SUPPLEMENTARY METHODS

Parallel Post-Translational Modification Scanning enhancing Hydrogen-Deuterium Exchange-Mass Spectrometry Coverage of Key Structural Regions

Hongda Liu,[†,#,∇] Dawei Wang,[†,#] Qiangmin Zhang,[†,#] Yang Zhao,[†] Tatyana Mamonova,[†] Lei Wang,[†] Cheng Zhang,[†] Sheng Li,[*,⊥] Peter A. Friedman,[†,‡] and Kunhong Xiao[*,†,§,∥]

[†]Laboratory for GPCR Biology, Department of Pharmacology and Chemical Biology, [‡]Department of Structural Biology, [§]Vascular Medicine Institute, and [∥]Biomedical Mass Spectrometry Center, University of Pittsburgh School of Medicine, 200 Lothrop Street, Pittsburgh, Pennsylvania 15213, United States
[⊥]Department of Medicine, University of California San Diego, La Jolla, California 92093, United States

ABSTRACT: Hydrogen-deuterium exchange-mass spectrometry (HDXMS) is a powerful technology to characterize conformations and conformational dynamics of proteins and protein complexes. HDXMS has been widely used in the field of therapeutics for the development of protein drugs. Although sufficient sequence coverage is critical to the success of HDXMS, it is sometimes difficult to achieve. In this study, we developed a HDXMS data analysis strategy that includes parallel post-translational modification (PTM) scanning in HDXMS analysis. Using a membrane-delimited G protein-coupled receptor (vasopressin type 2 receptor; V_2R) and a cytosolic protein (Na^+/H^+ exchanger regulatory factor-1; NHERF1) as examples, we demonstrate that this strategy substantially improves protein sequence coverage, especially in key structural regions likely including PTMs themselves that play important roles in protein conformational dynamics and function.

SUPPLEMENTARY METHODS

1. Hydrogen-deuterium exchange-mass spectrometry (HDXMS)

Digestion tuning. Digestion tuning experiments were first performed to optimize experimental conditions that can achieve the best peptide digestion pattern for NHERF1 or V_2R proteins. The protein concentrations were adjusted to ~ 2 mg/mL with Exchange Solution A (150 mM NaCl, 8.3 mM Tris-HCl, pH 7.2 in H_2O) using a 1:3 (V:V) protein: buffer ratio. For example, 36 µL of protein solution was diluted with 108 µL Exchange Solution A and then divided into equal sized aliquots (e.g. 24 µL). Each aliquot was then mixed individually at a 2:3 ratio (volume ratio of digestion tuning sample-to-quench solution, e.g. 24 µL : 36 µL) with a series of quench solutions containing 0.8% formic acid (FA), 16.6% glycerol, and various concentrations of Guanidine hydrochloride GuHCl (Sigma-Aldrich, cat. no. G4505) (ranging from 0 M to 4 M). For V_2R, Tris (2-carboxyethyl) phosphine hydrochloride (TCEP) (ThermoFisher Scientific, cat. no. 20491) (ranging from 15 mM-500 mM) was added to the quench buffer. The protein–buffer–quench solution mixtures were incubated at 0 ºC for 1 min, frozen on dry ice, and stored at -80 ºC until subjected to on-line pepsin digestion and liquid chromatography-tandem mass spectrometry (LC/MS/MS) analysis. Protein digestion maps under different conditions were generated and compared. The condition that produced the best peptide digestion pattern (the best coverage along the amino acid sequence, the most number of peptides, and the highest number of high-quality peptides) was used for all the HDXMS experiments described in this study. The optimal quench solution for NHERF was determined to be 0.8 M GuHCl, 0.8% FA, 16.6% Glycerol. The optimal quench solution for V_2R was determined to be 0.8 M GuHCl, 0.8% FA, 16.6% Glycerol, 20 mM TCEP.

