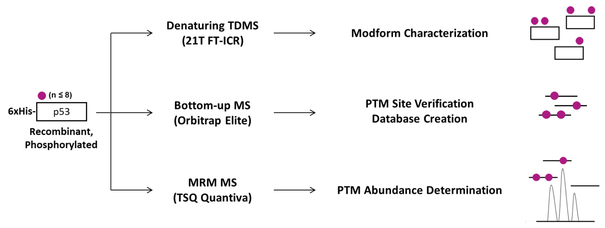
A multi-modal proteomics strategy for characterizing *in vitro* p53 modforms

Caroline J. DeHart[1,2], Luca Fornelli[1], Lissa C. Anderson[2], Jacek W. Sikora[1], Deepesh Agarwal[3], Dan Lu[3], Philip D. Compton[1], Christopher L. Hendrickson[2], Paul M. Thomas[1], Galit Lahav[3], Jeremy Gunawardena[3], and Neil L. Kelleher[1]

[1]National Resource for Translational and Developmental Proteomics, Northwestern University, Evanston, IL; [2]National High Magnetic Field Laboratory, Florida State University, Tallahassee, FL; [3]Harvard Medical School, Boston, MA

OVERVIEW

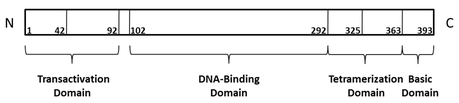
Objective:



Our primary objective was to apply a combined strategy of denaturing top-down MS, bottom-up MS, and multiple reaction monitoring (MRM) MS to identify and characterize the most abundant modforms (proteoforms bearing enzymatically reversible PTM that convey information as part of a signaling pathway) within a representative model population of intact recombinant human p53 (rp53) phosphorylated *in vitro* by Chk1 kinase. In doing so, we identified the modes of data required for the creation of a linear equation based algorithm to constrain the number of possible modforms and determine the actual modform distribution within this representative population.

INTRODUCTION

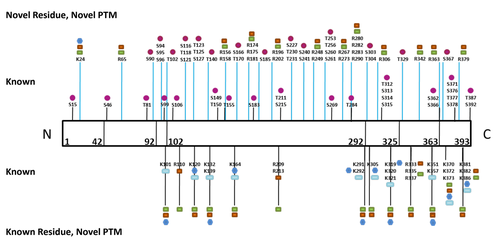
The p53 protein: a master cellular regulator



67 known p53-dependent cellular processes
458 clinically significant hp53 variations
24 known functions of the p53 protein
> 300 PTMs identified on p53
155,833 PubMed entries

Theoretical Mass: **43,653.08 Da**

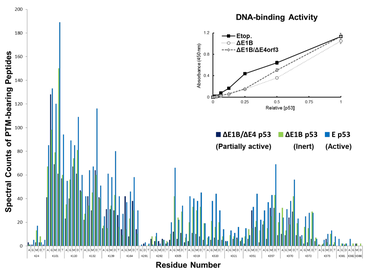
Endogenous human p53 is extensively modified



Summary of PTMs identified by tandem MS on endogenous human p53 isolated from primary fibroblasts (HFFs).

[Modified from DeHart C. J., Chahal, J. S., Flint, S. J., and Perlman, D. P., *Mol. Cell. Proteomics*, 2014 Jan; 13(1), 1-17.]

Endogenous p53 PTMs most likely comprise a combinatorial code

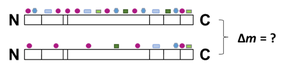


Summary and comparison of Lys modifications identified by tandem MS on three distinct populations of p53 isolated from HFFs: Active (**E p53**), Inert (**ΔE1B p53**), and Partially Active (**ΔE1B/ΔE4 p53**). Few differences in PTM location or identity could be determined, supporting the existence of a combinatorial PTM code governing p53 stability, localization, and transcriptional activity within the cell. **Inset:** The ability of the three p53 populations to bind to a synthetic consensus p53 sequence (one indicator for p53 transcriptional activity) was quantified via an ELISA-based assay.

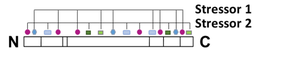
[Modified from DeHart C. J., Perlman, D. P., and Flint, S. J., *J. Virol.*, 2015 Mar;89(6):3209-20.]

