UNIVERSITY OF CALIFORNIA, SAN DIEGO

Mass spectrometric and bioinformatics approaches to characterizing of cyclic non-ribosomal peptides and ribosomally encoded peptide antibiotic

A thesis submitted in partial satisfaction of the
requirements for the degree Master of Science

in

Chemistry

by

Wei-Ting Liu

Committee in charge:

      Professor Pieter C. Dorrestein, Chair
      Professor Pavel A. Pevzner
      Professor Akif Tezcan

2009

The thesis of Wei-Ting Liu is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

<div align="right">Chair</div>

University of California, San Diego
2009

DEDICATION

To my parents, for everything.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SCHEMES

LIST OF TABLES

ACKNOWLEDGEMENTS

Chapter 1, in full, is a reprint of the material as it appears in Analytical Chemistry, 2009. Liu, Wei-Ting; Ng, Julio; Meluzzi, Dario; Bandeira, Nuno; Gutierrez, Marcelino; Simmons, Thomas L.; Schultz, Andrew W.; Linington, Roger G.; Moore, Bradley S.; Gerwick, William H.; Pevzner, Pavel A.; Dorrestein, Pieter C. The thesis author was the primary investigator and author of this paper.

VITA

2006	Bachelor of Science, in Biological Science, National Chiao Tung University, Hsinchu, Taiwan
2009	Master of Science, in Chemistry, University of California, San Diego

PUBLICATIONS

Ng, J.; Bandeira, N.; Liu, W. T.; Ghassemian, M.; Simmons, T. L.; Gerwick, W.; Linington, R.; Dorrestein, P.; Pevzner, P. Dereplication and De Novo Sequencing of Nonribosomal Peptides. *Nature Methods*. **2009**, accepted.

Hughes, C.; Yang, Y. L.; Liu, W. T.; Dorrestein, P. C.; Fenical, W.; La Clair, J. Natural Product Target Elucidation by Acyl Dye Transfer. *JACS communication*. **2009**, accepted.

Liu, W. T.; Ng, J.; Meluzzi, D.; Bandeira, Nuno.; Gutierrez, M.; Simmons, T. L.; Schultz A. W.; Linington, R. G.; Moore, B. S.; Gerwick, W. H.; Pevzner, P. A.; Dorrestein, P. C. The structural characterization of cyclic nonribosomal peptides by tandem mass spectrometry. *Anal. Chem.* **2009**, *81*, 4200-4209.

Burns, K. E.; Liu, W. T.; Boshoff, H. I. M.; Dorrestein, P. C., Barry, C. E. 3rd. Proteasomal Protein Degradation in Mycobacteria is Dependent upon a Prokaryotic Ubiquitin-Like Protein. *J. Biol. Chem.* **2009**, *284,* 3069-3075.

Chen, W. L.; Liu, W. T.; Yang, M. C.; Hwang, M. T.; Tsao, J. H.; Mao, S. J. T. A Novel Conformation-Dependent Monoclonal Antibody Specific to Native Structure of β-lactoglobulin and its Application. *J. Dairy Sci.* **2006**, *89*, 912-921.

CONFERENCE

Liu, WT; Ng, J.; Meluzzi, D.; Bandeira, N.; Gutierrez, M.; Simmons, TL.; Schultz AW.; Linington, R G.; Moore, BS.; Gerwick, WH.; Pevzner, PA.; Dorrestein, PC. The Interpretation and Annotation of Tandem Mass Spectrometry Data Obtained from Non-ribosomal Derived Cyclic Peptides. *The 2008 BIO International Convention.* Poster.

Chen WL, Liu WT, Hwang MT, and Mao SJT. A Novel Two-dimensional Gel Electrophoresis for Studying the Cross-linking Between β-lactoglobulin and Milk Proteins. *ADSA-ASAS-CSAS Meeting in Cincinnati.* Poster.

Chen WL, Liu WT, and Kao LP, Mao SJT. β-lactoglobulin Plays a Provactive Role in Lymphocyte Proliferation. *Experimental Biology 2006 San Francisco, USA.*

ABSTRACT OF THE THESIS

Mass spectrometric and bioinformatics approaches to characterizing of cyclic non-ribosomal peptides and ribosomally encoded peptide antibiotic

by

Wei-Ting Liu

Master of Science in Chemistry

University of California, San Diego, 2009

Professor Pieter C. Dorrestein, Chair

Natural products are a crucial component in drug discovery because of their considerable pharmaceutical properties. Cyclic non-ribosomally peptides are one category of natural products featured by containing non-standard amino acids and lactam or lactone structures, thus increasing the complexity of the resulting tandem mass spectrometry data. Cyclosporin, microcystins and nodularins all are well-known examples and have notable pharmacologically importance. In this current work, by collaborating with Dr. Pevzner bioinformatics group, a program was developed for the annotation and characterization of tandem mass spectra obtained from cyclic peptides. This program, MS-CPA is available as a web tool (http://lol.ucsd.edu/ms-cpa_v1/Input.py). The second half of this thesis will be focused on a ribosomally encoded peptide antibiotic trifolitoxin produced by a *rhizobium* strain. This peptide undergoes multiple unknown post-translational modifications results in a loss of 24 Da masses. Mass spectrometry based strategies as well as *in vivo* reconstitution experiments were used to partially characterize trifolitoxin structure.

Chapter I

Interpretation of Tandem Mass Spectra Obtained from

Cyclic Nonribosomal Peptides

Chapter 1, in full, is a reprint of the material as it appears in Analytical Chemistry, 2009.

It is a joint effort between thesis author, Julio Ng, and Dr. Nuno Bandeira.

1.1 INTRODUCTION

Ribosomally as well as non-ribosomally derived cyclic peptides are an important group of compounds because of their wide range of biological, toxic, and pharmacological activities, and they often exhibit unique chemical structures.[1,2] For example, the cyclic toxins microcystins and nodularins produced by cyanobacteria (blue-green algae) can wipe out entire fisheries and can cause death in humans.[3,4] In addition, it is now becoming increasingly clear that these naturally occurring cyclic peptides have biological roles in quorem sensing,[5,6] gliding,[7,8] prevention of aerial growth,[9] or cell adherence regulation[10] and that they can be used as a diagnostic markers for disease.[11] In addition, many cyclic peptides are used in the clinic. Well-known examples of cyclic natural products are cyclosporine, an immunosuppressant drug used to prevent organ rejection,[12] seglitide, a potent growth factor release inhibitor,[13] and ramoplanin, a novel antibiotic.[14] Because of the importance of their therapeutic applications, there is a continued development of strategies to generate cyclic libraries for drug screening programs.[15-18] In fact, many cyclic natural products with potent therapeutic properties are discovered every week.[19-24]

Therefore it is important to continue developing methods not only for isolating or preparing such cyclic peptides but also to characterize such peptides. Despite a lot of effort by mass spectrometrists,[25-37] we are still exploring the way cyclic peptides behave in a mass spectrometer, in particular during collision-induced dissociation (CID). Bioinformatics tools such as MASCOT, SEQUEST, and InsPecT are capable of robust interpretation of tandem MS spectra and also enable protein identification with the equipped database search engines.[38-40] However, few tools are designed for cyclic

peptides with a user friendly interface at a level accessible to non mass-spectrometrists. In addition most of the bioinformatics tools are based on somewhat refined fragmentation models, i.e., they may only annotate b and y ions. Both of these are the likely reasons why most scientists that isolate cyclic natural products and that develop cyclic peptide libraries for drug screening programs ignore all but only annotate a small amount of the ions that are typically observed from cyclic peptides in their structural elucidation efforts, leaving tens to hundreds of ions unaccounted for.[30] We became interested in this problem because when we attempted to annotate the tandem mass spectra of cyclic natural products isolated from marine organisms by manual means and discovered that a large proportion of the spectral intensity remained unaccounted for and that the annotation was very time-consuming.

Although a program that predicts theoretical fragmentation patterns such as PFIA may assist in manual annotation of cyclic peptides by providing all possible b ions,[41] MS-CPA is capable of direct annotation of the actual input cyclic peptide MS spectra and is also the first program that take into account the fragments that are a result of sequence-scrambling fragmentation pathways.

To improve our understanding of the fragmentation behavior of cyclic peptides we have developed a program that readily annotates a mass spectrum resulting from the collision-induced dissociation of cyclic peptides. In addition, we have created a user-friendly web interface so that other scientist that are non computer experts can easily use it to annotate their tandem mass spectra of cyclic peptides.

Using this program, we observed that much of the spectral intensity of a MS^2 mass spectra of a cyclic peptide could not be explained. Upon further analysis, we

realized that unanticipated fragmentation pathways were involved in cyclic peptides when the standard fragmentation rules were applied. The data suggested those unanticipated fragments resulted in scrambling of the sequence. These unusual fragments were first described by Harrisons et al., as nondirect sequence (NDS) ions based on the scrambling of the original peptide sequence in contrast to the direct sequence (DS) ions derived from typical fragmentation pathways.[30] While initially surprising to the authors that NDS are observed, the mechanistic details toward the formation of NDS ions have recently been described in detail.[34] We have included NDS in our annotations. Therefore, our program, MS-CPA, not only provides evidence for the existence of these NDS ions but also enables quantitative analysis of the spectral abundance that match to DS and NDS ions.

In order to demonstrate the utility of this program, we have not only applied it to the representative testing peptides, seglitide and the tyrocidines, but also used it to confirm the sequence of two newly discovered natural products, desmethoxymajusculamide C (DMMC) and dudawalamide A, both isolated from marine cyanobacteria *Lyngbya majuscule* (**Figure 1.1**). In addition, the program was used to verify the structure of desprenylcyclomarin C, a natural product isolated from a prenyltransferase mutant of the marine bacteria *Salinispora arenicola CNS-205*. This marine natural product could not be isolated in sufficient quantities to confirm its structure by NMR; therefore, this program was critical in the confirmation of its structure. Finally, during these studies we discovered three additional dehydrated cyclomarin analogues and used our program to localize the site of dehydration.
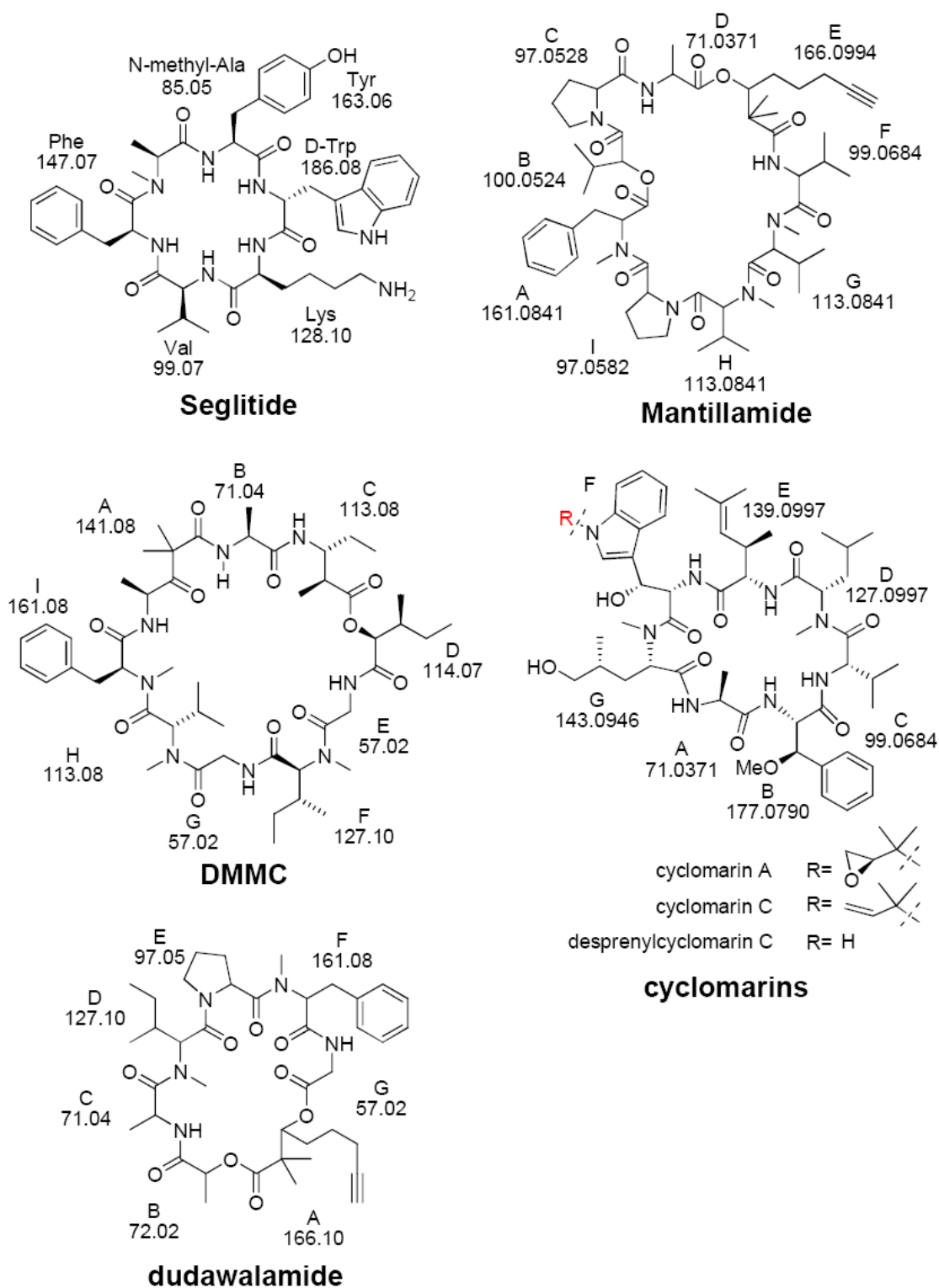
Figure 1.1: Structures of cyclic peptides discussed in this paper.

1.2 RESULTS AND DISCUSSION

1.2.1 Complexity of Cyclic Peptide Fragmentation

Because so many researchers work with cyclic peptides, the annotation of tandem mass spectra from cyclic peptides is important. The annotation, however, of tandem mass spectra of cyclic peptides is often difficult for mass spectrometrists and natural product scientists alike. The difficulty in the annotation of cyclic peptides arises from the nature of cyclic peptides itself. A cyclic peptide with n amino acid residues, theoretically, will yield n series of b ions but not any y ions.[32] If there are other ions such as a ions, internal fragments, and small neutral losses such as H_2O and NH_3, this complexity increases significantly. Therefore, it is difficult to annotate each and every ion in the spectrum of cyclic peptides and thus becomes an informatics problem. To overcome some of the complexity in the annotation of these peptides, we have developed a program that assists in the annotation of tandem mass spectrometry data based on input amino acid values and an experimental tandem mass spectrometric data set in .dta and .mzXML formats.

While we have presented, at a conference, that *de novo* sequencing of these nonribosomal peptides can be accomplished with near "perfect" mass spectral data sets using spectral alignments and a combination of de novo and database searching algorithms,[42] it quickly became clear that when we applied our first generation *de novo* sequencing algorithms to "nonperfect" mass spectrometry data sets typically encountered with more complex nonribosomally encoded peptides or symmetric cyclic peptides that these algorithms often identified a slightly different sequence. To improve the *de novo* sequencing algorithms that can be used to confirm the structures of isolated natural products, we need to improve our understanding of the resulting ions from a tandem mass

spectrometry experiment. This is, in particular, important when it comes to complex cyclic peptides.

1.2.2   Cyclic Peptide Annotation Program

To aid in the sequencing as well as to improve our understanding of the fragmentation behavior of cyclic peptides of nonribosomal origin, we developed a program named the MS-Cyclic Peptide Annotation program (MS-CPA) that readily annotates a mass spectrum resulting from the CID of a cyclic peptide. In particular, this program annotates b ions, a ions (losses of CO), and b_0 ions (losses of H_2O). However, y ions are not included, because cyclic peptides do not yield such ions.[32] The annotation program started as a Python script to mark b, a, and b_0 ions given a mass spectrum. The current implementation is capable of handling .dta and .mzXML file formats as this data format is becoming the standard format for reporting or depositing mass spectra and/or proteomic data sets[43,44] as spectrum inputs. For the reason that many cyclic peptides contain unusual or modified amino acids, we leave the freedom for users to input the amino acid masses manually. There is no size limitation to the mass of the amino acid that can be manually imported. Additionally, default standard amino acids masses are provided. Finally, the amino acid sequence is specified by the user in the order that they are encountered in the peptide.

For example, seglitide has a methylation on the nitrogen of alanine. This is a nonstandard amino acid; therefore, we can input 85.05280 for methyl-alanine rather than the alanine mass 71.03711.

In addition, once it was recognized that even for mass spectrometrically well behaved peptides, a large proportion of the ion intensity remained unexplained, and the capabilities of this program was expanded to consider neutral amino acid losses from the b ion ladder as well as evaluation of possible rearrangements based on the series of masses initially given. The current program has thousands of lines of code to annotate a spectrum for the generation of a graphical and tabular output on a web server.

We have made the MS-CPA program publicly available as a web tool at the UCSD center for computational mass spectrometry (http://lol.ucsd.edu/ms-cpa_v1/Input.py) and have also included a tutorial in the Experimental Section. In this paper, we demonstrate the utility of MS-CPA for the characterization of the cyclic peptides shown in **Figure 1.1**. The cyclic peptides in **Figure 1.1** are representative of the type of cyclic peptides encountered in drug screening programs.

1.2.3   Pre-analysis Data Processing of the Tandem Mass Spectrometry Input File

While the main code for this program is thousands of lines, the main challenge in the annotation process is actually the generation of a spectrum in which most peaks can be interpreted. Because of the great variance of experimental settings, instrumentation, and fragmentation properties of the compounds, preprocessing steps of the data that is required for each compound and experiment can vary a lot. To this end, we implemented a series of filters to enhance the signal-to-noise ratio of the experimental spectrum. Our current implementation regarding preprocessing includes centroid filtering, rank filtering, water filtering, isotope filtering, peak tolerance, and symmetrization. These preprocessing steps are detailed described in the Experimental Section, but the user can also choose not

to carry out any preprocessing. Given that noise peaks are unavoidable in a real mass spectrometry experiment, the main goal of the filters is to eliminate ions that are likely noise or ions that are uninformative without losing the important data. In addition, this gives the users of this program the flexibility to annotate their spectra in a manner they prefer. For example, the user may only want to annotate the top 10 ions in the spectrum. This is possible with this interface. In addition it is possible to annotate unfiltered spectra but results in a much longer computational processing time. In many cases, in natural product research, the samples are available in limited quantities or the peptide does not fragment well and therefore it is not always possible to produce the best mass spectra. The filters will allow us to work with these spectra, instead of repeating the experiment, which might not be possible in real world drug discovery applications where there is often a limited supply.

1.2.4   Nomenclature Used in This Paper

For discussion purposes of the results in this paper, we have adapted the nomenclature forwarded by Ngoka and Gross to describe the cyclic peptides in this paper.[45] The nomenclature developed by Ngoka and Gross describes the ions with a four-part descriptor with the general formula xnJZ, where "x" is the designation for the type of ion (b, a, etc.) and n is the number of amino acid residues that makes up the ion. J and Z are the one-letter codes for the two amino acid residues connecting the backbone amide bond, J-Z, which is broken to form the linear ion. J is the N-terminal amino acid residue and Z is the C-terminal amino acid residue. To illustrate the nomenclature, we use seglitide, a six-amino acid residue cyclic peptide illustrated in **Scheme 1.1** as an example.

In seglitide and tyrocidines, the one letter amino acid abbreviation was used to represent each residue, while in other compounds we assigned letters in order of their sequence using the standard alphabet since they contained too many modified residues. For example, in this paper we describe DMMC for which 6 out of 9 are modified or nonstandard amino acids, while dudawalamide A has 4 out of 7 that are nonstandard, mantillamide has 5 out of 9, and cyclomarins have 5 out of 7 (**Figure 1.1**). Because the alanine in seglitide has methylation in the nitrogen position, we use A′ to represent this methylated residue. Seglitide using this nomenclature would likely undergo random ring-openings following by the bn -> bn-1 pathway[45] resulting in the formation of 6 different series of b ions (**Scheme 1.1**).
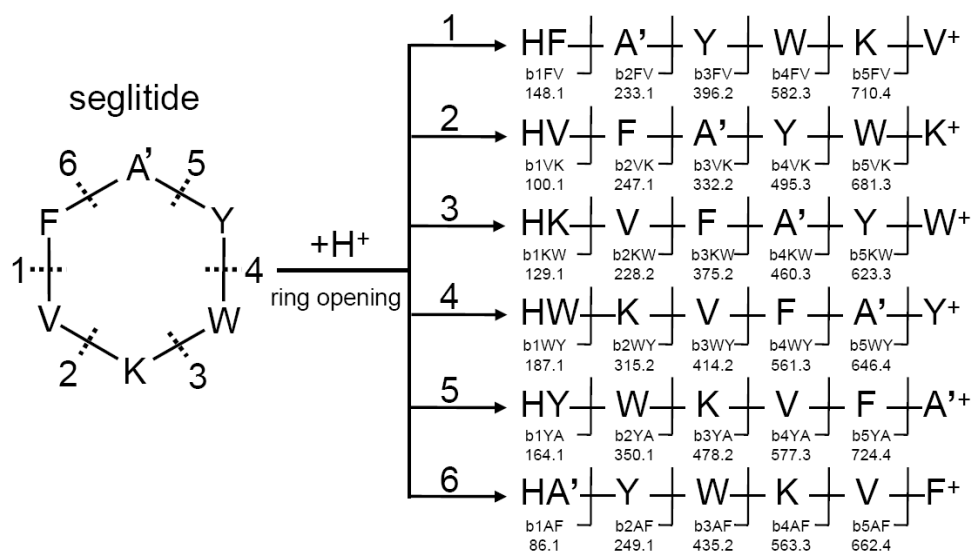
Scheme 1.1: Sequences of *b* ions from the fragmentation of seglitide. According to the conventional pathway for fragmentation of cyclic peptides, seglitide first undergoes random ring opening at each amide bond, yielding 6 different linear peptides. Sequential C-terminal amino acid cleavage results in six series of ions, for a total 30 *b* ions.

1.2.5 Cyclic Peptide Annotation Program Demonstration: Seglitide

We first illustrate the application and utility of MS-CPA using a simple cyclic peptide, seglitide, a somatostatin receptor antagonist consisting of six amino acids, and described the results using the nomenclature defined above (**Figure 1.1**). Seglitide was analyzed by Fourier-transform ion cyclotron resonance mass spectrometry (FTICR MS). A singly protonated ion was observed at 808.4247 Da, which is within 3 ppm of the theoretical mass of seglitide (808.4272 Da). This ion was subjected to CID in a linear ion trap, and the product ions were again analyzed by FTICR MS (**Figure 1.2**).

