Supporting information for: Peptigram: a web-based application for peptidomics data visualization

Jean Manguy,[†,‡,¶] Peter Jehl,[‡,¶] Eugène T. Dillon,[†,‡,§] Norman E. Davey,[‡,¶] Denis C. Shields,[*,†,‡,¶] and Thérèse A. Holton[†,‡,¶]

†*Food for Health Ireland,*

‡*UCD Conway Institute of Biomolecular and Biomedical Research,*

¶*School of Medicine,*

§*School of Biomolecular and Biomedical Science,*

University College Dublin, Belfield, Dublin 4, Ireland

Table of contents

Supplementary Text

Peptides tracks

A peptide track is generated for each sample, where the experimentally detected peptides for this sample are mapped to their position within their precursor protein. A green box is used to represent each peptide, where the peptide sequence is overlaid (see Figure 3c). The length and the position of the peptide boxes are representative of the length and position of the peptide within the precursor protein. The intensity of the green color is proportional to the relative intensity of the peptide. The numerical intensity of a peptide can be displayed by hovering over the peptide with the mouse cursor. Overlapping and adjacent peptide boxes are vertically arranged in a way that avoids confusion between distinct peptides.

Cleavage sites tracks

The position of known cleavages sites of common endopeptidases and chemicals can be displayed in dedicated tracks (see Figure 3d). Cleavage sites associated with a given peptidase are displayed in a unique color, with the name of the enzyme positioned around the cleavage site. As the peptidase color is consistent across the visualization, patterns of hydrolysis across samples can be assessed. A translucent box indicates instances where known cleavage sites of a peptidase are not observed in the input peptide data. A selection switch is available should the user wish to hide unobserved sites. Specific cleavage sites of interest to the user can be highlighted across the entire visualization by a dashed red line by simply clicking on such sites of interest.

Data display options

There are various options available to the user as to what is displayed in the peptide alignment map. The "Settings & filters" button opens a new box with "Controls" to help the user adjust the visualization as needed; see Supplementary Figure S5. The user can select which tracks

(e.g. samples, endopeptidases or other protein information) to display or not. Additionally the user can choose to filter the displayed peptide data by the observed peptide intensities (via a slider or by specifying the intensity minimum and maximum).

Peptide information box

A peptide information box is presented to the user when they click on a peptide of interest (see Supplementary Figure S4). In this box, the intensity of the peptide across samples is summarized in a table, allowing the user to quickly see variation between experimental conditions. Buttons are also provided in this view to allow the user to query external peptide databases and tools to gain more insight into the selected peptide of interest. Currently, two peptide databases, PeptideAtlas[S1] and PepBank[S2] are supported by Peptigram, as well as the peptide bioactivity prediction tool PeptideRanker.[S3]

Data searches and region selection

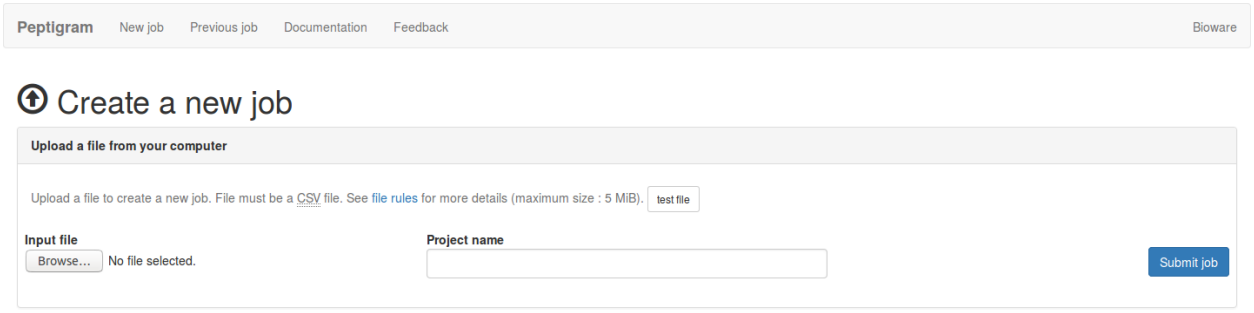
In order to locate a specific peptide of interest in the data the user can perform a search using the dedicated form. The peptide search form will accept either a peptide sequence or a regular expression. Peptides not matching the search criteria are obscured in order to highlight peptides that do (see Figure 4). For example, one could use regular expressions to highlight every peptide containing a motif or every peptide that starts or ends at a cleavage site. A separate search form allows the user to search for motifs in the precursor protein sequence or alignment (see Figure 4).

