

# Supplementary Materials for: Multiplex De Novo Sequencing of Peptide Antibiotics

Hosein Mohimani<sup>1</sup>, Wei-Ting Liu<sup>2</sup>, Yu-Liang Yang<sup>3</sup>, Susana P. Gaudêncio<sup>4</sup>, William Fenical<sup>4</sup>, Pieter C. Dorrestein<sup>2,3</sup>, and Pavel A. Pevzner<sup>5</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, UC San Diego

<sup>2</sup>Department of Chemistry and Biochemistry, UC San Diego

<sup>3</sup>Skaggs School of Pharmacy and Pharmaceutical Sciences, UC San Diego

<sup>4</sup>Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, UC San Diego

<sup>5</sup>Department of Computer Science and Engineering, UC San Diego ,  
Correspondence should be addressed to P.P. (ppezvner@cs.ucsd.edu).

April 4, 2011

## **Text S1. Multiplex Score.**

Let  $(S, S')$  be a spectral pair with *aligned*  $k$ -tags  $Tag = (m_1, \dots, m_k)$  and  $Tag' = (m'_1, \dots, m'_k)$ . Subtags  $m_i \dots m_j$  and  $m'_i \dots m'_j$  form a *matching subtag pair* if  $mass(m_i \dots m_j) \in S$  and  $mass(m'_i \dots m'_j) \in S'$ . The pair of masses of a matching subtag pair,  $mass(m_i \dots m_j)$  and  $mass(m'_i \dots m'_j)$ , form a *matching mass pair*. *RepeatTagPairScore* of the tag pair  $(Tag, Tag')$  is the number of matching subtag pairs, while *UniqueTagPairScore* is the number of matching mass pairs. *TagPairScore* is the average of *UniqueTagPairScore* and *RepeatTagPairScore*.

For example Consider sets  $S = \{ 72, 216, 244, 391, 405, 504, 505, 633, 796, 878, 1009, 1104 \}$  and  $S' = \{ 119, 216, 242, 244, 391, 405, 504, 522, 555, 631, 779, 796 \}$ . consisting of 12 highest intensity peaks from spectra of Tyrocidine A and A1. We illustrate how to score the tag pair  $Tag = (473, 391, 405)$  and  $Tag' = (487, 391, 405)$ , against the spectral pair  $S$  and  $S'$ .

Compound name	Cyclic Sequence										ParentMass
Tyrocidine A	Val	Orn	Leu	Phe	Pro	Phe	Phe	Asn	Gln	Tyr	1269
Tyrocidine A1	Val	<b>Lys</b>	Leu	Phe	Pro	Phe	Phe	Asn	Gln	Tyr	1283
Tyrocidine B	Val	Orn	Leu	Phe	Pro	<b>Trp</b>	Phe	Asn	Gln	Tyr	1308
Tyrocidine B1	Val	<b>Lys</b>	Leu	Phe	Pro	<b>Trp</b>	Phe	Asn	Gln	Tyr	1322
Tyrocidine C	Val	Orn	Leu	Phe	Pro	<b>Trp</b>	<b>Trp</b>	Asn	Gln	Tyr	1347
Tyrocidine C1	Val	<b>Lys</b>	Leu	Phe	Pro	<b>Trp</b>	<b>Trp</b>	Asn	Gln	Tyr	1361
Tyrocidine D	Val	Orn	Leu	Phe	Pro	<b>Trp</b>	<b>Trp</b>	Asn	Gln	<b>Trp</b>	1370
Tyrocidine E	Val	Orn	Leu	Phe	Pro	Phe	Phe	Asn	Gln	<b>Phe</b>	1253
Tryptocidine A	Val	Orn	Leu	Phe	Pro	Phe	Phe	Asn	Gln	<b>Trp</b>	1292
Tryptocidine B	Val	Orn	Leu	Phe	Pro	<b>Trp</b>	Phe	Asn	Gln	<b>Trp</b>	1331

**Table S1:** Variants of Tyrocidines. The four substitution positions relative to Tyrocidine A are shown in bold. Orn refers to the amino acid Ornithine.

Recall that, for  $Tag = (m_1, \dots, m_k)$ ,

$$Tag(s, f) = \sum_{j=s}^{f-1} m_j$$

and construct matrices for  $Tag(*, *)$  and  $Tag'(*, *)$ :

$$\begin{bmatrix} & 473 & 864 \\ \mathbf{796} & & \mathbf{391} \\ \mathbf{405} & \mathbf{878} & \end{bmatrix} \quad \begin{bmatrix} & 487 & 878 \\ \mathbf{796} & & \mathbf{391} \\ \mathbf{405} & 892 & \end{bmatrix}$$

where entries in bold in the left (right) matrix represent elements that are also present in the spectrum  $S$  ( $S'$ ), respectively. These bold elements are represented as 1s in the matrix  $M(S, Tag)$  of a spectrum  $S$  and a tag  $Tag$  as follows

$$M_{s,f} = \begin{cases} 1, & \text{if } Tag(s, f) \in S \\ 0, & \text{otherwise} \end{cases}$$

Then we have

$$M(S, Tag) = \begin{bmatrix} & 0 & 0 \\ 1 & & 1 \\ 1 & 1 & \end{bmatrix} \quad M(S', Tag') = \begin{bmatrix} & 0 & 0 \\ 1 & & 1 \\ 1 & 0 & \end{bmatrix}$$

Now *RepeatTagPairScore* is the number of ones in the elementwise product of  $M(S, Tag)$  and  $M(S', Tag')$ . In this case *RepeatTagPairScore* is 3. *UniqueTagPairScore* is also equal to 3, because none of the pairs  $(Tag(s, f), Tag'(s, f))$  is repeated. Finally, *TagPairScore* is equal to 3, the average of *UniqueTagPairScore* and *RepeatTagPairScore*.

A multitag  $\mathbf{Tag} = (Tag_1, \dots, Tag_m)$  is called an *aligned multitag* if  $Tag_j$  is aligned to  $Tag_1$  for each  $2 \leq j \leq m$ . *MultiplexScore* of an aligned  $k$ -multitag  $\mathbf{Tag} = (Tag_1, \dots, Tag_m)$  with

mass	explanation	mass	explanation	mass	explanation
72		389	FNQ	720	
73		391	FPF, PFF	734	
74		405	NQY	747	
90		408	FFN	748	
114	O, N	460		751	
119		476		762	
135		487		765	O ... F7, L ... N, F7 ... O
148		490		779	
166		504	LFPPF, NQYV, QYVO	796	P ... Y
213	VO	505	PPFN	862	
216		507		867	
225		522		876	
232		535		878	F7 ... L, N ... F4
242	NQ	538	FPFF	895	P ... V
244	FP,PF	552	FNQY	990	
260	LF	587		992	
261	FN	599		1007	O ... Q
266		600		1009	P ... O
291	QY	615		1025	F5 ... L, F6 ... F4
294	FF	616		1039	
334		618	O ... F6, N ... O	1056	L ... Y
342		633	P ... Q	1104	
357	LFP	650		1122	P ... L, F7 ... P, N ... F6
371		651	L ... F7, F7 ... V	1155	L ... V, Q ... F7
388		654		1170	O ... Y

**Table S2:** MALDI-TOF spectrum of Tyrocidine A represented as a set of integer masses. For masses that appear in the theoretical spectrum, the sequence of the corresponding subpeptide is shown. 31 out of 75 masses in the experimental spectrum appear in theoretical spectra. 45 out of 90 (possibly repeating) masses in the theoretical spectrum appear in the experimental spectrum. The three phenylalanines in VOLFPFNQY are shown by F4, F6 and F7 to avoid confusion. O ... F6 is a shorthand for OLFPPF subpeptide. For the details of transforming the experimental spectrum into the set of integer masses see **Results Section**.

2-tag	3-tag	4-tag	5-tag
(504, 765) =>	(244, 260, 765) =>	(244, 260, 230, 535) =>	(244, 260, 230, 388, 147) =>
(114, 1155) =>	(114, 260, 895) =>	(114, 260, 504, 391) =>	(114, 260, 213, 291, 391) =>
(504, 765) =>	(244, 260, 765) =>	(244, 260, 618, 147) =>	(244, 260, 230, 388, 147) =>
6-tag	7-tag	rank of 7-tag	
(244, 260, 111, 119, 388, 147) =>	(244, 260, 111, 119, 274, 114, 147)	1	
(114, 260, 213, 291, 244, 147) =>	(114, 260, 133, 80, 291, 244, 147)	2	
(244, 260, 111, 119, 388, 147) =>	(244, 260, 111, 119, 175, 213, 147)	3	

**Table S3:** Evolution of the three top score tags of Tyrocidine A, constructed by individual sequencing algorithm.

Tyc A			Tyc B			Tyc C		
mass	score	fragment	mass	score	fragment	mass	score	fragment
473	1	VOLF	213	1	VO	756	1	VOLFPW
717	0.98	VOLFPP	473	0.98	VOLF	99	0.94	V
864	0.94	VOLFPPF	99	0.96	V	213	0.86	VO
213	0.66	VO	1308	0.82	ParentMass	473	0.78	VOLF
1269	0.50	ParentMass	903	0.80	VOLFPWF	1347	0.76	ParentMass
102	0.42		1050	0.52		1056	0.60	VOLFPWWN
978	0.40	VOLFPPFN	620	0.36		942	0.40	VOLFPWW
1124	0.34		717	0.20		1186	0.20	
1238	0.24		756	0.16	VOLFPW	1184	0.14	VOLFPWWNQ
374	0.20		1145	0.14	VOLFPWFNQ	501	0.08	
1106	0.16	VOLFPPFNQ	1017	0.10	VOLFPWFN	682	0.08	
1252	0.16		773	0.08		903	0.08	
231	0.08		790	0.08		914	0.08	
1008	0.16		1119	0.06		1039	0.08	

**Table S4:** Spectral profiles for 200 highest-scoring 7-tags of Tyrocidines A, B and C with amino acids sequences VOLFPFFNQY, VOLFPWFNQY, and VOLFPWWNQY, correspondingly. Only 14 masses with the highest values of spectral profiles are shown.

