Supplementary Information: Increased diversity of peptidic natural products revealed by modification-tolerant database search of mass spectra

Alexey Gurevich[1], Alla Mikheenko[1], Alexander Shlemov[1], Anton Korobeynikov[1,2], Hosein Mohimani[3,4], and Pavel A. Pevzner[1,3,*]

[1]Center for Algorithmic Biotechnology, Institute of Translational Biomedicine, St. Petersburg State University, St. Petersburg, Russia
[2]Department of Mathematics and Mechanics, St. Petersburg State University, St. Petersburg, Russia
[3]Department of Computer Science and Engineering, University of California, San Diego, La Jolla, CA, USA
[4]Department of Computational Biology, Carnegie Mellon University, Pittsburgh, PA, USA
[*]ppevzner@ucsd.edu

Supplementary Tables

Supplementary Table 1. Information about spectral datasets. *Instrument* stands for mass-spectrometry hardware used for obtaining the dataset. *Size* of each dataset is given in initial state (downloaded from the GNPS storage [1]) and after the preprocessing procedure (peaks merging and filtering) performed by VarQuest, both are given in GB. *GNPS ID* stands for unique spectral dataset identifier at the GNPS storage.

Dataset	$Spectra_{PSEUD}$	$Spectra_{STREP_1}$	$Spectra_{STREP_2}$	$Spectra_{CYANO}$	$Spectra_{GNPS}$
# spectra	413234	178635	473135	11230344	$\approx$ 130 millions
Origin	Pseudomonas	Streptomyces	Streptomyces	Cyanobacteria	-
Instrument	microTOF-Q	LTQ Orbitrap	Q-TOF	maXis	-
Initial size	0.3	8.4	142.5	28.2	-
Prepr. size	0.1	0.3	2.5	12.0	-
GNPS ID	MSV-79450	MSV-78604	MSV-78839	MSV-78568	120 datasets

Supplementary Table 2. Performance of four PNP identification approaches on various spectral datasets against the PNPdatabase with 5021 PNPs. *Preprocessing* stands for the time needed for constructing the list of candidate peptides for all input spectra; *Processing* stands for the time needed for scoring and statistical significance estimation of all considered PSMs. *Peak RAM* stands for maximal RAM usage. $Spectra_{GNPS}$ was processed as 120 independent datasets, so the running time is a sum of 120 independent running times. Spectral networks were constructed using the GNPS interface (the Molecular Networking workflow [1]) with precursor and fragment ion tolerance set to 0.02 Da and all other parameters set to the default values. When processing $Spectra_{CYANO}$, the SpecNets tool crashed after a week of execution, and the BruteForce approach crashed due to exceeding memory limit. All tools were run on a 12 CPU Intel Xeon X7560 2.27GHz cluster node with 25 GB RAM.

Method	Time (hh:mm)		Peak RAM
	Preprocessing	Processing	(GB)
$Spectra_{PSEUD}$			
Standard	0:00	0:04	0.4
SpecNets	0:47	0:12	0.5
BruteForce	0:00	6:52	0.8
VarQuest	0:01	3:16	0.4
$Spectra_{STREP_1}$			
Standard	0:00	0:04	2.7
SpecNets	1:23	0:15	2.7
BruteForce	0:00	10:38	5.0
VarQuest	0:00	4:41	2.9
$Spectra_{STREP_2}$			
Standard	0:00	0:22	5.2
SpecNets	3:29	0:43	5.3
BruteForce	0:00	4:46	16.0
VarQuest	0:01	1:36	7.3
$Spectra_{CYANO}$			
Standard	0:01	0:37	14.5
VarQuest	0:29	19:48	15.8
$Spectra_{GNPS}$			
Standard	0:16	10:39	16.0
VarQuest	40:35	1681:06	18.9

Supplementary Table 3. Cyclosporin family in the PNPdatabase (47 cyclic PNPs, each of them 11 amino acid long). 29 of them were identified in $Spectra_{GNPS}$ as unmodified peptides (both by DEREPLICATOR and VarQuest). The remaining 18 variants were identified only as peptide variants (VarQuest only), each of them is represented by a single variant, mass shift is specified in a PNP name and the most likely position for modification/mutation is shown in *Structure* column by bold and underlined font. *Compound name* stands for the PNP description in its original database (35 came from DNP [2] and 12 from AntiMarin [3]). *Mass* is the integer mass in Da, modification/mutation mass shift is taken into account for the novel variants. *Structure* column shows rounded monoisotopic masses of amino acids (in Da) starting from the largest one (in most cases it is Bmt with mass 183 Da). Abbreviations: *Bmt* for butenyl-methyl-threonine, *pen* for pentanoic acid, *but* for butanoic acid.

