Citrullinome of Porphyromonas gingivalis Outer Membrane Vesicles
Confident Identification of Citrullinated Peptides

Larsen, Daniel Nyberg; Mikkelsen, Christian Engelbrecht; Kierkegaard, Mads; Bereta, Grzegorz P; Nowakowska, Zuzanna; Kaczmarek, Jakub Z; Potempa, Jan; Højrup, Peter

Published in:
Molecular & cellular proteomics : MCP

DOI:
10.1074/mcp.RA119.001700

Publication date:
2020

Document version
Accepted manuscript

Citation for published version (APA):
Larsen, D. N., Mikkelsen, C. E., Kierkegaard, M., Bereta, G. P., Nowakowska, Z., Kaczmarek, J. Z., Potempa, J., & Højrup, P. (2020). Citrullinome of Porphyromonas gingivalis Outer Membrane Vesicles: Confident Identification of Citrullinated Peptides. Molecular & cellular proteomics : MCP, 19(1), 167-180. https://doi.org/10.1074/mcp.RA119.001700

Terms of use
This work is brought to you by the University of Southern Denmark through the SDU Research Portal. Unless otherwise specified it has been shared according to the terms for self-archiving. If no other license is stated, these terms apply:

• You may download this work for personal use only.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim. Please direct all enquiries to puresupport@bib.sdu.dk

Download date: 14. May, 2021
Citrullinome of *Porphyromonas gingivalis* Outer Membrane Vesicles: Confident Identification of Citrullinated Peptides

Daniel Nyberg Larsen¹, Christian Engelbrecht Mikkelsen¹, Mads Kierkegaard¹, Grzegorz P. Bereta², Zuzanna Nowakowska²,³, Jakub Kaczmarek⁴, Jan Potempa²,⁵, Peter Højrup¹,*

1) University of Southern Denmark, Campusvej 55, Odense M, Denmark

2) Department of Microbiology, Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Krakow, Poland

3) Malopolska Center of Biotechnology, Jagiellonian University, Krakow, Poland

4) Research and Development Department, Ovodan Biotech A/S, 5000 Odense, Denmark

5) Department of Oral Immunology and Infectious Diseases, University of Louisville School of Dentistry, 501 S. Preston St., Louisville, KY, USA

* Communicating author. E-mail: php@bmb.sdu.dk

**Running title:** Confident identification of citrullinated peptides.

**Abbreviations:** (OMV) outer membrane vesicles, (WT) *P. gingivalis* W83 wild-type, (ΔPPAD) W83 knock-out mutant of peptidyl arginine deiminase, (C351A) W83 expressing PPAD with the active site cysteine mutated to alanine, (2D) two-dimensional HFBA-based, (LC-MS) liquid chromatography mass spectrometry, (PAD) humane peptidyl arginine deiminase, (PPAD) bacterial *P. gingivalis* peptidyl arginine deiminase, (RA) rheumatoid arthritis, (PGE₂) prostaglandin E2, (ACP A) anti-citrullinated protein antibodies, (RgpA/B) Arg-specific gingipains, (MS) mass spectrometry, (HFBA) heptafluorobutyric acid, (FA) formic acid, (HPLC) high performance liquid chromatography, (AAA) amino acid analysis, (mgf) Mascot Generic File format, (mgx) Mascot Generic eXtended, (RT) retention time, (TIC) total ion count, (T9SS) type IX secretion system,
Abstract

*Porphyromonas gingivalis* is a key pathogen in chronic periodontitis and has recently been mechanistically linked to the development of rheumatoid arthritis via the activity of peptidyl arginine deiminase generating citrullinated epitopes in the periodontium. In this project the outer membrane vesicles (OMV) from *P. gingivalis* W83 wild-type (WT), a W83 knock-out mutant of peptidyl arginine deiminase (ΔPPAD), and a mutant strain expressing PPAD with the active site cysteine mutated to alanine (C351A), have been analyzed using a two-dimensional HFBA-based (2D) separation system combined with LC-MS. For optimal and positive identification and validation of citrullinated peptides and proteins, high resolution mass spectrometers and strict MS criteria were utilized. This may have compromised the total number of identified citrullination but increased the confidence of the validation. A new 2D separation system proved to increase the strength of validation, and along with the use of an in-house build program, *Citrullia*, we establish a fast and easy semi-automatic (manual) validation of citrullinated peptides. For the WT OMV we identified 78 citrullinated proteins having a total of 161 citrullination sites. Notably, in keeping with the mechanism of OMV formation, the majority (51 out of 78) of citrullinated proteins were predicted to be exported via the inner membrane and to reside in the periplasm or being translocated to the bacterial surface. Citrullinated surface proteins may contribute to the pathogenesis of rheumatoid arthritis. For the C351A-OMV a single citrullination site was found and no citrullinations were identified for the ΔPPAD-OMV, thus validating the unbiased character of our method of citrullinated peptide identification.
Introduction

Citrullination is a deimination of arginine, which results in the loss of a single nitrogen and hydrogen along with the addition of an oxygen, resulting in a mass shift of 0.984 Da and loss of a single charge. Citrullination is a post-translational modification that can only occur on arginine residues, either on the N- or C-terminal of the peptides or internally.

Citrullination occurs in physiological and pathological conditions and is thought to play a range of different functions. The human peptidyl arginine deiminases (PAD) 1, 2, 3, 4, and 6 exert different roles as a result of expression in different cellular environments. PAD1 citrullinates keratin and filaggrin, which is important for terminal differentiation of keratinocytes.\(^1,2\) PAD2 citrullinates myelin basic protein, which is involved in the myelin sheath formation.\(^3,4\) PAD3 is involved in hair growth by citrullination of trichohyalin.\(^5,6\) PAD4 has been assigned more adverse functions, from regulation of gene-expression\(^7\)–\(^11\), through immune modulation\(^12\), to auto-citrullination for regulation of citrullinated protein\(^13\). Finally, the PAD6 function is not completely understood due to lack of substrate, but it is thought to have a role in reproduction\(^14\).

Only few species of bacteria belonging to genus *Porphyromonas* have been found to express peptidyl arginine citrullinating enzyme closely related on the amino acid sequence level, if not identical.\(^15\) The best characterized is *Porphyromonas gingivalis* peptidyl arginine deiminase (PPAD). The sequence identity between PPAD and the human PADs is low, approximately 30%\(^16\), however, similar activity is preserved. There are some differences, however, and while human PADs activity is dependent on calcium and targets internal Arg residues, PPAD preferentially citrullinates C-terminal Arg in a calcium-independent manner. Apart from the preference for C-terminal Arg citrullination, specificity with respect to preceding residues has not been identified for PPAD. Conversely a study on PAD2 and PAD4 showed very broad specificity with PAD2 favoring Tyr in the +3 position (Assohou-Luy et al.\(^17\)).

Citrullinations have been found to contribute to the pathogenicity of various diseases including rheumatoid arthritis (RA) and periodontitis. *P. gingivalis*, while absent or at the low level in dental biofilm of
periodontally healthy subjects, occurs in high numbers in the mouth of periodontitis patients and is thought to be one of the primary causes of periodontitis.\textsuperscript{18–20} Apart from PPAD \textit{P. gingivalis} expresses Arg-specific gingipains, RgpA and RgpB\textsuperscript{16}, which are important enzymes for citrullination, as they generate peptides and protein fragments with C-terminal Arg, instant substrates for modification by PPAD\textsuperscript{21}. In line with concerted action of Rgps and PPAD the Rgp-null mutant shows very little citrullination in comparison to the parental \textit{P. gingivalis} W83 strain. Nevertheless, PPAD can modify internal Arg residues\textsuperscript{22,23} but this reaction occurs at the rate thousand times slower than citrullination of C-terminal Arg\textsuperscript{24}. This is in accordance with the topology of the substrate binding site perfectly shaped to accommodate Arg at the C-terminus with no room for an extended peptide chain.\textsuperscript{25}