The resulting MS2 spectra were then analyzed by MS-CPA. The spectrum was subject to standard filtering procedures to increase the signal-to-noise ratio. First, because the raw spectrum was collected in profile mode, only the top peak was retained in a window of ±0.05 Da. Second, the top 200 most intense peaks were retained. Lastly, isotopic and water-loss peaks were filtered out, yielding 146 final peaks. As shown in **Figure 1.3**, the output of MS-CPA includes input residues and the parent mass that is obtained as user input or directly obtained from the input. dta or. mzXML file (A), summary of input filtering parameters and resulting ions counts (B), quantitative statistics of cleavage and total explainable ion intensity (C), a spectrum with color-coded matches (b ions are showed in red; water loss are green; a ions are cyan; NDSs are blue; unannotated ions are yellow) (D), a plot of mass errors of the annotated ions (E), and a list of matched fragment ions in tabular format (E). For seglitide, the MS-CPA output indicates that 28 out of the 30 possible b ions were matched to observed masses. The explainable ion intensity of the b ions combined with possible a ions and loss-of-water ions was 71.5% of the total ion intensity. The absolute difference between the calculated

and the experimental masses was less than 0.004 Da. Among the annotations, some of the ions with high intensity contained water loss even though there was no serine or threonine in the sequence. In addition, masses corresponding to addition of 28 Da (plus CO) were observed. These ions were not expected, thus we subjected these ions to additional rounds of tandem mass spectrometry (MS^3 and MS^4) to verify if the annotations were real or not. With these additional rounds of fragmentation, the authenticity of MS-CPA annotations was verified and these ions are indeed correctly annotated. Although the mechanisms behind the formation of these unusual fragments are still elusive, MS-CPA enabled us to discover the existence of these ions.
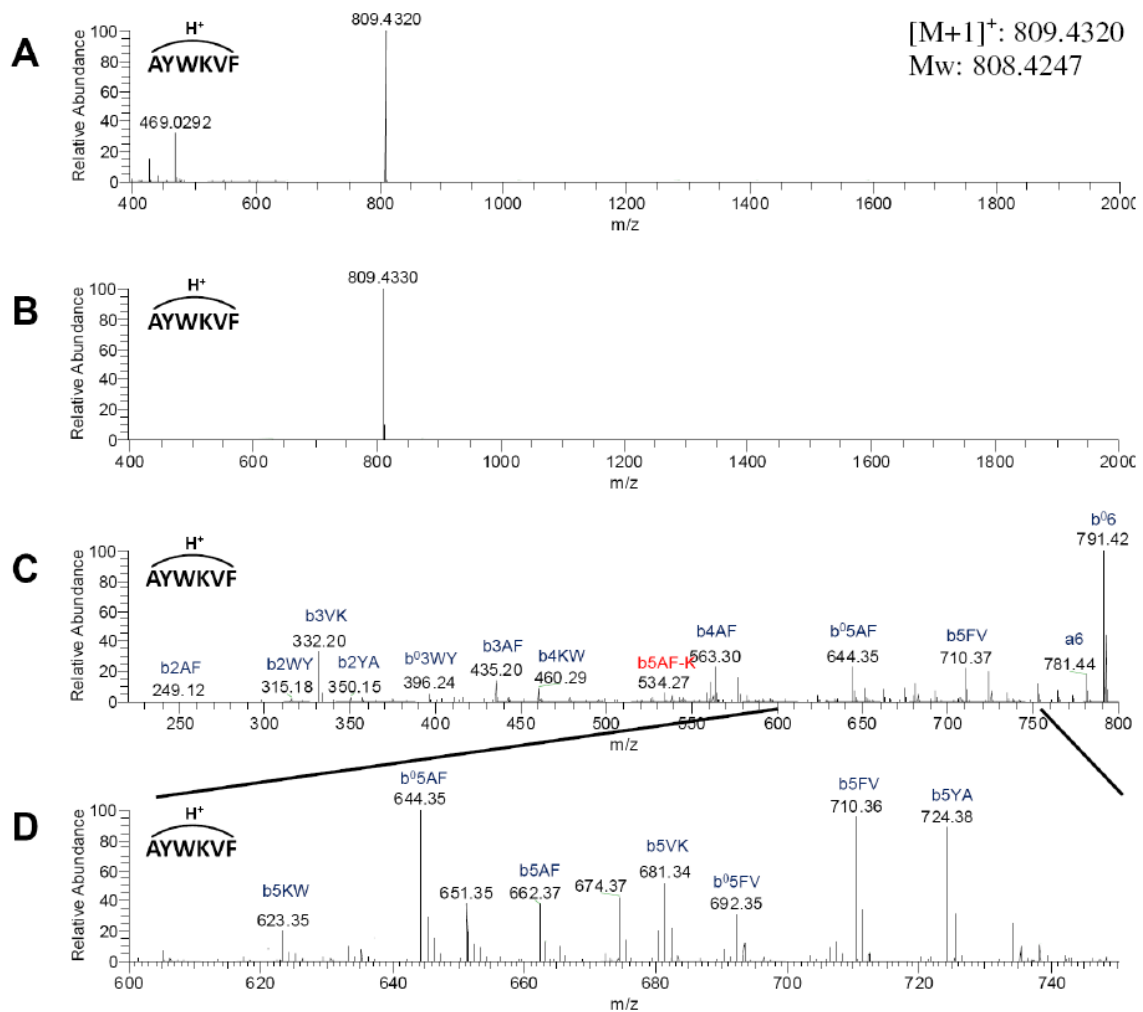
Figure 1.2: Seglitide MS and MS^2 spectrum. MS and MS^2 spectrum were collected by ESI-LTQ-FTICR MS. A: Broadband spectrum. B: Spectra obtained with a isolation window set for seglitide parent ion (M+H)$^+$ C: MS^2 spectrum of seglitide. D: Zoom in spectrum of 600~750 m/z region.

A

Parameter	Value
Peak tolerance	0.006
Amino acids	A = 85.053 Y = 163.063 W = 186.079 K = 128.095 V = 99.068 F = 147.068
Modifications	-18.011 +18.011 -27.995 +27.995

C

Description	Cuts	Fraction Int
A * Y * W * K * V * F	5/5	0.1418
F * A * Y * W * K * V	5/5	0.1452
V * F * A * Y * W * K	5/5	0.1322
K * V * F * A * Y * W	5/5	0.0852
W * K * V * F * A . Y	4/5	0.0557
Y . W * K * V * F * A	4/5	0.1545
Total explainable	28/30	0.7147
Total explainable by normal and NDS ions		0.8206

Experimental Parent Mass	808.4272
Theoretical Parent Mass	808.4272

B

Filtering Method	Peak Count
Original	27912
centroid filtering width = 0.10 top = 1	547
Take top intensity peaks	200
Isotope filtering (optional)	146
Symmetrization	270

F

top	mass	intensity	percent	annotation
1	332.197	35908.24	9.538	{332.197} VFA
2	644.354	21833.30	5.799	{644.356} AYWKV[-18.01]
3	563.297	19337.46	5.136	{563.298} AYWK
4	710.365	15516.38	4.121	{710.367} FAYWK
5	435.202	14833.12	3.940	{435.203} AYW
6	577.313	13561.04	3.602	{577.314} YWKV
7	396.239	12899.44	3.426	{396.240} WKV[-18.01]
8	559.302	12740.19	3.384	{559.303} YWKV[-18.01]
9	724.380	12244.54	3.252	{724.382} YWKVF
10	764.412	9253.73	2.458	{764.413} AYWKVF[-27.99]-[17.03]
11	460.291	8626.17	2.291	{460.292} KVFA
12	662.365	8156.27	2.166	{662.367} AYWKV
13	692.354	7673.47	2.038	{692.356} FAYWK[-18.01]
14	752.375	7271.79	1.932	{752.377} YWKVF[+27.99]
15	681.339	6730.78	1.788	{681.340} VFAYW
16	674.365	6699.22	1.779	{674.367} WKVFA[+27.99]
17	651.349	6295.16	1.672	{651.351} KVFAY[+27.99]
18	534.270	5998.01	1.593	{534.272} AYWV
19	734.365	4694.99	1.247	
20	350.150	4640.29	1.233	{350.150} YW
21	582.270	4379.05	1.163	{582.272} FAYW
22	544.291	3995.58	1.061	
23	527.297	3928.30	1.043	{527.298} AWKV[+27.99]
24	442.281	3905.33	1.037	{442.282} KVFA[-18.01]
25	623.354	3787.63	1.006	{623.356} KVFAY
26	357.228	3757.14	0.998	{357.229} KVF[-18.01]
27	665.344	3747.00	0.995	
28	416.229	3646.87	0.969	
29	315.181	3521.48	0.935	{315.182} WK
30	538.302	3083.50	0.819	{538.303} YKVF

Figure 1.3: MS-CPA output from analysis of seglitide MS^2 data. MS-CPA input parameters summary (A,B), number of cleavages and explainable intensity. The * indicates an ion that cleaves here was annotated in the spectrum. (C), annotated spectrum (D), accuracy analysis (E), annotated ions list (unsymmetrized) (to save space, only the 30 most intense ions were displayed) (The annotation is the recommended arrangement for the provided input sequence and specified parameters. Users should verify the annotations manually.) (F). In the output spectrum and annotation list, the *b* ions are showed in red; H$_2$O loss are green; *a* ions are cyan; NDS's are blue; unannotated ions are yellow in the spectrum and unlisted in the table. Symmetric ions are not shown in D or F.

1.2.6   Observation of NonDirect Sequence Ions in Seglitide

Because more than 28% of the ion intensity remained unexplained, we explored the nature and significance of the remaining ion intensity. Because these data were acquired with high resolution, the molecular mass of each ion could be determined. First, we analyzed these for alternate combinations of amino acids that would result from peptide residues rearrangements. We found 58 such ions comprising roughly 10% of the total ion intensity.

Each of these scrambled sequence ions had mass errors within 0.004 Da, in agreement with all of the other masses we had annotated. The fact that so many of the ions could be explained by a rearrangement of the amino acid sequence is unlikely be coincidental or due to noise. In fact, some of these scrambled ions are of relatively high abundance. In seglitide, the most abundant NDS ion was up to 16% of the normalized ion intensity when the most intense ion was set to 100%. These kinds of scrambled sequence ions have previously been observed in peptides and described as nondirect sequence ions.[28,30-32,34-37] Because of their relatively high abundance, they are included into our annotation program MS-CPA. By their inclusion, the accountable signal intensity increases from 71.5% to 82.1%. Notably, some ions still remain unannotated, these ions are likely a result of sidechain fragmentations, unknown fragmentations, or noise inherently present in the mass spectrometry data set.

To confirm the presence of NDS ions from seglitide, the two most intense of these ions, AYWV and YKVF (b_{5AF-K}, b_{5YA-W}), and each b_5 ion (i.e., the parent ion minus one amino acid) were isolated and subjected to an additional round of CID. The b_5 ions were chosen for comparison and were anticipated to be linear by conventional fragmentation

pathways bx → bx-1.[46] Surprisingly, the MS^3 spectra indicated that none of these selected ions simply followed the conventional rules for fragmentation which state that cyclic peptides sequentially lose amino acid residues from the C-terminus after the initial ring opening event (**Figure 1.4**).[26] Instead, we observed a mixed series of b ions (**Scheme 1.1**) which suggest that the precursors for the MS^3 experiment are still cyclic. For example, if the b_5 ion FAYWK was of linear structure, only the b_{nFK} ion series should be present in the associated MS^3 spectrum (**Figure 1.4A**); however, we observed the relatively intense b_{nYA} and b_{2WY} ions. These additional ion fragments most likely originate from cyclic peptide precursors.
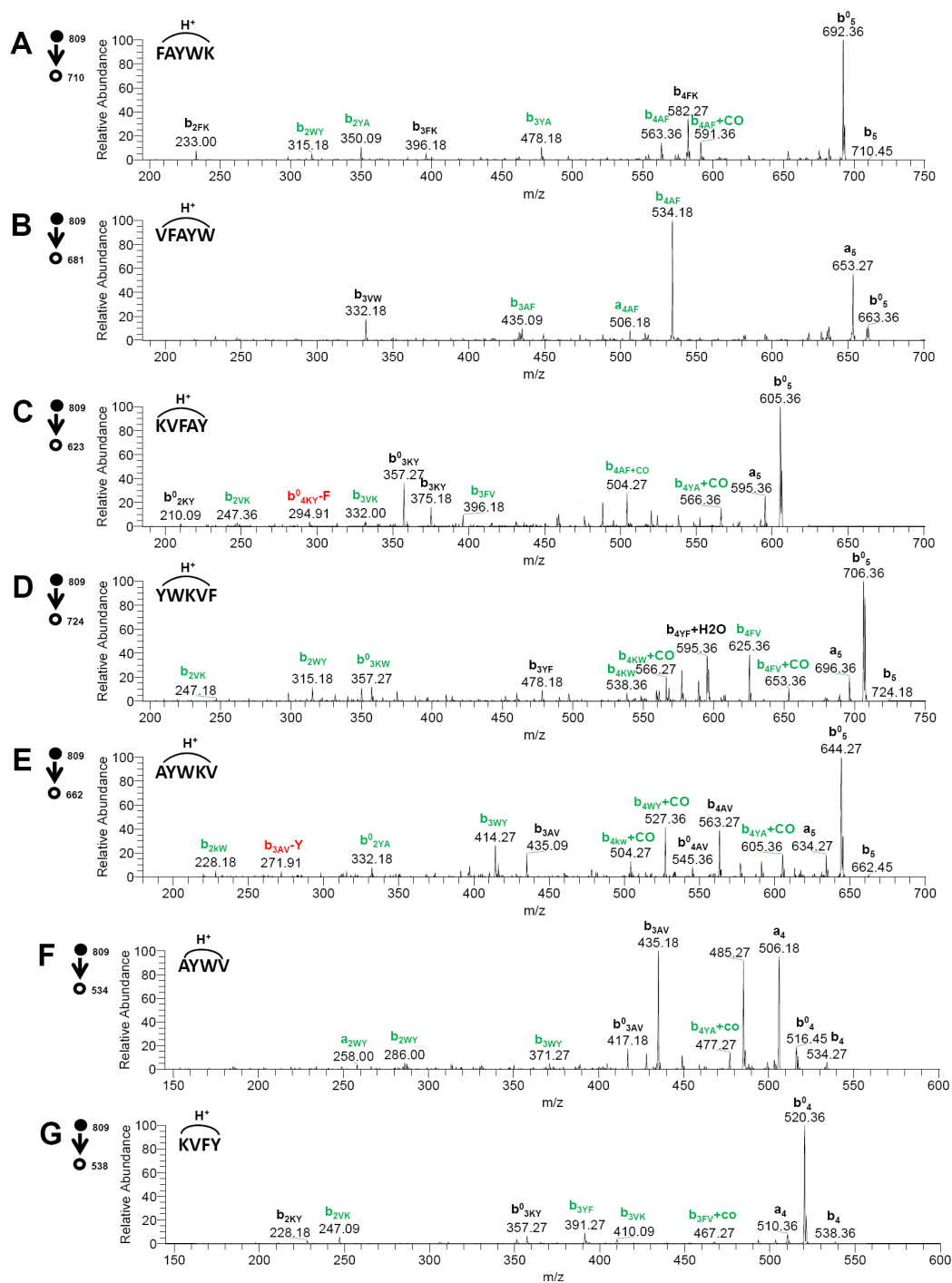
Figure 1.4: MS³ spectra of representative seglitide sequence ions. The presence of daughter *b*-ions from different linearized parent ions suggests that the parent ion is cyclic (as opposed to linear, as initially assumed). MS³ spectra were collected by ESI-LTQ MS. A-E: b_5 ions. F-G: top two NDS ions observed. Expected sequence ions of a linear peptide are showed in black color. Expected ions for cyclic peptide are showed in green color combined with black color ones. Red color represents for NDS ions.

To explore these NDS ions behavior, we first compared the total ion intensities explained by assuming a linear precursor with those explained by assuming a circular precursor. For example, CID on the ion b_{5KW} (KVFAY) would yield K, KV, KVF, KVFA, Y, AY, FAY, and VFAY fragments if the b_{5KW} ion was linear. However, if this ion was circular, we would observe 20 possible fragments. In the case of b_{5KW} (**Figure 1.4C**), the ions annotated as AYKV, FAYK, and YKVF show high intensity and are easily explained if the precursor ion is considered to be circular. In fact, 88% of total ion intensity can be explained by assuming a circular precursor, while only 42% can be explained by assuming a linear precursor. **Table 1.1** summarizes the analysis of the seven MS^2 ions that were subjected to additional CID and annotated as either linear or circular. Among these seven MS^2 ions, the only one that gave poor fragmentation is b_{5VK} (VFAYW), with 12 cleavages out of 20. However, this ion produced a very intense peak (b_{4AF}) that corresponds to loss of phenylalanine. This peak would not have been the most intense ion in the MS3 spectrum if the initial cyclic peptide had first undergone linearization and then eliminated the C-terminal residue (i.e., tryptophan) as predicted by conventional fragmentation rules.[26,45] While all of the foregoing results strongly support the cyclic nature of the MS^2 ions resulting from CID of seglitide, it is likely that a mixture of cyclic and linear forms ultimately contribute to the MS^3 spectrum.

Table 1.1: MS-CPA analysis[b] of the two most intense NDS ions and *b5* ions of seglitide.

	Linear annotation		Circular annotation		Number of critical fragments[a]
	cuts	explained intensity	cuts	explained intensity	
b_{5FV}	6/8	64.75%	13/20	89.24%	3
b_{5VK}	5/8	20.46%	12/20	88.82%	4
b_{5KW}	4/8	42.39%	11/20	88.36%	4
b_{5YA}	6/8	46.15%	14/20	92.54%	6
b_{5AF}	5/8	55.19%	8/20	82.01%	1
b_{5AF-K}	4/6	73.16%	7/12	85.38%	2
b_{5YA-W}	4/6	50.47%	8/12	76.81%	3
[a] Fragments that cover the linear breakpoint. [b] Results were analyzed by isotope removal, water removal, NH3 removal and window filtering with width 10, top 10, unsymmetric.					

Although the formation of these NDS ions have been recognized since 2003,[30] the actually mechanisms behind them are still a hot research topic. Several groups argued the importance of understanding this phenomenon in the development of *de novo* sequence programs. Therefore, a few mechanisms have been proposed to account for NDS ions.[28,30-32,34] The general consensus involves a cyclic intermediate occurring by recyclization. The presence of recyclized intermediates have been verified by Riba-Garcia and co-workers using ion-mobility MS.[36,37] The tendency of generating NDS ions was also studied under N-acetylation modification or various activation energy.[35] Recently, just after this current manuscript was submitted, a more thorough mechanism and pathway was published by Bleiholder et al., in which a sequence-scrambling fragmentation pathway was proposed describing the mechanism of NDS ions based on experimental and energetic calculations in agreement with the cyclic NDS ions we

observed.[34] Therefore, our program, MS-CPA, provides solid evidence showing the existence and abundance of these NDS ions with nonribosomally derived cyclic peptides.

1.2.7 Dependence of the Intensity of NDS Ions on Activation Time and Activation q

Because we anticipated that changing the activation time and energy would provide control of the intensities of NDS ions, we analyzed the effects of the different CID parameters on NDS ion abundance (**Figure 1.5**). Surprisingly, and somewhat unsatisfying, the amount of these NDS ions did not demonstrate a significantly change with increased activation time and energy (q). This phenomenon can be attributed to the fact that when the activation q increases, the m/z range of a frequency sweep decreases (**Figure 1.6**) because low m/z product ions start to lose stable trajectories as activation energy q is rising.[47,48] This is a well documented flaw with linear ion traps. Because of the large activation q, the fragment ions and NDS ions no longer fall within the acquired scan. Therefore, the cleavage coverage and NDS abundance decrease drastically (**Figure 1.5B**). In addition, there does not appear to be significantly added benefit from changing the activation time and activation q in the increased coverage when the spectra are merged. Unfortunately we do not have PQD (Plused-Q Dissociation) on our instrument that could partially overcome this limitation found with ion traps.[49,50] Although not tested in this work, as the authors do not have such instrumentation in their laboratories, it is anticipated that Q-TOFs and triple quadrupoles do not have this limitation.
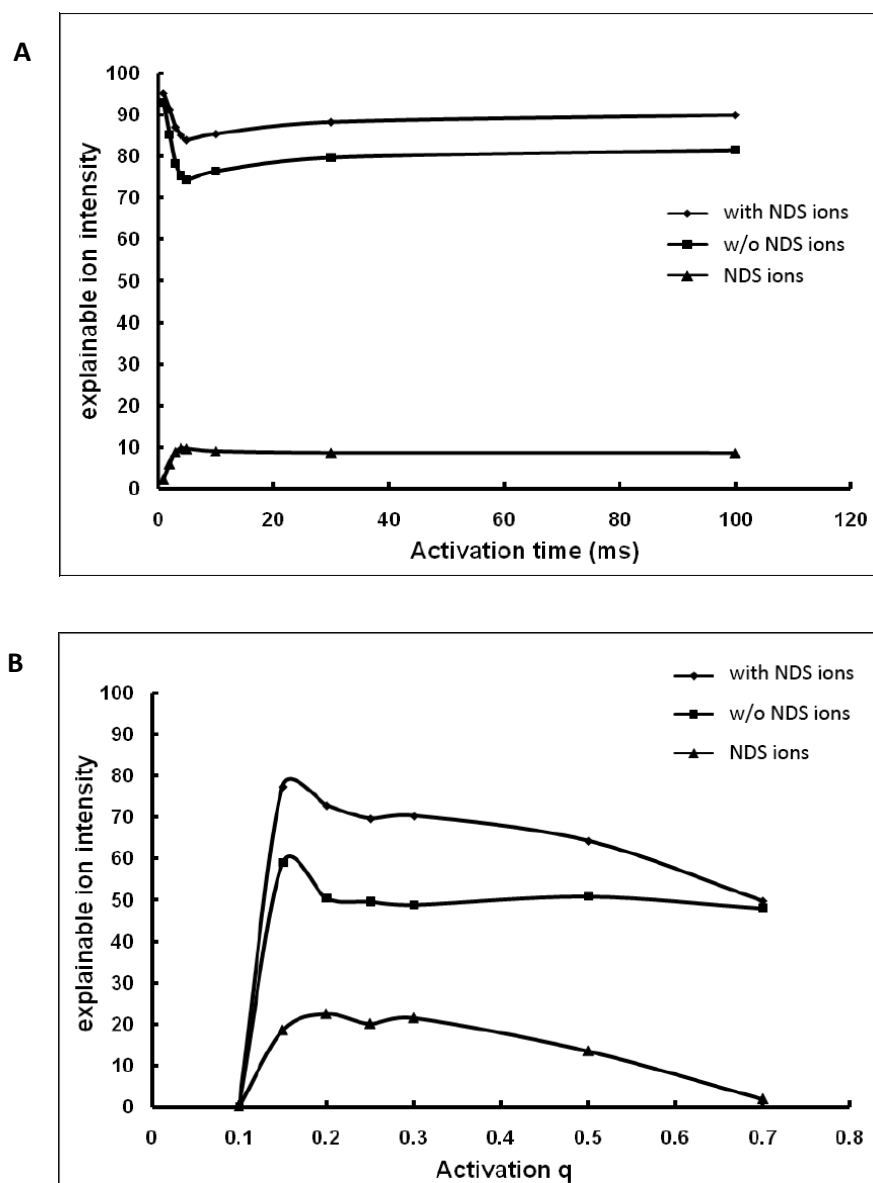
Figure 1.5: Effect of activation time and activation q on NDS ion formation. In activation time experiment (A), Seglitide MS^2 spectrum was collected using ESI-LTQ MS at different activation time (1-100ms), activation q was set as 0.25 (Thermo parameter) and subjected to MS-CPA analysis. The results showed before 5ms NDS ions formation increase correlated to activation time, and then reach a plateau, suggesting fragmentation can be done in 5ms. In activation q experiment (B), Seglitide MS^2 spectrum was collected using ESI-LTQ MS at different activation q (0.05-0.7 Thermo parameter), while activation time was set as 100ms. Resulting spectra were again subjected to MS-CPA analysis. The results showed below q 0.15 NDS ions formation increase correlated to activation q, and fall drastically due to the collecting m/z range decrease and thus sequence ions including NDS ions fall out the measuring region.
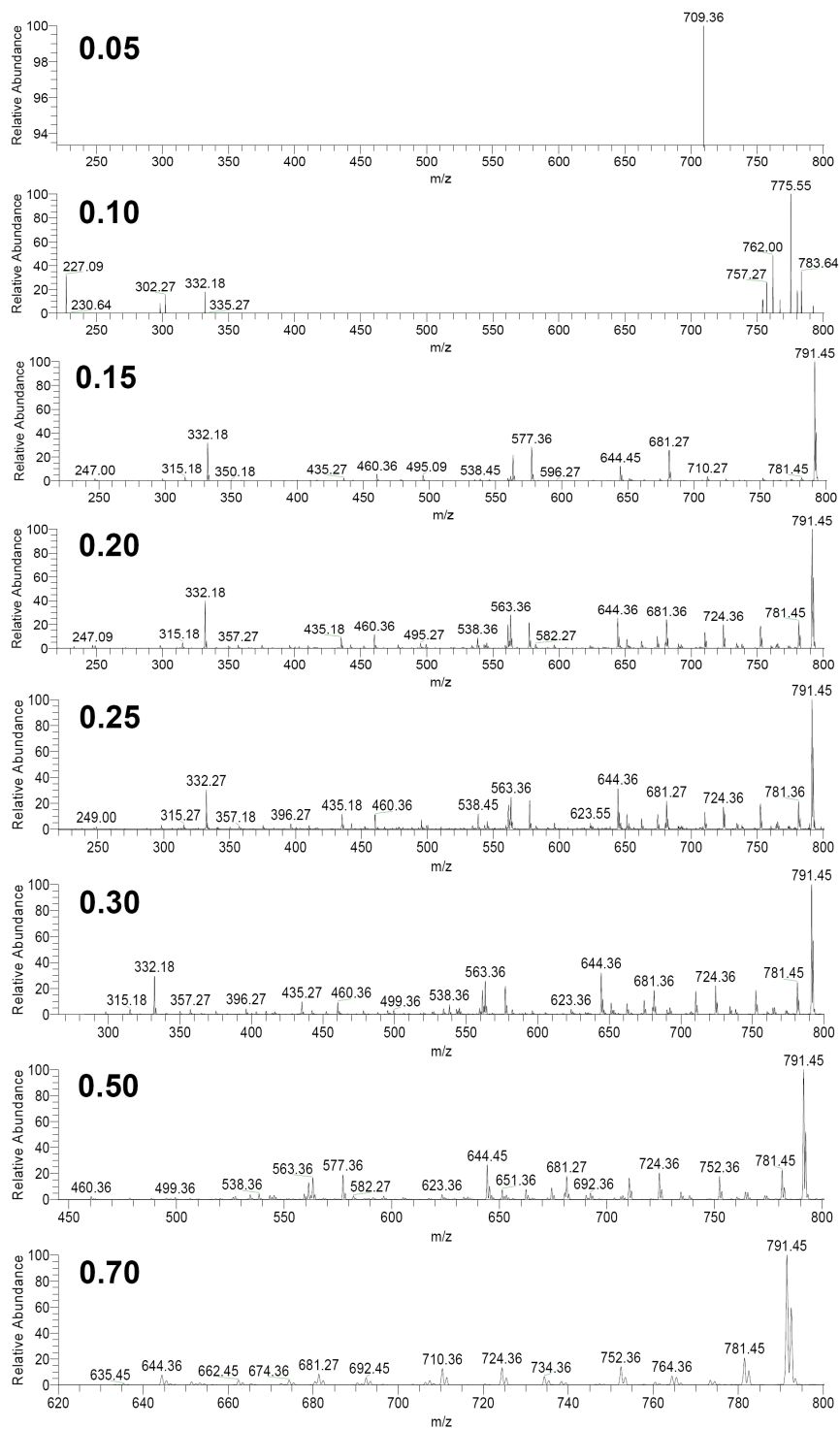
Figure 1.6: Effect of activation energy q on fragmentation. Seglitide MS^2 spectra were collected using ESI-LTQ MS at different activation q (0.05-0.7 Thermo parameter) with activation time set as 100ms. Spectra showed measuring m/z regions drastically narrow down as activation q increase over 0.3.