# Compound name	Mass	Structure
Identified in $Spectra_{GNPS}$ as known peptides		
1 NVA2-hydroxy-MeL4-cyclosp	1232	183 113 127 127 71 71 127 99 143 71 99
2 (dihyd-MeBmt)1-(hydroxy-MeL)4-cyclosp	1220	185 113 127 127 71 71 127 99 143 71 85
3 Ser-cyclosp	1218	183 113 127 127 87 71 127 99 127 71 85
4 (8-hydroxy-MeBmt)1-cyclosp	1218	199 113 127 127 71 71 127 99 127 71 85
5 Antibiotic FR 901459	1218	183 113 113 127 71 71 127 113 127 71 101
6 Cyclosporin C	1218	183 113 127 127 71 71 127 99 127 71 101
7 hydroxy-MeL4-cyclosp	1218	183 113 127 127 71 71 127 99 143 71 85
8 Cyclosporin V	1216	183 113 127 127 71 85 127 99 127 71 85
9 (MeVal)5-cyclosp	1216	183 113 127 127 71 71 127 113 127 71 85
10 Isocyclosp D	1216	183 113 127 127 71 71 127 99 127 71 99
11 Cyclosporin P	1204	169 113 127 127 71 71 127 99 127 71 101
12 Cyclosporin W	1204	183 99 127 127 71 71 127 99 127 71 101
13 Cyclosporin Y	1202	183 113 127 127 71 71 113 99 127 71 99
14 Cyclosporin X	1202	183 113 127 113 71 71 127 99 127 71 99
15 Cyclosp-N*-De-Me	1202	183 113 113 127 71 71 127 99 127 71 99
16 Cyclosp-9-(N-Met-Ile)	1202	183 113 127 127 71 71 127 99 127 71 85
17 (MeThr)4-cyclosp	1190	183 113 127 127 71 71 127 99 115 71 85
18 Cyclosporin E	1188	183 99 127 127 71 71 127 99 127 71 85
19 Cyclosporin L	1188	169 113 127 127 71 71 127 99 127 71 85
20 Cyclosporin U	1188	183 113 127 127 71 71 113 99 127 71 85
21 Cyclosporin B	1188	183 113 127 127 71 71 127 99 127 71 71
22 Cyclosp-4-(2-(Metamino)pen)	1188	183 113 113 127 71 71 127 99 127 71 85
23 Cyclosp-10-(2-Aminobut)	1188	183 113 127 127 71 71 127 85 127 71 85
24 Cyclosp-N9-De-Me	1188	183 113 127 127 71 71 127 99 113 71 85
25 Cyclosp-Deoxy	1186	167 113 127 127 71 71 127 99 127 71 85
26 Cyclosporin Z	1174	155 113 127 127 71 71 127 99 127 71 85
27 Cyclosp-N5-De-Me	1174	183 99 127 127 71 71 127 99 127 71 71
28 Cyclosporin R	1174	183 113 113 127 71 71 113 99 127 71 85
29 Cyclosporin Q	1174	183 113 127 127 71 71 127 99 99 71 85
Identified in $Spectra_{GNPS}$ only as peptide variants		
30 4,8-Di-hydroxy-MeL-Cycl A **+131Da**	1407	183 113 **127** 127 129 71 127 99 143 71 85
31 (Hiv-L-T)-Cycl A 10-L **+104Da**	1351	183 113 113 127 100 71 **127** 113 127 71 101
32 (Hiv-L-T)-Cycl A 8-G,10-L **+106Da**	1339	183 113 113 127 **100** 71 127 113 127 57 101
33 Cyclosp-10-Ile **+8Da**	1240	183 113 127 127 71 71 127 **113** 127 71 101
34 (Hiv-L-T)-Cycl A **+7Da**	1240	183 113 **113** 127 100 71 127 99 127 71 101
35 Cyclosp-7-Hydroperoxide **-11Da**	1237	215 113 127 127 71 71 127 99 127 71 **99**
36 Cyclosp-2-(O-(2-hydroxyeth)-Ser) **-27Da**	1235	183 113 127 **187** 71 71 127 99 127 71 85
37 (Hiv-L-T)-Cycl A 8-hydroxy,10-L **-28Da**	1235	199 113 113 127 100 71 **127** 113 127 71 101
38 (Hiv-L-T)-Cycl A 9,10-Di-Leu **+1Da**	1234	183 113 113 **127** 100 71 127 113 113 71 101
39 Allylgly-cyclosp **+12Da**	1226	**183** 113 127 127 71 71 127 99 127 71 97
40 (g-hydroxy-MeLeu)8-cyclosp **-41Da**	1219	183 113 127 127 129 71 127 99 **127** 71 85
41 Cyclosporin S **+28Da**	1218	183 113 **127** 127 71 71 127 99 99 71 101
42 Cyclosp-4-Leu,-10-Ala **+42Da**	1218	183 113 113 127 71 71 127 71 **127** 71 101
43 (Thr(2),Leu(5),Ala(10))cyclosp **+14Da**	1190	183 113 71 127 71 71 127 113 **127** 71 101

# Compound name	Mass	Structure
44 (MeOT-2,MeA-3,MeV-5)cyclosp **-76Da**	1184	183 113 127 **127** 71 71 127 113 127 85 115
45 Cyclosporin O **-67Da**	1093	127 113 127 127 71 71 127 99 **127** 71 99
46 Cyclosporin K **-114Da**	1086	167 113 127 127 71 71 127 99 **127** 71 99
47 Cyclosporin J **-124Da**	1022	**127** 113 127 127 71 71 127 99 127 71 85

Supplementary Table 4. List of common mass offsets identified by VarQuest in $Spectra_{GNPS}$. Only offsets identified in at least 30 distinct peptides are shown. The total offset range $-300 \ldots 300$ Da was divided into non-overlapping intervals of size 0.1 Da. All PSMs were grouped according to modified amino acid mass and offset. The groups were sorted by the number of unique peptides. *Offset* is the mass difference between a known PNP and its identified variant (in Da). *Residue* is the amino acid corresponding to the most likely location of the modification in a PNP graph. The fraction of the amino acid among all amino acids with the same mass offset is given in the parenthesis. Numbers of distinct PNPs, PNP families and PSMs identified with specified mutations/modifications are given in *# PNPs (families)* and *# PSMs* respectively. Abbreviations (in addition to standard amino acid identifiers): m.X for methylated version of X amino acid, *Hiv, Leuc, NMA, and Thz* are explained in details in Supplementary Table 19.

Offset	Top-3 residues (fractions)			# PNPs	# PSMs
	R1	R2	R3	(families)	
14.0	L/I (20.3%)	V (18.9%)	A (6.6%)	279 (101)	2851
−14.0	L/I (28.3%)	V (11.5%)	m.L/I (7.9%)	230 (94)	3289
18.0	L/I (43.1%)	P (10.3%)	V (7.5%)	137 (72)	1255
28.0	L/I (19.8%)	V (15.8%)	m.L/I (9.6%)	130 (59)	735
−28.0	L/I (21.4%)	V (11.9%)	m.L/I (6.0%)	116 (58)	1124
22.0	L/I (30.4%)	V (11.6%)	P (8.3%)	112 (45)	482
16.0	V (19.4%)	L/I (18.5%)	m.L/I (14.5%)	90 (50)	357
8.0	L/I (28.0%)	V (13.6%)	m.A (12.7%)	89 (44)	785
30.0	L/I (17.6%)	V (10.2%)	m.L/I (9.3%)	88 (49)	230
15.0	V (17.4%)	L/I (13.8%)	m.A (9.2%)	87 (52)	325
−16.0	L/I (13.8%)	m.L/I (11.0%)	V (11.0%)	79 (52)	3435
32.0	L/I (30.0%)	V (20.0%)	P (13.0%)	78 (53)	1028
−15.0	L/I (13.7%)	A (10.8%)	m.L/I (6.9%)	78 (46)	185
−13.0	L/I (20.2%)	P (11.7%)	A (8.5%)	77 (46)	269
36.0	L/I (22.5%)	V (16.9%)	P (9.0%)	71 (37)	166
−6.0	L/I (24.7%)	V (12.4%)	T (8.2%)	71 (38)	186
42.0	L/I (24.7%)	V (12.9%)	P (6.5%)	68 (41)	551
6.0	L/I (21.1%)	T (10.0%)	m.L/I (8.9%)	67 (45)	509
12.0	L/I (16.7%)	V (11.9%)	Hiv (6.0%)	66 (44)	379
−2.0	V (12.0%)	L/I (12.0%)	A (6.0%)	66 (48)	147
5.0	L/I (16.2%)	m.A (11.2%)	V (10.0%)	64 (41)	172
19.0	L/I (23.6%)	V (12.5%)	P (11.1%)	61 (47)	190
−30.0	L/I (11.7%)	A (7.8%)	V (7.8%)	61 (42)	133
−18.0	L/I (12.3%)	m.L/I (9.9%)	F (7.4%)	61 (46)	517
13.0	V (13.2%)	L/I (10.5%)	P (9.2%)	60 (40)	156
34.0	L/I (21.5%)	P (19.0%)	A (7.6%)	59 (41)	285
−42.0	L/I (15.3%)	T (9.7%)	A (6.9%)	57 (36)	386
23.0	L/I (26.9%)	m.A (10.3%)	P (9.0%)	56 (40)	198
29.0	L/I (16.0%)	V (14.7%)	P (8.0%)	55 (41)	139
−12.0	L/I (36.2%)	P (14.5%)	m.L/I (7.2%)	55 (40)	217
−8.0	L/I (17.4%)	P (10.1%)	T (7.2%)	55 (38)	129
44.0	P (10.6%)	L/I (10.6%)	m.A (9.1%)	54 (36)	335
−95.1	L/I (46.0%)	P (20.6%)	V (7.9%)	54 (39)	407
−27.0	L/I (17.1%)	A (14.3%)	P (11.4%)	54 (35)	210
2.0	L/I (20.6%)	P (11.8%)	m.L/I (10.3%)	52 (32)	295
4.0	L/I (30.6%)	m.L/I (12.9%)	F (6.5%)	51 (37)	102
17.0	L/I (23.9%)	m.L/I (9.9%)	P (9.9%)	51 (34)	140
−44.0	L/I (15.5%)	G (8.5%)	P (7.0%)	51 (40)	255

