

mMass as a Software Tool for the Annotation of Cyclic Peptide Tandem Mass Spectra

Timo H. J. Niedermeyer^{1,2*}, Martin Strohal³

1 Institute of Pharmacy, Ernst-Moritz-Arndt University, Greifswald, Germany, **2** Cyano Biotech GmbH, Berlin, Germany, **3** Institute of Microbiology, v.v.i., Academy of Sciences of the Czech Republic, Czech Republic

Abstract

Natural or synthetic cyclic peptides often possess pronounced bioactivity. Their mass spectrometric characterization is difficult due to the predominant occurrence of non-proteinogenic monomers and the complex fragmentation patterns observed. Even though several software tools for cyclic peptide tandem mass spectra annotation have been published, these tools are still unable to annotate a majority of the signals observed in experimentally obtained mass spectra. They are thus not suitable for extensive mass spectrometric characterization of these compounds. This lack of advanced and user-friendly software tools has motivated us to extend the fragmentation module of a freely available open-source software, mMass (<http://www.mmass.org>), to allow for cyclic peptide tandem mass spectra annotation and interpretation. The resulting software has been tested on several cyanobacterial and other naturally occurring peptides. It has been found to be superior to other currently available tools concerning both usability and annotation extensiveness. Thus it is highly useful for accelerating the structure confirmation and elucidation of cyclic as well as linear peptides and depsipeptides.

Citation: Niedermeyer THJ, Strohal M (2012) mMass as a Software Tool for the Annotation of Cyclic Peptide Tandem Mass Spectra. PLoS ONE 7(9): e44913. doi:10.1371/journal.pone.0044913

Editor: John Matthew Koomen, Moffitt Cancer Center, United States of America

Received: May 15, 2012; **Accepted:** August 9, 2012; **Published:** September 13, 2012

Copyright: © 2012 Niedermeyer, Strohal. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the Czech Science Foundation (P206/12/1150), the Ministry of Education, Youth and Sports of the Czech Republic (ME10013). The funders had no role in study design, data collection and analysis, or preparation of the manuscript. No additional external funding received for this study.

Competing Interests: Timo Niedermeyer is an employee of Cyano Biotech GmbH. However, the described software has not been developed at Cyano Biotech, and the company does not have any financial interests in it. There are no patents, products in development or marketed products to declare. This does not alter the authors' adherence to all the PLoS ONE policies on sharing data and materials, as detailed online in the guide for authors. Cyano Biotech GmbH has given TN the permission to publish the manuscript.

* E-mail: timo.niedermeyer@uni-greifswald.de

Introduction

Cyclic peptides are a group of natural products that attracts the interest of a large number of researchers due to their intriguing structures and powerful and diverse bioactivities. Cyclic peptides are found in many bacterial and fungal strains [1,2]. Especially cyanobacteria have long been known for the wealth of cyclic peptides and cyclic depsipeptides they produce [3–6]. These compounds can be of ribosomal origin (e.g. the microviridins found in *Microcystis* strains) [7,8], but predominantly they are synthesized by non-ribosomal peptide synthetases (NRPS) [9–12]. Non-ribosomal peptides are fascinating for the natural product chemist, as they are not only composed of proteinogenic amino acids but also contain unusual and highly modified amino acids or monomers constructed by polyketide synthases (PKS). These compounds often derive from a mixed PKS/NRPS biosynthetic pathway as described in great detail e.g. for the cyanobacterial microcystins [13–15]. Opposed to 20 proteinogenic amino acids, over 500 different monomers can be compiled from known non-ribosomal peptides and are publicly available in the NORINE database [2,16]. Still, additional monomers are being described in the literature.

Among the most prominent examples of pharmaceutically exploited non-ribosomal cyclic peptides are cyclosporine, an immunosuppressant from *Tolypocladium inflatum* [17,18], and vancomycin, an antibiotic from *Amycolatopsis orientalis* [19,20]. Several other compounds or compound mixtures are either used

directly (e.g. tyrotricin, daptomycin and actinomycin D) or serve as lead structures for current drug development (e.g. the cryptophycins) [21].

The structural characterization of newly isolated or synthesized bioactive cyclic peptides is the basis for their further exploitation. Mass spectrometry often serves as a valuable tool for the first steps in dereplication, structure confirmation and elucidation of cyclic peptides, when limited amount of material complicates direct characterization by NMR spectroscopy.

To validate proposed structures, theoretical *in-silico* fragmentation is commonly used for comparison with experimental data. Surveying freely available software that can be used for *in-silico* fragmentation of cyclic peptides and the annotation of their tandem mass spectra, we have found that despite the general interest in these compounds, currently available bioinformatic tools do not allow convenient and effective assignment of cyclic peptide tandem mass spectra. Driven by the needs of proteomics scientists, many software tools for *in-silico* fragmentation and annotation of linear peptides have been developed based on the observed gas phase fragmentation behavior of peptides [22–26] and theoretical aspects of fragmentation calculation [27–31]. These tools, however, are not necessarily suited for the natural product or synthetic chemist working with cyclic peptides.

Most of the available tools with a graphical user interface (GUI; e.g. ProteinProspector [32], InSilicoSpectro [33,34], Proteomics Toolkit [35], PeptideART [36], mMass [37]) are suited for the *in-*

silico fragmentation and annotation of linear peptides but do not offer the same functionality for cyclic peptides. In addition, mainly the proteinogenic amino acids can be used to assemble a peptide sequence, and support for unusual monomers is very limited. Although some modifications can be applied to the selectable amino acids, this feature is typically restricted to commonly found modifications and does not reflect the wealth of the non-proteinogenic monomers encountered in non-ribosomal peptides. Last but not least, several of the programs mentioned above are web-based tools, showing innate disadvantages as discussed below.

A wealth of knowledge has been accumulated on fragmentation pathways of linear peptides [22–26]. Although not equally well understood, the fragmentation of cyclic peptides has also been reasonably well studied and will be described in more detail below. Translation of cyclic peptide fragmentation pathways into a software algorithm should therefore be feasible. To the best of our knowledge, only three software tools capable of *in-silico* fragmentation of cyclic peptides are available for free today.

massXpert2 is a powerful software for polymer processing [38], and suitable for the *in-silico* fragmentation of cyclic peptides, as well. However, its feature-rich user interface can be discouraging for a novice user. In addition, the software requires deep understanding of the fragmentation mechanisms, since the user needs to manually define possible fragmentation pathways.

PFIA was the first software specifically designed for *in-silico* fragmentation of cyclic peptides [39]. It is an open-source software that can be used as an off-line tool. The software is easy to install and the user interface is plain and intuitive. Peptide sequences can be entered and saved. Amino acids can be modified with a restricted set of predefined modifications. A major drawback is that – in addition to only 12 predefined non-proteinogenic amino acids – at most four custom amino acids can be added by the user. This restricts the usability for non-ribosomal peptide analysis, as in many cyclic peptides of this type, more than four unusual amino acids are found. Another disadvantage of the software is that only a-series, b-series and immonium ions can be calculated, while neutral losses and other fragmentation series are not taken into account. Lastly, experimental data cannot be imported, so that automatic matching of calculated fragments to acquired data is not possible.

