there is an user protein compilation bar (2.5). Statistics concerning TOPPR can be found here (this takes sometime to calculate!).

2.1 Search summary

Every search result window starts with a search summary. The number of substrates found and the number of corresponding processed sites are indicated. A peptide can be mapped on two different proteins. When this happens, TOPPR will store two different processed sites. The number of unique peptides found in a search result is counted and is given in the summary. The search parameters are shown and differ depending on the used search method. Links to the other parts of the search result are also present in the search summary.

2.2 Motif section

Search results obtained using the parameter or motif search method have a motif section. Also, proteins in the user compilation that are viewed have a motif section.

In the motif section different tools can be used to analyse the found substrates.

- **Automatic generation of an iceLogo.** An iceLogo (see 3.1) needs both a positive or experimental set, and a negative or background set. For both the experimental and background set, one or more treatments can be selected. The background set can also be set as the natural occurrence of amino acids from the substrate species. A subset of these processed sites will be used to generate the iceLogo if the “Only selected peptides?”

check box is selected, and these peptides or processed sites can be selected in the peptide section. The default P value used in the generation of an iceLogo is 0.05. The sites in the processed site list can be used to create other iceLogo visualization on `http://iomics.ugent.be/icelogoserver/logo.html`.

- **Automatic generation of a weblogo.** One or more treatments can be selected. The processed sites that are present in the search result and that are cleaved by the selected treatments will be used in the generation of the weblogo (see 3.2). A subset of these processed sites will be used to generate the weblogo if the "Only selected peptides?" checkbox is selected. The peptides or processed site can be selected in the peptide section. The height of the Y-axis in the weblogo can be changed (default is 4.0 bits) by selecting a different number in the drop down menu "Weblogo bits". The width of the weblogo can be changed (default is from P15 to P15) in the "Weblogo width" drop down menu. A choice can be made between 15, 10, 8, 4 and 2 amino acids pre (non-prime sites) and post (prime sites) processed site.

  The unchangeable default settings for the creation of a weblogo are:

  - picture format = PNG
  - weblogo width = 500 px
  - weblogo height = 300 px
  - small sample correction = on
  - color scheme = default
  - frequence plot = off
  - show fineprint = on

- **Starting the Jalview applet.** One or more treatments can be selected. The processed sites that are present in the search result and that were cleaved by the selected treatments will be visualized with Jalview. A subset of these processed sites will be visualized if the "Only selected peptides?" check box is selected, and peptides are selected.

- **Generating a POPS model.** One or more treatments can be selected. The processed sites that are present in the search result and that are cleaved by the selected treatments will be used in generating the Pops model (see 3.3). A subset of these processed sites will be used to generate the Pops model if the "Only selected peptides?" checkbox is selected. The width of the Pops model can be changed (default from P8 to P8) in the "Pops model width" drop-down menu. For every position, all the amino acids will be counted. This number will then be divided by the total number of amino acids on that position. This number will be multiplicated by 10 after a subtraction of 0.5. An example of how a Pops model is calculated is shown in the following table.

Amino acid	#	# /total	(# /total)-0.5	((# /total)-0.5)*10
A	7	7/55 = 0.1272	-0.3727	-3.727
C	1	1/55 = 0.0181	-0.4818	-4.818
D	30	30/50 = 0.5454	0.0454	0.454
...	17	...	...	...
Y	0	0	-0.5	-5

Table 2.1: The numbers in the last column will be used in the Pops model

Only distinct, non-redundant peptides (with the processed sites in the middle) will be used to generate iceLogos, weblogos and are viewable in the processed site list. This is done because otherwise peptides that are mapped to multiple proteins (isoforms) will get an unwanted higher weight and influence in the logos. *Example: The peptide SGFTSLLMERL is mapped to three different proteins (Tubulin alpha-1B chain - Homo sapiens, Tubulin alpha-1C chain - Homo sapiens and Tubulin alpha chain-like 3 - Homo sapiens). Only two peptides (with the processed sites in the middle) will be used to generate the weblogo. The first and the second processed peptide are the same (both LQGFLVFHSFGGGTGSGFTSLLMERLSVDY). Only one of those two will be selected. The third peptide is not equal to the first two (the amino acids that differ are indicated in bold). The last peptide will therefore also be used to generate the weblogo although it was mapped by the same peptide (SGFTSLLMERL).*

