Supplemental Experimental Procedures

Comprehensive Mass Spectrometric Mapping of the Hydroxylated Amino Acid residues of the α1(V) Collagen Chain*

Chenxi Yang[‡§], Arick C. Park[¶], Nicholas A. Davis[¶], Jason D. Russell[‡§], Byoungjae Kim[¶], David D. Brand[‖], Matthew J. Lawrence[¶], Ying Ge[¶**], Michael S. Westphall[‡‡], Joshua J. Coon[‡§‡‡], and Daniel S. Greenspan[¶1]**

[‡]Departments of Chemistry, [§]Biomolecular Chemistry, and [¶]Cell and Regenerative Biology, and the [**]Human Proteomics Program, and [‡‡]Genome Center, University of Wisconsin, Madison, Wisconsin; [¶]Research Service, Veterans Affairs Medical Center, University of Tennessee Health Science Center, Memphis, Tennessee.

Running Title: Mapping of Hydroxylated Amino Acids of α1(V) collagen

Digestion of Pro-α1(V) and α1(V) Samples for Mass Spectrometry. For each protease, 50 μg protein aliquots of Pro-α1(V) or α1(V) were digested. To produce peptides cleaved N-terminal to Asp, protein was incubated overnight at 37 °C in 50 mM Tris, 2.5 mM Zinc sulfate, pH 8.0, with 5 μg of the protease AspN (New England Biolabs, Ipswich, MA). Peptides cleaved C-terminal to Glu were prepared by incubating protein with 5 μg of GluC (Roche Diagnostics, Indianapolis, IN) overnight in 50 mM Tris and 25 mM ammonium bicarbonate (pH ~ 7.8) at room temperature. Peptides predominately cleaved C-terminal to Trp, Leu, Tyr, Phe (with slower kinetics for cleavage at Met, Asp, Glu, Ala) was achieved with 5 μg of chymotrypsin (Promega, Madison, WI) incubated overnight in 50 mM Tris, 1 mM $CaCl_2$ at room temperature. Protein digested overnight in 90 mM Tris, 8.5 mM $CaCl_2$, 5 mM DTT, 0.5 mM EDTA, pH 7.6 at 37 °C, with 5 μg of ArgC (Roche Diagnostics) produced peptides cleaved C-terminal to Arg. Tryptic peptides (cleavage C-terminal to Lys and Arg) were produced by digesting protein with 5 μg of trypsin (Promega) overnight in 50 mM Tris (pH 8.0), and 1 mM $CaCl_2$ at 37 °C. Sequential digestion using GluC and AspN was carried out by first digesting protein with GluC for 6 hrs at room temperature, followed by the addition of AspN and incubation overnight at 37 °C. Each digest was quenched by freezing at -80 ^{0}C, and desalted on a 50 mg, tC18 SepPak cartridge

(Waters Corp.). Each eluate was then dried under vacuum and resuspended in 50 µl 0.2% formic acid resulting for MS analysis (~ 1 µg/µl peptide concentration).

Manual validation of peptide modification assignments. Peptides identified through database correlation with P-value scores larger than $1x10^{-9}$ were discarded and were not considered for manual validation. Although OMSSA usually identified the correct peptide sequence, the number, type, and specific location of modifications were occasionally incorrect. Manual verification of the peptide sequence and site(s) of modification was carried out for every modified peptide species discussed in this manuscript (supplemental information S4 and S5). Spectral validation was carried out as follows: the mass of the unmodified peptide sequence was compared to the modified peptide sequence identified with OMSSA. This mass difference represents the mass of the PTM(s) on the peptide (i.e. a mass difference of +15.995 Da relative to the unmodified peptide suggests hydroxylation or oxidation). The high mass accuracy of the Orbitrap limited the potential number of modifications that needed to be considered. Each peptide was iteratively evaluated considering all combinations of PTMs on residues that can accommodate the modifications. Theoretical product ion masses were produced using the MS-Product program available in the web version of Protein Prospector (http://prospector.ucsf.edu). Product ions were confidently mapped to peptide sequences if the production ions had a signal-to-noise ratio > 3 and mass error < 10 ppm from the expected product ion monoisotopic mass. Modifications indicated as localized had product ion matches that unambiguously mapped the PTM to a specific residue in the peptide sequence. In situations where product ions did not support PTM assignment to a specific residue, the modification was considered unlocalized and is clearly indicated.