Due to the successive layers of information provided by the Peptigram tracks, it can become difficult to examine a protein region of interest in the peptide alignment map. To alleviate this, the "region selection" tool from ProViz is implemented in Peptigram (see Figure 3b). Through the use of a slider at the top of the display, or by specifying the positions in the "Focus" form, users can select the range of interest from the precursor protein. Users can also specify a focus range by clicking on a particular component of the visualization (e.g.

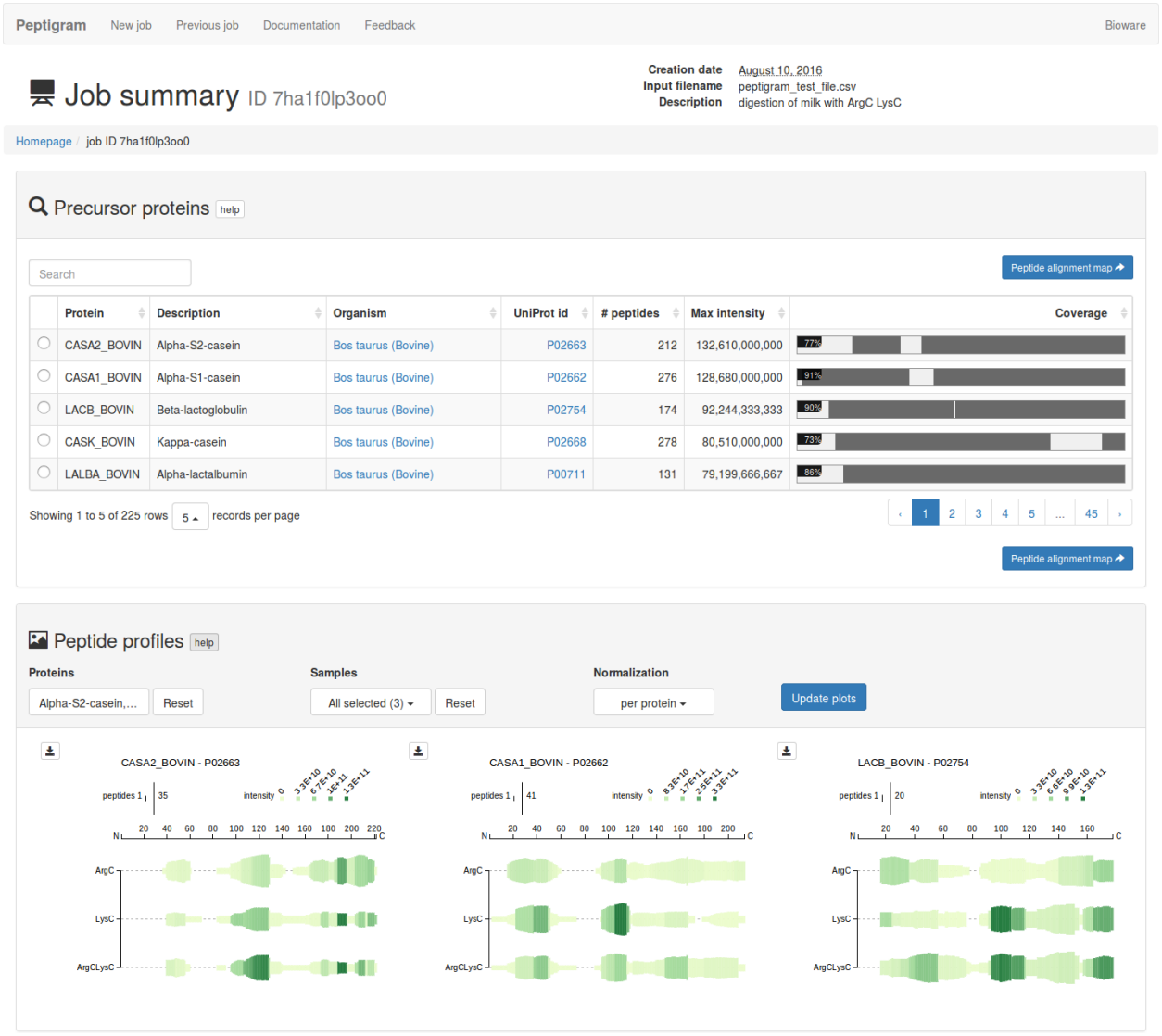
a peptide or sequence variant) and simultaneously pressing the "Ctrl" key. Once a region is selected the user can crop the visualization before exporting it to a PDF. Additionally the peptide sequences within this region can be extracted for querying in external databases or for download. This can also be achieved by extracting the sequences from the alignment.