Offset	Top-3 residues (fractions)			# PNPs	# PSMs
	R1	R2	R3	(families)	
−29.0	L/I (14.8%)	E (9.3%)	A (7.4%)	48 (36)	203
9.0	L/I (20.8%)	P (13.2%)	V (9.4%)	47 (33)	77
−81.1	V (33.3%)	L/I (27.8%)	P (13.0%)	47 (31)	247
20.0	L/I (23.8%)	P (14.3%)	V (12.7%)	46 (32)	163
−23.0	P (20.0%)	L/I (12.0%)	T (10.0%)	45 (36)	70
−32.0	L/I (16.7%)	F (13.0%)	Leuc (7.4%)	44 (37)	194
−87.0	F (11.8%)	S (11.8%)	V (11.8%)	43 (36)	88
7.0	L/I (23.5%)	V (9.8%)	F (7.8%)	42 (33)	63
37.0	V (16.7%)	L/I (14.6%)	m.A (12.5%)	42 (30)	132
−34.0	F (14.5%)	L/I (14.5%)	P (7.3%)	42 (36)	636
−26.0	P (19.4%)	L/I (17.7%)	A (6.5%)	42 (31)	305
45.0	L/I (30.6%)	P (10.2%)	m.A (10.2%)	41 (30)	174
−99.1	L/I (31.4%)	V (19.6%)	m.L/I (13.7%)	41 (35)	24638
−56.0	A (15.7%)	P (9.8%)	T (7.8%)	41 (31)	199
−57.0	A (15.2%)	P (10.9%)	G (8.7%)	41 (33)	78
26.0	V (17.8%)	L/I (17.8%)	T (8.9%)	40 (33)	187
−6.1	L/I (47.1%)	m.L/I (13.7%)	V (3.9%)	40 (21)	96
−72.0	F (12.5%)	P (12.5%)	L/I (12.5%)	40 (35)	1180
−55.0	A (10.6%)	W (10.6%)	S (8.5%)	40 (34)	111
−40.0	V (15.6%)	A (11.1%)	L/I (11.1%)	40 (31)	65
−39.0	L/I (18.8%)	A (10.4%)	m.L/I (6.2%)	40 (27)	78
−38.0	L/I (22.6%)	P (16.1%)	G (9.7%)	40 (29)	1917
−36.0	L/I (17.4%)	Y (8.7%)	V (6.5%)	40 (30)	79
10.0	V (17.6%)	L/I (13.7%)	m.L/I (9.8%)	39 (29)	134
43.0	P (16.3%)	L/I (14.0%)	A (11.6%)	39 (32)	69
46.0	L/I (28.0%)	P (14.0%)	V (10.0%)	39 (30)	117
48.0	L/I (15.6%)	P (11.1%)	V (11.1%)	39 (32)	430
50.0	L/I (22.7%)	A (11.4%)	P (9.1%)	39 (28)	95
−5.0	L/I (18.0%)	V (12.0%)	m.L/I (8.0%)	39 (26)	3469
33.0	L/I (23.4%)	P (12.8%)	A (8.5%)	38 (30)	527
64.0	L/I (21.4%)	m.L/I (11.9%)	P (9.5%)	38 (32)	64
−43.0	L/I (15.7%)	A (9.8%)	T (9.8%)	38 (27)	225
−19.0	P (12.2%)	m.L/I (10.2%)	T (8.2%)	37 (24)	157
−11.0	A (11.4%)	V (11.4%)	m.A (9.1%)	37 (31)	78
27.0	L/I (18.2%)	V (11.4%)	A (9.1%)	36 (32)	135
40.0	L/I (19.1%)	m.L/I (10.6%)	T (10.6%)	36 (27)	224
−9.0	L/I (26.7%)	P (15.6%)	m.L/I (8.9%)	36 (28)	91
−113.1	L/I (32.4%)	F (13.5%)	m.L/I (13.5%)	36 (28)	151
−84.0	V (18.2%)	L/I (18.2%)	P (11.4%)	36 (32)	88
−53.0	A (22.0%)	L/I (17.1%)	P (9.8%)	36 (23)	292
−22.0	L/I (13.7%)	P (11.8%)	T (11.8%)	36 (30)	1748
11.0	L/I (22.7%)	V (13.6%)	Q (11.4%)	35 (32)	230
24.0	L/I (18.6%)	P (14.0%)	V (11.6%)	35 (27)	1905
35.0	L/I (21.4%)	P (19.0%)	V (11.9%)	35 (27)	75
57.0	L/I (29.3%)	A (9.8%)	F (7.3%)	35 (30)	322
−109.1	L/I (39.0%)	m.L/I (19.5%)	F (7.3%)	35 (27)	127
131.1	L/I (20.0%)	P (17.5%)	S (7.5%)	35 (27)	211
−41.0	L/I (16.0%)	A (8.0%)	P (8.0%)	35 (27)	468
−46.0	L/I (15.0%)	F (12.5%)	A (12.5%)	35 (29)	82
31.0	L/I (20.5%)	T (10.3%)	m.L/I (7.7%)	34 (25)	94
52.0	P (13.9%)	V (13.9%)	L/I (13.9%)	34 (22)	47
56.0	L/I (18.4%)	V (12.2%)	K (10.2%)	34 (22)	119
14.1	V (46.3%)	L/I (19.5%)	Thz (4.9%)	34 (20)	143
28.1	E (20.0%)	A (12.5%)	L/I (10.0%)	34 (17)	139
−70.0	V (14.0%)	A (11.6%)	T (11.6%)	34 (29)	95
−10.0	P (14.6%)	L/I (14.6%)	T (7.3%)	34 (27)	70

