Published in final edited form as:

Proteomics. 2013 September ; 13(17): 2682–2691. doi:10.1002/pmic.201300064.

Identification and Characterization of Citrulline-modified Brain Proteins by Combining HCD and CID Fragmentation

Zhicheng Jin^{1,*}, Zongming Fu², Jun Yang³, Juan Troncosco⁴, Allen D. Everett³, and Jennifer E. Van Eyk^{1,5}

¹Department of Medicine, Cardiology, Johns Hopkins University, Baltimore, MD

²Department of Pediatrics, Hematology, Johns Hopkins University, Baltimore, MD

³Department of Pediatrics, Cardiology, Johns Hopkins University, Baltimore, MD

⁴Department of Pathology, Neuropathology, Johns Hopkins University, Baltimore, MD

⁵Department of Biological Chemistry, Johns Hopkins University, Baltimore, MD

Abstract

Citrullination is a protein post-translational modification of arginine residues catalyzed by peptidylarginine deiminase. Protein citrullination has been detected in the central nervous system and associated with a number of neurological diseases. However, identifying citrullinated proteins from complex mixtures and pinpointing citrullinated residues has been limited. Using reversed-phase liquid chromatography and high resolution mass spectrometry, this study determined *in vitro* citrullination sites of glial fibrillary acid protein (GFAP) and myelin basic protein (MBP) and *in vivo* sites in brain protein extract. Human GFAP has five endogenous citrullination sites, R30, R36, R270, R406, and R416, and MBP has fourteen *in vivo* citrullination sites. Human neurogranin (NRGN/RC3) was found citrullinated at residue R68. The sequence of citrullinated peptides and citrullination sites were confirmed from peptides identified in trypsin, Lys-C, and Glu-C digests. The relative ratio of citrullination was estimated by simultaneous identification of citrullinated and unmodified peptides from Alzheimer's and control samples. The site occupancy of citrullination at the residue R68 of NRGN ranged from 1.6% to 9.5%. Compared to collision-induced dissociation (CID), higher-energy collisional dissociation (HCD) mainly produced protein backbone fragmentation for citrullinated peptides. CID triggered HCD fragmentation is an optimal approach for the identification of citrullinated peptides in complex protein digests.

Introduction

Protein citrullination is an irreversible protein post-translational modification of incorporated arginine residues converting the guanidinium group to an ureido group. This protein modification is catalyzed by peptidylarginine deiminase (PAD) in the presence of calcium (Figure 1) [1,2]. Five PAD enzymes are expressed in humans (PAD 1–4, and PAD 6). PAD

Corresponding Author: Jennifer E. Van Eyk, Johns Hopkins University Bayview Proteomics Center, Mason F. Lord Bldg, Center Tower, Room 602, Johns Hopkins University, 5200 Eastern Ave, Baltimore MD 21224, jvaneyk1@jhmi.edu.

*Current address for Zhicheng Jin: Genzyme Corporation, 1 Mountain Road, Framingham, MA 01701

The authors have declared no conflict of interest.

activity and protein citrullination have been associated with various human diseases, including rheumatoid arthritis [3], multiple sclerosis [4], Alzheimer's disease [5], and recently heart disease [6]. Accumulating evidence suggests that proteins, for example, filaggrin, vimentin, histones, and collagens, can be citrullinated under physiological or in disease states [7,8]. Citrullination has been shown to be disruptive for a few proteins, altering their structure and destabilizing the protein, affecting protein-protein interactions and other biological functions [2]. In the context of inflammatory disease, citrullinated proteins are more immunogenic and potentially play a role in the pathogenesis of autoimmune diseases. Commercial assays were developed to measure autoantibody of citrullinated proteins/peptides for the diagnosis of rheumatoid arthritis [9,10]. However, the number of modified proteins and importantly, the actual citrullinated amino acid residues remains limited.

It is this potential for increased antigenicity with the development of auto-antibodies that has led investigators to explore the role of citrullination of brain proteins as a potential autoimmune mechanism for degenerative brain diseases such as Parkinson's and Alzheimer's diseases. To date, endogenous citrullination has been identified in a few brain specific or enriched proteins including glial fibrillary acid protein (GFAP) [4,11] and myelin basic protein (MBP) [2]. It was reported previously that eleven arginine residues of human MBP may be citrullinated [12,13] while bovine MBP was reported to have at least two citrullination sites [14,15]. While for GFAP, the specific amino acid residues citrullinated are still unknown. Based on the chemical derivatization of citrulline with 2,3-butanedione and antipyrine and subsequent western blot to the modified protein citrulline, the level of citrullinated GFAP has been shown to be increased in the brain of patients with Alzheimer's disease [5]. Again, using chemical modification and western blot approach, compared to control there was hyper-citrullination of MBP and GFAP in spinal cord in mice with experimental autoimmune encephalomyelitis, a disease resembling multiple sclerosis in humans [11]. Citrullinated GFAP and MBP have been shown to be increased in the brain and spinal cord of patients with multiple sclerosis [2]. These proteins can be found (and in some cases, elevated) in tissue, cell and various body fluids (e.g. serum, cerebral spinal fluid) in adults and children. However, exact modified sites have not been determined, nor if the increases observed with disease are due to increase site occupancy at the same modified residues or if new disease-specific sites occur. Nor, is it known which of the five PAD isoforms is responsible for modification. PAD2 is the predominantly expressed isoform in the central nervous system and lacks significant amino acid sequence homology (28% amino acid homology) with other PAD enzymes 1, 3, 4, and 6. However, the other PAD isoforms are also present at a lesser and varying degree in the brain and nervous system (Eg. human protein atlas, www.proteinatlas.org).

