Journal of proteome research

Technical Note

pubs.acs.org/jpr

PepShell: Visualization of Conformational Proteomics Data

Elien Vandermarliere,[†,‡,∥] Davy Maddelein,[†,‡,∥] Niels Hulstaert,[†,‡] Elisabeth Stes,[†,‡] Michela Di Michele,[†,‡] Kris Gevaert,[†,‡] Edgar Jacoby,[§] Dirk Brehmer,[§] and Lennart Martens*[,†,‡]

[†]Department of Medical Protein Research, VIB, Albert Baertsoenkaai 3, B-9000 Ghent, Belgium

[‡]Department of Biochemistry, Faculty of Medicine and Health Sciences, Ghent University, Albert Baertsoenkaai 3, B-9000 Ghent, Belgium

[§]Oncology Discovery, Janssen Research and Development, A Division of Janssen Pharmaceutica NV, Turnhoutseweg 30, B-2340 Beerse, Belgium

Supporting Information

ABSTRACT: Proteins are dynamic molecules; they undergo crucial conformational changes induced by post-translational modifications and by binding of cofactors or other molecules. The characterization of these conformational changes and their relation to protein function is a central goal of structural biology. Unfortunately, most conventional methods to obtain structural information do not provide information on protein dynamics. Therefore, mass spectrometry-based approaches, such as limited proteolysis, hydrogen−deuterium exchange, and stable-isotope labeling, are frequently used to characterize protein conformation and dynamics, yet the interpretation of these data can be cumbersome and time consuming. Here, we present PepShell, a tool that allows interactive data analysis of mass spectrometry-based conformational proteomics studies by visualization of the identified peptides both at the sequence and structure levels. Moreover, PepShell allows the comparison of experiments under different conditions, including different proteolysis times or binding of the protein to different substrates or inhibitors.

KEYWORDS: *Comparison of peptides, conformational proteomics, visualization of peptides, PepShell, protein structure*

INTRODUCTION

Proteins are not merely rigid entities but dynamic molecules that can undergo drastic conformational changes induced by post-translational modifications or binding to cofactors, substrates, inhibitors, or other biomolecules. Because these changes are often essential for their biological function, the analysis and visualization of the dynamic properties of a protein are important aspects of the exploration of the working mechanism of a protein. Nevertheless, the classic methods to analyze protein structures are not able to resolve these dynamic features. The outcome of a protein crystallographic experiment is essentially a static image with little information on the flexibility of the protein because the protein molecules are frozen in fixed positions within the crystal lattice.[1] This limitation is overcome with nuclear magnetic resonance (NMR). In an NMR experiment, the structure is determined in solution and therefore flexibility is retained and can thus ultimately be measured.[2] Unfortunately, NMR is limited to a protein size of 50 kDa;[3] thus, many multidomain proteins are excluded from NMR analysis. Yet, it is especially in these multidomain proteins that conformational plasticity is very important: signal transfer within multidomain proteins frequently occurs through structural alterations.

To overcome the hurdles of the classic methods, the study of protein structure with mass spectrometry, called conformational proteomics, is frequently used to characterize protein conformations and dynamics. Conformational proteomic approaches, such as limited proteolysis,[4] hydrogen−deuterium exchange,[5] MS footprinting,[6] and stable-isotope labeling,[7] are becoming increasingly popular as alternative tools to unravel the structural dynamics of a protein or even whole proteomes.[8] All of these methods exploit the surface accessibility of amino acids within the native protein conformation or, more specifically, the differences in protein surface accessibility in different situations within a protein structure. Limited proteolysis exploits the accessibility of cleavable positions that are specific to the protease used.[9] Depending on the conformational state of the protein of interest, different cleavage sites of the protein will be exposed, which results in a difference in released peptides when subjected to limited proteolysis. A hydrogen−deuterium exchange experiment takes advantage of the exchange rate between deuterium atoms from D_2O and hydrogen atoms from amide groups. This rate of exchange depends both on the solvent accessibility and