Several findings implicate PPAD as an important virulence factor of \textit{P. gingivalis}. Through citrullination of C-terminal residues in epidermal growth factor\textsuperscript{23} and C5a anaphylatoxin\textsuperscript{26} the enzyme can contribute to the periodontal tissue damage and attenuation of innate immune responses, respectively. Furthermore, \textit{P. gingivalis}' citrullinome modulates neutrophil activity\textsuperscript{27}, constrains \textit{P. gingivalis} biofilm development\textsuperscript{28}, affects epithelial cells transcriptome\textsuperscript{29}, contributes to formation of dual-species biofilm with opportunistic fungus Candidia\textsuperscript{30}, and is responsible for stimulation of prostaglandin E2 (PGE\textsubscript{2}) secretion by gingival fibroblasts.\textsuperscript{31} The latter activity can be directly linked to periodontitis and RA pathogenicity as PGE\textsubscript{2} promotes bone resorption. Apart from that, C-terminally citrullinated peptides generated by concerted action of Rgp and PPAD are considered pivotal in breaking the immunotolerance leading to production of specific anti-citrullinated protein antibodies (ACPA) directly responsible for development of RA.\textsuperscript{32} This theory fits well, with the findings of heightened levels of ACPAs in periodontitis and RA patients.\textsuperscript{19,20,33–36} Likewise, it is supported by data from animal models of RA and \textit{P. gingivalis} infection in which RA severity is dependent on PPAD expression.\textsuperscript{32} In this context, determination of \textit{P. gingivalis} citrullinome is very important but a challenging task.
The detection of citrullinations started with a color development reagent assay (COLDER assay) according to Clancy et al., which depends on chemical derivatization of the urea group of citrulline. This method is mostly used in in vitro assays, due to its poor sensitivity and need of large amounts of citrullinations.

Antibody based methods also depend on chemical derivatization, the first anti-citrulline assay was developed by Senshu et al. These antibody-based assays have a major disadvantage as they have been found to cross-react with carbamylation, which is a chemical modification of lysine into homocitrulline. Currently the most promising technique for the identification of citrullinations with regard to sensitivity and specificity is mass spectrometry (MS).

The development during the last decades within the MS field of protein research has made the investigation of clinical samples and whole-cell lysates possible, particularly when fractionated prior to injection. Furthermore, less complex samples can be run directly, leading to significant reduction of manual work prior to analysis. The major problem with MS based methods for the single Dalton mass shift is the possibility of misinterpretation of a deamination of asparagine or glutamine or a wrong isotope picking, the loss of a single charge, giving rise to poor ionization and fragmentation, and small retention time shift. We addressed these problems in the present paper by the use of heptafluorobutyric acid (HFBA) as the ion-pairing reagent during two-dimensional fractionation, as well as optimized mass spectrometric data acquisition, and development of specific software, Citrullia, designed for identification and validation of citrullinations.
Experimental Procedures

Bacterial fraction preparation

Cultures of *P. gingivalis* strain W83 and its isogenic mutants; C351A (with a point mutation of the catalytic cysteine residue, C351A, in PPAD rendering the enzyme catalytically inactive) and ΔPPAD (with the *ppad* gene deleted) were maintained on TSB agar plates with 5% defibrinated sheep blood and supplements: yeast extract (5 mg/ml), L-cysteine (0.5 mg/ml), hemin (5 µg/ml), and menadione (1 µg/ml). Liquid cultures were inoculated from 5–6 days old plates to liquid TSB medium with supplements and cultured for 18–20 hours at 37 °C in an anaerobic chamber. Cultures were then diluted to OD$_{600}$ = 0.1 in fresh medium and cultured as before for 20–22 hours. Aldrithiol-4 (1.5 mM) was added to cultures immediately after incubation. Cultures were then centrifuged (7 500 rcf, 15 minutes, 4 °C), supernatant collected and filtered through 0.45 µm membrane filter. The filtrate was ultracentrifuged at 70,000 rcf for 2 hours at 4 °C. The collected sediment encompassing OMV was washed and then suspended in PBS with 1 mM TLCK by gentle sonication. The concentration of protein was determined using the Bradford method with bovine albumin as a standard.

Sample preparation

OMV were reduced by the addition of dithiothreitol (DTT) to a final concentration of 10 mM. The samples were then incubated for 30 min at 50-57°C, followed by alkylation with iodoacetamide (IAA) added to a final concentration of 24 mM and incubation in the dark for 20 min. Excess IAA was removed by treatment with DTT and proteins in the sample were digested by overnight incubation at 37°C with 2% w/w in-house methylated trypsin.

For complete analysis without fractionation, the samples were micro purified essentially as described, dried down and resuspended in 0.1% formic acid (FA).
For analysis by high performance liquid chromatography (HPLC) fractionation, each sample was dried down and resuspended in 0.05 % heptafluorobutyric acid (HFBA) prior to off-line separation.

Off-line HPLC separation

Samples of 100 µg in a volume of 50 µl 0.05 % HFBA were injected into an Agilent 1260 Infinity HPLC system equipped with a Phenomenex Aeris™ 2.6 µm PEPTIDE XB-C18 100 Å, LC Column 150 x 2.1 mm, Ea (00F-4505-AN) running at a flow of 0.200 ml/min. Buffer A = 0.05% HFBA, buffer B = 0.05% HFBA and 90% acetonitrile.

The method used a loading time of 2 min with 5% phase B and separated with a gradient of 12.5% B / 5 min; 50.0% B / 30 min, 60% B / 35 min, 90%B / 35.5 min, 90% B / 36.5 min, 5% B / 37 min, followed by equilibration at 5% B for 13 min.

Mass spectrometry

Samples were run on an EASY-nLC1000 Liquid Chromatography system (Thermo Fisher Scientific), using a 3 µm trap column (100 µm inner diameter, 5 µm Reprosilpur 120 C18, Dr. Maisch GmbH, Germany) and an 18 cm analytical column (75 µm inner diameter, 3µm Reprosilpur 120 C18) coupled online to a Q Exactive HF Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Fisher Scientific). The methods applied on the mass spectrometer had the following settings in common: positive mode, an MS1 resolution of 120,000, AGC target of 3e6, maximum injection time of 100 ms, and a scan rage of 300-1400 m/z. The common MS2 settings were: resolution of 30,000, AGC target of 1e6, isolation window of 0.8 m/z, and a fixed first mass of 110.0 m/z. Furthermore, peptides with charges ranging from +1 to +6 were included, while +7 and above were excluded along with isotopes.

For the duplicate experiment (30-minute gradient) the following specific parameters were used: dynamic exclusion of 5 seconds, maximum injection time of 200 ms, loop count of 5. For the triplicate experiment (30-minute gradient) the parameters were changed to: dynamic exclusion of 5 seconds, maximum injection
time of 100 ms, loop count of 10. For the triplicate experiment with long 120-minute gradient the parameters were: dynamic exclusion of 15 seconds, maximum injection time of 100 ms, loop count of 10.

**Amino acid analysis**

To determine protein amounts and composition, amino acid analysis (AAA) was applied essentially as described by Højrup 2015. Samples of 2-4 µg protein were dried in small polypropylene tubes, lids were punctured, and they were placed in 25 ml glass vials along with 200 µl of 6N HCl, 0.1 % phenol, 0.1 % 2-thioglycolic acid and closed with a MinInert valve (VICI) after being covered by argon and evacuated to <1 mBar pressure. After overnight hydrolysis at 110 °C, samples were dried, re-dissolved, and analyzed on a BioChrom 30+ amino acid analyzer using recommended conditions.

**Data handling**

Citrullination is the exchange of an amino group with oxygen resulting in a mass increase of 0.984 Da. This difference is readily identified by modern MS search engines, but as the mass difference is identical to the commonly occurring deamidation of asparagaine and glutamine and can further be mistaken by wrongly picked isotope in the ms1 spectrum, we developed a program for extraction and display of spectra identified as potentially containing citrullinated residues by the search engine.

In order to improve the validation of citrullinations and enable manual validation within a reasonable timeframe, we developed a program called Citrullia in C# version 7.2 for Windows. In addition, the .NET framework v. 4.7.2 was used in Visual Studio 2019 integrated development environment and the Metro Modern UI (Dennis Magno) was used for visual user interface elements. As search engine we chose the publicly available X! Tandem program (www.thegpm.org).

As we deemed it necessary also to validate that the correct parent ion isotope had been picked, we used a new mass file format based on the Mascot Generic File format (mgf) but extended with ms1 information. The format named mgx (Mascot Generic eXtended) has been developed by MassAI Bioinformatics
and the mgx conversion program **MGF Filter** can be freely downloaded from www.massai.dk/download.html. The mgx format contains the MS1 level information interspersed with the MS2 level data, but each spectrum is recognized by a new keyword “MSLEVEL=”. These files can easily be separated in ms1 and ms2 information for generating the standard mgf format files suitable for standard search engines.