Hydrogen-deuterium exchange experiments. Purified NHERF1 or V_2R protein is used for hydrogen-deuterium exchange experiments. For each protein, three sets of samples were prepared: (1) non-deuterated (Non-D), (2) fully deuterated (FD), and (3) time-dependent on-exchange samples. A protein:buffer:quench solution ratio of 1:3:6 (volume) was used for all sample preparations. The FD samples were prepared to measure the back-exchange rates during on-line pepsin digestion and LC/MS analysis. To prepare the FD sample sets, the protein samples were mixed with D_2O buffer (0.8% FA in 100% D_2O) at 1:3 (volume) ratio and incubated at room temperature for 12 h before quenching. After incubation, 6 volumes of optimal quench solution were added to quench the hydrogen-deuterium exchange reaction. The Non-D samples were used to normalize retention times for identified peptides in HDX analysis. To prepare Non-D sample sets, the protein samples were mixed with H_2O buffer (150 mM NaCl, 8.3 mM Tris-HCl, pH 7.2 in H_2O) at 1:3 ratio (volume) without the incubation step. The on-exchange sample sets were prepared by adding 3 volumes of

D_2O buffer (150 mM NaCl, 8.3 mM Tris-HCl, pH7.2 in D_2O) at 0 °C and incubating for varying time points (10, 100, 1000, 10,000, and 100,000 sec). Six volumes of ice-cold quench solution were then added to each on-exchange sample, followed by snap-freezing on dry ice and transferring to -80 °C freezer.

On-line pepsin digestion, LC/MS/MS analysis, and data processing. Samples prepared above were thawed right before the on-line pepsin digestion at 0 °C using an AS3000 cryogenic autosampler and immediately passed through a Poroszyme™ Immobilized Pepsin Cartridge (1 mm x 20 mm) by utilizing a Shimadzu LC-10AD HPLC pump (0.05% TFA at 20 µL/min). The proteolytic products are collected subsequently on a C_{18} trap column (Magic C_{18} AQ, 0.2x 2 mm), and desalted for 2 min. The pepsin digested peptides were then separated on a Magic C_{18} AQ reverse phase analytical column (Michrom Bioresource, 0.2 x 50 mm, 3 µm, 200Å) using a linear acetonitrile gradient of 8%-48 % B over 30 min (2 µL/min; Solvent A: 0.05% TFA; Solvent B: 80% acetonitrile, 20% water, 0.01% TFA). The column effluent is directed into an Orbitrap Velos or Elite mass spectrometer (Thermo Fisher Scientific) for MS analysis. Both MS1 and MS2 spectra were collected using the data-dependent acquisition mode. Each scan cycle is initiated with a full MS1 scan of high mass accuracy over the range of 150-2,000 *m/z*, followed by MS/MS scans on the top five most abundant precursor ions from preceding MS1 scan. Peptide identification was performed using LC/MS/MS data sets collected from Non-D samples and the SEQUEST database search engine using Proteome Discoverer (Thermo Fisher Scientific). The SEQUEST database search results are submitted to HDExaminer (Sierra Analytics) and filtered by score filters (e.g. Xcorr) to create an initial peptide pool. The quality of the MS1 data for each filtered peptide is then assigned an initial quality score by HDExaminer. The isotopic envelope of each peptide is manually investigated and adjusted based on the theoretical isotopic envelope to improve the quality score. Only peptides with high quality in the MS1 spectra are kept in the final peptide pool to ensure data quality. The retention times and *m/z* ranges of each peptide from the final peptide pool were manually verified and adjusted across all LC/MS/MS data sets from on-exchange samples and FD samples to ensure that HDExaminer had selected the correct peptide for all experiments. The centroids of isotopic envelopes of non-deuterated, partially deuterated, and fully deuterated peptides are automatically calculated by HDExaminer. The deuteration level is also displayed by the software, with corrections for back-exchange using the methods of Zhang and Smith[1, 2]. Results from FD samples were used to calculate the back-exchange rates during on-line pepsin digestion and LC/MS/MS analysis. A deuterium accumulation plot was created for each peptide as a further quality check and data refinement process. Rainbow maps were generated by HDExaminer.

