Journal of
Proteomics & Bioinformatics

Bennike et al., J Proteomics Bioinform 2013, 6:12
DOI: 10.4172/jpb.1000293

Research Article Open Access

Optimizing the Identification of Citrullinated Peptides by Mass Spectrometry: Utilizing the Inability of Trypsin to Cleave after Citrullinated Amino Acids

Tue Bennike[1]*, Kasper B Lauridsen[1], Michael Kruse Olesen[2], Vibeke Andersen[3,4], Svend Birkelund[1] and Allan Stensballe[1]

[1]*Department of Health Science and Technology, Aalborg University, Fredrik Bajers Vej 3B, 9220 Aalborg, Denmark*
[2]*Department of Rheumatology, and Center for Clinical Research, Vendsyssel Teaching Hospital, Bispensgade 37, 9800 Hjoerring, Denmark*
[3]*Institute of Regional Health Services Research, University of Southern Denmark, Winslowparken 19, 3, 5000 Odense C, Denmark*
[4]*Organ Center, Hospital of Southern Jutland, Kresten Philipsens Vej 15, 6200 Aabenraa, Denmark*

Abstract

Citrullinated proteins have been associated with several diseases and citrullination can most likely function as a target for novel diagnostic agents and unravel disease etiologies. The correct identification of citrullinated proteins is therefore of most importance. Mass spectrometry (MS) driven proteomics can with bottom up strategies analyze protein profiles and PTMs in complex samples. However, the site-specific characterization of citrullination using MS remains problematic, especially in complex samples where no sensitive chemical modification technique exists. A tryptic missed cleavage after citrulline is therefore often used as a marker for citrullination post processing. However, C-terminal tryptic citrullinated peptides have also been reported. In this study, we therefore aimed at optimizing the identification of citrullinated peptides in complex samples.

To assess the cleavage properties of trypsin, digestion was performed on synthetic peptide sets containing either arginine or citrulline. The peptide sequences originated from disease-associated *in vivo* citrullinated proteins; some reported as being C-terminal tryptic citrullinated peptides. Furthermore, the proteolytic activity was verified using digested synovial fluid samples from a rheumatoid arthritis patient. The samples were analyzed using liquid chromatography/tandem MS with electrospray ionization.

Our *in vivo* and *in vitro* studies clearly demonstrate the inability of trypsin to cleave after citrulline residues. Based on our findings, we present a strategy for verifying citrullinated sites in complex samples post processing, in proteomics shotgun experiments. By requiring a missed cleavage for the identification of citrullinated peptides, we demonstrate that 64% of false-positively annotated citrullination sites could be removed. We furthermore demonstrated likely pitfalls of applying the strategy.

In conclusion, manual annotation of citrullinated peptide spectra remains essential to ensure correct annotation. Implementing a missed cleavage requirement significantly reduces the number of spectra needing manual verification with minimal loss. This method may help future proteomics studies identify citrullinated proteins in complex samples.

Keywords: Citrulline; Citrullination; Missed cleavage; Mass spectrometry; Trypsin; Digestion; Synovial fluid

Abbreviations: Aa: Amino Acid; Anti-CCP2: Second Generation Test for Anti-Citrullinated Protein Antibodies; Arg: Arginine; Cit: Citrulline; ESI: Electrospray Ionization; FA: Formic Acid; FDR: False Discovery Rate; LC: Liquid Chromatography; m/z: Mass-to-Charge; MALDI: Matrix-Assisted Laser Desorption/Ionization; MS: Mass Spectrometry; PAD: Peptidylarginine Deiminases; PTM: Posttranslational Modification; RA: Rheumatoid Arthritis; RT: Retention Time; SP: Synthetic Peptide; XIC: Extracted Ion Chromatogram

Introduction

The modification of proteins is a common biological process. After translation of the messenger RNA into protein, most proteins are covalently modified at least once [1]. These posttranslational modifications (PTMs) are often crucial to ensure the correct physiological function of the given protein. More than 200 distinct PTMs have been identified [2]. Furthermore, the PTMs can determine the activity state, localization, turnover, and interaction with other proteins and substrates [1,3-5]. In this study, we optimized the mass spectrometry (MS) driven identification of the PTM citrullination. Citrullinated proteins and auto-antibodies against these have been associated with several diseases including: rheumatoid arthritis (RA), Alzheimer's disease, and cancer [6-8]. MS remains the only method for identifying the exact site of citrullination, however, the correct identification of citrullinated peptides from MS data by automated search engines remains troublesome [4,9].

The posttranslational modification citrullination

Citrullination is the deimination of the amino acid (aa) arginine. In the reaction, one of the side-chain nitrogen atoms of arginine is hydrolyzed, yielding the non-standard amino acid citrulline (Figure 1) along with ammonia as a side product. The reaction was first described by Fearon [10] in 1939, and citrullination can take place in

***Corresponding author:** Tue Bennike, Department of Health Science and Technology, Aalborg University, Fredrik Bajers Vej 3B, 9220 Aalborg, Denmark, Tel: +45 2613 9003; Fax: +45 9814 1818; E-mail: tbe@hst.aau.dk

Received October 27, 2013; **Accepted** December 11, 2013; **Published** December 15, 2013

Citation: Bennike T, Lauridsen KB, Olesen MK, Andersen V, Birkelund S, et al. (2013) Optimizing the Identification of Citrullinated Peptides by Mass Spectrometry: Utilizing the Inability of Trypsin to Cleave after Citrullinated Amino Acids. J Proteomics Bioinform 6: 288-295. doi:10.4172/jpb.1000293