X! Tandem was called with the following parameters: Parent ion precision: 10 ppm, ms2 precision: 0.02 Da, enzyme specificity: Cleavage after Lys and Arg, no cleavage before Pro, number of missed cleavages: <= 2. Fixed modifications: carbamidomethyl [C], variable modifications: deamidation [N, Q], oxidation [M], and citrullination [R]. The database searched was downloaded on 01.10.2018 from UniProt (www.uniprot.org) and contained all sequences from the *P. gingivalis* strains W83 and ATCC. The total number of proteins in the database was 4,417 sequences. The identified sequences were mapped to the PG locus numbers by comparing to the *P. gingivalis* W83 protein database from the Comprehensive Microbial Resource Web site (cmr.jcvi.org, June 2008).
Results

Citrullia

The citrullination of an arginine residue results in a mass change of +0.9840 Da. This change can be readily identified by modern mass spectrometers, but as pointed out by Küster and co-workers\textsuperscript{46} identification is fraught with danger of misinterpretation. As we thus anticipated that all potential citrullinated peptides had to be manually validated, we developed a program, Citrullia, for fast identification and easy manual validation. For fast and easy validation, the data needed is presented within a single window containing filename, retention time (RT), sequence, charge-state, e-value from the X! Tandem mass search, parent mass, accession number, table of determined ions, MS1, and MS2 spectra. It further provides information on the isotopic distribution of the MS1 peak, neutral losses, immonium ions, and an elution position for both the first- (HFBA) and the second-dimension (FA) separation. Furthermore, within this single window, each spectrum can be marked as validated. All the validated spectra are saved into a list, which can be extracted as one single list. This makes the validation a relatively straightforward process, as a researcher can quickly scroll through all potential candidates and mark them as validated or non-validated. Citrullia thereby ensures identification and semi-automatic validation of citrullinated peptides.

Raw files were first converted into mzXML using MS converter (part of the ProteoWizard package), peak picking was the only filter used, prior to conversion into the MGX file format. When loaded into Citrullia, the MS1 and MS2 data were separated and the MS2 data saved in a standard mgf file format. Each MS2 datafile was then searched individually using the X! Tandem search engine, with all parameters set by Citrullia. The 26-30 result files from a given experiment were loaded into Citrullia, and a multiple path search was performed (Fig 1). Initially peptides identified as citrullinated were extracted, and the entire X! Tandem run was searched for matching peptides with an MS1 mass difference of -0.984. If paired peptides were found, they were forwarded for validation. A second run was then performed using arginine-containing peptides, which were searched for potential citrulline-containing peptides using a mass
difference of +0.984. For validation, matching citrullinated and non-citrullinated peptides were displayed with one spectrum mirrored. This enables an easy comparison and visualization of mass shifts and differences in the fragmentation pattern. Validated peptides were then marked as such. Finally, non-matched citrullinated peptides were also validated individually but marked as *singles*. The two main criteria for validation were:

1) To establish that citrullinated arginine residues were delineated by fragment ions in the MS2 spectra in order to unambiguously distinguish it from potential deamidated asparagine and glutamine residues.

2) Verify that the correct monoisotopic ion was picked for parent ion fragmentation.

Three additional criteria were used for the validation:

1) The fragmentation pattern in agreement with the charge localization in the peptide, e.g. a C-terminally citrullinated tryptic peptide will not have a C-terminally located positive charge which results in a subsequent change in the fragmentation pattern (supplemental Fig. 4).

2) In reversed phase FA-based chromatography a citrullinated peptides will show delayed retention time relative to the non-citrullinated peptide.47

3) In reversed phase HFBA-based chromatography a citrullinated peptide will show the reversed (leading) retention time behavior.48

Citrullinations were thus evaluated both on MS1 specificity, fragment ions, fragmentation pattern, and retention time behavior in one or two dimensions. The user interface of *Citrullia* is presented in supplemental Fig. 1.
Identification of citrullination in the OMV of *P. gingivalis* W83 strains (WT, C351A, and ΔPPAD)

In order to improve the detection of citrullinated peptides, we decided to evaluate a two-dimensional chromatographic system. This was based on off-line separation of peptides using HFBA as modifier in the first dimension and an on-line FA-based system, both using C18 reversed phase column material. In a publication by Mant et al.\textsuperscript{48} it was reported that HFBA as a modifier bound strongly to positively charged residues, in particular arginines, causing arginine containing peptides to be retained longer in a reversed phase system compared to same peptides with citrulline(s). This was corroborated using synthetic peptides [results not shown]. We therefore decided to evaluate whether the HFBA modifier could be used in a standard two-dimensional proteomics setup and introduce a validation step for identification of citrullinated peptides. The first-dimension separation of *P. gingivalis* OMV tryptic peptides resulted in 26-30 fractions, each of which was subsequently analyzed using a 30-minute gradient in FA.

The isolated OMV from three different *P. gingivalis* strains were analyzed for citrullinations in technical duplicates. The preparations were from wild-type *P. gingivalis* W83 (WT), the W83 strain with a mutated PPAD, where the active cysteine was replaced with alanine (C351A), and a knockout of PPAD (ΔPPAD) in the W83 strain. Based on previous results on the *P. gingivalis* OMV secretome\textsuperscript{49} we expected WT to contain the majority of citrullinations, while fewer were expected for C351A and ΔPPAD. All samples were analyzed using our two-dimensional setup with HFBA as modifier in the first dimension and FA in the second dimension. The chromatogram and the total ion count (TIC) for all 30 fractions of the separation of WT replicate 1 are shown in Fig. 2. The orthogonality of the HFBA separation can be observed in the FA-based second dimension, as a broad elution pattern from 10 to 22 minutes was typical for the first 20 fractions from the first-dimension separation. The last 10 first-dimension fractions are shifted towards later elution in the second dimension. Separation of WT replicate 2, C351A replicate 1 and 2, and ΔPPAD replicate 1 and
revealed almost identical separation patterns (supplemental Fig.2A-E) indicating that the difference in content of citrullinated peptides was not sufficient for a measurable difference in the TIC.

The majority of the identified proteins and peptides were found in fractions collected between 12- and 30-minute elution from the first-dimension separation. The profile of identified peptides in each fraction was reproducible, and each fraction showed similar identification numbers across the different samples and replicates. Citrullinated peptides were found across the entire fractionation, but the majority of citrullinations were identified in fractions collected between 18 and 26 min. Only a few fractions showed large standard deviations in peptide or protein identification, which could be explained by lack of material or too low resuspension buffer volume (Fig. 3).

For the majority of fractions, the number of peptides and proteins identified in each OMV fraction was C531A >ΔPPAD> WT (Table 1). The number of unique peptides identified in C351A was approximately 15% higher than in ΔPPAD and twice that of WT. The number of identified unique proteins was almost identical for C351A and ΔPPAD, but a quarter less for WT (Fig. 4A). This may indicate that the 115 proteins determined in C351A are close to the total number measurable in the current dynamic range and MS settings. However, while the overlap between technical replicates was fairly high, the overlap of peptides between OMV derived from different strains was low, resulting in only 626 common peptides for all three strains (Fig. 4B). On the other hand, the overlap between identified proteins was relatively high (i.e. only 12-18 additional unique proteins identified per strain-fraction OMV), indicating that many of the unique identified proteins are of low abundance.

When searching for citrullinations, the vast majority was identified in the WT replicates, as only a single citrullination was identified in replicate 1, fraction 15 of the P. gingivalis PPAD$^{C351A}$ strain and none in ΔPPAD. Of the total of 52 unique citrullinations, 13 were found in pairs with the non-citrullinated peptide and the remainders were found as singles, i.e. the corresponding Arg-containing peptide could not be found. Based on the paired citrullinations, fraction shifts in the first dimension HFBA separation can be
calculated. Of the 13 paired citrullinations, 12 were found to elute in an earlier fraction and one was found to elute both earlier as well as later, depending on the replicate. The RT shift was calculated for the second dimension FA separation, where the average RT shift has been calculated to 50 seconds of delayed elution for a citrullinated peptide. This is in accordance with the previous observations.