Received: November 24, 2014
Published: March 2, 2015

DOI: 10.1021/pr5012125
J. Proteome Res. 2015, 14, 1987−1990

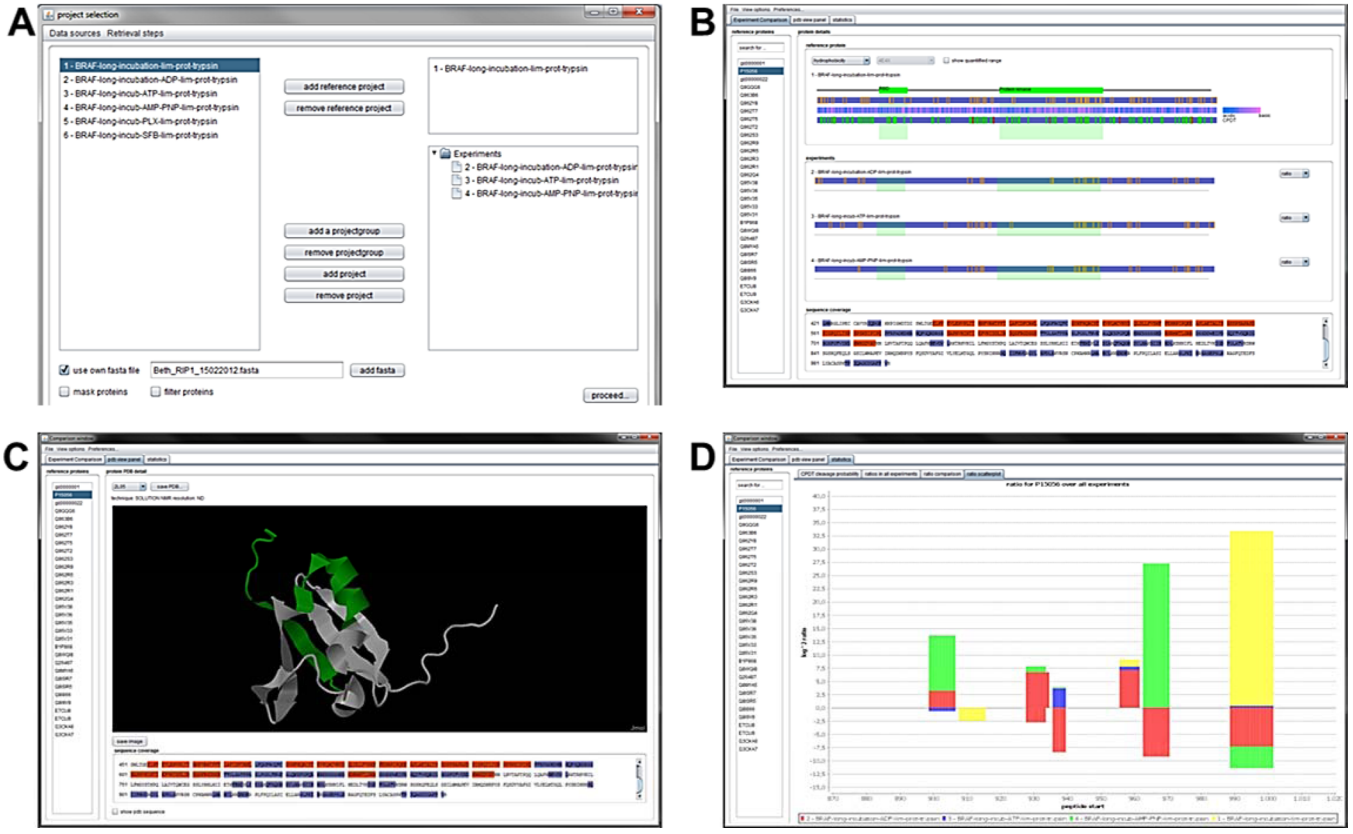

Figure 1. (A) PepShell project selection interface. Here, the user can choose the reference project and any other projects that need to be analyzed and compared to each other. (B) Experiment comparison panel in PepShell. This panel allows comparison of the selected experiments at the sequence level. (C) PDB view panel in PepShell. Here, the user can compare experiments at the structure level. (D) Statistics panel in PepShell. This panel allows the user to compare the enzymatic cleavage probability as well as the relative quantification of the peptides identified across the experiments.