GFAP is an intermediate filament protein that is expressed in the CNS system. It is a brain-specific cytoskeleton protein that can be detected in the blood stream following brain injury. As such, GFAP has been well-studied as a biomarker of brain injury [16]. It was reported previously that GFAP protein was citrullinated in the human autoimmune disease multiple sclerosis [4,11], where the citrullinated GFAP was detected using an antibody against chemically modified citrulline. However, the citrullinated residue(s) of GFAP was not identified in the previous report. Recently, our group has shown that the protein

concentration of GFAP is increased in the blood of children with birth related hypoxic-ischemic encephalopathy on extracorporeal cardiopulmonary support [17,18]. GFAP has been shown as a potential marker providing valuable information for the diagnosis of stroke. As well, circulating levels of MBP has also been associated with traumatic brain injury [19,20]. In addition, the brain specific protein neurogranin (NRGN) which plays a central role in learning is present in cerebral spinal fluid with Alzheimer's disease [21]. It is interesting to speculate whether changes in the citrullination status of brain specific proteins, like GFAP could occur with brain injury and damage. Considering the critical role of intracellular calcium and potential to activate PAD during ischemic reperfusion injury of the brain, increased citrullination of specific residues on selective proteins may provide additional information than, just the concentration of the protein itself.

Detection of citrullinated peptides has been difficult using standard chemical derivatization and immunostaining techniques. A recent technical innovation is the use of accurate m/z of peptide precursor ions [22], reversed-phase chromatographic separation, and a unique fragmentation pathway of citrullinated peptides [23] upon collision induced dissociation (CID). In addition, higher-energy collisional dissociation (HCD) fragmentation has been employed in the analysis of large-scale phosphoproteomics [24] and identification of proteins with O-GlcNAc modification [25]. HCD spectrum contains low-mass ions that can be missed in the ion trap CID spectrum. Because of this advantage, HCD fragmentation was tested in our study for its utility in the identification of citrullinated peptides from complex protein digests. Since neutral loss of isocyanic acid occurs for citrullinated peptides upon CID fragmentation, an optimal approach would be CID triggered HCD fragmentation [25]. The goal of this study is to determine the citrullination sites of GFAP, MBP, and NRGN *in vitro* and *in vivo*.

Experimental Procedures

Materials

Bovine myelin basic protein (MBP) and peptidylarginine deiminase type 2 (PAD2) were obtained from Sigma-Aldrich (St. Louis, MO). Because bovine MBP was readily available commercially and due to it high amino acid homology (78 %) to the human protein, we selected bovine MBP as a model protein to validate our methods for the identification of citrullinated peptides from complex mixtures.

Full length human NRGN cDNA was released from human cDNA ORF clone (Origene, RC201209) by digested with Sgf I and Mlu I, then cloned into bacterial expression vector pEX-N-His (Origene, PS100030) between Sgf I and Mlu I. The expression construct (pEX-N-His-NRGN) was verified by DNA sequencing. pEX-N-His-NRGN was then transformed into Rosetta 2 (DE3) E. coli strain (EMD, 71403), the resulting strain was grown in auto induction TB media (EMD, 70491) supplemented with ampicillin overnight at 37 °C to produce recombinant protein. The overexpressed His-NRGN protein was purified from above bacteria lysate using conventional Ni-NTA agarose method. The purified His-NRGN protein was visualized by SDS-PAGE followed by commassie staining, and validated by MS (data not shown). Additional amino acid residues were introduced in the recombinant

protein NRGN, including His tag at N-terminus, T79 and R80 at C-terminus (Supplement Figure 1).

Bovine NRGN was identified within the bovine MBP sample by mass spectrometry (MS). GFAP isolated from human brain tissue was purchased from Calbiochem (EMD Chemical, Inc., Gibbstown, NJ). Sequencing grade modified trypsin and endoproteinase Glu-C were purchased from Promega (Madison, WI). Endoproteinase Lys-C was obtained from Roche Diagnostics (Indianapolis, IN). RapiGest surfactant was purchased from Waters (Milford, MA). Other chemicals were from Sigma-Aldrich (St. Louis, MO).

Human samples

Human brain tissues from six individuals were obtained from the Johns Hopkins Brain Resource Center for this study with Institutional Board Review approval. All brain samples were from cases >65 years of age and obtained < 18 hours post mortem and immediately frozen. Three samples were no evidence of Alzheimers (CERAD 0 and BRAAK 0-2) and served as controls and three samples were from Alzheimers cases (CERAD C and BRAAK 6).

Peptidylarginine deiminase treatment and proteolysis

Analysis was carried out on bovine MBP, human GFAP, and human NRGN untreated or treated with PAD2 to convert peptidylarginine to peptidylcitrulline. Protein concentration was determined using a Pierce bicinchoninic acid (BCA) protein assay kit (Pierce, Rockford, IL). Each protein (2 µg) was treated with 0.15 µg of PAD2 (specific activity, 0.254 unit/µg; one unit will produce 1 µmole of N-α-benzoyl-citrulline ethyl ester from N-benzoyl-L-arginine ethyl ester per hr at 55°C and at pH 7.2) in a total volume of 40µL of a buffer composed of 20mM CaCl₂, 200mM Tris-HCl pH 7.5, 10mM dithiothreitol, at 55°C for 2 hours. Untreated samples were dissolved in the same buffer. Samples with and without PAD2 treatment were denatured with 0.1% RapiGest surfactant, reduced with 5 mM tris(2-carboxyethyl)phosphine (TCEP) at 50 °C for 30 minutes, and alkylated with 10 mM iodoacetamide at room temperature for 30 min (in the dark). Sequencing grade modified trypsin, Lys-C, or Glu-C was added to protein samples at a ratio of 1:20 (enzyme to substrate). The samples were incubated at 37 °C for 16 hours and digestion was complete. Samples were desalted by solid phase extraction with Oasis reversed-phase HLB cartridges (30mg/30µm, Waters, Milford, MA) according to the manufacture's instruction and were eluted in 80% acetonitrile with 0.1% formic acid. The protein digests were vacuum dried and reconstituted in 20% acetonitrile with 0.1% formic acid for further analysis.