2. PTM scanning

First step --- Search for possible PTMs using available PTM databases or regular proteomic experiments. Literature search was conducted to look for possible PTMs for NHERF1 and V_2R. Some well-curated PTM websites provide excellent resources for this purpose. These websites include PhosphoSitePlus[3] (https://www.phosphosite.org/), dbPTM[4] (dbptm.mbc.nctu.edu.tw/), UniProtKB[5-7] (http://www.uniprot.org/), HPRD[8, 9] (http://www.hprd.org/index_html), PhosphoELM[10] (http://phospho.elm.eu.org/index.html) and others. In addition, a number of bioinformatic tools/websites can be used to predict PTMs for proteins of interest. These PTM prediction tools/websites include Post-Translational Modification Site Prediction–User Interface page[11](http://csbl.bmb.uga.edu/~ffzhou/PTMP-UI/), FindMod (https://web.expasy.org/findmod/), GPS (http://gps.biocuckoo.org/), NetPhos (http://www.cbs.dtu.dk/services/NetPhos/), NetPhosK (http://www.cbs.dtu.dk/services/NetPhosK/), NetAcet[12] (http://www.cbs.dtu.dk/services/NetAcet/), and Sulfinator (http://www.expasy.org/tools/sulfinator/).

In this study, acetylation, phosphorylation, methylation, S-nitrosylation, and ubiquitylation were reported for NHERF1 (**Table S4**) and palmitoylation, phosphorylation for V_2R (**Table S1**). Glycosylation was also reported for V_2R[13], although it is not included in the PhosphoSitePlus® and dbPTM.

In some cases, if no PTM information is available for the proteins of interest in literature, regular proteomic experiments can be performed to search for possible PTMs. An aliquot of protein sample prepared for HDXMS is removed and digested by trypsin or other proteases. The digested peptides are submitted to LC/MS/MS and data analysis using regular procedures for PTM studies[14-19]. PTM enrichment is generally not required because the sensitivities of many modern mass spectrometers are capable of characterizing these PTMs from purified protein preparation. In addition, PTM enrichment may make it possible to identify post-translationally modified peptides that are at low stoichiometry and are not detected in HDXMS experiments.

Second step --- Identify PTM-containing peptic peptides from the LC/MS/MS data collected from HDXMS experiments. After the possible PTM types and sites have been identified in the first step, one can proceed to the second step of the PTM scanning. In this step, the SEQUEST or other database search engines are used to search for possible PTM-containing peptic peptides in LC/MS/MS data sets collected from Non-D samples in HDXMS experiments. In this study, MS/MS spectra were searched against NHERF1 or V_2R protein sequence using non-enzymatic search. Search parameters allowed for a mass tolerance of ±10 ppm for precursor ion, and a mass tolerance of ± 0.8 Da for product ion, and up to six total dynamic modifications (**Table S5**). For NHERF1, search parameters for dynamic modifications were set as 42.010565 Da (acetylation) on Lys and N-term, 79.966331 Da (phosphorylation) on Ser, Thr, and Tyr, 14.01565 Da

(methylation) on Arg, Lys, and C-term, and 15.994915 Da (oxidation) on methionine. For V_2R, search parameters for dynamic modifications were set as 79.966331 Da (phosphorylation) on Ser, Thr, and Tyr, 238.229666 Da (palmitoylation) on Cys, and 15.994915 Da (oxidation) on methionine. Ubiquitination-containing peptides were not considered since they are typically difficult to be identified in HDXMS experiments (**Table S5**).

Search results were filtered with XCorr thresholds and the PTM-containing peptides identified were manually confirmed by checking both MS1 and MS2 spectra. The confirmed PTM-containing peptic peptides were used in HDXMS data analysis.