The most recent software tool, MS-CPA [40], has been refined since its initial publication in 2009, and been renamed to NRP-Annotation (P.A. Pevzner, personnel communication on Sept. 07, 2011). For this program no local installation is possible but it is an exclusively web-based tool, which poses several disadvantages for the user. The program accessibility cannot be guaranteed at all times, as it requires an active Internet connection and a working server hosting the program. The user cannot rely on reproducibility as the software version can be changed by the maintainer without any notice, and previous versions of the program are lost for the user (e.g. the published version of MS-CPA is not available anymore). Raw data of the experiments have to be uploaded to the server, which may hinder scientists to analyze more sensitive data they do not wish to upload to a third party. NRP-Annotation allows for free input of monomer masses to compose a sequence, which enables all imaginable monomers to be entered. However, as the program is a web-based tool, the monomer data has to be entered for each and every use anew. It is not possible to save and reuse entered monomers or sequences for future sessions. This is most inconvenient if e.g. fragmentation patterns for structurally related compounds need to be calculated and all monomer masses of all structures have to be entered individually and repeatedly. The program is not open-source, thus it cannot be altered to suit individual needs. A mass spectrum must be submitted for the calculation, however, only files in the .dta format are accepted,

whereas the previous version, MS-CPA, accepted the.mzXML format as well. Moreover, processing of uploaded mass spectra cannot be influenced by the user. After initial processing, calculated fragments are matched to the peaks observed in the spectrum. This results in some statistical data, an annotated spectrum, a list of peaks and matches, and an error plot as output. a-Series and b-series as well as sequence scrambling reactions, neutral losses of water and ammonia, and carbon monoxide adducts are taken into account. It might be confusing that in all reports provided by the software, the mass values of uncharged fragments (i.e. without the mass of any additional charging protons) are used instead of the measured *m/z* values.

Thus all available software solutions had significant drawbacks and could not be used efficiently for our purpose. The following features have been established to be necessary for the desired annotation tool:

- off-line tool running on multiple operating systems,
- spectrum manipulation and processing abilities (e.g. support for various data formats, baseline correction, peak picking etc.),
- ability to save spectra, sequences and interpretation results for future use,
- monomer database and editor for easy monomer and sequence management,
- inclusion of as many of the known cyclic peptide fragmentation pathways as possible,
- customizable neutral losses allowing for inclusion of side-chain losses,
- allowance of multiple neutral losses from one fragment,
- matching to and annotation of experimental data.

Materials and Methods

Compounds and Sample Preparation

Microcystin LF, microcystin LR, cryptophycin-1, and microginin FR1, compounds isolated from cyanobacteria, were received from the pure natural products library of Cyano Biotech GmbH, Berlin, Germany. The compounds were dissolved to a concentration of 5 µg/ml in 50:50 methanol:water. Experimental data for seglitide and cyclomarin A were downloaded from the NRP webpage (<http://bix.ucsd.edu/nrp/>; IT (auto) data for seglitide, TOF data for cyclomarin A) on September 15, 2011. The structures of the compounds are shown in figure 1.

Liquid Chromatography – Mass Spectrometry

MS data have been acquired using an HPLC coupled to an IT-TOF mass spectrometer (Shimadzu Europe GmbH, Duisburg, Germany) using electrospray ionization in positive mode and were evaluated using the vendor's software LCMSsolution version 3.60.361 with Formula Predictor version 1.13. The compounds were separated on a Kinetex C₁₈ column (2.6 µm, 100×3 mm, phenomenex, Torrance, USA) using a gradient ranging from 5 to 80% acetonitrile in water over 25 min (0.1% formic acid added as modifier). Precursor ions corresponding to [M+H]⁺ were isolated in the ion trap, fragmented by collision induced dissociation (CID) using argon as collision gas (collision energy set to 150%, collision gas to 100%, and q(Frequency) to 45.0 kHz), and separated in the TOF analyzer. MS/MS scans were averaged and converted to the mzXML format using the vendor's software. For the calculation of sum formulae, the monoisotopic mass averaged from at least three scans has been used. The Δ values indicated in the text are the

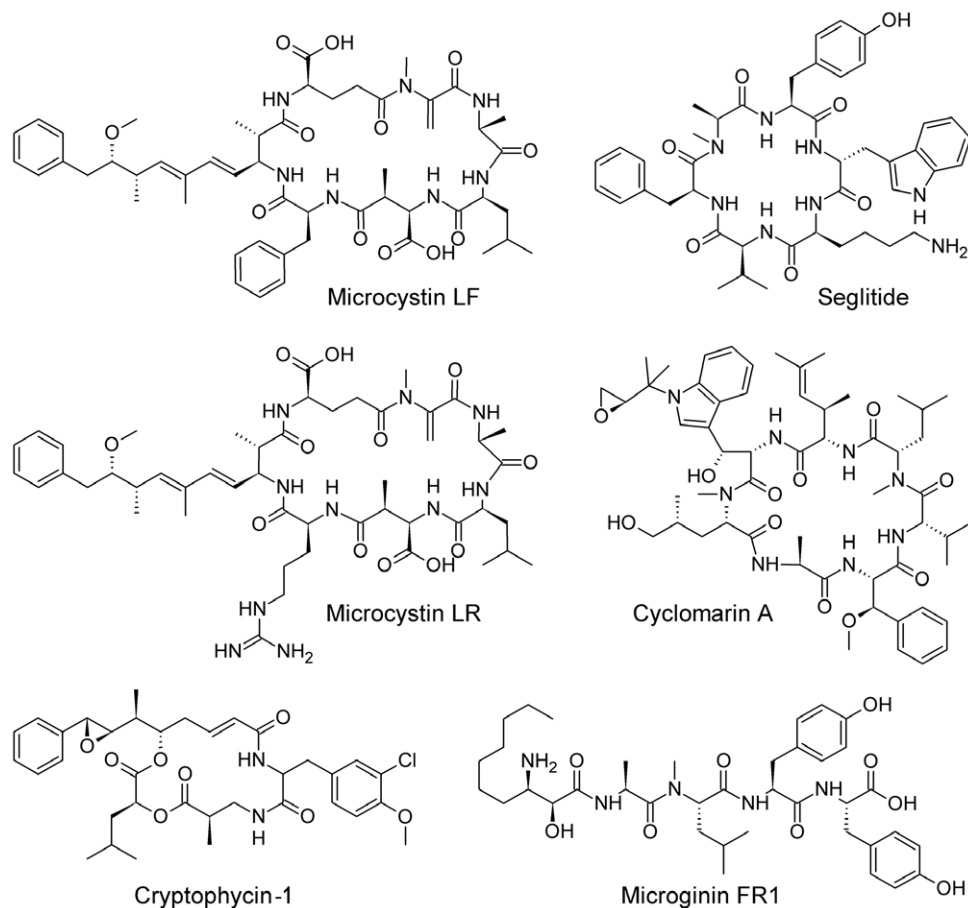


Figure 1. Compounds discussed in this paper.
doi:10.1371/journal.pone.0044913.g001

deviations of the theoretical masses of the discussed fragments and the experimentally determined masses in ppm.

Comparison with NRP-Annotation

MS data were imported into mMass and peaks were manually picked. The resulting peak list was saved in the .dta file format and uploaded to the NRP-Annotation server. NRP-Annotation filters peaks according to a non-described algorithm. To ensure a fair comparison between the two software tools, the peak list in mMass was manually reduced to only contain the peaks that were also evaluated by NRP-Annotation.