Tubulin alpha-1B chain - Homo sapiens 131-LQGFLVFHSFGGGTG (146) SGFTSLLMERLSVDY-161

Tubulin alpha-1C chain - Homo sapiens 131-LQGFLVFHSFGGGTG (146) SGFTSLLMERLSVDY-161

*Tubulin alpha chain-like 3 - Homo sapiens 138-LQGFL*___I___*F*___R___*SFGGGTG (153) SGFTSLLMERL*___TGEY___*-168*

2.3 Peptide section

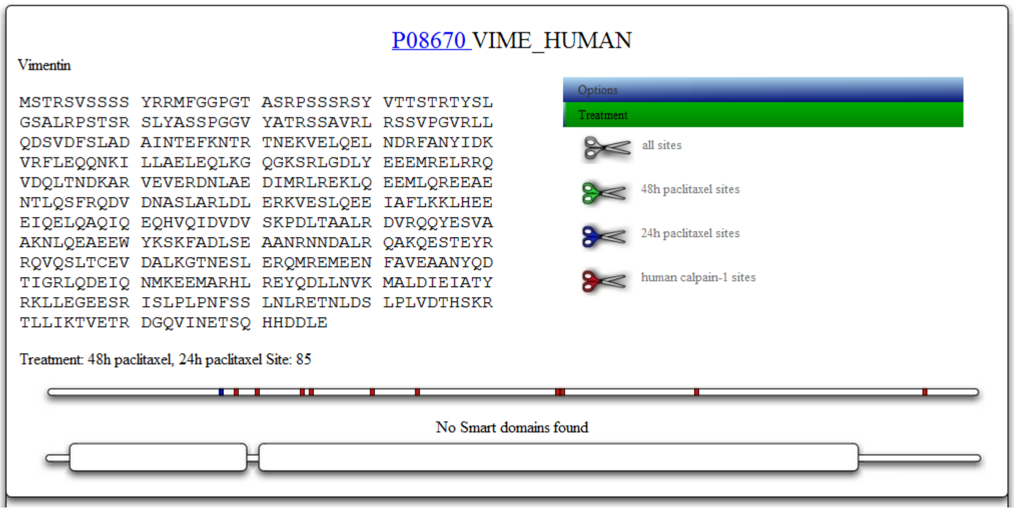
A peptide section is generated by a parameter search, motif search and a user protein compilation. All the found processed sites are listed in the peptide section. These processed sites can be grouped by treatment or by protein. Ordering by protein can give more information about the number of processed sites in a protein. Processed sites cleaved by multiple treatments can be easily spotted in this type of listing. The number of processed sites found for a specific treatment can best be seen in the treatment ordering. In both ordering methods processed sites or peptides can be selected. These selected peptides can be used to generate a subset of processed sites. This subset can be analysed by different tools in the motif section. Substrates in the peptide section can then be added to a "user protein compilation". This compilation can be viewed by opening the performed searches field.

2.4 Protein section

All the search methods expect the parameter search method generate a protein section. For every substrate there is a box with three fields: the sequence, the processed sites and the extra information field. This box will be red if an orthologue or paralogue was found (See UniProtKB/Swiss-Prot search 1.2 for more information).