Tyc A tag							Score	Tyc A1 tag							Score	PS
								<b>227</b>	260	97	147	147	114	291	21	22.5
								213	<b>274</b>	97	147	147	114	291	18	21
								213	260	<b>113</b>	147	147	114	291	16	20
213	260	97	147	147	114	291	24	213	260	97	<b>161</b>	147	114	291	14	19
								213	260	97	147	<b>161</b>	114	291	12	18
								213	260	97	147	147	<b>128</b>	291	14	19
								213	260	97	147	147	114	<b>305</b>	17	20.5

**Table S5:** Illustration of the pairwise score for  $Tag = (213, 260, 97, 147, 147, 114, 291)$  and spectra  $S$  and  $S'$  of Tyc A/A1 spectral pair. The substitution residues in seven correspondig tags for  $Tag$  are shown in bold.  $PS$  stands for  $PairwiseScore(Tag, S, S')$ .

Peptide $P$	Peptide $P'$	$MXS(S)$	$MXS(S')$	$MXPS(S, S')$	$MXPS(S', S)$	$\Delta(S, S')$	# $(S, S')$ -shared
Tyc A	Tyc A1	28	23	24	24	3	57
Tyc B	Tyc B1	26	27	25	25	3	66
Tyc C	Tyc C1	24	23	22.5	22.5	2	63
Tyc A	Tyc B	28	26	24.5	24.5	5	57
Tyc A1	Tyc B1	23	27	23.5	24	2.5	54
Tyc B	Tyc C	26	24	23.5	22.5	3	66
Tyc B1	Tyc C1	27	23	23.5	24	2.5	59
Tyc A	Tyc C	28	24	21	22.5	8.5	48
Tyc A1	Tyc C1	23	23	22	21.5	2.5	43
Tyc A	Tyc B1	28	27	23.5	23.5	8	32
Tyc A1	Tyc B	23	26	21	20.5	7.5	29
Tyc A	Tyc C1	28	23	20	20.5	10.5	26
Tyc A1	Tyc C	23	24	19	20	8	24
Tyc B	Tyc C1	26	23	19	20	10	31
Tyc B1	Tyc C	27	24	22	22	7	35

**Table S6:** Identification of putative spectral pairs  $S/S'$  in Tyrocidine family (spectra that represent peptides  $P/P'$  differing by a single amino acid). Peptides in the first seven pairs differ from each other by a single substitution, while peptides in the remaining 8 pairs differ from each other by multiple substitutions.  $MXS$  and  $MXPS$  stand for *MaxScore* and *MaxPairwiseScore*. The average  $\Delta$  is 3 and 7.75 for the first 7 peptide pairs and the last 8 peptide pairs, correspondingly. The number of shared peaks,  $|S \cap S'|$ , plus the number of  $\delta$ -shared peaks,  $|S \cap (S' - \delta)|$ , are also shown, where  $\delta$  is the difference between parent masses of  $S$  and  $S'$ . The average number of such  $(S, S')$ -shared peaks on first seven peptide pairs is 60.2 and 33.5 for the last eight peptide pairs.

Tyc A tag								$Score$	$PS$	Tyc A1 tag								$Score$	$PS$
111	119	243	128	260	244	161	<b>28</b>	<b>23</b>	100	147	243	128	260	244	161	<b>23</b>	17.5		
111	119	175	213	147	244	260	27	<b>23</b>	100	144	261	147	243	128	260	<b>23</b>	17.5		
111	119	388	147	130	114	260	27	21	97	147	260	128	262	225	164	<b>23</b>	<b>18</b>		
102	111	119	388	147	244	158	27	20.5	99	128	260	244	261	147	144	22	<b>18</b>		
80	133	114	260	147	244	291	27	<b>23</b>	98	147	243	128	260	244	163	22	16		

$$\Delta(S, S', TagList, TagList') = MaxScore(TagList, S) + MaxScore(TagList', S') - MaxPairwiseScore(TagList, S, S') - MaxPairwiseScore(TagList', S, S') = 28 + 23 - 23 - 18 = 10$$

**Table S7:** Illustration of maximum score, maximum pairwise score, and  $\Delta$  distance for a list of five top scoring tags of individual sequencing of Tyc A and Tyc A1.  $PS$  stands for *PairwiseScore*.

respect to spectra  $\mathbf{S} = (S_1, \dots, S_m)$  and spectral network  $G$  is defined as:

$$\text{MultiplexScore}(\mathbf{Tag}, \mathbf{S}, G) = \sum_{(j,r) \in E(G)} \text{TagPairScore}(\text{Tag}_j, \text{Tag}_r, S_j, S_r)$$

where  $E(G)$  is the set of edges of the spectral network  $G$ .

### Text S2. Adaptive Spectral Network.

The described multiplex sequencing algorithm assumes that the spectral network was constructed before multiplex sequencing started. Below we introduce a different approach based on an adaptive construction of the spectral network in each step of the multiplex sequencing algorithm. To score  $(k+1)$ -tags, the algorithm uses the best estimation of the spectral network from the highest scoring  $k$ -tags.

At every step we consider a complete graph on all spectra (with edges connecting all pairs of spectra). We define a non-negative *weight function*  $w(j, r)$  on all pairs of spectra  $S_j$  and  $S_r$  from the set of spectra  $\mathbf{S} = (S_1, \dots, S_m)$  and further require that the total weights of all pairs equals to 1:  $\sum_{1 \leq j < r \leq m} w(j, r) = 1$ . Define the *weighted multiplex score* of a multitag  $\mathbf{Tag} = (\text{Tag}_1, \dots, \text{Tag}_m)$ , with respect to spectra  $\mathbf{S}$  and weighting  $w$ , as

$$\begin{aligned} \text{MultiplexScore}_w(\mathbf{Tag}, \mathbf{S}) \\ = \sum_{(j,r) \in E(K_m)} w(j, r) \cdot \text{TagPairScore}(\text{Tag}_j, \text{Tag}_r, S_j, S_r) \end{aligned}$$

One can see that there is a relationship between *MultiplexScore* defined in the main text (for a spectral network  $G$ ) and *MultiplexScore<sub>w</sub>* defined here. In fact, by defining

$$w_G(e) = \begin{cases} 1/|E(G)|, & e \in E(G) \\ 0, & \text{otherwise} \end{cases}$$

the weighted multiplex score of a multitag with respect to weighting  $w_G$  is equal to its multiplex score with respect to spectral network  $G$  divided by  $|E(G)|$ .

Now for a  $k$ -tag  $\text{Tag}$  of spectrum  $S_u$ , define

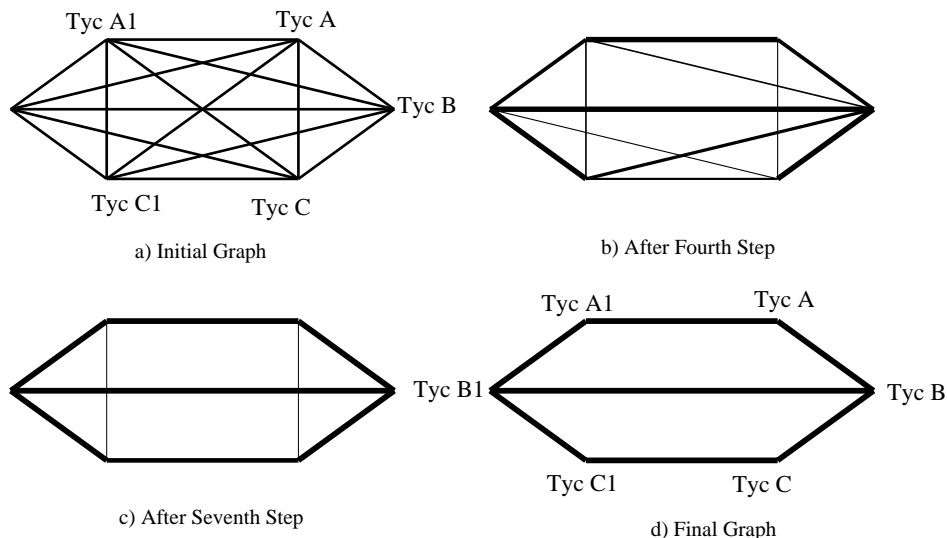
$$\text{MultiplexScore}_w(\text{Tag}, u, \mathbf{S}) = \text{MultiplexScore}_w(\mathbf{multitag}(\text{Tag}, u, \mathbf{S}, K_m), \mathbf{S})$$

We start with a uniform weighting on all edges of complete graph with  $w(j, r) = \frac{2}{m \cdot (m-1)}$ . Similar to previous sections, the algorithm tries to find high scoring tags by using weighted multiplex score. However, after finding high score  $k$ -tags for each  $k$ , it updates the weighting  $w$  as follows. Let  $\mathbf{Tag}_1 \dots \mathbf{Tag}_t$  be the multitags returned by the algorithm at step  $k$ , where  $\mathbf{Tag}_l = (\text{Tag}_l^1, \dots, \text{Tag}_l^m)$ . For each pair of vertices  $(j, r)$ , we define

$$\text{CorrespondenceCount}(j, r) = \#\{l \mid \text{Tag}_l^j \text{ and } \text{Tag}_l^r \text{ are corresponding}\}$$

Then we update weights as follows:

$$w(e) = \frac{\text{CorrespondenceCount}(j, r)}{\sum_{1 \leq j < r \leq m} \text{CorrespondenceCount}(j, r)}$$



**Fig. S1:** Weighting of the spectral network of the Tyrocidine family at different stages of the algorithm. The width of each edge is proportional to its weight (edges with weight zero are not shown). Spectral Networks are shown at (a) the initial step ( $k = 2$ ) with uniform weights on all edges, (b)  $k = 4$ , (c)  $k = 7$ , and (d) the final step ( $k = 10$ ).

and use new weights for scoring  $(k + 1)$ -tags.