Design of the Software and Implemented Fragmentation Pathways

Extension of an existing mass spectrometry program. On a search for a software tool satisfying as many as possible of the criteria discussed above, one of the authors (TN) identified the software mMass to be the most suitable basis for further development. mMass is a portable and cross-platform open-source software for mass spectra processing, written mainly in Python [37], which already offered a GUI, spectrum manipulation capabilities, the possibility to enter and save peptide sequences and do *in-silico* fragmentation and spectrum annotation. However, entering and fragmentation of peptides was limited to linear sequences and proteinogenic amino acids, and a user-editable monomer database as well as several fragmentation options important for cyclic peptides were missing. Since mMass is

open-source, the software could be modified to meet all of the criteria described above.

First of all, a monomer library editor has been designed, giving the user the possibility to conveniently enter, edit and save monomers needed for the composition of peptide sequences. The library has been filled with all monomers compiled from the NORINE database [16], facilitating its use for new users and allowing for easy addition of own variants. Furthermore, the monomer editor allows for the definition of possible neutral losses from individual monomers.

Secondly, the sequence organization and handling has been extended to enable the composition of peptides containing non-proteinogenic amino acids. For this purpose, an additional sequence editor has been designed that allows the convenient composition of peptides by typing or dragging and dropping monomers from the monomer library right into the editor. In addition, a peptide can be set as being linear or cyclic.

Finally, the fragmentation module has been improved to handle all possible known fragmentation pathways of cyclic peptides, to allow for the loss of custom neutrals, and to allow for multiple neutral losses from one fragment.

To assess the capabilities of the resulting software package, several cyanobacterial natural product MS² spectra have been annotated, and the annotations have been compared to those made by NRP-Annotation.

Backbone chain fragmentation pathways. Cyclic peptides are known for initial cleavage of the amide bonds, leading to the

formation of all possible b-ion linear sequences having a free N-terminus and, as is known for b-ions, an oxazolone “C-terminus” [25,41,42]. As all initially formed b-ions have an oxazolone “C-terminus”, x-, y-, and z-ions possibly resulting from further fragmentation of these precursors do not have the free carboxylic acid C-terminus normally found in these ion types. Therefore in this publication, they are called y-like, x-like and z-like ions. A summary of possible backbone fragmentation pathways is depicted in figure 2, taking Microcystin LF as an example.

b- and y-like fragment ions. The linear b-ions resulting from the initial cleavage can – as is well known for linear b-ions – undergo further fragmentation via the b_x - b_{x-1} pathway to form other ions of the respective b-series. Fragmentation can also continue via the b_x - y_z pathway, resulting in the respective y-series [26]. However, as the source fragment of the y-series, like b-ions, has an oxazolone “C-terminus” and not a free carboxylic acid, the y-series ions are not distinguishable from identical b-ions originating from a b_x - b_{x-1} fragmentation of another linear precursor of the same cyclic peptide. For the sake of completeness, the calculation of y-like ions has been included in mMass, although for every y-like ion there is a structurally corresponding b-ion.

a- and x-like fragment ions. b-Ions are known to lose CO to form the corresponding a-ions via the b_x - a_x pathway [26], which is also often described for cyclic peptides [42–45]. Formation of x-like ions is mechanistically possible, therefore this ion series has also been included in the software, although to the best of our knowledge these ions have not yet been described for cyclic peptides.

c- and z-like fragment ions. c-Type fragmentation of the precursor b-ions is mechanistically possible and has been described for microcystins [46,47]. The same holds true for z-like fragments [47], which in the case of cyclic peptides are equal to y-like-NH₃ or b-NH₃ ions with respect to their sum formula. However, they have been included in the software because they should be brought to the attention of the user also if the options to calculate y-like ions or NH₃-losses from b-ions are not chosen, and because the NH₃ loss from y-like or b-ions is not restricted to the N-terminal amino group, whereas in the case of z-like ions, the terminal amino group is always lost.

Alternative initial ring opening. b+CO- and c+CO-ions have frequently been described [40,45,48,49]. Although the

mechanism of their formation has not yet been extensively studied, Eckart [42] has proposed the following pathway: In addition to the initial formation of b-ions, the cleavage of the C_α-C_{carbonyl} bond in an a/x-like manner is also possible, resulting in the corresponding N-formyl immonium ions. Subsequent loss of amine fragments leads to ions that formally can be described as “CO adducts”. The proposed mechanism of the formation of a +CO ion observed within this study is shown in figure 3. The formyl group is attached to the N-terminus, thus only a-, b-, and c-type ions can show this formal addition of CO. Addition of CO to an a-type ion leads to an ion with the same sum formula as the corresponding b-ion and can thus not be discriminated in MS² experiments.

Sequence scrambling. Sequence scrambling of a- and b-ions is a phenomenon that has recently been described and studied in more detail [50–56]. It has also been observed in the case of cyclic peptides [40,45,57]. The mechanism of such sequence scrambling according to the cited publications would be as follows. 1) Initial break-up of the cyclic peptide to a b-ion as described above. 2) Fragmentation of the resulting linear peptide. 3) Recyclization of the resulting b- or a-ion fragment by nucleophilic attack of the N-terminal nitrogen on a C-terminal carbonyl carbon in the case of b-ions or on the carbon center of the C-terminal immonium ion in the case of a-ions. 4) Reopening of the resulting cyclic peptide at another peptide bond. 5) Fragmentation of the resulting linear peptide. Sequence scrambling can result in a lot of identical sequences, e.g. for a cyclic peptide with seven monomers the formation of 14 identical b₂-ions is possible over different pathways. For the sake of clarity, a filtering algorithm has been implemented that removes all ions with identical sequence but the ion formed over the simplest pathway. As is shown in Example S1, the *in-silico* sequence scrambling of YAGFL-NH₂ using our algorithm fits to the experimental fragmentation described in the literature [51,58]. The software does not take into account possible sequence scrambling reactions via the side chains [59].

Neutral Losses. Loss of H₂O or NH₃ is a common phenomenon found for cyclic peptides, even if no hydroxyl groups are present in the molecule [40,44,46,48,60–64]. Loss of H₂O and NH₃ at the same time is possible, and also side chain neutral losses in conjunction with H₂O or NH₃ losses have been observed [27,45,47]. Side chain neutral losses are commonly observed upon