2.4.1 The sequence field

At the top of the field the UniProtKB/Swiss-Prot accession and the UniProtKB/Swiss-Prot entry name are given. The next line shows the UniProtKB/Swiss-Prot description for this protein. At the bottom of this field a bar is created. This bar represents the primary protein structure. Blue and red squares indicate the relative position of the processed sites in this protein. The squares are blue when two or more treatments induced a processed event on that position. If only one treatment cleaves the protein there, the square is red. Hovering over the squares will generate visualization of text, which gives the processed site position and the treatment(s). Clicking the square will generate the visualization of the processed site in the protein sequence by a colored scissor. The scissors can be made invisible by clicking the square again. Figure 2.1 gives an impression of the sequence field.



P08670 VIME_HUMAN

Vimentin

```
MSTRSVSSSS YRRMFGGPGT ASRPSSSRSY VTTSTRTYSL
GSALRPSTSR SLYASSPGGV YATRSSAVRL RSSVPGVRLL
QDSVDFSLAD AINTEFKNTR TNEKVELQEL NDRFANYIDK
VRFLEQQNKI LLAELEQLKG QGKSRLGDLY EEEMRELRRQ
VDQLTNDKAR VEVERDNLAE DIMRLREKLQ EEMLQREEAE
NTLQSFRQDV DNASLARLDL ERKVESLQEE IAFLKKLHEE
EIQELQAQIQ EQHVQIDVDV SKPDLTAALR DVRQQYESVA
AKNLQEAEEW YKSKFADLSE AANRNNDALR QAKQESTEYR
RQVQSLTCEV DALKGTNESL ERQMREMEEN FAVEAANYQD
TIGRLQDEIQ NMKEEMARHL REYQDLLNVK MALDIEIATY
RKLLEGEESR ISLPLPNFSS LNLRETNLDS LPLVDTHSKR
TLLIKTVETR DGQVINETSQ HHDDLE
```

Treatment: 48h paclitaxel, 24h paclitaxel Site: 85

No Smart domains found

Figure 2.1: The sequence field.

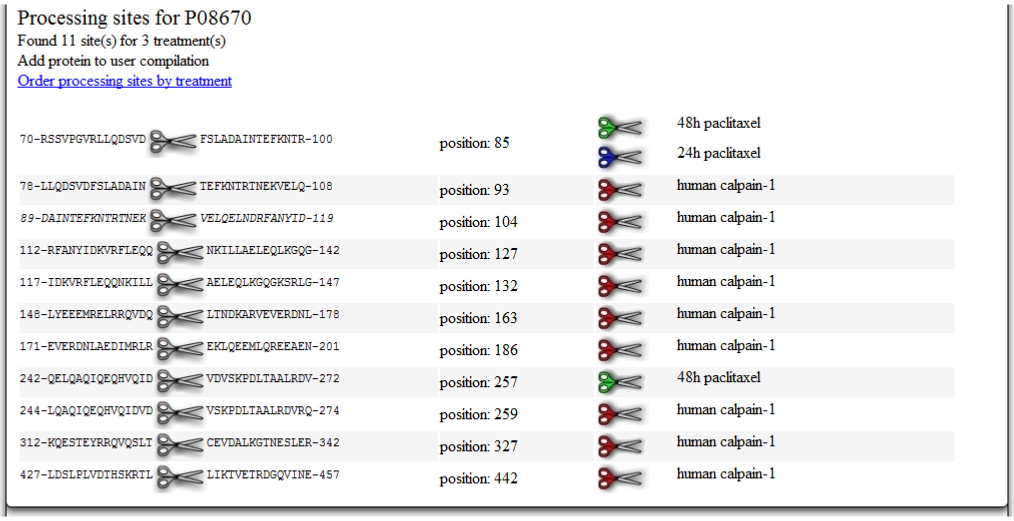
Next to the sequence two different drawers give additional functions. In the first "Options" drawer, three different actions can be initiated.