hydrogen-bond status of the amide group.[10] An MS footprinting experiment also labels solvent-accessible residues, but here, the hydroxyl radical is most frequently used as labeling agent and the accessible side chains are labeled.[11] Stable-isotope labeling enables the labeling of proteins *in vivo*, but it also allows chemical labeling of specific, solvent-accessible amino acids. Examples are the labeling of cysteines and lysines with *N*-ethylmaleimide and succinic anhydride, respectively.[7]

The experimental setup and subsequent workflow of a conformational proteomics experiment do not deviate drastically from that of a classic mass spectrometry-based experiment in which the peptides present in a complex peptide mixture are identified. In the case of hydrogen−deuterium exchange, MS footprinting, and stable-isotope labeling, solvent-accessible regions of the protein are first labeled, followed by protein denaturation and digestion. In limited proteolysis experiments, the denaturation step is skipped, and only partial digestion is allowed by drastically reduced digestion times or protease to substrate ratios. In all technologies, proteolysis is followed by LC−MS/MS analysis and subsequent peptide identification and quantification. Any introduced modifications must, of course, be taken into account during peptide identification. The final outcome of a conformational proteomics experiment is a list of peptides and, depending on the specific approach used, their associated properties, such as the degree of deuteration or identification of labeled amino acids. Moreover, the data typically span multiple experimental conditions across which the structural observations are to be compared. To be able to gain information about the dynamics and conformation of a protein of interest, the peptide lists from these different experiments finally have to be combined and, if available, mapped onto the structure of the protein.

To fulfill these latter steps, we developed PepShell, a tool that allows interactive data analysis of mass spectrometry-based conformational proteomics experiments by visualization of the generated peptides both at the sequence and structure levels. PepShell allows the comparison of protein structural changes under different experimental conditions, for example, different durations of proteolysis or exposure to a labeling agent, binding to different substrates or inhibitors, or different environmental conditions. PepShell is thus ideally suited to guide the interpretation of mass spectrometry-based proteomics data and their interpretation in the context of protein structure and dynamics.

■ TOOL DESCRIPTION AND FUNCTIONALITY

PepShell aids the user in interpreting the outcome of conformational proteomics experiments and is composed of three panels: the experiment comparison panel, the PDB view panel, and the statistics panel. PepShell is an open source Java application that is released under the permissive Apache2 license. The PepShell binaries, source code, and documentation can be found at the project Web site at http://code.google.com/p/pepshell. The online PepShell manual, which contains a comprehensive example, is also made available in the Supporting Information.

DOI: 10.1021/pr5012125
J. Proteome Res. 2015, 14, 1987−1990

The Data To Analyze

Usually, a conformational proteomics experiment is composed of several parallel experiments: different substrates or inhibitors might be tested, different time points can be used, or different conditions might be explored. The ultimate purpose is to analyze the differences between these experiments and, as such, gain structural information. PepShell allows the input from a broad range of structural proteomics experiments such as limited proteolysis, hydrogen−deuterium exchange, MS footprinting, and stable-isotope labeling, requiring only that these data are presented in a comma-separated text file format. In addition to the text-based input format, PepShell also allows the user to connect to an in-house data management system such as ms-lims.[12] Data can be added separately for each experiment or can be grouped into a single file that contains the merged data.