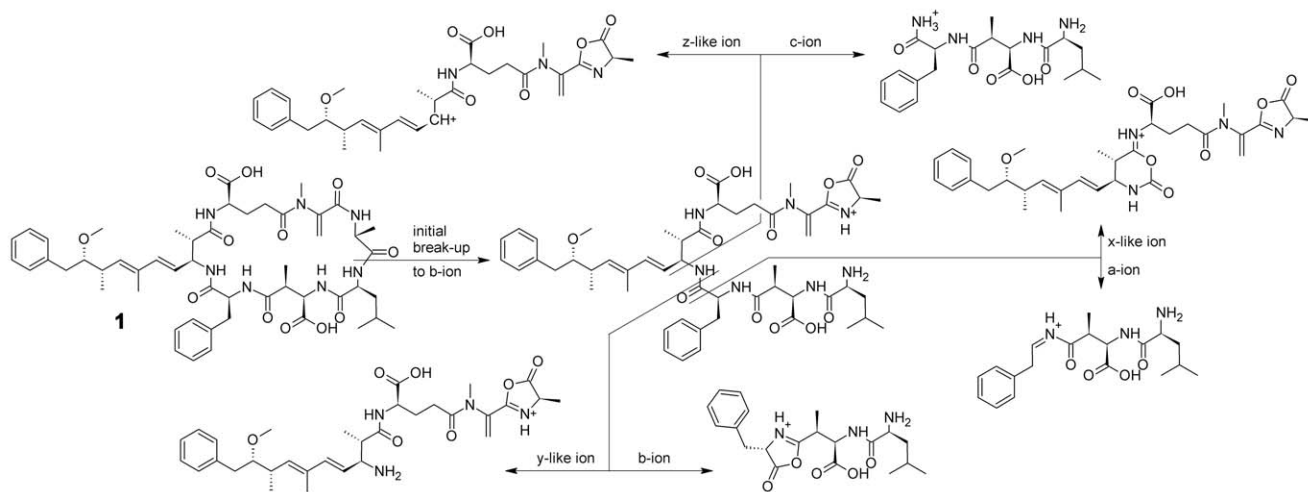


Figure 2. Exemplary backbone fragmentation pathways for a cyclic peptide. After initial cleavage of a random amide bond of Microcystin LF (1), the resulting b-ion can undergo further fragmentation, resulting in a/x-like, b/y-like or c/z-like ions. doi:10.1371/journal.pone.0044913.g002

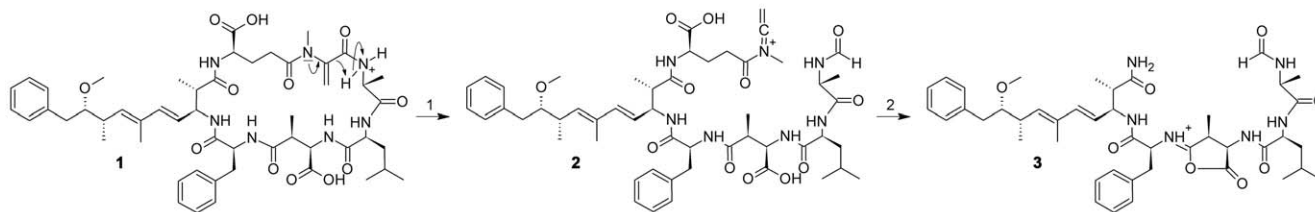


Figure 3. Proposed pathway for the formation of an observed N-formyl fragment ion. Initial D-alanine protonation of microcystin LF (1) and subsequent reactions 1 and 2 result in the formation of the observed N-formyl fragment ion (3). 1) Break-up to N-formyl ion (2). 2) c-like fragmentation and water loss.

doi:10.1371/journal.pone.0044913.g003

Electron Capture Dissociation (ECD), Electron-Transfer Dissociation (ETD) and high-energy Collision Induced Dissociation (CID) [65–68]. They are less frequently observed for low-energy CID [23,69,70], but several side chain neutral losses have also been described for this dissociation technique, e.g. neutral losses from arginine [47,71–73], the non-proteinogenic amino acid Adda ((2S,3S,8S,9S)-3-amino-9-methoxy-2,6,8-trimethyl-10-phenyl-deca-4,6-dienoic acid) [46,47] or other modified amino acids [74]. Due to the possible occurrence of these losses, mMass now gives the user the possibility to freely define side chain neutral losses. However, as the authors do not use ECD/ETD, losses typical for these techniques have not been included in the monomer database yet. Users are invited to mine the mentioned references and enter the losses using the build-in monomer editor.

Gains. The formal addition of H₂O to b-ions has been described and mechanistically discussed [57,73,75]. Possible pathways for the formation of these ions in the case of cryptophycin-1 are discussed below and shown in Figure 4.

Nomenclature. The nomenclature for cyclic peptide fragments proposed by Ngoka and Gross [76], based on Biemanns modification [77] of Roepstorffs nomenclature [78], is suitable for cyclic peptide fragments containing mainly proteinogenic amino acids. However, if a peptide is mainly composed of non-proteinogenic amino acids, the labeling becomes confusing due to multi-letter monomer abbreviations. Thus this software

indicates the initial cleavage site using numbers, similar to the approach proposed by Jegorov et al. [79]. Because the software already used the symbol “-” in a different context, we chose to use the symbol “|” instead, optically indicating a bond cleavage between two monomers of the cyclic peptide. Thus e.g. the 6 monomers containing b-ion of a cyclic peptide initially cleaved between the first and second monomer is labeled b_{6[1|2][1–6]}, 1 being the novel C-terminal oxazolone and 2 the novel N-terminus. Sequence scrambling complicates the nomenclature, as two ring openings and an interjacent fragmentation step must be described. If e.g. the b_{6[2|1][1–6]} fragment recycles, reopens, is subsequently cleaved after the new position 2 and loses one further amino acid, the fragment is labeled b_{5[1|2][1–6][2|3][1–5]}. Thus our fragment ion index nomenclature intuitively summarizes the complete fragment formation history. In the user interface, the complete fragment sequences are given in the fragment table using the respective monomer abbreviations as well.

Results and Discussion: Evaluation of the Software

Experimental and *in-silico* Fragmentation of Peptides

Several compounds were subjected to MSⁿ analyses as described above, and the resulting tandem mass spectra were annotated using both NRP-Annotation and the novel version of

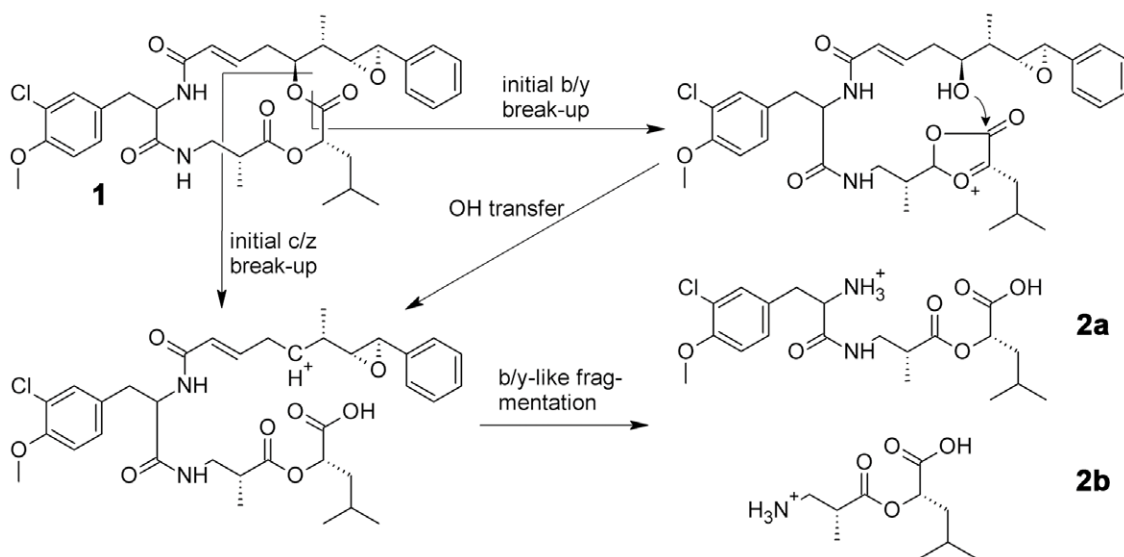


Figure 4. Proposed mechanism for the formation of observed b+H₂O ions. b+H₂O fragment ions (2a) and (2b) of cryptophycin-1 (1) can result from two ring opening pathways: direct c/z-like cleavage of the cyclic peptide or b/y-like cleavage followed by OH transfer.

doi:10.1371/journal.pone.0044913.g004

mMass. The results of this comparison are described in the following and summarized in table 1. The tandem mass spectra are shown in figure 5. Raw data (Data S1) as well as annotation analysis reports of the compounds discussed in the following (Example S2) can be found in the Supporting Information.