- **Show domains** This action will visualize the know domains in the PFAM and SMART database by visualizing a domain protein bar under the unfolded protein picture. Hovering over the domains will indicate the name and position of the protein domain. Clicking it will open a new tab with the information about the domain in the PFAM or SMART database.
- **Show secondary structure** A very simple secondary structure predication (*via* StrBioLib) will be indicated on the sequences. Beta sheet predications are indicated by a green arrow whereas alpha helices are indicated by a blue loop.
- **Show Gene Ontologies** The know gene ontology terms linked to this protein will be visualized beneath the sequence. Clicking on the link will open a new tab with the gene ontology information on the gene ontology web site.

The second drawer "Treatments" holds a list of buttons that can be used to show and hide the processed sites of specific treatments in the protein sequence.

2.4.2 The processed site field

Every processed site and the peptide with the sequence surrounding this site is listed and ordered by treatment. The processed sites can also be ordered by processed position in the protein. Ordering by position can be used to find positions where more than one treatment cleaved the protein. The processed site sequence is in italics if different isoforms were found for the MS/MS identified peptide, mapped on the substrate. Hovering over such a peptide will generate a small box displaying the number of isoforms found. Clicking it will start a motif search for this peptide and thus generates a search result with all the isoforms. Figure 2.2 gives an impression of the processed site field.



Processing sites for P08670

Found 11 site(s) for 3 treatment(s)

Add protein to user compilation

Order processing sites by treatment

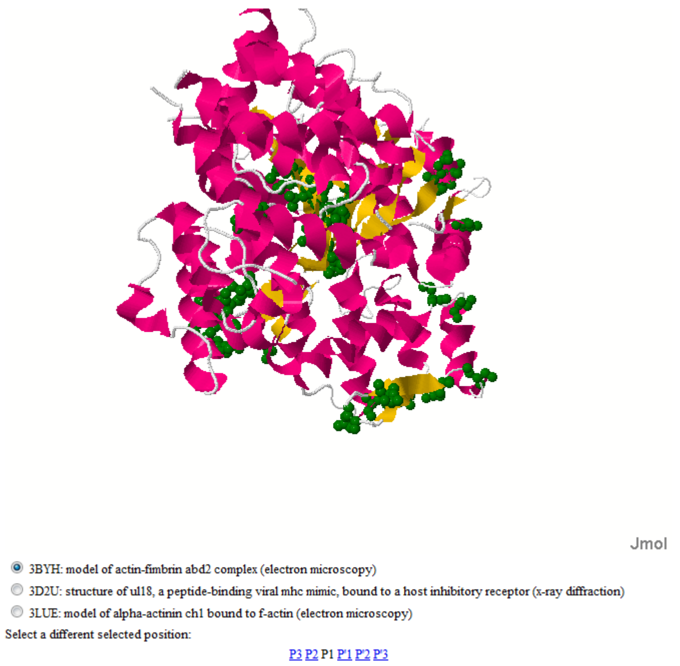
70-RSSVPGVRLLQDSVD ✂ FSLADAINTEFKNTR-100	position: 85	✂ ✂	48h paclitaxel 24h paclitaxel
78-LLQDSVDFSLADAIN ✂ TEFKNTRTNEKVELQ-108	position: 93	✂	human calpain-1
89-DAINTEFKNTRTNEK ✂ *VELQELNDRFANYID-119*	position: 104	✂	human calpain-1
112-RFANYIDKVRFLEQQ ✂ NKILLAELEQLKGQG-142	position: 127	✂	human calpain-1
117-IDKVRFLEQQNKILL ✂ AELEQLKGQGKSRLG-147	position: 132	✂	human calpain-1
148-LYEEEMRELRRQVDQ ✂ LTNDKARVEVERDNL-178	position: 163	✂	human calpain-1
171-EVERDNLAEDIMRLR ✂ EKLQEEMLQREEAEN-201	position: 186	✂	human calpain-1
242-QELQAQIQEQHVQID ✂ VDVSKPDLTAALRDV-272	position: 257	✂	48h paclitaxel
244-LQAQIQEQHVQIDVD ✂ VSKPDLTAALRDVRQ-274	position: 259	✂	human calpain-1
312-KQESTEYRRQVQSLT ✂ CEVDALKGTNESLER-342	position: 327	✂	human calpain-1
427-LDSLPLVDTHSKRTL ✂ LIKTVETRDGQVINE-457	position: 442	✂	human calpain-1

Figure 2.2: The processed sites field.