1.2.8 Capability of MS-CPA in Analyzing an Antibiotic Mixture

In addition to seglitide, we investigated the antibiotic mixture tyrothricin, which contains more than 28 different compounds and is readily available commercially due to its clinical utility as a typical antibiotic. Some of these compounds, individually called tyrocidines, are known to be cyclic peptides.[51] We used MS-CPA to analyze several ions from this mixture (**Figure 1.7**, **Table 1.2**). In the case of tyrocidine A, the program successfully annotated 74 b ions out of 90 possible. In contrast, only 17 b ions were identified through manual annotation of tandem mass spectra from tyrocidine A, despite this being one of the most thorough studies of cyclic peptides available to date in the literature demonstrating a significant advantage of spectra using our approach.[52]
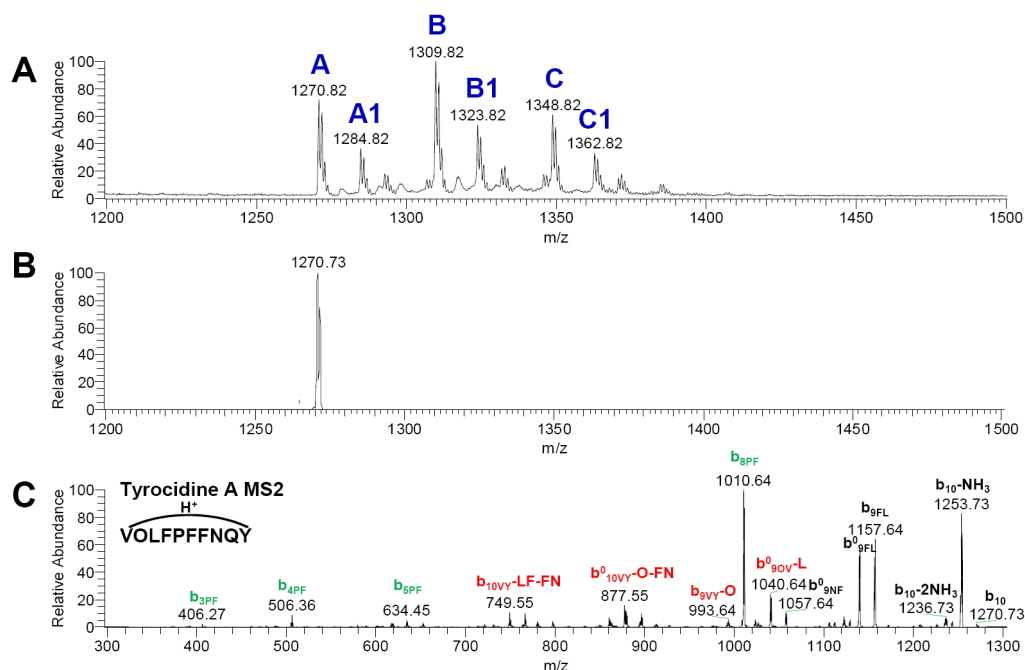
Figure 1.7: Tyrocidines MS and MS^2 spectra. MS and MS^2 spectra were collected by ESI-LTQ MS. A: Broadband spectrum showed different species of tyrocidines in tyrothricin antibiotic mixture. B: Isolation of tyrocidine A (protonated form) C: MS^2 spectrum of tyrocidine A.

Table 1.2: Summary of tyrocidines sequence, molecule weight and MS-CPA results collected by low resolution MS (LTQ).

|	sequence	obs Mw	cal Mw	ppm	cleavage	explainable ion intensity w/o NDS ions	with NDS ion
A	cyclo [VOLFPFFNQY]	1269.72	1269.65	52	74/90	59.37%	85.73%
A1	cyclo [VKLFPFFNQY]	1283.72	1283.63	39	80/90	61.18%	77.36%
B	cyclo [VOLFPWFNQY]	1308.72	1308.65	42	66/90	53.29%	84.90%
B1	cyclo [VKLFPWFNQY]	1322.72	1322.68	29	74/90	59.60%	75.84%
C	cyclo [VOLFPWWNQY]	1347.72	1347.67	32	60/90	51.80%	85.83%
C1	cyclo [VKLFPWWNQY]	1361.72	1361.69	18	76/90	54.63%	77.73%

1.2.9 Using MS-CPA to Annotate Cyclic Peptides Containing Nonstandard Subunits.

Seglitide and the tyrocidines have a uniform peptidic backbone with standard amino acids. However, many nonribosomal cyclic peptides are cyclized via lactone formation and include nonstandard amino acids.[53] Theoretical calculations suggested that cyclic peptides favor a lactone bond as the initial ring-opening site, and also the fragmentation pathway of cyclic peptides differs when lactone bond(s) were involved.[29] It is therefore important to establish how these other structural features impact the fragmentation data and the results analyzed by the MS-CPA program. Thus, we analyzed several nonribosomal cyclic peptide natural products containing lactone linkages and nonstandard amino acids by tandem MS followed by MS-CAP (**Table 1.3** and **Figures 1.8-1.11**). These included three marine cyanobacterial depsipeptides: desmethoxymajusculamide C (DMMC), mantillamide, and dudawalamide A, all three of which were isolated because of their biological activity to cancer cells or malaria parasites (**Figure 1.1**).[54-56]

Analysis of DMMC by MS-CPA uncovered 36 of the 72 b ions expected from standard fragmentation. Including NDS ions, the proportion of explained total ion intensity increased from 71.1% to 78.3%. Similar results were obtained for mantillamide. These data indicate that nonstandard residues and ester linkages do not diminish the program's ability to insightfully annotate a tandem mass spectrum.

Dudawalamide A was isolated from the marine cyanobacterium *Lyngbya majuscula*, and its structure was determined by NMR methods. A full report on the structure and bioactivity of dudawalamide will be published elsewhere.[55] A high-resolution MS2 spectrum of this compound was submitted to MS-CPA for annotation. The program was also provided with the masses of the dudawalamide subunits determined by NMR (**Figure 1.10**). The fragmentation behavior of dudawalamide, also a lactone, was found to be very different from the fragmentation behavior of mantillamide and DMMC. Although 96.0% of the total ion intensity was explained by b ions with absolute mass errors smaller than 0.008 Da, only 18 of the predicted 42 b ions were identified by the program. Thus, a high proportion of total ion intensity was accounted for by a small fraction of the expected b ions. This phenomenon can be explained by the presence of labile connections between residues within dudawalamide. Such weak connections are represented in normal peptides by amides N-terminal to prolines, amides C-terminal to Asp and Glu, or amides involving tertiary amines.[28] Three such linkages are present in dudawalamide: one at the N-terminus of proline and the other two at the N-termini of the *N*-methylated phenylalanine and the *N*-methylated isoleucine. Because of these three labile connections, the fragmentation of dudawalamide produced only a few

ions, which were consistent with the known structure of dudawalamide but provided little sequence coverage.

Lastly, we used MS-CPA to investigate the structures of cyclomarin A, cyclomarin C, and desprenylcyclomarin C. The natural products cyclomarin A and C were originally isolated, based on their strong anti-inflammatory activity, from the marine bacterium *Streptomyces* sp. CNB-982.[57] Subsequently, desprenylcyclomarin C was isolated from a prenyltransferase mutant of *Salinispora arenicola* CNS-205 but could not be produced in amounts sufficient to enable structural characterization by NMR.[20] We therefore subjected all three cyclomarins to mass spectrometry and acquired MS2 spectra of each analogue. The broadband mass spectra of each of these cyclomarins showed a protonated ion species and a even much more stronger species corresponding to dehydrated forms (**Figure 1.11 A,D,G**), providing evidence that these natural products are prone to water loss. The MS2 spectra of both the protonated and dehydrated forms of each cyclomarin analogue were collected and subjected to MS-CPA.

Analysis by MS-CPA consistently revealed the presence of strong b5GF and b4AG ions the MS2 spectra in all of these cyclomarin species (**Table 1.3, Figure 1.11**), thus confirming that desprenylcyclomarin C is structurally related to cyclomarin A and C. Overall, these analysis of cyclomarins identified from 8 to 16 b ions out of 34 possible b ions. The fraction of explained total ion intensity ranged from 37.0 to 50.3% when NDS ions were excluded and from 47.2 to 72.5% when NDS ions were included. On the other hand, this fraction was much higher for the dehydrated forms of cyclomarins, ranging from 72.2 to 79.1% without NDS ions and from 76.4 to 91.8% with NDS ions. In addition, we have successfully localized the dehydration site to the tryptophan-derived

residue. Because cyclomarins are so prone to dehydration, it is possible that this is the form that provides its anti-inflammatory activity. The most likely path leading to dehydration is the formation of an imine on the tryptophan residue, yielding a conjugated system upon loss of water. These examples highlight the usefulness of MS-CPA to assist in the structural characterization of cyclic nonribosomally encoded natural products even when limited quantities are available.

Table 1.3: Summary of MS-CPA analysis of cyclic peptide natural products discussed in text.

name	cuts	explainable ion intensity	
		w/o NDS ions	with NDS ions
DMMC	36/72	71.10%	78.30%
Dudawalamide A	18/42	96.00%	97.30%
Mantillamide	34/72	65.21%	81.99%
Cyclomarin A	12/42	50.34%	72.48%
Cyclomarin C	16/42	36.98%	47.20%
Des[a]-cyclomarin C	8/42	46.74%	55.48%
Dehy[b] Cyclomarin A	16/42	73.79%	87.88%
Dehy[b] Cyclomarin C	12/42	76.35%	91.83%
Dehy[b] Des[a]-cyclomarin C	12/42	72.16%	79.09%

[a] Des: Desprenyl. [b] Dehy: dehydrated. [c] For the cyclomarins, manual annotations for ions reflecting loss of MeOH were included in the calculations of explainable ion intensity.
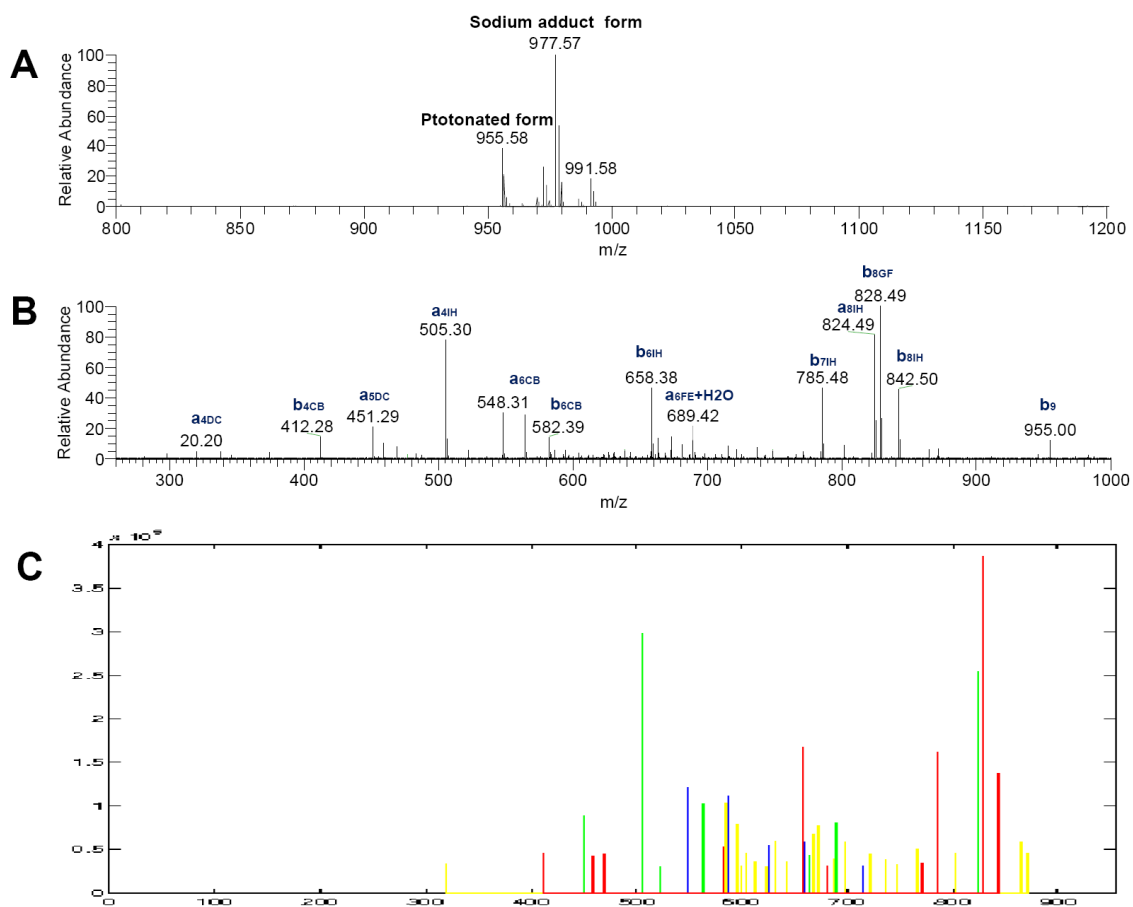
Figure 1.8: DMMC MS and MS^2 spectra. MS and MS^2 spectra were collected by ESI-LTQ-FTICR MS. A: Boardband DMMC spectrum. B: MS^2 spectrum of DMMC protonated form (954.57 m/z). C: MS-CPA output of annotated DMMC spectrum.
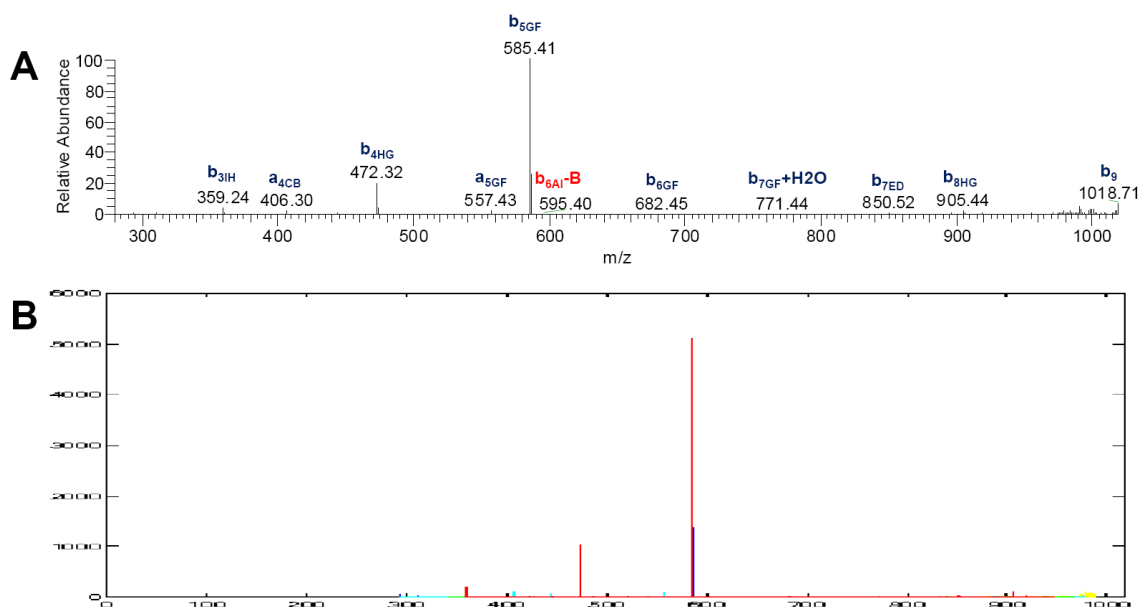
Figure 1.9: Mantillamide MS and MS² spectra. MS and MS² spectra were collected by ESI-LTQ MS. A: MS² spectrum of mantillamide A protonated form. B: MS-CPA output of annotated Mantillamide A spectrum.
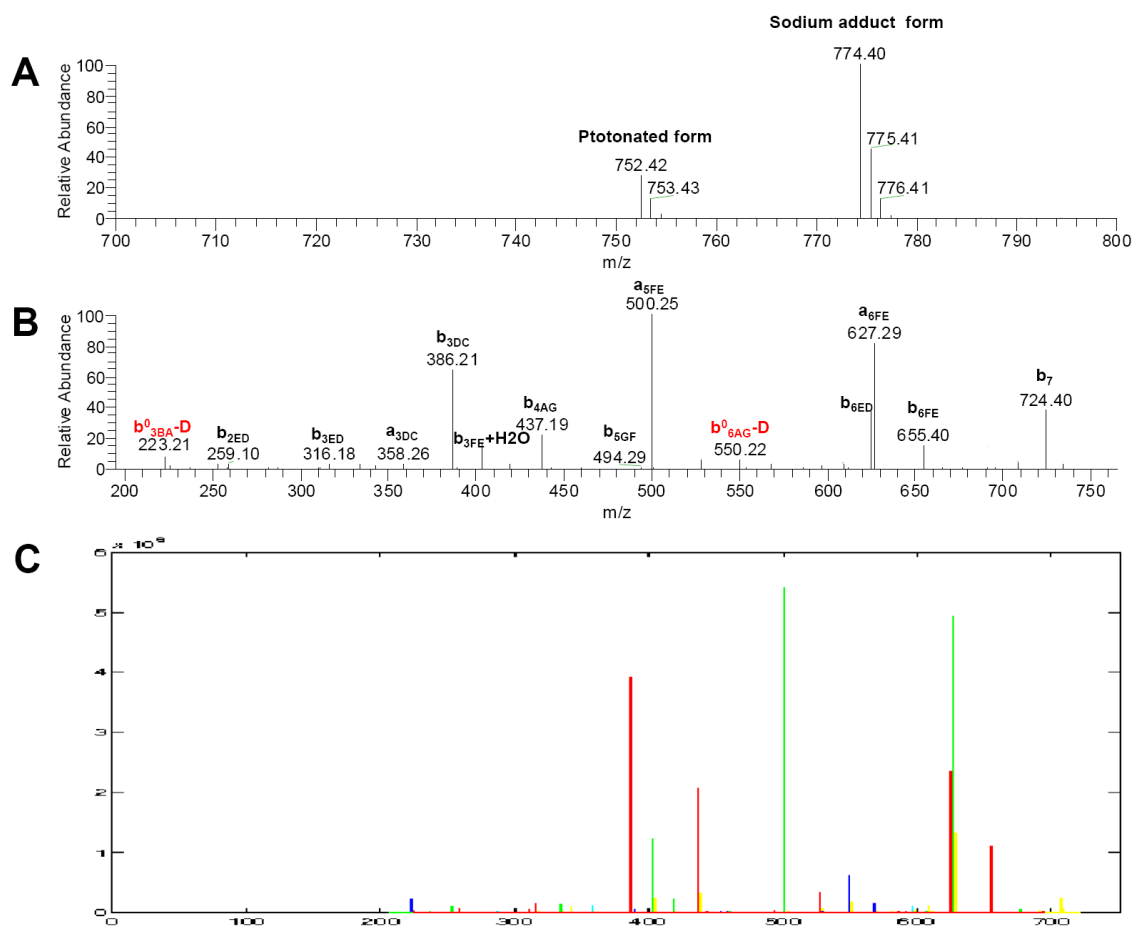
Figure 1.10: Dudawalamide A MS and MS² spectra. MS and MS² spectra were collected by ESI-LTQ-Orbitrap MS. A: Broadband spectrum of Dudawalamide A. B: MS² spectrum of dudawalamide A protonated form (752.42 m/z). C: MS-CPA output of annotated dudawalamide A spectrum.
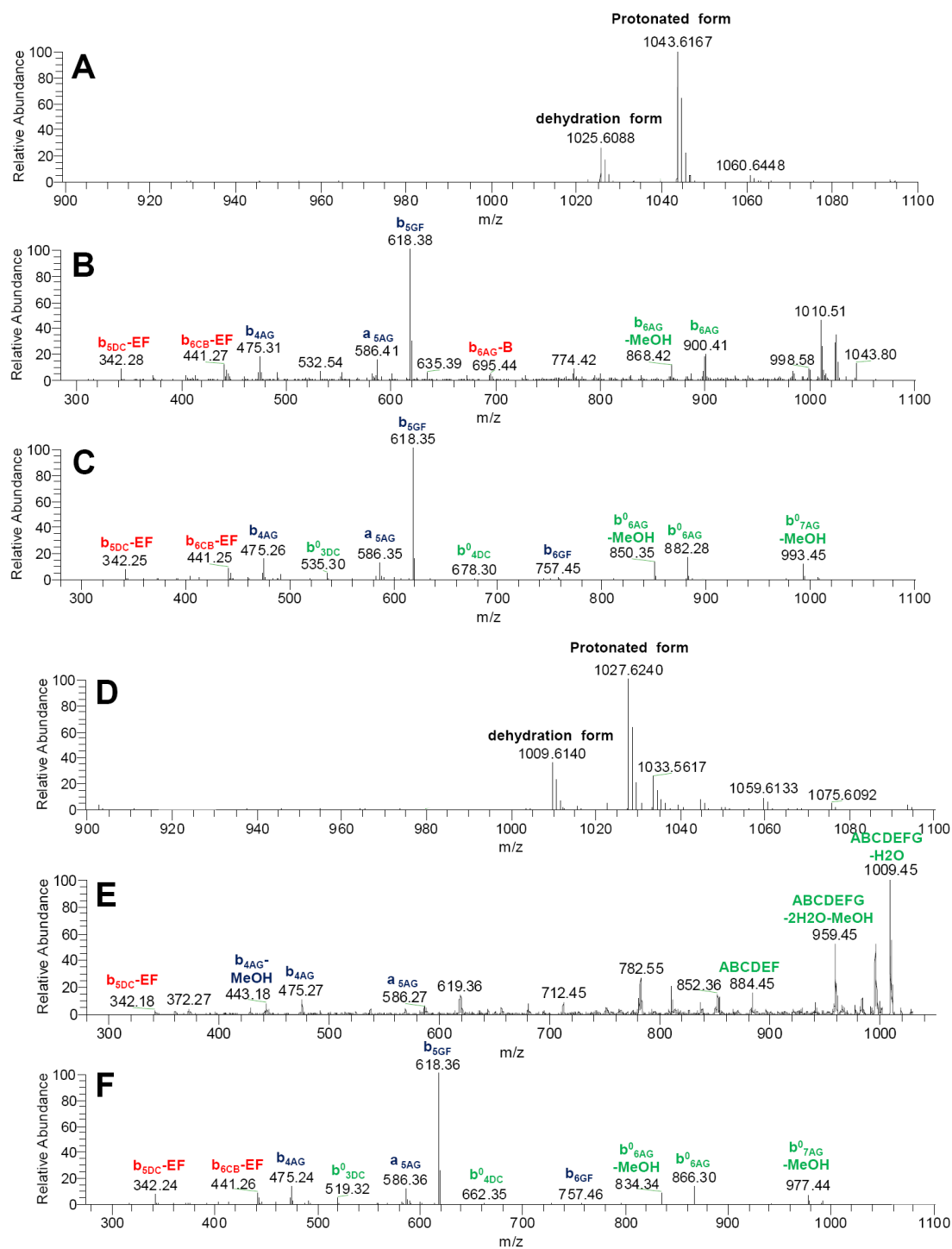
Figure 1.11: Cyclomarin MS and MS^2 spectra. Cyclomarins MS spectra were collected by ESI-LTQ-FTICR MS. MS^2 spectrum were collected by ESI-LTQ MS. A-C: cyclomarin A spectrums; D-F: cyclomarin B spectrums; G-I: desprenylcyclomarin C spectrums. Broadband (A, D, G); MS^2 of cyclomarin protonated form (B, E, H). MS^2 spectrums of cyclomarin dehydration form (C, F, I).
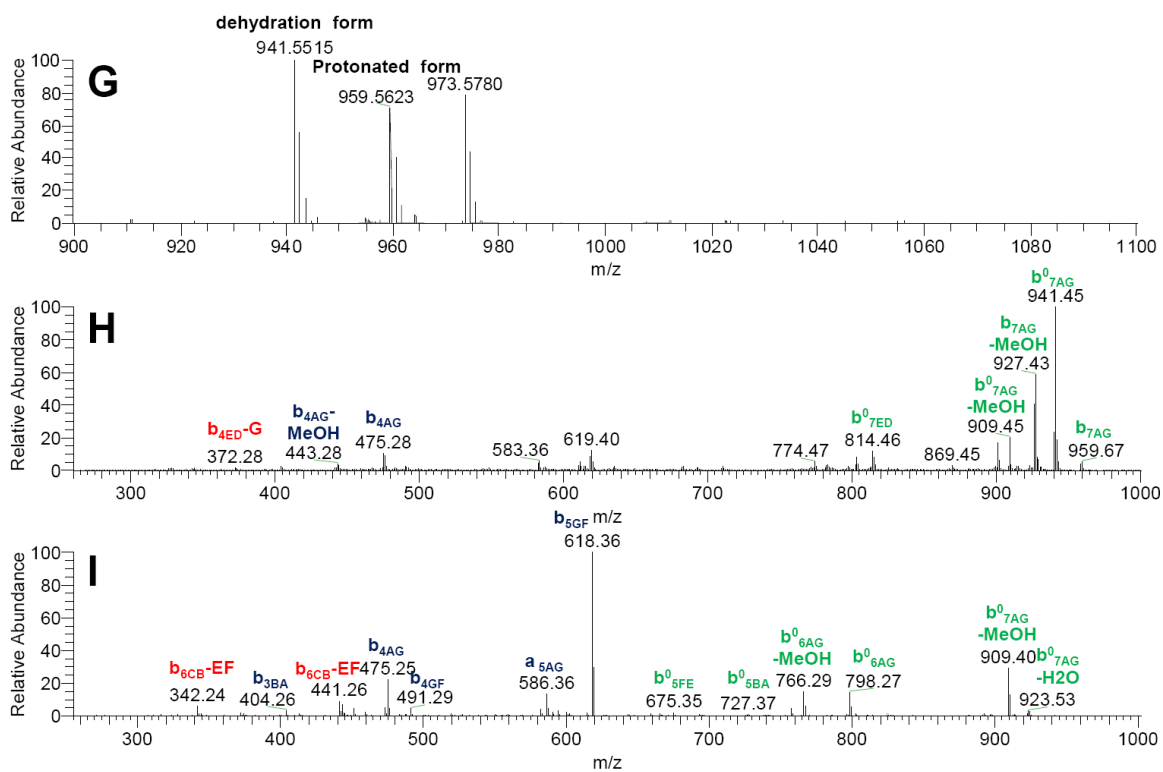
Figure 1.11: Cyclomarin MS and MS^2 spectra, continued. Cyclomarins MS spectra were collected by ESI-LTQ-FTICR MS. MS^2 spectrum were collected by ESI-LTQ MS. A-C: cyclomarin A spectrums; D-F: cyclomarin B spectrums; G-I: desprenylcyclomarin C spectrums. Broadband (A, D, G); MS^2 of cyclomarin protonated form (B, E, H); MS^2 spectrums of cyclomarin dehydration form (C, F, I).