Protein extraction from human brain tissue samples

The tissue sample (~30 mg) was homogenized in 0.2 mL 8 M urea, Amberlite IRN 150L, 2 M thiourea, 4% Chaps, 1% DTT and proteins were extracted with Sample Grinding Kit (GE Healthcare, Piscataway, NJ) according to the product instruction. The homogenate was centrifuged at 12,000 rpm for 10 min and the supernatant was saved in -80 °C freezer. The supernatant (100 µL) was cleaned with 2-D clean-up kit (GE Healthcare, Piscataway, NJ) according to the manufacture's protocol. After precipitation and centrifugation, the pellet was dissolved in 0.4 mL 6 M urea in 50 mM of Tris-HCl buffer (pH 7). Total protein

concentration of the crude extract was determined using Pierce BCA protein assay kit (Pierce, Rockford, IL). The protein extract was reduced, alkylated, and digested with trypsin, Lys-C, or Glu-C as described above. After desalting and vacuum drying, the protein digests were reconstituted in 20% acetonitrile with 0.1% formic acid for further analysis.

LC-MS/MS and Data Analysis

Digested protein samples (~50 ng for pure protein and ~400 ng for tissue) were analyzed either by nanoflow HPLC on a LTQ-Orbitrap or LTQ-Orbitrap Elite mass spectrometers (Thermo Scientific, San Jose, CA). For the LTQ-Orbitrap, an Agilent 1200 series nanoflow LC system (Agilent, Santa Clara, CA) was used for chromatographic separation with Solvent A (0.1% formic acid) and Solvent B (90% acetonitrile in 0.1% formic acid). Samples were loaded onto reversed-phase capillary columns, 75 μm inner diameter PicoFritSelf/P column (New Objective, Woburn, MA), in house packed with 10cm of Magic C18AQ packing material (5 μm diameter particles, 200 Å pore size) from Michrom Bioresources, Inc. (Auburn, CA). Separation started with 2% solvent B for 8 min at a flow rate of 2 $\mu\text{L}/\text{min}$. A 36 min linear gradient from 10% to 45% solvent B was followed by a 10min linear gradient from 45% to 95% solvent B at a flow rate of 300 nL/min. For the LTQ-Orbitrap Elite, chromatography was performed using an Easy-nLC (Thermo Scientific, San Jose, CA) and a BioBasic C18 PicoFrit column (300 Å, 5 μm , 75 μm $\times$ 10 cm, 15 μm tip, New Objective). Peptides were separated using a 90 min gradient from 2% to 30% solvent B followed by 15 min wash at 90% solvent B and the flow rate was 300 nL/min.

Both MS instruments were run in positive mode. The LTQ-Orbitrap was operated with the electrospray voltage at +1.7 kV. MS survey scan (from m/z 250–1800) was acquired at a resolution of 60,000 (fwhm at m/z 400). The top ten most intense ions were isolated and fragmented by CID (normalized collision energy: 35%) for the LTQ Orbitrap. The Orbitrap ELITE was operated in positive mode with the electrospray voltage at +3 kV and the top twenty most intense ions were isolated and fragmented by CID (normalized collision energy: 28%) in the linear ion trap. The HCD was triggered by neutral loss of isocyanic acid and fragment ions were detected in the Orbitrap at a resolution of 30,000 (fwhm at m/z 400).

Peptides and proteins were initially identified using the Sorcerer 2 SEQUEST (version 3.5, Sage-N Research, Milpitas, CA) search engine with post-search analysis using Scaffold 3 (Proteome Software Inc., Portland, OR). MS/MS spectra were searched against the UniProtKB human sequence database (147700 entries) or IPI bovine sequence database (v3.62, 83947 entries). Search parameters included full-enzymatic digestion with up to three missed cleavages, fixed modification being carbamidomethylation at cysteine residues, and variable modifications being asparagines or glutamine deamination, and arginine deamination. Mass tolerance was 20 ppm for precursor ions and 1.0 Da for fragment ions for both instruments. Peptides with asparagines (N) or glutamine (Q) deamination have the same mass increment (0.98402 amu) and similar shift in retention time. The positive assignment rate by search engine was usually less than 10%. As a result, manual verification to MS/MS spectra of citrullinated peptides and their unmodified counterpart was needed to remove the false assignment of citrulline. Manual validation includes confirming that experimental mass-to-charge ratio of precursor ion was within the specified mass tolerance;

MS/MS spectra matched theoretical fragment ions of the peptides. Ideally, chromatographic peaks of citrullinated peptides and their unmodified counterpart should be identified from raw data. Each experiment was repeated three times.

Results

1. In vivo analysis of human GFAP using CID triggered HCD

To assess the potential for the citrullination of GFAP, the purified human protein was treated with the PAD2 enzyme in presence of activating calcium. Trypsin, Lys-C or Glu-C were used to generate appropriate peptides of the untreated and treated sample prior to mass spectrometry analysis. In the untreated sample, there were 5 natural citrullination sites, residues R30, R36, R270, R406, and R416, identified for human GFAP protein (Table 1, MS/MS spectra of the citrullinated peptides were shown in Supplement Figure 2). Each MS/MS spectrum was manually verified by comparing it with theoretical fragment ions. The peptide sequences were also confirmed by high-resolution Orbitrap MS analysis (Supplement Table 1).