Supplementary Figures

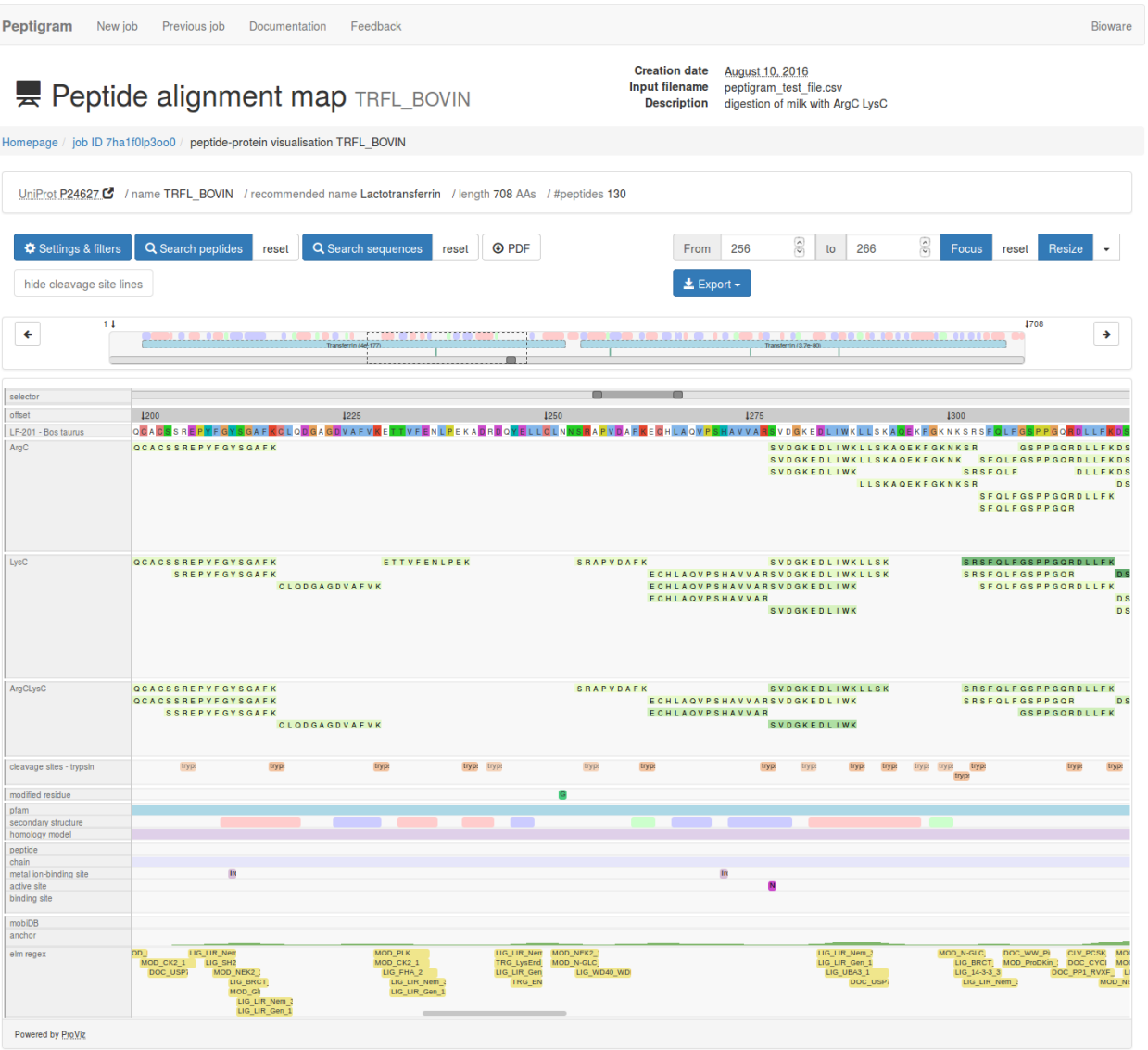
Supplementary Figure S1: The form to create a new job requires two inputs: the user's data file and a short description of the project. An example test file is also available here in order to test Peptigram.