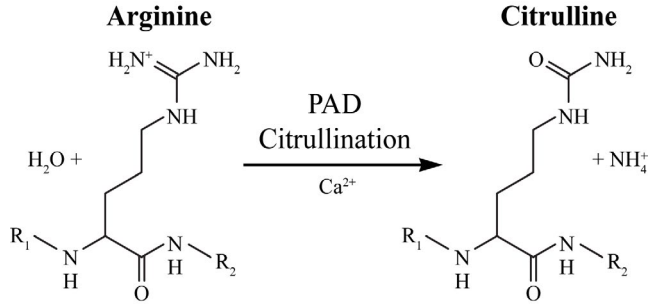

Figure 1: Citrullination of Arginine, *in vivo* catalyzed by the PAD enzyme family.

alkaline solutions at ambient temperature. Protein citrullination is *in vivo* catalyzed by a family of calcium-binding enzymes, namely the peptidylarginine deiminases (PADs). At physiological pH, arginine has a +1 charge due to the guanidine group, whereas citrulline is neutral. Thus, each citrullination event lowers the overall charge of the protein [5,11]. Essentially, the PTM leads to the loss of one positive charge and a monoisotopic mass difference of +0.984016 [4]. The loss of positive charge influences the overall charge of the protein. This in turn affects the charge distribution, isoelectric point, and hydrogen bond forming abilities of the protein [5]. The exact role of the modification remains largely unknown, but it is believed to alter the protein fold, change the protein polarity, or lead to denaturation in order to render the protein more prone to enzymatic degradation [5,11,12]. An example of a citrullinated protein is myelin basic protein, which accounts for up to 35% of the total myelin protein. Myelin is a major part of the central nervous system, where it functions as a non-conducting isolator between the nerve fibers [13,14]. Myelin exists in different charge isomers that differ by the degree of citrullination [11,15]. Finally, citrullination is known to play a role in the disease etiology of the joint disease rheumatoid arthritis (RA), and citrullinated proteins and peptides have previously been identified in the synovial fluid from these patients [12,16,17]. Moreover, RA is today diagnosed by detecting the presence of anti-citrullinated protein antibodies in the serum, with a sensitivity of 71% and a specificity of 95% [16]. The second generation test for anti-citrullinated protein antibodies (anti-CCP2) is regarded as the golden standard in diagnosing RA, however, the identification of novel disease associated citrullinated peptides may improve the anti-CCP2 test and allow for identifying clinically distinct RA patient subgroups [16]. Based on what is known about citrullinated proteins and RA, disease associated citrullinated proteins may be relevant for other inflammatory diseases as well. The knowledge of such specific citrullinated proteins will allow for analyzing disease pathways, unravel disease etiologies, and function as targets for novel diagnostic and therapeutic agents.

Detection of citrullinated peptides by mass spectrometry

In bottom up proteomic strategies, proteins are typically enzymatically digested by the protease trypsin prior to mass spectrometric analysis. Trypsin cleaves the C-terminal side of the basic aa arginine or lysine, unless either is succeeded by a proline [18]. As citrulline does not have the positive charge of arginine, the general assumption is that citrullination will result in a missed cleavage by trypsin [19-22]. This correlates well with the kinetic function of trypsin, and observations of trypsin cleavage properties [18,23]. However, tryptic peptides without a missed cleavage after citrulline have also been reported [4,24]. This could likely be due to over cleavage by trypsin or incorrect annotations made by the automated search engines. Therefore, as stated by De Ceuleneer et al. [4] in 2012, caution has to be taken when interpreting a missed cleavage as an indication of a citrullinated peptide.

In the present study, we decided to analyze the end product of tryptic digested peptides citrulline containing peptides and proteins using liquid chromatography (LC)-tandem MS sequencing. The sequences were chosen from identified disease relevant citrullinated proteins found in literature. We, furthermore, analyzed synovial fluid from a RA patient to make a similar assessment on an *in vivo* sample. This also allowed us to assess the number of citrullinated proteins found in the synovial fluid. Finally, we assessed the quality of automated citrullination annotation in the Mascot search engine, common to most proteome laboratories.

Materials and Methods

Digestion of peptides

A total of 24 synthetic peptide sets containing either arginine or citrulline (Table 1) were designed and purchased with carbamoylmethyl modified cysteine residues (JPT Peptide Technologies GmbH, Berlin, Germany). The aa sequences originated from reported citrullinated tryptic peptides from *in vivo* modified proteins, and some sites were reported as C-terminal citrullinated. The peptide sequences were selected to ensure that nearly all peptides had at least one lysine or arginine to confirm the successful digestion. The freeze-dried peptides were resuspended according to the manufacturer's instructions, and the masses were verified by matrix assisted laser desorption/ionization (MALDI) MS (data not shown). Tryptic digestion was performed in 5% acetonitrile (ACN) and 50 mM ammonium bicarbonate. 0.5 µg crudepeptide was added to 0.05 µg sequencing grade trypsin (Promega, Fitchburg, USA). The samples were digested overnight at 37°C and acidified with formic acid (FA) to a concentration of 5% followed by nanoLC-MS/MS analysis. The undigested samples were treated identically, but the addition of trypsin was omitted. The peptides were LC-MS analyzed one sample at a time, except in one case where two samples were injected simultaneously to verify the observed change in retention time (RT). All mass-calculations were performed in GPMAW 9.02 (Lighthouse Data, Odense, Denmark).

Synovial fluid sample preparation

Synovial fluid was obtained from a 69 year old female RA patient, who was tested positive for anti-cyclic citrullinated peptide antibodies and rheumatoid factor. The digestion of synovial fluid prior to MS analysis was performed using the filter aided sample preparation Protein Digestion Kit (Expedeon, San Diego, USA) according to manufacturer's instructions using 30 kDa cutoff spin filters. 90 µg total SF protein was digested using two µg sequencing grade modified trypsin (Promega), and the samples were digested overnight at 37°C. After trypsin digestion, the samples were acidified with trifluoroacetic acid (TFA), desalted with TAGRA C18 columns (Nest Group, Southborough, USA), and finally resuspended in 2% acetonitrile (ACN) 1% FA.

