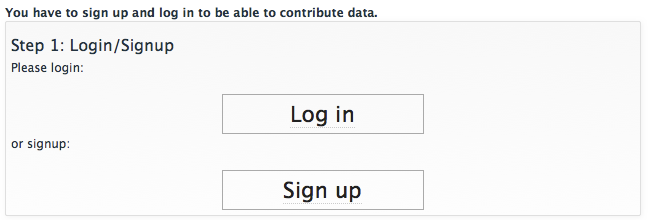
**TopFIND – data contribution HOWTO**

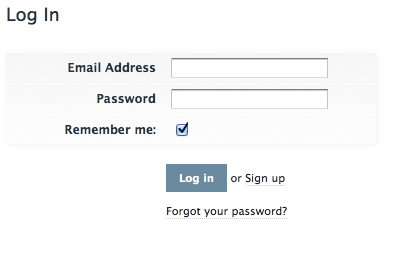
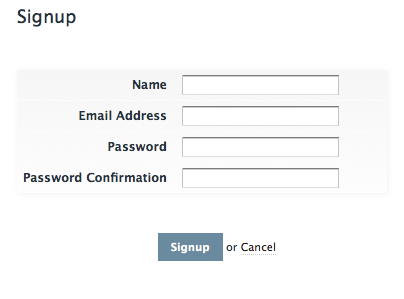
Your are highly encouraged to submit your data on protein termini and protein processing. This guide provides a step by step introduction to the simple batch submission process.

Please visit <http://clipserve.clip.ubc.ca/topfind/contribute/> to start the three-step process.

**I – Signup and login**

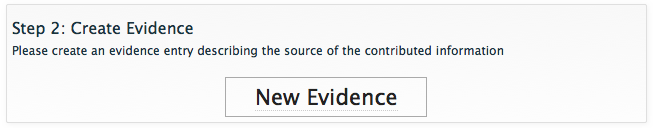


To limit spam we require you to create a custom account and login. If you have not done so far, please signup by providing your name, e-mail address and choosing a password.



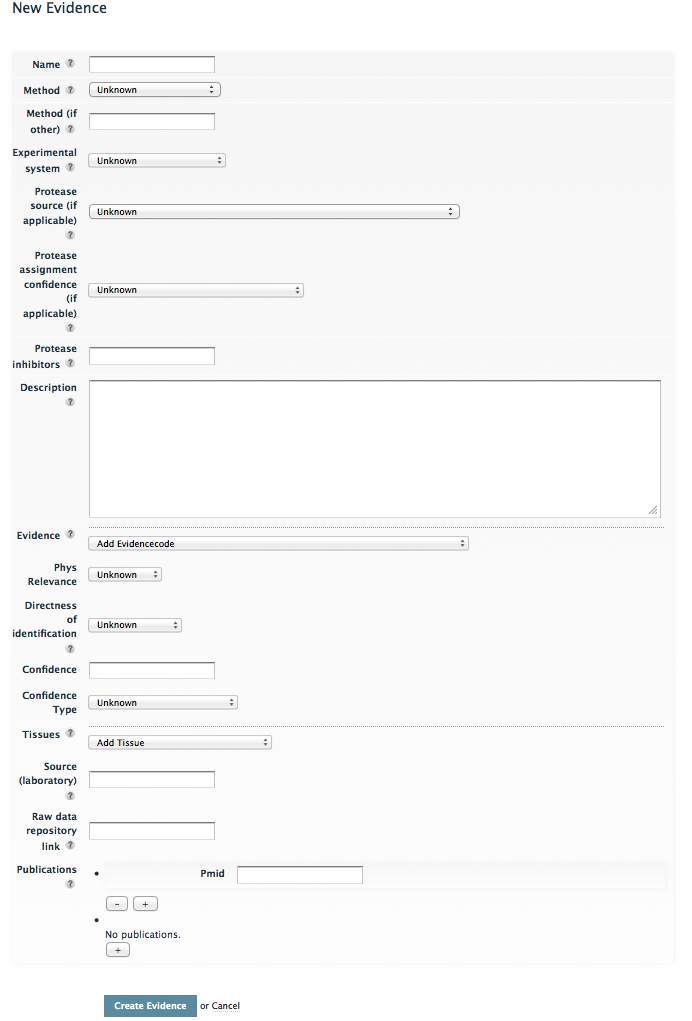
**II – Evidence**

For each experiment, experimental condition or theoretical approach you have to create a new evidence record. This record is used to collect the information on how your data has been generated.