Offset	Top-3 residues (fractions)			# PNPs	# PSMs
	R1	R2	R3	(families)	
−4.0	L/I (18.6%)	F (11.6%)	V (11.6%)	34 (29)	109
66.0	P (24.4%)	m.A (17.1%)	L/I (9.8%)	33 (27)	76
7.9	L/I (27.8%)	m.L/I (13.9%)	V (11.1%)	33 (18)	50
−48.0	L/I (13.5%)	m.F (13.5%)	F (10.8%)	33 (30)	50
−17.0	L/I (14.6%)	G (12.2%)	F (7.3%)	33 (31)	63
59.0	L/I (36.6%)	m.L/I (12.2%)	A (9.8%)	32 (26)	105
70.0	m.L/I (15.8%)	L/I (13.2%)	V (10.5%)	32 (27)	80
−28.1	L/I (24.3%)	V (18.9%)	m.L/I (16.2%)	32 (21)	220
141.1	L/I (15.4%)	F (10.3%)	V (10.3%)	32 (26)	121
113.1	L/I (25.0%)	S (16.7%)	V (12.5%)	32 (22)	792
−60.0	P (11.6%)	L/I (11.6%)	T (11.6%)	32 (27)	90
−20.0	L/I (17.5%)	F (10.0%)	P (10.0%)	32 (24)	55
3.0	L/I (20.0%)	m.L/I (17.1%)	P (8.6%)	31 (22)	53
38.0	L/I (21.1%)	m.L/I (13.2%)	P (10.5%)	31 (23)	97
47.0	L/I (25.0%)	V (20.0%)	P (10.0%)	31 (26)	769
42.1	L/I (25.6%)	E (15.4%)	V (12.8%)	31 (18)	73
−100.1	L/I (38.2%)	m.L/I (14.7%)	F (8.8%)	31 (24)	70
−42.1	L/I (29.4%)	m.L/I (8.8%)	V (8.8%)	31 (21)	59
−24.0	L/I (12.8%)	T (10.3%)	Q (7.7%)	31 (28)	4865
−21.0	NMA (9.5%)	F (7.1%)	P (7.1%)	31 (24)	200
62.0	A (12.2%)	P (12.2%)	m.L/I (9.8%)	30 (26)	169
91.0	L/I (19.4%)	m.L/I (16.1%)	P (16.1%)	30 (25)	61
−111.1	L/I (37.1%)	m.L/I (34.3%)	Y (2.9%)	30 (21)	374
−69.0	A (14.3%)	L/I (14.3%)	m.L/I (11.4%)	30 (24)	63
−67.0	A (14.6%)	L/I (12.2%)	V (9.8%)	30 (27)	71

Supplementary Table 5. Distribution of $SPCScore(P_{known}, S)$ in the brute-force method PSMs. Values in cells are the number of PSMs reported by the brute-force method at the specified significance level and with the stated number of shared peaks (η).

P-value	$\eta \leq 1$	$\eta = 2$	$\eta = 3$	$\eta = 4$	$\eta \geq 5$
$Spectra_{PSEUD}$					
10^{-5}	144	461	1440	2594	12832
10^{-10}	67	157	572	1188	5561
10^{-15}	0	6	20	129	1728
$Spectra_{STREP_1}$					
10^{-5}	196	635	1658	2654	9947
10^{-10}	28	64	195	249	2112
10^{-15}	1	0	7	16	624
$Spectra_{STREP_2}$					
10^{-5}	68	181	361	536	1528
10^{-10}	2	10	25	36	185
10^{-15}	0	0	1	1	51

Supplementary Table 6. Estimation of FDR using various decoy ratios. Decoy ratio D is defined as a fraction of the decoy database size to the target database size. For all datasets FDR is computed for VarQuest identifications with the specified significance level and specified in %. *Time* stands for the total computation time on a 12 CPU Intel Xeon X7560 2.27GHz cluster node. Decoy ratio equals to 1 gives rather good estimate of FDR in comparison to more accurate values at $D = 10$, the absolute difference in the values is within 1-2% in the most cases. At the same time, the overall computation is 1.5 times faster with $D = 1$ than with $D = 10$, so we choose $D = 1$ for our benchmarks.

P-value	$D = 1$	$D = 5$	$D = 10$
$Spectra_{PSEUD}$			
$\leq 10^{-10}$	3.2	3.0	3.0
$\leq 10^{-15}$	1.8	1.5	1.5
Time (hh:mm)	3:17	3:53	4:16
$Spectra_{STREP_1}$			
$\leq 10^{-10}$	6.2	5.7	4.7
$\leq 10^{-15}$	1.9	3.8	3.3
Time (hh:mm)	4:41	5:48	7:31
$Spectra_{STREP_2}$			
$\leq 10^{-10}$	8.5	5.9	6.1
$\leq 10^{-15}$	3.6	3.0	3.7
Time (hh:mm)	1:37	1:55	2:26

Supplementary Table 7. Estimation of FDR using various decoy generation strategies. *Classical* stands for shuffling of amino acids in the fixed PNP structure, *DEREPLICATOR* is for rearrangement of the PNP mass between its nodes, and *VarQuest* is for shuffling of amino acids and displacement of one random edge. *Experiment 1* is VarQuest run on a highly reliable spectral dataset against the full PNP database (expected FDR is 0%). *Experiment 2* is VarQuest run on the same spectral dataset against the reduced PNP database without "correct" peptides and all their known variants (expected FDR is 50%). FDR is given in percentage.

Strategy	*Experiment 1*	*Experiment 2*
Classical	7.5	63.0
DEREPLICATOR	0.0	2.0
VarQuest	0.5	55.0
Expected	0.0	50.0

Supplementary Table 8. Number of compounds with the specified number of generalized peptide bonds in AntiMarin [3], DNP [2], MIBiG [4], StreptomeDB [5] and the combined database, denoted as *Combined*. *Initial* DB size is the number of compounds downloaded from the corresponding database website, *Filtered* DB size is the number of compounds after removing isomorphic molecules and structures with more than one connected component. *Combined* database does not include peptides with less than 4 bonds. *Class* stands for the automatic taxonomy of compounds with ClassyFire [6] software tool: *none* is for unclassified molecules; *peptidic* is for structures reported as Amino acids, peptides, and analogues, Depsipeptides, Hybrid peptides, Peptoid-peptide hybrids, or Polypeptides; *other* is for all other chemical classes (non peptidic). ClassyFire was run on compounds with at least 4 generalized peptide bonds in each database.