The four highest scoring  $k$ -tags of Tyrocidine family reconstructed with weighted multiplex sequencing are the same four highest scoring tags reconstructed by multiplex sequencing (see **Table 1 (c)**). The scores, however, are different. WMS column of **Table 1 (c)** shows weighted multiplex scores of these tags.

**Figure S1** shows how the spectral network improves with each iteration and finally reveals all spectral pairs as edges with positive weights.

### **Text S3. Multiplex sequencing of unknown variants of Tyrocidines.**

We construct the spectral network for spectra of Tyrocidines, connecting spectra  $S$  and  $S'$  by an edge if the number of  $(S, S')$ -shared peaks is greater than a threshold (using  $\Delta$ -scores results in similar spectral networks). Recall that  $\#(S, S')$ -shared peaks is equal to the number of shared peaks,  $|S \cap S'|$ , plus the number of  $\delta$ -shared peaks,  $|S \cap (S' - \delta)|$ , where  $\delta$  is the difference between parent masses of  $S$  and  $S'$ . The resulting spectral network is very complex reflecting the complexity of crude microbial extracts. We therefore choose a very stringent threshold (at least 60  $(S, S')$ -shared peaks) to construct edges in the corresponding spectral network and focus our attention on the connected component of the spectral network that contains known Tyrocidines. Such stringent threshold (see **Table S10**) is chosen to ensure that all spectra in this component represent members of the Tyrocidine family (some of them may represent chemical adducts). **Figure S2** reveals a spectral network of 41 spectra that contains spectra of both known and unknown tyrocidines. However, some vertices in this spectral network correspond to very similar spectra (with parent masses differing by 1-2 Da) that most likely

	Rank	Tag							
(a)	1	99	127	139	268	193	101	115	
	2	99	127	139	93	193	175	216	
	3	99	127	139	75	193	193	216	
	Rank	Tag							
(b)	1	71	113	113	128	113	147	113	113
	2	71	113	128	113	113	147	113	113
	3	71	113	113	128	113	113	147	113

**Table S8:** (a) Top three multiplex sequencing reconstructions of spectra of Cyclomarin A, Cyclomarin C, Dehydro Cyclomarin A and Dehydro Cyclomarin C (projected on Cyclomarin A). (b) Top three multiplex sequencing reconstructions of MALDI-TOF spectra of Reginamides (projected on Reginamide A). Exactly the same results can be achieved by multiplex sequencing of ten ESI-IT spectra of variants of Reginamide with masses 897, 911 (Reginamide A), 925, 939, 953, 967, 981, 995, 1009, 1023.

Sequence								Score
71	113	113	128	113	147	113	113	35.3
71	113	113	128	113	113	147	113	35.0
71	113	128	113	147	113	113	113	33.0
71	113	113	128	147	113	113	113	32.7
71	113	113	113	128	147	113	113	32.3
71	113	113	113	128	113	147	113	32.3
71	113	113	113	128	113	113	147	32.3
71	128	113	113	113	147	113	113	31.7
71	113	113	128	113	113	113	147	31.7
71	128	113	113	147	113	113	113	31.3
71	113	128	113	113	113	147	113	31.0
71	113	113	113	113	128	113	147	28.0
71	128	113	113	113	113	147	113	27.0
71	113	113	113	113	128	147	113	26.7
71	113	128	113	113	113	113	147	26.3
71	128	113	147	113	113	113	113	26.0
71	113	128	147	113	113	113	113	25.0
71	128	113	113	113	113	113	147	24.0
71	128	147	113	113	113	113	113	23.7
71	113	113	113	113	113	128	147	23.0

**Table S9:** All possible sequences of masses formed by 113 (with multiplicity 5), 71, 128, and 147, along with their multiplex score against the Reginamide A spectrum. The three top scoring peptides are the same as the three top peptides returned by the multiplex sequencing algorithm.



Peptide $P$	Peptide $P'$	# $(S, S')$ -shared peaks
Tyc A	Tyc A1	72
Tyc B	Tyc B1	68
Tyc C	Tyc C1	72
Tyc A	Tyc B	60
Tyc A1	Tyc B1	59
Tyc B	Tyc C	61
Tyc B1	Tyc C1	58
Tyc A	Tyc C	61
Tyc A1	Tyc C1	59
Tyc A	Tyc B1	32
Tyc A1	Tyc B	32
Tyc A	Tyc C1	29
Tyc A1	Tyc C	36
Tyc B	Tyc C1	32
Tyc B1	Tyc C	34

**Table S10:** Number of shared peaks of Tyrocidines (analogous of **Table S7** for data-dependent Ion Trap spectra).

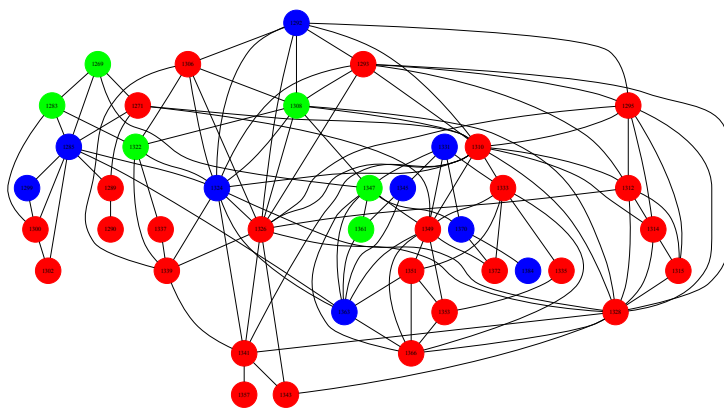
represent isotopic variants or artifacts caused by parent mass errors. We therefore cluster similar spectra (with close parent masses) and represent them as a single vertex in the spectral network. **Figure S3** shows the resulting spectral network and reveals 13 distinct tyrocidine-like peptides.

Since the spectra in this network corresponding to Tyrocidine A, A1, B, B1, C and C1 are well established, we dereplicate [6] the remaining 7 spectra from these 6 spectra using the **multitag** function described in **Figure 3**. The only difference is that, instead of a single input sequence, we input the sequence of all six Tyrocidines A, A1, B, B1, C and C1, and **multitag** algorithm choose the highest scoring possible dereplication for all the other variants. **Table S11** shows the dereplicated peptides (compare with known tyrocidines shown in **Table S12**). Among 5 parents masses that are shared between our results and previous results, in four cases our dereplication matches previously known sequence, and in one case we have a different sequence, probably due to poor fragmentation. In fact, we see peptides with masses 1292, 1331, 1345, 1370 and 1384, which are known to be result of the substitution of Tyrosine by Tryptophan in tenth residue of Tyrocidine A, B, B1, C and C1. We can observe an additional mass, 1306, which is probably result of substitution of Tyrosine by Tryptophan in Tyrocidine A1. However, our current derpelication does not confirm this.

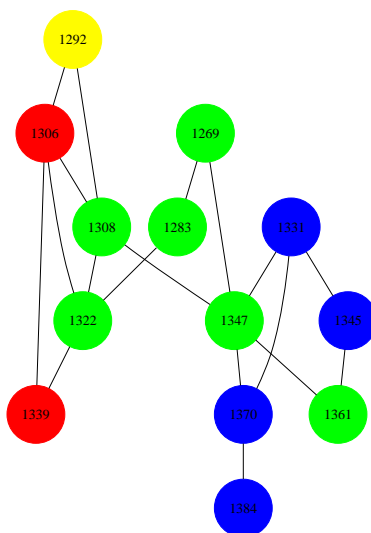
#### **Text S4. Multiplex sequencing of Reginamides.**

Similar to Tyrocidines, we construct the spectral network of Reginamides and sequenced the candidate peptides by dereplication. The results are shown in **Figure S4**, **Figure S5** and **Table S13**.

**Table S9** shows multiplex score of all sequences formed by five 113 residues, 71, 147 and 128 residues. The three top scoring sequences in this table are the same as the three top scoring sequences returned by multiplex sequencing, which confrims the correct reconstruction of order



**Fig. S2:** The spectral network of Tyrocidine family, where two spectra are connected if  $\#(S, S')$ -shared peaks is greater than 60 (each spectrum contains 75 peaks). Green nodes are Tyrocidine A, A1, B, B1, C and C1. Blue nodes are other known variants of tyrocidines. Red nodes stand for candidate variants of Tyrocidines that are not yet reported in the literature. Some of these masses might stand for adducts.

