Published in final edited form as:

Nat Biotechnol. 2009 July ; 27(7): 600–601. doi:10.1038/nbt0709-600.

NCBI Peptidome: a new public repository for mass spectrometry peptide identifications

Douglas J Slotta, Tanya Barrett, and Ron Edgar

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA

To the Editor

Peptidome (available at <http://www.ncbi.nlm.nih.gov/peptidome/>) is a new public resource that archives and freely distributes tandem mass spectrometry (MS) peptide and protein identification data generated by the scientific community (Fig. 1). The database is administered by the National Center for Biotechnology Information (NCBI; Bethesda, MD, USA) at the National Institutes of Health. The core structure of the Peptidome database is based on NCBI's Gene Expression Omnibus (GEO^{1,2}), and uses much of the same source code for handling submissions and accessing the database. Peptidome, like GEO, was constructed to promote sharing and dissemination of experimental data at a level of detail that is useful to both the specialized community and to the general biological community.

Great emphasis has been placed on making the deposit procedures as simple as possible, yet still supporting a high level of experimental annotation. Although data from all stages of the experiment are captured, the burden of submission is minimized by accepting standard file formats and using spreadsheet templates for metadata. This method was chosen to allow submissions to be generated either programmatically or by hand, using tools familiar to everyone, without the delay of waiting for the perfect file format and slick graphical user interface to be developed. As the field matures, so will the data submission process. A Peptidome submission includes the following components:

1. Biological and methodological metadata. This information gives context to the experimental data and facilitates understanding of the goals of the study, the material under investigation, the instrumentation employed and the protocols used to generate and analyze the data.
2. The original raw spectra data, converted to open format. Raw data includes all of the original MS1 and MS2 data before any post-processing, regardless of whether it was identified or not. The availability of raw data is critical for independent verification and validation of the reported study results. Furthermore, collections of raw data greatly benefit the MS community in wider applications, such as preparing custom spectral libraries, conducting large-scale statistical studies of peptide fragmentation and developing new peptide-matching algorithms.
3. Peptide identification output files. These files contain matched peptide sequence information derived from the observed spectra using any search algorithm.
4. Conclusion-level results. These include the final, filtered list of identified proteins, and the peptides leading to those assignments, as determined by the submitter. This finalized list has much flexibility in that it allows inclusion of identifications from

new algorithms or *de novo* identifications. It also allows the submitter to provide complete results independent of the methodology used, for example, cutoff scores, reverse database searches for false positives or agreement between multiple search engines. Peptidome processing procedures augment these lists with auxiliary information extracted from the peptide identification files, including scores and post-translational modifications. This augmented and enriched list is a convenient mechanism to access the relevant data discussed in an associated manuscript.

Submissions are organized into studies and samples. A sample includes all results from the same biological source, regardless of the number of instrument runs required to collect the data. A study is a collection of samples related in some fashion, for example, one sample may be the control and another the diseased state, or each sample may be one slice of a time series. Each sample has its own list of identified proteins, peptides and spectra. All studies and samples are assigned unique and stable accession numbers that may be used to cite and retrieve the records. Data may remain private for a limited time, typically pending manuscript publication, and during this period submitters may grant journal reviewers anonymous access to their data. Public data may be browsed by studies or by samples, and each sample may be viewed either protein-centric or peptide-centric. The evidence for each peptide may also be explored down to the level of individual spectra. The current focus is on providing a complete view of all the data; additional functionality will be added in the future to aid in the analysis of the data, such as annotating spectra displays with theoretical fragment ions. In addition to displaying results through the website, all originally submitted files may be downloaded for offline analysis.

Peptidome incorporates several measures to increase the quality, value and utility of submitted data. One of the distinguishing features of Peptidome is the high level of human support and curation in the submission process intended to ensure sufficient metadata and completeness of the submission. In addition, an attempt is made to match submitted identified protein names to NCBI protein designations, which improves accessibility to the larger biological community; this allows the experimental results to be retrieved by an independent search for a given protein or gene locus, without a priori knowledge of the pertinent experiment. Users can locate studies and proteins of interest by querying NCBI's powerful Entrez search system with relevant keywords, protein names, accessions and more. Data navigation is further enhanced through inter-database links that reciprocally connect Peptidome to other NCBI resources such as PubMed, GenBank and Gene.

In the spirit of collaborative data-sharing, Peptidome is entering into data exchange agreements with established MS resources including PRIDE3, PeptideAtlas4, Tranche5 and other ProteomExchange members. We hope to expand this network in the future by collaborating with other, more specialized proteomic resources, such as Human ProteinPedia6. In spite of any apparent duplication of effort and design of these resources, we believe the community will benefit from alternative analysis or presentation styles and the cooperation of disparate repositories in data deposits and dissemination options.