1.3   CONCLUSION

Because cyclic peptides are an important class of therapeutics and toxins, we have developed a program, MS-CPA, to facilitate the structural characterization of these types of natural products. Users can easily access the program on the World Wide Web in order to annotate their tandem mass spectra of cyclic peptides. Using this program, we solidified the amino acid sequence of several recently discovered bioactive natural products, such as dimethoxymajuscalide (DMMC), mantillamide, dudawalamide A, and verified the structure of desprenycyclomarin C as well as dehydro-desprenylcyclomarin C that were isolated from a desprenyltransferase knockout *S. arenicola* CNS-205 strain. This analysis demonstrates the strength of this program when combined with tandem mass spectrometry, as well as a candidate structure enables the structural characterization of cyclic peptides produced in such low quantities that normally prohibit the use of other structural methods such as NMR.

Using our annotation program, we observed that cyclic nonribosomal peptides fragment in unusual ways. This kind of sequence-scrambling fragmentations results in a spontaneous recyclization event. The observation of NDS ions makes the problem of *de novo* sequencing of cyclic peptides even more challenging than was previously anticipated. Therefore, the annotation and understanding of the fragmentation patterns will, undoubtly, facilitate and improve *de novo* sequencing algorithm developments.

In summary, our current developed program provides a rapid annotation platform for tandem MS spectra of cyclic peptides. Also, although not designed for this, it can likely also be used to analyze the cyclization phenomenon of linear peptides. We are currently using this program to annotate peptides that have been isolated from marine

organisms that have potent cancer, malarial, and antibiotic resistant bacterial inhibitory activities. The approach described in this paper should be useful to the studies of cyclic peptide virulence factors, the chemical ecology of cyclic peptides, as well as cyclic peptides in drug screening programs.[1,58-61]

1.4   EXPERIMENTAL SECTION

Sample Preparation

Seglitide was purchased from Aldrich and was dissolved to a concentration of 20 μg/mL in 50:50 methanol (MeOH)/water with 1.0% acetic acid (AcOH). Dudawalamide A and DMMC were isolated from cyanobacteria and prepared in a solution of 50 μg/mL concentration in 50:50 MeOH/water with 1.0% AcOH and was infused into the mass spectrometer. Cyclomarins were isolated from a marine actinomycete and desalted with C18 ZipTip pipet tips (Millipore) following the manufacturer's protocol to a final concentration of 50 μg/mL.

Mass Spectrometry

All samples were subjected to electrospray ionization on a Biversa Nanomate (Advion Biosystems, Ithaca, NY) nanospray source (pressure, 0.3 psi; spray voltage, 1.4-1.8 kV). Seglitide, tyrothricin, and DMMC were analyzed a Finnigan LTQ-FTICR MS instrument (Thermo-Electron Corporation, San Jose, CA) running Tune Plus software version 1.0 and Xcalibur software version 1.4 SR1. Dudawalamide A was analyzed on a Thermo LTQ-Orbitrap-MS instrument (Thermo) running Tune Plus and Xcalibur software version 2.0. Activation time and q experiments, low-resolution spectra of

seglitide, tyrothricin, and cyclomarins were acquired on a Finnigan LTQ-MS (Thermo-Electron Corporation, San Jose, CA) running Tune Plus software version 1.0. The final spectrum was obtained by averaging MS2 scans with QualBrowser software version 1.4 SR1 (Thermo). Generally, the instrument was first autotuned on the m/z value of the ion to be fragmented. Then, the [M + H]+ ion of each compound was isolated in the linear ion trap and fragmented by collision induced dissociation (CID). Sets of consecutive, high-resolution, full MS/MS scans were acquired in centroid or profile mode and averaged using QualBrowser software (Thermo). The Thermo-Finnigan RAW files containing the average spectra were then converted to mzXML file format using the program ReAdW (tools.proteomecenter.org).

Pre analysis data-processing

The pre analysis data–processing effect was demonstrated in **Table 1.4** and **Figure 1.12**, and was discussed in below:

<u>Filtering:</u> The main challenge in the annotation process is actually the generation of a spectrum in which most peaks can be interpreted. Because of the great variance of experimental settings, instrumentation, and fragmentation properties of the compounds, the pre-processing steps required for each compound and experiment vary a lot. To this end, we implemented a series of filters to enhance the signal to noise ratio of the experimental spectrum. Given that noise peaks are unavoidable, the main goal of the filters is to eliminate ions that are likely noise without losing the actual data. In addition this gives the user of this program the flexibility to annotate their spectra in a manner they prefer. For example the user may only want to annotate the top 10 in the spectrum.

This is possible with this interface. In many cases in natural product research the samples are available in limited quantities or the peptide does not fragment well and therefore it is not always possible to produce the best mass spectra. The filters will allow us to work with these spectra, instead of repeating the experiment, which might not be possible.

The filters can be summarized as follow: Centroid, Window filtering: Centroid and Window are typically designed by the same manner, with Centroid filtering previous to Window filtering. These filters are equivalent of sliding a window of a certain width and keeping only the allowed numbers of peaks selected by intensity rank within the window. There are 2 parameters that are given for this filter, namely, width which is the width of the sliding window, and top, which is the minimum rank (inclusive) of peaks within the window, others will be discarded. For example, for Fourier Transform (FT) mass spectra that is collected in profile mode, a centroid filtering procedure with width=0.1 and top=1, keeps the tallest peak in a given profile. Then widows filtering with width=0.5 and top=2 keeps the top 2 peaks within every 0.5 m/z region. In practice, this works as well as (if not better) than complicated centroiding algorithms that try to "match" the peak obtained in a mass spectrum, in terms of speed and accuracy.

Rank filtering: this filtering keeps the peaks which intensity rank is better than the user specified top parameter.

Isotopic filtering: this filtering takes advantage of the observation that most true peaks also have isotopic peaks (resulting from the incorporation of N^{15} and C^{13}) in the mass spectrum. We can use this observation in two ways. First, we can require all peaks in the resulting spectrum after the filtering to have a supporting isotopic peak, keeping only those peaks that have isotopic peaks and discarding them otherwise. Or, second, we

can remove the isotopic peaks only and still keep those peaks that do not have isotopic peaks. The first approach is more stringent and works well for mass spectra that have a lot of peaks, but it might filter out too many peaks in spectra which most true peaks do not have isotopic counterparts. There are two parameters that can be set for this filter. Error is the maximum distance in Daltons allowed between the experimental isotopic peak and the computed peak for the experimental peak to be considered an isotopic peak. Ratio is the minimum normal peak to isotopic peak intensity ratio for the pair to be considered in this relation. For example, ratio=2 requires that the monoisotpic peak must have twice the intensity than its isotopic peak.

Water filtering: This filtering is very similar to the isotopic filtering, except that it evaluates normal peaks and their corresponding water losses, instead of normal peaks and their isotopic counterparts.

These filters are by no means the only computational methods to decrease the complexity of a mass spectrum, but they are very effective in practice.

Peak tolerance: Peak tolerance filtering keeps the peaks with observed mass within allowed mass difference between theoretical mass.

Symmetrization: Additionally, when the parent mass of a spectrum is known, the complementary set of ions can be computed. We call this procedure symmetrization of the mass spectrum, and it can be done by creating a new peak for each difference of the parent mass and an experimental peak in the initial spectrum. In the best case scenario, the number of annotations can be doubled by this procedure.

The filters, as well as the symmetrization can be applied to the spectrum in different order yielding can be optimized to provide different results. The web interface

will allow the user to select different filters, and generate dynamically a preview of the results.

Table 1.4: Filtering effect. Seglitide MS^2 spectra were analyzed by MS-CPA using different parameters. The table shows the parameter setting and analysis results.

	filtering parameters				explainable ion intensity	
	centroid width, top	rank	Symmetric	cleavages	w/o NDS ions	with NDS ions
A	0.1, 1	100	Yes	30/30	64.87%	72.56%
B	1, 1	100	Yes	16/30	81.69%	88.18%
C	0.1, 1	30	Yes	12/30	82.49%	86.01%
D	0.1, 1	100	No	22/30	64.87%	72.56%
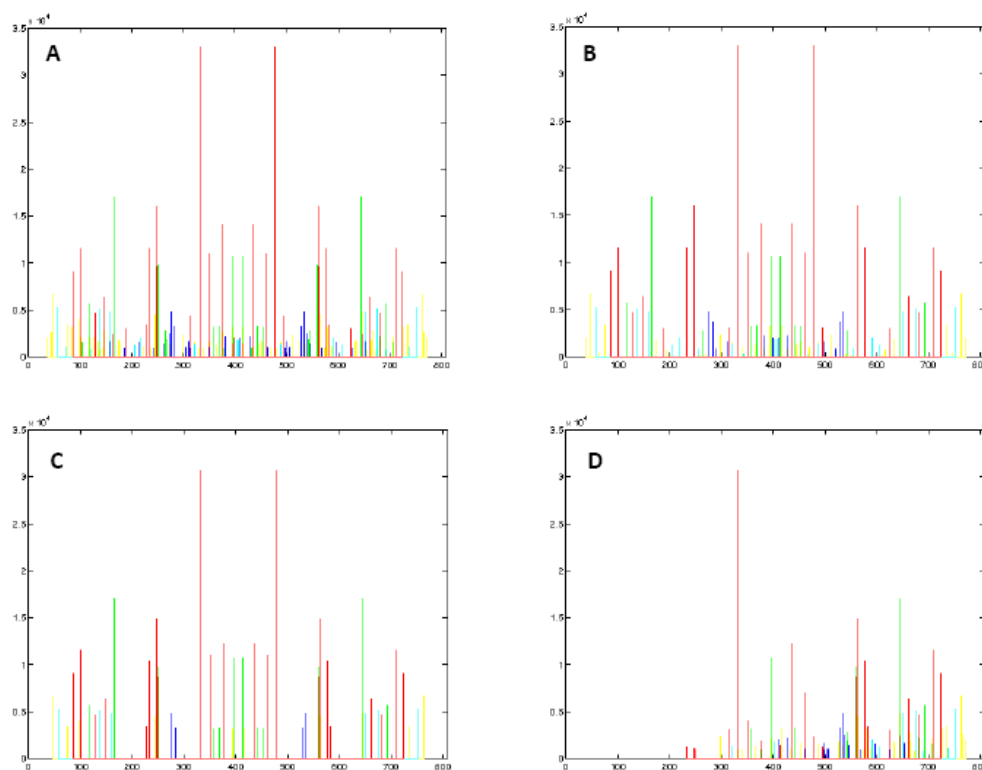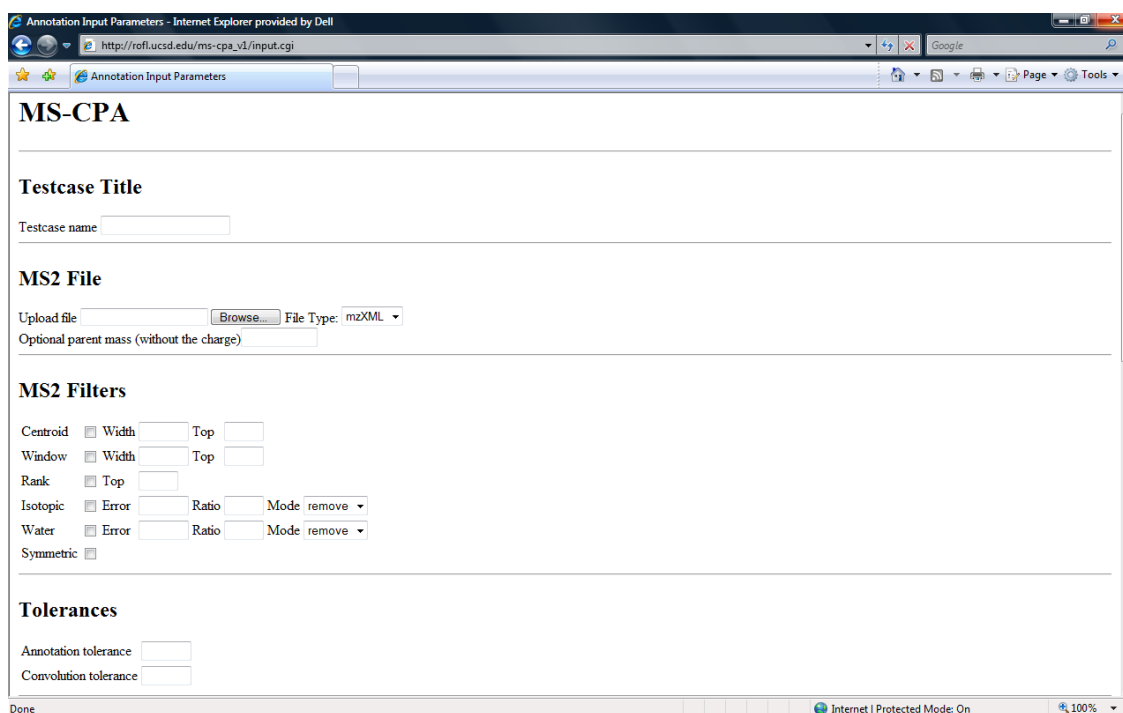
Figure 1.12: Filtering effect. Seglitide MS^2 spectra were analyzed by CAP using different parameters. The MS-CPA output spectra were showed.

MS-CPA Tutorial

This tutorial overviews the steps for using MS-CPA to analyze cyclic peptides. MS-CPA is available at (http://lol.ucsd.edu/ms-cpa_v1/Input.py).

This program currently is capable of handling .dta and .mzXML file formats as this data format is becoming the standard format for reporting or depositing mass spectra and/or proteomic data-sets[3] as spectrum inputs. Additionally, this program provides several filters which can be selected to apply if desired. Followed by specify the residue mass and sequence, and it is ready to go!

Figure 1.13: MS-CPA input setting.

Steps using seglitide spectrum as an example are described as follow (**Figure 1.13**):

1.   Simply specify a name for Testcase Title. In this case, we typed in seglitide.

2.   Upload a file either in .mzXML or .dta format.

3.   If symmetrization is desired, specify the parent mass that enables generating the complementary set of ions base on the input mass which should be more accurate. If not, MS-CPA will automatically pick up the experimental parent mass for symmetrization.

4.   Choose filters that are desired to apply and specify the parameters. How each filter function was detailed discussed in filters section. If a filter is desired to add in, users can click the box right behind the filter name and also specify parameters. In this case we choose centroid filtering- width 0.1, top 1; top- 200; isotope- error 0.01, ratio 10, remove; and symmetric.

5.   Tolerance specifies the allowed mass difference between observed mass and theoretical mass. In this case, since the seglitide spectrum was collected using FT-ICR, we set the annotation tolerance and convolution tolerance as 0.006.

Figure 1.14: Input amino acid setting.

6.   Then users need to specify residue masses. For the reason that many cyclic peptides contain unusual or modified amino acid, we leave the freedom for users to input the amino acid masses manually **(Figure 1.14)**. Also, default standard amino acids mass are provided by simply click the default bottom.
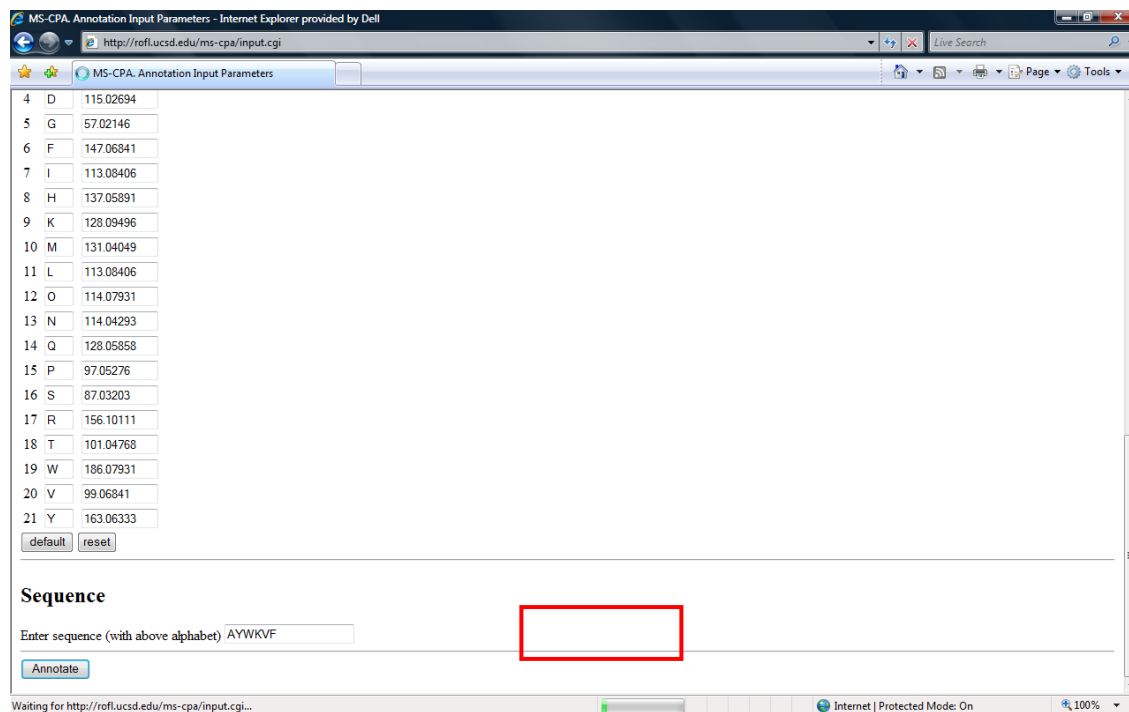
Figure 1.15: Input sequence and launch bottom.