Mass spectrometry

The synthetic peptide samples were analyzed by automated LC-electrospray ionization (ESI) MS/MS using a 1200 series Agilent nanoflow HPLC (Agilent Technologies, Santa Clara, USA) system coupled online by a nanospray ion source (Proxeon, Odense, Denmark)

SP#	Protein Name	Protein ID	Sequence	Ref.	% RT shift
1	α-enolase	P06733.2	13-DS **R/Cit** GNPTVEVDLFTS**K**GLF**R**-32	[20]	2.4
2	-	P06733.2	266-DPS **R/Cit** YISPDQLADLY**K**SFI**K**-285	[20]	3.0
3	-	P06733.2	415-EELGS**K**A**K**FAG **R/Cit** NF **R/Cit** NPLA**K**-434	[20]	1.6
4	Vimentin	AAH66956	31-VTTST**R**TYSLGSAL **R/Cit** PSTS**R**-50	[21]	1.4
5	-	AAH66956	136-EQL**K**GQG**K**S **R/Cit** LGDLYEEEM**R**-155	[21]	2.6
6	-	AAH66956	371-NM**K**EEMA**R**HL **R/Cit** EYQDLLNV**K**-390	[21]	1.4
7	-	AAH66956	371-NM**K**EEMA **R/Cit** HL**R**EYQDLLNV**K**-390	[21]	2.0
8	protein-arginine deiminase type-4	NP_036519.2	201-LHVA **R/Cit** SEMD**K**V **R/Cit** VFQAT **R/Cit** G**K**-220	[30]	3.6
9	-	NP_036519	375-GL**K**EFPI**K** **R/Cit** VMGPDFGYVT**R**-394	[30]	3.4
10	-	NP_036519	480-PAPD**RK**GF**R**LLLASP **R/Cit** SCY**K**-499	[30]	2.0
11	C-X-C motif chemokine 10	P02778.2	24-LS **R/Cit** TV**R**CTCISISNQPVNP**R**-43	[31]	1.6
12	tubulin polymerization promoting protein p25 α	ACB10579	72-EMHG**K**NWS**K**LC **R/Cit** DCQVIDG**R**-91	[22]	1.6
13	-	ACB10579	154-SGVT**K**AISSPTVS **R/Cit** LTDTT**K**-173	[22]	1.4
14	Fibrinogen α-chain	P02671	29-AEGGGV **R/Cit** GP**R**VVE **R/Cit** HQSAC**K**-48	[32]	0
15	-	P02671	562-SHHPGIAEFPS **R/Cit** G**K**SSSYS**K**-581	[32]	0.4
16	-	P02671	583-FTSSTSYN **R/Cit** GDSTFES**K**SY**K**-602	[32]	1.0
17	Myelin basic protein	P02686.3	158-A **R/Cit** HGFLP **R/Cit** H**R**DTGILDSIG**R**-177	[33]	2.6
18	-	P02686.3	245-LS**R**FSWGAEGQ **R/Cit** PGFGYGG**R**-264	[33]	2.0
19	-	P02686.3	257-PGFGYGG **R/Cit** ASDY**K**SAH**K**GF**K**-276	[33]	0.6
20	-	P02686.3	284-LS**K**IF**K**LGG **R/Cit** DS**R**SGSPMA**R**-303	[33]	0.6
21	Fibrinogen β-chain	P02675	266-Y **R/Cit** VYCDMNTENGGWTVIQN**R**-285	[24]	0
22	-	P02675	254-MYLIQPDSSV**K**PY **R/Cit** VYCDM**R**-273	[24]	3.4
23	-	P02675	55-EAPSL **R/Cit** PAPPPISGGGY**R**AR-74	[24]	2.0
24	Fibrinogen γ-chain	P02679.3	132-SI **R/Cit** YLQEIYNSNNQ**K**IVNL**K**-151	[24]	3.0

Table 1: The complete list of synthetic peptides (SP) derived from human proteins, along with protein accession number, amino acid position, the reference to the citrullinated protein, and the observed RT shift between the citrulline and arginine version of the peptide. Bold letters indicates expected tryptic cleavage sites, and Cit indicates citrulline.

to a hybrid microQTOF mass spectrometer (Bruker, Bremen, Germany). One pmol sample was loaded on an in-house packed 10 cm reversed phase C18 column using a single column system (Dr. Maisch, Germany; Reprosil-pur C18-AQ). The sample was eluted with a linear gradient of 98% solvent A (0.1% FA, 0.005% heptafluorobutyric acid) and 2% solvent B (90% ACN, 0.1% FA, 0.005% heptafluorobutyric acid), which was increased to 40% solvent B on a 30 minutes ramp gradient at a constant flow rate of 200 nL/min. The mass spectrometer was used in data dependent mode to automatically switch between MS and MS/MS acquisition, selecting the three most abundant ions. The resulting raw files were analyzed using Bruker Daltonics Data Analysis v 3.4 (Build 192), and extracted ion chromatograms (XIC) were constructed with all predicted tryptic peptides ± m/z 0.01, under the assumption that trypsin cleaves after arginine, lysine, and citrulline.