The Cyclic Peptide Microcystin LF

The fragmentation behavior of microcystins, cyclic heptapeptides found in several cyanobacterial genera, has been well studied [46,47,60,80,81]. Thus microcystin LF has been used to examine the accuracy and comprehensiveness of our fragmentation algorithms. To our surprise, NRP-Annotation was filtering several significant and important peaks out of the uploaded raw data. Furthermore, only relatively few peaks were annotated (summing up to 23% of the overall peak intensity after filtering). Several of the most abundant signals, including the base peak, were not assigned. Thus we did a comparative study using the novel algorithms of mMass. Including only a- and b-ions, sequence scrambling and simple neutral losses of H₂O and NH₃ in our calculations in analogy to NRP-Annotation, we obtained identical assignments. Including also M-, c-, x-, and z-type ions matched an additional 26% of the overall peak intensity. Including the option for N-formyl ions brought a further 10% increase of matched intensity. Allowance of user defined neutral losses, described for the Adda moiety as discussed above, raised the matching quota by further 26%. The inclusion of multiple neutral losses gained 10%, resulting in a total annotated peak intensity of 96%.

The tandem mass spectrum of microcystin LF shows several significant peaks that had not been assigned by NRP-annotation, but were annotated by mMass (figure 5A). The fragment ions observed at m/z 478.2644 and m/z 561.3017 were annotated as c-ions c₄[7₁][1-4] (Δ 3.3 ppm) and c₅[6₁7][1-5] (Δ 2.5 ppm), resp. A z-like ion with a neutral loss, z₅[2₁3][3-7]-C₉H₁₀O, at m/z 559.3127 (Δ 0.2 ppm) is often described for Adda-containing microcystins [47,60]. The respective ion without neutral loss, z₅[2₁3][3-7] is observed at m/z 693.3816 (also assigned by NRP-Annotation as b-NH₃-ion; Δ 3.1 ppm). The base peak of the spectrum at m/z 852.4486 has been annotated as M-C₉H₁₀O by mMass (Δ 1.9 ppm), showing how important it is to allow for user definable neutral losses and to include M ions in the calculations. Another significant peak was observed at m/z 835.4252, annotated by mMass as M-C₉H₁₀O-NH₃ (Δ 1.9 ppm), indicating that fragments with multiple neutral losses can make up a significant proportion of the total peak intensity. N-formyl ions have been assigned to both b- and c-like ion series by mMass; the proposed formation mechanism of a prominent N-formyl c-ion at m/z 801.4549, c₅[7₁][1-5]-H₂O+CO (Δ 0.5 ppm), is shown in figure 3. To confirm the annotations made, selected ions were

subjected to MS³. The resulting product ion spectra and their interpretation are presented in Figure S1.

The Cyclic Depsipeptide Cryptophycin-1

NRP-Annotation again filters many significant peaks out of the uploaded MS² spectrum using its non-described algorithm. The base peak of the spectrum as well as other peaks of significant intensity are not annotated (figure 5B). Analysis by mMass shows that the base peak at m/z 218.1384 is the water adduct of a b-series ion, b₂[2₁3][1-2]+H₂O (Δ 1.4 ppm). In fact, also the water adduct of the fragment ion b₃[1₁2][1-3], containing one more monomer, is found in the spectrum at m/z 429.1782 (Δ 1.2 ppm), although with lower intensity. The free carboxylic acid C-terminus can result from two hypothetical ring opening pathways, either by direct c/z-like cleavage of the cyclic peptide or by b/y-like cleavage followed by OH transfer as proposed by Fang et al. [75] and depicted in figure 4. As all depsipeptides after initial b/y-type cleavage possess a free OH group that can serve as OH donor, or a free carboxylic acid after c/z-type cleavage, we expect that b+H₂O ions will be commonly observed ions for all depsipeptides. Thus it is important for the interpretation of depsipeptide MS² spectra to have this option available in the software used to annotate the spectra. The third most intense peak at m/z 410.1512 has also not been assigned by NRP-Annotation. mMass assigned this peak as a₂[4₁][1-2]-H₂O (Δ 1.2 ppm). This underlines the importance to allow neutral losses from all types of initial fragment ions. The corresponding a-ion without water loss is also observed with lower abundance at m/z 428.1628 (Δ 1.4 ppm). The MS³ spectra of the main signals discussed and their interpretation are shown in Figure S1.

Other Cyclic Peptides

Because our assignments of the fragments observed for microcystins LF and cryptophycin-1 using mMass were rather extensive, we also compared the mass spectra annotations of further cyclic peptides, namely microcystin LR, seglitide, and cyclomarin A. For microcystin LR, we again found a more concise annotation using mMass. However, NRP-Annotation did assign a higher percentage of the peak intensity for this compound than for microcystin LF; due to the arginine-localized charge, b-ions containing arginine show a high abundance. The experimental data available for seglitide are low-resolution ion trap data; this leads to a very diverse annotation when using mMass at a peak tolerance of 0.2 Da, resulting in a peak intensity annotation of close to 100%. Data with higher resolution would possibly avoid the ambiguous annotation of the signals, but this data was not available from the NRP-Annotation website. The tandem mass spectrum of seglitide is much easier to interpret than those of the

Table 1. Comparison of fragment ion annotation using NRP-Annotation or mMass.

	MC-LF ¹	MC-LR ¹	Seglitide ²	CM A ³	CRY-1 ¹	MG FR1 ¹
NRP-An.	23%, 13 of 57	52%, 22 of 66	74%, 38 of 53	73%, 36 of 99	32%, 10 of 36	--
mMass*	23%, 13 of 57	52%, 22 of 66	74%, 38 of 53	73%, 36 of 99	32%, 10 of 36	--
mMass**	96%, 51 of 57	79%, 55 of 66	98%, 51 of 53	100%, 90 of 99	80%, 21 of 36***	99%, 12 of 14

Indicated are the percentage of the ion intensity annotated as well as the number of ions matched by the software vs. the total ion number.

MC = Microcystin, CM = Cyclomarin, CRY = Cryptophycin, MG = Microginin.

*options chosen that reflect the fragments NRP-Annotation can calculate (a- and b-ions, -NH₃ and -H₂O, Allow Scrambling, including b₁ ions).