Database	DB size		Compounds with # bonds					Class ($\geq$ 4 bonds)		
	Initial	Filtered	≤ 2	3	4	5	≥ 6	none	peptidic	other
AntiMarin	60908	47654	39611	2238	1402	674	3729	424	4121	1260
DNP	254735	185546	167112	8155	4122	1958	4199	1903	3421	4955
MIBiG	963	888	611	55	33	15	174	47	69	106
StreptomeDB	3992	3554	2909	225	94	77	249	102	83	235
Combined	-	10067	-	-	3244	1592	4663	701	5021	4345

Supplementary Table 9. Classification of PNPs and PNP families in the PNPdatabase. *Class* refers to chemical class reported by ClassyFire [6]. *# PNPs (families)* stands for the number of PNPs (PNP families) in the PNPdatabase. *# identified PNPs (families)* is the number of PNPs (PNP families) identified by the standard identification algorithm (Standard) and VarQuest in $Spectra_{GNPS}$ at 5% FDR. *Total* number of families does not sum up because in a few cases PNPs from different chemical classes may form a single PNP family.

Class	# PNPs (families)	# identified PNPs (families)	
		Standard	VarQuest
Amino acids, peptides, etc	2219 (900)	114 (64)	630 (307)
Depsipeptides	1541 (372)	204 (68)	907 (242)
Hybrid peptides	656 (168)	37 (15)	183 (50)
Peptoid-peptide hybrids	37 (5)	18 (1)	34 (4)
Polypeptides	568 (244)	47 (9)	271 (91)
Total	5021 (1582)	420 (143)	2025 (648)

Supplementary Table 10. Distribution of PNP origins in the PNPdatabase. *Collection* refers to originating collection where PNPs were taken from (AntiMarin [3], DNP [2], MIBiG [4], and StreptomeDB [5] or their various combination). *# PNPs (families)* stands for the number of PNPs (PNP families) in the PNPdatabase. *# identified PNPs (families)* stands for the number of PNPs (PNP families) identified by the standard identification algorithm (Standard) and VarQuest in $Spectra_{GNPS}$ at 5% FDR. *Total* number of families does not sum up because in some cases PNPs from different collections may form a single PNP family. All rows are exclusive, e.g. # PNPs in "Antimarin" counts the number of PNPs present in AntiMarin only and not present in 3 other databases, while # PNPs in "Antimarin and MIBiG" counts the number of PNPs present simultaneously in AntiMarin and MIBiG but absent in 2 other databases.

Collection	# PNPs (families)	# identified PNPs (families)	
		Standard	VarQuest
AntiMarin	1518 (662)	134 (48)	647 (268)
MIBiG	69 (49)	1 (1)	17 (14)
DNP	1653 (587)	114 (57)	599 (245)
StreptomeDB	132 (58)	6 (2)	42 (22)
AntiMarin and DNP	1470 (551)	140 (58)	668 (252)
AntiMarin and MIBiG	4 (4)	1 (1)	2 (2)
AntiMarin and StreptomeDB	32 (15)	4 (2)	8 (6)
DNP and MIBiG	6 (5)	0 (0)	2 (2)
DNP and StreptomeDB	6 (5)	4 (3)	4 (3)
MIBiG and StreptomeDB	5 (2)	0 (0)	0 (0)
All except StreptomeDB	32 (29)	5 (4)	15 (13)
All except MIBiG	71 (44)	6 (4)	16 (12)
All except DNP	3 (3)	0 (0)	0 (0)
All except AntiMarin	1 (1)	0 (0)	0 (0)
All four	19 (17)	5 (5)	5 (5)
Total	5021 (1582)	420 (143)	2025 (648)

Supplementary Table 11. Distribution of PNP family sizes in the PNPdatabase (5021 PNPs from 1582 PNP families). *# PNP families* stands for the number of PNP families containing *Size* distinct compounds. *# identified (complete)* is the number of PNP families having at least one member (or all members) identified by the standard identification algorithm (Standard) and VarQuest in $Spectra_{GNPS}$ at 5% FDR

Size	# PNP families	# identified (complete)	
		Standard	VarQuest
1	976	26 (26)	296 (296)
2	225	26 (13)	111 (76)
3	86	16 (2)	54 (27)
4	64	8 (0)	35 (17)
5	46	13 (0)	24 (13)
6—10	97	22 (1)	66 (22)
11—20	54	17 (0)	37 (7)
21—30	12	6 (0)	10 (1)
≥ 31	22	9 (0)	15 (3)

Supplementary Table 12. Distribution of PNP structures in the PNPdatabase. *Structure* refers to structure of chemical compound (linear, cyclic, branch-cyclic, and all others referred as *complex*). *# PNPs (families)* stands for the number of PNPs (PNP families) in the PNPdatabase. *# identified PNPs (families)* is the number of PNPs (PNP families) identified by the standard identification algorithm (Standard) and VarQuest in $Spectra_{GNPS}$ at 5% FDR. Structures for families are calculated based on a family member with median mass.

Structure	# PNPs (families)	# identified PNPs (families)	
		Standard	VarQuest
linear	1265 (450)	90 (25)	333 (109)
cyclic	1257 (387)	209 (70)	903 (305)
branch-cyclic	1250 (279)	100 (42)	605 (170)
complex	1249 (466)	21 (6)	184 (64)

Supplementary Table 13. Distribution of the number of the generalized peptide bonds in PNPs in the PNPdatabase. *# bonds* stands for the number of generalized peptide bonds. *# PNPs (families)* stands for the number of PNPs (PNP families) in the PNPdatabase. *# identified PNPs (families)* stands for the number of PNPs (PNP families) identified by the standard identification algorithm (Standard) and VarQuest in $Spectra_{GNPS}$ at 5% FDR. The number of bonds for families is computed based on a family member with median mass.

# bonds	# PNPs (families)	# identified PNPs (families)	
		Standard	VarQuest
4	790 (282)	6 (6)	40 (21)
5	422 (126)	20 (8)	96 (36)
6	619 (215)	63 (23)	249 (78)
7	678 (213)	39 (21)	328 (120)
8	622 (183)	72 (25)	304 (110)
9	360 (121)	21 (5)	165 (66)
10	336 (78)	50 (13)	168 (47)
11	220 (52)	63 (9)	169 (36)
12	230 (56)	23 (10)	133 (30)
13	108 (39)	6 (4)	51 (25)
14	134 (47)	6 (5)	81 (20)
15	85 (30)	6 (4)	31 (9)
16	53 (11)	9 (4)	32 (8)
17	21 (14)	1 (1)	9 (8)
18	100 (14)	19 (2)	70 (8)
19	65 (26)	4 (2)	24 (10)
≥ 20	178 (75)	12 (1)	75 (16)

Supplementary Table 14. Most frequent residues in the PNPdatabase. Chemical formulas and fractions for top 50 most frequent residues are shown. *AA* is 1-letter amino acid code for chemical formulas matching standard amino acids or their methylated forms (prefixed with "m."). *Mass* is residue monoisotopic mass (in Da). *Fraction* of a residue *per compounds* is the number of PNPs (families) containing this residue at least once divided by the total number of PNPs (families). *Fraction* of a residue *per all amino acids* is the total number of this residue (may be present multiple times in a single PNP) divided by the total number of residues in all PNPs (families). Fractions for families are calculated based on a family member with median mass. Common fragments C_2H_3O, CO, NH, etc. correspond the components of the PNP graph formed by break of N-C bonds of sides chains such as acetyl (C_2H_3O), carboxyl (CO) and nitrogen.