The synovial fluid sample was analyzed on an automated LC-ESI MS/MS setup using an UltiMate™ 3000 UPLC system upgraded with a RSLC nanopump module. The system was coupled online with an emitter for nanospray ionization (New objective 360-20-10) to a Q Exactive mass spectrometer (Thermo Scientific, Waltham, USA). Five µg sample were loaded onto a C18 reversed phase column (Dionex Acclaim PepMap RSLC C18, 2 µm, 100 Å, 75 µm×50 cm) and eluted with a linear gradient of 96% solvent A (0.1% FA, 0.005% heptafluorobutyric acid) and 4% solvent B (90% ACN, 0.1% FA, 0.005% heptafluorobutyric acid), which was increased to 35% solvent B on a 120 minutes ramp gradient at a constant flow rate of 300 nL/min. The mass spectrometer was used in a data dependent mode, selecting the 12 ions with the highest intensity for HCD fragmenting. Fragmented ions were dynamically excluded for 30 sec. The resulting raw files were analyzed using Thermo Proteome Discoverer v 1.4 (build 288) connected to a Mascot server with a local Homo sapiens Uniprot database. The raw files were searched against the database using the following parameters: maximum three missed cleavages, 10 ppm precursor mass tolerance, 30 mmu fragment mass tolerance, variable modifications: Citrullination (R), Deamidation (NQ), Oxidation (M), static modifications: carboxymethyl (C), and percolator with 1% false discovery rate (FDR). All peptides annotated as citrullinated or deamidated were subsequently manually assessed. For a successful annotation of citrullination, the PTM should either be visible on the fragment ion, or by the subsequent fragment ion series. Raw files were inspected in Qual Browser Thermo Xcalibur v 2.2 (build 42).

Results and Discussion

Tryptic cleavage of citrulline containing peptides

To demonstrate the overall effectiveness of the protease trypsin to cleave at modified residues, we investigated a total of 24 peptide pairs (Table 1). In the first set of experiments, the theoretical outcome by *in situ* digestion were examined and compared to the empirical data. Protein-arginine deiminase type-4 can be citrullinated at position 495_{Arg}. A synthetic peptide pair SP 10 and SP 10_{Cit} with arginine or citrulline respectively were made with the aa sequence 480-PAPD**RK**GF**R**LLLASP**R/Cit**SCY**K**-499 (Table 1). The digestion experiment demonstrated that trypsin cleaves after arginine, but not citrulline (Figure 2). For SP 10, prior to digestion, only the synthetic peptide was detected in the XIC. After digestion, peptides corresponding to PAPD**R**, LLLASP**R** and SCY**K** were detected, corresponding to a complete cleavage after 495_{Arg} (Figure 3). This was not the case for the citrullinated peptide, SP 10_{Cit}, where peptides corresponding to PAPD**R** and LLLASP**Cit**SCY**K** were detected after digestion. All investigated peptides demonstrated this behavior, and tryptic cleavage after a citrulline residue was never observed.

Commonly, in solution trypsin digestion protocols, a trypsin-

to-protein ratio of 1:50 to 1:100 is employed and a digestion time varying from a few hours to overnight digestion [18]. To ensure a complete tryptic digestion, we employed a significantly higher trypsin concentration in the performed tryptic digestions (1:10) combined with 12 h incubation at 37°C [25]. However, cleavage after citrulline could not be detected for any of the investigated peptides, whereas the arginine and lysine cleavage sites were fully cleaved. Our study thereby demonstrates that tryptic cleavage after citrulline is not occurring to a detectable extent under standard digestion conditions.

Citrullination and retention time shift

Citrullination results in an increased hydrophobicity of the modified peptide, and as an outcome citrullinated peptides will have a longer RT on a reversed phase high-performance LC column [26]. The XIC of the coinjection of SP 3 and $SP3_{cit}$ (Figure 4) demonstrates that the RT shift between these two peptides on this system and gradient is 3.0%. The RT shifts for the peptide sets were extracted across different LC runs (Table 1). For 22 of the 24 peptide sets, the RT shift was greater than 0.4% (7 s on the 30 min gradient), which in this experiment was enough to ensure baseline separation. The results demonstrate that coelusion for most peptides will not occur, which confirms the findings by Raijmakers et al. [26]. This becomes important when considering the +0.984016 Da mass shift caused by citrullination (and deamidation) which is close to that of a single neutron +1.008665 Da. As a result, if a citrulline and an arginine containing peptide are coeluting, the m/z signal from the citrullinated peptide will fall within the isotopic cluster of the non-citrullinated peptide. This will cause an apparent change in the isotopic cluster [27]. Due to the typical isolation window used on mass spectrometers (m/z 1 to 3), both the modified and unmodified peptide would in this scenario be selected for fragmentation, which most likely would hinder a successful identification of the PTM [4]. An alternative solution to dealing with coelusion would be to run the samples on a MS with ion-mobility, as this would likely separate the citrullinated peptides from the non-citrullinated peptides regardless of LC separation.

In a tryptic digest containing citrullinated proteins and non-citrullinated proteins, a likely scenario in *in vivo* samples, the site of citrullination will in the non-modified peptide be a tryptic cleavage site. The mass of the citrullinated peptide will, therefore, in most cases not overlap with the mass of the non-citrullinated peptide. However, when using other proteases such as LysC the RT shift becomes important for the successful identification of both peptides [4]. This is relevant when assessing the ratio of citrullinated peptides, as e.g. LysC will result in two comparable peptides. De Ceuleneer et al. [27] demonstrated in 2012 that the degree of skewed isotope pattern can be used to quantify the amount of citrullinated peptide, by designing gradients where the two peptides do coelute. However, in order to have an overlap of elusion, the LC gradient must be designed accordingly as at least 22 of our 24 peptide sets did not fulfill this requirement on our setup. This might limit the feasibility of the approach for complex samples.