7. Followed by specify the expected sequence and click annotate bottom, MS-CPA will start to process input spectrum.

8. Users can monitor the running status by looking at the status bar circled by the red box (**Figure 1.15).** Usually MS-CPA takes a few seconds to a few minutes to process data depending on the size of input file and the number of signals.

9. The outputs of MS-CPA are showed in **Figure 1.1**.

A continued study featured *de novo* sequence of Nonribosomal peptides has recently been accepted for publication to Nature Method, 2009. Ng, Julio; Bandeira, Nuno; Liu, Wei-Ting; Ghassemian, Majid; Simmons, Thomas L.; Gerwick, William; Linington, Roger; Dorrestein, Pieter; Pevzner, Pavel. The thesis author was the third author of this paper.

Chapter 1, in full, is a reprint of the material as it appears in Analytical Chemistry, 2009. Liu, Wei-Ting; Ng, Julio; Meluzzi, Dario; Bandeira, Nuno; Gutierrez, Marcelino; Simmons, Thomas L.; Schultz, Andrew W.; Linington, Roger G.; Moore, Bradley S.; Gerwick, William H.; Pevzner, Pavel A.; Dorrestein, Pieter C. The thesis author was the primary investigator and author of this paper.

Chapter II

The initial characterization of a ribosomally encoded antibiotic trifolitoxin

and its biosynthetic pathway

2.1   INTRODUCTION

Natural products are a group of compounds found in nature sources and have pharmaceutical values. Most of the natural products are secondary metabolites of living creatures playing important roles in chemical defense and communication. This group of compound showed a broad range of activities; well-known examples are tetracycline and penicillin, wildly used antibiotics; lovastatin, a cholesterol-lowering agent; cyclosporine A, an immunosuppressant drug used to prevent organ rejection[12]; microcystins, nodularins, toxins that can wipe out the entire fisheries[3,4], and taxol, curacin A, potent cancer drugs[62-64].

Today, more than 60% clinical drugs in use are, or inspired by natural products[65-67]. Although emerging combinatorial chemistry with advantages of time efficient and low cost have once challenged natural product-based drug discovery approach, however, the lack of novel chemistry coming from combinatorial chemistry approach making only one compound (Nexavar) was brought up to clinical use by this method[66]. In the last decade, with the growing knowledge and availability of genome sequencing and the increasing number of drug resistant pathogenic microorganisms, natural products have regained interest.

In the present research, we focused on a natural product antibiotic, trifolitoxin (TFX) which is a ribosomally synthesized, post-translationally modified peptide antibiotic that inhibits a specific group of organisms within the α-proteobacteria[68-69]. TFX inhibits most strains of α-proteobacteria including nitrogen fixing symbiotic bacteria in the genera *Rhizobium, Sinorhiozbium* and *Mesorhizobium,* plant pathogens in the genus *Agrobacterium* and animal pathogens *Brucella*[69-71].

Mature TFX is one of the smallest, most active post-translationally modified antibiotics[70]. It is reported that mature TFX has at least 4 isoforms sharing same molecule weight, making structure characterization very difficult. The two most abundant forms are referred to as TFX1 and TFX2. TFX2 is far more active than TFX1, having a Ki of 10 ng/ml while the Ki of TFX1 is 150 ng/ml[72].

The biosynthesis especially how the post-translational modification acts on TFX has merely been understood; most of the current assumptions are based on bioinformatics preditions. It was demonstrated that a cassette of genes, tfxABCDEFG is involved in TFX production. These genes, when incorporated into a wide range of α-proteobacteria can confer the TFX production activity[69,73]. It is proposed that the structural gene, TfxA, encodes TFX pro-peptide which then undergoes post-translational modification through the actions of TfxB, TfxC and TfxF[68,73]. TfxE is involved in TFX resistance while TfxG is important in optimal TFX production through the production of TFX isomers[74]. Mature TFX molecules are modified form of linear peptides DIGGSRQGCVA comprising the c-terminal 11 amino acids of TfxA. A study done by Scupham and Triplett that use site-direct mutagenesis approach identified six mutants D32A, I33A, G34A, G35A, S36A, V41A still retain full TFX activity. Therefore the TFX active sites were mapped to RQGC region[72]. It is suggested that this antibiotic contains a UV-absorbing chromophore derived from residues R37 and Q38 and a thiazoline ring formed from residue C40[75]. Synthetic TFX peptide without posttranslational modification does not have antibiotic activity [72].

In this current study, we first used tandem mass spectrometry to localize the modification sites and then adopted *in vivo* reconstitution approach aiming to identify the

actions of TFX modification enzymes. In tandem MS characterization, high accuracy tandem MS were acquired, isotope feeding assay, as well as mutant analogs analysis were performed to assist TFX structure characterization. Capture of y4-H_2O suggested cyclization on G39, C40 forming thiazoline and eliminated the possibility of cyclization on G35, S36. Also, the losses of the other six hydrogens were localized to a more specific residue Q38. Based on these findings we have proposed a preliminary structure which agrees well with all current evidences including NMR, UV absorption, and fluorescence features. In addition we have laid down the foundation to study the function of each of the enzymes on the TFX gene cluster.

2.2 RESULTS and DISSCUSSION

2.2.1   Post-translational modifications of TFX

Seven genes, tfxABCDEFG are involved in TFX production. The first gene tfxA encoded a 42 amino acid length peptide with its c-terminal 11 amino acid similar to matured TFX. The calculated mass of this 11-mer peptide is 1061.49239, whereas mature TFX was measured to have a mass of 1037.43488. After considering all possible chemical compositions, this 24.05751 Da difference is considered as losses of 8 hydrogens and 1 oxygen (0 ppm), which most likely was caused by a water loss and another three desaturations. Combined with NMR evidence obtained from Art Edison's lab, the most reasonable attribution of this water loss comes from cyclodehydration which are often seen in natural products when cysteine or serine is adjacent to a glycine residue. The mechanisms of this type of heterocyclization are well characterized. Generally it is triggered by the nucleophile attack from the oxygen or the sulfur on serine

or cysteine side chains to the neighboring carbonyl carbon generating thiazoline or oxazoline[76]. In many cases, thiazoline and oxazoline will further undergo oxidation and introduce a double bond generating thiazole or oxazole structure. A well known example is microcin B17 which undergoes extensive PTMs and yields 8 thiazole or oxazole rings on its mature structure. In TFX pro-peptides sequence, one serine and one cysteine were found adjacent to glycine, both of these are considered as candidates for the water loss.

To localize the PTM sites, mature TFX were subjected to OrbiTrap mass spectrometry, and the resulting spectrum and fragmentation ion map were showed in **figure 2.1A, figure 2.2, and table 2.1**. Observing b7 ion with losses of 24 Da originally excited us since this suggest TFX may actually has an oxazoline rather than previous suggested thiazoline structure. We thus set up to perform alkylation experiments because if cysteine is unmodified it will have a free thiol and will be able to be alkylated. However, alkylation experiment did not support a free thiol structure. Therefore, we suspected b7 ion has an unknown rearrangement involved, thus showing at an unexpected mass.

In the other hand, having observed y4 ion with 18 Da losses localized the water loss on G39-C40 residue, mostly possible forming a thiazoline structure. This ruled out the possibility of forming oxazoline on G35-S36 and also revealed that this modification do not undergo further oxidation after cyclodehydration. Ion fragmentation map (**Figure 2.2**) localized the other 6 Da losses to S36-R37-Q38 region. To confirm the annotations of the molecular formulas of these ion fragments, we conducted N15 feeding experiment. The way we used the N15 to confirm the annotation is by first predicting the number of nitrogens presented and then matched the mass shift corresponding to the number of

nitrogens. The resulting spectra of the N15 feeding experiment and ion table were showed in **figure 2.1B** and **table 2.2.** All ions were shifted by expected mass and their correctness was validated.

MS data inferring that mature TFX has thiazoline structure is enlighting and suggests that S36 is not involved in cyclization. If no other PTMs act on Serine 36, it should be able to be fragmented. However, no fragments were observed in this region leaving the possibility that the serine may has some sort of modifications or its the peptide property making it hard to be fragmented. We suspected residue changes in the pro-peptide sequence should change the fragmentation patterns and may help in identifying more informatics cleavages. Scupham et al., in 2006 has done site-mutagenesis to each residue on TFX pro-peptides, we anticipated these strains will help in characterization of TFX structure. We obtained these strains from the Triplett laboratory, isolated the mutant analogs and subjected the TFX analogs to MS, and the resulting spectra and annotations tables are showed in **figure 2.1 C, D, E** and **table 2.3**. Some of the mutant TFX analogs yielded much richer fragmentations, especially having observing y4 (-18) and b9 (-24) ion provides robust information that the loss of 18 Da is at G39-C40, again pointing to the thiazoline formation. Most interestingly, covering series ions from y4 to y8 makes the localization of the loss of 6 Da to a more precise position R37-Q38. This also suggested that TFX contain unmodified amide bond between S36 and R37. In here, using tandem MS, we successfully localized the modification sites (**figure 2.2**) and provided important piece of structural information.
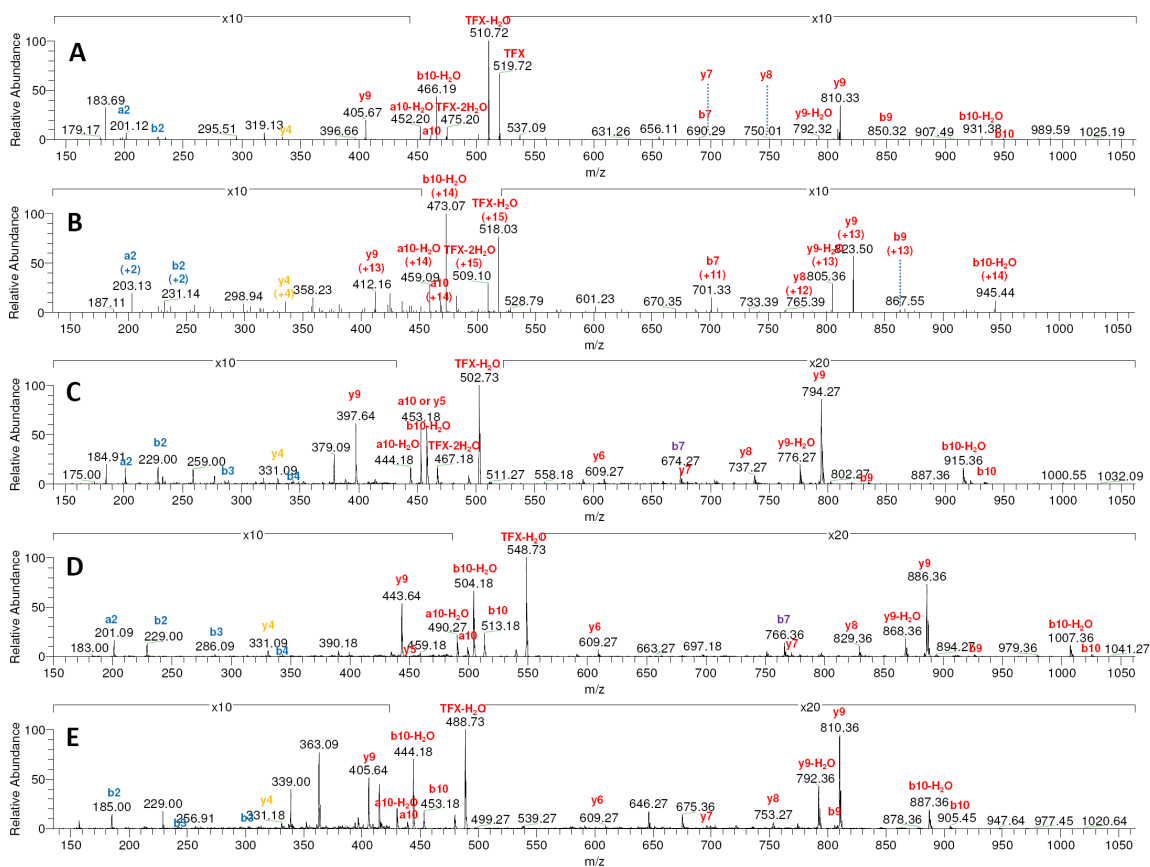
Figure 2.1: Tandem MS analysis of TFX and its mutant analogs. A: MS/MS spectra of TFX obtained by OrbiTrap MS. B: LTQ MS/MS spectra of extracted TFX grown on N15 labeling media. C, D, E: LTQ MS/MS spectra of TFX mutant analogs S36A, S36Y, D32A, respectively.

Table 2.1: Summary of the ions observed in TFX MS/MS spectra obtained from OrbiTrap (related to figure 2.1A).

	charge	Observed ion	Observed Mw	Theoretical Mw	Delta	ppm
*TFX	2	519.7224	1037.4302	1037.43488	-0.0046	-4.47
*TFX-H_2O	2	510.7172	1019.4198	1019.42431	-0.0045	-4.38
*TFX-2H_2O	2	501.7122	1001.4098	1001.41375	-0.0039	-3.90
*b10	1	949.3911	948.3838	948.38720	-0.0034	-3.56
*b10	2	475.1989	948.3832	948.38720	-0.0040	-4.18
*b10-H_2O	1	931.3803	930.3730	930.37663	-0.0036	-3.88
*b10-H_2O	2	466.1936	930.3726	930.37663	-0.0040	-4.29
*b9	1	850.3228	849.3155	849.31878	-0.0033	-3.84
*b7	1	690.2926	689.2853	689.28814	-0.0028	-4.09
b2	1	229.1175	228.1102	228.11101	-0.0008	-3.46
*y9	1	810.3285	809.3212	809.32387	-0.0027	-3.27
*y9	2	405.668	809.3214	809.32387	-0.0024	-3.00
*y9-H_2O	1	792.3181	791.3108	791.31331	-0.0025	-3.15
*y8	1	753.3073	752.3000	752.30241	-0.0024	-3.18
*y7	1	696.2858	695.2785	695.28094	-0.0024	-3.48
*a10	2	461.2003	920.3860	920.39228	-0.0062	-6.78
*a10-H_2O	2	452.1964	902.3782	902.38172	-0.0035	-3.86
a2	1	201.1226	200.1153	200.11609	-0.0008	-3.85
y4-H_2O	1	331.1425	330.1352	330.13618	-0.0010	-2.9

Table 2.2: Summary of the ions observed in LTQ MS/MS spectra of extracted TFX grown on N15 labeling media (related to figure 2.1B).

	charge	N14 ion	Formula	Nitrogen number	Expect ion	Signal observed
Intact TFX			C41H71N15O16S			
TFX	2	519.7224	C41H63N15O15S	15	527.22	+
TFX-H20	2	510.7172	C41H61N15O14S	15	518.22	+
TFX-2H2O	2	501.7122	C41H59N15O13S	15	509.22	+
b10	1	949.3911	C38H56N14O13S	14	963.39	-
b10	2	475.1989	C38H56N14O13S	14	482.20	+
b10-H20	1	931.3803	C38H54N14O12S	14	945.38	+
b10-H20	2	466.1936	C38H54N14O12S	14	473.19	+
b9	1	850.3228	C33H47N13O12S	13	863.32	+
b7	1	690.2926	C28H39N11O10	11	701.29	+
*b2	1	229.1175	C10H16N2O4	2	231.12	+
y9	1	810.3285	C31H47N13O11S	13	823.33	+
y9	2	405.6680	C31H47N13O11S	13	412.17	+
y9-H20	1	792.3181	C31H45N13O10S	13	805.32	+
y8	1	753.3073	C29H44N12O10S	12	765.31	+
y7	1	696.2858	C27H41N11O9S	11	707.29	-
a10	2	461.2003	C37H56N14O12S	14	468.20	+
a10-H2O	2	452.1964	C37H54N14O11S	14	459.20	+
a2	1	201.1226	C9H16N2O3	2	203.12	+

Table 2.3: Summary of ions observed in LTQ MS/MS spectra of TFX mutant analogs S36A, S36Y, D32A (related to figure 2.1C, D, E.)

| strain	S36A			S36Y			D32A		
	cal Mw	obs Mw	error	cal Mw	obs Mw	error	cal Mw	obs Mw	error
b1	115.03	-	-	115.03	-	-	71.04	-	-
b2	228.11	227.99	-0.12	228.11	227.99	-0.12	184.12	183.99	-0.13
b3	285.13	285.08	-0.05	285.13	285.08	-0.05	241.14	241.08	-0.06
b4	342.15	342.08	-0.07	342.15	342.08	-0.07	298.16	298.08	-0.08
b5	429.19	-	-	429.19	-	-	429.19	-	-
b6	585.29	-	-	585.29	-	-	585.29	-	-
b7	691.35	673.26	-18.09	783.38	765.35	-18.02	663.36	645.26	-18.09
b8	730.37	-	-	822.40	-	-	702.38	-	-
b9	833.38	833.35	-0.03	925.41	925.35	-0.05	805.39	805.26	-0.12
b10	932.45	932.08	-0.37	1024.48	1024.44	-0.03	904.45	904.44	-0.01
b11	1003.49	-	-	1095.51	-	-	975.49	-	-
y1	89.05	-	-	89.05	-	-	89.05	-	-
y2	188.12	-	-	188.12	-	-	188.12	-	-
y3	273.13	-	-	273.13	-	-	273.13	-	-
y4	330.15	330.08	-0.06	330.15	330.17	0.03	330.15	330.17	0.03
y5	452.21	-	-	452.21	-	-	452.21	-	-
y6	608.31	608.35	0.05	608.31	608.26	-0.04	608.31	608.26	-0.04
y7	679.34	679.26	-0.08	771.37	771.35	-0.02	695.34	695.26	-0.08
y8	736.36	736.26	-0.10	828.39	828.35	-0.04	752.36	752.26	-0.10
y9	793.39	793.26	-0.12	885.41	885.35	-0.06	809.38	809.35	-0.03
y10	906.47	-	-	998.50	-	-	922.47	-	-
y11	1021.50	-	-	1113.52	-	-	1017.50	-	-
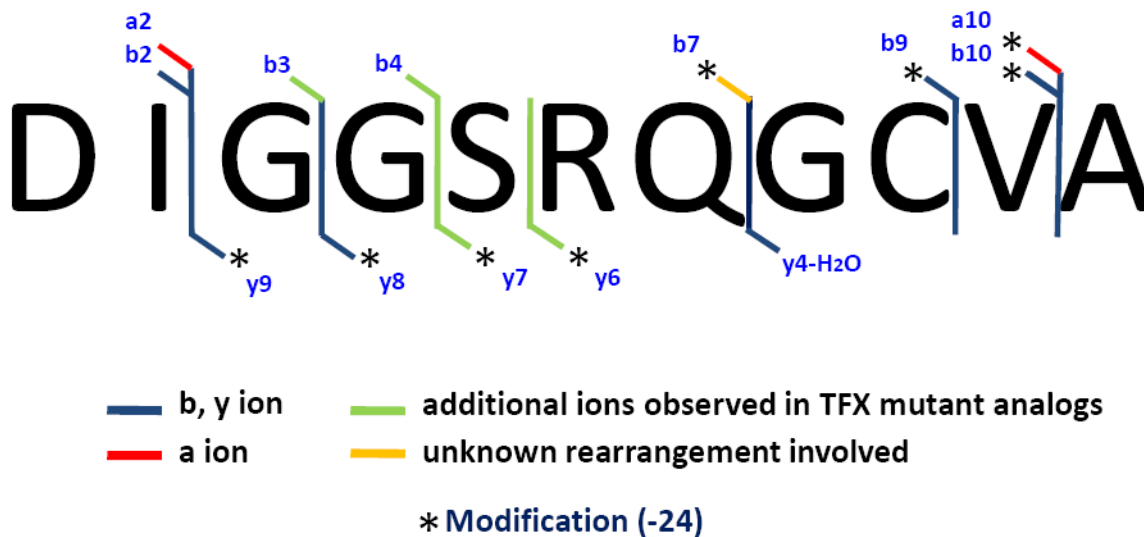
Figure 2.2: Fragmentation map of MS/MS spectra derived from TFX, and its mutant analogs.

Although localization of 6 hydrogen loss on glutamine is quite exciting, the complete structure characterization of TFX still requires more information. In fact, the isolation of TFX can be traced back to 20 years ago and has been attempted by at least 5 labs on three continents[70]. Since that, lots of efforts have been done on TFX characterization, including full NMR characterization, site-direct mutagenesis, transposon mutagenesis, gene deletion and so on, but the complete structure remains unsolved. Nonetheless at this time it is possible to forward a candidate structure.

Figure 2.3 showed the UV absorption and emission profiles of TFX. Besides thiazoline, the chromophore were showed to have the maxium UV absorbance at 302nm and emission at 393nm which belonged to a purple-blue visible light region consisting with the observation of blue fluorescence. This UV absorbance and fluorescene features suggested that the chomophore should have extensive conjugated double bond system.
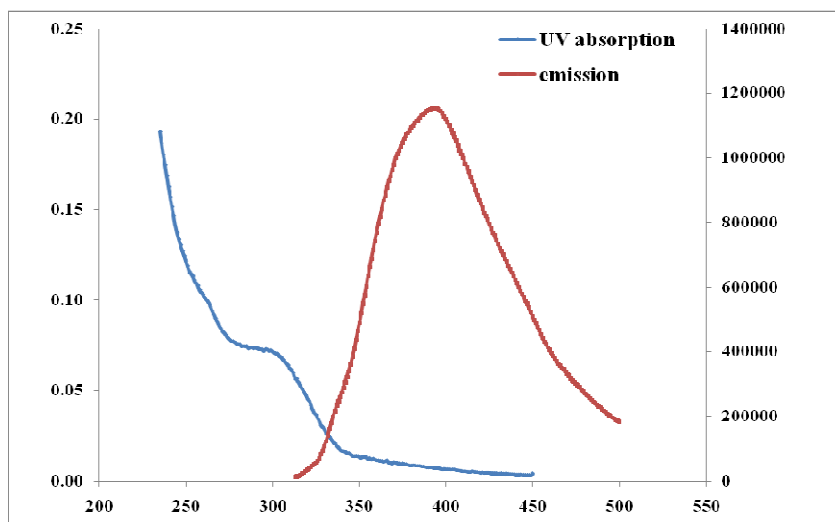
Figure 2.3: UV absorption and emission profiles of TFX. Emission spectra was obtained by excited TFX at 302nm.