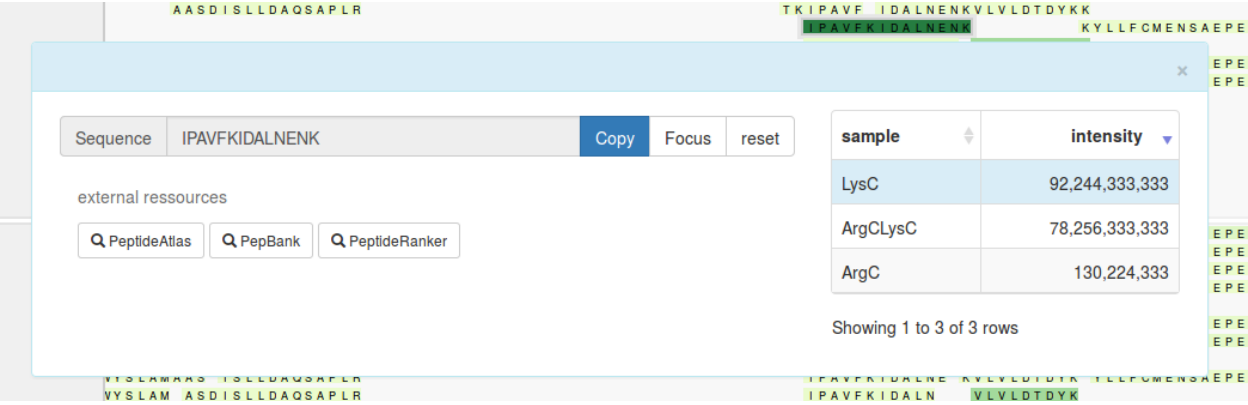
Supplementary Figure S2: The "Job summary" page is displayed once the extraction of data is complete. A table summarizes the precursor proteins, displays the peptide coverage of each protein and allows the user to select a precursor protein in order to display the relevant peptide alignment map (see Supplementary Figure S3). Below, the user can select which peptide profile visualization to display and can download them as a SVG file. Here, from our test data, the peptide profiles of α_{S2}–casein, α_{S1}–casein and β–lactoglobulin are displayed.