Tryptic missed cleavage as a citrullination verification strategy

The identical mass shift of deamidation of asparginine or glutamine causes it to be mistaken for citrullination and vice versa by automated protein identification search engines, such as Mascot (Matrix Science, Boston, USA). This is especially pronounced if only one of the modifications is allowed as a variable modification in the search (data not shown). Resultantly, a manual verification of citrullinated peptide candidate tandem mass spectra remains necessary for unambiguous site specific identification. However, our findings support that a missed cleavage can be used to significantly reduce the number of reported citrullinated peptide spectra, which needs to be manually verified, since a deamidation of asparagine or glutamine will not lead to a tryptic missed cleavage. Combining the correct mass increment with the MS/

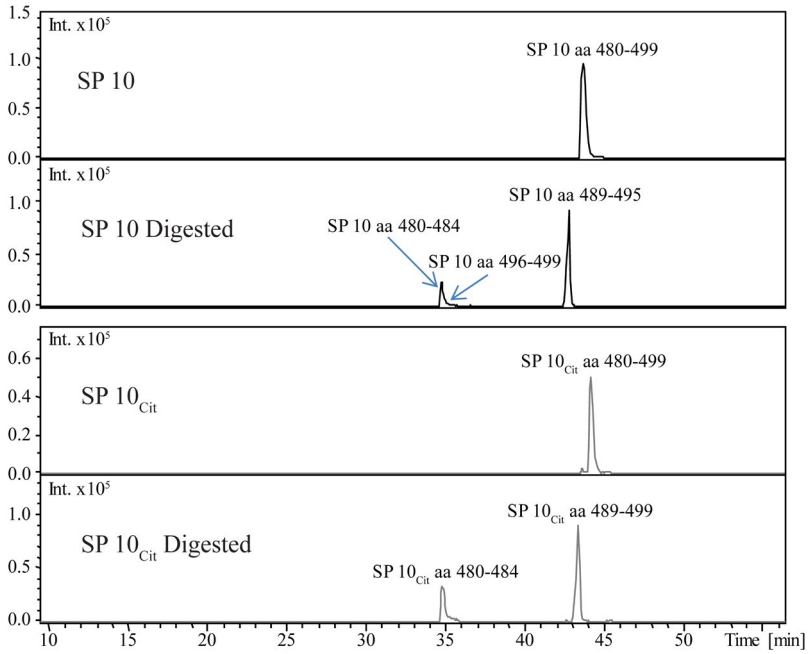

Figure 2: XIC of predicted citrullinated and non-citrullinated tryptic peptides +/- m/z 0.01, before and after tryptic digestion of SP 10 and $SP\ 10_{Cit}$, with the sequence aa 480-499. For digested SP 10, the fragments aa 489-495 and aa 496-499 are detected, confirming a cleavage after aa 495_{Arg}. For digested $SP\ 10_{cit}$, the fragment aa 489-499 is detected, confirming a missed cleavage after aa 495_{Cit}.

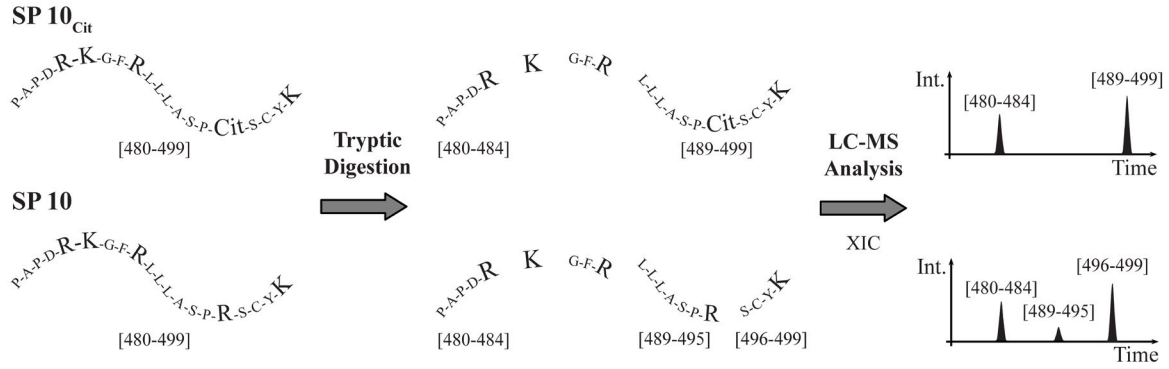

Figure 3: Citrullination protects the natural cleavage site in SP 10_{Cit}, where the entire fragment aa 489-499 is detected post digestion. For SP 10, the fragments aa 489-499 is cleaved into aa 489-495 and aa 496-499.

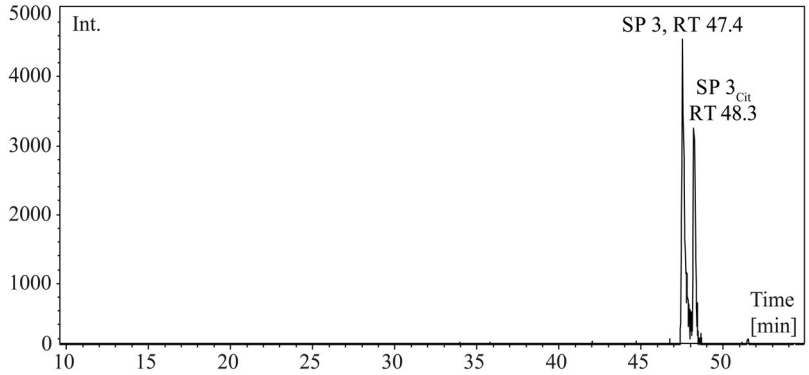

Figure 4: XIC of a co-injection of SP 3 and SP 3_{Cit} +/- m/z 0.01 demonstrating the RT shift caused by citrullination. The RT shift is calculated to be 3.0% (0.9 min) within this 30 minute gradient.

MS information, RT shift (if the non-citrullinated peptide is available) and hindered cleavage by trypsin seems to be a valid approach for a reliable identification of citrullinated peptides. It should be noted that *in vivo* samples may contain tryptic peptides with a citrullinated C-terminal, if the C-terminal of the original protein ends on a modified arginine. Using the above described identification parameters these peptides will be dismissed as false positives.

