LinX 2.0

User's Guide

Introduction

LinX is a Java language-based program designed for fast assignment, evaluation, and validation of mass spectrometric datasets. It has been primarily developed for mass spectrometric data (MS1) with really high resolution. However, currently, LinX also enables to analyze data from other instruments (MS2) when it's MS1 data are exported separately. In addition, newly this software enables to process data from ^{15}N labeling of proteins (paragraph **Check N labeling**). LinX is highly suitable for wide range of users for it's simple and intuitive graphical user interface (GUI) as well as the maximal independence on mass spectrometric instrument or software.

However, this software tool was designed for analysis of single proteins and for small protein-protein or protein-nucleic acid (paragraph **Protein-RNA/DNA**) complexes, therefore authors don't recommend to use it for large complexes. It would lead to rapid increase of computational time and misassignments. Moreover, in order to avoid wrong assignment of MS data authors recommend MS2 analysis of identified cross-links. Authors also recommend manual check all results in spectra even if they are confirmed by MS2 analysis. Beginners in evaluation of chemical cross-linking data should read the guidelines (*C. Iacobucci et al, AnalChem. 2017;89: 7832-7835*) for it. In case you would be lost or confused even after reading this manual, all examples of inputs and outputs for BSA search are placed in folder .**Test_data\BSA**.

Installation and Launching Info

Windows

The presence of java software (java) is prerequisite for successful installation. The program itself is distributed as a compressed ZIP file containing all necessary components in a folder **LinX**. Immediately after unpacking, program is ready to use with no need of further installations.

Program Launching:

- o   Windows Explorer
  - •   Go to the **LinX** home folder
  - •   Double click on **LinX.jar**

    Launching in this way, the program restore last used settings from the **input.sen** data file.

- o   Command line options
  - •   Launch the program without any arguments
    - o   Run the following command from the **LinX** home folder:

      java -jar LinX.jar

      This command initiates the program GUI with setting stored in the **input.sen** data file (as in the case described above).

  - •   Perform assignment using the file containing desired settings as the first argument:

    java -jar LinX.jar example.sen

The program runs in the background using settings from the entered file, in this case **example.sen** file. If all computing conditions (modifications, cross-links, data paths, etc.) are correct, new window with results will appear. Otherwise, the GUI displays on the panel where an error is occurred.
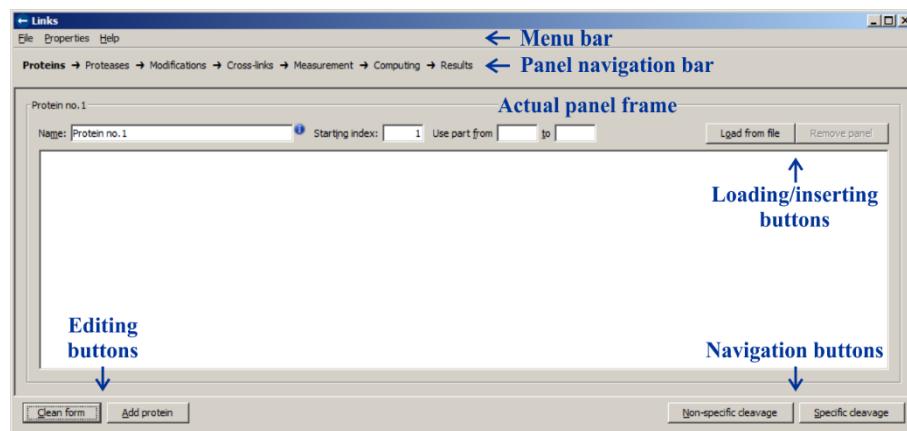
Linux and Mac OS

Currently, LinX doesn't work on Linux or Mac operating systems.

Graphical User Interface (GUI)

Main Window

LinX launches in the main window, where the content can be modified depending on current computing step. Basic elements of the window are described below.

Menu bar – identical for all panels; it provides access to the file processing, editable libraries and supporting information

Panel navigation bar – identical for all panels, actual panel is **highlighted** in bold; for visual orientation among panels only; each panel represents one step of the calculation query

Actual panel frame – editable area of each panel

Loading/inserting buttons – enable to insert specified requirements for calculation or load information from files

Editing buttons – enable editing of previously entered inputs in actual panel

Navigation buttons – perform shifting among individual steps of calculation query

Main Menu Structure

File

- ○ Load settings   [CTRL+O]

    This function enables the computing settings loading from a formerly created setting file or output (***.sen**). Settings are always loaded for actual and all following panels, **not for previous**.

- ○ Save settings   [CTRL+S]

   Actual settings and parameters of the current and all previous panels are saved into the new setting file (***.sen**).
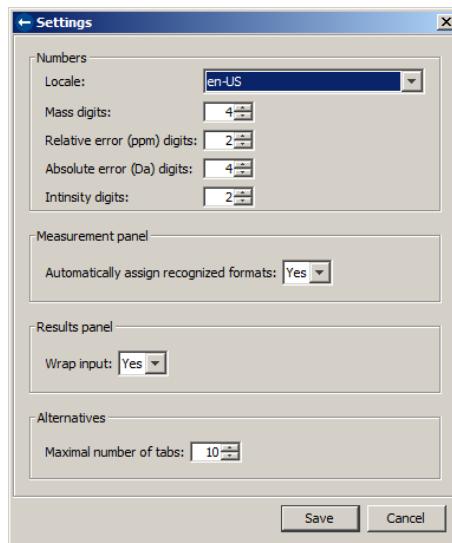
Properties

Each section of the menu is also accessible via a keyboard shortcut. For detailed information, see the section Editable Libraries.

- ○ Elements   [CTRL+E]
- ○ Proteases   [CTRL+P]
- ○ Modifications   [CTRL+M]
- ○ Bonds   [CTRL+B]
- ○ Measurement formats   [CTRL+F]

- ○ Options   [CTRL+T]

    In a settings window, format of numbers displayed in the **Results** panel can be edit there. General requirements for number format (separators of thousands, decimal point/commas) depends on user locale, which can be manually changed.

    There are also more alterable parameters available via this window. Additionally, automatic recognition of an input file format can be enabled/disabled here, as well as options for wrapping the input files.

Help
- o User guide
- o About

Individual panels

Proteins

The panel **Proteins** consists of a single window with editable text fields *Name*, *Starting index*, *Use part from-to* and the main field *Sequence*.