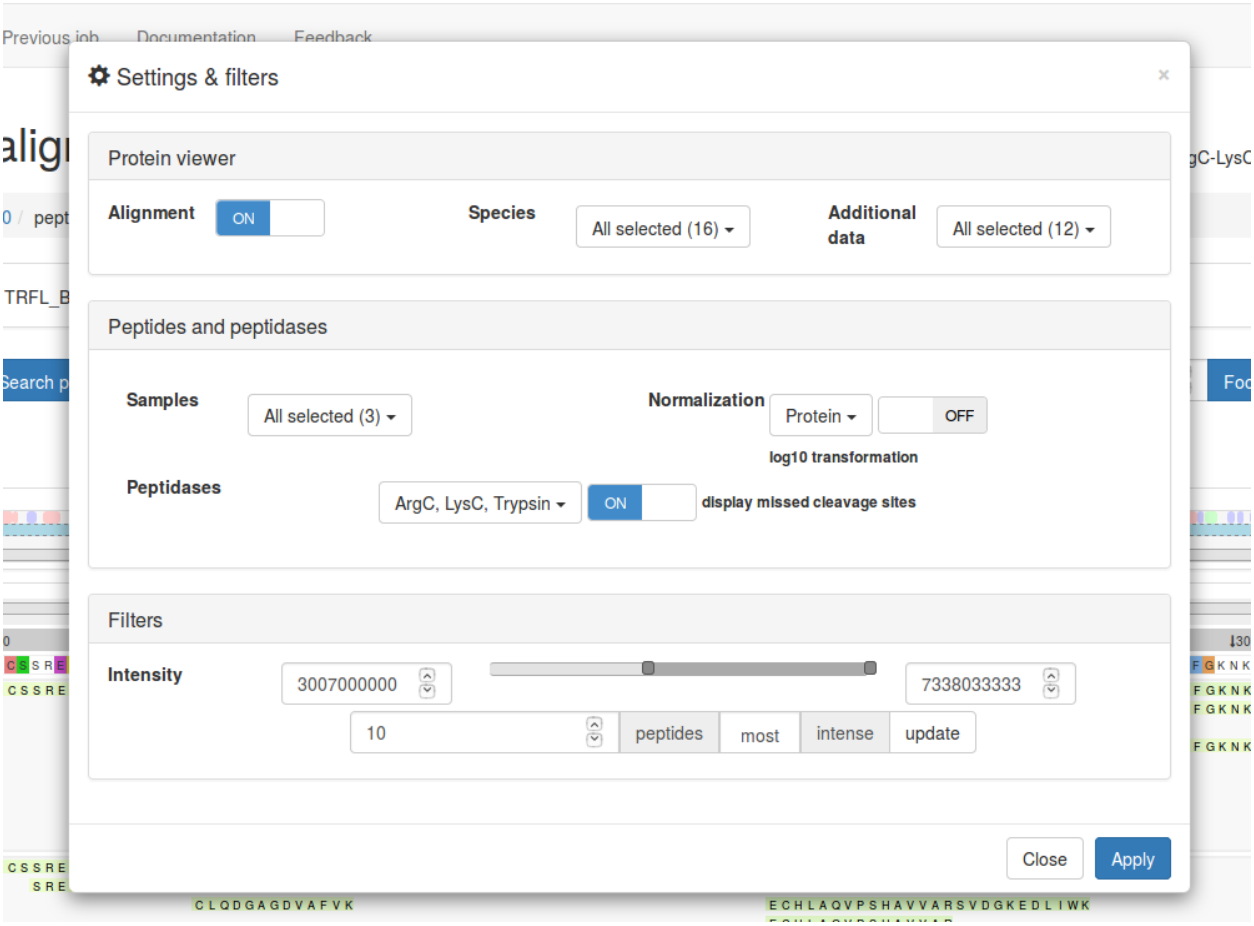
Supplementary Figure S3: The peptide alignment map page contains the sequence data visualization, in addition to an interface to control it and information about the precursor protein. The interface contains numerous buttons and input fields such as: the "Settings & filters" button to open a panel of the same name (see Supplementary Figure S5), buttons to highlight peptides or motifs in the precursor protein using regular expressions; or buttons to mark a part of the precursor protein. Below, a box summarizes the precursor protein and allows the user to quickly access a part of the protein without needing to scroll through the entire peptide alignment map. The dashed line represents the part of the alignment map currently visible in the browser.