Formula	AA	Mass	Fraction (%)			
			per compounds		per all amino acids	
			PNPs	Families	PNPs	Families
$C_6H_{11}ON$	L/I	113.1	46.15	45.20	9.63	9.49
C_3H_5ON	A	71.0	35.99	32.11	6.82	6.38
C_5H_9ON	V	99.1	33.86	33.75	6.01	6.01
$C_4H_7O_2N$	T	101.0	30.95	28.57	4.75	4.69
C_5H_7ON	P	97.1	27.44	28.57	4.29	4.93
C_2H_3ON	G	57.0	24.16	28.13	4.46	6.30
$C_3H_5O_2N$	S	87.0	22.39	25.73	3.90	4.45
C_9H_9ON	F	147.1	17.51	20.80	2.46	3.05
C_4H_7ON	m.A	85.1	14.84	8.41	5.66	2.34
$C_5H_8O_2N_2$	Q	128.1	12.99	12.52	2.32	1.95
$C_4H_6O_2N_2$	N	114.0	11.79	16.18	1.83	2.60
$C_5H_7O_3N$	E	129.0	11.35	11.44	1.69	1.74
$C_6H_{12}ON_2$	K	128.1	10.93	13.53	1.93	2.62
C_2H_3O	-	43.0	10.64	8.66	1.39	1.28
$C_9H_9O_2N$	Y	163.1	10.56	14.66	1.56	2.25
$C_4H_5O_3N$	D	115.0	10.36	12.77	1.62	1.99
$C_7H_{13}ON$	m.L/I	127.1	8.78	7.40	1.64	1.19
CO	-	28.0	7.87	4.42	1.09	0.55
$C_{11}H_{10}ON_2$	W	186.1	7.83	8.98	1.07	1.24
$C_6H_{12}ON_4$	R	156.1	6.95	9.67	1.03	1.61
C_4HONS	-	111.0	6.19	5.25	1.02	1.02
C_4H_5ON	-	83.0	5.44	5.06	0.77	0.76
$C_{10}H_{11}ON$	m.F	161.1	5.12	4.74	0.68	0.68
C_6H_9ON	m.P	111.1	4.48	2.53	0.65	0.40
HN	-	15.0	4.06	1.26	0.48	0.17
$C_5H_{10}ON_2$	-	114.1	4.06	2.91	0.55	0.39
$C_{10}H_{11}O_2N$	m.Y	177.1	3.96	2.34	0.49	0.28
$C_5H_7O_2N$	-	113.0	3.92	2.40	0.57	0.36
$C_6H_{10}O_2$	-	114.1	3.70	2.21	0.49	0.36
$C_6H_{11}O_2N$	-	129.1	3.70	1.96	0.49	0.26
$C_4H_5O_4N$	-	131.0	3.37	2.72	0.46	0.36
CH_3O	-	31.0	3.25	4.68	0.43	0.59
$C_5H_9O_2N$	m.T	115.1	3.09	1.96	0.38	0.26
$C_5H_8O_2$	-	100.1	2.93	3.73	0.53	0.71
C_5H_9ONS	M	131.0	2.87	6.26	0.39	0.86
$C_4H_8ON_2$	-	100.1	2.85	2.15	0.68	0.41
CH_2ON	-	44.0	2.73	1.33	0.32	0.15
$C_6H_7ON_3$	H	137.1	2.57	4.42	0.33	0.57
C_3HNS	-	83.0	2.51	1.64	0.31	0.25
CHO	-	29.0	2.47	2.15	0.29	0.25
C_2H_5N	-	43.0	2.29	2.15	0.28	0.28
$C_{13}H_{11}O_3N_3$	-	257.1	2.21	1.64	0.26	0.19
$C_5H_9ON_2$	-	113.1	2.21	1.45	0.27	0.19

Formula	AA	Mass	Fraction (%)			
			per compounds		per all amino acids	
			PNPs	Families	PNPs	Families
$C_{11}H_{13}O_2N$	-	191.1	2.15	1.14	0.26	0.14
$C_5H_9O_2N_2$	-	129.1	2.15	2.34	0.25	0.27
C_3H_3ON	-	69.0	2.13	2.28	0.44	0.44
$C_{12}H_{18}O_3N_7$	-	308.1	2.11	0.38	0.24	0.04
$C_6H_{10}O_3N_2$	-	158.1	2.09	1.01	0.43	0.17
$C_7H_{12}O_3N_2$	-	172.1	2.07	1.64	0.32	0.23
$C_6H_{14}ON$	-	116.1	2.01	0.88	0.23	0.10
Total entries			5021	1582	43385	13760

Supplementary Table 15. Peptide variants identified by VarQuest in the search of $Spectra_{PSEUD}$, $Spectra_{STREP_1}$ and $Spectra_{STREP_2}$ against the PNPdatabase. Identifications of known PNPs (zero offset and isotopic shifts) are not shown. *Peptide* refers to a known PNP, *Offset* is identified modification mass and the amino acid corresponding to the most likely location of the modification in a PNP graph. *M* stands for suggested modification or mutation type: "I" for insertion, "D" for deletion, "?" indicates uncertainty (modification mass does not match standard amino acid masses), "!" shows high confidence in predicting a deletion (negative modification mass exactly matches the mass of an amino acid). *Mass* is the total variant PNP monoisotopic mass (in Da). *Sc* stands for SPCScore. *Strain* column indicates a specific Pseudomonas ($Spectra_{PSEUD}$), or Streptomyces ($Spectra_{STREP_1}$ and $Spectra_{STREP_2}$) strain associated with spectra file, "N/A" stands for unknown strand. In *SN* column "+" indicates that the variant was identified by SpecNets and "−" otherwise. The boldfaced and underlined PSMs are analyzed in details (## 44, 139, and 227). PNPs with complex names are abbreviated with "$*^N$" suffix and described in details in Supplementary Table 18. Abbreviations (in addition to standard amino acid identifiers): m.X for methylated version of X amino acid, *lt* for lipid tail, the rest are explained in details in Supplementary Table 19.