2.4.3 The extra information field

Three additional forms of information can be visualized in the extra information field.

- **View the 3D structure in JMol** This action can only be done if (a part) of the structure of the protein is known. A different structure can be selected if more than one pdb structure is found by choosing one in the radio button list. By default the P1 position (the amino acid just before the cleavage site) is visualized in the structure by a green ball and stick representation of the corresponding amino acid. Other (non) primed positions can be selected below the structure radio button list (see figure 2.3).



● 3BYH: model of actin-fimbrin abd2 complex (electron microscopy)

○ 3D2U: structure of ul18, a peptide-binding viral mhc mimic, bound to a host inhibitory receptor (x-ray diffraction)

○ 3LUE: model of alpha-actinin ch1 bound to f-actin (electron microscopy)

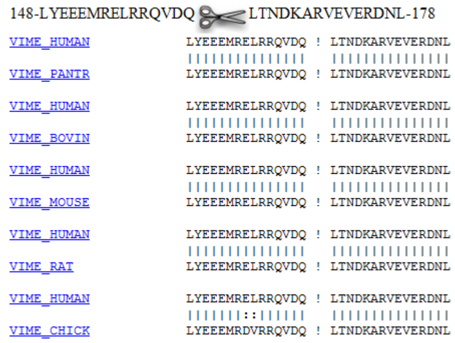
Select a different selected position:

P3 P2 P1 P'1 P'2 P'3

Figure 2.3: Example 3D visualization in TOPPR.

- **View information about the spectra and peptides** Every processed site is identified by either the N- or C-terminal COFRADIC technology. This implies that every processed site is linked to an identified MS/MS spectrum. Information about the identifications and spectra can be found by pressing this link. The (B and Y ion annotated) spectrum can be visualized by making use of the Pride spectrum viewer.

- **View the conservation of the processed sites** The orthologues (found by the method described in 1.2) will be aligned by the Needleman-Wunsch pairwise global alignment algorithm with default parameters. An example output of this is given in figure 2.4.

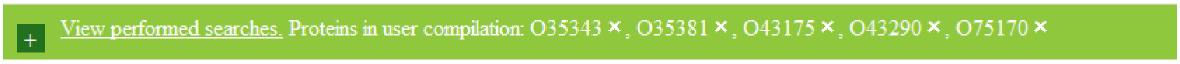


```
148-LYEEEMRELRRQVDQ ✂ LTNDKARVEVERDNL-178
VIME_HUMAN          LYEEEMRELRRQVDQ ! LTNDKARVEVERDNL
                    |||||||||||||||   |||||||||||||||
VIME_PANTR          LYEEEMRELRRQVDQ ! LTNDKARVEVERDNL

VIME_HUMAN          LYEEEMRELRRQVDQ ! LTNDKARVEVERDNL
                    |||||||||||||||   |||||||||||||||
VIME_BOVIN          LYEEEMRELRRQVDQ ! LTNDKARVEVERDNL

VIME_HUMAN          LYEEEMRELRRQVDQ ! LTNDKARVEVERDNL
                    |||||||||||||||   |||||||||||||||
VIME_MOUSE          LYEEEMRELRRQVDQ ! LTNDKARVEVERDNL

VIME_HUMAN          LYEEEMRELRRQVDQ ! LTNDKARVEVERDNL
                    |||||||||||||||   |||||||||||||||
VIME_RAT            LYEEEMRELRRQVDQ ! LTNDKARVEVERDNL

VIME_HUMAN          LYEEEMRELRRQVDQ ! LTNDKARVEVERDNL
                    |||||||::||||||   |||||||||||||||
VIME_CHICK          LYEEEMRDVRRQVDQ ! LTNDKARVEVERDNL
```

Figure 2.4: Example output for the processed site conservation.