Based on NMR characterizaiton done by Max Tate (The University of Adelaide-Australia), Art Edison (University of Floridia), Victor Rumjanek (Universidade Federal Rural do Rio de Janeiro UFRRJ, Brazil), and a series of discussions between Triplett's lab and Jodie Johnson from University of Floridia, we forwarded two candidate TFX structures. These are shown in **figure 2.4** (compound 2, 3). However, since we observed the cleavage between S36 and R37 suggesting the peptide backbone connecting these two residues are unmodified. Therefore, we antipicated that compound 3 are more likely to be the mature TFX structure. This structure is consistent with all the structural data thusfar obtained for TFX.
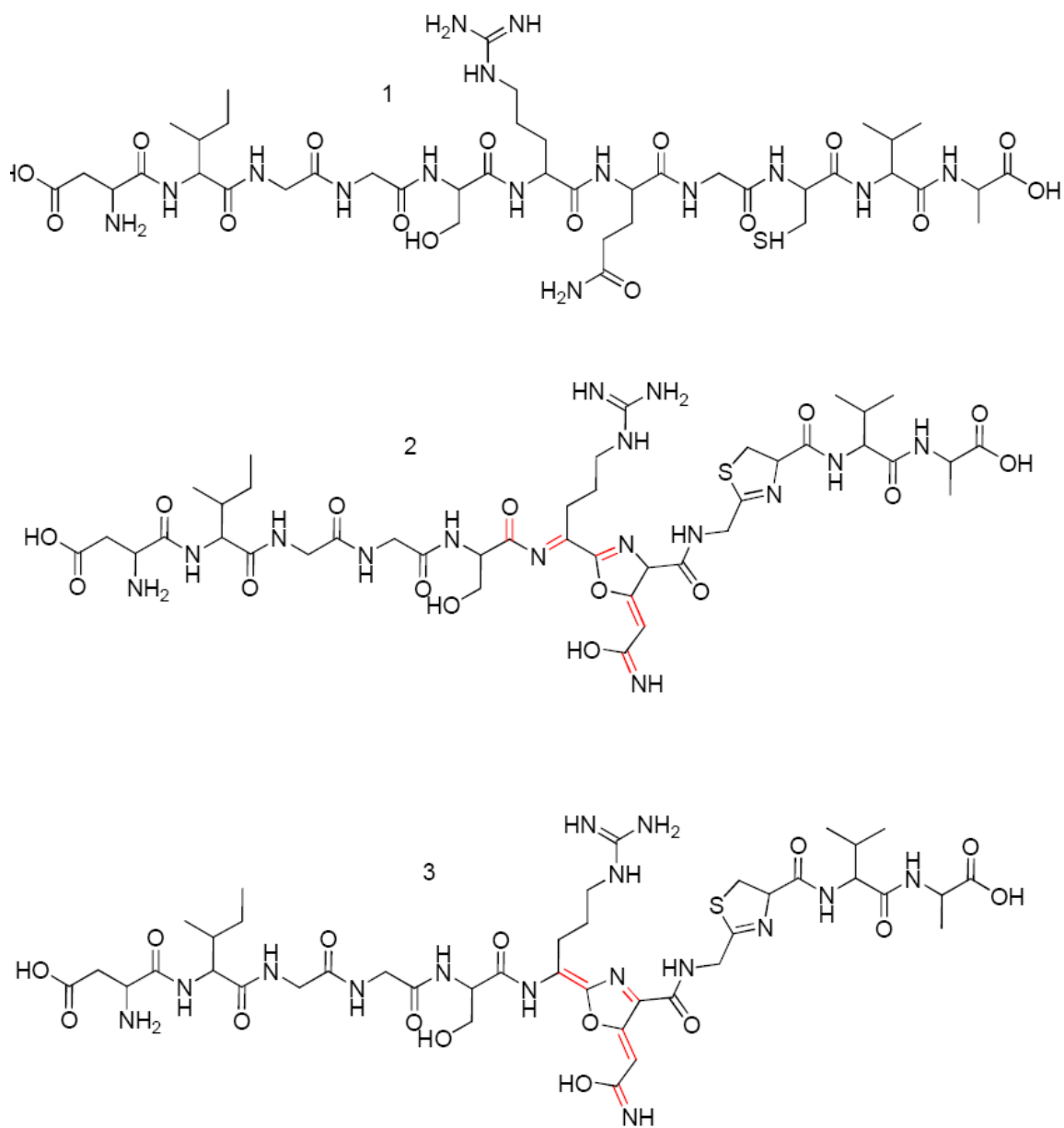
Figure 2.4: Proposed TFX structure. Compound 1 is the unmodified TFX peptide sequence. Compound 2 and 3 are proposed TFX structure candidates.

2.2.2 Bioinformatics analysis of TFX gene cluster

The above structure is a proposed structure and needs to be solidified. Therefore, we aimed to go in structure elucidation from a different angle. We anticipated that if we can pinpoint the biosynthetic mechanism of TFX that it would help in understanding how this molecule was modified and would refute or solidify the structural annotation. To start this part of work we performed a detailed biosynthetic analysis of the genes proposed to be involved in the biosynthesis of TFX (**Table 2.4**). TFX is a ribosomally encoded peptide antibiotic. Generally, a biosynthetic gene cluster for ribosomally encoded peptides are composed of a structure gene, and other genes responsible for immunity, export, and PTMs. TfxA has the last 11 amino acid similar to mature TFX thus was suspected to be the structure gene which encodes for TFX pro-peptide. It contains a 42 amino acids leader peptide which will be cleaved during TFX maturation process. The timing of when the leader peptide is removed and the function of the leader sequence are still unknown. Processing a leader sequence is often seen in ribosomal encoded antibiotics or toxins. Generally, it is involved in stabilization of the antibacterial peptide by preventing intracellular degradation[77], acts as a chaperone folding the molecule so that it is recognized by the maturation machinery[78], or serves as a recognition sequence for the maturation or export machineries[79]. TFX leader sequence may take similar functions.

Table 2.4: Homology and Predicted Function of the Trifolitoxin Biosynthetic Gene Cluster

Gene	Size[a]	Homolog and origin	Position	Identity/Similarity[b]	Predicted function
tfxA	42				Structure gene
tfxB	373	NADH oxidase [Bacillus cereus E33L] nitroreductase [Actinosynnema mirum DSM 43827]	213-353 110-361	34/51 25/40	McbC like dedydrogenase
tfxC	356	putative NADH oxidase [Pyrococcus furiosus DSM 3638] nitroreductase [Desulfatibacillum alkenivorans AK-01]	153-356 125-356	26/44 27/44	McbC like dedydrogenase
tfxD	416	putative short-chain alcohol dehydrogenase [uncultured bacterium]	210-289	31/53	11 transmembrane domain
tfxE	252	HAD family hydrolase [Burkholderia ambifaria MC40-6]	1-133	30/33	YcaO superfamily Cyclodehydratase
tfxF	258	NADH oxidase [Bacillus cereus E33L]	48-224	25/37	McbC like dedydrogenase
tfxG	261	phosphotransferase family protein [Vibrio angustum S14] aminoglycoside phosphotransferase [Salinispora arenicola CNS-205]	7-210 2-206	35/47 37/51	isomerzation

[a] numbers are in amino acids

[b] numbers are in percent

As for the tailoring enzymes, the C-terminal end three genes (tfxBCF) that have 25-34% identity, 37-51% similarity to NADH oxidase appear to be candidates. These three genes are highly homologous to the mcbC enzyme responsible for oxidation of the thiazoline/oxazoline to thiazole/oxazole in microcin B17 system (mechanism showed in **scheme 2.1A and C**). Although mcbC-like dehydrogenase is a conserved enzyme in natural products, having three genes that are encoded for mcbC-like dehydrogenase in biosynthetic gene cluster is uncommon. **Figure 2.5** showed a few well known gene cluster that contain genes homolog to either tfxB, tfxC or tfxF. Investigating the mature products of these gene clusters, one can easily found that all these molecules contain thiazole or oxazole moiety (**Figure 2.6**). However, no clues showing that thiazole/oxazole structures were involved in TFX mature structure. Therefore, we suspected that tfxBCF may be a new class of enzyme that is responsible for catalyzing reaction other than thiazole/oxazole formation. Since there are 6 Da losses in TFX maturation which correspond to three oxidations, these three genes may possible each catalyze an oxidation introducing a double bond and cause a loss of 2 hydrogens.
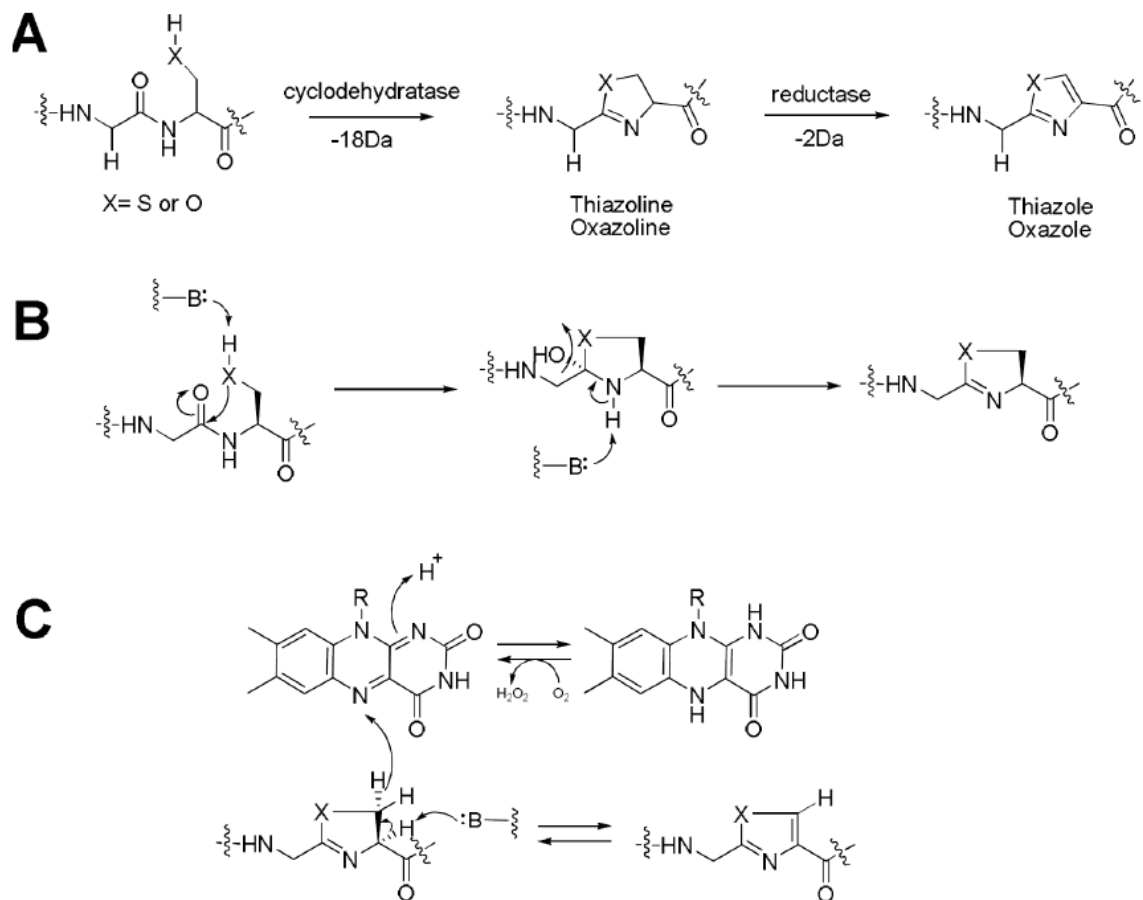
Scheme 2.1: Thiazole/oxazole formation pathway. A: Outline of thiazole/oxazole formation pathway. B: Mechanism of heterocyclization of serine and threonine generating thiazoline and oxazoline. C: Further oxidation generated by a FMN dependent pathway to create thiazole/oxazole.
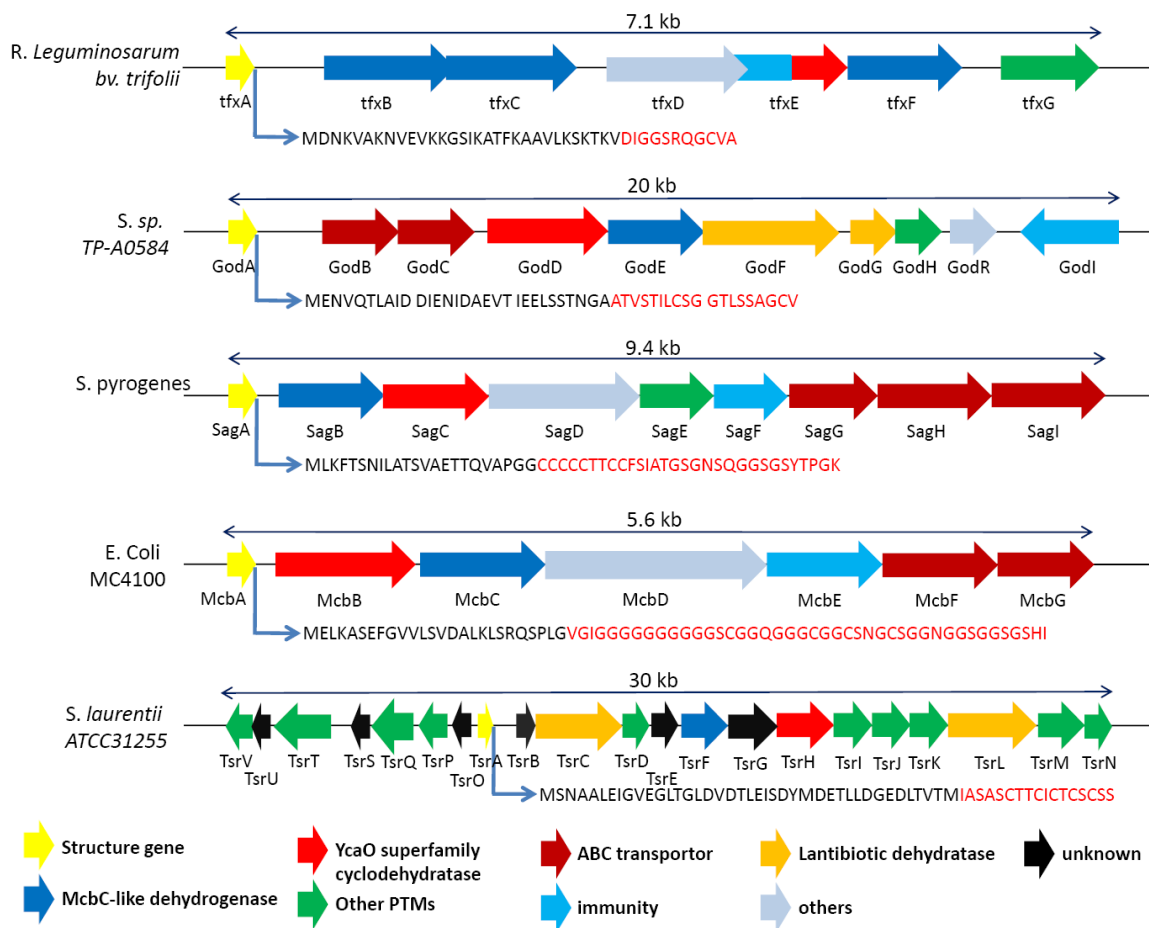
Figure 2.5: Representative biosynthetic gene clusters that producing thiazole/oxazole-containing peptides.
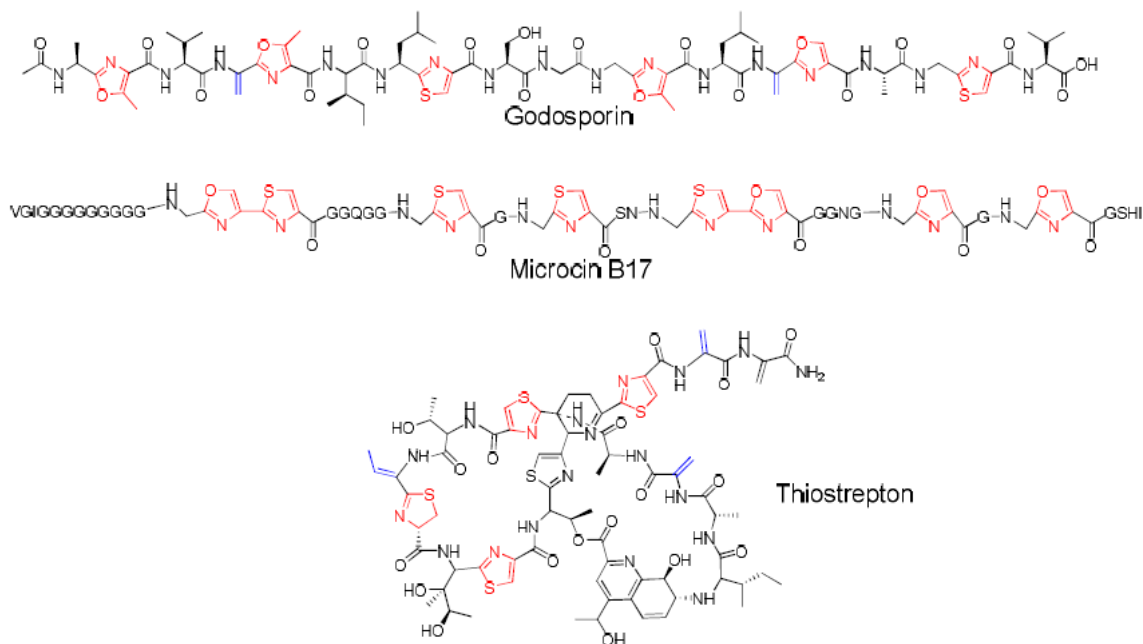
Figure 2.6: Known mature thiazole/oxazole-containing peptide structure encoded by the gene clusters shown in figure 2.5.

TfxD is a membrane protein and is predicted to have 11 transmembrane domains based on a transmembrane predictor server (DAS)[80]. We suspected TfxD is a membrane protein that possible work as molecule pump and responsible for TFX efflux and may also associate with immunity. In most microcins system, at least one membrane protein were encoded as export system, such as ATP transporter in microcin B17, microcin J25; and multidrug efflux transporter for microcin C7/C51. In some microcins system, an ATP-binding transporter with photolytic domain was found. These types of membrane proteins are responsible for export as well as leader peptides removal. Therefore, in here, we suspected TfxD can act as an exporter and also leaves possibility that it may associated with leader peptide removal.

The left two genes, tfxE and tfxG were once thought to be involved in TFX resistance and immunity based on bioinformatics analysis. TfxG contains a eukaryotic-type serine/threonine kinase motif and is highly homolog to aminoglycoside phosphotransferase. Aminoglycoside phosphotransferase are associated with the deactivation of aminoglycoside type antibiotics such as kanamycin, amikacin, gentamicin and neomycin. It inactivated the antibiotics by phosphorylation. Also, a report showing failure of getting viable phenotype with tfxE or tfxF insertion by transposon mutagenesis also pointed that both tfxE and tfxG are immunity genes[73]. However, later on a deletion studies done by removal tfxG gene showed no effect on TFX resistance whereas deletion of tfxE reduced the strain's resistance to exogenous TFX[74]. TfxG were then found to involve in TFX isomer formation.

Although tfxE were demonstrated to be immunity gene, a bioinformatics searching of tfxE revealed it belongs to YcaO superfamily. This family is recently classified as a potentially cyclodehydratase (mechanism showed in **scheme 1A and 1B**) and are likely to catalyze the conversion of all 6 cysteines in the thiocillin backbone to thiazole rings[81]. This family of enzyme is also found in streptolysin S gene clusters[82]. This homolog leaves the possibility that TfxE can also be a cyclodehydratase.

2.2.3 In vivo reconstitution

Bioinformatics analysis provides putative assumptions that each TFX gene may be function as. To decipher the actions of each biosynthetic gene, generally two approaches are being used (outline in **figure 2.7**). Using *in vitro* strategy, TFX complementation is conducted by incubating TFX precursor and tailoring enzymes alone

with some cofactors. For *in vivo* strategy, precursor-encoding gene is co-transformed with tailoring enzyme gene or genes. TFX processing can be monitored by the mass changes using mass spectrometry.
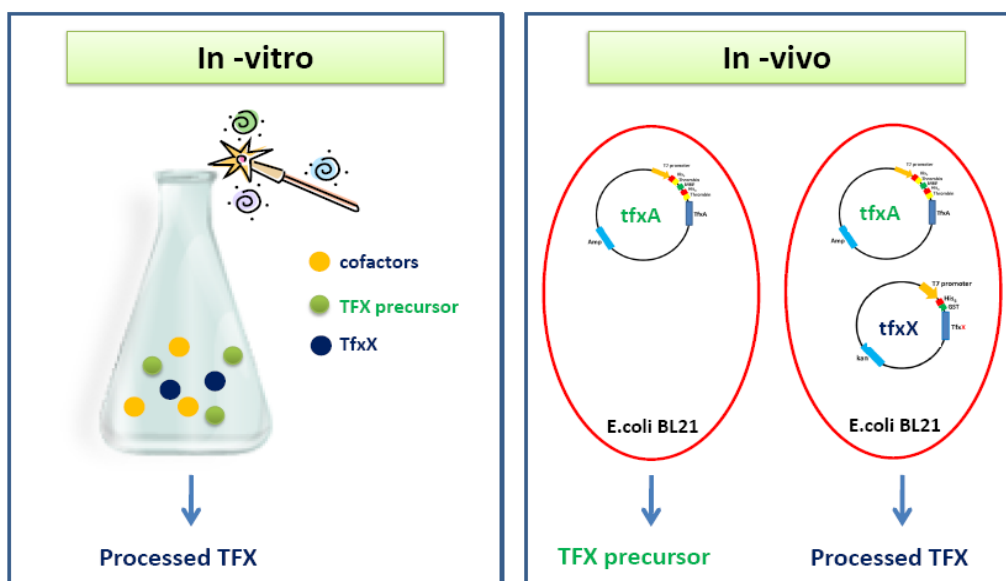
Figure 2.7: Outline of *in vitro* and *in vivo* approaches of TFX complementation studies.

Traditionally, scientists tend to purified out gene products and use *in vitro* reconstitution approach to pinpoint the action. Although *in vitro* reconstitution approach benefits in having more control to the reactions which enable kinetic and regio-selection characterization, however, one needs to be able to predict the cofactors beforehand. Since the understanding of TFX are still at early stage, having no or little clues about the reactions, not to mention figuring out the corresponding cofactors, we set out to use *in vivo* reconstitution approaches and adopt *E.coli* as a heterologous host system.

We suspected that each TFX gene may carry out a biosynthetic step or at least some genes in TFX biosynthetic gene cluster is capable of taking TFX pro-peptide as substrate and work independently without coordinating other TFX gene products. Enzymes that are capable of working alone were found in PKS and NRPS system as well as ribosomally encoded peptides. For example, in microcin C7/C51 system, mccB alone is capable of adenylating C7-C51 propeptides[83] and mccD alone is responsible of transferring of propylamine on the phosphate group of mccB modified pro-MccC7-MccC51[84]. Thus, we designed an experiment to pinpoint the action of each gene on TFX pro-peptide. The strategies were outlined in **figure 2.8**.
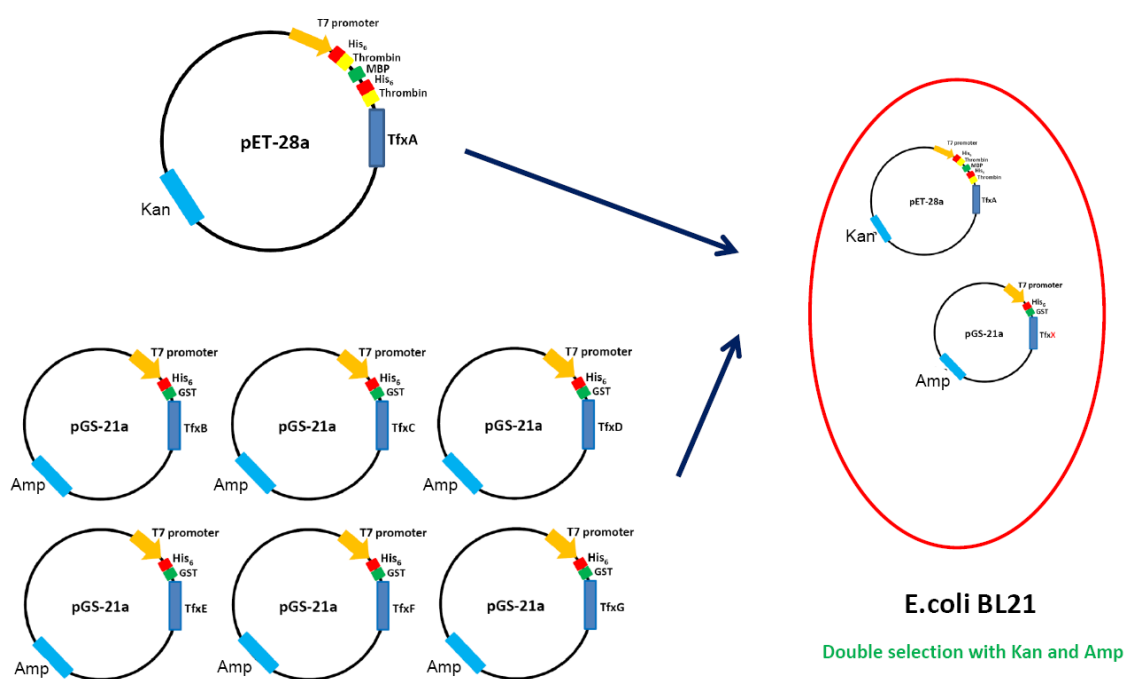

Figure 2.8: Outline of trifolitoxin *in vivo* reconstitution studies. TfxA was subcloned to a pET-28a plasmid, whereas other TFX gene was each assembled in one pGS-21a plasmid. Two plasmids were co-transformed and the cells that successfully carry both two plasmids are selected by containing both antibiotics markers.