Pseudolocalization of modifications. Some modifications that were classified as unlocalized were sub-classified as pseudolocalized. In these cases, there were insufficient product ions to localize the modification to a specific residue in the presence of other residues that could bear the same modification. However, the location of the modification on the peptide

may be inferred from well-characterized, high fidelity modification motifs for collagen-derived proteins. For example, the bovine α1(V) peptide D^{708}GPQGPPGGIGNPGAVGE^{725} was found to have two hydroxylation modifications based upon accurate mass measurements and partial sequence analysis. One modification was localized to the proline residue at sequence position 720. Product ions support the assignment of the second hydroxylated residue on residues 713 through 718 (PPGGI), but this hydroxylation could not be definitively localized to either of the modifiable proline residues (i.e., P^{713} or P^{714} could be modified but we lack product ion data to support one modification over the other.). Based upon the motif Gly-X-Y, if the Y position is a proline residue it is the most likely position of modification as Gly-X-Hyp is a common modification motif in collagen. If both the X and Y positions are proline residues, it is still more likely that the Y position is hydroxylated based upon current knowledge of collagen hydroxylation patterns. We developed several additional guidelines used to rank the likelihood of a particular modification.

The ranking of likely modifications for pseudolocalization based upon the Gly-X-Y modification motif (where positions X and Y can by hydroxylated or glycosylated) was as follows:

(1) Gly-X-Hyp > Gly-Hyp-Y

- Y-position proline hydroxylation is more likely than X-position proline hydroxylation

(2) Gly-Hyp-Hyp > Gly-Hyp-Y + Gly-X-Hyp (or Gly-Hyp-Y)

- X- and Y-position proline hydroxylations are more likely than an X-position hydroxylation plus an additional X- or Y-position hydroxylation on anther region of the peptide. Such a situation can arise when multiple hydroxylation modifications are unlocalized with more than one G-X-Y motif.

(3) Gly-X-Hyl > Gly-Hyp-Y

- Y-position hydroxylation of lysine is more likely than X-position proline hydroxylation

(4) Gly-X-Hyl.Gal.Glc > Gly-Hyl.Gal.Glc-Y

- Y-position hydroxylysine glycosylation is more likely than X-position hydroxylysine glycosylation

Because pseudolocalization lacks spectral support, we do not have high confidence in assigning one modifiable position over another. However, we find the pseudolocalization classification useful for comparison when a position has been previously reported as modified in the literature.

Supplemental Information Table and Spectra Captions

Supplemental Table S1. Modifications of bovine α1(V) collagen. A table of modified peptides identified by tandem mass spectrometry indicating localized modifications, unlocalized modifications, and pseudolocalized modifications. A sequence/modification map of α1(V) collagen indicating sequence coverage and localized modifications.

Supplemental Table S2. Modifications of recombinant human α1(V) collagen. A table of modified peptides identified by tandem mass spectrometry indicating localized modifications, unlocalized modifications, and pseudolocalized modifications. A sequence/modification map of α1(V) collagen indicating sequence coverage and localized modifications.

Supplemental Spectra S1. Tandem mass spectra of peptides derived from bovine α1(V) collagen. Each spectrum is accompanied by a table of sequence-specific product ions identified from the spectrum along with product ion mass errors. Localized modifications are indicated above the table and are identified as bold letters in the peptide sequence. Unlocalized and pseudolocalized sites of modification are also given. Note: each product ion in the table met the requirements of having < 10 ppm mass error and a signal-to-noise ratio > 3.

Supplemental Spectra S2. Tandem mass spectra of peptides derived from recombinant α1(V) collagen. Each spectrum is accompanied by a table of sequence-specific product ions identified from the spectrum along with product ion mass errors. Localized modifications are indicated above the table and are identified as bold letters in the peptide sequence. Unlocalized and pseudolocalized sites of modification are also given. Note: each product ion in the table met the requirements of having < 10 ppm mass error and a signal-to-noise ratio > 3.