The project selection interface (Figure 1a) subsequently allows the user to select a reference project from the different performed experiments and to indicate which setups need to be compared with each other. For example, the reference project can be a conventional digestion experiment: the protein of interest is denatured and subsequently fully labeled and digested. This allows the user to gain insight into the regular labeling and digestion pattern of the protein, revealing commonly missed cleavages as well as undetectable peptides.

Experiment Comparison within PepShell

The experiment comparison panel (Figure 1b) of PepShell allows the comparison of the selected experimental setups at the sequence level. The reference project is displayed in the top of this panel. For each protein (selected in the list on the left), the user can display the peptides generated in this reference project. Moreover, relevant sequence-specific information can be visualized, including protein domain information extracted from UniProt,[13] the hydrophobicity of the amino acids, or the trypsin cleavage propensity as predicted by the CP-DT tool.[14] It should be noted that the latter property provides the user information with the probability of cleavage at the given site by the protease trypsin and is therefore relevant only if trypsin was used in the experiment.

The middle part of the experiment comparison panel displays information related to the main objective of PepShell: a visualization of the data output from different experimental setups. For each experimental condition, the identified and quantified peptides are mapped onto the sequence of the protein of interest. Depending on the quantitation strategy of the experiment, the user can opt to visualize either the ratio of observed peptides (e.g., heavy versus light labeling) or their peak intensity. This display allows for a straightforward comparison of the differences in peptide quantity over the different experiments, which, in turn, can be interpreted in terms of surface accessibility to finally yield structural information.

In the lower part of the panel, the protein sequence is provided. The coloring scheme of the sequence provides information on the location of the peptides in the sequence and shows the delineation of the different domains.

PepShell's PDB View Panel

This panel (Figure 1c) maps the detected peptides on the protein structure. Here, the main requirement is the availability of a 3D structure of the protein of interest. Structures that were resolved via protein crystallography or NMR can be retrieved from the Protein Data Bank (PDB),[15] but in-house available structures also can be loaded via a connection to a local structure database. Alternatively, a single flat pdb file can be used. PepShell allows the user to select a specific structure if several structures are available for the protein, and the peptides are subsequently mapped onto this structure. Visualization is performed using Jmol,[16] an interactive Java-based viewer for structures.

Statistics within PepShell

The statistics panel (Figure 1d) allows the user to analyze the peptides of interest in more detail. A first type of data that is provided is the outcome of the CP-DT[14] prediction for tryptic cleavage probability for each tryptic cleavage position. Another tab allows the detailed comparison of the peptide ratios over the different experimental setups. In a limited proteolysis experiment, for example, different peptides will be generated depending on the experimental conditions, the small molecule, inhibitor, or cofactor bound, or the proteolysis conditions (protease, protease/substrate ratio, and duration). For example, solvent-accessible residues are expected to get processed even at very brief proteolysis times; hence, their derived peptides will be available in high amounts under most conditions tested. On the other hand, less accessible residues require more time before they will be proteolyzed and therefore their derived peptides will be available in smaller amounts upon brief proteolysis, whereas their amount will increase with increasing proteolysis time. This will give an indication on the different protein conformations adopted by a protein under the different experimental conditions. In labeling experiments, such as MS footprinting, stable-isotope labeling, and hydrogen−deuterium exchange experiments, the degree of labeling of a peptide will change over time. Peptides that originate from a region within the protein that is highly accessible will already be strongly labeled after a brief labeling time, whereas peptides that contain less accessible residues will require longer labeling times or might even not get fully labeled. Comparison of these differences for one peptide over the different experimental setups can reflect structural changes in a specific region and, as such, allows the mapping of protein dynamics.