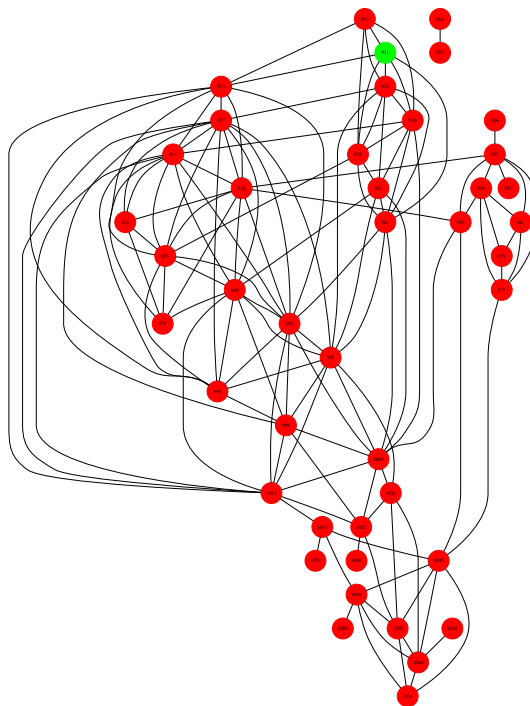


(a)

**Fig. S3:** The spectral network of Tyrocidine family after clustering similar spectra. The sequences are predicted in **Table S11** using the function **multitag** described in **Figure 3**, and sequence of tyrocidines A, A1, B, B1, C and C1 (green nodes) as the inputs. For four of the spectra, predicted sequence matches the previous works (blue nodes). For one of the spectra, the parents mass matches, but the sequence is different from previous works, possibly due to poor fragmentation (yellow node). We also have two spectra, with masses that are not reported previously to our best knowledge (red nodes).

PM	Tag										Score	Comment
1269	99	114	113	147	97	147	147	114	128	163	21	Tyrocidine A
1283	99	128	113	147	97	147	147	114	128	163	26	Tyrocidine A1
1291	99	114	113	147	97	186	147	97	128	163	18	New
1292	99	114	113	147	97	186	131	114	128	163	22	PM matches Tyrocidine A[1]
1306	99	128	113	147	97	186	147	114	112	163	23	New
1308	99	114	113	147	97	186	147	114	128	163	25	Tyrocidine B
1322	99	128	113	147	97	186	147	114	128	163	32	Tyrocidine B1
1331	99	114	113	147	97	186	147	114	128	186	24	Tryptocidine B[1]
1345	99	128	113	147	97	186	147	114	128	186	27	previously reported[1]
1347	99	114	113	147	97	186	186	114	128	163	24	Tyrocidine C
1361	99	128	113	147	97	186	186	114	128	163	30	Tyrocidine C1
1370	99	114	113	147	97	186	186	114	128	186	26	Tyrocidine D[1]
1384	99	128	113	147	97	186	186	114	128	186	24	previously reported[1]

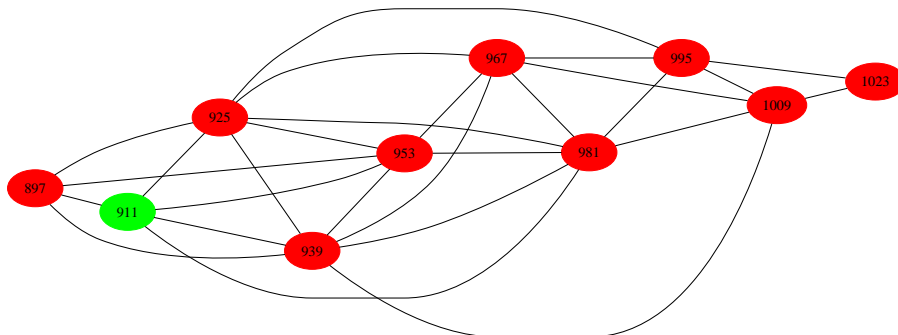
**Table S11:** Reconstructed peptides from the spectra corresponding to vertices in the spectral network shown in **Figure S3**. The spectra were dereplicated using (known) Tyrocidines A, A1, B, B1, C and C1 by applying the **multitag** algorithm described in **Figure 3**. Four of the sequences are reported previously (see **Table S12**). For one spectrum with previously reported parent mass, 1292 Da, our reconstruction slightly differs from that of [1].



**Fig. S4:** The spectral network of Reginamides, where two spectra are connected if  $\#(S, S')$ -shared peaks is greater than 25 (each spectra consists of 40 peaks.) Green node is Reginamide A. Some of the nodes may represent chemical adducts.

Parent Mass	Tag									
1237	99	114	113	147	97	115	147	114	128	163
1251	99	128	113	147	97	115	147	114	128	163
1253	99	114	113	147	97	147	147	114	128	147
<b>1269</b>	<b>99</b>	<b>114</b>	<b>113</b>	<b>147</b>	<b>97</b>	<b>147</b>	<b>147</b>	<b>114</b>	<b>128</b>	<b>163</b>
<b>1283</b>	<b>99</b>	<b>128</b>	<b>113</b>	<b>147</b>	<b>97</b>	<b>147</b>	<b>147</b>	<b>114</b>	<b>128</b>	<b>163</b>
1285	99	114	113	147	97	147	163	114	128	163
1285	99	114	113	147	97	163	147	114	128	163
<b>1292</b>	99	114	113	147	97	147	147	114	128	186
1292	99	114	113	147	97	186	147	114	128	147
1299	99	128	113	147	97	163	147	114	128	163
1299	99	128	113	147	97	147	163	114	128	163
1308	99	114	113	147	97	147	186	114	128	163
<b>1308</b>	<b>99</b>	<b>114</b>	<b>113</b>	<b>147</b>	<b>97</b>	<b>186</b>	<b>147</b>	<b>114</b>	<b>128</b>	<b>163</b>
1322	99	128	113	147	97	147	186	114	128	163
<b>1322</b>	<b>99</b>	<b>128</b>	<b>113</b>	<b>147</b>	<b>97</b>	<b>186</b>	<b>147</b>	<b>114</b>	<b>128</b>	<b>163</b>
1324	115	114	113	147	97	186	147	114	128	163
1324	99	114	113	147	97	163	186	114	128	163
<b>1331</b>	<b>99</b>	<b>114</b>	<b>113</b>	<b>147</b>	<b>97</b>	<b>186</b>	<b>147</b>	<b>114</b>	<b>128</b>	<b>186</b>
1336	99	114	113	147	97	147	214	114	128	163
1338	115	128	113	147	97	186	147	114	128	163
1338	99	128	113	147	97	163	186	114	128	163
<b>1345</b>	<b>99</b>	<b>128</b>	<b>113</b>	<b>147</b>	<b>97</b>	<b>186</b>	<b>147</b>	<b>114</b>	<b>128</b>	<b>186</b>
<b>1347</b>	<b>99</b>	<b>114</b>	<b>113</b>	<b>147</b>	<b>97</b>	<b>186</b>	<b>186</b>	<b>114</b>	<b>128</b>	<b>163</b>
1350	99	128	113	147	97	147	214	114	128	163
<b>1361</b>	<b>99</b>	<b>128</b>	<b>113</b>	<b>147</b>	<b>97</b>	<b>186</b>	<b>186</b>	<b>114</b>	<b>128</b>	<b>163</b>
1363	99	114	113	163	97	186	186	114	128	163
<b>1370</b>	<b>99</b>	<b>114</b>	<b>113</b>	<b>147</b>	<b>97</b>	<b>186</b>	<b>186</b>	<b>114</b>	<b>128</b>	<b>186</b>
<b>1384</b>	<b>99</b>	<b>128</b>	<b>113</b>	<b>147</b>	<b>97</b>	<b>186</b>	<b>186</b>	<b>114</b>	<b>128</b>	<b>186</b>

**Table S12:** All known variants of Tyrocidines, along with their sequences reported in [1]. The variants which are correctly reconstructed by multiplex sequencing are shown in bold.



**Fig. S5:** The spectral network of Reginamides after clustering similar spectra, with threshold of 25. Their sequences are dereplicated from Reginamide A in **Table S13** using the **multitag** algorithm.

Parent Mass	Tag									Score
897	71	99	113	128	113	147	113	113	113	31
911	71	113	113	128	113	147	113	113	113	31
925	71	113	113	142	113	147	113	113	113	25
939	71	113	113	156	113	147	113	113	113	31
953	71	113	113	170	113	147	113	113	113	29
967	71	113	113	184	113	147	113	113	113	28
981	113	85	113	184	113	147	113	113	113	28
995	71	113	113	212	113	147	113	113	113	24
1009	113	113	113	184	113	147	113	113	113	26
1023	71	113	113	240	113	147	113	113	113	20

**Table S13:** Dereplication results of Reginamide variants represented by the spectral network in **Figure S5** from Reginamide A, using **multitag** algorithm.

of residues.

### **Text S5. Accurate reconstruction of amino acid masses.**

Fractional parts of amino acid masses in the peptide sequence are reconstructed using the derived integer masses and high resolution spectra (when available).

Let  $m_i$ ,  $1 \leq i \leq k$ , be the (derived) integer amino acid masses and  $x_i$ ,  $1 \leq i \leq k$ , be the (unknown) fractional parts of amino acid masses. We represent each mass in the experimental spectrum as  $s_j + y_j$  ( $1 \leq j \leq N$ ), where  $s_j$  and  $y_j$  represent integer and fractional parts of the masses, respectively. Our goal is to reconstruct  $x_i$ 's from  $m_i$ 's,  $s_j$ 's and  $y_j$ 's.