Protein *Name* or *Sequence* could be written directly into the field, copy-pasted from clipboard or imported from a simple ***.fasta** file using *Load from file* button. The use of one-letter code for amino acid sequences is required. Expected amino acid mutations can be defined using character "/" between. The entry "MUT/ATE" designates possible (**NOT** strict) change of amino acid threonine to alanine.

The *Name* field (***.fasta** file header) cannot contain expressions "peptide", "all proteins" or "within one molecule" (upper and lowercase independent), neither any brackets or dashes. The length of the protein name is limited to 30 characters.
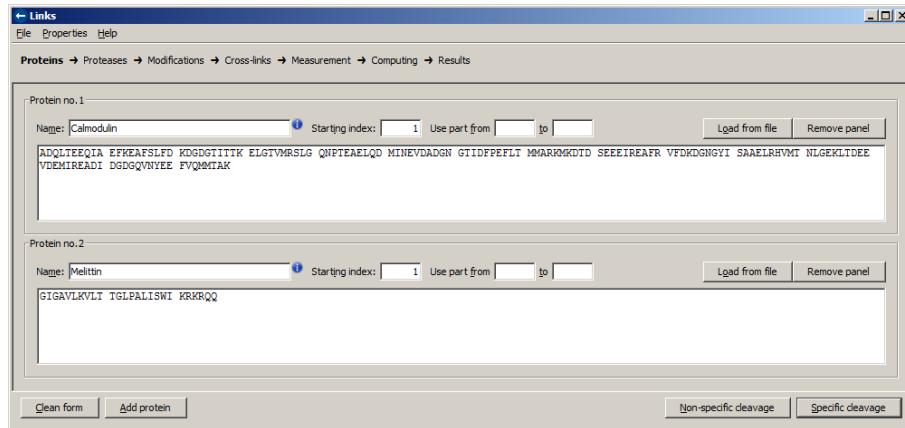
The *Starting index* field indicates number of the first amino acid in the *Sequence* field. It is possible to use either positive or negative number.

Use part from-to fields enable to select only a part of the inserted sequence. Entered numbers must be consistent with the *Starting index* and the length of corresponding sequence.

Program was primarily designed for cross-linked peptides assignment. For this reason, calculating algorithm accepts more than one sequence for library computation. The Add protein button can call another complete protein frame;
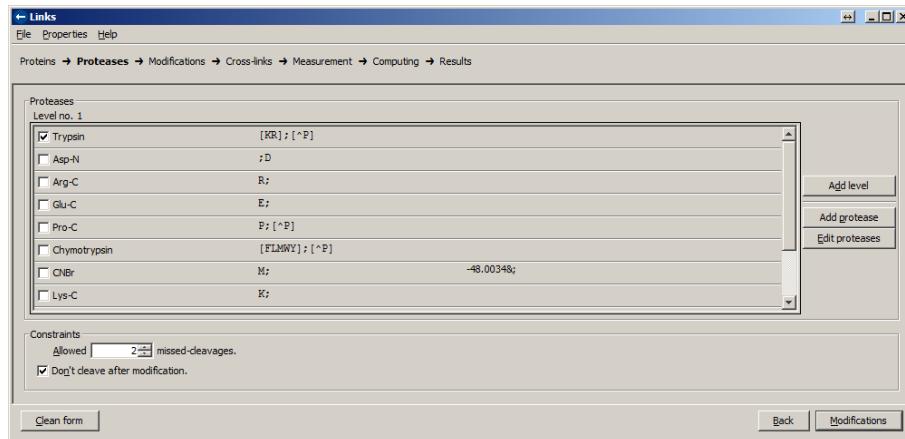
conversely, existing frame can be removed with the *Remove panel* button. The *Clean form* button resets all fields of the panel.

The navigation button *Specific cleavage* shifts the program to the **Proteases** panel, while button *Non-specific cleavage* switches program directly into the panel **Modifications**.

Proteases

The panel **Proteases** contains a checkbox list of common specific and semi-specific proteases used in most proteomic studies. If a protease is checked, program generates peptide library according to the defined cleavage rules. More than one protease can be used in each cleavage level.

The second frame of this panel contains cleavage constraints settings. The maximum number of allowed missed cleavage sites can be set as well as possibility to skip the cleavage site after/before modified amino acids. The number of missed cleavage sites is related to each selected protease.

Additionally, the program supports multilevel cleavage function, which is available by the button *Add level*. For instance, protein is cleaved with a protease in the first level. After that, acquired peptides can be modified or tagged (e.g., new peptide N-terminus labeling) and then cleaved by another protease again.

Cleavage constraints are active for both (or more) cleavage levels. Specific constraint setting for each digestion level is not allowed.

Moreover, editing or inserting new proteases is allowed directly from this panel. The button **Add protease** opens simple floating panel with two editable fields. The first field called *Name* cannot contain control characters and spaces in this field; digestion specificity is characterized in the second field *Cleavage rules*. To express protease specificity, use following control characters.

Control characters

- One-letter amino acid code – express targeted amino acid(s)

- "." (dot) – represents any amino acid

- ";" (semicolon) – defines cleavage site of the selected amino acid
  - e.g., "K;" - cleavage on the C-terminus of lysine

- "[]" (square brackets) – in cases where protease cleaves before or after more than one amino acid, write them all into the brackets
  - e.g., "[FLMWY];" – cleavage after phenylalanine or leucine or methionine …)

- "[^X]" (caret with aminoacid inside square bracktets) – is used for specific cleavage with specific amino acid-dependent restrictions
  - e.g. "[KR]; [^P]" – trypsin digest specificity; cleavage after lysine and arginine, but only when the following amino acid is not proline

The list of proteases is stored in the file **.\Properties\proteases.prs**. There are several ways to access this library from a program window. The library is accessible via Properties in the Main Menu, using a shortcut CTRL+P or by clicking the **Edit protease** button on the **Protease** panel. For more information, see the section Editable Libraries.

The **Clean form** button disables all checkboxes of actual panel and sets a number of missed cleavages to the zero value. Navigation button *Modification* switches the program to the **Modification** editing panel, meanwhile button **Back** returns the program to previous panel. Switching between the panels does not lead to any changes of their content.

Modifications

Panel **Modification** is divided into two parts. On the top, there are several editable drop-down lists which allow to define almost unlimited number of modifications. Active modifications included in a calculation are listed in the second part of the panel. Columns in the resulting table of active modifications are sortable according to the text chain in the cells.