■ CONCLUSIONS

The increasing popularity of structural proteomics is in stark contrast with the availability of efficient tools to visualize this multitude of new data. There are, however, some tools available that aid data interpretation. MTMDAT[17] supports the analysis of limited proteolysis data and provides the changes in relative abundance of the peptide fragments over time. Its extension, MTMDAT-HADDOCK,[18] allows the visualization of protein interactions. HDX data can be analyzed with the aid of QUDeX-MS,[19] which allows the estimation of deuterium incorporation, or HDX workbench,[20] Hydra,[21] or Hexicon,[22] which allow management and visualization of HDX data. All of these tools are, however, approach-specific and are aimed primarily at mass spectrometrists with a specific focus on the experimental mass spectrometry data and their processing and interpretation. We have therefore developed PepShell, a Java-based tool that allows overarching interactive data analysis of mass spectrometry-based conformational proteomics experiments by visualization of the generated peptides both at the sequence and structural levels and does so for a broad range of structural proteomics experimental setups. In contrast to the above-mentioned tools, PepShell is thus intended to support downstream users, typically not mass spectrometry specialists,

DOI: 10.1021/pr5012125
J. Proteome Res. 2015, 14, 1987−1990

to interpret the results obtained from a variety of conformational proteomics approaches through the integrative visualization of these results with data collected from local or remote resources.

Where most tools available for mass spectrometry-based proteomics data analysis provide merely a list of measured peptides and their associated properties, PepShell enables the analysis of differential experimental setups. It uses these peptide lists to compare different experimental conditions, and, importantly, it allows the visualization of these differences onto the structure of the protein. As such, PepShell bridges the gap between mass spectrometry-based proteomics data and their interpretation in the context of protein structure and dynamics.

■ ASSOCIATED CONTENT

Supporting Information

PepShell manual. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: lennart.martens@ugent.be; Tel.: +32 92649358; Fax: + 32 92649484.

Author Contributions

‖E.V. and D.M. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by a research project grant (IWT 110431) of the Flemish government agency for Innovation by Science and Technology (IWT). E.S. is a postdoctoral researcher of the Research Foundation-Flanders (FWO). L.M. and K.G. acknowledge the support of Ghent University (Multidisciplinary Research Partnership "Bioinformatics: From Nucleotides to Networks"). Part of this work was funded through the EU FP7 PRIME-XS project (grant 262067). We thank Dr. Herman Van Vlijmen for the critical assessment of the manuscript.

■ ABBREVIATIONS

LC, liquid chromatography; MS, mass spectrometry; NMR, nuclear magnetic resonance; PDB, Protein Data Bank; SILAC, stable-isotope labeling by amino acids in cell culture

■ REFERENCES

(1) McPherson, A. *Introduction to Macromolecular Crystallography.*; John Wiley & Sons: Hoboken, NJ, 2003.

(2) Ishima, R.; Torchia, D. A. Protein dynamics from NMR. *Nat. Struct. Biol.* **2000**, *7*, 740−743.

(3) Wider, G.; Wüthrich, K. NMR spectroscopy of large molecules and multimolecular assemblies in solution. *Curr. Opin Struct. Biol.* **1999**, *9*, 594−601.

(4) Milanovic, M.; Kracht, M.; Schmitz, M. The cytokine-induced conformational switch of nuclear factor kB p65 is mediated by p65 phosphorylation. *Biochem. J.* **2014**, *457*, 401−413.

(5) West, G.; Pascal, B.; Ng, L.; Soon, F.; Melcher, K.; Xu, H.; Chalmers, M.; Griffin, P. Protein conformation ensembles monitored by HDX reveal a structural rationale for abscisic acid signaling protein affinities and activities. *Structure* **2013**, *21*, 229−235.

(6) Gau, B.; Garai, K.; Frieden, C.; Gross, M. Mass spectrometry-based protein footprinting characterizes the structures of oligomeric apolipoprotein E2, E3, and E4. *Biochemistry* **2011**, *50*, 8117−8126.

(7) Kahsai, A.; Rajagopal, S.; Sun, J.; Xiao, K. Monitoring protein conformational changes and dynamics using stable-isotope labeling and mass spectrometry. *Nat. Protoc.* **2014**, *9*, 1301−1319.