3. Glycopeptide prediction by GlycoMod

The LC-MS/MS raw data sets collected from non-deuterated samples in HDXMS experiments of V_2R were manually inspected to check for low mass CID (Collision-induced dissociation)-induced carbohydrate marker ions in MS/MS spectra. Several MS/MS spectra have peaks with masses corresponding to those of certain glycan fragments, for example, $[(Hex)_1(HexNAc)_1]^{1+}$ (m/z=366.14), $[(Hex)_2(HexNAc)_1]^{1+}$ (m/z=528.21), $[(Hex)_3(HexNAc)_1]^{1+}$ (m/z=690.06), and $[(Hex)_4(HexNAc)_1]^{1+}$ (m/z=852.30). These MS/MS spectra were considered to be fragmented spectra from glycans or possible glycopeptides. The monoisotopic masses of these possible glycopeptides were uploaded to GlycoMod (*https://web.expasy.org/glycomod/*).[20] A glycopeptide "^{18}PSLPS*NSSQE^{27}" was predicted where *N is an N-linked glycosylation site. The region between residues 18-27 was recovered after this glycopeptide was included (**Figure S5**).

REFERENCES

1. Zhang, Z. & Smith, D.L. Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Sci* **2**, 522-531 (1993).
2. Thevenon-Emeric, G., Kozlowski, J., Zhang, Z. & Smith, D.L. Determination of amide hydrogen exchange rates in peptides by mass spectrometry. *Anal Chem* **64**, 2456-2458 (1992).
3. Hornbeck, P.V. et al. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic acids research* **43**, D512-520 (2015).
4. Huang, K.Y. et al. dbPTM 2016: 10-year anniversary of a resource for post-translational modification of proteins. *Nucleic acids research* **44**, D435-446 (2016).
5. UniProt Consortium, T. UniProt: the universal protein knowledgebase. *Nucleic acids research* **46**, 2699 (2018).
6. The UniProt, C. UniProt: the universal protein knowledgebase. *Nucleic acids research* **45**, D158-D169 (2017).
7. Apweiler, R. et al. UniProt: the Universal Protein knowledgebase. *Nucleic acids research* **32**, D115-119 (2004).
8. Keshava Prasad, T.S. et al. Human Protein Reference Database--2009 update. *Nucleic acids research* **37**, D767-772 (2009).

9. Peri, S. et al. Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome research* **13**, 2363-2371 (2003).
10. Dinkel, H. et al. Phospho.ELM: a database of phosphorylation sites--update 2011. *Nucleic acids research* **39**, D261-267 (2011).
11. Zhou, F., Xue, Y., Yao, X. & Xu, Y. A general user interface for prediction servers of proteins' post-translational modification sites. *Nature protocols* **1**, 1318-1321 (2006).
12. Kiemer, L., Bendtsen, J.D. & Blom, N. NetAcet: prediction of N-terminal acetylation sites. *Bioinformatics* **21**, 1269-1270 (2005).
13. Innamorati, G., Sadeghi, H. & Birnbaumer, M. A fully active nonglycosylated V2 vasopressin receptor. *Molecular pharmacology* **50**, 467-473 (1996).
14. Larsen, M.R., Trelle, M.B., Thingholm, T.E. & Jensen, O.N. Analysis of posttranslational modifications of proteins by tandem mass spectrometry. *BioTechniques* **40**, 790-798 (2006).
15. Chicooree, N., Unwin, R.D. & Griffiths, J.R. The application of targeted mass spectrometry-based strategies to the detection and localization of post-translational modifications. *Mass spectrometry reviews* **34**, 595-626 (2015).
16. Silva, A.M., Vitorino, R., Domingues, M.R., Spickett, C.M. & Domingues, P. Post-translational modifications and mass spectrometry detection. *Free radical biology & medicine* **65**, 925-941 (2013).
17. Doll, S. & Burlingame, A.L. Mass spectrometry-based detection and assignment of protein posttranslational modifications. *ACS chemical biology* **10**, 63-71 (2015).
18. Zhao, Y. & Jensen, O.N. Modification-specific proteomics: strategies for characterization of post-translational modifications using enrichment techniques. *Proteomics* **9**, 4632-4641 (2009).
19. Jensen, O.N. Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Current opinion in chemical biology* **8**, 33-41 (2004).
20. Cooper, C.A., Gasteiger, E. & Packer, N.H. GlycoMod--a software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics* **1**, 340-349 (2001).