The first drop-down list called **_Target protein_** (1) enables to select modified protein(s). Modification can occur on all proteins, on the specific one, or on the newly created peptide (useful for multilevel digestion).

The second drop-down menu named **_Target group_** (2) contains a list of all amino acids and C- and N-terminal functional groups found in the protein sequence(s) defined on the panel **Proteins**. Specific amino acid can be selected from the list by a mouse cursor or by writing a one-letter amino acid code into the field. To switch between fields, use a Tab key.

After selection of the amino acid, the specific position of modification can be chosen using the third field **_Sequence position_** (3). Modifications can be defined for all amino acids present in the sequence, or for one specific amino acid.
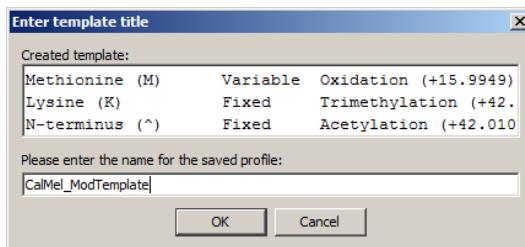
Menus *Target group* and *Sequence position* permit to define modification for more than one specific aminoacid at the one step. Multiple values (amino acid code or sequence positions) must be written manually and separated by commas.

Options of the fourth field *Occurrence* (4) enable to calculate the modification as fixed or variable. The last editable field on this panel involves a list of *Modifications* (5) defined in the file **.\Properties\modifications.prs**. The list of modifications is also accessible via a shortcut CTRL+M or the Main Menu bar. All editing processes are described in the section Editable Libraries.
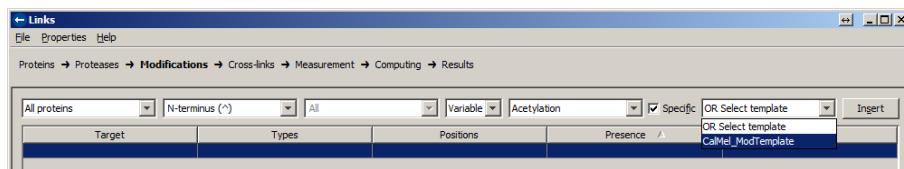
The checkbox *Specific* serves as a filter of the modification list according to the specificity defined in the library. If enabled, only specific modification for selected amino acid will appear. After clicking the *Insert* button, modifications and their parameters are copied to the *Active modifications* list. This list is editable via buttons *Remove selected* and *Edit selected*. Editing is performed using upper editable field(s). The text is highlighted in red when it is being edited. Changes are accepted after *Insert button* click.

Frequently used list of modifications can be saved as a template, which is stored in the file **.\Properties\modifications.template.prs**. To create a new template, select required lines in the active modification list, then click by a right mouse button on the selected line and choose the option **Create template** in the context menu. A window summarizing selected modifications is then displayed. Enter the name of the template and confirm its creating by OK.

Templates can be uploaded via the *Template* (6) field **for selected protein** and activated by clicking on the *Insert* button.

Created templates don't contain the information about protein or specific positions in the sequence. Entries from the template are inserted for selected protein only. In some cases, table of active modifications will have to be checked and edited manually after template uploading.
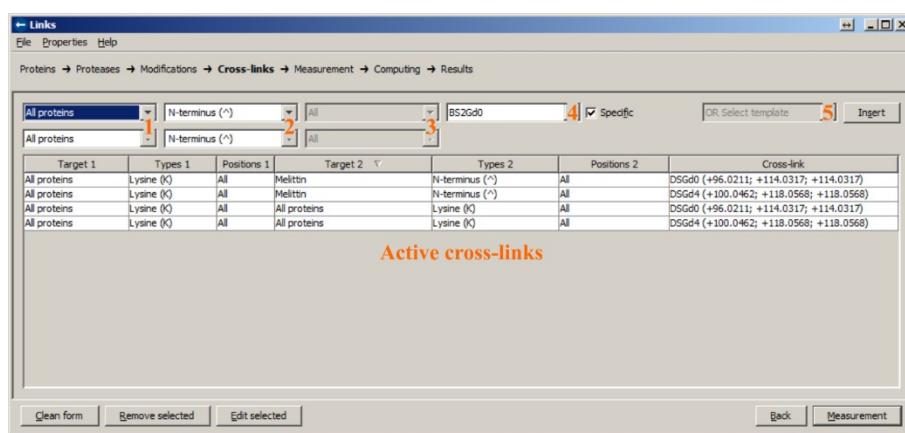
The panel **Modifications** contains three navigation buttons. Button *Back* switches the program to the previous panel **Proteases**. *No cross-links* button shifts the process directly into a panel **Measurements**, while *Cross-links* button to the **Cross-links** panel.

Cross-links

The panel **Cross-links** enables to add disulfide or chemically initiated bonds between two selected amino acids. Computational algorithm then creates additional library where masses of two peptides are calculated according to the previously considered parameters (digestion, modifications …).

The composition of cross-link list is very similar to the **Modification** list. Panel is divided into two sections again. Cross-linking module is designed mainly for bifunctional chemical reagents and disulfide bonds. For this reason, the editable fields *Target protein* (1), *Target group* (2) and *Sequence position* (3) are present in pairs on this panel. Field *Bond* (4) contains a list of pre-defined cross-links saved in the file **.\Properties\Bonds.prs**. The bond library is accessible via a shortcut CTRL+B or by the Main Menu bar. Detailed description of editing the library is thoroughly explained in the section Editable Libraries.

Checkbox *Specific* can be used for filtering the bonds. If enabled, only bonds relevant for selected amino acids are displayed in the field *Bond*. After clicking the *Insert* button, bonds and their parameters are copied to the *Active cross-links* list. The list is editable with buttons *Remove selected* and *Edit selected*. Editing is performed by a drop-down list in the upper part of the window; the actually edited line is highlighted in red. Changes are accepted by *Insert button* click again.
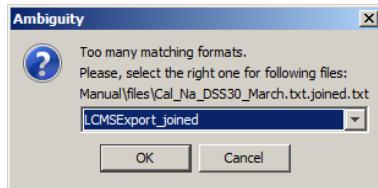

Button *Back* shifts the program to the previous panel **Modification,** while *Measurement* button switches the screen to the last panel.