(8) Feng, Y.; De Franceschi, G.; Kahraman, A.; Soste, M.; Melnik, A.; Boersema, P. J.; de Laureto, P. P.; Nikolaev, Y.; Oliveira, A. P.; Picotti, P. Global analysis of protein structural changes in complex proteomes. *Nat. Biotechnol.* **2014**, *32*, 1036−1044.

(9) Fontana, A.; de Laureto, P. P.; Spolaore, B.; Frare, E.; Picotti, P.; Zambonin, M. Probing protein structure by limited proteolysis. *Acta Biochim. Polym.* **2004**, *51*, 299−321.

(10) Yan, X.; Watson, J.; Ho, P. S.; Deinzer, M. L. Mass spectrometric approaches using electrospray ionization charge states and hydrogen−deuterium exchange for determining protein structures and their conformational changes. *Mol. Cell. Proteomics* **2004**, *3*, 10−23.

(11) Guan, J.; Chance, M. Structural proteomics of macromolecular assemblies using oxidative footprinting and mass spectrometry. *Trends Biochem. Sci.* **2005**, *30*, 583−592.

(12) Helsens, K.; Colaert, N.; Barsnes, H.; Muth, T.; Flikka, K.; Staes, A.; Timmerman, E.; Wortelkamp, S.; Sickmann, A.; Vandekerckhove, J.; Gevaert, K.; Materns, L. ms_lims, a simple yet powerful open source laboratory information management system for MS-driven proteomics. *Proteomics* **2010**, *10*, 1261−1264.

(13) Jain, E.; Bairoch, A.; Duvaud, S.; Phan, I.; Redaschi, N.; Suzek, B. E.; Martin, M. J.; McGarvey, P.; Gasteiger, E. Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinf.* **2009**, *10*, 136.

(14) Fannes, T.; Vandermarliere, E.; Schietgat, L.; Degroeve, S.; Martens, L.; Ramon, J. Predicting tryptic cleavage from proteomics data using decision tree ensembles. *J. Proteome Res.* **2013**, *12*, 2253−2259.

(15) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(16) Herraez, A. Biomolecules in the computer: Jmol to the rescue. *Biochem Mol. Biol. Educ.* **2006**, *34*, 255−261.

(17) Hennig, J.; Hennig, K.; Sunnerhagen, M. MTMDAT: Automated analysis and visualization of mass spectrometry data for tertiary and quaternary structure probing of proteins. *Bioinformatics* **2008**, *24*, 1310−1312.

(18) Hennig, J.; de Vries, S.; Hennig, K.; Randles, L.; Walters, K.; Sunnerhagen, M.; Bonvin, A. MTMDAT-HADDOCK: high-throughput, protein complex structure modeling based on limited proteolysis and mass spectrometry. *BMC Struct. Biol.* **2012**, *12*, 29.

(19) Salisbury, J.; Liu, Q.; Agar, J. QUDeX-MS: hydrogen/deuterium exchange calculation for mass spectra with resolved isotopic fine structure. *BMC Bioinf.* **2014**, *15*, 403.

(20) Pascal, B.; Willis, S.; Lauer, J.; Landgraf, R.; West, G.; Marciano, D.; Novick, S.; Goswami, D.; Chalmers, M.; Griffin, P. HDX workbench: software for the analysis of H/D exchange MS data. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1512−1521.

(21) Slysz, G.; Baker, C.; Bozsa, B.; Dang, A.; Percy, A.; Bennett, M.; Schriemer, D. Hydra: software for tailored processing of H/D exchange data from MS or tandem MS analyses. *BMC Bioinf.* **2009**, *10*, 162.

(22) Lindner, R.; Lou, X.; Reinstein, J.; Shoeman, R.; Hamprecht, F.; Winkler, A. Hexicon 2: automated processing of hydrogen-deuterium exchange mass spectrometry data with improved deuteration distribution estimation. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 1018−1028.

DOI: 10.1021/pr5012125
J. Proteome Res. 2015, 14, 1987−1990