As shown in Figure 2A, neutral loss of isocyanic acid (HNCO, accurate mass 43.0058 amu) is the dominant fragmentation pathway upon collision induced dissociation [23]. This unique fragment ion was monitored for identification and verification of citrullinated peptides. For HCD fragmentation (Figure 2B), the abundance of the neutral loss fragment ions was reduced significantly and peptide amide backbone cleavage dominates MS/MS spectra of citrullinated peptides. This can be explained by additional collision events with nitrogen molecules in the collision cell. The formation of sequence-informative b and y ions greatly facilitates the identification of citrullinated peptides and proteins from complex protein digests. At low-mass region, b₂, b₃, and y₂ ions are present in HCD spectrum, but not in CID spectrum. Because nominal resolution of HCD was 15,000, the charge state of fragment ions was determined. As a result, HCD spectra are complementary to CID spectra for the identification of citrullinated peptides.

After PAD2 treatment of GFAP, a total of seventeen citrullinated arginine residues were identified (Table 1). Note this includes site that were detected with trypsin, Lys-C, and Glu-C digestion. These included the five *in vivo* citrullination sites of GFAP observed in the untreated sample as well as an additional twelve citrullinated residues. The peptide sequences were confirmed by CID spectra (Supplement Figure 2) and the high accurate m/z values of the precursor ions.

2. LC-MS/MS analysis of bovine MBP protein

A total of five arginine residues of bovine MBP were found citrullinated in the untreated sample whether digested with trypsin (Table 1) or digested with Lys-C (Supplement Table 2). The observation of citrullinated peptides NIVTPR*TPPPSQGK (residue 91–104), PGFGYGGR*ASDYK (residue 122–134), and their unmodified counterpart NIVTPR (residue 91–96), PGFGYGGR (residue 122–129) indicated that citrullination may inhibit proteolytic digestion by trypsin at the C-terminal of citrulline residue. However, one citrullinated peptide of MBP does contain citrulline at the C-terminal (Table 1 and

Supplement Table 2). The peptide NIVTPRTPPPSQGK can also undergo deamination at residue Q102. This deamination was reported previously in chicken MBP protein [26]. As the neutral loss of the isocyanic acid is only present in the CID spectrum of the citrullinated peptide and not the deaminated peptide, the neutral loss peak can be used as a diagnostic ion to differentiate between peptide that has an N or Q deamination from that of the R citrullinated peptide. However, MS/MS spectrum of peptides with N or Q deamination may not be collected due to low ion intensities. In this case, high resolution HCD spectrum was helpful in identifying the peptides with citrullination at arginine residues.

In the bovine MBP sample treated *in vitro* with the PAD2 enzyme, five novel (not detected previously) citrullination sites were identified, and all eleven citrullination sites that had previously reported were confirmed (Table 1). The CID spectra of all peptides were shown in Supplement Table 1 and Supplement Figure 3. The C-terminal peptide LGGR*DSR*SGSPMAR*R* contains four potential citrullination sites and truncated C-terminal peptides were also observed (Table 1).

3. Citrullinated Peptide of NRGN

Bovine neurogranin was identified within in the commercial bovine MBP sample, as a contaminant which co-purified with MBP. This bovine NRGN protein was found both as the R68 unmodified and citrullinated forms (Supplement Table 3). The sequence of the citrullinated peptides containing R68 were confirmed by MS/MS spectra upon CID (Supplement Figure 4) and accurate m/z value of precursor ions. To locate all possible citrullination sites in human NRGN, which is highly conserved with bovine, recombinant human NRGN was treated with PAD2 enzyme and the untreated and treated samples were digested with either trypsin, Lys-C, or Glu-C. Not unexpected, residue R68 of human NRGN was also citrullinated in human NRGN after PAD2 treatment. Five additional citrullination sites were observed for recombinant protein NRGN after treatment (Table 1 and Supplement Table 3). This included an additional R80 amino acid not found in the endogenous protein but had been introduced during cloning of the NRGN. These results demonstrated that all five arginine residues of human NRGN can be citrullinated *in vitro*. MS/MS spectra of the citrullinated peptides of human NRGN were shown in Supplement Figure 4.

4. Citrullinated proteins in human brain samples

Endogenous citrullinated sites of three brain proteins GFAP, MBP, and NRGN were identified and quantified in brain tissue samples obtained from non-Alzheimer controls (n=3) and with Alzheimer's disease (n=3) (Supplement Table 5 for clinical description). GFAP, MBP, and NRGN were identified in all of the human brain tissue samples with maximum amino acid sequence coverage of 64%, 54%, and 44% respectively. The peptide sequences and the citrullinated residues are listed in Table 2. *In vivo*, GFAP contained five citrullination sites at residues R30, R36, R270, R406, and R416 while MBP had fourteen residues citrullinated, including six novel arginine residues (R32, R44, R50, R92, R189, and R196). Mapping of citrullinated arginine residues demonstrated conserved *in vivo* citrullination of MBP R92 and R124, GFAP R406 and 416 in both the non-AZ and AZ samples and MBP R32 and GFAP R270 detected only in AZ samples. Residue R68 of NRGN was found citrullinated (representative chromatographic peak of the citrullinated and

unmodified peptides containing residue R68 of human NRGN see Figure 3). With mass tolerance of 10 ppm, the citrullinated peptide can be differentiated from the C13 peak of the unmodified peptide based on the m/z value of the precursor ion.