Measurement

Panel **Measurement** contains three checkboxes. The checkbox frame *Filter peptides* allows to choose, which type of results program returns. If this checkbox is **disabled**, program returns complete library of theoretically generated peptides. All settings, restrictions, modifications and cross-links are included in this library.
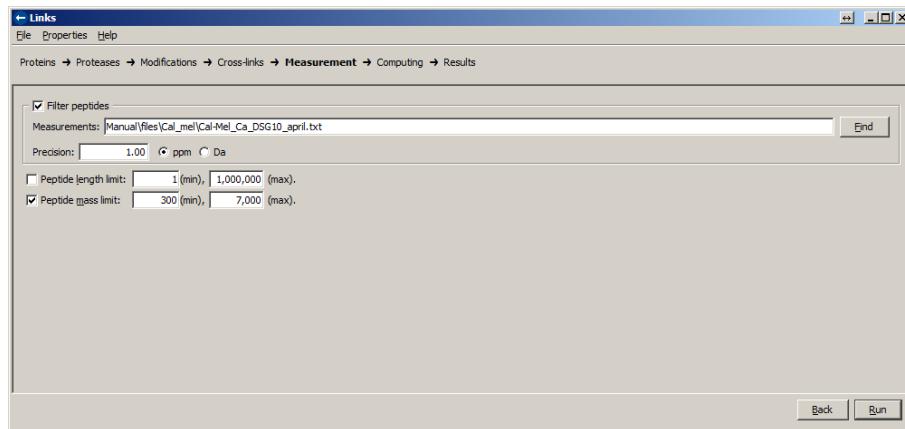
If the checkbox is **enabled**, fields *Measurements*, *Precision* and the button *Find* become accessible. In this setup, the theoretical library of peptides is compared with a list of deconvoluted mass spectrometric data. The input file of

MS data must contain plain text values in tabulator or space separated columns. Overview of supported formats and new formats definition can be found in the section Editable Libraries. If the data format recognition is enabled (see Options in menu **Properties**) and concurrently only one defined format is possible to fit on the data file, the format will be assigned automatically. Otherwise, if some ambiguity in the format assignment is determined, a new window for manual format election will appear.

Additionally, program enables to compare more than one mass list file in the same assignment process.

The range of positive assignment is limited by a value entered in the field *Precision*, which enables to choose between two mostly used units (ppm or Da).
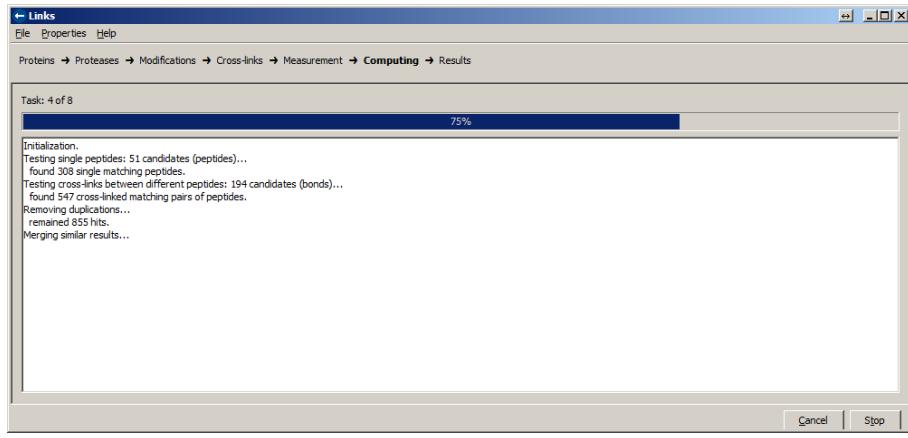
Two other checkboxes on the panel **Measurement** facilitate restriction of length or mass of peptides, which are included in the calculation. These options are very useful for large protein calculations, because it significantly reduces the calculation time and computer memory workload.

The button *Back* switches the program to the previous panel **Cross-links,** while the button *Run* starts the calculation.

Computing

The panel **Computing** substitutes a command line or console, as it allows to track the calculation and MS data assignment progress.

The first screenshot shows the Computing panel:

Links

File   Properties   Help

Proteins → Proteases → Modifications → Cross-links → Measurement → **Computing** → Results

Task: 4 of 8

75%

Initialization.
Testing single peptides: 51 candidates (peptides)...
   found 308 single matching peptides.
Testing cross-links between different peptides: 194 candidates (bonds)...
   found 547 cross-linked matching pairs of peptides.
Removing duplications...
   remained 855 hits.
Merging similar results...

Cancel   Stop

The **Stop** button interrupts the calculation and the program returns results what have been calculated until interruption, whereas the **Cancel** button annuls ongoing calculation and switches the screen back to the **Measurement** panel. Stop and Cancel processes can take a while.

Results

The panel **Results** contains all calculation parameters listed on the top of the window, and also a table editor below, where all processed data are shown. All columns in this table are sortable, while lines can be selected for the **Alternative search** or **Grouping**. These functions are mentioned in the section Function, modules and extensions.

Results can be saved to the text file with the extension ***.sen** using buttons **Save all** or **Save current**. The **Save all** button saves all tabs to the separate files, while **Save current** saves only the current tab.

To improve productivity and effectivity, **Function, modules and extensions** were added to the results panel.

Because of requirements for editability, table editor on **Results** panel contains **variables** in the cells, **not** simple text. It means that data shown in this panel is not possible to copy out using clipboard. To export data, use buttons *Save current* or *Save all*.

Function, modules and extensions

Grouping

The grouping utility is accessible directly from the results window via button *Grouping*. Grouping enables to join several lines into the one according to the specified modifications, bonds or masses. The grouping task can be used just for selected lines via a right mouse button and the context menu, whereas using the *Grouping* button submenu affects all lines in all tabs of the results window. Grouped lines can be ungrouped back, via the right mouse button and the context menu for selected lines or by the *Grouping* button submenu for all the entries.
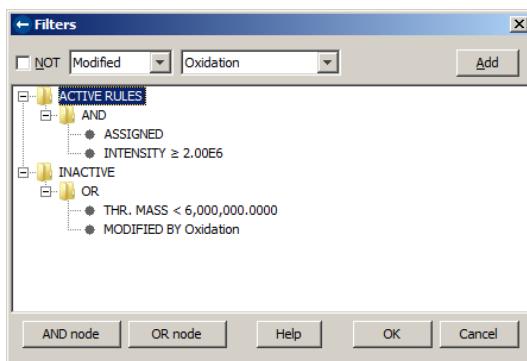
If the Grouping was performed according to the modifications or bonds, individual values in grouped lines are separated by a character "|". In case of mass-based grouping, the content of individual cells is closed into curly brackets and after that separated by the "|" character.