**Citrullination of OMV derived from WT *P. gingivalis***

Based on the above presented results, we decided to analyze the wild type *P. gingivalis* W83 in more detail, in order to obtain the best characterization of the citrullinome and to evaluate our two-dimensional separation strategy. Using a single off-line HFBA separation of a tryptic digest of the WT resulted in 26 fractions, technical triplicate injections were performed on the mass spectrometer and was compared to a triplicate injection of the entire tryptic digest. For the off-line separation short 30-minute gradients were used, while a two-hour gradient was used when analyzing the complete mixture.

For the detailed analysis of the WT OMV, mass spectrometric data acquisition was optimized by decreasing the loading time (from 200 to 100 ms) and using a top 10 instead of a top 5 method, thus doubling the number of ions being fragmented and analyzed. For the long 120 minutes gradient, two microgram of WT OMV tryptic digest was micro-purified for each injection, as this was the estimated amount injected from the two-dimensional separation method over the central 20 fractions.

As C-terminally citrullinated peptides were expected mainly to be singly charged, these ion species were included for all sample runs. This inclusion had a greater impact for the long gradients where 15 out of 95 identified citrullinated peptides were singly charged. For the two-dimensional system the same occurred in only 6 out of 206 peptides (supplemental Table 3). The reason for the lower number of singly charged peptides in the 2D system is likely that the same peptide has a higher probability of being observed multiple times. If a peptide occurs as a doubly charged species, the resulting improved fragmentation would result in a higher e-value which would be more favorable, and thereby be selected for further analysis. This can also be seen from the observation that in the long gradient 44% of identified citrullinated
peptides were observed without a positively charged side-chain residue (primarily histidine, as tryptic digest takes place after lysine and arginine), which was only the case for 29% of the two-dimensional method.

For the triplicate HFBA fractionated samples, between 1312 and 1452 unique peptides were identified (Table 2). This is approximately four times the number of peptides identified in the first duplicate analysis. This resulted in the identification of 133 unique proteins, which is approximately 35% more than in the first experiment. While the number of identified citrullinated peptides also increased, the increase was approximately 48% compared to the number obtained in the first experiment. A total of 99 citrullinated peptides were identified for the triplicate, resulting in 56 proteins being positively identified as citrullinated. Of the citrullinated peptides, approximately half were identified along with the corresponding arginine-containing peptide (e.g. paired) while the rest were found as singles. The use of HFBA as modifier in the first dimension also showed the majority of the paired peptides (85) eluting in leading fractions with an average difference of 1.7 fractions. Eleven peptides eluted in the same and 2 in lagging fractions. As fractions were collected for either 5 (first and last fractions) or 1 min, an even better separation could probably be obtained by collecting smaller fraction (for shorter time windows) in the first dimension, at the expense of additional MS runs and less material per run. A single citrullinated peptide (AGNHTVQGATR) was found in both lagging and leading fractions, and twice in the same fraction, indicating that it may be sticky and elute over a large part of the chromatogram. In the second dimension, the opposite retention behavior of the citrulline/arginine peptides was observed, with an average separation time of approximately one minute using a 30-minute gradient and four minutes for the long gradient. This shows that both the first- and second-dimension RT can be used for validation, if both the citrullinated and non-citrullinated peptide are present and have been matched.

In the second dimension, the fractionation showed peptide elution across a wide time period. For the first 20 out of the 30 total fractions, the peptides eluted over a period of ≈12 min (from 10 to 22 min) on a 30
min gradient, while the elution for the last 10 fractions shifted towards a later start of peptide elution (17 min). Thereby the pooling of fractions, which is often performed when using HILIC\textsuperscript{57} or High pH\textsuperscript{56} is not possible, which increases the number of fractions to run in our HFBA system, but likely results in better separation of the sample. Furthermore, pooling of fractions would decrease the obtained resolution and obscure information on the exact fraction in which the peptide is eluting in the first dimension.

For the experiment using a long gradient, the total number of identified peptides and proteins was half of that observed in the two-dimensional separation (Table 2). However, the number of identified citrullinated peptides and proteins were very similar. A major difference was that only a small number of paired citrulline/arginine peptides were found. These were found with an average delayed elution of the citrulline peptide of 4 minutes in the 2h FA-based gradient. The difference in retention compared to the two-dimensional separation is due to the four times longer gradient.

Large differences in the number of observed proteins were observed between the WT OMV and the mutant strains when analyzed by the two-dimensional method. In total 115 proteins were identified for C351A and 112 for ΔPPAD, while only 88 were identified for the WT. This is likely caused by many WT citrullinated peptides being singly charged and thus more difficult to detect due to less fragmentation relative to multiply charged ions. As protein content and composition in the samples, except for the presence of PPAD, should be identical, the number of lysine terminated peptides should thus remain constant. However, the number of arginine/citrulline-terminated peptides may vary depending on the level of citrullination, hereby shifting the ratio. Although the number of identified peptides varied between the samples, the most striking difference is that the arginine/lysine terminated peptide ratio was very low in WT OMV (0.18) while almost identical for C351A (0.82) and ΔPPAD (0.79) (Table 3). Similar ratios were observed for the WT OMV long gradient triplicates (0.16) while the two-dimensional method increased the ratio to 0.28 (Table 4) showing that optimization of the data acquisition and the two-dimensional separation had improved the depth of analysis regarding citrullination.
Discussion

*P. gingivalis* secretes a very active peptidyl arginine deiminase (PPAD) along with Arg-specific gingipains (Rgps) using type IX secretion system (T9SS). During translocation across the outer membrane conserved C-terminal domains are cleaved off by sortase (PorU), and an anionic lipopolysaccharide is attached, anchoring the enzymes to the *P. gingivalis* cell surface.\(^{21,50}\) In this way PPAD and gingipains are in very close proximity since they are major components of the surface electron dense layer composed of circa 30 proteins secreted via T9SS.\(^ {51}\) Apparently in this environment Rgps generate C-terminal arginine residues on peptides and protein fragments, which are efficiently converted into citrulline by PPAD.\(^ {21}\) By budding of the outer membrane, the OMV coated with the surface electron dense layer are released into the environment carrying inside some periplasmic proteins.\(^ {52}\) Formation of the OMV is not a random process and is driven by a mechanism selectively sorting virulence factors into OMV excluding at the same time abundant outer membrane proteins.\(^ {53}\) In this way OMVs are very important for host-pathogen interactions extending the outreach of *P. gingivalis* virulence factors, including gingipains, PPAD, and citrullinated proteins into periodontal tissues. Therefore, taking into account pathogenic potency of citrullinated proteins it was important to develop a technique allowing unbiased identification of citrullination sites and delineation of the citrullinome of the *P. gingivalis* OMV.

Since PPAD mainly citrullinate C-terminal arginines\(^ {54}\), a large part of the citrullinated peptides in a tryptic digest were expected to be singly charged, due to the peptide size and citrullination. Therefore, we included singly-charged ions in our MS method and database searches. Such an approach increases the number of possible citrullination sites but is usually neglected due to potential false-positive issues and inclusion of a large number of non-peptide ions. In the most comprehensive study to date of the human citrullinome performed by Küster and coworkers\(^ {46}\), the authors excluded all potential C-terminal citrullinations and thereby ignored an estimated 10 - 48% of their data, which could potentially contain true positives.
In order to characterize the citrullinome at the *P. gingivalis* surface and the importance of PPAD, we analyzed proteins located in the OMV of the wild type W83 strain (WT), the mutant with the active cysteine mutated to alanine (C351A), and the knockout mutant (ΔPPAD). The proteomic analysis of the OMV derived from WT and mutants of the W83 using a two-dimensional strategy, identified a combined set of 88 to 115 proteins displaying a common core of 69 proteins (Table 1, Fig. 4). The *P. gingivalis* OMV proteome has previously been determined by Reynolds and co-workers\textsuperscript{51}, where they established a set of 151 proteins. Comparing our set of identified proteins to theirs, revealed only an approximately fifty percent overlap (supplemental Table 1). The reason for this can either be some differences in the purification of the OMVs as a bacterial growth phase, gel separation and slicing vs. 2D chromatography, or different settings during the database search. Setting our database search parameters similar to the ones presented in Veith et al.\textsuperscript{51} increased our identifications to >650 proteins, but only increased the overlap to approximately 60 percent.

In a more recent paper\textsuperscript{55} the same group using the same strain and preparation, but a slightly different gel and MS strategy, identified 181 proteins. Here 27 proteins from the earlier paper (18%) were not found, showing that the exact details for preparation and analysis are essential for comparison.