Based on integrated peak area (Figure 3C), occupancy rate of citrullination can be estimated by calculating the ratio of modified peptide vs. the sum of modified and unmodified peptides. It was estimated that occupancy rate of citrullinated peptide at R68 residue was 8.6% for the AD3 brain tissue sample without considering different ionization efficiency of the two peptides (Figure 3D). The occupancy rate of the endogenous citrullination site of NRGN (residue R68) and MBP (residues R92 and R124) in all brain tissue samples were determined for citrullinated peptides identified in Lys-C digests. For peptide DSHHPAR*TAHYGSLPQK of MBP that contains residue R92, peak area of both charge +4 and charge +3 ions are included. The occupancy rates of three sites were significantly higher for AD3 and Control_1 samples. Citrullination of NRGN at residue R68 was detected in AD1, AD3, and C1 control samples. Statistical assessment (t-test) of the occupancy level of the three sites for NRGN and MBP between AD and controls samples didn't reveal significant difference with such a small sample size.

Discussion

GFAP is a brain-specific cytoskeleton protein produced by astrocytes. Following traumatic brain injury and stroke, GFAP breaks down and is released to CSF and blood, which makes it a potential biomarker for a range of brain injury or diseases [16]. Our group has shown that GFAP is elevated and detected in the blood of children with birth related hypoxic-ischemic encephalopathy on extracorporeal cardiopulmonary support [18]. We propose that citrullination may be additionally useful as a disease-induced PTM with both biological and biomarker importance. Although the brain protein GFAP has been reported in several studies to be citrullinated by chemical modification and Western blot [4,11,27], the actual citrullinated residues have not been identified. This stimulated our interest in identifying the arginine residues of GFAP that are citrullinated *in vivo*. For the identification of citrullinated proteins, the common method is chemical derivatization of citrulline residues followed by antibody or mass spectrometric detection. However, derivatization is time consuming and requires strong cation-exchange and reversed-phase chromatographic separation to remove the excess reagents, 2,3-butanedione and antipyrine [13]. Based on previous studies, reaction efficiency of chemical modification is a concern and a great extent of citrullinated peptides may not be derivatized [13,28]. Considering that protein citrullination may not be abundant, this can be problematic for the identification of citrullinated proteins or peptides from protein digests. Mass spectrometric methods may provide the best means for the detection of citrulline modified arginine as exemplified by Hermansson *et al.* that reported identification of citrullinated peptides of fibrinogen based on high resolution mass spectrometry and retention time shift on reversed-phase HPLC [22]. In addition, neutral loss-triggered electron transfer dissociation has been reported recently for the identification of citrullinated peptides from simple protein mixtures [29]. Compared to ETD, HCD was a universal MS/MS method and involved minimum tuning of instrument parameters.

Because database search engine may not differentiate N or Q deamination from citrullination at arginine residue [22], citrullinated peptides can be differentiated from peptides with N or Q deamination based on unique fragmentation pathway, neutral loss of isocyanic acid upon CID. When Lys-C or Glu-C is used in proteolysis, identifying both peptides with and without citrullination is an effective way to remove false identification. When peptides are either too short or too long and can't be identified by database search, MS/MS spectra of targeted peptides were manually identified from raw data based on high accurate m/z values of precursor ions (<5 ppm). With direct mass spectrometric approach, we identified twelve endogenous citrullination sites for human MBP protein, four endogenous citrullination sites for GFAP, and one for NRGN. After treatment with PAD2 *in vitro*, seventeen arginine residues of human GFAP were citrullinated and all arginine residues present in bovine MBP can be citrullinated *in vitro*.

Citrullination of R41, R47, and R63 in bovine MBP protein have not been reported previously. Citrullination at residue R129 (R130 in human MBP) was previously reported in human MBP protein and citrullination of residue R96 was reported previously in bovine MBP [14,26]. It was reported previously that R106 residue of bovine MBP can be mono- or di-methylated [30]. The peptide, **GRGLSLSR**FSWGAEGQK (residue 105–121) without methylation, with mono- or di-methylation at residue R106 were observed in our study. After PAD2 treatment, **GR***GLSLSR*FSWGAEGQK (residue 105–121) with citrullination at residues R106 and R112, **GRGLSLSR***FSWGAEGQK with citrullination at R112 and with mono- or dimethylation at R106 were observed. This result suggests that arginine methylation prevents PAD2 enzyme catalyzed citrullination at the same residue. Our finding was consistent with a previous report indicating that this the case [31].

Bovine NRGN was identified in the bovine MBP sample and was found citrullinated at R68 residue. NRGN protein was identified in total protein extract from three human brain tissues and *in vivo* citrullination at residue R68 was confirmed in this study. These results demonstrated that direct mass spectrometric approach and CID triggered HCD fragmentation can be applied in the identification of citrullinated peptides from complex biological samples. This is the first time that citrullination of NRGN was discovered and the citrullinated residues of GFAP were determined.

NRGN/RC3 plays an important role in the long-term potentiation, learning, and memory [32]. NRGN binds calmodulin (CaM) at low intracellular calcium concentration and regulates calmodulin-dependent signaling pathway. The affinity of NRGN-CaM interaction is decreased when NRGN was phosphorylated at serine 36 (a protein kinase C or PKC site). In respond to influx of calcium, NRGN releases calmodulin and activates downstream signal transduction. Biochemical characterization of NRGN suggests that it may play a significant role in Ca²⁺ signaling and PKC-mediated signaling pathways. Recently, NRGN gene was found to be associated with schizophrenia and bipolar disorder in genome wide association studies. As well, the protein level of NRGN was found elevated in cerebrospinal fluid of Alzheimer's disease [21]. Our results showed that all arginine residues of human NRGN can be citrullinated *in vitro* and residue R68 was an endogenous citrullination site for human and bovine NRGN with variable citrullination in small number of control and Alzheimer brain samples.