Filtering

Purpose of the **Filtering** function is to organize and simplify the table of results. This function is accessible from the **Results** panel using button *Filtering* and it shows in the separate window. The filtering is performed using a tree-structured environment, where folder icons represent individual nodes of the tree. There are two basic nodes - **Active** and **Inactive** rules. Formerly created and unused conditions can be stored in the folder **Inactive rules**. Particular rules can be transferred between nodes using clipboard shortcuts CTRL+C (V, X) or by a right mouse button and context menu. Filtering panel does not support drag and drop gesture.

New rules are established using drop-down menu in the upper part of the window. If the condition is negative, the **NOT** checkbox must be enabled. All filterable objects are listed in the first drop-down menu. Depending on the selected object, other options specific for this choice are displayed in a following menu. Final conditions must be accepted with the *Add* button.

If more than one rule is required in the same level, use *AND node* or *OR node* buttons to create a new parent node (folder). The AND (OR) node determines a relationship between rules inside the parent node.

OR node is set as a default root node in the Filtering module!

Filterable values:

- **Assigned**
  - Rows with assigned records only are displayed in the results window.

- **Modified**
  - Assigned records containing any modification, including Type 0 cross-link (nomenclature according *A. Sinz J.Mass Spectrom. 2003; 38: 1225–1237*), are displayed.
  - When the grouping function is performed before filtering, at least one joined records contains modified peptide.

- **X-linked**
  - Assigned records containing any disulfide or cross-link (Type 1 or Type 2) are displayed.
  - If grouping is performed, a least one of grouped records contains disulfide, Type 1 or Type 2 cross-link.

- **Interpeptide**
  - Assigned records containing any interpeptide disulfide or cross-link Type 2 are displayed.
  - If grouping is used, at least one of joined records contains disulfide or Type 2 cross-link between two peptides.

- **Intrapeptide**
  - Assigned peptides (modified or unmodified), intrapeptide disulfide, Type 0, and Type 1 cross-link are displayed.
  - If grouping is used, at least one of joined records contains single peptide (modified or unmodified), intrapeptide disulfide, Type 0 or Type 1 cross-link.

- **Ambiguous**
  - Ambiguity filter is applicable only for records grouped by masses. It displays masses for which more than one assignment exists.

- **Contains**
  - Filtering conditions are defined by the **regular expressions** and only appropriate records are displayed.
  - If grouping is used before, at least one of joined records fulfills the defined rule.

- **Arithmetic**
  - Assigned records are displayed if numeric values in selected columns are consistent with defined inequality.

Draw

Draw is a utility, which enables to visualize protein sequence coverage including modified peptides.

Key to the colors of peptides:
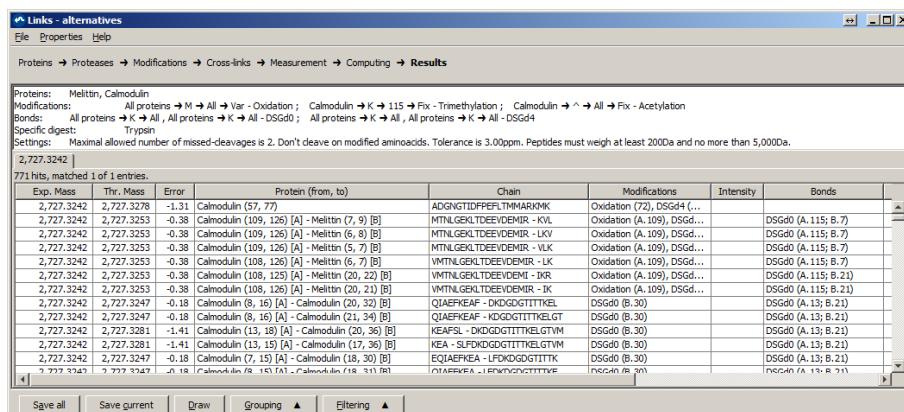
- ○ white - unmodified peptide
- ○ **red** - peptide with any modification including Type 0 cross-link
- ○ **blue** - Type 1 and Type 2 cross-linked peptide
- ○ **purple** - Type 1 and Type 2 cross-linked peptide with modification

Alternatives

The alternative search serves as a control mechanism of selected masses assignment. The search is executable from the **Results** window by right-clicking the selected line and choosing the option **Find alternatives** in the context menu. Then the new search starts in an external window.

Whereas proteins are digested according to the selected rules in a standard search, Alternative search employs **non-specific** digest of all sequences from the protein panel. All the other parameters in following panels remain active and are included in the new calculation. The Alternative search can be launch even for unassigned masses.

The Alternative search returns the list of assigned masses in the standard **Results** window.
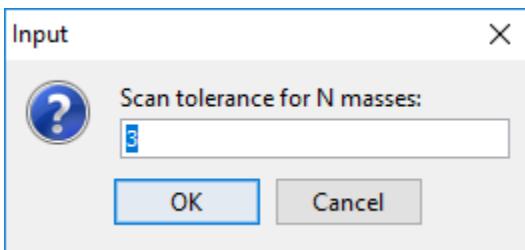
It is possible to launch alternative search for more masses in a single run. However, the calculation of alternative theoretical library is compute- and memory-intensive process, thus it is recommended to search for selected masses individually. But in some cases, especially when selected masses for calculation are close to each other, is conversely recommended to calculate their alternatives in the single search.

Check N labeling

The **Check N labeling** is function that enables to assign masses to cross-linked peptides containing ^{15}N isotope. This utility was constructed for specific experiments where interaction interface of homodimers is studied using chemical cross-linking. In this experiment, the natural form of monomer (containing 99.6% of ^{14}N and 0.4% of ^{15}N) is cross-linked with monomer labeled by ^{15}N (ideally 100% of ^{15}N). The calculation is based on the fact the resulting Type 2 cross-links from dimerization interface must exist in 3 or 4 forms ($^{14}N/^{14}N$, $^{14}N/^{15}N$, $^{15}N/^{14}N$ and $^{15}N/^{15}N$) and therefore 3 or 4 different masses should be visible in the spectra (*T. Taverner et al, JBC. 2002;277: 46487-46492*). The **Check N labeling** is executable from the **Results** window by via button *Check N labeling*. Input for the calculation is list of assigned masses for peptides modified by cross-linking reagent in the standard **Results** window. If you press **Check N labeling** button, the LinX ask first for the range of scans that will be applied for the search. Default value is 3 scans. This value enable to find the corresponding masses in three scans before and after scan of

original mass from **Results** window. The LinX also ask for the minimal intensity and maximal difference from median value of assigned cross-links for calculation of *correction factor*. More details about it you can find in paragraph *correction factor*.