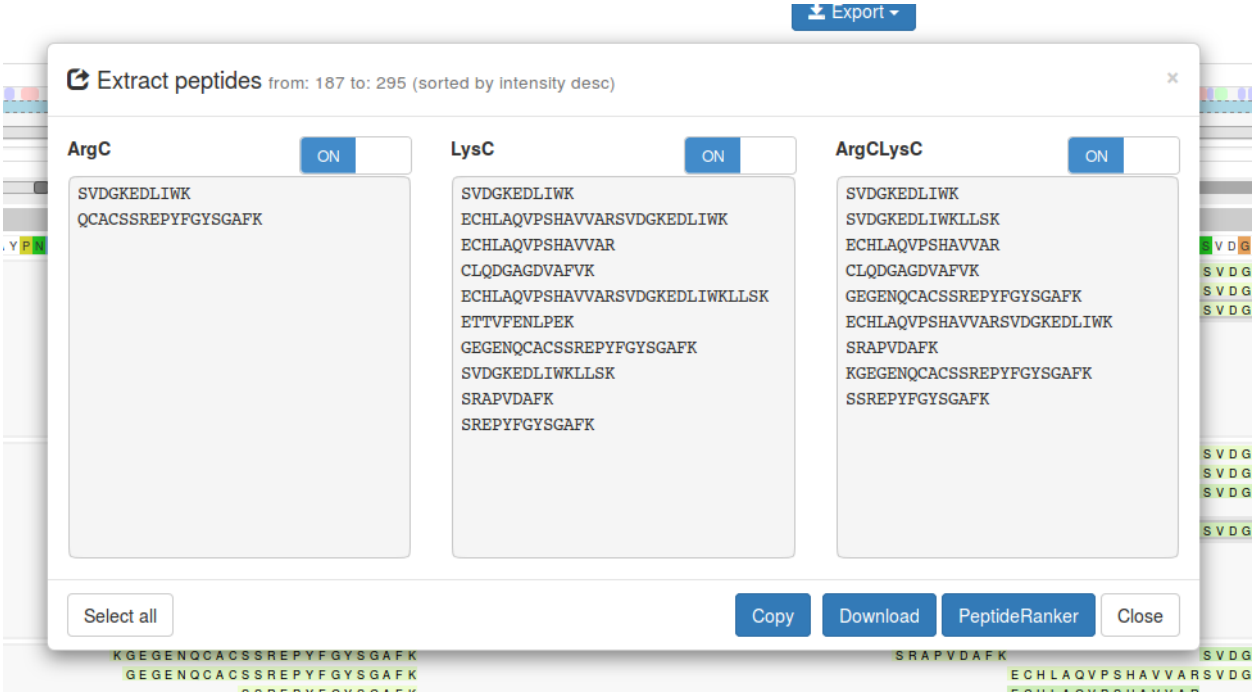
Supplementary Figure S4: The "peptide information box" is displayed when the user clicks on a peptide box. The first field on the left allows the user to copy the peptide sequence or to highlight the position of the selected peptide with the "Focus" button. Below, buttons open a new window where the user can launch searches in external databases (i.e. PeptideAtlas[S1] and PepBank[S2]) and a prediction tool (PeptideRanker[S3]). On the right, the table summarizes the intensities of the selected peptide in each sample. The row highlighted in blue corresponds to the peptide box the user clicked on to display this panel. Here, the panel was displayed by clicking on the "IPAVFKIDALNENK" peptide box in the LysC peptide track of the β–lactoglobulin peptide alignment map.



Supplementary Figure S5: The "Settings & filters" panel allows the user to dynamically modify peptide alignment map visualizations. In the top panel named "Protein viewer", the user can modify settings relating to the precursor protein sequence (or the alignment if available) and the sequence information supplied by ProViz.[S4] The second panel, named "Peptides and peptidases" can be used to select which samples to display, how to normalize the peptide intensities and which cleavage sites to display. The last panel, named "Filters" allows the user to restrict the displayed peptides by intensity. To do this, the user can either directly input the intensity thresholds, use the dedicated slider or specify a number of most (or least) intense peptides. After selecting their parameters the user can click on the "Apply" button to implement their selections.



Supplementary Figure S6: The "Extract peptides" panel can be accessed via the "Extract" button once a region of the precursor protein is selected (by default the full protein is extracted). The user can select samples and then copy or download the list of associated peptide sequences. From here the user can also launch PeptideRanker.[S3]



References

(S1) Desiere, F.; Deutsch, E. W.; King, N. L.; Nesvizhskii, A. I.; Mallick, P.; Eng, J.; Chen, S.; Eddes, J.; Loevenich, S. N.; Aebersold, R. The PeptideAtlas project. *Nucl. Acids Res.* **2006**, *34*, D655–D658.

(S2) Shtatland, T.; Guettler, D.; Kossodo, M.; Pivovarov, M.; Weissleder, R. PepBank - a database of peptides based on sequence text mining and public peptide data sources. *BMC Bioinformatics* **2007**, *8*, 280.

(S3) Mooney, C.; Haslam, N. J.; Pollastri, G.; Shields, D. C. Towards the improved discovery and design of functional peptides: common features of diverse classes permit generalized prediction of bioactivity. *PLoS ONE* **2012**, *7*, e45012.

(S4) Jehl, P.; Manguy, J.; Shields, D. C.; Higgins, D. G.; Davey, N. E. ProViz - a web-based visualization tool to investigate the functional and evolutionary features of protein sequences. *Nucl. Acids Res.* **2016**, *44*, W11–W15.