Citrullination of arginine residues in CNS proteins may in addition to being biologically important provide a unique marker to differentiate disease from normal patient or to determine progression of the disease. As citrullination of proteins has been proposed to increase immunogenicity of several proteins, it is possible that auto-antibodies against the various citrullinated forms of these proteins could act as auto-antigens following brain injury and exacerbating brain injury. Thus, detection (or blocking) of these auto-antibodies could be used for improved prognosis, risk stratification, or therapy. This study demonstrated that CID triggered HCD fragmentation and high resolution mass spectrometry coupled with reversed-phase HPLC separation allowed identification of citrullinated peptide from complex matrices with great confidence.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NHLBI proteomics contract NHLBI-HV-10-05 (J. Van Eyk) and the CTSA grant 1U54RR023561-01A1.

Abbreviation

PAD	Peptidylargininedeiminase
RA	Rheumatoid arthritis
GFAP	Glial fibrillary acid protein
MBP	Myelin basic protein
NRGN/RC3	Neurogranin
CNS	Central nervous system
MS	Mass spectrometry
CID	Collision-induced dissociation
HCD	Higher-energy collisional dissociation

References

1. Arita K, Hashimoto H, Shimizu T, Nakashima K, Yamada M, et al. Structural basis for Ca²⁺-induced activation of human PAD4. *Nat Struct Mol Biol.* 2004; 11:777–783. [PubMed: 15247907]
2. Harauz G, Musse AA. A tale of two citrullines - Structural and functional aspects of myelin basic protein deimination in health and disease. *Neurochem Res.* 2007; 32:137–158. [PubMed: 16900293]
3. Wegner N, Lundberg K, Kinloch A, Fisher B, Malmstrom V, et al. Autoimmunity to specific citrullinated proteins gives the first clues to the etiology of rheumatoid arthritis. *Immunol Rev.* 2010; 233:34–54. [PubMed: 20192991]
4. Nicholas AP, Sambandam T, Echols JD, Tourtellotte WW. Increased citrullinated glial fibrillary acidic protein in secondary progressive multiple sclerosis. *J Comp Neurol.* 2004; 473:128–136. [PubMed: 15067723]
5. Ishigami A, Maruyama N. Importance of research on peptidylarginine deiminase and citrullinated proteins in age-related disease. *Geriatr Gerontol Int.* 2010; 10:S53–S58. [PubMed: 20590842]

6. Giles JT, Fert-Bober J, Park JK, Bingham CO III, Andrade F, et al. Myocardial citrullination in rheumatoid arthritis: a correlative histopathologic study. *Arthritis Res Ther.* 2012; 14:R39. [PubMed: 22364592]
7. Suzuki A, Yamada R, Yamamoto K. Citrullination by peptidylarginine deiminase in rheumatoid arthritis. *Autoimmunity, Pt D.* 2007; 1108:323–339.
8. Jang B, Shin HY, Choi JK, Du PTN, Jeong BH, et al. Subcellular Localization of Peptidylarginine Deiminase 2 and Citrullinated Proteins in Brains of Scrapie-Infected Mice: Nuclear Localization of PAD2 and Membrane Fraction-Enriched Citrullinated Proteins. *J Neuropath Exp Neur.* 2011; 70:116–124. [PubMed: 21343880]
9. Pruijn GJ, Wiik A, van Venrooij WJ. The use of citrullinated peptides and proteins for the diagnosis of rheumatoid arthritis. *Arthritis Res Ther.* 2010; 12:203–210. [PubMed: 20236483]
10. Wiik AS, van Venrooij WJ, Pruijn GJ. All you wanted to know about anti-CCP but were afraid to ask. *Autoimmun Rev.* 2010; 10:90–93. [PubMed: 20727426]
11. Raijmakers R, Vogelzangs J, Croxford JL, Wesseling P, van Venrooij WJ, et al. Citrullination of central nervous system proteins during the development of experimental autoimmune encephalomyelitis. *J Comp Neurol.* 2005; 486:243–253. [PubMed: 15844172]
12. Wood DD, Moscarello MA. The isolation, characterization, and lipid-aggregating properties of a citrulline containing myelin basic protein. *J Biol Chem.* 1989; 264:5121–5127. [PubMed: 2466844]
13. Stensland M, Holm A, Kiehne A, Fleckenstein B. Targeted analysis of protein citrullination using chemical modification and tandem mass spectrometry. *Rapid Commun Mass Spectrom.* 2009; 23:2754–2762. [PubMed: 19639564]
14. Zand R, Li MX, Jin XY, Lubman D. Determination of the sites of posttranslational modifications in the charge isomers of bovine myelin basic protein by capillary electrophoresis mass spectroscopy. *Biochemistry.* 1998; 37:2441–2449. [PubMed: 9485392]
15. Schulz P, Cruz TF, Moscarello MA. Endogenous Phosphorylation of Basic-Protein in Myelin of Varying Degrees of Compaction. *Biochemistry.* 1988; 27:7793–7799. [PubMed: 2462908]
16. Schiff L, Hadker N, Weiser S, Rausch C. A literature review of the feasibility of glial fibrillary acidic protein as a biomarker for stroke and traumatic brain injury. *Mol Diagn Ther.* 2012; 16:79–92. [PubMed: 22497529]
17. Ennen CS, Huisman TA, Savage WJ, Northington FJ, Jennings JM, et al. Glial fibrillary acidic protein as a biomarker for neonatal hypoxic-ischemic encephalopathy treated with whole-body cooling. *Am J Obstet Gynecol.* 2011; 205:251–257. [PubMed: 21784396]
18. Bembea MM, Savage W, Strouse JJ, Schwartz JM, Graham E, et al. Glial fibrillary acidic protein as a brain injury biomarker in children undergoing extracorporeal membrane oxygenation. *Pediatr Crit Care Me.* 2011; 12:572–579.
19. Giacoppo S, Bramanti P, Barresi M, Celi D, Cuzzola VF, et al. Predictive Biomarkers of Recovery in Traumatic Brain Injury. *Neurocrit Care.* 2012; 16:470–477. [PubMed: 22528282]
20. Greene DN, Schmidt RL, Wilson AR, Freedman MS, Grenache DG. Cerebrospinal Fluid Myelin Basic Protein Is Frequently Ordered but Has Little Value A Test Utilization Study. *Am J Clin Pathol.* 2012; 138:262–272. [PubMed: 22904139]
21. Thorsell A, Bjerke M, Gobom J, Brunhage E, Vanmechelen E, et al. Neurogranin in cerebrospinal fluid as a marker of synaptic degeneration in Alzheimer's disease. *Brain Res.* 2010; 1362:13–22. [PubMed: 20875798]
22. Hermansson M, Artemenko K, Ossipova E, Eriksson H, Lenggqvist J, et al. MS analysis of rheumatoid arthritic synovial tissue identifies specific citrullination sites on fibrinogen. *Proteom Clin Appl.* 2010; 4:511–518.
23. Hao G, Wang DC, Gu J, Shen QY, Gross SS, et al. Neutral Loss of Isocyanic Acid in Peptide CID Spectra: A Novel Diagnostic Marker for Mass Spectrometric Identification of Protein Citrullination. *J Am Soc Mass Spectr.* 2009; 20:723–727.
24. Nagaraj N, D'Souza RC, Cox J, Olsen JV, Mann M. Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation. *J Proteome Res.* 2010; 9:6786–6794. [PubMed: 20873877]

