# Identifying NPPs and Neuromodulators - A test case

INPUT:

1. Amphibian set from UniProtKB (only full-length proteins, proteins contains the sequence feature of propeptide, excluding the feature of of transmembrane domain). The sequences are provided as an example set.

2. Collect 40 sequences from the list as input to NeuroPID.

3. Activate the predictor by activating the SP local application.

RESULTS:

1. <u>Result Summary:</u> Out of 40 proteins - 2 proteins were removed due to the lack of SP. All the other sequences are marked as candidates. Note that only 40% of the input set agreed with 3/4 predictions.
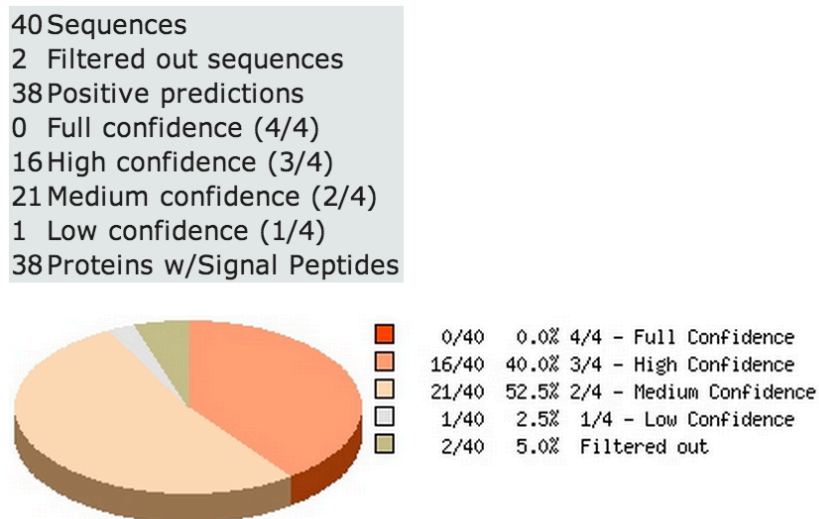
2. <u>Detailed Results:</u> A list of 38 sequences reported in a sortable table with a graphical representation, the location of basic resides (R, K) and the presence of dibasic and tribasic residues is indicated. The user can analyse in depth each prediction by forwarding the sequence

- FASTA – download the sequence
- ProtoNet - classification scheme for clusters
- UniProt – detailed database of proteins
- SignalP -activating the SignalP 4.0 application
- NeurPred – Predictor of processing of NPPs to mature NPs.

Internal Score is presented this score captures the density of the dibasic and tribasic residues in the sequence following cleavage of its SP.

3. High Quality Predictions: A sub-list of the of all candidate sequences (as in the Detailed Results) reported in a sortable table. The table is filtered by the Internal Score.

4. Download Results: The download options are focused on several formats including FASTA format, XSL table and a list of ID. These lists can be further analysed by other tolls (e.g., NeuroPred, Pandora, David enrichment etc).

5. Features Histograms: There are a number of dominating features that can be analysed. Each protein is shown as a dashed line in view of the distribution of the values of the negative and positive training sets. The feature units are transformed to standard deviations. See example for a specific protein and a specific feature (Aromaticity, for Sequence O42350).
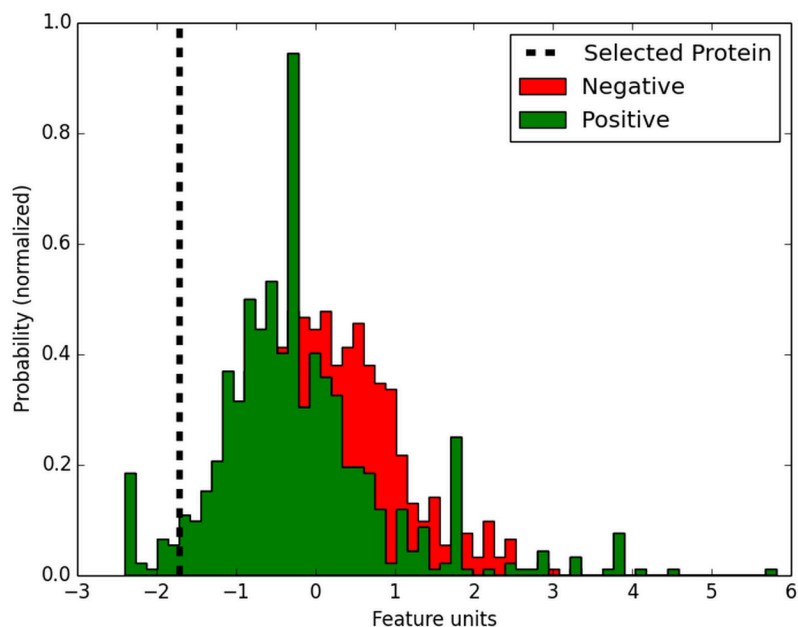
6. Sequences Graphics: An illustration of the protein graphics and several counts along the sequence such as the number and density of basic resides, dibasic and triple basic sequences. The protein ID, names and length are also shown.

BIOLOGICLA INFERENCE:

The high quality results support the view that most of the proteins (35/40) are related to NPPs and cell modulators. Inspection of the biological knowledge supported the enrichment of short protein modulators, neurohormones, and NPPs protein among the amphibian. They are composed of only 7 prototype proteins.

The dominating proteins include Brevinins (21/35 sequences). These are short peptides that expressed on the skin of the frog. Brevinins are antimicrobial peptides against Gram-positive bacteria and fungi. These peptides have low hemolytic activity against human erythrocytes.

The other proteins that belong to the high quality predictions include:

- C-type natriuretic peptides (2). These are hormone that plays a role in regulation of chondrocytes proliferation and differentiation. May also be vasoactive and natriuretic.
- Dybowskin (3). These are antimicrobial peptide with activity against the Gram-positive bacterium
- Esculentin (3). These are antibacterial peptides against representative Gram-negative and Gram-positive bacterial species. Esculentins exhibit hemolytic activity.
- Gastrin/cholecystokinin-like peptide (1). In the brain, it is strongly expressed in the pituitary gland.
- Pro-opiomelanocortin (2). The NPP that belong to the POMC family. The rich NPs products cover biologically active peptides that involved in diverse cellular functions.
- Ranakinin-N (1). These peptides induce contraction of intestinal smooth muscle. May induce relaxation of arterial smooth muscle. It lacks antibacterial activity.

In summary – the NeuroPID identified many short active peptides in amphibians. Many of these products are antimicrobial and others are classical NPs.

Disclaimer and Comments:

1. The false positives includes fragment of receptors, protease negative regulators and antimicrobial peptides (see examples above).

2. The length of the NPPs and the modulators are often <300 amino acids. I can be used as a filter for the input.

3. The Internal Score is a unified feature that is used as a strong filter. We use the threshold on the Internal Score and the consistency of > 2 prediction methods as a mark for high quality prediction.