We have mapped the identified citrullinated proteins to their predicted subcellular location within the bacteria using the presence of signal peptide and the PG locus database. Three main groups were identified: lipoproteins (24%), T9SS secreted (28%), and others including cytoplasmic and inner membrane proteins (48%), (supplemental Fig. 4). Comparing our distribution of the citrullinated proteins to the complete OMV proteome of Veith et al.,\textsuperscript{55} revealed large differences particularly with regard to the cytoplasmic and inner membrane proteins. Whereas these proteins constituted only a few percent of recognized protein in the Veith et al. study they were found in much larger numbers in our experiment. Also, in contrast to abundance of the T9SS secreted proteins (almost 2/3 of identified proteins) and rarity of lipoproteins identified by Veith et al., in our study, we found an equal number of lipoproteins and T9SS cargo proteins. As outlined above, the differences may be caused by several analytical differences and is likely further skewed by comparing all found proteins against citrullinated proteins. The T9SS secreted
proteins are likely more resistant towards proteolysis and citrullination by Rgps and PPAD, respectively, while lipoproteins, cytoplasmic, and inner membrane proteins are more sensitive.

In another publication by van Dijl and coworkers\textsuperscript{49}, the authors aimed at determining the secreted citrullinome of \textit{P. gingivalis}. This study showed a similar number of identified proteins (64) as a core set, with a relatively large difference in the secretome among five different strains of \textit{P. gingivalis}. Note that the procedure used in this work did not exclusively purify the OMV, and apparently both soluble proteins and those associated with OMV were analyzed, which probably contributed to the observed differences among analyzed strains.

A striking difference between the data presented here and the data of van Dijl\textsuperscript{49} is the number of citrullinated proteins identified with high confidence which is 34 versus only 2, respectively. Furthermore, they found 11 citrullinated proteins in the ΔPPAD W83 strain, which originated in our laboratory. Although none of these peptides was identified with high confidence, finding such peptides in the PPAD-null strain undermines their approach of unbiased identification of citrullination. Apparently, the van Dijl’s group either could not distinguish Gln/Asn deamidations from Arg deimination or picked up a wrong isotope, or have other technical problems precluding confident identification of citrullinated peptides. For the C351A PPAD mutant we only identified a single citrullination (supplemental Fig.4), showing that the replacement of the cysteine in the active site may not completely remove all activity, but severely diminish it.

As the analysis of the various mutants revealed that the WT OMV was the only sample-type producing a significant number of citrullinations, we used this for for setting up a two-dimensional gradient and MS data acquisition method. Experiments comparing the HFBA-sed separation system with a standard high Ph chromatography showed similar separation power. Combined with the option of additional validation of citrullination, we decided to validate it by comparing to a triplicate analysis of the WT OMV in a 2-hour one-dimensional gradient. The total number of validated citrullinated peptides was 99 and 95 for the two- and one-dimensional method, respectively, resulting in the identification of 39 and 35 citrullinated proteins.
(Table 2). Although the total number of identified citrullinations does not vary much between the two methods, the confidence in identification of the citrullinations differed. A major difference was, however, that the two-dimensional system managed to identify twice as many proteins as the simple gradient. This difference in identification efficiency of citrullinated and normal peptides may be caused by citrullinated peptides ionizing as singly charged species, which need a higher ion count in order to fragment sufficiently for identification. Although the vast majority of identified citrullinated peptides were found as C-terminal citrullinations, we did identify a few peptides having internal citrullinations (spectra presented in supplemental Fig 4.1 and 4.2). As internally citrullinated peptides generally ionize as doubly charged ions with resulting higher detection efficiency, it supports the contention that PPAD has a strong to almost exclusive preference for C-terminal arginine residues.

While identified citrullinated peptides varied somewhat between technical duplicates, the number of identified unique proteins varied surprisingly little (Fig. 5 panel A to F). In contrast to this, the difference between both identified citrullinated peptides as well as citrullinated proteins varied more (Fig. 5 panel G and H). The variation could be due to the ion pairing reagent utilized (HFBA) in the first-dimension, where it enriches for specific peptides. However, the two-dimensional method doubled the number of identified peptides and proteins. Whether this increase was related to this particular two-dimensional system and could be reproduced by a high/low pH system or similar was not tested. Even though the number of identified citrullinations was similar between our two systems, our two-dimensional separation clearly shows potential advantages with regard to validation of citrullinations and the depth of analysis.

On the other hand, the number of identified peptides and proteins was much lower for WT OMV than for C351A and ΔPPAD. As the amount of sample was identical for all samples, the difference is likely caused by citrullination of the WT peptides resulting in a lower detection efficiency. This assumption was verified by measuring of the ratio between lysine terminated and arginine terminated peptides in the various samples as shown in Table 3 and 4. Taken together these results indicate that numerous P. gingivalis proteins were...
cleaved by RgpA/B, but almost all generated C-terminal arginines were modified by PPAD to citrulline. As expected, however, OMV-associated proteins were not degraded but rather a relatively small fraction of molecules was nicked at some sites, as SDS-gel electrophoresis of the various samples showed clear bands for intact proteins [results not shown].

The present results clearly show that limitations in the detection of C-terminally citrullinated peptides lie in the detection by mass spectrometry, as most of these peptides are ionized as singly charged species. A number of ways to alleviate this can be suggested. The most straightforward way would be to increase the net charge of the peptides, either at the N-terminus (e.g. by TMT labeling58), or at the C-terminus (e.g. using techniques from C-terminomics59). A purely MS-based method could be to increase the general charge by supercharging60 or use a faster MS instrument in combination with a decision tree for optimal data acquisition of singly charged species. A biochemical way of increasing the charge could be to digest with endopeptidase Lys-N which will generate lysine N-terminals at the expense of larger peptides, which may be more difficult to identify. An additional complexity arises when the sample has a low level of citrullination, which is often the case for mammalian citrullination analysis. For these analyses the sample has to be enriched. Here a reaction with diols61 is a possibility, particularly when the diol is coupled to a biotin group, which enables isolation with streptavidin62. This method has not been generally used, probably due to difficult synthesis of the reagent or problems with the identification. In this way there is ample room for improvements on different levels of the analysis, from MS method optimization to sample preparation.

Conclusion

The citrullinations identified and validated by our method are characterized by high overall confidence due to the manual validation steps and no exclusion of data prior to analysis. One drawback is the large amount of data that must be handled manually. This clear downside of the method was mitigated by use of the in-house developed program Citrullia, allowing analysis of hundreds to thousands of spectra in a fair amount
of time. Taken together, our results show that the citrullinome of *P. gingivalis* is much larger than the 6-25 proteins estimated earlier\(^4^9\). Here we have identified 78 proteins as being citrullinated and having a total of 161 citrullination sites. Nevertheless, our data also show that we most likely identified only a fraction of the actual number of citrullinated peptides present in *P. gingivalis* OMV. This under-detection of modified peptides is due to the removal of a positive charge by citrullination resulting in lower ionization potential and insufficient fragmentation. While our results are obtained without chemical derivatization, enrichment or other techniques, in order to obtain the full citrullinome of *P. gingivalis* new methods need to be implemented, where either a better way of analyzing singly charged peptide ions are applied or citrullinated peptides are derivatized resulting in multiply charged species. Despite these limitations this study reveals that at least 78 proteins of *P. gingivalis* OMV are citrullinated. Notably the clear majority (51 proteins) possessed a signal peptide targeting proteins to periplasm. Among these 17 are predicted to be lipoproteins and are 9 proteins secreted via T9SS and known to be enriched in the OMV. Apart from a few exceptions of internal or N-terminal citrulline, all identified citrullination sites were at the C-terminus. Interestingly however, only one of them was mapped to the native Arg occurring at the C-terminus of circa 200 proteins encoded in the *P. gingivalis* genome (Suppl. Table 3). This finding confirms our contention that Rgps and PPAD work in concert in generation of C-terminally citrullinated peptides derived from both bacteria and host proteins. Some of modified peptides/proteins apparently contribute to the autoimmune response leading to RA\(^3^2\), others stimulate the PGE\(_2\) synthesis pathway\(^2^1\), affect inter- and intra-microbial interactions in biofilm\(^2^7,^2^9\) and finally modify responses of neutrophils\(^2^7\) and gingival epithelial cells.\(^2^9\) Recognizing which citrullinated proteins/peptides are responsible for which activity is a challenge and the method described in this report is a first step in the direction of understanding the pathobiological meaning of the *P. gingivalis* citrullinome.
Acknowledgements

JP and GB were supported by NIH (NIDCR DE 022597 to JP) and National Science Center (Krakow, Poland): UMO-2018/30/A/NZ5/00650(to JP) and 2018/29/N/NZ1/00992 (to GB). Morten I Rasmussen is thanked for supplying software for the conversion of raw files to mgx format. Dr. Paul Veith is appreciated for providing the PG locus database.