2.5 User protein compilation bar

At the bottom of the result screen, a green bar is shown (see figure 2.5). A table with the performed searches is visualized if “View performed searches” is clicked. The result of these searches can be viewed in a new window or a new tab if the search parameters in the first column are clicked. In this green bar, the proteins that are in the user protein compilation are shown. Proteins can be added by clicking “add” in the peptide section. Proteins can be deleted again from this compilation by clicking the white cross next to the protein UniProtKB/Swiss-Prot accession in the green bar. The user protein compilation is a useful tool to compare different proteins of interest. The compilation can be viewed by going to the performed searches table and clicking on the compilation. Performed searches can also be visualized by clicking the search log button in the upper bar.



+ View performed searches. Proteins in user compilation: O35343 ×, O35381 ×, O43175 ×, O43290 ×, O75170 ×

Figure 2.5: The user compilation bar.

Chapter 3

Additional terms

3.1 iceLogo

IceLogo builds on probability theory to visualize significant conserved sequence patterns in multiple peptide sequence alignments against background (reference) sequence sets that can be tailored to the studied system and the used protocol. The unique advantages of iceLogo compared to other sequence logo creating tools is that iceLogo has a more dynamic nature and is correcter and completer in the analysis of conserved sequence patterns. Click here for more information.

Improved visualization of protein consensus sequences by iceLogo. Colaert N, Helsens K, Martens L, Vandekerckhove J, Gevaert K. Nat Methods. 2009 Nov;6(11):786-7

3.2 Weblogo

Weblogo is a web-ased application designed to make the generation of sequence logos as easy and painless as possible. Sequence logos are a graphical representation of an amino multiple sequence alignment developed by Tom Schneider and Mike Stephens. Each logo consists of stacks of symbols, one stack for each position in the sequence. The overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of each amino or nucleic acid at that position. In general, a sequence logo provides a richer and more precise description of a processed site than would a consensus sequence. Click here for more information.

Sequence logos: a new way to display consensus sequences.Schneider TD, Stephens RM. Nucleic Acids Res. 1990 Oct 25;18(20):6097-100.

WebLogo: a sequence logo generator. Crooks GE, Hon G, Chandonia JM, Brenner SE. Genome Res. 2004 Jun;14(6):1188-90.

3.3 PoPS

PoPS is a software tool for modeling protease activity. Its main aim is to provide an environment in which the user can learn and reason about the activity of any protease. PoPS allows the user to create a model of activity for an arbitrary protease by specifying general parameters that are commonly used to describe the function of a protease, for example, the number of important subsites, their amino acid preferences and the relative importance of the subsites. The user can then supply a substrate (protein sequence) and PoPS will use the protease model to predict where the substrate will be cleaved. A simple prediction of cleavages is very useful in itself, particularly if the protease has a well-defined activity. In addition, the predictions do not just include what will happen, but also what will not happen. However, PoPS's utility does not end here. Once a prediction has been made, PoPS allows the user to adjust parameters and the original protease model to perform a deeper investigation of how the protease is functioning. For example, in the case of a protease with a poorly defined function, predictions can be compared to experimental results, and the model can be altered to gain a better understanding of the true model of the protease's activity.

PoPS also aims to provide other information about the substrate that might improve the accuracy of predictions. This includes tertiary structure and secondary structure predictions, but more options are currently being implemented. Click here for more information.

PoPS: a computational tool for modeling and predicting protease specificity. Boyd SE, Pike RN, Rudy GB, Whisstock JC, Garcia de la Banda M. J Bioinform Comput Biol. 2005 Jun;3(3):551-85.