**all available options chosen.

***if ³⁷Cl isotope peaks are manually deleted: 88%, 21 of 32.

m/z tolerances used for matching calculated to experimentally observed fragments: ¹0.015 Da, ²0.2 Da, ³0.1 Da.

doi:10.1371/journal.pone.0044913.t001

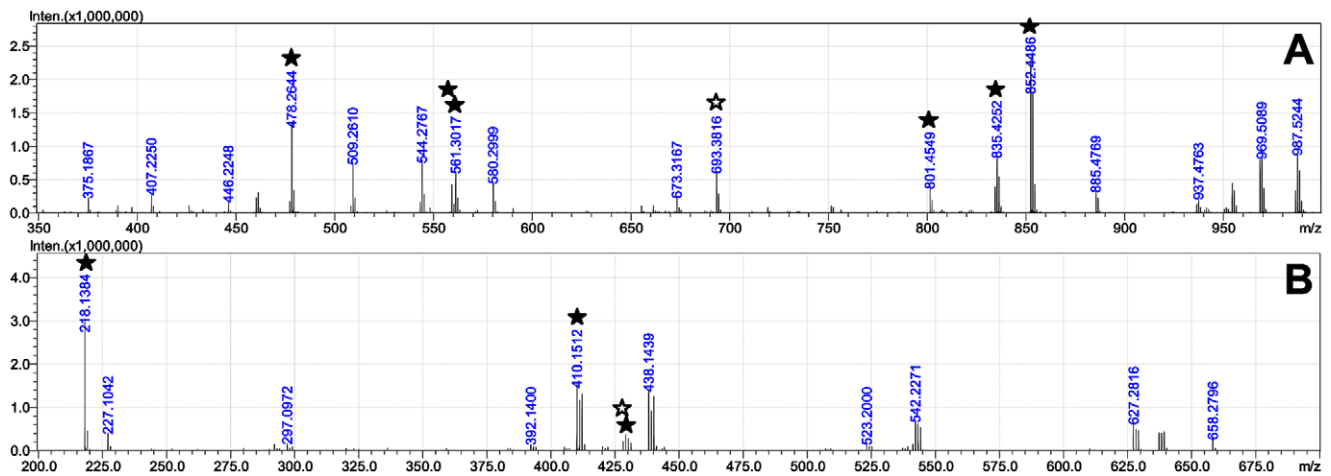


Figure 5. Tandem mass spectra of microcystin LF and cryptophycin-1. (A) microcystin LF, (B) cryptophycin-1. Signals discussed in the text are indicated with a star (filled star: not annotated by NRP-Annotation). doi:10.1371/journal.pone.0044913.g005

other compounds, because seglitide contains only one non-proteinogenic amino acid. In the case of cyclomarin A, 0.1 Da mass tolerance was used by the authors of MS-CPA although the data had been obtained using a TOF analyzer. Again, mMass was able to annotate almost every peak in the spectrum.

The Linear Peptide Microginin FR-1

To study the suitability of the software for the *in-silico* fragmentation of linear non-ribosomal peptides, microginin FR-1 served as a model compound. The software efficiently annotated almost all observed peaks. Due to the fact that this compound is a linear peptide, the fragmentation is much simpler and could also have been accomplished with other software suitable for the *in-silico* fragmentation of linear peptides allowing for the definition of custom monomers. However, mMass offers more extensive fragmentation options than most of the other software available and enables very convenient non-ribosomal peptide monomer handling and sequence composition, and thus can also be the software of choice for the annotation of linear peptide tandem mass spectra.

Conclusions

Because the currently available software tools for cyclic peptide tandem mass spectrum annotation could not be used for comprehensive spectrum annotation, we have significantly extended a freely available and open-source software, mMass, to allow for the necessary functionalities. The program, to the best of our knowledge, is the most concise tool available today for cyclic as well as non-ribosomal peptide tandem mass spectrum annotation, and has proven its usability. Due to the open nature of the software, users are able to contribute. This especially holds true for the monomers library, where individual users are invited to share their own extended monomer databases.

Using own experimental data sets as well as data obtained from other scientists that studied *in-silico* fragmentation of cyclic peptides, we have shown that for all examined peptides, our algorithms were able to assign >80% of the total observed peak intensity and that the software is equally well suited for the annotation of cyclic depsipeptide and linear peptide fragment ions. The assignments made by mMass are more comprehensive than those of all other currently available software tools for this purpose.

mMass is not suitable for the prediction of tandem mass spectra of cyclic peptides as no fragmentation ion intensities can be calculated.

But as the knowledge about the intensity patterns of non-ribosomal peptide fragments, and even more for cyclic peptide fragments, is rudimentary and highly depending on the instrument type used, this is not easy to implement at the time being. Furthermore, the software is not able to check if calculated fragments are likely to be observed under experimental conditions. The resulting annotations thus need the attentive assessment of the analyst – but this holds true for all software solutions. The fragmentation prediction has not been extended yet to also include branched cyclic peptides or peptides with more than one cycle.

In the used datasets, some signals still remained unassigned. A closer examination of these peaks might reveal novel fragmentation mechanisms for cyclic peptides, which can be added in any future version of mMass.

The software can conveniently be used in early steps of structure elucidation/dereplication (matching of *in-silico* fragmentation of compounds in databases with experimental tandem MS spectra) and as one of the last steps of structure confirmation (extent of assignable peaks). A typical workflow for a tandem mass spectrum annotation is presented in Example S3.

Availability

mMass 5.2.0. The Python source code, executables for MS Windows, Mac OS X and Linux as well as test data can be downloaded free of charge from the project webpage <http://www.mmass.org>. The executables do not require installation. All download packages contain a comprehensive manual.

Supporting Information

Figure S1 Product ion spectra of selected fragments discussed in the publication.

(PDF)

Data S1 Raw and processed tandem MS data.

(ZIP)

Example S1 Sequence Scrambling using mMass.

(PDF)

Example S2 Analysis and Annotation Reports of the compounds discussed.

(PDF)

Example S3 Workflow for the annotation of mass spectra. (PDF)

Acknowledgments

We thank Maude Pupin and the team of the NORINE database for providing us with the monomer dataset. TN thanks Robert Gieseke for Python support.