Please fill the provided form with as much information as possible. Click on the question marks to get detailed information on the meaning of the individual parameters.

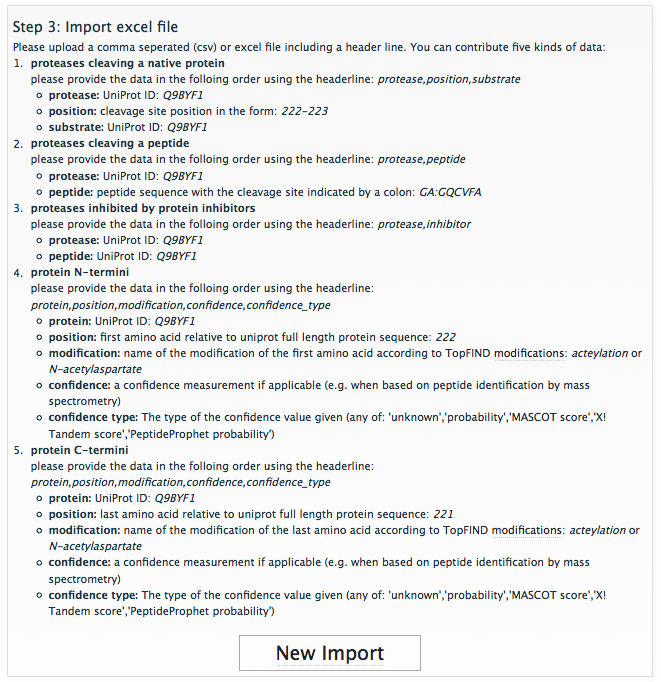
The quality and completeness of this information is key to allow other scientists to make sense of your submitted data. The provided information is the basis of the filtering mechanism (right panel in on every protein page) provided by TopFIND.