To verify the *in vitro* findings, synovial fluid from a RA patient was analyzed. The data was processed using Proteome Discoverer with Mascot as search engine, and a total of 364 proteins were identified at 1% FDR. All peptides annotated as being deamidated (NQ) or citrullinated (R) and contained at least one arginine were inspected.

Ten peptides were annotated as being deamidated (and not citrullinated), and contained at least one arginine. Of these, one was found to actually be citrullinated.

A total of 58 peptides were annotated as being citrullinated. 37 of these 58 peptides (64%) could be annotated directly and unambiguously from the fragment mass, and 7 of the 58 peptides (12%) by the masses of flanking fragments. However, 14 of the 58 peptides (24%) were found not to be citrullinated (Figure 5A).

When investigating non-citrulline peptides that were incorrectly annotated in greater detail, 10 of the 14 peptides (64%) were marked as having a citrullinated arginine on the C-terminal (Figure 5B). The remaining 4 of the 14 peptides (36%) contained an internal arginine residue. The detected mass shift was a result of the secondary isotope peak having been selected for MS/MS fragmentation, and not the monoisotopic peak. This was also the case for some of the peptides being incorrectly annotated as C-terminal citrullinated. The secondary isotope leads to a +1.008665 Da mass shift, which can be mistaken for a citrullination or a deamidation, especially in low resolution fragment spectra.

One of the 44 correctly annotated citrullinated peptides (2%) contained a citrullination on the C-terminal. Upon further analysis, the peptide was identified as the Complement C3f fragment (PMID P01024). The citrullination was found to be on position 1320, the C-terminal aa of the C3f fragment. C3f is a peptide that is released *in vivo* when C3b is converted to C3ib by the serine protease Factor I during complement activation [28,29]. Therefore, it seems likely that this is an example of the citrullinations that might be missed when implementing the no C-terminal citrullination limitation. Thus, such peptides will be lost in the data analysis when searching for citrullinated peptides and not allowing C-terminal citrullinations.

Finally, 30 of the 37 citrullinated peptides that could be unambiguously annotated, belonged to proteins previously identified as being citrullinated in synovial fluid by Van Beers et al. [17] (Table 2). Seven of the synovial fluid proteins have to the best of our knowledge not prior been identified as being citrullinated.

C-terminal peptides such as C3f are present; however, as demonstrated, by far the majority of the C-terminal citrullinated

peptides identified were incorrectly annotated. It therefore seems plausible that many peptides assigned as being C-terminal citrullinated in literature are due to deamidations being mistaken for citrullinations or that the secondary isotope peak of the peptide was selected for fragmentation. Hence, a manual verification of the fragment spectra seems necessary if a C-terminal citrulline is to be confidently annotated. Another means of unambiguously identifying a citrullination, is to use a chemical modification, as described by De Ceuleneer et al. [30].

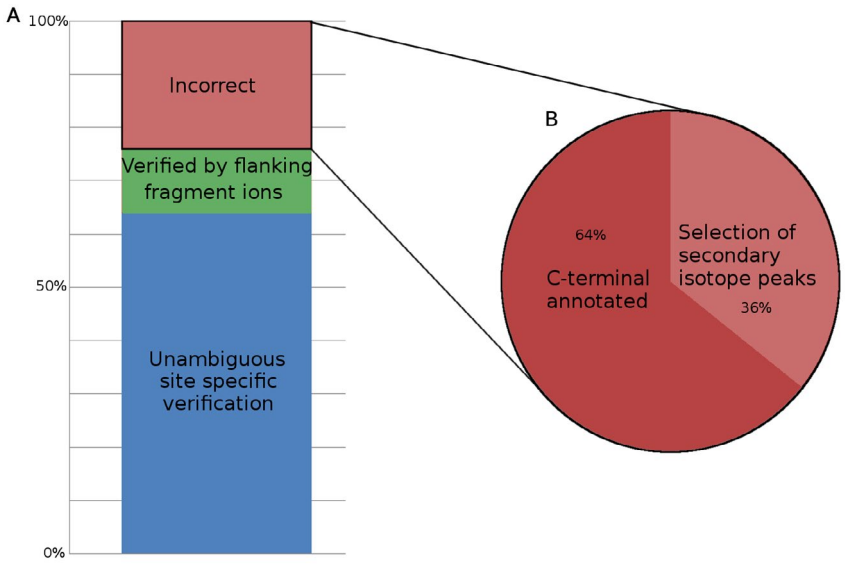

Figure 5: A: Distribution result of the manual PTM analysis of the peptides, annotated as being citrullinated by Mascot. **B:** Distribution of cause for incorrect citrullination annotation by Mascot.