Questions remaining after analysis of p53 by bottom-up mass spectrometry

How many combinatorial patterns of p53 PTM are realistically possible? (How many endogenous p53 modforms exist?)

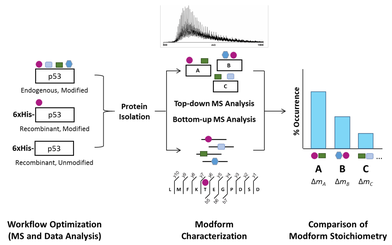


How extensively modified is any single p53 modform?



How can the logic underlying p53 regulation via combinatorial PTM be deduced in the absence of modform identification?

Proposed analysis of endogenous p53 modforms by top-down mass spectrometry

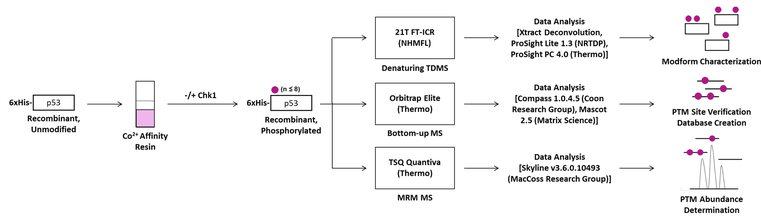


We hypothesized that the examination of endogenous and modified human p53 by top-down MS (TDMS) would enable both the generation of a comprehensive modform profile, comprising all modforms of p53 present within a specific cellular context, and the preliminary elucidation of the combinatorial regulatory logic underlying observed patterns of p53 PTM. To that end, we have developed a method by which training populations of recombinant and endogenous intact p53 protein can be isolated and analyzed for the presence of PTM by a combined strategy of high-resolution tandem MS and TDMS. Complete workflow optimization will enable the analysis and characterization of endogenous human p53 modforms from a broad range of cellular contexts, including multiple stress response pathways (e.g. DNA damage) and disease states (e.g. human cancers).

METHODS

Analysis of representative p53 modforms by denaturing top-down mass spectrometry

Experimental strategy:

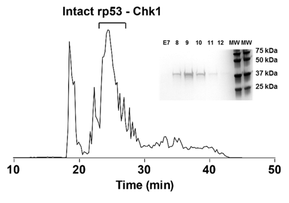


Recombinant human p53 (rp53) was purified by cobalt affinity chromatography and phosphorylated *in vitro* with Chk1 kinase. Modified rp53 populations and kinase-free controls were analyzed in parallel by denaturing top-down MS and *de facto* middle-down MS on a 21T FT-ICR mass spectrometer (NHMFL), as well as by tryptic digest and subsequent bottom-up MS on an Orbitrap Elite mass spectrometer (Thermo). Multiple reaction monitoring (MRM) analysis of the resulting unmodified and Chk1-treated rp53 tryptic peptides was performed on a TSQ Quantiva mass spectrometer (Thermo). Phosphorylation sites identified by bottom-up MS were used to create a custom database against which the top-down and *de facto* middle-down MS data were then searched using ProSight PC (Thermo). The resulting identified modforms (comprising the C-terminal region of phosphorylated rp53) were manually validated and assembled into a coverage map using Adobe Illustrator.

RESULTS

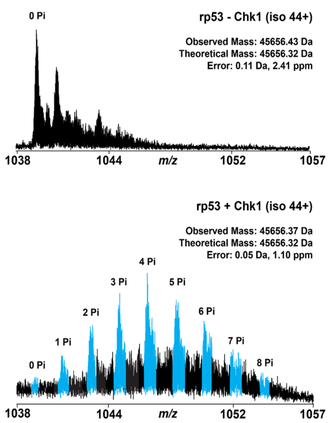
Analysis of unmodified rp53 by denaturing top-down mass spectrometry

Example Chromatogram (21T FT-ICR):



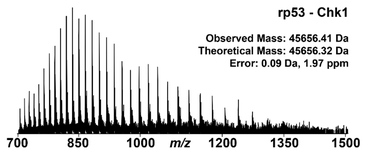
Top: Example FTMS1 total ion chromatogram (TIC) of intact and unmodified rp53. **Top Inset:** $AgNO_3$- stained gel showing eluted fractions of Co^{2+}-purified rp53. **Bottom:** Example FTMS1 TIC of intact and phosphorylated rp53. **Bottom Inset:** $AgNO_3$- and ProQ-stained gel showing the presence of phosphoryl groups only on Chk1-treated rp53.