References

- Tiburzi F, Visca P, Imperi F (2007) Do Nonribosomal Peptide Synthetases Occur in Higher Eukaryotes? *IUBMB Life* 59: 730–733.
- Caboche S, Leclère V, Pupin M, Kucherov G, Jacques P (2010) Diversity of monomers in nonribosomal peptides: towards the prediction of origin and biological activity. *J Bacteriol* 192: 5143–5150.
- Sainis I, Fokas D, Vareli K, Tzakos AG, Kounnis V, et al. (2010) Cyanobacterial cyclopeptides as lead compounds to novel targeted cancer drugs. *Mar drugs* 8: 629–657.
- Namikoshi M, Rinehart KL (1996) Bioactive compounds produced by cyanobacteria. *J Ind Microbiol Biotechnol* 17: 373–384.
- Jaspars M, Lawton LA (1998) Cyanobacteria - a novel source of pharmaceuticals. *Curr Opin Drug Discov Dev* 1: 77–84.
- Sielaff H, Christiansen G, Schwecke T (2006) Natural products from cyanobacteria: Exploiting a new source for drug discovery. *IDrugs* 9: 119–127.
- Ziemert N, Ishida K, Liaimer A, Hertweck C, Dittmann E (2008) Ribosomal Synthesis of Tricyclic Depsipeptides in Bloom-Forming Cyanobacteria. *Angew Chem Int Ed Engl* 47: 7756–7759.
- Velásquez JE, van der Donk WA (2011) Genome mining for ribosomally synthesized natural products. *Curr Opin Chem Biol* 15: 11–21.
- Marahiel MA (2009) Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis. *J Pept Sci* 15: 799–807.
- Schwarzer D, Finking R, Marahiel MA (2003) Nonribosomal peptides: from genes to products. *Nat Prod Rep* 20: 275–287.
- Strieker M, Tanović A, Marahiel MA (2010) Nonribosomal peptide synthetases: structures and dynamics. *Curr Opin Struct Biol* 20: 234–240.
- Finking R, Marahiel MA (2004) Biosynthesis of nonribosomal peptides. *Annu Rev Microbiol* 58: 453–488.
- Tillett D, Dittmann E, Erhard M, von Döhren H, Börner T, et al. (2000) Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide-polyketide synthetase system. *Chem Biol* 7: 753–764.
- Dittmann E, Neilan BA, Börner T (2001) Molecular biology of peptide and polyketide biosynthesis in cyanobacteria. *Appl Microbiol Biotechnol* 57: 467–473.
- Christiansen G, Fastner J, Erhard M, Börner T, Dittmann E (2003) Microcystin Biosynthesis in *Planktothrix*: Genes, Evolution, and Manipulation. *J Bact* 185: 564–572.
- Caboche S, Pupin M, Leclère V, Fontaine A, Jacques P, et al. (2008) NORINE: a database of nonribosomal peptides. *Nucleic Acids Res* 36: 326–331.
- Dreyfuss M, Härrli E, Hofmann H, Kobel H, Pache W, et al. (1976) Cyclosporin A and C: New Metabolites from *Trichoderma polysporum*. *Microbiol* 133: 125–133.
- von Wartburg A, Traber R (1988) Cyclosporins, fungal metabolites with immunosuppressive activities. *Prog Med Chem* 25: 1–33.
- McCormick MH, Stark WM, Pittenger GE, Pittenger RC, McGuire JM (1956) Vancomycin, a new antibiotic. I. Chemical and biologic properties. *Antibiot Annu* 3: 606–611.
- Nagarajan R (1991) Antibacterial Activities and Modes of Action of Vancomycin and Related Glycopeptides. *Antimicrob Agents Chemother* 35: 605–609.
- Rohr J (2006) Cryptophycin Anticancer Drugs Revisited. *ACS Chem Biol* 1: 747–750.
- Wysocki VH, Resing KA, Zhang Q, Cheng G (2005) Mass spectrometry of peptides and proteins. *Methods* 35: 211–222.
- Papayannopoulos IA (1995) The Interpretation of Collision-Induced Dissociation Tandem Mass Spectra of Peptides. *Mass Spectrom Rev* 14: 49–73.
- Barlow CK, Hair RAJO (2008) Gas-phase peptide fragmentation: how understanding the fundamentals provides a springboard to developing new chemistry and novel proteomic tools. *J Mass Spectrom* 43: 1301–1319.
- Harrison AG (2009) TO b OR NOT TO b: The Ongoing Saga of Peptide b Ions. *Mass Spectrom Rev* 28: 640–654.
- Paizs B, Suhai S (2005) Fragmentation pathways of protonated peptides. *Mass Spectrom Rev* 24: 508–548.
- Arnold RJ, Jayasankar N, Aggarwal D, Tang H, Radivojac P (2006) A machine learning approach to predicting peptide fragmentation spectra. *Pac Symp Biocomput* 11: 219–230.
- Frank AM (2009) Predicting Intensity Ranks of Peptide Fragment Ions. *J Proteome Res* 8: 2226–2240.
- Webb-Robertson B-JM, Cannon WR (2007) Current trends in computational inference from mass spectrometry-based proteomics. *Brief Bioinformatics* 8: 304–317.
- Li S, Arnold RJ, Tang H, Radivojac P (2011) On the Accuracy and Limits of Peptide Fragmentation Spectrum Prediction. *Anal Chem* 83: 790–796.
- Zhang Z (2004) Prediction of Low-Energy Collision-Induced Dissociation Spectra of Peptides. *Anal Chem* 76: 3908–3922.
- UCSF Mass Spectrometry Facility, ProteinProspector. Available: <http://prospector2.ucsf.edu/prospector/ms/home.htm>. (accessed September 15, 2011).
- Masselot A (2006) InSilicoSpectro. Available: <http://insilicospectro.vital-it.ch/>. (accessed September 15, 2011).
- Colinge J, Masselot A, Carbonell P, Appel RD (2006) InSilicoSpectro: An Open-Source Proteomics Library. *J Proteome Res* 5: 619–624.
- Institute for Systems Biology (ISB), Proteomics Toolkit. Available: <http://db.systemsbiology.net:8080/proteomicsToolkit/FragIonServlet.html>. (accessed September 15, 2011).
- Li S (2010) PeptideART. Available: <http://peptideart.sourceforge.net/>. (accessed September 15, 2011).
- Strohalm M, Kavan D, Novák P, Volný M, Havlicek V (2010) mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data. *Anal Chem* 82: 4648–4651.
- Rusconi F (2009) massXpert 2: a cross-platform software environment for polymer chemistry modelling and simulation/analysis of mass spectrometric data. *Bioinformatics* 25: 2741–2742.
- Jagannath S, Sabareesh V (2007) Peptide Fragment Ion Analyser (PFIA): a simple and versatile tool for the interpretation of tandem mass spectrometric data and de novo sequencing of peptides. *Rapid Commun Mass Spectrom* 21: 3033–3038.
- Liu W-T, Ng J, Meluzzi D, Bandeira N, Gutierrez M, et al. (2009) The interpretation of tandem mass spectra obtained from cyclic non-ribosomal peptides. *Anal Chem* 81: 4200–4209.
- Poller NC, Oomens J, Suhai S, Paizs B (2005) Spectroscopic and theoretical evidence for oxazolone ring formation in collision-induced dissociation of peptides. *J Am Chem Soc* 127: 17154–17155.
- Eckart K (1994) Mass Spectrometry of Cyclic Peptides. *Mass Spectrom Rev* 13: 23–55.
- Nakamura T, Nahaki H, Kinoshita T (1986) Amino Acid Sequence Determination of Cyclic Peptides by Tandem Mass Spectrometry Coupled with Fast Atom Bombardment Ionization. *Mass Spectrosc* 34: 307–319.
- Jegorov A, Paizs B, Kuzma M, Zabka M, Landa Z, et al. (2004) Extraribosomal cyclic tetradepsipeptides beauverolides: profiling and modeling the fragmentation pathways. *J Mass Spectrom* 39: 949–960.
- Schäfer M, Fuchs R, Budzikiewicz H, Springer A, Meyer J-M, et al. (2006) Structure elucidation of cyclic pyoverdins and examination of rearrangement reactions in MS/MS experiments by determination of exact product ion masses. *J Mass Spectrom* 41: 1162–1170.
- Kubwabo C, Vais N, Benoit FM (2005) Characterization of microcystins using in-source collision-induced dissociation. *Rapid Commun Mass Spectrom* 19: 597–604.
- Yuan M, Namikoshi M, Otsuki A, Rinehart KL, Sivonen K, et al. (1999) Low-energy collisionally activated decomposition and structural characterization of cyclic heptapeptide microcystins by electrospray ionization mass spectrometry. *J Mass Spectrom* 34: 33–43.
- Mayumi T, Kato H, Kawasaki Y, Harada K-ichi (2007) Formation of diagnostic product ions from cyanobacterial cyclic peptides by the two-bond fission mechanism using ion trap liquid chromatography/multi-stage mass spectrometry. *Rapid Commun Mass Spectrom* 21: 1025–1033.
- Eckart K, Schwarz H, Tomer KB, Gross ML (1985) Tandem Mass Spectrometry Methodology for the Sequence Determination of Cyclic Peptides. *J Am Chem Soc* 107: 6765–6769.
- Yagüe J, Paradelo A, Ramos M, Ogueta S, Marina A, et al. (2003) Peptide rearrangement during quadrupole ion trap fragmentation: added complexity to MS/MS spectra. *Anal Chem* 75: 1524–1535.
- Harrison AG, Young AB, Bleiholder C, Suhai S, Paizs B (2006) Scrambling of sequence information in collision-induced dissociation of peptides. *J Am Chem Soc* 128: 10364–10365.
- Jia C, Qi W, He Z (2007) Cyclization reaction of peptide fragment ions during multistage collisionally activated decomposition: an inducement to lose internal amino-acid residues. *J Am Soc Mass Spectrom* 18: 663–678.
- Harrison AG (2008) Peptide sequence scrambling through cyclization of b(5) ions. *J Am Soc Mass Spectrom* 19: 1776–1780.
- Jia C, Qi W, He Z (2007) Cyclization reaction of peptide fragment ions during multistage collisionally activated decomposition: an inducement to lose internal amino-acid residues. *J Am Soc Mass Spectrom* 18: 663–678.