Corresponding to each match between theoretical and experimental spectrum,  $\sum_{i_1}^{i_2} m_i = s_j$ , we consider an equation  $\sum_{i_1}^{i_2} x_i = y_j$  on fractional parts of amino acid masses. The number of equations is equal to number of matches, and the number of unknowns is equal to the number of amino acids in the peptide sequence. The system of equations is usually over-determined (the number of matches is greater than the number of amino acids in the peptide sequence). Usually, an overdetermined system of equations is inconsistent, and can not be solved exactly. We find the solution of this system that gives the minimum sum square error.

**Table S14** shows high resolution estimation of masses of **Reginamide A**, and Reginamide variant with integer masses 939 and 953, which are obtained using high resolution spectra. In case of Reginamide A, we also utilize the value of parent mass given by high resolution Orbitrap ESI-MS (912.6282  $[M + H]^+$ ). **Table S14** shows the estimated mass of each amino acid, its estimated elemental composition, and the difference between the estimated mass and the mass of an amino acid with the derived elemental composition.

### **Text S6. NMR Analysis of Reginamide A.**

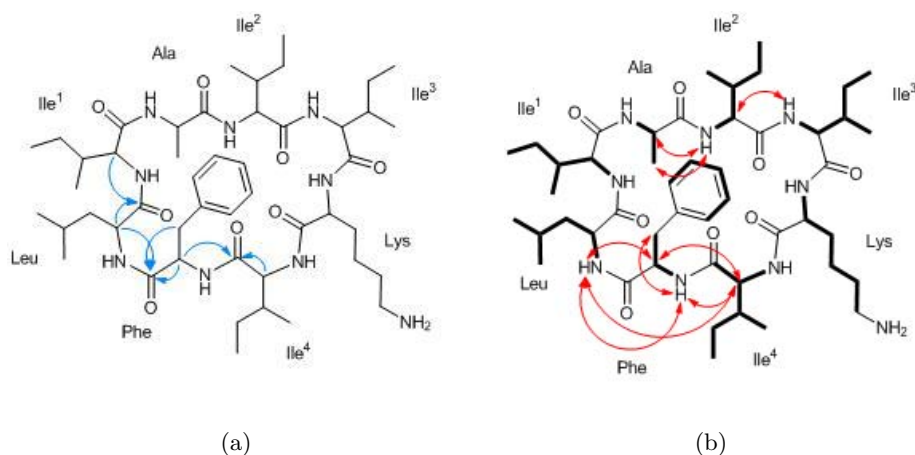
In order to independently validate the derived sequence of Reginamide, we conducted NMR analysis. Reginamide A, obtained as white powder, has the molecular formula  $C_{48}H_{81}N_9O_8$  from the analysis by high resolution ESI-MS (Orbitrap MS with  $m/z$  912.6282  $[M + H]^+$ , with 0.4 ppm error). In the  $^{13}C$  NMR spectrum, eight amide carbonyl carbons were observed, consistent with the 8-mer cyclic peptide that was predicted from multiplex sequencing using mass spectrometry. In addition, eight broad NH signals and eight  $H\alpha$  signals of amino acids were observed in  $^1H$  NMR spectrum, which indicated Reginamide A is a peptide consisting of eight amino acid residues. The amino acid compositions of Reginamide A, deduced from 1D and 2D TOCSY and COSY NMR spectra (in  $CD_3OD/302K$  and  $C_5D_5N/320K$ ) revealed that Reginamide A is formed by one Leu, four Ile, one Phe, one Lys, and one Ala, also in agreement with masses predicted from multiplex sequencing.

Several deuterated solvent have been applied on Reginamide A but none is suitable for collecting both HMBC and ROESY at the same time.  $CD_3OD$  provided the highest quality NMR spectra and the spin-spin splitting or chemical shift of NH and  $H\alpha$  were annotated. However, the fast hydrogen deuterium exchange in  $CD_3OD$  results in the loss of NH signals and their relative correlations in 2D spectrum and therefore we were not able to obtain all the correlations. Nonetheless, the HMBC spectrum (**Figure S6**, in  $CD_3OD/302K$ ) displayed some important  $^3J$  correlations between amide carbonyl carbons and neighboring  $H\alpha$  protons ( $\delta_H$  4.19/ $\delta_{C=O}$  176.0;  $\delta_H$  4.37/ $\delta_{C=O}$  174.0;  $\delta_H$  4.61/ $\delta_{C=O}$  172.3), enabling the partial sequence  $Ile^4$ -Phe-Leu-Ile $^1$ , in agreement with the algorithmically predicted sequence. Subsequently Reginamide A was dissolved in  $C_5D_5N$  to obtain a ROESY spectrum so that

all amide NH signals can be observed. NMR temperature was optimized as 320 K to collect ROE correlations with mixing time 400ms. Two partial sequences,  $Ile^4$ -Phe-Leu and

Reginamide A								
Estimated Mass (Da)	71.03729	113.08406	113.08405	128.09500	113.08404	147.06849	113.08397	113.08402
Estimated Composition	$C_3H_5NO$	$C_6H_{11}NO$	$C_6H_{11}NO$	$C_6H_{12}N_2O$	$C_6H_{11}NO$	$C_9H_9NO$	$C_6H_{11}NO$	$C_6H_{11}NO$
Amino Acid	Ala	Ile	Ile	Lys	Ile	Phe	Leu	Ile
Composition Mass (Da)	71.03711	113.08406	113.08406	128.09496	113.08406	147.06841	113.08406	113.08406
Mass Error (Da)	0.00018	0.00000	-0.00001	0.00004	-0.00002	0.00008	-0.00009	-0.00004
Reginamide 939								
Estimated Mass (Da)	71.03724	113.08442	113.08438	156.09046	113.08427	147.06902	113.08442	113.08452
Estimated Composition	$C_3H_5NO$	$C_6H_{11}NO$	$C_6H_{11}NO$	$C_7H_{12}N_2O_2$	$C_6H_{11}NO$	$C_9H_9NO$	$C_6H_{11}NO$	$C_6H_{11}NO$
Amino Acid	Ala	Ile	Ile	non-standard	Ile	Phe	Leu	Ile
Composition Mass (Da)	71.03711	113.08406	113.08406	156.08980	113.08406	147.06841	113.08406	113.08406
Mass Error (Da)	0.00013	0.00036	0.00032	0.00066	0.00021	0.00061	0.00036	0.00046
Reginamide 953								
Estimated Mass (Da)	71.03695	113.08461	113.08444	170.10551	113.08426	147.06895	113.08397	113.08462
Estimated Composition	$C_3H_5NO$	$C_6H_{11}NO$	$C_6H_{11}NO$	$C_8H_{14}N_2O_2$	$C_6H_{11}NO$	$C_9H_9NO$	$C_6H_{11}NO$	$C_6H_{11}NO$
Amino Acid	Ala	Ile	Ile	non-standard	Ile	Phe	Leu	Ile
Composition Mass (Da)	71.03711	113.08406	113.08406	170.10643	113.08406	147.06841	113.08457	113.08406
Mass Error (Da)	-0.00016	0.00055	0.00038	-0.00092	0.00020	0.00054	-0.00060	0.00056
Reginamide 897								
Estimated Mass (Da)	71.03699	99.06911	113.08412	128.09495	113.08459	147.06835	113.08455	113.08427
Estimated Composition	$C_3H_5NO$	$C_5H_9NO$	$C_6H_{11}NO$	$C_6H_{12}N_2O$	$C_6H_{11}NO$	$C_9H_9NO$	$C_6H_{11}NO$	$C_6H_{11}NO$
Amino Acid	Ala	Val	Ile	Lys	Ile	Phe	Leu	Ile
Composition Mass (Da)	71.03711	99.06840	113.08406	128.09496	113.08406	147.06841	113.08406	113.08406
Mass Error (Da)	-0.00012	0.00071	0.00006	0.00001	0.00053	-0.00006	0.00049	-0.00021

**Table S 14:** High resolution mass of residues of Reginamide A and Reginamides with precursor masses 939, 953 and 897. The closest elemental composition to each estimated mass is also shown. The elemental composition is calculated using the Elcomp software (<http://library.med.utah.edu/masspec/elcomp.htm>), by setting the tolerance to 0.001, and maximum number of Carbon, Hydrogen, Nitrogen and Oxygen atoms to twenty. In all the cases, software returned a unique composition. The non-standard elemental compositions  $C_7H_{12}N_2O_2 + H_2O$  and  $C_8H_{14}N_2O_2 + H_2O$  match the compositions of amino acids Acetyl-Ornithine and Acetyl-Lysine, correspondingly (the only compositions in PDBeChem matching these elemental compositions). For Reginamide A and Reginamide 897, we also have calculated  $[M + H]^+$  of 912.6282 and 898.6149 respectively, which matches our results here with error of 0.4 and 2.8 ppm respectively.