#	Peptide	Offset	M	Mass	*P*-value	Sc	Strain	SN
				$Spectra_{PSEUD}$				
PNP family: Orfamides; Producer: Pseudomonas [7]								
1	Orfamide A	$V+14.0$	-	1308.9	1.5×10^{-28}	19	*P. sp. LCBR*	+
2	Orfamide A	$V+17.0$	-	1311.9	6.8×10^{-21}	15	*P. sp. LCBR*	+
3	Orfamide A	$lt+26.0$	-	1320.9	2.5×10^{-19}	16	*P. sp. LCBR*	−
4	Orfamide A	$lt+28.0$	-	1322.9	7.8×10^{-21}	18	*P. sp. LCBR*	+
5	Orfamide A^{*1}	$V+28.0$	-	1308.9	1.1×10^{-17}	14	*P. sp. LCBR*	+
6	Orfamide B	$T-18.0$	-	1280.8	1.1×10^{-23}	9	*P. sp. LCBR*	+
7	Orfamide B	$V-4.0$	-	1294.8	1.0×10^{-27}	11	*P. sp. LCBR*	+
PNP family: Xantholysins and similar; Producer: Pseudomonas [8]								
8	Antib. MA026	$L/I+1.0$	-	1776.1	3.7×10^{-18}	9	*P. fluorescens*	+
9	Antib. MA026	$V+12.1$	-	1787.1	4.3×10^{-18}	14	*P. fluorescens*	−
10	Antib. MA026	$L/I+23.0$	-	1798.1	1.1×10^{-18}	9	*P. fluorescens*	−
11	Xantholysin A^{*1}	$V+12.1$	-	1773.1	9.7×10^{-20}	16	*P. sp. BW18*	+
12	Xantholysin A^{*1}	$L/I+14.1$	-	1775.2	2.1×10^{-21}	12	*P. fluorescens*	+
13	Xantholysin A^{*1}	$V+15.1$	-	1776.1	1.2×10^{-22}	15	*P. fluorescens*	+
14	Xantholysin A^{*2}	$L/I-10.9$	-	1790.2	1.1×10^{-19}	13	*P. sp. BW18*	−
15	Xantholysin A^{*2}	$L/I+5.1$	-	1806.2	1.6×10^{-19}	17	*P. fluorescens*	−
16	Xantholysin A^{*2}	$lt+12.0$	-	1813.1	5.0×10^{-24}	14	*P. fluorescens*	−
17	Xantholysin A	$L/I+8.0$	-	1783.1	1.4×10^{-20}	11	*P. fluorescens*	−
18	Xantholysin A	$L/I+12.0$	-	1787.1	2.5×10^{-25}	15	*P. fluorescens*	−
19	Xantholysin A	$L/I+14.1$	-	1789.2	2.8×10^{-18}	15	*P. fluorescens*	−
20	Xantholysin A	$L/I+15.1$	-	1790.2	4.7×10^{-19}	14	*P. fluorescens*	−
PNP family: Tolaasins; Producer: Pseudomonas [9]								
21	Tolaasin^{*1}	$L/I-3.0$	-	2001.2	3.8×10^{-24}	14	*P. tolaasii*	+
22	Tolaasin B	$T+14.0$	-	1986.2	3.9×10^{-32}	18	*P. tolaasii*	+
23	Tolaasin B	$K+15.0$	-	1987.2	3.4×10^{-25}	17	*P. tolaasii*	+
24	Tolaasin D	$V+15.0$	-	2001.2	2.2×10^{-23}	16	*P. tolaasii*	+
25	Tolaasin D	$V+18.0$	-	2004.2	8.4×10^{-35}	17	*P. tolaasii*	+
26	Tolaasin D	$V+19.0$	-	2005.2	3.6×10^{-26}	16	*P. tolaasii*	+
27	Tolaasin D	$L/I+23.0$	-	2009.2	8.4×10^{-23}	13	*P. tolaasii*	+
28	Tolaasin D	$L/I+28.0$	-	2014.2	2.1×10^{-23}	16	*P. tolaasii*	−