The **Check N labeling** returns the list of assigned masses in the new window similar to the standard **Results** window (example below).

Unlike **Results** window, The **Check N labeling** window contains four extra columns labeled 14N/14N, 14N/15N , 15N/14N and 15N/15N. Masses in these four column are highlighted by different color. The green (light blue in version for color blind people) color means the charge mass for precalculated mass (both written in the cell; charge mass is in brackets) was found in the inputed masslist and therefore corresponding form of monomer was presented in solution. In contrary, the red color means opposite. The shade of red/green color correlates with intensities of masses. The color is getting dark with increasing intensity of mass. Logaritm of intensity of charge mass is visible when you tap on the cell by cursor. The possible "N/A" value means the intrapeptide cross-link is analysed and thus, it is impossible to assign the ^{14}N/^{15}N and ^{15}N/^{14}N forms. The cells with masses can be also dark blue in special cases when the LinX would have found more than one mass for the search in in the predefined range of scans and masses. Then the cell contains number of assigned masses and mass with highest intensity. However all these assigned masses are written into separate column after export to txt.file. The complete **Check N labeling** table can be saved into the .txt file using button *Export*. Test data including all inputs and results for testing this utility are in the.**Test_data\15N_labeled_protein** folder.

Inter/Intra ratio

Unfortunately, in real cross-linking experiment, isotopic patterns of intramolecular and intermolecular cross-links often overlap and two populations of restrains are formed and thus only relaying on MS/MS information is not sufficient to find out what population of cross-links is more dominant in its mixture. The cross-linked ion intensities must be taken in to the account. Therefore the **Check N labeling** window contains also the column labeled *Inter/Intra ratio* that shows if intermolecular and intramolecular cross-linked peptides are presented in sample at same time. The function enable to distinguish if the identified…

NKRP1B_15N.txt (MSe export)

Cross-links	Protein	Modif...	Bonds	Mass	Scan	Charge	14N/14N	14N/15N	15N/14N	15N/15N	Inter/Intra Ratio
DSIK - DSIK	NKRP1...		DSG (...	1019,...	141...	2	1019,5255 (510...	1024,5107 (512...	1024,5107 (512...	1029,4959 (515...	0,12
DSIK - DKVTS	NKRP1...		DSG (...	1106,...	109...	2	1106,5576 (553...	1112,5398 (556...	1111,5427 (556...	1117,5249 (559...	0,07
DKVTS - DKVTS	NKRP1...		DSG (...	1193,...	83-83	2	1193,5896 (597...	1199,5718 (600...	1199,5718 (600...	1205,5540 (603...	0,43
DKCFR - DSIK	NKRP1...		DSG (...	1225,...	143...	2	1225,5881 (613...	1230,5733 (615...	1234,5615 (617...	1239,5466 (620...	---
MLSVQKSSVQK	NKRP1...		DSG (...	1330,...	221...	2	1330,7035 (665...	N/A	N/A	1345,6590 (673...	---
MLSVQKSSVQK	NKRP1...		DSG (...	1330,...	277...	2	1330,7035 (665...	N/A	N/A	1345,6590 (673...	---
MLSVQKSSVQK	NKRP1...	Oxidat...	DSG (...	1346,...	191...	2	1346,6984 (673...	N/A	N/A	1361,6540 (681...	---
MLSVQKSSVQK	NKRP1...	Oxidat...	DSG (...	1346,...	197...	2	1346,6984 (673...	N/A	N/A	1361,6540 (681...	---
DSIKEK - DKVTS	NKRP1...		DSG (...	1363,...	77-79	2	1363,6951 (682...	1369,6773 (685...	1371,6714 (686...	1377,6536 (689...	---
WICQK - WIC...	NKRP1...		DSG (...	1449,...	331...	2	1449,7017 (725...	1457,6780 (729...	1457,6780 (729...	1465,6543 (733...	---
MLSVQK - DK...	NKRP1...		DSG (...	1468,...	577...	1	1468,7287 (146...	1477,7020 (147...	1476,7050 (147...	1485,6783 (148...	---
DSIK - DVLKI...	NKRP1...		DSG (...	1502,...	293...	2	1502,8312 (751...	1512,8016 (756...	1507,8164 (754...	1517,7867 (759...	---
KGATLLLIQ - ...	NKRP1...		DSG (...	1513,...	335...	2	1513,8836 (757...	1518,8688 (759...	1524,8510 (762...	1529,8361 (765...	0,89
DVLKITGVT - ...	NKRP1...		DSG (...	1589,...	271...	2	1589,8632 (795...	1595,8455 (798...	1599,8336 (800...	1605,8158 (803...	---
KGATLLLIQ - ...	NKRP1...		DSG (...	1600,...	317...	2	1600,9156 (800...	1606,8978 (803...	1611,8830 (806...	1617,8652 (809...	0,10
KGATLLLIQ - ...	NKRP1...		DSG (...	1771,...	287...	3	1771,0211 (591...	1778,9974 (593...	1781,9885 (594...	1789,9648 (597...	---
DCGRKGATLL...	NKRP1...		DSG (...	2032,...	291...	3	2032,0743 (678...	2038,0565 (680...	2050,0209 (684...	2056,0032 (686...	---
MLSVQK - EEL...	NKRP1...		DSG (...	2163,...	655...	1	2163,1730 (216...	2178,1285 (217...	2171,1492 (217...	2186,1048 (218...	---
DKVTSESCST ...	NKRP1...		DSG (...	2190,...	247...	3	2190,9819 (730...	2204,9404 (735...	2201,9493 (734...	2215,9078 (739...	---
ENGSCATISG...	NKRP1...		DSG (...	2241,...	483...	2	2241,0162 (112...	2248,9925 (112...	2257,9658 (112...	2265,9421 (113...	---
ESCSTDNRWI...	NKRP1...		DSG (...	2342,...	263...	3	2342,0362 (781...	2350,0125 (784...	2361,9769 (787...	2369,9532 (790...	---
ESCSTDNRWI...	NKRP1...		DSG (...	2342,...	357...	3	2342,0362 (781...	2350,0125 (784...	2361,9769 (787...	2369,9532 (790...	---
WICQKELNHK...	NKRP1...		DSG (...	2433,...	115...	4	2433,1980 (609...	2447,1564 (612...	2450,1475 (613...	2464,1060 (616...	---
DKCFR - DNR	NKRP1		DSG (	2447	373	2	2447,1707 (122...	2471,0995 (123...	2456,1440 (122...	2480,0728 (124...	---

Export   Load masslist   Close

…cross-link is strictly intermolecular or if it's the mixture of intermolecular and intramolecular cross-links and what type of them is dominating. The *Inter/Intra ratio* is calculated from intensities of isotopic patterns written in experimental data. Precisely is calculated as the sum of intensities for 14N/15N, 15N/14N forms divided by the sum of intensities for 14N/14N , 15N/15N forms. The *Inter/Intra ratio* can achieve any number from range 0.01 to 1.00 where 1.00 means only intermolecular cross-link is formed while value 0.01 means only intramolecular cross-link is formed.