Isolated MS1 Spectra (21T FT-ICR):



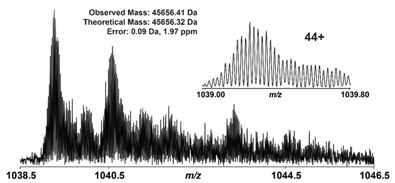
Top: Isolated MS1 spectrum of the 44+ charge state of intact and unmodified rp53. **Bottom:** Isolated MS1 spectrum of the 44+ charge state of intact and phosphorylated rp53, showing the presence of up to 8 phosphorylated sites (consistent with the number previously reported in the literature). Both spectra were acquired at 300,000 FTRP on an LC timescale.

Broadband MS1 Spectrum (21T FT-ICR):



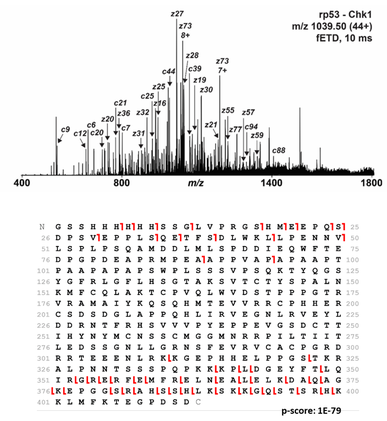
Example broadband MS1 spectrum of intact and unmodified rp53 acquired at 150,000 FTRP on an LC timescale.

Isolated MS1 Spectrum (21T FT-ICR):



Example isolated MS1 spectrum of the 44+ charge state of intact and unmodified rp53 acquired at 300,000 FTRP on an LC timescale. **Inset:** Zoomed-in view of most abundant peak, showing the near-baseline isotopic resolution obtained by averaging two scans.

Example MS2 Spectrum (21T FT-ICR):

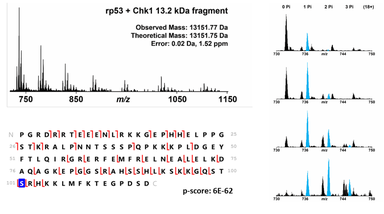


Top: Example MS2 spectrum of intact and unmodified rp53, acquired by targeting the 44+ charge state for fragmentation by front-end ETD (fETD), with selected *c* and *z* ions indicated within the spectrum. **Bottom:** Fragment ion coverage map of the rp53 sequence, showing that confident identification can be achieved on an LC timescale.

RESULTS II

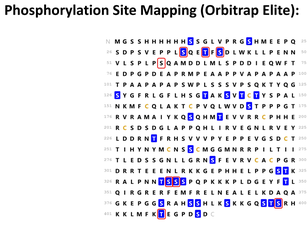
Characterization of C-terminal p53 modforms by combined top-down and bottom-up mass spectrometry

13.2 kDa C-terminal Fragment (21T FT-ICR):



Top Left: Example MS1 spectrum of the 13.2 kDa rp53 C-terminal fragment, showing the presence of up to 4 phosphorylated sites. **Right:** Zoomed-in view of the 18+ charge state of the 13.2 kDa fragment over 4 sequential MS1 scans. **Bottom Left:** Example sequence coverage map with localized phosphorylation site obtained by targeting the 18+ charge state of the singly phosphorylated 13.2 kDa fragment species for fETD MS2 fragmentation.

Phosphorylation Site Mapping (Orbitrap Elite):



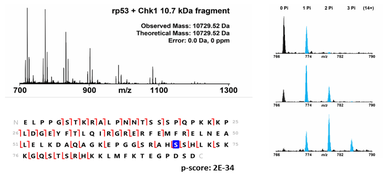
Phosphorylation sites identified on Chk1-treated rp53 by bottom-up MS. (**Red:** described in previous studies.)

Craig AL, Chrystal JA, Fraser JA, Sphyris N, Lin Y, Harrison BJ, Scott MT, Domreiter I, Hupp TR., *Mol Cell Biol.* 2007 May;27(9):3542-55.