The initial exemplar submissions in Peptidome have been culled from selected submissions to PeptideAtlas, with the metadata enriched from papers associated with the experiments. Sharing and linking between resources will maximize the visibility and utility of public MS peptide identification data. We invite the proteomics community to support and participate in these data-sharing endeavors7 by contributing to the Peptidome resource, and we welcome input and suggestions.

Acknowledgments

The authors would like to thank D. Rudnev, D. Troup, R. Muerter, A. Soboleva, M. Tomashevsky and O. Ayanbule for their aid in the development of this project. Advice and support was provided by L. Geer, S. Sechi, S. Markey and

the rest of the NIMH Laboratory of Neurotoxicology. This research was supported by the Intramural Research Program of the National Library of Medicine and the Roadmap, National Institutes of Health.

References

1. Edgar R, Domrachev M, Lash AE. *Nucleic Acids Res* 2002;30:207–210. [PubMed: 11752295]
2. Barrett T, et al. *Nucleic Acids Res* 2009;37:D885–D890. [PubMed: 18940857]
3. Jones P, et al. *Nucleic Acids Res* 2008;36:D878–D883. [PubMed: 18033805]
4. Deutsch EW, Lam H, Aebersold R. *EMBO Rep* 2008;9:429–434. [PubMed: 18451766]
5. Andrews, PC. 56th ASMS Conference on Mass Spectrometry and Allied Topics, American Society for Mass Spectrometry; Denver, CO. 2008.
6. Mathivanan S, et al. *Nat Biotechnol* 2008;26:164–167. [PubMed: 18259167]
7. *Nat Biotechnol* 2007;25:262.

The image is a screenshot of the Peptidome NCBI Peptide Data Resource homepage. At the top, there is a navigation bar with the NCBI logo on the left and the Peptidome logo on the right. Below the logos, there is a horizontal menu with links: Peptidome Home, Browse Data, Search Data, Submit Data, Submission Guidelines, Contact Us, and FAQ. The main content area is titled "About Peptidome" and contains a paragraph describing the resource as a public repository for tandem mass spectrometry peptide and protein identification data. Below this paragraph is a list of four bullet points: Lists of identified proteins, Lists of identified peptides used in protein identification, Mass spectra that provide evidence for these identifications, and Descriptive information about the biological samples, instrumentation, and informatics. To the right of the main content area is a table titled "Total Holdings" with four rows: Studies (15), Samples (87), Proteins (30108), and Peptides (128538). Below the "About Peptidome" section is a section titled "Presentations and posters" with a link to view the ASMS 2009 poster. Below that is a "Browse" section with a link to browse studies. Below that is a "Submit" section with a link to submission guidelines. At the bottom of the page, there is a footer with contact information, copyright, privacy, and accessibility links, and logos for the National Center for Biotechnology Information, US National Library of Medicine, and the USA.gov logo.

NCBI • Peptidome

Not logged in | login

About Peptidome

Peptidome is a public repository that archives and freely distributes tandem mass spectrometry peptide and protein identification data generated by the scientific community. Several layers of data are captured to promote understanding of the experiment and analysis of the underlying data, including:

- Lists of identified proteins
- Lists of identified peptides used in protein identification
- Mass spectra that provide evidence for these identifications
- Descriptive information about the biological samples, instrumentation, and informatics

[More information »](#)

Total Holdings

Studies	15
Samples	87
Proteins	30108
Peptides	128538
Spectra	6896500

Presentations and posters

[Click here to view our ASMS 2009 poster](#)

Browse

In addition to archiving data, web-based interfaces are provided to help users browse and explore studies, peptides and proteins.

[Browse studies »](#)

Submit

Peptidome accepts tandem mass spectrometry data submissions from the proteomics community. Submission procedures are flexible and made as simple as possible by accepting native file formats and straightforward spreadsheet templates. Submissions are assigned stable accession numbers that may be cited in manuscripts describing the experiments. Submitters may hold their Peptidome records private until published.

[Submission guidelines »](#)

Last modified: Friday, 12-Jun-2009 10:22:52 EDT

[Contact](#) | [Copyright](#) | [Privacy](#) | [Accessibility](#)

National Center for Biotechnology Information, US National Library of Medicine
8600 Rockville Pike, Bethesda, MD USA 20894

Figure 1.
The homepage of Peptidome, NCBI's peptide data resource.