TFX biosynthetic genes were amplified from a TFX borne plasmid pT2TFXK obtained from the Triplet laboratory using Pfu polymerase, and the resulting DNA fragments were analyzed by agarose gel to check the right DNA size (**Figure 2.9**). TfxA was subcloned to a pET28-a plasmid carrying a MBP tag on its C-terminal end. The other TFX genes were each subcloned into a pGS-21a plasmid. Plasmids pET-28a and pGS-21a contain different origins of replication and different selective markers; therefore can be maintained together in the same *E. coli* cell. TfxA and each TFX gene were co-transformed to *E. coli* BL21 (DE3) and the gene products were co-expressed.
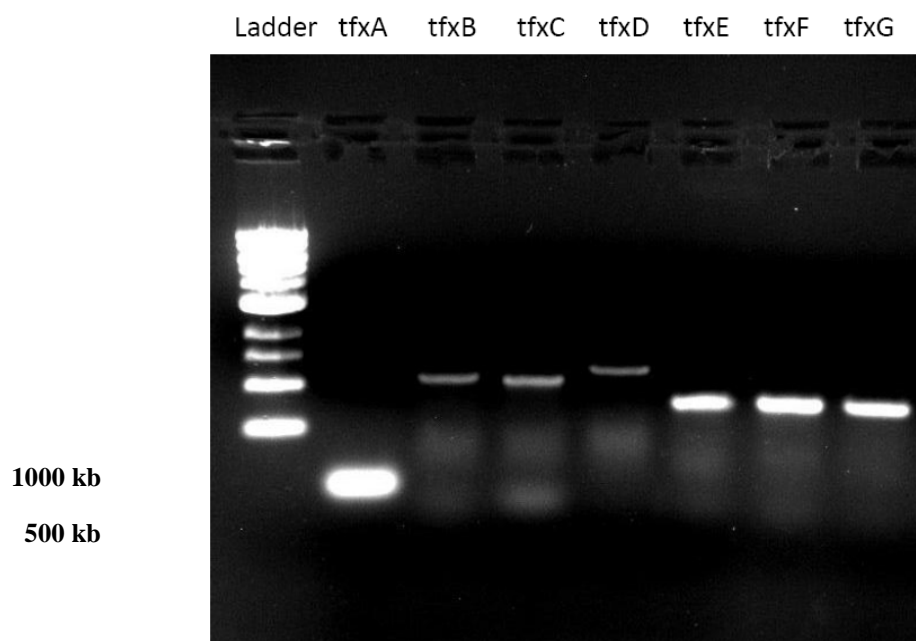

Figure 2.9: PCR amplification of TFX biosynthetic genes.

Expressed proteins were analyzed by SDS-PAGE showed in **figure 2.10**. Except TfxA, all the other TFX gene products, when expressed alone is highly insoluble. However, it is observed here TfxB when co-expressed with TfxA, becomes soluble

suggesting there are some interactions between these two proteins. In our previous expression tests, TfxA is highly soluble; however, **figure 2.10** shows that when co-expressed with other TFX proteins, the solubility decrease drastically.

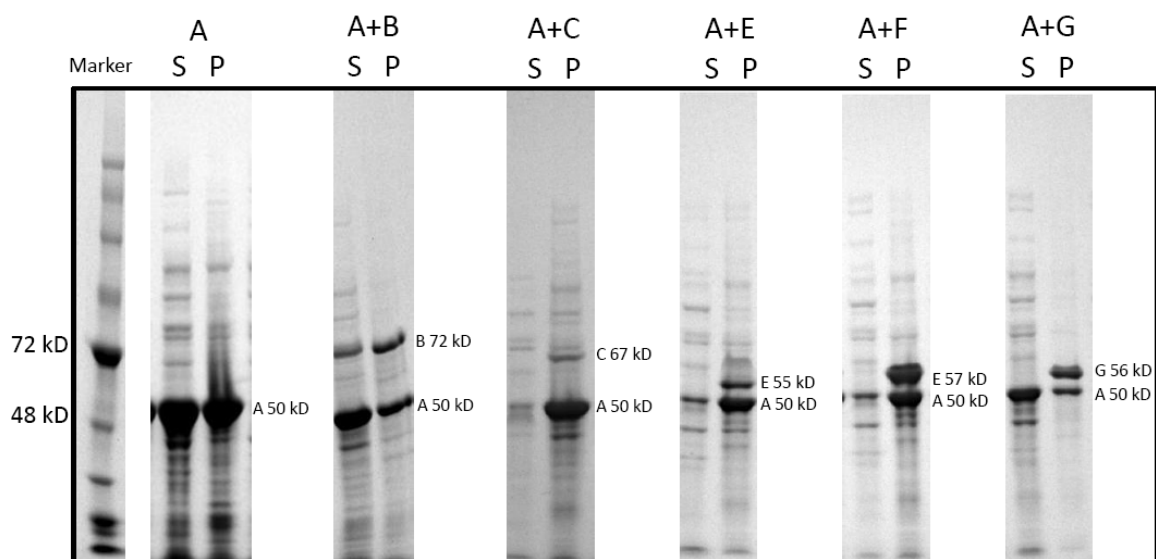**Figure 2.10: SDS-PAGE analysis of co-expressed proteins.** A: TfxA was expressed alone. A+B, C, E, F, G: Co-express TfxA with TfxB, TfxC, TfxE, TfxF, or TfxG. S represents the supernatant portion of the protein homogenous solution after sonication which suggesting the soluble portion. P represents the re-suspended cell pellet of the protein homogenous solution after sonication which suggesting the insoluble portion.

Table 2.5: Co-expression status of TFX biosynthetic gene products.

Clones	Expression cell	Vector	Resistance	Expression property	Glycerol stock
tfxA^K	BL21(RIPL) (DE3)	MBP-pet28	Km	Soluble	Yes
tfxB^A	BL21(DE3)	PGS-21a	Amp	Insoluble	Yes
tfxA^K+B^A	BL21(DE3)	MBP-pet28 PGS-21a	Km + Amp	soluble	Yes
tfxA^K+C^A	BL21(DE3)	MBP-pet28 PGS-21a	Km + Amp	Insoluble	Yes
tfxA^K+D^A	BL21(DE3)	MBP-pet28 PGS-21a	Km + Amp	Sequence confirmed, No expression of tfxD	Yes
tfxA^K+E^A	BL21(DE3)	MBP-pet28 PGS-21a	Km + Amp	Insoluble	Yes
tfxA^K+F^A	BL21(DE3)	MBP-pet28 PGS-21a	Km + Amp	Insoluble	Yes
tfxA^K+G^A	BL21(DE3)	MBP-pet28 PGS-21a	Km + Amp	Insoluble	Yes

To check if TfxA is modified by these biosynthetic gene products, TfxA was purified out by amylose resin. After digested with thrombin, 200 ug of MBP-TfxA were subjected to HPLC separation and analyzed by FT-MS. MBP-tfxA alone was also transformed and followed by same preparation method and subjected to FT-MS as a control.

MBP-TfxA subjected to thrombin digestion was predicted to yield three fragments by cleaving at the two red arrow position (LVPR_GS) shown in **figure 2.11B**. **Figure 2.11A** shows the HPLC trace of MBP-TfxA, the most abundant peak at around 23 min was demonstrated to be the MBP moiety by FT-MS analysis (**Figure 2.11C**). However, HPLC revealed more than two fragments other than MBP moiety. MS characterization found five fragments and later on the cause was pinpointed to an unknown protease that cleave at the two alanine located at C-terminal portion (showed as blue arrow) thus yielding two extra fragments (**figure 2.11B**).
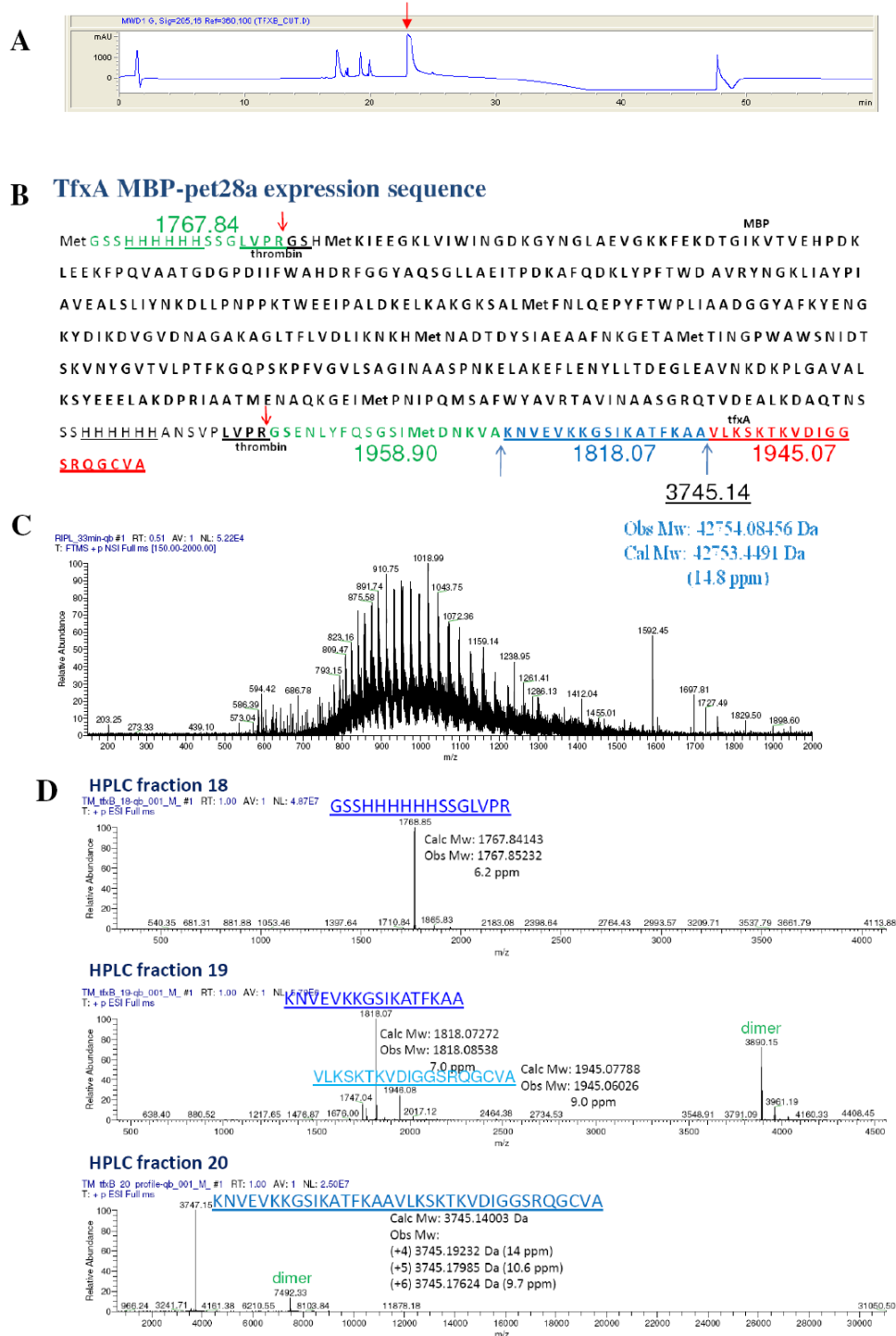
Figure 2.11: MS characterization of *E.coli* expressed MBP-TfxA. A: HPLC trace of purified MBP-TfxA. B: Expression sequence of MBP-TfxA. C: MS characterization of HPLC fraction eluted at 23 min and was mapped to the MBP portion with cleavage at the two thrombin recognition site. D: MS characterization of HPLC fractions eluted at 18 to 20 min.

The C-terminal part fragments with a calculated mass of 3745 were analyzed by FT-MS. This peptide was observed in +5 state (750.44 m/z), +6 state (625.53 m/z), +7 state (536.31 m/z), and +8 state (469.40 m/z) (**Figure 2.12**). MBP-TfxA purified from *E.coli* that harbored other TFX gene was analyzed by same procedure. Unfortunately, no interesting modifications were found (**Figure 2.13**). A few reasons may account for this unsatisfying result. One possible reason is that TFX modification enzymes need to work cooperatively to carry out certain biochemical transformations. Such a phenomenon has been observed before. In microcin B17, although McbB was responsible for cyclodehydration whereas McbC is in charge of oxidation to form thiazole and oxazole, however, their function was abolished when tested individually. Another possibility is that the 7 genes of the TFX gene cluster is not sufficient to make TFX and that there are other genes or factors involved in TFX biosynthesis that are conserved in closely related strains but are not present in *E.coli.*
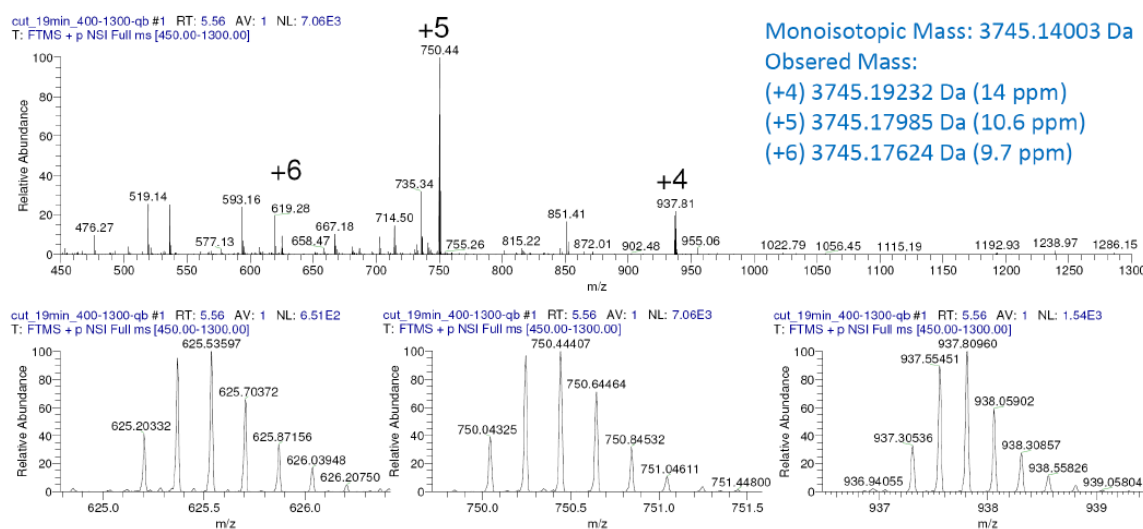

Figure 2.12: FT-MS characterization of the c-terminal end region of TfxA.
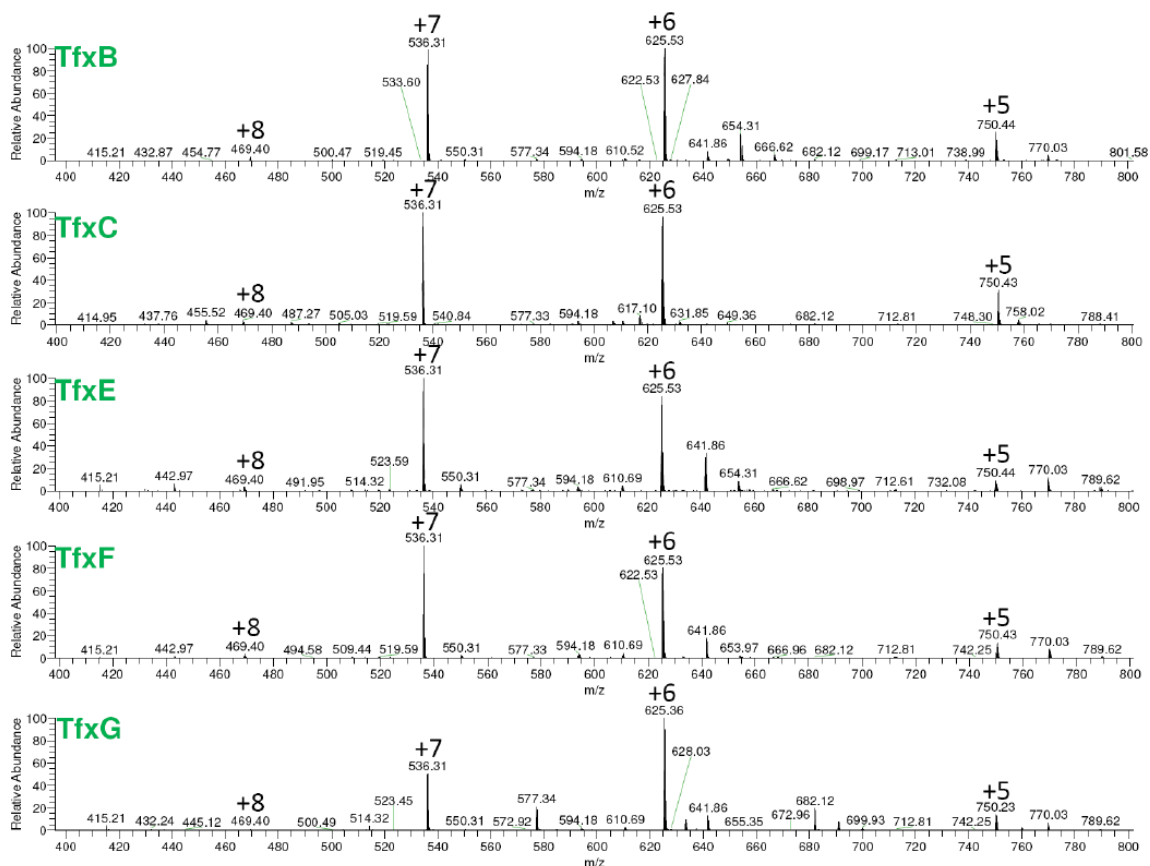
Figure 2.13: FT-MS analysis of TfxA which was co-expressed with TFX biosynthetic gene products (specify on upper-left corner). MBP-TfxA was co-expressed with TFX genes. After MBP affinity column purification and HPLC separation, fractions that contained TfxA c-terminal end were analyzed by FT-MS.

Before stepping into co-expressing of combination of TFX biosynthetic genes to eliminate the concern that *E.coli* may not capable of acting as a heterologous system for TFX biosynthesis studies, we set out to assemble and express all the TFX biosynthetic genes in one *E.coli* cell. This approach was proposed and its feasibility was proved by Watanabe et al., in 2006[85]. By introducing all the biosynthetic involved genes of echinimycin to *E.coli*, they were able to retrieve mature echinomycin from expression culture. To avoid the toxicity during the gene construction, we design to use two separate

plasmids, one carries tfxA gene, and the other carries tfxBCDEFG genes. Since this is a heterologous expression, we manually introduced T7 promoter, ribosome binding site, and T7 terminator alone with each TFX gene. This construct was carefully prepared by Dr. Yu-Quan Xu (PostDoc fellow, Dorrestein Lab). The assembly of tfxBCEDFG was showed in **figure 2.14**. First, this plasmid was subjected to PCR using corresponding primers to check successfully incorporation of all TFX biosynthetic genes (**Figure 2.15**).

Figure 2.14: Assembly of TFX biosynthetic gene cluster.

After confirming the incorporation of TFX genes, protein expressions were first checked by transforming plasmid that contain single TFX gene (constructs that carrying each gene before assembling together), and the expression results were showed in **figure 2.16A.** It can be noticed that TfxB, C, E, G showed correct sized protein expression. TfxD and TfxF did not show protein expression. TfxD is a membrane protein and may be the reason that hinders it to be expressed. Since in our assumption, TfxD is more likely to be involved in TFX export and our primary goal is to understand the PTMs, therefore, we suspect the failure of expression TfxD will not hamper PTM characterization. We are more concerned about the expression of TfxF, but since the sequence of insertion genes

are checked and the incorporating of genes is confirmed, at current time, we cannot explain why tfxD and F were not expressed. Thus we moved forward to the co-expression of TFX gene cluster. **Figure 2.16B** was done by co-transformed TfxA and TfxBCDEFG plasmids. Red box highlighted TfxA position and blue boxes circled the other TFX proteins. Unfortunately, TfxA seems not very soluble thus was not able to be purified out. Currently, we are trying to put a MBP tag in front of TfxA since this tag can increase the solubility of the expressed proteins. Once we get that construct we can purify out TfxA and look for PTMs by mass spectrometry. We are also planning on doing target identification. Since TFX is a fluorescence molecule, this feature can be used to localize its distribution on living cells.
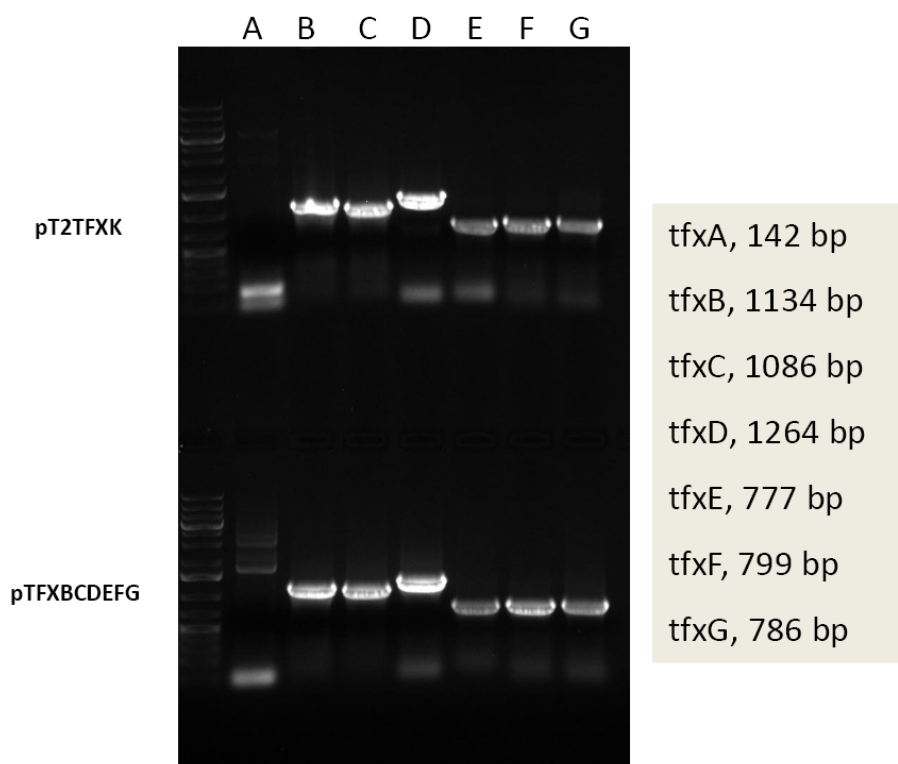
Figure 2.15: PCR check for successfully insertion of all TFX biosynthetic genes.
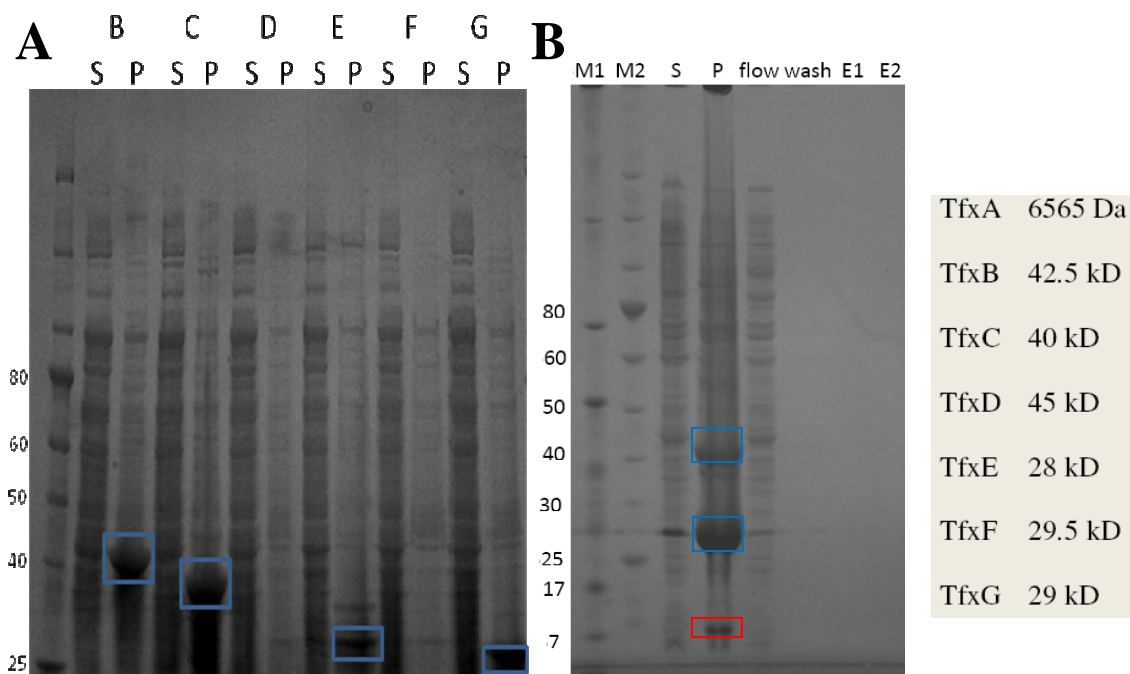
Figure 2.16: Expression of TfxABCDEFG. Panel A was done by transforming and expressing separate gene to check the protein expression. TfxB, C, E, G showed protein expression at the right size. TfxD and TfxF did not show protein expression. Panel B was done by co-transforming TfxA and TfxBCDEFG. Red box highlighted the TfxA and Blue box circled the other TFX proteins. S represents the supernatant portion of the protein homogenous solution after sonication which suggesting the soluble portion. P represents the re-suspended cell pellet of the protein homogenous solution after sonication which suggesting the insoluble portion.