25. Zhao P, Viner R, Teo CF, Boons GJ, Horn D, et al. Combining high-energy C-trap dissociation and electron transfer dissociation for protein O-GlcNAc modification site assignment. *J Proteome Res.* 2011; 10:4088–4104. [PubMed: 21740066]
26. Kim J, Zhang R, Strittmatter EF, Smith RD, Zand R. Post-translational Modifications of Chicken Myelin Basic Protein Charge Components. *Neurochem Res.* 2009; 34:360–372. [PubMed: 18618245]
27. Jang B, Kim E, Choi JK, Jin JK, Kim JI, et al. Accumulation of citrullinated proteins by up-regulated peptidylarginine deiminase 2 in brains of scrapie-infected mice - A possible role in pathogenesis. *Am J Pathol.* 2008; 173:1129–1142. [PubMed: 18787103]
28. De Ceuleneer M, De Wit V, Van Steendam K, Van Nieuwerburgh F, Tilleman K, et al. Modification of citrulline residues with 2,3-butanedione facilitates their detection by liquid chromatography/mass spectrometry. *Rapid Commun Mass Spectrom.* 2011; 25:1536–1542. [PubMed: 21594927]
29. Creese AJ, Grant MM, Chapple LLC, Cooper HJ. On-line liquid chromatography neutral loss-triggered electron transfer dissociation mass spectrometry for the targeted analysis of citrullinated peptides. *Anal Methods.* 2011; 3:259–266.
30. Brame CJ, Moran MF, McBroom-Cerajewski LDB. A mass spectrometry based method for distinguishing between symmetrically and asymmetrically dimethylated arginine residues. *Rapid Commun Mass Spectrom.* 2004; 18:877–881. [PubMed: 15095356]
31. Pritzker LB, Joshi S, Harauz G, Moscarello MA. Deimination of myelin basic protein. 2. Effect of methylation of MBP on its deimination by peptidylarginine deiminase. *Biochemistry.* 2000; 39:5382–5388. [PubMed: 10820009]
32. Diez-Guerra FJ. Neurogranin, a Link Between Calcium/Calmodulin and Protein Kinase C Signaling in Synaptic Plasticity. *Iubmb Life.* 2010; 62:597–606. [PubMed: 20665622]

Figure 1.
Conversion of peptidylarginine to peptidylcitrulline by PAD enzymes in the presence of calcium

Figure 2. Comparison of CID and HCD spectra of a citrullinated peptide of human GFAP in Glu-C digests. A: CID spectrum of the citrullinated peptide, **GHLKR***NIVVKTVE of human GFAP at m/z 747.44, z : +2; B: HCD spectrum of this citrullinated peptide at m/z 747.44.

Figure 3.

Extracted ion chromatogram of the in vivo citrullinated peptide and intact peptide of NRGN in Lys-C digests of the AD3 brain sample. A: HPLC peak of the unmodified peptide at m/z 909.9275 with mass tolerance of 10 ppm. Right: isotopic clusters of the +2 ion. B: HPLC peak of the citrullinated peptide at m/z 910.4195 with mass tolerance of 10 ppm. Right: isotopic clusters of the +2 ion. C: Overlap of HPLC peaks in A and B. From integrated peak area, occupancy rate of citrullinated peptide at R68 residue was 8.6% for the AD3 sample. D: Citrullination occupancy rate for NRGN at R68, MBP at R92 and R124. Occupancy rate was calculated based on integrated peak area of citrullinated peptide vs. the sum of citrullinated and intact peptides.