Data Availability

The current version of Citrullia can be downloaded from https://github.com/Citrullia-SDU/Citrullia.

All the ms raw files used for searching are available on PRIDE with the accession number PXD014537.
References

1. Nachat R, Méchin MC, Takahara H, et al. Peptidylarginine deiminase isoforms 1-3 are expressed in the epidermis and involved in the deimination of K1 and filaggrin. J Invest Dermatol. 2005;124(2):384-393. doi:10.1111/j.0022-202X.2004.23568.x

2. Ishida-Yamamoto A, Senshu T, Eady RAJ, et al. Sequential reorganization of cornified cell keratin filaments involving filaggrin-mediated compaction and keratin 1 deimination. J Invest Dermatol. 2002;118(2):282-287. doi:10.1046/j.0022-202x.2001.01671.x

3. Lamensa JWE, Moscarello MA. Deimination of Human Myelin Basic Protein by a Peptidylarginine Deiminase from Bovine Brain. J Neurochem. 1993;61(3):987-996. doi:10.1111/j.1471-4159.1993.tb03612.x

4. Finch PR, Wood DD, Moscarello MA. The presence of citrulline in a myelin protein fraction. FEBS Lett. 1971;15(2):145-148. doi:10.1016/0014-5793(71)80043-1

5. Steinert PM, Parry DAD, Marekov LN. Trichohyalin mechanically strengthens the hair follicle: Multiple cross-bridging roles in the inner root sheath. J Biol Chem. 2003;278(42):41409-41419. doi:10.1074/jbc.M302037200

6. Tarcsa E, Marekov LN, Mei G, Melino G, Lee S-C, Steinert PM. Protein Unfolding by Peptidylarginine Deiminase. J Biol Chem. 1996;271(48):30709-30716. doi:10.1074/jbc.271.48.30709

7. Cuthbert GL, Daujat S, Snowden AW, et al. Histone deimination antagonizes arginine methylation. Cell. 2004;118(5):545-553. doi:10.1016/j.cell.2004.08.020

8. Wang Y, Wysocka J, Sayegh J, et al. Human PAD4 regulates histone arginine methylation levels via demethylimination. Science. 2004;306(5694):279-283. doi:10.1126/science.1101400

9. Li P, Yao H, Zhang Z, et al. Regulation of p53 target gene expression by peptidylarginine deiminase 4.
10. Yao H, Li P, Venters BJ, et al. Histone Arg modifications and p53 regulate the expression of OKL38, a mediator of apoptosis. J Biol Chem. 2008;283(29):20060-20068. doi:10.1074/jbc.M802940200

11. Zhang X, Gamble MJ, Stadler S, et al. Genome-Wide Analysis Reveals PADI4 Cooperates with Elk-1 to Activate c-Fos Expression in Breast Cancer Cells. PLOS Genet. 2011;7(6):1-15. doi:10.1371/journal.pgen.1002112

12. Proost P, Loos T, Mortier A, et al. Citrullination of CXCL8 by peptidylarginine deiminase alters receptor usage, prevents proteolysis, and dampens tissue inflammation. J Exp Med. 2008;205(9):2085-2097. doi:10.1084/jem.20080305

13. Andrade F, Darrah E, Gucek M, Cole RN, Rosen A, Zhu X. Autocitrullination of human peptidyl arginine deiminase type 4 regulates protein citrullination during cell activation. Arthritis Rheum. 2010;62(6):1630-1640. doi:10.1002/art.27439

14. Wright PW, Bolling LC, Calvert ME, et al. ePAD, an oocyte and early embryo-abundant peptidylarginine deiminase-like protein that localizes to egg cytoplasmic sheets. Dev Biol. 2003;256(1):73-88.

15. Gabarrini G, Chlebowicz MA, Vega Quiroz ME, et al. Conserved Citrullinating Exoenzymes in Porphyromonas Species. J Dent Res. 2017. doi:10.1177/0022034517747575

16. Mangat P, Wegner N, Venables PJ, Potempa J. Bacterial and human peptidylarginine deiminases: targets for inhibiting the autoimmune response in rheumatoid arthritis? Arthritis Res Ther. 2010;12(3):209. doi:10.1186/ar3000

17. Assohou-Luty C, Raijmakers R, Benckhuijsen WE, et al. The human peptidylarginine deiminases type 2 and type 4 have distinct substrate specificities. Biochim Biophys Acta - Proteins Proteomics. 2014;1844(4):829-836. doi:10.1016/j.bbapap.2014.02.019
18. Montgomery AB, Kopec J, Shrestha L, et al. Crystal structure of Porphyromonas gingivalis peptidylarginine deiminase: Implications for autoimmunity in rheumatoid arthritis. *Ann Rheum Dis.* 2016;75(6):1255-1261. doi:10.1136/annrheumdis-2015-207656

19. Reichert S, Schlumberger W, Dähnrich C, et al. Association of levels of antibodies against citrullinated cyclic peptides and citrullinated α-enolase in chronic and aggressive periodontitis as a risk factor of Rheumatoid arthritis: a case control study. *J Transl Med.* 2015:1-10. doi:10.1186/s12967-015-0625-7

20. Quirke A, Lugli EB, Wegner N, et al. Heightened immune response to autocitrullinated Porphyromonas gingivalis peptidylarginine deiminase: a potential mechanism for breaching immunologic tolerance in rheumatoid arthritis. *Ann Rheum Dis.* 2014;73(1):263-269. doi:10.1136/annrheumdis-2012-202726

21. Wegner N, Wait R, Sroka A, et al. Peptidylarginine deiminase from Porphyromonas gingivalis citrullinates human fibrinogen and alpha-enolase: implications for autoimmunity in rheumatoid arthritis. *Arthritis Rheum.* 2010;62(9):2662-2672. doi:10.1002/art.27552

22. Abdullah SN, Farmer EA, Spargo L, Logan R, Gully N. Porphyromonas gingivalis peptidylarginine deiminase substrate specificity. *Anaerobe.* 2013;23:102-108. doi:10.1016/j.anaerobe.2013.07.001

23. Pyrc K, Milewska A, Kantyka T, et al. Inactivation of Epidermal Growth Factor by Porphyromonas gingivalis. *Infect Immun.* 2013;81(1):55-64. doi:10.1128/IAI.00830-12

24. Goulas T, Mizgalska D, Garcia-ferrer I, et al. Structure and mechanism of a bacterial host-protein citrullinating virulence factor, Porphyromonas gingivalis peptidylarginine deiminase. *Sci Rep.* 2015;(June):1-17. doi:10.1038/srep11969

25. Bereta G, Goulas T, Madej M, et al. Structure, function, and inhibition of a genomic / clinical variant of Porphyromonas gingivalis peptidylarginine deiminase. *Protein Sci.* 2019;28(3):478-486.
26. Bielecka E, Scavenius C, Kantyka T, et al. Peptidyl arginine deiminase from Porphyromonas gingivalis abolishes anaphylatoxin C5a activity. *J Biol Chem.* 2014;289(47):32481-32487. doi:10.1074/jbc.C114.617142

27. Kriebel K, Hieke C, Engelmann R, et al. Porphyromonas gingivalis Peptidyl Arginine Deiminase Can Modulate Neutrophil Activity via Infection of Human Dental Stem Cells. *J Innate Immun.* 2018;10(4):264-278. doi:10.1159/000489020

28. Vermilyea DM, Ottenberg GK, Davey ME. Citrullination mediated by PPAD constrains biofilm formation in *P. gingivalis* strain 381. *NPJ biofilms microbiomes.* 2019;5:7. doi:10.1038/s41522-019-0081-x

29. Aliko A, Kaminska M, Bergum B, et al. Impact of Porphyromonas gingivalis Peptidylarginine Deiminase on Bacterial Biofilm Formation, Epithelial Cell Invasion, and Epithelial Cell Transcriptional Landscape. *Sci Rep.* 2018;8(1):14144. doi:10.1038/s41598-018-32603-y