Author Contributions

Conceived and designed the experiments: THJN. Performed the experiments: THJN. Analyzed the data: THJN. Wrote the paper: THJN MS. Programmed prototype extension of mMass: THJN. Refined the software code and implemented it into the new version of mMass: MS.

55. Polfer NC, Oomens J, Suhai S, Paizs B (2007) Infrared Spectroscopy and Theoretical Studies on Gas-Phase Protonated Leu-enkephalin and Its Fragments: Direct Experimental Evidence for the Mobile Proton. *J Am Chem Soc* 129: 5887–5897.
56. Bythell BJ, Maitre P, Paizs B (2010) Cyclization and rearrangement reactions of a(n) fragment ions of protonated peptides. *J Am Chem Soc* 132: 14766–14779.
57. Fuchs R, Budzikiewicz H (2001) Rearrangement reactions in the electrospray ionization mass spectra of pyoverdins. *Int J Mass Spectrom* 210–211: 603–612.
58. Bleiholder C, Osburn S, Williams TD, Suhai S, Van Stipdonk M, et al. (2008) Sequence-scrambling fragmentation pathways of protonated peptides. *J Am Chem Soc* 130: 17774–17789.
59. Tang XJ, Thibault P, Boyd RK (1993) Fragmentation reactions of multiply-protonated peptides and implications for sequencing by tandem mass spectrometry with low-energy collision-induced dissociation. *Anal Chem* 65: 2824–2834.
60. Yuan M, Namikoshi M, Otsuki A (1998) Effect of amino acid side-chain on fragmentation of cyclic peptide ions: differences of electrospray ionization/collision-induced decomposition mass spectra of toxic heptapeptide microcystins containing ADMAdda instead of Adda. *Eur J Mass Spectrom* 4: 287–298.
61. Schilling B, Wang W, McMurray JS, Medzihradzky KF (1999) Fragmentation and sequencing of cyclic peptides by matrix-assisted laser desorption/ionization post-source decay mass spectrometry. *Rapid Commun Mass Spectrom* 13: 2174–2179.
62. Sabareesh V, Ranganayaki RS, Raghobama S, Bopanna MP, Balaram H, et al. (2007) Identification and Characterization of a Library of Microheterogeneous Cyclohexadepsipeptides from the Fungus *Isaria*. *J Nat Prod* 70: 715–729.
63. Tilvi S, Naik CG (2007) Tandem mass spectrometry of kahalalides: identification of two new cyclic depsipeptides, kahalalide R and S from *Elysia grandifolia*. *J Mass Spectrom* 42: 70–80.
64. Tomer KB, Crow FW, Gross ML, Kopple KD (1984) Fast atom bombardment combined with tandem mass spectrometry for the determination of cyclic peptides. *Anal Chem* 56: 880–886.
65. Cooper HJ, Hudgins RR, Håkansson K, Marshall AG (2002) Characterization of amino acid side chain losses in electron capture dissociation. *J Am Soc Mass Spectrom* 13: 241–249.
66. Savitski MM, Nielsen ML, Zubarev RA (2007) Side-chain losses in electron capture dissociation to improve peptide identification. *Anal Chem* 79: 2296–2302.
67. Xia Q, Lee MV, Rose CM, Marsh AJ, Hubler SL, et al. (2011) Characterization and diagnostic value of amino acid side chain neutral losses following electron-transfer dissociation. *J Am Soc Mass Spectrom* 22: 255–264.
68. Leymarie N, Costello CE, O'Connor PB (2003) Electron Capture Dissociation Initiates a Free Radical Reaction Cascade. *J Am Chem Soc* 125: 8949–8958.
69. Lioc H, O'Hair RAJ (2007) Comparison of collision-induced dissociation and electron-induced dissociation of singly protonated aromatic amino acids, cystine and related simple peptides using a hybrid linear ion trap-FT-ICR mass spectrometer. *Anal Bioanal Chem* 389: 1429–1437.
70. Medzihradzky KF (2005) Peptide Sequence Analysis. *Methods Enzymol* 402: 209–244.
71. Deery MJ, Summerfield SG, Buzy A, Jennings KR (1997) A Mechanism for the Loss of 60 u from Peptides Containing an Arginine Residue at the C-Terminus. *J Am Soc Mass Spectrom* 8: 253–261.
72. Gehrig PM, Hunziker PE, Zahariev S, Pongor S (2004) Fragmentation pathways of N(G)-methylated and unmodified arginine residues in peptides studied by ESI-MS/MS and MALDI-MS. *J Am Soc Mass Spectrom* 15: 142–149.
73. Bythell BJ, Csonka IP, Suhai S, Barofsky DF, Paizs B (2010) Gas-phase structure and fragmentation pathways of singly protonated peptides with N-terminal arginine. *J Phys Chem B* 114: 15092–15105, and references cited therein.
74. Swiderek KM, Davis MT, Lee TD (1998) The identification of peptide modifications derived from gel-separated proteins using electrospray triple quadrupole and ion trap analyses. *Electrophoresis* 19: 989–997.
75. Fang S, Takao T, Satomi Y, Mo W, Ahimionishi Y (2000) Novel Rearranged Ions Observed for Protonated Peptides Via Metastable Decomposition in Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. *J Am Soc Mass Spectrom* 11: 345–351, and references cited therein.
76. Ngoka LCM, Gross ML (1999) A Nomenclature System for Labeling Cyclic Peptide Fragments. *Am Soc Mass Spectrom* 10: 360–363.
77. Biemann K (1992) Mass Spectrometry of Peptides and Proteins. *Annu Rev Biochem* 61: 977–1010.
78. Roepstorff P, Fohlman J (1984) Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* 11: 601.
79. Jegorov A, Havlicek V (2001) Spontaneous N → O acyl shift in the [M+H]⁺ ions of [MeBmt1]-cyclosporins in an ion trap. *J Mass Spectrom* 36: 633–640.
80. Mayumi T, Kato H, Imanishi S, Kawasaki Y, Hasegawa M, et al. (2006) ORIGINAL ARTICLE Structural Characterization of Microcystins by LC/MS/MS under Ion Trap Conditions. *J Antibiot* 59: 710–719.
81. Namikoshi M, Rinchart KL, Sakai R, Stotts RR, Dahlem AM, et al. (1992) Identification of 12 hepatotoxins from a Homer Lake bloom of the cyanobacteria *Microcystis aeruginosa*, *Microcystis viridis*, and *Microcystis wesenbergii*: nine new microcystins. *J Org Chem* 57: 866–872.