**Fig. S6:** Key 2D NMR correlations of Reginamide A. Blue is  $^1\text{H} - ^{13}\text{C}$  HMBC correlation (left, observed in  $\text{CD}_3\text{OD}$  302 K); red is  $^1\text{H} - ^1\text{H}$  ROESY correlation (right, observed in  $\text{C}_5\text{D}_5\text{N}$  320 K); black bold line is  $^1\text{H} - ^1\text{H}$  TOCSY correlation (right, observed in  $\text{C}_5\text{D}_5\text{N}$  320 K)

$\text{Ala-Ile}^2\text{-Ile}^3$ , were confirmed by strong inter-residues  $\text{H}\alpha/\text{NH}$  ROE correlations in agreement with multiplex sequencing analysis. These data, together with multiplex sequencing results indicate that Reginamide A is cyclic- $\text{Ala-Ile}^2\text{-Ile}^3\text{-Lys-Ile}^4\text{-Phe-Leu-Ile}^1$ . Based on the mass spectrometry analysis it is anticipated (based on the 14 Da shifts) that the lysine is the residue that is modified or preferentially substituted with a alkane-chain functionality.

**Table S15** show  $^1\text{H}$  (600 MHz) and  $^{13}\text{C}$  (125 MHz) NMR data of Reginamide A. ( $\text{CD}_3\text{OD}$ , 302 K). Asterisks means exchangeable. **Table S16** shows  $^1\text{H}$  (600 MHz) NMR data of Reginamide A ( $\text{C}_5\text{D}_5\text{N}$ , 320 K). Additional NMR data are shown in **Figures S7-S20**.

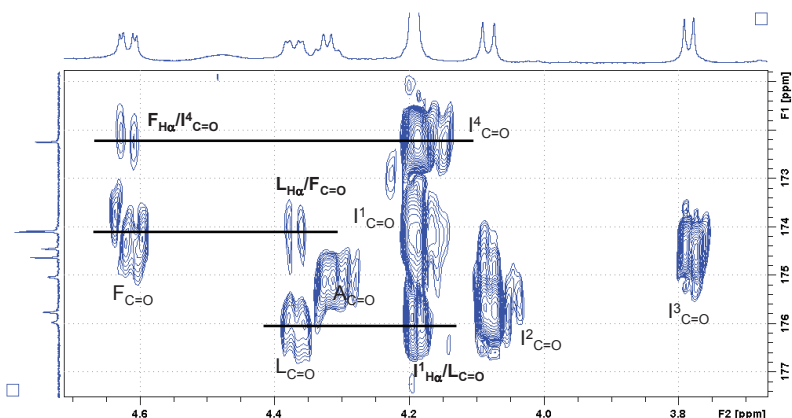


	$\delta_H$	$\delta_C$		$\delta_H$	$\delta_C$
<b>Ile<sup>1</sup></b>			<b>Lys</b>		
<i>NH</i>	6.89, <i>brs</i> *		<i>NH</i>	7.50, <i>brs</i>	
$\alpha$	4.19, <i>m</i>	59.7	$\alpha$	4.48, <i>brs</i>	53.3
$\beta$	1.30, <i>m</i>	26.9	$\beta$	1.65, <i>m</i>	32.5
	1.37, <i>m</i>		$\beta$	1.73, <i>m</i>	
$\gamma$	0.97, <i>m</i>	12.1	$\gamma$	1.30, <i>m</i>	30.8
	2.02, <i>m</i>	37.1	$\gamma$	1.52, <i>m</i>	27.6
$\delta$	0.97, <i>m</i>	16.1	$\gamma$	1.60, <i>m</i>	
<i>C = O</i>		174.1	$\epsilon$	2.94, <i>m</i>	40.8
			<i>C = O</i>		174.7
<b>Ala</b>			<b>Ile<sup>4</sup></b>		
<i>NH</i>	7.74, <i>d</i> , $J = 4.2Hz$		<i>NH</i>	8.33, <i>d</i> , $J = 5.9Hz^*$	
$\alpha$	4.32, <i>brt</i> , $J = 5.9Hz$	50.1	$\alpha$	4.19, <i>m</i>	60.2
$\beta$	1.4, <i>d</i> , $J = 6.8Hz$	19.3	$\beta$	1.30, <i>m</i>	27.3
<i>C = O</i>		175.1		1.37, <i>m</i>	
			$\gamma$	0.97, <i>m</i>	11.9
<b>Ile<sup>2</sup></b>				2.02, <i>m</i>	37.2
<i>NH</i>	8.45, <i>brs</i>		$\delta$	0.97, <i>m</i>	15.6
$\alpha$	4.08, <i>d</i> , $J = 10.3Hz$	60.3	<i>C = O</i>		172.3
$\beta$	1.28, 1.67	26.2			
			<b>Phe</b>		
$\gamma$	0.95, <i>m</i>	10.7	<i>NH</i>	8.51, <i>d</i> , $J = 7.2Hz$	
	1.88, <i>m</i>	36	$\alpha$	4.61, <i>m</i>	56.8
$\delta$	0.97, <i>m</i>	15.4	$\beta$	2.76, <i>dd</i> , $J = 13.7, 12.7Hz$	37.6
			<i>Ph</i>	7.23, <i>m</i>	127.9
<b>Ile<sup>3</sup></b>				7.23, <i>m</i>	127.9
<i>NH</i>	8.11, <i>brs</i>			7.23, <i>m</i>	129.6
$\alpha$	3.78, <i>d</i> , $J = 7.7Hz$	60.8		7.3, <i>m</i>	130.1
$\beta$	0.97, <i>m</i>	26.5	<i>C = O</i>		138.9
	1.45, <i>m</i>				174.0
$\gamma$	0.49, <i>d</i> , $J = 6.8Hz$	15.3	<b>Leu</b>		
	1.6, <i>m</i>	36.6	<i>NH</i>	8.05, <i>d</i> , $J = 7.0Hz$	
$\delta$	0.81, <i>t</i> , $J = 7.3Hz$	11.3	$\alpha$	4.37, <i>brd</i> , $J = 7.6Hz$	54.6
<i>C = O</i>		174.5	$\beta$	1.65, <i>m</i>	25.9
				1.83, <i>m</i>	
			$\gamma$	2.19, <i>m</i>	40.9
			$\delta$	0.98, <i>m</i>	21.6
				1.05, <i>d</i> , $J = 6.5Hz$	23.9
			<i>C = O</i>		176.0

**Table S15:**  $^1H$  (600 MHz) and  $^{13}C$  (125 MHz) NMR data of Reginamide A. ( $CD_3OD$ , 302 K). Asterisk means exchangeable.

	$\delta_H$		$\delta_H$
<b>Ile<sup>1</sup></b>		<b>Lys</b>	
NH	7.37, <i>brs</i>	NH	9.38, <i>brs</i> *
$\alpha$	4.83, <i>m</i>	$\alpha$	5.06, <i>m</i>
$\beta$	1.70, <i>m</i>	$\beta$	2.10, <i>m</i>
$\gamma$	1.19, <i>d</i> , $J = 6.9Hz$	$\gamma$	1.47, <i>m</i>
	2.45, <i>m</i>	$\delta$	1.69, <i>m</i>
$\delta$	0.98, <i>t</i> , $J = 7.2Hz$	$\epsilon$	3.10, <i>m</i>
<b>Ala</b>		<b>Ile<sup>4</sup></b>	
NH	8.31, <i>brs</i>	NH	7.47, <i>brs</i> *
$\alpha$	5.18, <i>q</i> , $J = 6.7Hz$	$\alpha$	4.23, <i>dd</i> , $J = 8.0, 4.5Hz$
$\beta$	1.77, <i>d</i> , $J = 6.7Hz$	$\beta$	0.97, <i>m</i>
			1.52, <i>m</i>
<b>Ile<sup>2</sup></b>		$\gamma$	0.67, <i>d</i> , $J = 6.8Hz$
NH	9.74, <i>d</i> , $J = 5.1Hz$		1.85, <i>m</i>
$\alpha$	4.45, <i>m</i>	$\delta$	0.57, <i>t</i> , $J = 6.7Hz$
$\beta$	1.33, <i>m</i>		
	1.69, <i>m</i>	<b>Phe</b>	
$\gamma$	0.94, <i>d</i> , $J = 6.5Hz$	NH	9.48, <i>brs</i>
	2.20, <i>m</i>	$\alpha$	5.29, <i>ddd</i> , $J = 11.5, 8.0, 3.2Hz$
$\delta$	0.70, <i>t</i> , $J = 7.4Hz$	$\beta$	3.11, <i>dd</i> , $J = 14.6, 8.0Hz$
			3.89, <i>dd</i> , $J = 14.6, 3.2Hz$
<b>Ile<sup>3</sup></b>		<i>Ph</i>	7.22, <i>t</i> , $J = 7.5Hz$
NH	9.20, <i>d</i> , $J = 5.4Hz$		7.28, <i>t</i> , $7.5Hz$
$\alpha$	4.70, <i>dd</i> , $J = 5.4, 4.4Hz$		7.37, <i>d</i> , $J = 7.5Hz$
$\beta$	1.52, <i>m</i>		
	1.70, <i>m</i>	<b>Leu</b>	
$\gamma$	1.21, <i>d</i> , $J = 7.1Hz$	NH	8.60, <i>d</i> , $J = 7.3Hz$
	2.39, <i>m</i>	$\alpha$	4.99, <i>m</i>
$\delta$	0.91, <i>t</i> , $J = 7.4Hz$	$\beta$	1.98, <i>m</i>
		$\gamma$	2.26, <i>ddd</i> , $J = 13.8, 10.0, 3.7Hz$
			2.71, <i>m</i>
		$\delta$	0.94, <i>d</i> , $J = 6.7Hz$
			1.01, <i>d</i> , $J = 6.7Hz$

**Table S16:**  $^1H$  (600 MHz) NMR data of Reginamide A. ( $C_5D_5N$ , 320 K). Asterisk means exchangeable.



**Fig. S7:** Key HMBC correlations (600MHz,  $CD_3OD$ , 302 K,  $^{2,3}J = 7$  Hz) of Reginamide A

### Text S7. Parameter Setting.

The proposed method requires setting few parameters, such as the peptides length  $k$ , and the threshold on  $\delta(S, S')$  for accepting an edge in the spectrum graph. Choice of  $\delta(S, S')$  can always be avoided by using adaptive approach explained in **Text S2**. Setting peptide length can be achieved by the following procedure. By starting from a small peptide length, by each unit of increase in peptide length the optimal peptide score jumps significantly. However, after taking more steps, the score remains constant. Optimal peptide length should be chosen at this phase transition.