Protein name	Protein ID	Sequence	Mod.Res.
Actin, cytoplasmic 1	P60709	85-IWHHTFYNEL **Cit** VAPEEHPVLLTEAPLNPK-113	96
α-1-antitrypsin	P01009	218-WE **Cit** PFEVKDTEEEDFHVDQVTTVK-241	220
α-2-macroglobulin	P01023	704-VGFYESDVMG **Cit** GHAR-719	715
Apolipoprotein A-I	P02647	231-AKPALEDL **Cit** QGLLPVLESFK-250	239
Apolipoprotein A-IV	P06727	276-GNL **Cit** GNTEGLQK-287	279
Apolipoprotein C-III	P02656	45-DALSSVQESQVAQQA **Cit** GWVTDGFSSLK-71	60
Apolipoprotein E	P02649	186-EGAE **Cit** GLSAIR-198	190
Apolipoprotein L1	O14791	321-VNEPSILEMS **Cit** GVK-334	331
Complement C1r subcomponent	P00736	472-MGNFPWQVFTNIHG **Cit** GGGALLGDR-494	484
Complement C3	P01024	1285-DAPDHQELNLDVSLQLPS **Cit**-1303	1303
Complement C4-A	P0C0L4	1350-QI **Cit** GLEEELQFSLGSK-1365	1352
Fibrinogen α-chain	P02671	582-QFTSSTSYN **Cit** GDSTFESK-599	591
Fibrinogen α-chain	P02671	481-EVVTSEDGSDCPEAMDLGTLSGIGTLDGF **Cit** HR-512	510
Fibrinogen α-chain	P02671	259-MELERPGGNEIT **Cit** GGSTSYGTGSETESPR-287	272
Fibrinogen α-chain	P02671	115-GDFSSANN **Cit** DNTYNR-129	123
Fibrinogen β-chain	P02675	53-**Cit** EEAPSLRPAPPPISGGGYR-72	53
Fibronectin	P02751	253-GNLLQCICTGNG **Cit** GEWK-269	265
Fibronectin	P02751	2335-RPGGEPSPEGTTGQSYNQYSQ **Cit** YHQR-2360	2354
Haptoglobin	P00738	117-L **Cit** TEGDGVYTLNNEK-131	118
Ig κ-chain V-III region GOL	P04206	46-LLMYGASS **Cit** ATGIPDRFSGSGSGTDFTLTISR-78	55
Inter-α-trypsin inhi-bitor heavy chain H2	P19823	136-TVG **Cit** ALYAQAR-146	139
Inter-α-trypsin inhi-bitor heavy chain H4	Q14624	658-MNF **Cit** PGVLSSR-669	661
Plasminogen	P00747	88-M **Cit** DVVLFEK-96	89
Protein AMBP	P02760	294-GPC **Cit** AFIQLWAFDAVK-309	297
Proteoglycan 4	Q92954	187-NSAAN **Cit** ELQK-196	191
Proteoglycan 4	Q92954	1307-AIGPSQTHTI **Cit** IQYSPARLAYQDK-1330	1317
Proteoglycan 4	Q92954	1285-RPALNYPVYGETTQV **Cit** R-1301	1300
Prothrombin	P00734	453-YNW **Cit** ENLD **Cit** DIALMK-467	456,461
Prothrombin	P00734	434-YE **Cit** NIEK-440	436
Serum albumin	P02768	97-LCTVATL **Cit** ETYGEMADCCAK-117	105

Table 2: The complete list of proteins where citrulline has been annotated directly and unambiguously on the fragment. Bold protein name indicates newly identified citrullinated RA synovial fluid proteins. All other proteins have been found by van Beers JJ et al. [17] to contain citrulline.

The citrulline residue is modified by the addition of an imidazolone derivative, which results in a mass shift of +50 Da. This method has the advantage of being highly specific for citrulline residues however, as noted in the article the method requires a high concentration of the citrullinated peptide(s) and the method is not suitable for complex protein samples. The missed cleavage strategy is applied post acquisition and search hence no chemical modification is required and the method is compatible with standard shotgun proteomics on complex samples.

Conclusion

Citrullination has been associated with several diseases, and auto-antibodies against citrullinated proteins are today used as an important clinical diagnostic biomarker for characterizing rheumatoid arthritis. The exact physiological role of citrullination in relation to diseases is incomplete, and specific analyses are needed to expand upon current knowledge. However, the essential MS driven verification of the exact site of modification remains problematic.

Using 24 sets of synthetic peptides containing citrulline and arginine, with sequences from previously identified citrullinated proteins reported in the literature, we have demonstrated the inability of trypsin to cleave after a citrulline residue. Furthermore, our study confirms that the RT shift between a citrullinated and a non-citrullinated peptide on a short reversed phase (C18) column and a 30 minutes gradient in most cases is enough to ensure that the two peptides will not coelute. C-terminal citrullinated tryptic peptides can, therefore, only occur if a given protein or protein fragment ends on a citrullinated arginine residue, which is the case for the citrullinated protein fragment C3f. Deamidation of asparagine or glutamine will not lead to a missed cleavage. Therefore, a missed cleavage can be used as a marker to reduce the number of reported citrullinated peptide spectra which needs to be manually validated. In an *in vivo* synovial fluid sample from a RA patient, 64% of the false positively reported citrullinations could be readily dismissed using this strategy. In addition, 7 proteins, not previously reported as being citrullinated, have been identified in the RA synovial fluid.