Table 1

Citryllination sites of MBP, GFAP, and NRG1

Protein, UniProt ID	Peptide	Sequence number ^a	Enzyme	<i>In vitro</i> site	<i>In vivo</i> site
MBP, bovine P02687	YLASASTMDHAR*HGFLPR	12-29	Trypsin	R23	
	YLASASTMDHAR*HGFLPR*HR	12-31	Trypsin	R23, R29	
	YLASASTMDHAR*HGFLPR*HR*	12-31	Trypsin	R23, R29, R31	
	DTGILDSLR*FFGSDR	32-47	Trypsin	R41	R41
	FFGSDR*GAPK	42-51	Trypsin	R47	R47
	R*SGSGKDGHHAAAR*TTHYGSLPQK	52-73	Lys-C	R52, R63	R63
	DGHHAAR*TTHYGSLPQK	57-73	Trypsin, Lys-C	R63	R63
	AQGHAR*PDENPVVHFFK	74-90	Trypsin, Lys-C	R78	
	NIVTPR*TPPPSQGK	91-104	Trypsin, Lys-C	R96	R96
	GR*GLSLSR*FSWGAEGQK	105-121	Trypsin, Lys-C	R106, R112	
	PGFGYGGRR*ASDYK	122-134	Trypsin, Lys-C	R129	R129
	LGGRR*DSR*SGSPMA(-) ^c	155-167	Trypsin, Lys-C	R158, R161	
	LGGRR*DSR*SGSPMAR*(-)	155-168	Trypsin, Lys-C	R158, R161, R168	
	LGGRR*DSR*SGSPMAR*RR*(-)	155-169	Trypsin, Lys-C	R158, R161, R168, R169	
	R*LPGGTR*LSLAR	30-41	Trypsin	R30, R36	R30, R36
GFAP, Human P14136	LPGGTR*LSLAR	31-41	Trypsin	R36	R36
	VR*FLEQNK	87-95	Lys-C	R88	
	ALAAELNQLR*AK	96-107	Lys-C	R105	
	LR*LRLDQLTANSAR*LE	123-138	Glu-C	R124, R126, R136	
	AENNLAAAYR*QEADE	165-178	Glu-C	R173	
	LQEQLAR*QQVHVE	211-223	Glu-C	R217	
	WYR*SKFADLTDAAR*NAE	256-273	Glu-C	R258, R270	
	FADLTDAAR*NAELLR	261-276	Trypsin	R270	R270
	FADLTDAAR*NAELLR*QAK	261-279	Lys-C	R270, R276	
	ANDYR*RR*QLQSLTCDLE	282-297	Glu-C	R286, R287	
	LALDIEIATYR*K	357-368	Lys-C	R367	

Protein, UniProt ID	Peptide	Sequence number ^a	Enzyme	<i>In vitro</i> site	<i>In vivo</i> site
NRGN, human, Q92686	GHLKR*NIVVKTVE	402–414	Glu-C	R406	R406
	TVEMR*DGEVIK	412–422	Trypsin	R416	R416
	GPGPGGPGGAGVARR*GGAGGGPSGD(-)	55–78	Trypsin, Lys-C	R68	R68
	RGRKGPGPGGGAGVARR*GGAGGGPSGD(-)	51–78	Glu-C	R68	R68
NRGN, recombinant protein ^b	IQASFR*GHMAR*K	33–44	Lys-C	R38, R43	
	GPGPGGPGGAGVARR*GGAGGGPSGDTR*(-)	55–80	Lys-C	R68, R80 ^b	
	R*GR*KGPGPGGGAGVARR*GGAGGGPSGDTR*(-)	51–80	Glu-C	R51, R53 R68, R80 ^b	

^a sequence number includes the initiating Met at position 1 for all proteins

^b T79 and R80 are introduced into recombinant protein NRGN during subcloning

^c (-) indicates the C-terminal of this protein

Table 2*In vivo* citrullination sites of human MBP, GFAP, and NRG1

Protein, UniProt ID	Peptide	Enzyme	<i>In vivo</i> site	Control ^a			AD ^a		
				1	2	3	1	2	3
MBP, human P02686-3	YLATASTMDHAR*HGFLPR	Trypsin	R26	+	+		+		+
	YLATASTMDHAR*HGFLPR*HR	Trypsin	R26, R32				+		+
	HR*DTGILDSIGR	Trypsin	R34	+	+	+	+		+
	DTGILDSIGR*FFGGDR	Trypsin	R44	+	+		+		+
	FFGGDR*GAPK	Trypsin	R50	+	+		+	+	+
	DSHHPAR*TAHYGSLPQK	Trypsin, Lys-C	R92	+	+	+	+	+	+
	SHGR*TQDENPVVHFFK	Trypsin, Lys-C	R106	+	+	+	+		+
	NIVTPR*TPPPSQGK	Trypsin, Lys-C	R124	+	+	+	+	+	+
	FSWGAEGQR*PGFGYGGR	Trypsin	R149	+	+		+		+
	PGFGYGGR*ASDYK	Trypsin	R157	+	+		+	+	+
GFAP, Human P14136	LGGR*DSR*SGSPMAR*R*(-)	Trypsin, Lys-C	R186, R189 R196, R197	+		+	+		+
	R*LGPGTR*LSLAR	Trypsin	R30, R36	+		+	+	+	
	LGPGTR*LSLAR	Trypsin	R36	+		+	+	+	
	FADLTDAAR*NAELLR	Trypsin	R270					+	+
	GHLKR*NIVVKTVE	Glu-C	R406	+	+	+	+	+	+
NRG1, Human Q92686	TVEMR*DGEVIK	Trypsin	R416	+	+	+	+	+	+
	GPFGPGGGAGVAR*GGAGGPGSGD(-)	Trypsin, Lys-C	R68	+			+		+
	RGRKGPFGPGGAGVAR*GGAGGPGSGD(-)	Glu-C	R68	+					+

^a Refer to Supplement Table 5 for the details of brain tissue samples