Correction factor

The **Check N labeling** window also give information about original mixture of unlabeled (^{14}N) and ^{15}N-labeled protein as well. Information is represented by *correction factor* that is computed from intensities of all assigned peptides for each protein written in the Input 1 file. The *correction factor* serves as double check if unlabeled and labeled protein are equimolar. If not the LinX warns user(s) the isotopic patterns in raw spectra will be affected and recommended manual verification of all identified intermolecular cross-links in raw data must be done. The factor is calculated for each assigned peptide whose intensity is higher than $1 \cdot 10^7$ ($1 \cdot 10^7$ is default value that can be changed by user in specific window that occur after initiation of **Check N labeling** function). Further data are statistically processed and 20% of outliers are discarded (20% is default value that can be changed by user in specific window that occur after initiation of **Check N labeling** function). The final *correction factor* is a weighted average of the data after reduction.

Protein-RNA/DNA
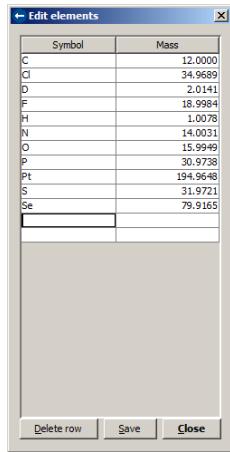
LinX can be also used for identification of cross-links from protein-nucleic acid complexes. Unfortunately, since the aminoacids (alanine, threonine, glycin and cystein) and nucleobases (adenine, thymine, guanine and cytosine) has according IUPAC nomenclature same symbols (ATGC) it is required to exchange the symbols of nucleobases (J for adenine, O for thymine, Z for guanine and B cytosine) in the input file and inn the parameters of script. This means exchange the folder **Properties** from main root with the folder **Properties** in folder **.\Protein_DNA** and rewrite the nucleobases code in fasta file. For this could be used excel file called **New_DNA_Code_Generator** which is placed in **.\Protein_DNA .** All other function and utilities are same as for protein-protein complexes. Test data including all inputs and results for testing this utility are in the **.\Test_data\Foxo4_DNA** folder.

Editable Libraries

Elements $[$**CTRL+E**$]$

Atomic weights of chemical elements, fundamental building blocks of all molecules (amino acids, modification reagents, cross-linkers), are defined according to the standard atomic weights available on the [IUPAC website](#). Elements and their weights are stored in the file **.\Properties\elements.prs**. Library is accessible via a CTRL+E shortcut and the menu option Properties. To add a new element, double click into the blank cell and then insert its name and mass (atomic weight). Changes must be confirmed by the button **Save**.

Edit elements

Symbol	Mass
C	12.0000
Cl	34.9689
D	2.0141
F	18.9984
H	1.0078
N	14.0031
O	15.9949
P	30.9738
Pt	194.9648
S	31.9721
Se	79.9165

Delete row   Save   Close

Amino acids

The amino acid molecular mass is computed as a sum of atomic weights of elements in a chemical formula, molecules of water are excluded. Amino acids composition and their naming (use one-letter code) are stored in the file **.\Properties\monomers.prs**. Remaining alphabet letters can be defined as a special or noncoding amino acid or any other molecule (synthetic aminoacids, nucleic acid bases, etc.). Amino acid library is accessible using the menu Properties.

Proteases   [CTRL+P]

Pre-defined proteases and their preferences are saved in the **.\Properties\proteases.prs** file. The library is accessible via a shortcut CTRL+P, using the Main menu or the button *Edit Proteases* on the panel **Proteases**. In the Edit protease window, proteases in a list can be ordered using navigation buttons in the upper right corner or edited by the second group of buttons. Buttons *Add protease* and *Edit protease* open two identically looking windows, where the fields *Name* and *Cleavage rules* can be added/edited according to the rules described in the panel **Proteases**. If the checkbox **Simple protease** is disabled, static *Modification of termini* generated after a cleavage can be defined. Additionally, there is an option to protect a newly created N- or C-terminus against following modifications by the proper checkbox.

Edit proteases

Trypsin_all	[KRY];	
TrypsinP	[KR];	
Trypsin	[KR];[^P]	
Asp-N	;D	
Arg-C	R;	
Glu-C	E;	
Pro-C	P;[^P]	
Chymotrypsin	[FLMWY];[^P]	
CNBr	M;	-48.0034&;
Lys-C	K;	
TEV	E...Y.Q;[GS]	
Pepsin	;[FLWY] [FLWY];	
nonspecific	[A-Z];	-18.0106;

Move to top
Move up
Move on
Move down
Move to end

Edit protease
Add protease
Remove protease

Close

Edit protease

Name of protease: CNBr
Cleavage rules: M;

Modifications of termini
Left side: -48.003371   Right side:
☑ Lock C-terminus   ☐ Lock N-terminus

☐ Simple protease   OK   Cancel

Modifications [CTRL+M]

The library of pre-defined modifications is stored in the **.\Properties\modifications.prs** file. It is also accessible by a shortcut CTRL+M or by the Main menu bar. The library table contains four editable columns. The **Name** column can contain letters and numbers, special characters are not allowed. A **mass** shift caused by modification can be entered either as an explicit numeric value (to introduce negative shift, use mathematical sign "-") or as a loss/addition of elements. In cases, where a modification is mediated by isotopically labeled reagents, enter an isotopic shift to the column **Check Diff**. Program then searches for a **Mass** shift and returns positive assignment, if a sum of both **Mass** and **Check Diff** values is included in the mass list. The **Specificity** column is consisted of a list of amino acids, which can be potentially modified. Symbols "^" and "$" represent N-terminus and C-terminus respectively.