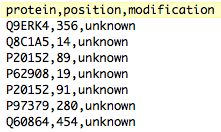
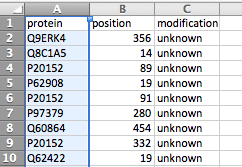
1. **Name:** a unique (short) name of your choice
2. **Method**: A dropdown list of the current main approaches [electronic annotation, COFRADIC, N-TAILS, C-TAILS, ATOMS, Edman sequencing, enzymatic biotinylation, chemical biotinylation, MS (gel based), MS (semi-tryptic peptide identification), MS (other), mutation based analysis]. If your method is not listed, please choose other and state in free text in the next field.
3. **Experimental system:** please select one of [cell free, cell culture, organ or tissue culture, in vivo sample]
4. **Protease activity and modulation thereof:** please select one of [natural expression, recombinant, over expression, knockdown (expression), genetic knockout, knockdown (functional, protease drug or blocking antibody or inhibitor used)].
5. **Certainty of protease assignment:** A qualitative measure that the protease assigned to a cleavage is correct. This reflects the possibility of additional proteolytic activities other than the one studied and reported being present in the experimental system: [I, no other proteolytic activities present (e.g. a simple in vitro cleavage assay); II, proteolytic system present but abolished (e.g. denatured); III, proteolytic system present but impaired (e.g. inactivated by inhibitors); IV, proteolytic system present and active (e.g. cell culture or *in vivo* studies); or unknown].
6. **Protease inhibitors:** Due to the critical effect of the activity of any other proteases in the system on the data generated and its interpretation it is explicitly asked if protease inhibitors were used during sample preparation and if so which ones.
7. **Description:** Free form text field to add important information, e.g. detailed method descriptions, that does not fit elsewhere.
8. **Evidence code**: Categorization [e.g. inferred from electronic annotation] according to the controlled evidence vocabulary from the Obo foundry.
9. **Physiological relevance**: A qualitative measure of the likelihood that the identified terminus, cleavage or inhibition is present and relevant in vivo [physiologically relevant, likely physiologically relevant, likely no physiological relevance, no physiological relevance, unknown]. Examples that can be considered in this classification include whether the substrate was identified in tissue samples or only in biochemical assays, is native or not, whether the protease is normally present in a certain cell, tissue or development stage, or whether it is even present in a given species.
10. **Directness**: A qualitative measure [direct, likely direct, likely indirect, indirect, unknown] describing if the reported cleavage or terminus has either been directly observed (e.g. an N-terminus identified by Edman sequencing or proteomics approaches such as ATOMS, or terminomics approaches such as TAILS or COFRADIC or is indirectly inferred (e.g. by inferring a C-terminus from an experimentally observed neo-N terminus).
11. **Confidence value and type**: A quantitative measure describing the confidence that the given observation is correct. For example, the results generated from proteomics data base search programs such as Mascot includes a peptide identification confidence score, which can be filtered upon by TopFIND. *Type* refers to the program or system used to generate these scores. This differentiation allows for the evaluation and filtering of confidence values from algorithms and methodologies that cannot necessarily be directly compared, like for example a peptide identification by mass spectrometry and sequence determination by Edman sequencing.
12. **Tissue distribution**: If relevant, the cells or tissues in which the observation was made according to the controlled UniProtKB tissue and cell line vocabulary.
13. **Laboratory**: State your laboratory or the laboratory that originally created the data
14. **Raw data repository**: Links to the raw data supporting containing related data
15. **Publications**: Add related publications by PubMed ID

**III – Formatting your data and submission**

Next you have to bring your actual data into a defined comma separated format. The exact content of each column depends on the type of data you want to submit and is explained in detail on the TopFIND contribution and help pages as depicted below.



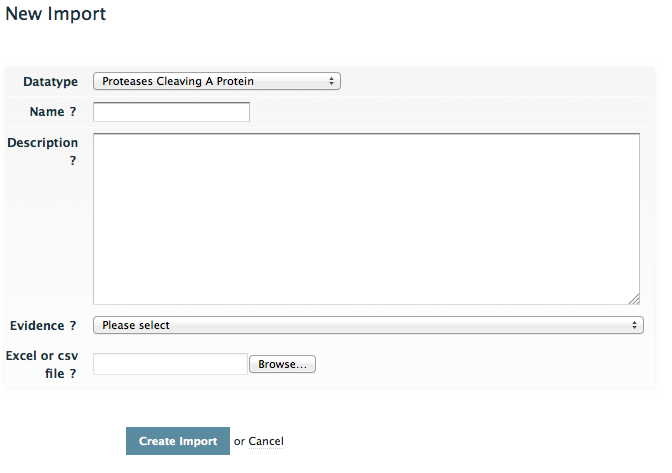
The easiest way to generate the required file is to assemble the data in your favorite spreadsheet application like Microsoft Excel and save it as comma separated file.  
Please make sure that the header names are exactly as specified and that the entries are separated by commas and not any other character.



After after putting your data into the right format you can proceed with uploading the file to the TopFIND server. Please press “New Import” and fill the provided form. The question marks provided detailed help for every section.

First you have to pick the type of data you are submitting (e.g. N-termini or protease cleavage sites), next you provide a custom name and description of your choice (used to reference this dataset). Finally you select the corresponding evidence report. This can be the one that you just created, however you can also link multiple files to the same evidence if one evidence report exactly matches the information for all.

Finally you select the file containing your data and submit the information.



Following this you will receive a confirmation email. At the same time the TopFIND curators will be informed about your submission. They will check data integrity and completeness and finally make your data available to the scientific community.

**Thank you very much for your contribution!**