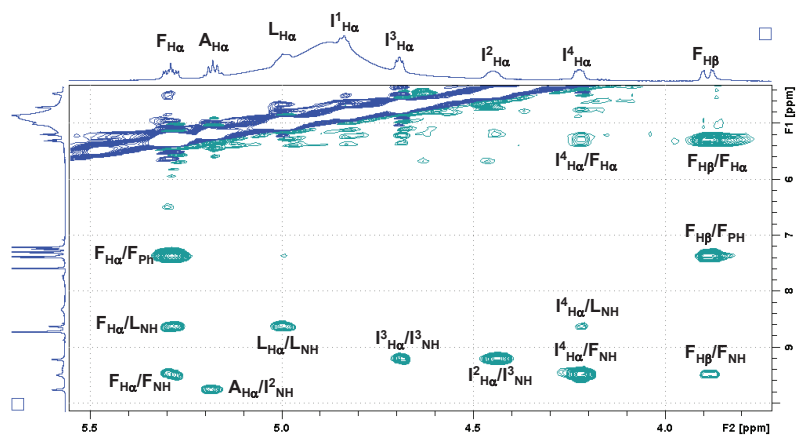


Fig. S8: Key ROSEY correlations (600MHz,  $C_5D_5N$ , 320 K, mixing time 400ms) of Reginamide A

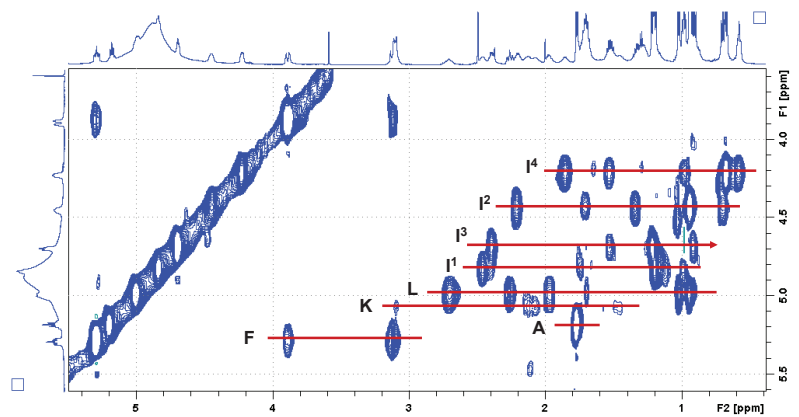
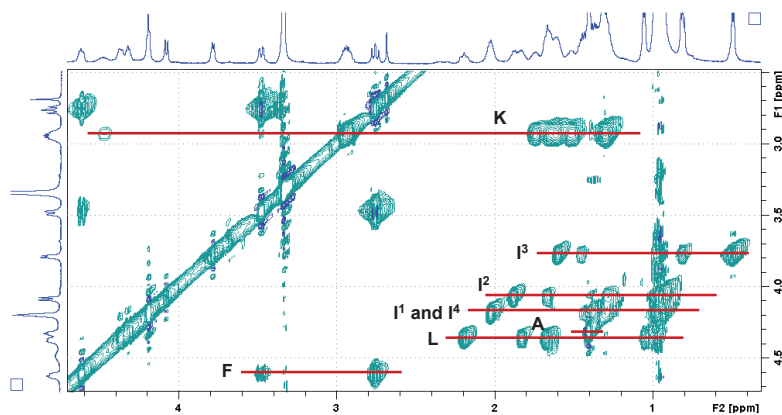
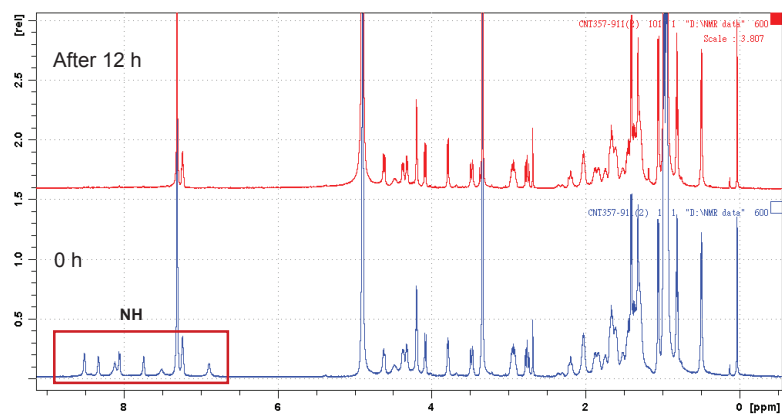


Fig. S9: Key TOCSY correlations (600MHz,  $C_5D_5N$ , 320 K, mixing time 90ms) of Reginamide A



**Fig. S10:** Key TOCSY correlations (600MHz,  $CD_3OD$ , 302 K, mixing time 90 ms) of Reginamide A



**Fig. S11:**  $^1H$  spectra (600MHz,  $CD_3OD$ , 302 K) of Reginamide A

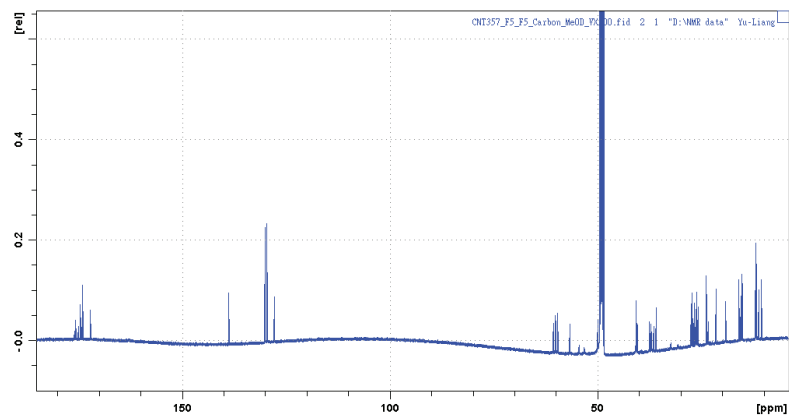


Fig. S12:  $^{13}\text{C}$  NMR spectra (125MHz,  $\text{CD}_3\text{OD}$ , 302 K) of Reginamide A

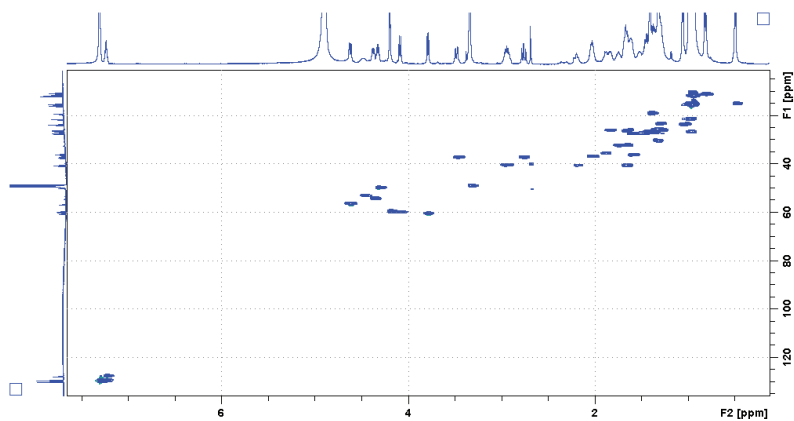
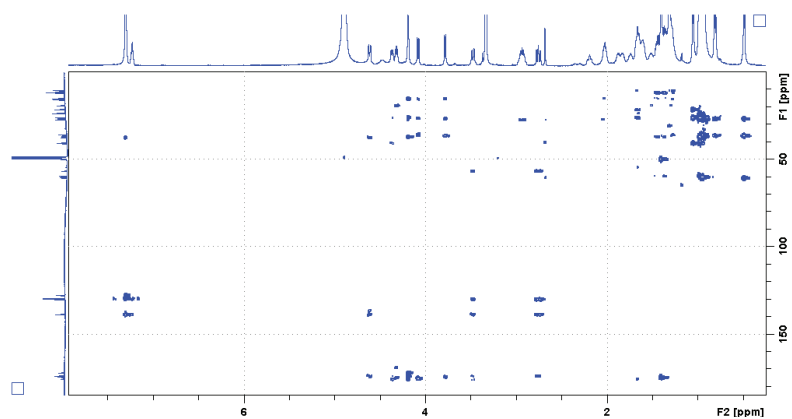
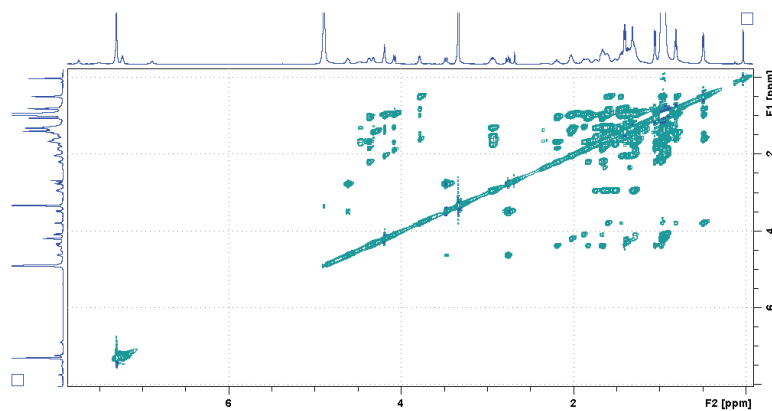