After each table alteration, changes must be confirmed by the *Save* button.

Name	Mass	Check Diff	Specificity
Acetylation	-H+COCH3 (+42.0106)	+0.0000	^, S
Amidation	-OH+NH2 (-0.9840)	+0.0000	$
Ammonia loss	-NH3 (-17.0265)	+0.0000	^, K, N, Q, R
Carbamidation	+57.0215	+0.0000	C
Carboxymethyl	-H+CH2COOH (+58.0055)	+0.0000	^, K
Cystein desulfatation	-S (-31.9721)	+0.0000	C
Formylation	-H+CHO (+27.9949)	+0.0000	
Fucosylation	-OH+C6H11O5 (+146.0579)	+0.0000	S
GlcNAc	C8H15NO6-H2O (+203.07...	+0.0000	^, N
GlcNAc2Man5	+C46H76N2O35 (+1,216....	+0.0000	N
Gluconoylation	+178.0477	+0.0000	^
HexNAc	-H2O+C8H15N1O6 (+203...	+0.0000	N, S, T, X
Hydration	+H2O (+18.0106)	+0.0000	H, R
Hydroxylation	-H+OH (+15.9949)	+0.0000	K, P
IAA	-H+C2N1H4O (+57.0215)	+0.0000	C
MMTS	-H+SCH3 (+45.9877)	+0.0000	C
Methylation	-H+CH3 (+14.0157)	+0.0000	
Oxidation	+O (+15.9949)	+0.0000	M
Phosphorylation	-OH+H2PO4 (+79.9663)	+0.0000	S, T, Y
Sulfation	-OH+HSO4 (+79.9568)	+0.0000	
Trimethylation	-H3+C3H9 (+42.0470)	+0.0000	
Urea loss	-CO(NH2)2 (-60.0324)	+0.0000	$
Water loss	-H2O (-18.0106)	+0.0000	S, T
hanging ADHd0	-OH+C6H13O2N4 (+156....	+0.0000	$, D, E
hanging ADHd0x	-OH+C6H13O2N4 (+156....	+8.5020	$, D, E

Delete row   Save   Close

Bonds [CTRL+B]

The pre-defined library of bonds is stored in the **.\Properties\bonds.prs** file. This library is accessible by a shortcut CTRL+B or by the Main menu bar and is consisted of nine editable columns. As well as in the table of modifications, the column **Name** can also contain only letters or numbers; mass shift of cross-link can be entered as an explicit numeric value or loss/addition of elements. If isotopically labeled reagents are used for cross-linking, isotope shift can be entered to the column **Check Diff** to avoid unnecessary definition of a deuterated reagent in the column Mass. The program searches for the specific **Mass** shifts and returns positive assignment if a sum of **Mass** and **Check Diff** values is found in the mass list.

Amino acids, which can be modified by a cross-linker are written in the column **Specificity 1**; symbols "^" and "$" mark N-terminus and C-terminus. Additionally, the column **Specificity 2** was included to this library due to different reacting specificities of heterobifunctional reagents. The blank filed in any **Specificity** column describes unlimited specificity of the reagent, whereas symbol "---" in the **Specificity 2** column represents the same target amino acids as written in the first one (homobifunctional reagents).

If a cross-linking reagent is influenced by a partial hydrolysis (Type 0 cross-link) or other degrading process, altered Mass shift of Type 0 cross-link can be entered to the column **Mass 1** (again with **Check Diff 1** if required). For another mass shift caused by the different behavior of the second functional group of reagent, columns **Mass 2** and **Check Diff 2** were introduced to the library.

All performed changes must be confirmed by the ***Save*** button.

Name	Mass	Check Diff	Specificity 1	Specificity 2	Mass 1	Check Diff 1	M...	C...
ADHd0	-O2H4+C6O2N4H14 (+138.0905)	+0.0000	$, D, E	---	-OH2+C6O2N4H14 (+156.1011)	+0.0000		
BS2Gd0	-H2+C5H6O2 (+96.0211)	+0.0000	^, K, Y	---				
BS2Gd4	-H2+C5H2O2D4 (+100.0462)	+0.0000	^, K, Y	---				
BS3d0	-H2+C8H12O2 (+138.0681)	+0.0000	^, K, Y	---				
BS3d4	-H2+C8H8O2D4 (+142.0932)	+0.0000	^, K, Y	---				
DSGd0	-H2+C5H6O2 (+96.0211)	+0.0000	^, K, Y	^, K, S, Y	-H+C5H7O3 (+114.0317)	+0.0000	-H...	+...
DSGd4	-H2+C5H2O2D4 (+100.0462)	+0.0000	^, K, Y	^, K, S, Y	-H+C5H3O3D4 (+118.0568)	+0.0000	-H...	+...
DSSd0	-H2+C8H12O2 (+138.0681)	+0.0000	^, K, Y	^, K, S, Y	-H+C8H13O3 (+156.0786)	+0.0000	-H...	+...
DSSd4	-H2+C8H8O2D4 (+142.0932)	+0.0000	^, K, Y	^, K, S, Y	-H+C8H9O3D4 (+160.1038)	+0.0000	-H...	+...
Disulfide bond	-H2 (-2.0157)	+0.0000	C	---				
EDC	-H2O (-18.0106)	+0.0000	^, $, D, E, K	---				

Delete row   Save   Close

Measurement formats [CTRL+F]

To maximize the independence of the output data formats on an instrument or instrument-dependent software, LinX enables to define your own type of delimiter-separated data format. The Edit format window contains six columns. The **Name** of the format can contain letters, numbers, and in this case special characters as well. In the column called **Pattern**, the delimiter of individual values in the mass list file must be specified (use regular expressions). **Extensions** connected to your format can improve its recognition during loading the data files. The **Mass**, **Intensity** and **Retention time columns** contain numbers, what represent the position of the specific column in the input data file. Zero value in columns is used for cases, where input data file doesn't contain any of optional values.

Contact and support

Please feel free to send us any comments or bug reports. For this purpose please use mail address zdenek.kukacka@biomed.cas.cz.

Acknowledgement

The authors namely MSc. Jan Jelinek, MSc. Michal Rosulek, MSc. Zdenek Kukacka, Ph.D., Dr. Daniel Kavan, Ph.D., and Dr. Petr Novak, Ph.D. thank all collaborators for helping during never-ending period of development and you for choosing LinX.