30. Karkowska-Kuleta J, Bartnicka D, Zawrotniak M, et al. The activity of bacterial peptidylarginine deiminase is important during formation of dual-species biofilm by periodontal pathogen *Porphyromonas gingivalis* and opportunistic fungus *Candida albicans*. *Pathog Dis.* 2018;76(4). doi:10.1093/femspd/fty033

31. Gawron K, Bereta G, Nowakowska Z, et al. Peptidylarginine deiminase from *Porphyromonas gingivalis* contributes to infection of gingival fibroblasts and induction of prostaglandin E2 -signaling pathway. *Mol Oral Microbiol.* 2014;29(6):321-332. doi:10.1111/omi.12081

32. Potempa J, Mydel P, Koziel J. The case for periodontitis in the pathogenesis of rheumatoid arthritis. *Nat Rev Rheumatol.* 2017;13(10):606-620. doi:10.1038/nrrheum.2017.132

33. Lundberg K, Wegner N, Yucel-lindberg T, Venables PJ. Periodontitis in RA—the citrullinated enolase
Laugisch O, Wong A, Sroka A, et al. Citrullination in the periodontium - a possible link between periodontitis and rheumatoid arthritis. *Clin Oral Investig.* 2016;20(4):675-683. doi:10.1007/s00784-015-1556-7

Haag S, Schneider N, Mason DE, et al. Identification of new citrulline-specific autoantibodies, which bind to human arthritic cartilage, by mass spectrometric analysis of citrullinated type ii collagen. *Arthritis Rheumatol.* 2014;66(6):1440-1449. doi:10.1002/art.38383

Pratesi F, Panza F, Paolini I, et al. Fingerprinting of anti-citrullinated protein antibodies (ACPA): Specificity, isotypes and subclasses. *Lupus.* 2015;24(4-5):433-441. doi:10.1177/0961203314560005

Clancy KW, Weerapana E, Thompson PR. ScienceDirect Detection and identification of protein citrullination in complex biological systems. *Curr Opin Chem Biol.* 2016;30:1-6. doi:10.1016/j.cbpa.2015.10.014

Knipp M, Vasák M. A Colorimetric 96-Well Microtiter Plate Assay for the Determination of Enzymatically Formed Citrulline. *Anal Biochem.* 2000;264:257-264. doi:10.1006/abio.2000.4805

Senshu T, Sato T, Inoue T, Akiyama K, Asaga H. Detection of citrulline residues in deiminated proteins on polyvinylidene difluoride membrane. *Anal Biochem.* 1992;203(1):94-100. doi:10.1016/0003-2697(92)90047-B

Verheul MK, van Veelen PA, van Delft MAM, et al. Pitfalls in the detection of citrullination and carbamylation. *Autoimmun Rev.* 2017;17:136-141. doi:10.1016/j.autrev.2017.11.017

Heissel S, Frederiksen SJ, Id JB, Peter H. Enhanced trypsin on a budget : Stabilization , purification and high-temperature application of inexpensive commercial trypsin for proteomics applications. 2019;(Cid):1-16.
42. Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc*. 2007;2:1896. https://doi.org/10.1038/nprot.2007.261.

43. Højrup P. Analysis of Peptides and Conjugates by Amino Acid Analysis. In: Houen G. (Eds) Peptide Antibodies. *Methods in Molecular Biology*. 1348th ed. Humana Press, New York, NY; 2015. doi:https://doi.org/10.1007/978-1-4939-2999-3_8

44. Magno D. Metro Modern UI. https://www.nuget.org/packages/MetroModernUI/.

45. Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics*. 2004;20(9):1466-1467. doi:10.1093/bioinformatics/bth092

46. Lee C, Wang D, Wilhelm M, et al. Mining the Human Tissue Proteome for Protein. *Mol Cell Proteomics*. 2018:1378-1391. doi:10.1074/mcp.RA118.000696

47. Fert-bober J, Venkatraman V, Hunter CL, et al. Mapping Citrullinated Sites in Multiple Organs of Mice Using Hypercitrullinated Library. *J Proteom Res*. 2019;18:2270-2278. doi:10.1021/acs.jproteome.9b00118

48. Mant CT, Hodges RS. Context-dependent effects on the hydrophilicity/hydrophobicity of side-chains during reversed-phase high-performance liquid chromatography: Implications for prediction of peptide retention behaviour. *J Chromatogr A*. 2006;1125(2):211-219. doi:10.1016/j.chroma.2006.05.063

49. Stobernack T, Glasner C, Junker S, et al. Extracellular Proteome and Citrullinome of the Oral Pathogen Porphyromonas gingivalis. *J Proteome Res*. 2;15(12):4532-4543. doi:10.1021/acs.jproteome.6b00634

50. Gorasia DG, Veith PD, Chen D, et al. Porphyromonas gingivalis Type IX Secretion Substrates Are Cleaved and Modified by a Sortase-Like Mechanism. *PLoS Pathog*. 2015;11(9):e1005152.
51. Veith PD, Chen Y, Gorasia DG, et al. Porphyromonas gingivalis Outer Membrane Vesicles Exclusively Contain Outer Membrane and Periplasmic Proteins and Carry a Cargo Enriched with Virulence Factors. *J Proteome Res*. 2014;2;13(5):2420-2432. doi:10.1021/pr401227e

52. Chen Y-Y, Peng B, Yang Q, et al. The outer membrane protein LptO is essential for the O-deacylation of LPS and the co-ordinated secretion and attachment of A-LPS and CTD proteins in Porphyromonas gingivalis. *Mol Microbiol*. 2011;79(5):1380-1401. doi:10.1111/j.1365-2958.2010.07530.x

53. Haurat MF, Aduse-Opoku J, Rangarajan M, et al. Selective sorting of cargo proteins into bacterial membrane vesicles. *J Biol Chem*. 2011;286(2):1269-1276. doi:10.1074/jbc.M110.185744

54. Goulas T, Mizgalska D, Garcia-ferrer I, et al. Structure and mechanism of a bacterial host-protein citrullinating virulence factor , Porphyromonas gingivalis peptidylarginine deiminase. *Nat Publ Gr*. 2015;[April]:1-17. doi:10.1038/srep11969

55. Veith PD, Luong C, Tan KH, Dashper SG, Reynolds EC. Outer Membrane Vesicle Proteome of Porphyromonas gingivalis Is Differentially Modulated Relative to the Outer Membrane in Response to Heme Availability. *J Proteome Res*. 2018;17:2377-2389. doi:10.1021/acs.jproteome.8b00153

56. Batth TS, Olsen J V. Offline High pH Reversed-Phase Peptide Fractionation for Deep Phosphoproteome Coverage. In: von Stechow L, ed. *Phospho-Proteomics: Methods and Protocols*. New York, NY: Springer New York; 2016:179-192. doi:10.1007/978-1-4939-3049-4_12

57. Alpert AJ. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J Chromatogr*. 1990;499:177-196. doi:10.1016/S0021-9673(00)96972-3

58. Hainard A, Licker V, Turck N, et al. Relative Quantification of Proteins in Human Cerebrospinal Fluids by MS / MS Using 6-Plex Isobaric Tags. *Anal Chem*. 2008;80(8):2921-2931. doi:10.1021/ac702422x
59. Tanco S, Gevaert K, Van Damme P. C-terminomics: Targeted analysis of natural and posttranslationally modified protein and peptide C-termini. *Proteomics*. 2015;15(5-6):903-914. doi:10.1002/pmic.201400301

60. Kjeldsen F, Giessing AMB, Ingrell CR, Jensen ON. Peptide Sequencing and Characterization of Post-Translational Modifications by Enhanced Ion-Charging and Liquid Chromatography Electron-Transfer Dissociation Tandem Mass Spectrometry. *Anal Chem.* 2007;79(24):9243-9252. doi:10.1021/ac701700g

61. 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(17):2754-2762. doi:10.1002/rcm

62. Tutturren AE V, Holm A, Fleckenstein B. Specific biotinylation and sensitive enrichment of citrullinated peptides. *Anal Bioanal Chem*. 2013;405(29):9321-9331. doi:10.1007/s00216-013-7376-1
Tables