Ou YH, Chung PH, Sun TP, Shieh SY., *Mol Biol Cell.* 2005 Apr;16(4):1684-95.

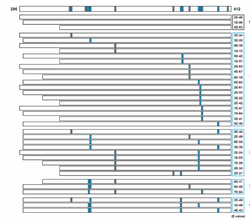
Shieh SY, Ahn J, Tamai K, Taya Y, Prives C., *Genes Dev.* 2000 Feb 1;14(3):289-300.

10.7 kDa C-terminal Fragment (21T FT-ICR):



Top Left: Example MS1 spectrum of the 10.7 kDa rp53 C-terminal fragment, showing the presence of up to 3 phosphorylated sites. **Right:** Zoomed-in view of the 14+ charge state of the 10.7 kDa fragment over 3 sequential MS1 scans. **Bottom Left:** Example sequence coverage map with localized phosphorylation site obtained by targeting the 14+ charge state of the singly phosphorylated 10.7 kDa fragment species for fETD MS2 fragmentation.

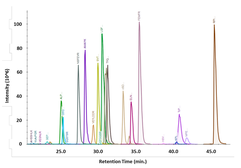
C-terminal Modforms Identified:



Coverage map showing the 22 distinct combinations of phosphorylation sites identified on the C-terminus of Chk1-treated rp53 by combined top-down and bottom-up MS. 5 of the 16 sites predicted by bottom-up MS were not observed in the top-down MS data.

Analysis of model p53 proteoforms by Multiple Reaction Monitoring (MRM)

Example Chromatogram (TSQ Quantiva):

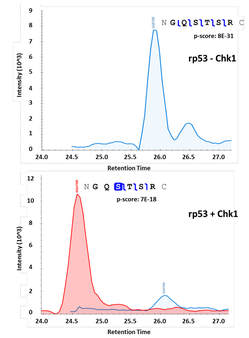


Example TIC showing the retention times of rp53 peptides included in the scheduled MRM method for analysis of PTMs on both rp53 and endogenous p53.

Example Coverage Map (TSQ Quantiva):

Coverage map showing the regions of rp53 sequence represented by the peptide peaks shown in the above chromatogram.

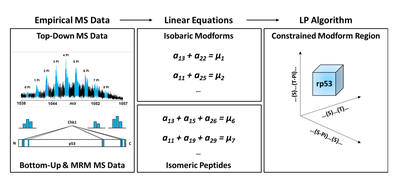
Example Phosphopeptide (TSQ Quantiva):



Top: Example peptide peak from MRM analysis of unmodified rp53, showing the fragment ion coverage resulting from CID MS2. **Bottom:** Example peptide peaks from MRM analysis of Chk1-treated rp53, showing both the different relative abundance of both peptide species and the phosphorylation site localized by CID MS2.

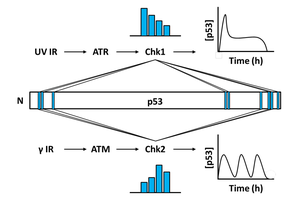
Future Directions

Model to Determine Actual Modform Distributions:



We will develop a network of linear equations to accurately describe both p53 modforms and their relative abundances within a population indicated by empirical top-down, bottom-up, and MRM MS data. We will then apply these equations to an LP algorithm to constrain the position of the p53 modform distribution to lie within a polyhedral region of the high-dimensional space of all possible modforms, thus elucidating the actual p53 modform distribution within the population.

Comparison of Endogenous Modform Distributions:



We will compare the modform distributions between CHK1 and CHK2 phosphorylated p53, in order to address whether it is the differences in specific PTMs or the distribution of PTMs on p53 that cause distinct dynamic and phenotypic outcomes.

Acknowledgements: The studies shown here were supported in part by **NIH P41 GM108569** for the National Resource for Translational and Developmental Proteomics (NRTDP) based at Northwestern University, and by **NIH R01 GM105375**, "Information Processing by Post-translational Modification" (to J.G., G.L., and N.L.K.). A portion of this work was performed at the National High Magnetic Field Laboratory, which is supported by **NSF DMR-1157490** and the State of Florida.

NATIONAL RESOURCE FOR TRANSLATIONAL AND DEVELOPMENTAL PROTEOMICS