References

1. Mann M, Jensen ON (2003) Proteomic analysis of post-translational modifications. Nat Biotechnol 21: 255-261.
2. Thelen JJ, Miernyk JA (2012) The proteomic future: where mass spectrometry should be taking us. Biochem J 444: 169-181.
3. Jensen ON (2004) Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. Curr Opin Chem Biol 8: 33-41.
4. De Ceuleneer M, Van Steendam K, Dhaenens M, Deforce D (2012) In vivo relevance of citrullinated proteins and the challenges in their detection. Proteomics 12: 752-760.
5. György B, Tóth E, Tarcsa E, Falus A, Buzás EI (2006) Citrullination: a posttranslational modification in health and disease. Int J Biochem Cell Biol 38: 1662-1677.
6. Schellekens GA, de Jong BA, van den Hoogen FH, van de Putte LB, van Venrooij WJ (1998) Citrulline is an essential constituent of antigenic determinants recognized by rheumatoid arthritis-specific autoantibodies. J Clin Invest 101: 273-281.
7. Ishigami A, Ohsawa T, Hiratsuka M, Taguchi T, Kobayashi S, et al. (2005) Abnormal accumulation of citrullinated proteins catalyzed by peptidylargininedeiminase in hippocampal extracts from patients with Alzheimer's disease. J Neurosci Res 80: 120-128.
8. Chang X, Fang K (2010) PADI4 and tumourigenesis. Cancer Cell Int 10: 7.
9. Slade DJ, Subramanian V, Fuhrmann J, Thompson PR (2013) Chemical and biological methods to detect posttranslational modifications of arginine. Biopolymers.
10. Fearon WR (1939) Thecarbamidodiacetyl reaction: a test for citrulline. Biochem J 33: 902-907.
11. Curis E, Nicolis I, Moinard C, Osowska S, Zerrouk N, et al. (2005) Almost all about citrulline in mammals. Amino Acids 29: 177-205.
12. Hermansson M, Artemenko K, Ossipova E, Eriksson H, Lengqvist J, et al. (2010) MS analysis of rheumatoid arthritic synovial tissue identifies specific citrullination sites on fibrinogen. Proteomics ClinAppl 4: 511-518.
13. Hartline DK, Colman DR (2007) Rapid conduction and the evolution of giant axons and myelinated fibers. Curr Biol 17: R29-35.
14. Sjostrand FS (1953) Thelamellated structure of the nerve myelin sheath as revealed by high resolution electron microscopy. Experientia 9: 68-69.
15. Kim JK, Mastronardi FG, Wood DD, Lubman DM, Zand R, et al. (2003) Multiple sclerosis: an important role for post-translational modifications of myelin basic protein in pathogenesis. Mol Cell Proteomics 2: 453-462.
16. vanVenrooij WJ, van Beers JJ, Pruijn GJ (2011) Anti-CCP antibodies: the past, the present and the future. Nat Rev Rheumatol 7: 391-398.
17. van Beers JJ, Schwarte CM, Stammen-Vogelzangs J, Oosterink E, Božič B, et al. (2013) The rheumatoid arthritis synovial fluid citrullinome reveals novel citrullinated epitopes in apolipoprotein E, myeloid nuclear differentiation antigen, and ß-actin. Arthritis Rheum 65: 69-80.
18. Olsen JV, Ong SE, Mann M (2004) Trypsin cleaves exclusively C-terminal to arginine and lysine residues. Mol Cell Proteomics 3: 608-614.
19. Kinloch A, Tatzer V, Wait R, Peston D, Lundberg K, et al. (2005) Identification of citrullinated alpha-enolase as a candidate autoantigen in rheumatoid arthritis. Arthritis Res Ther 7: R1421-1429.
20. Bang H, Egerer K, Gauliard A, Lüthke K, Rudolph PE, et al. (2007) Mutation and citrullination modifies vimentin to a novel autoantigen for rheumatoid arthritis. Arthritis Rheum 56: 2503-2511.
21. Kleinnijenhuis AJ, Hedegaard C, Lundvig D, Sundbye S, Issinger OG, et al. (2008) Identification of multiple post-translational modifications in the porcine brain specific p25alpha. J Neurochem 106: 925-933.
22. Ordóñez A, Martínez-Martínez I, Corrales FJ, Miqueo C, Miñano A, et al. (2009) Effect of citrullination on the function and conformation of antithrombin. FEBS J 276: 6763-6772.
23. Vandermarliere E, Mueller M, Martens L (2013) Getting intimate with trypsin, the leading protease in proteomics. Mass Spectrom Rev 32: 453-465.
24. Zhao X, Okeke NL, Sharpe O, Batliwalla FM, Lee AT, et al. (2008) Circulating immune complexes contain citrullinated fibrinogen in rheumatoid arthritis. Arthritis Res Ther 10: R94.
25. Hustoft HK, Reubsaet L, Greibrokk T, Lundanes E, Malerod H (2011) Critical assessment of accelerating trypsination methods. J Pharm Biomed Anal 56: 1069-1078.
26. Raijmakers R, van Beers JJ, El-Azzouny M, Visser NF, BožičB, et al. (2012) Elevated levels of fibrinogen-derived endogenous citrullinated peptides in synovial fluid of rheumatoid arthritis patients. Arthritis Res Ther 14: R114.
27. De Ceuleneer M, Van Steendam K, Dhaenens M, Elewaut D, Deforce D (2012) Quantification of citrullination by means of skewed isotope distribution pattern. J Proteome Res 11: 5245-5251.
28. Harrison RA, Lachmann PJ (1980) The physiological breakdown of the third component of human complement. Mol Immunol 17: 9-20.
29. Ganu VS, Müller-Eberhard HJ, Hugli TE (1989) Factor C3f is a spasmogenic fragment released from C3b by factors I and H: the heptadeca-peptide C3f was synthesized and characterized. Mol Immunol 26: 939-948.
30. De Ceuleneer M, De Wit V, Van Steendam K, Van Nieuwerburgh F, Tilleman K, et al. (2011) Modification of citrulline residues with 2,3-butanedione facilitates their detection by liquid chromatography/mass spectrometry. Rapid Commun Mass Spectrom 25: 1536-1542.
31. Andrade F, Darrah E, Gucek M, Cole RN, Rosen A, et al. (2010) Autocitrullination of human peptidyl arginine deiminase type 4 regulates protein citrullination during cell activation. Arthritis Rheum 62: 1630-1640.
32. Loos T, Mortier A, Gouwy M, Ronsse I, Put W, et al. (2008) Citrullination of CXCL10 and CXCL11 by peptidylarginine deiminase: a naturally occurring posttranslational modification of chemokines and new dimension of immunoregulation. Blood 112: 2648-2656.

33. Kubota K, Yoneyama-Takazawa T, Ichikawa K (2005) Determination of sites citrullinated by peptidylargininedeiminase using 18O stable isotope labeling and mass spectrometry. Rapid Commun Mass Spectrom 19: 683-688.

34. Wood DD, Moscarello MA (1989) The isolation, characterization, and lipid-aggregating properties of a citrulline containing myelin basic protein. J Biol Chem 264: 5121-5127.