**Table 1.** Replicates of outer membrane vesicles (OMV) were analyzed for wild-type (WT) *P. gingivalis* the isogenic mutant in the same strain expressing inactive PPAD (PPADC351A), and in the knockout (ΔPPAD) W83 strain. Each replicate is represented with the combined results of unique proteins and the total number of peptides. For each replicate and sample, number of proteins, peptides, proteins with citrullination, identified citrullinated peptides, paired citrullinations (peptide identified with both arginine and citrulline), and single citrullinations (citrullinated only) are presented. For samples including paired citrullinations, fraction shift and retention time shift are shown.

| Sample                  | OMV W83 WT | OMV W83 C351A | OMV W83 ΔPPAD | Found in all |
|-------------------------|------------|---------------|---------------|--------------|
|                         | Rep. 1     | Rep. 2        | Comb.         | Rep. 1       | Rep. 2       | Comb.         | Rep. 1      | Rep. 2       | Comb. | 3/3 |
| Identified unique proteins | 77         | 79            | 88            | 113          | 107          | 115           | 93          | 101          | 112   | 69  |
| Identified unique peptides | 912        | 886           | 1073          | 1969         | 1760         | 2185          | 1549        | 1514         | 1879  | 626 |
| Citrullinated proteins  | 30         | 24            | 34            | 1            | 0            | 1             | 0           | 0            | 0     | 0   |
| Identified citrullinated peptides | 43         | 34            | 52            | 1            | 0            | 1             | 0           | 0            | 0     | 0   |
| Paired citrullinations  | 9          | 8             | 13            | 0            | 0            | 0             | 0           | 0            | 0     | 0   |
| Single citrullinations  | 34         | 26            | 39            | 1            | 0            | 1             | 0           | 0            | 0     | 0   |
| ±Cit earlier fraction   | 8          | 8             | 13            | -            | -            | -             | -           | -            | -     | -   |
| ±Cit same fraction      | 0          | 0             | 0             | -            | -            | -             | -           | -            | -     | -   |
| ±Cit later fraction     | 1          | 0             | 1             | -            | -            | -             | -           | -            | -     | -   |
| RT shift from Arg to Cit (Sec) | 49.1     | 50.3          | 49.7          | -            | -            | -             | -           | -            | -     | -   |

**Table 2.** Number of proteins and peptides of the WT-OMV identified in a standard two-hour LC-MS separation and an off-line HFBA separation into 30 fractions combined with 30-minute LC-MS separations of
each fraction. Each LC-MS separation was performed in triplicates. In addition, the number of citrullinated proteins and peptides, as well as the number of paired (both native and citrullinated peptide identified) and single (only citrullinated peptide) are shown. For the HFBA fractionated sample, the number of citrullinated peptides found in a leading, in the same, and in lagging HPLC fraction is noted along with the fraction shift.

| Sample       | Long Gradient | Fractionated | Common for both methods |
|--------------|---------------|--------------|-------------------------|
|              | WT_T 1 | WT_T 2 | WT_T 3 | Unique 1 | WT_T 2 | WT_T 3 | Unique 1 |                     |
| # Unique proteins identified | 61 | 61 | 62 | 68 | 119 | 119 | 121 | 133 | 62 |
| # Unique peptides identified | 693 | 668 | 634 | 849 | 1452 | 1312 | 1371 | 1757 | 522 |
| # Citrullinated proteins | 38 | 36 | 39 | 48 | 44 | 39 | 44 | 56 | 31 |
| # Citrullinations | 69 | 73 | 65 | 95 | 74 | 60 | 72 | 99 | 46 |
| # Paired | 3 | 2 | 2 | 4 | 36 | 27 | 35 | 45 | 0 |
| # Single | 66 | 71 | 63 | 92 | 38 | 33 | 37 | 59 | 26 |
| ±Cit leading fraction | - | - | - | - | 31 | 24 | 30 | - | - |
| ±Cit same fraction | - | - | - | - | 4 | 3 | 4 | - | - |
| ±Cit lagging fraction | - | - | - | - | 1 | 0 | 1 | - | - |
| Fraction shift | - | - | - | - | -1.67 | -1.81 | -1.69 | - | - |
| RT shift (sec) | 254 | 230 | 288 | 257,3 | 75 | 56.6 | 63.2 | 64.9 | - |

**Table 3.** The number of peptides including a C-terminal arginine or lysine residue found in the different replicates, along with the ratio of the two. Clear difference is seen from WT towards Mutant (C351A) and ΔPPAD.
| Sample       | #Arginines | #Lysines | R/K  |
|--------------|------------|----------|------|
| WT_2         | 314        | 1672     | 0.188|
| WT_3         | 287        | 1625     | 0.177|
| C351A_2      | 2140       | 2598     | 0.824|
| C351A_3      | 1881       | 2314     | 0.813|
| dPPAD_1      | 1368       | 1679     | 0.815|
| dPPAD_2      | 1321       | 1707     | 0.774|

Table 4. The number of peptides including a C-terminal arginine or lysine along with their ratios as determined in each replicate using the long gradient or the two-dimensional HFBA-based method

| Sample                  | #Arginines | #Lysines | R/K  |
|-------------------------|------------|----------|------|
| WT_1_Long               | 129        | 817      | 0.158|
| WT_2_Long               | 125        | 786      | 0.159|
| WT_3_Long               | 118        | 721      | 0.164|
| WT_1_Two-dimensional_method | 834    | 2971     | 0.281|
| WT_2_Two-dimensional_method | 716    | 2578     | 0.278|
| WT_3_Two-dimensional_method | 782    | 2707     | 0.289|
Figures and Figure Legends

**Figure 1.** Flowchart of data handling in Citrullia. MS2 data are initially analyzed by the X! Tandem search engine. Peptides identified as citrullinated are then paired with all potential arginine-containing peptides for validation. All not-identified arginine peptides are then compared to potential citrullinated peptides for validation. Finally, all citrullinated peptides that have not found a match to an arginine peptide are forwarded for validation. All validated citrullinations are then saved in a table with relevant information.

**Figure 2.** Two-dimensional separation of a tryptic digest of P. gingivalis OMV proteins. Left: Second-dimension total ion count traces of the 30 fractions resulting from the first dimension HFBA fractionation. The first set of fractions was obtained from 5 to 10 min elution time. From 10 to 38 minutes, fractions were collected in 1 min intervals. The last set was collected from 38 to 43 minutes elution time. Each first-dimension fraction is shown from 10 to 30 min elution. Right: First dimension UV trace (214nm) of HFBA/acetonitrile separation of 100 µg WT P. gingivalis OMV.

**Figure 3.** Distribution of identified peptides and proteins from the first-dimension separation of the three P. gingivalis strains. A) Number of peptides identified per fraction. B) Number of proteins identified per fraction. C) number of citrullinated peptides identified per fraction, were each sample duplicate is treated as a single sample. Fraction 1 and fraction 30 have been collected for 5 min, while fractions from 2 through 29 were collected for 1 min each during the HPLC gradient development. The error bars indicate the variation between duplicates.

**Figure 4.** Venn diagrams of the overlap of peptides and proteins between the different OMV isolates. A) Overlap of the combined unique peptides and B) overlap of the combined unique proteins (Table 1).

**Figure 5.** Venn diagrams, showing overlap between three replicates for the long gradient: A) unique proteins identified, B) unique citrullinated proteins, and C) unique citrullinated peptides. Three replicates for the fractionated samples: E) unique proteins identified, F) unique citrullinated proteins, and G) unique
citrullinated peptides. Furthermore, comparison between the long gradient and the fractionated sample H) unique citrullinated proteins and I) unique citrullinated peptides.

FIGURE 1

![Diagram of MS1 and MS2 data processing流程图](image1)

FIGURE 2

![Heatmap of MS1 and MS2 data processing](image2)
FIGURE 4

Unique Peptides Identified

C351A 601
ΔPPAD 395
626 59
799

Unique Proteins Identified

C351A 601
ΔPPAD 395
18 69

C351A 601
ΔPPAD 395
69

WT 229
159
626

WT 12
5
23
18

WT 2
18
12
Figure 5

A Unique Proteins Identified
B Long Gradient Citrullinated Proteins
C Citrullinated Peptides

E Unique Proteins Identified
F Fractionated Citrullinated Proteins
G Citrullinated Peptides

Long Gradient Vs. Fractionated
H Citrullinated Proteins
I Citrullinated Peptides