2.3 EXPERIMENTAL SECTION

Strains growth condition

Rhizobium strains were cultured on Bergersen's synthetic media (BSM) (Bergersen 1961) containing 15 g agar/L. Selection antibiotics kanamycin 50 µg/ml or tetracycline 1.25 µg/ml were added to solid media. *Escherichia coli* strains were grown on Luria-Bertani medium. Antibiotics were used at the following concentrations: kanamycin 50 µg/ml; ampicillin 100 µg/ml; and chloramphenicol 25 µg/ml.

Sample preparation Mass spectrometry

MALDI-TOF MS spectra was generated by scraping approximately 1µl colony directly from agar plate and spotted on MALDI plate. All samples were analyzed by mixed with 1µl of matrix formulated by 35mg/mlα-cyano-4-hydroxycinnamic acid and 15mg DHB solubilized in 78% acetonitrile, 1% TFA. The samples were analyzed by MALDI-TOF mass spectrometer (Bruker). For Ion trap, OrbiTrap, and FT-MS analysis, a 10 µL sample was desalted with C18 ZipTip pipette tips (Millipore) following the manufacturer's protocol, diluted to 10 µL in 50:49:1 MeOH/H_2O/AcOH, infused by nano-electrospray ionization with a Biversa Nanomate (Advion Biosystems, Ithaca, NY) (pressure: 0.3 psi, spray voltage: 1.4-1.8 kV) and analyzed by Finnigan LTQ-MS, LTQ-OrbiTrap-MS, and LTQ-FTICR-MS, respectively (Thermo-Electron Corporation, San Jose, CA) running Tune Plus software version 1.0 and Xcalibur software version 1.4 SR1. For MS2 experiment, the interested ion was isolated in the linear ion trap and fragmented by collision induced dissociation with a isolation width 3-5 m/z and collision energy of

15 (activation q: 0.25; activation time: 100ms). MS/MS scans were acquired and averaged using QualBrowser software (Thermo).

Preparing of tfx constructs.

TfxA gene was amplified from plasmid pT2TFXK using Pfu polymerase and the following primers: Forward 5'- AAGGATCCATAATGGATAACAAGGTTGCGAAG AATGTCG -3'. Reverse 5'- AAGCGGCCGCATATTAAGCGACGCAGCCCTGAC - 3'. The amplified fragment was digested with BamHII and NotI and cloned into MBP-pET28a to generate plasmid MBP-tfxA. The other tfx genes were amplified from plasmid pT2TFXK in similar manners using the following primers:

tfxB

Forward 5' AAAAGCTTATAATGGACTTCGTCCAACGATTCG 3'

Reverse 5' AACTCGAGATATCAATCATGGTGGTTCTCG 3'

tfxC

Forward 5' AAAAGCTTATAATGATTGAACTGCGCCCGCTTC 3'

Reverse 5'AACTCGAGATACTAGCCTGCCGTGACGAGATAAAGC 3'

tfxD

Forward 5' CGACAGATCTAGCATGAGCGACGAAAACCA 3'

Reverse 5' GGTGCTCGAGGGTTTATCTGTTCGGTAGTGC 3'

tfxE

Forward 5' AAAAGCTTATAATGCACTACCGAACAGATAAACC 3'

Reverse 5' AACTCGAGATATCATGAATAACTCACCGCCTGTATC 3'

tfxF

Forward 5' AAAAGCTTATAATGAGAGCAAGCAAAACACC 3'

Reverse 5' AACTCGAGATACTATGGATTAAATCCATCAACG 3'

tfxG

Forward 5' AAAAGCTTATAATGAATGATGAGATTTGCCTGAC 3'

Reverse 5' AACTCGAGATACTATGCCAGCGCTGCTTC 3'

The amplified fragment of tfxD was digested with BglII and XhoI whereas other TFX genes were digested with BamHI and XhoI, all these genes were cloned into pGS-21a to generate plasmid tfxB, tfxC, tfxD, tfxE, tfxF, tfxG.

Protein Expression and Protein Purification

MBP-tfxA plasmid were pre-mixed with other TFX gene plasmid and co-transformed into *E. coli* BL21(DE3). Colonies contain both plasmids were selected by plating onto LB-agar plates containing 100 µg/ml ampicillin, and 50 µg/ml kanamycin and then subcloned. 5ml overnight cultures were used to inoculate 1L of LB media supplied with ampicillin and kanamycin which was grown to an OD600 of 0.6-0.8 at 37 °C. The cells were induced with 50mg/L IPTG and grown overnight at 18 °C. The cells were harvested by centrifugation and resuspended in 20ml buffer A (50 mM Tris, pH 7.5, 500 mM NaCl, 2.5% glycerol 0.5 mM TCEP). After sonicate for 2min (2s on, 2s off), and cellular debris was removed by centrifugation. The supernatant was applied to an amylose column (2 mL resin/L of cultured cells) pre-equilibrated with buffer A. The column was washed with 20 column volumes of buffer A, and then the protein was eluted with buffer B (50 mM Tris, pH 7.5, 150 mM NaCl, 2.5% glycerol 0.5 mM TCEP, 10mM maltose). Fractions were analyzed by a pre-casted 4-12% PAGEGel.

REFERENCES

(1)   Schmidt, E. W.; Nelson, J. T.; Rasko, D. A.; Sudek, S.; Eisen, J. A.; Haygood, M. G.; Ravel, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7315–7317.

(2)   Pomilio, A. B.; Battista, M. E.; Vitale, A. A. *Curr. Org. Chem.* **2006**, *10*, 2075–2121.

(3)   Rinehart, K. L.; Harada, K.; Namikoshi, M.; Chen, C.; Harvis, C. A. *J. Am. Chem. Soc.* **1988**, *110*, 8557–8558.

(4)   Gupta, N.; Pant, S. C.; Vijayaraghavan, R.; Lakshmana Rao, P. V. *Toxicology.* **2003**, *188*, 285–296.

(5)   Holden, M. T. G.; Chhabra, S. R.; de Nys, R.; Stead, P.; Bainton, N. J.; Hill, P. J.; Maneeld, M.; Kumar, N.; Labatte, M.; England, D.; Rice, S.; Givskov, M.; Salmond, G. P. C.; Stewart, G. S. A. B.; Bycroft, B. W.; Kjelleberg, S; Williams, P. *Mol. Microbiol.* **1999**, *33*, 1254–1266.

(6)   Ibrahim, M.; Guillot, A.; Wessner, F.; Algaron, F.; Besset, C.; Courtin, P.; Gardan, R.; Monnetl, V. *J. Bacteriol.* **2007**, *189*, 8844–8854.

(7)   Poupel, O.; Tardieux, I. *Microb. Infect.* **1999**, *1*, 653–662.

(8)   Branda, S. S.; Chu, F.; Kearns, D. B.; Losick, R.; Kolter, R. *Mol. Microbiol.* **2006**, *59*, 1229–1238.

(9)   Straight, P. D.; Willey, J. M.; Kolter, R. *J. Bacteriol.* **2006**, *188*, 4918–4925.

(10)   Sturme, M. H. J.; Nakayama, J.; Molenaar, D.; Murakami, Y.; Kunugi, R.; Fujii, T.; Vaughan, E. E.; Kleerebezem, M.; de Vos, W. M. *J. Bacteriol.* **2005**, *187*, 5224–5235.

(11)   Jegorov, A.; Hajduch, M.; Sulc, M.; Havlicek, V. *J. Mass Spectrom.* **2006**, *41*, 563–576.

(12)   Italia, J. L.; Bhardwaj, V.; Ravi Kumar, M. N. V. *Drug Discovery Today.* **2006**, *11*, 846–854.

(13)   Hannon, J. P.; Nunn, C.; Stolz, B.; Bruns, C.; Weckbecker, G.; Lewis, I.; Troxler, T.; Hurth, K.; Hoyer, D. *J. Mol. Neurosci.* **2002**, *18*, 15–27.

(14)   Gerding, D. N.; Muto, C. A.; Owens, R. C., Jr. *Clin. Infect. Dis.* **2008**, *46*, S43–S49.

(15)   Kofoed, J.; Reymond, J. L. *J. Comb. Chem.* **2007**, *9*, 1046–1052.

(16)   Fluxa, V. S.; Reymond, J. L. *Bioorg. Med. Chem.* **2009**, *17*, 1018–1025.

(17)   Berkovich-Berger, D.; Lemcoff, N. G. *Chem. Commun.* **2008**, *14*, 1686–1688.

(18)   Liu, T.; Joo, S. H.; Voorhees, J. L.; Brooks, C. L.; Pei, D. *Bioorg. Med. Chem.* **2009**, *17*, 1026-1033.

(19)   Schultz, A. W.; Oh, D. C.; Carney, J. R.; Williamson, R. T.; Udwary, D. W.; Jensen, P. R.; Gould, S. J.; Fenical, W.; Moore, B. S. *J. Am. Chem. Soc.* **2008**, *130*, 4507–4516.

(20)   Zhang, Y.; Zhou, S.; Wavreille, A. S.; DeWille, J.; Pei, D. *J. Comb. Chem.* **2009**, *17*, 1026–1033.

(21)   Feng, Y.; Carroll, A. R.; Pass, D. M.; Archbold, J. K.; Avery, V. M.; Quinn, R. J. *J. Nat. Prod.* **2008**, *71*, 8–11.

(22)   Linington, R. G.; Edwards, D. J.; Shuman, C. F.; McPhail, K. L.; Matainaho, T.; Gerwick, W. H. *J. Nat. Prod.* **2008**, *71*, 22–27.

(23)   Shimokawa, K.; Mashima, I.; Asai, A.; Ohno, T.; Yamada, K.; Kita, M.; Uemura, D. *Chem. Asian J.* **2008**, *3*, 438–446.

(24)   Shindoh, N.; Mori, M.; Terada, Y.; Oda, K.; Amino, N.; Kita, A.; Taniguchi, M.; Sohda, K. Y.; Nagai, K.; Sowa, Y.; Masuoka, Y.; Orita, M.; Sasamata, M.; Matsushime, H.; Furuichi, K.; Sakai, T. *Int. J. Oncol.* **2008**, *32*, 545–555.

(25)   Krishnamurthy, T.; Szafraniec, L.; Hunt, D. F.; Shabanowitz, J.; Yates, J. R.; Hauert, C. R.; Carmichael, W. W.; Skulberg, O.; Coddii, G. A.; Missler, S. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 770–774.

(26)   Ngoka, L. C. M.; Gross, M. L. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 732–746.

(27)   Jegorov, A.; Paizs, B.; Zabka, M.; Kuzma, M.; Havlıcek, V.; Giannakopulos, A. E.; Derrick, P. J. *Eur. J. Mass Spectrom.* **2003**, *9*, 105–116.

(28)   Yagu¨e, J.; Paradela, A.; Ramos, M.; Ogueta, S.; Marina, A.; Barahona, F.; Lo´pez de Castro, J. A.; Va´zquez, J. *Anal. Chem.* **2003**, *75*, 1524–1535.

(29)   Jegorov, A.; Paizs, B.; Kuzma, M.; Zabka, M.; Landa, Z.; Sulc, M.; Barrow M. P.; Havlicek, V. *J. Mass Spectrom.* **2004**, *39*, 949–960.

(30)   Harrison, A. G.; Young, A. B.; Bleiholder, C.; Suhai, S.; Paizs, B. *J. Am. Chem. Soc.* **2006**, *128*, 10364–10365.

(31)   Jia, C.; Qi, W.; He, Z. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 663–678.

(32)   Qi, W.; Jia, C.; He, Z.; Qiao, B. *Acta Chim. Sin.* **2007**, *65*, 233–238.

(33)   Tilvi, S.; Naik, C. G. *J. Mass Spectrom.* **2007**, *42*, 70–80.

(34)   Bleiholder, C.; Osburn, S.; Williams, T. D.; Suhai, S.; Van Stipdonk, M.; Harrison, A. G.; Paizs, B. *J. Am. Chem. Soc.* **2008**, *130*, 17774–89.

(35)   Harrison, A. G. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1776–1780.

(36)   Riba-Garcia, I.; Giles, K.; Bateman, R. H.; Gaskella, S. J. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1781–1787.

(37)   Riba-Garcia, I.; Giles, K.; Bateman, R. H.; Gaskell, S. J. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 609–613.

(38)   Eng, J. K.; McCormack, A. L.; Yates, J. R., III *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976.

(39)   Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551–3567.

(40)   Tanner, S.; Shu, H.; Frank, A.; Wang, L. C.; Zandi, E.; Mumby, M.; Pevzner, P. A.; Bafna, V. *Anal. Chem.* **2005**, *77*, 4626–4639.

(41)   Jagannath, S.; Sabareesh, V. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 3033–3038.

(42)   Bandeira, N.; Ng, J.; Meluzzi, D.; Linington, R. G.; Dorrestein, P.; Pevzner, P. A. *Proceedings of the Twelfth Annual International Conference in Research in Computational Molecular Biology*; Springer-Verlag: Berlin, Germany, 2008; pp 181-195.

(43)   Lin, S. M.; Zhu, L.; Winter, A. Q.; Sasinowski, M.; Kibbe, W. A. *Exp. Rev. Proteomics* **2005**, *2*, 839–845.

(44)   Pedrioli, P. G. A.; Eng, J. K.; Hubley, R.; Vogelzang, M.; Deutsch, E. W.; Raught, B.; Pratt, B.; Nilsson, E.; Angeletti, R. H.; Apweiler, R.; Cheung, K.; Costello, C. E.; Hermjakob, H.; Huang, S.; Julian, R. K.; Kapp, E.; McComb, M. E.; Oliver, S. G.; Omenn, G.; Paton, N. W.; Simpson, R.; Smith, R.; Taylor, C. F.; Zhu, W.; Aebersold, R. *Nat. Biotechnol.* **2004**, *22*, 1459–1466.

(45)   Ngoka, L. C. M.; Gross, M. L. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 360–363.

(46)   Paizs, B.; Suhai, S. *Mass Spectrom. Rev.* **2005**, *24*, 508–548.

(47)   Payne, A. H.; Glish, G. L. *Anal. Chem.* **2001**, *73*, 3542–3548.

(48)   Racine, A. H.; Payne, A. H.; Remes, P. M.; Glish, G. L. *Anal. Chem.* **2006**, *78*, 4609–4614.

(49)   Schwartz, J. C.; Syka, J. E. P.; Quarmby, S. T. *Proceedings of the 53rd ASMS Conference on Mass Spectrometry and Allied Topics*, San Antonio, TX, June 5-9, **2005**.

(50)   Schlabach, T.; Zhang, T.; Miller, K.; Kiyonami, R. *The 2006 ABRF Conference*; Long Beach, CA, Feb 11-14, **2006**.

(51)   Eckart, K. *Mass Spectrom. Rev.* **1994**, *13*, 23–55.

(52)   Pittenauer, E.; Zehl, M.; Belgacem, O.; Raptakis, E.; Mistrik, R.; Allmaier, G. *J. Mass Spectrom.* **2006**, *41*, 421–447.

(53)   Kopp, F.; Marahiel, M. A. *Nat. Prod. Rep.* **2007**, *24*, 735–749.

(54)   Simmons, T. L. Ph.D. Thesis, University of California at San Diego, San Diego, CA, **2008**.

(55)   Gutie´ rrez, M.; Gerwick, W. H. Manuscript in preparation.

(56)   Linington, R. G. Manuscript in preparation.

(57)   Renner, M. K.; Shen, Y. C.; Cheng, X. C.; Jensen, P. R.; Frankmoelle, W.; Kauffman, C. A.; Fenical, W.; Lobkovsky, E.; Clardy, J. *J. Am. Chem. Soc.* **1999**, *121*, 11273–11276.

(58)   Selim, S.; Negrel, J.; Govaerts, C.; Gianinazzi, S.; van Tuinen, D. *Appl. Environ. Microbiol.* **2005**, *71*, 6501–6507.

(59)   Nair, S. S.; Romanuka, J.; Billeter, M.; Skjeldal, L.; Emmett, M. R.; Nilsson, C. L.; Marshall, A. G. *Biochim. Biophys. Acta* **2006**, *1764*, 1568–1576.

(60)   Greve, H.; Kehraus, S.; Krick, A.; Kelter, G.; Maier, A.; Fiebig, H. H.; Wright, A. D.; Ko¨nig, G. M. *J. Nat. Prod.* **2008**, *71*, 309–312.

(61)   Adams, B.; Po¨rzgen, P.; Pittman, E.; Yoshida, W. Y.; Westenburg, H. E.; Horgen, F. D. *J. Nat. Prod.* **2008**, *71*, 750–754.

(62)   Wani, M. C.; Taylor, H. L.; Wall, M. E.; Coggon, P.; and McPhail, A. T. *J. Am. Chem. Soc.* **1971**, *93*, 2325-2327.

(63)   Blokhin, A. V.; Yoo, H. D.; Geralds, R. S.; Nagle, D. G.; Gerwick, W. H.; and Hamel, E. *Mol. Pharmacol.* **1995**, *48,* 523-531.

(64)   Singh, R.; Sharma, M.; Joshi, P.; and Rawat, D. S. *Anticancer Agents Med Chem.* **2008**, *8*, 603-617.

(65)   Harvey, A.L. *Drug Discovery Today.* **2008**, *13*, 894-901.

(66)   Newman, D. J.; and Cragg, G. M. *J. Nat. Prod.* **2007**, *70,* 461-477.

(67)   Butler, M. S. *Nat. Prod. Rep.* **2008**, *25*, 475–516.

(68)   Breil, B. T.; Borneman, J.; Triplett, E. W. *J. Bacteriol.* **1996**, *178*, 4150–4156.

(69)   Triplett, E. W.; Breil, B. T.; Splitter, G. A. *Appl. Environ. Microbiol.* **1994**, *60*, 4163–4166.

(70)   Triplett, E. W.; Barta, T. M. *Plant Physiol.* **1987**, *85*, 335–342.

(71)   Triplett, E. W. *Proc Natl Acad Sci USA.* **1988**, *85*, 3810–3814.

(72)   Scupham, A. J.; Triplett, E. W. Journal of Applied Microbiology. **2006**, *100*, 500–507.

(73)   Breil, B. T.; Ludden, P. W.; Triplett, E. W. *J. Bacteriol.* **1993**, *175*, 3693–3702.

(74)   Scupham, A. J.; Dong, Y.; Triplett, E. W. *Appl. Environ. Microbiol.* **2002**, *68*, 4334–4340.

(75)   Lethbridge, B. J. PhD, University of Adelaide, Australia. **1989**.

(76)   Roy, R. S.; Gehring, A. M.; Milne, J. C.; Belshaw, P. J.; Christopher T. Walsh. *Nat. Prod. Rep.*, **1999**, *16*, 249–263.

(77)   Xu, Z. J.; Moffett, D. B.; Peters, T. R.; Smith, L. D.; Perry, B. P.; Whitmer, J.; Stokke, S. A.; Teintze, M. *J. Biol. Chem.* **1995**, *270*, 24858–24863.

(78)   Hubbard, B. R.; Jacobs, M.; Ulrich, M. M.; Walsh, C.; Furie B.; Furie, B. C. *J. Biol. Chem.*, **1989**, *264*, 14145–14150.

(79)   Huber, P.; Schmitz, T.; Griffin, J.; Jacobs, M.; Walsh, C.; Furie, B.; B. C. Furie, *J. Biol. Chem.*, **1990**, 265, 12467–12473.

(80)   Cserzo, M.; Wallin, E.; Simon, I.; von Heijne, G.; Elofsson, A. *Prot. Eng.* **1997**, *10*, 673-676.

(81)   Brown, L. C. W.; Acker, M. G.; Clardy, J.; Walsh, C. T.; Fischbach, M. A. *PNAS*. **2009**, *106*, 2549–2553.

(82)   Lee, S. W.; Mitchell, D. A.; Markley, A. L.; Hensler, M. E.; Gonzalez, D.; Wohlrab, A.; Dorrestein, P. C.; Nizet, V.; Dixon, J. E. *PNAS*. **2008**, *105*, 5879–5884.

(83)   Roush, R. F.; Nolan, E. M.; Lohr, F.; Walsh, C. T. *J. Am. Chem. Soc.* **2008**, *130*, 3603-3609.

(84)   Fomenko, D. E.; Metlitskaya, A. Z.; Peduzzi, J.; Goulard, C.; Katrukha, G. S.; Gening, L. V.; Rebuffat, S.; Khmel, I. A. *Antimicrobial Agents and Chemotherapy.* **2003**, *47*, 2868–2874.

(85)   Watanabe, K.; Hotta, K.; Praseuth, A. P.; Koketsu, K.; Migita, A.; Boddy, C. N.; Wang, C. C. C.; Oguri, H.; Oikawa, H. *Nature Chemical Biology.* **2006***, 2*, 423-428.