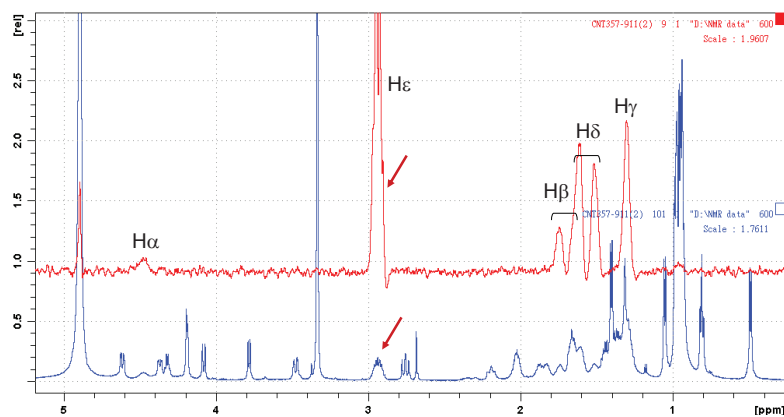
Fig. S13: HSQC spectrum (600MHz,  $\text{CD}_3\text{OD}$ , 302 K,  $^1J = 145$  Hz) of Reginamide A



**Fig. S14:** HMBC spectrum (600MHz,  $CD_3OD$ , 302 K,  $^{2,3}J = 7$  Hz) of Reginamide A

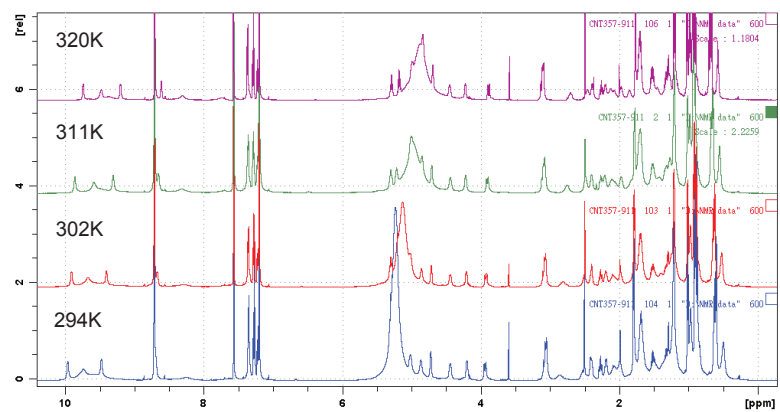


**Fig. S15:** TOCSY spectrum (600MHz,  $CD_3OD$ , 302 K, mixing time 90 ms) of Reginamide A

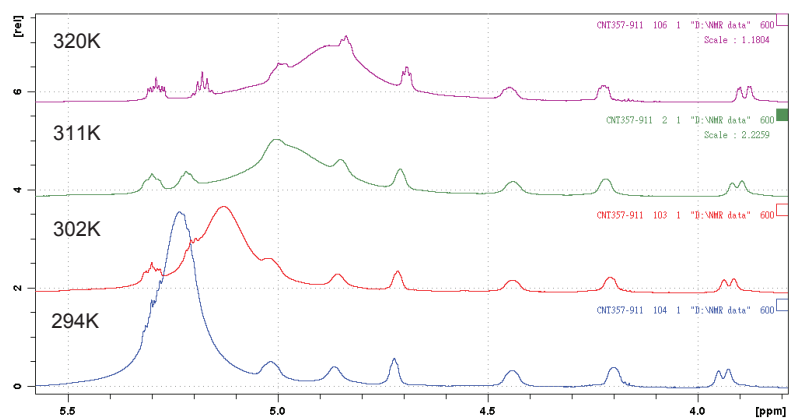


**Fig. S16:** 1D TOCSY spectrum (red) and  $^1\text{H}$  NMR spectrum (blue) of Reginamide A.  $H_\epsilon$  was irradiated by 90 ms mixing time

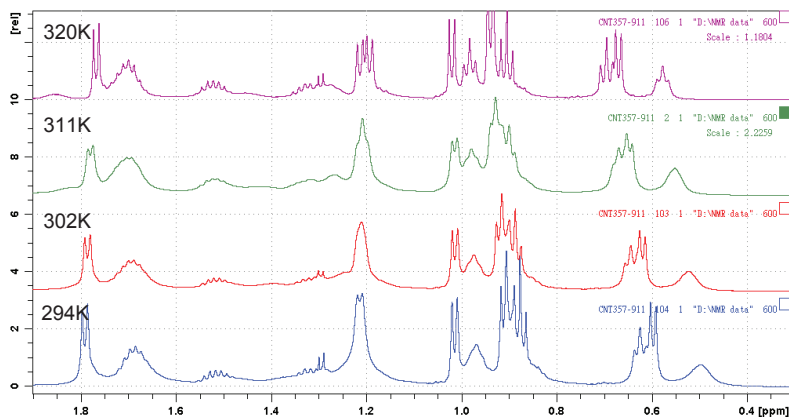




(a)

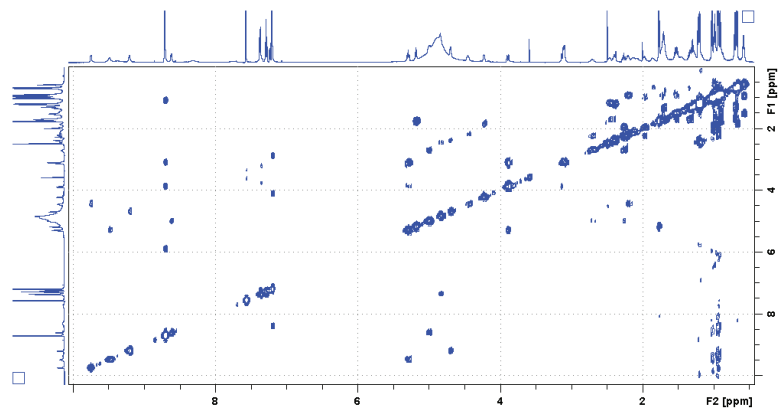


(b)

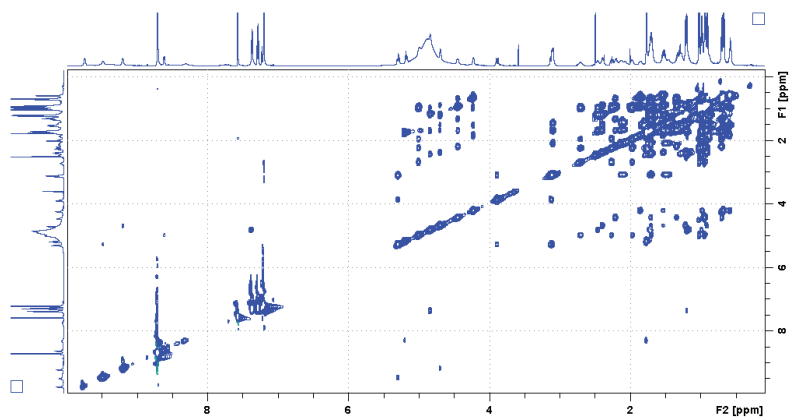


(c)

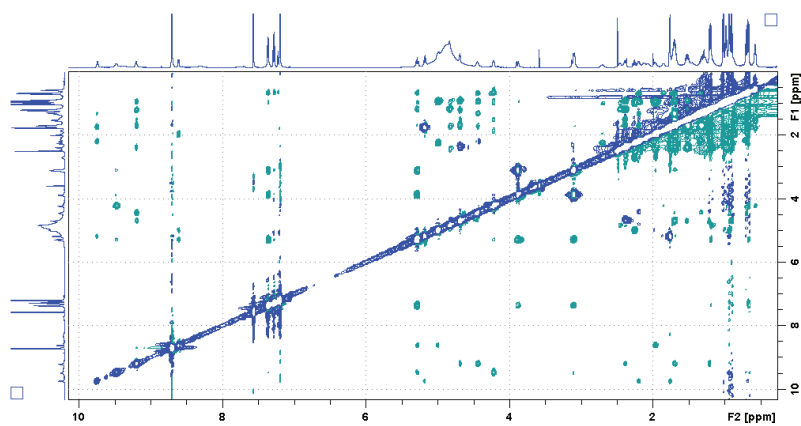
Fig. S17:  $^1\text{H}$  NMR spectra (600MHz,  $\text{C}_5\text{D}_5\text{N}$ ) of Reginamide A at different temperatures.



**Fig. S18:** COSY spectrum (600MHz,  $C_5D_5N$ , 320 K) of Reginamide A



**Fig. S19:** TOCSY spectrum (600MHz,  $C_5D_5N$ , 320 K, mixing time 90ms) of Reginamide A



**Fig. S20:** ROSEY spectrum (600MHz,  $C_5D_5N$ , 320 K, mixing time 400ms) of Regimamide A