#	Peptide	Offset	M	Mass	P-value	Sc	Strain	SN
29	Tolaasin D	$V+29.0$	-	2015.2	1.2×10^{-22}	15	*P. tolaasii*	+
30	Tolaasin II	$Q+31.0$	-	1973.2	2.8×10^{-27}	15	*P. tolaasii*	+
31	Tolaasin II	$\Delta But+44.0$	-	1986.2	5.4×10^{-29}	13	*P. tolaasii*	+
32	Tolaasin II	$V+73.0$	I?	2015.2	1.3×10^{-17}	13	*P. tolaasii*	+
PNP family: Putisolvins; Producer: Pseudomonas [10]								
33	Putisolvin I*[1]	$S-14.0$	-	1379.8	4.9×10^{-45}	18	*P. putida*	+
34	Putisolvin I*[1]	$S+8.0$	-	1401.8	7.1×10^{-23}	11	*P. putida*	+
35	Putisolvin I*[1]	$S+22.0$	-	1415.8	3.2×10^{-31}	14	*P. putida*	−
36	Putisolvin I	$S+14.0$	-	1393.8	2.6×10^{-35}	17	*P. putida*	+
37	Putisolvin I	$S+17.0$	-	1396.9	3.6×10^{-27}	13	*P. putida*	+
38	Putisolvin I	$S+18.0$	-	1397.8	4.1×10^{-35}	16	*P. putida*	+
39	Putisolvin I	$S+19.0$	-	1398.8	4.8×10^{-21}	9	*P. putida*	−
40	Putisolvin I	$S+22.0$	-	1401.8	7.7×10^{-32}	16	*P. putida*	−
41	Putisolvin I	$S+36.0$	-	1415.8	1.9×10^{-27}	14	*P. putida*	−
42	Putisolvin I	$S+40.0$	-	1419.8	2.7×10^{-21}	15	*P. putida*	−
PNP family: Massetolides; Producer: Pseudomonas [11, 12]								
43	Massetolide*[1]	$lt+141.1$	I?	1308.9	3.0×10^{-18}	14	*P. sp. LCBR*	+
44	**Massetolide A**	$L/I+113.1$	I	1252.8	4.2×10^{-19}	19	*P. synxantha*	+
45	Massetolide E	$V+14.0$	-	1125.7	3.4×10^{-20}	14	*P. putida*	+
46	Massetolide F	$L/I+14.0$	-	1139.7	9.0×10^{-25}	16	*P. fluorescens*	+
47	Massetolide F	$L/I+18.0$	-	1143.7	5.9×10^{-22}	9	*P. sp. PGSB*	−
48	Massetolide F	$lt+26.0$	-	1151.7	7.2×10^{-19}	16	*P. fluorescens*	+
49	Massetolide G	$lt+14.0$	-	1153.7	1.2×10^{-19}	18	*P. fluorescens*	+
50	Pseudodesmin B	$V+15.0$	-	1125.7	1.0×10^{-17}	12	*P. fluorescens*	+
51	Massetolide I	$L/I+28.0$	-	1139.7	5.2×10^{-19}	14	*P. fluorescens*	+
Singletons; Producers: Pseudomonas [13]								
52	Amphisin	$D+14.0$	-	1408.8	5.3×10^{-19}	18	*P. fluorescens*	+
PNP family: Surfactins and similar; Producer: Bacillus [14–16]								
53	Pumilacidin F	$L/I-14.0$	-	1035.7	4.0×10^{-19}	14	*P. fluorescens*	+
54	Esperin	$E-14.0$	-	1021.7	2.7×10^{-19}	18	*P. reactans*	+
55	Esperin	$V+14.0$	-	1049.7	4.7×10^{-20}	16	*P. rhodesiae*	+
56	Lipopeptide NO	$E+42.0$	-	1035.7	1.1×10^{-19}	19	*P. reactans*	+
57	Surfactin*[1]	$CH_2-14.0$	-	1035.7	7.3×10^{-23}	20	*P. rhodesiae*	+
58	Surfactin*[2]	$L/I+18.0$	-	1053.7	1.3×10^{-19}	9	*P. reactans*	−
59	Surfactin B1	$L/I+18.0$	-	1039.7	1.4×10^{-27}	11	*P. reactans*	−
60	Surfactin C1	$L/I+18.0$	-	1053.7	5.9×10^{-26}	16	*P. reactans*	−
61	Surfactin*[3]	$CH_2-14.0$	-	1035.7	1.9×10^{-25}	13	*P. reactans*	+
62	Surfactin*[4]	$CH_2-14.0$	-	1021.7	8.5×10^{-19}	14	*P. reactans*	+
63	Surfactin*[5]	$CH_2-14.0$	-	1021.7	7.9×10^{-27}	18	*P. reactans*	+
64	Surfactin*[6]	$CH_2-14.0$	-	1035.7	8.4×10^{-20}	11	*P. tolaasii*	+
PNP family: Xentrivalpeptides; Producer: Xenorhabdus [17]								
65	Xentrivalpep. Q	$T-90.0$	D?	670.4	1.9×10^{-23}	11	*P. fluorescens*	−
66	Xentrivalpep. A	$T-90.0$	D?	769.5	5.4×10^{-20}	15	*P. tolaasii*	−
67	Xentrivalpep. B	$V+241.1$	I?	1024.6	3.9×10^{-19}	16	*P. sp. C52*	+
68	Xentrivalpep. F	$V+171.0$	I?	1024.6	1.6×10^{-24}	16	*P. putida*	+
69	Xentrivalpep. G	$T-70.0$	D	769.5	2.9×10^{-21}	14	*P. fluorescens*	−
70	Xentrivalpep. G	$V+185.1$	I?	1024.6	8.9×10^{-20}	14	N/A	+
71	Xentrivalpep. K	$Iva-22.0$	-	769.5	4.6×10^{-18}	12	*P. fluorescens*	−
PNP family: Bacillomycins and similar; Producer: Bacillus [18, 19]								
72	Bacillomycin F1	$N-27.0$	-	1044.6	1.9×10^{-20}	17	*P. reactans*	+
73	Bacillomycin F6	$N-27.0$	-	1072.6	2.1×10^{-23}	14	*P. rhodesiae*	+
74	Bacillopeptin B	$S+10.0$	-	1044.6	1.3×10^{-20}	11	*P. rhodesiae*	+
75	Bacillomycin*[1]	$E-14.0$	-	1044.5	3.9×10^{-22}	13	*P. rhodesiae*	+
PNP family: SNA-60-367; Producer: Bacillus [20]								
76	SNA-60-367*[1]	$E+30.0$	-	1476.8	8.9×10^{-31}	22	*P. rhodesiae*	+
77	SNA-60-367*[2]	$E+2.0$	-	1476.8	1.1×10^{-17}	15	*P. reactans*	+
78	SNA-60-367*[3]	$E+2.0$	-	1490.8	9.3×10^{-28}	18	*P. rhodesiae*	+
Singletons; Producers: NOT Pseudomonas, mostly marine sponges [21–36]								
79	Antib. BK230	$m.Y-118.0$	D?	1007.7	2.3×10^{-19}	19	*P. rhodesiae*	+
80	Axinastatin 4	$L/I+6.0$	-	812.5	8.2×10^{-22}	10	*P. fluorescens*	−
81	$\alpha Subst.-I_B$	$R-156.1$	D!	529.3	3.0×10^{-19}	8	*P. fluorescens*	+
82	Callyaerin H	$Dia+80.0$	I?	1123.6	6.2×10^{-21}	16	*P. fluorescens*	+
83	Euryjanicin B	$P+48.0$	-	757.4	3.7×10^{-19}	14	*P. aeruginosa*	−
84	Gramicidin S2	$Orn-102.1$	D	1024.6	1.1×10^{-18}	19	*P. sp. Z*	+

#	Peptide	Offset	M	Mass	P-value	Sc	Strain	SN
85	Hymenamide F	$R-138.1$	D?	626.3	5.8×10^{-18}	12	*P. fluorescens*	−
86	Hymenamide F	$R-106.1$	D?	658.3	2.6×10^{-20}	12	*P. fluorescens*	−
87	Hymenamide F	$R-7.0$	-	757.4	8.5×10^{-25}	14	*P. ici*	−
88	Hymenamide F	$R+73.0$	I?	837.4	4.6×10^{-22}	15	*P. putida*	−
89	Hymenamide H	$P-91.1$	D?	812.4	1.9×10^{-21}	13	*P. sp. F*	−
90	Kulokainalide 1	$Lac+233.1$	I?	1123.6	1.1×10^{-18}	14	N/A	+
91	Lobocyclamide C	$T+11.0$	-	1366.8	1.1×10^{-19}	18	*P. putida*	+
92	Malaysiatin	$V-85.1$	D?	667.4	4.6×10^{-18}	9	*P. fluorescens*	−
93	Omphalotin F	$m.L/I-103.0$	D	1294.8	5.1×10^{-18}	19	*P. fluorescens*	+
94	Phakellistatin 11	$F+51.1$	-	1024.6	2.7×10^{-18}	16	*P. putida*	+
95	Rolloamide A	$L/I-112.1$	D?	651.4	2.0×10^{-19}	14	*P. sp. F*	+
96	Stylissamide B	$P-22.0$	-	789.4	1.9×10^{-18}	11	*P. putida*	−
97	V-L-P-V-P	$V-99.1$	D!	552.3	1.2×10^{-20}	8	*P. rhodesiae*	−
98	Verrucamide B	$m.F-7.0$	-	1393.8	1.1×10^{-18}	20	*P. putida*	+
99	Wainunuamide	$H-119.1$	D?	626.3	2.7×10^{-18}	10	*P. fluorescens*	−
100	Wainunuamide	$H+92.0$	I?	837.4	2.8×10^{-18}	11	*P. resinovor.*	−
				$Spectra_{STREP_1}$				
PNP family: Surugamides and similar; Producer: Streptomyces [37–39]								
101	Champacyclin	$F-113.1$	D	784.5	2.8×10^{-21}	20	*S. albus*	−
102	Champacyclin	$L/I-99.9$	D	797.7	3.8×10^{-19}	12	*S. albus*	−
103	Champacyclin	$K-99.1$	D	798.5	1.7×10^{-18}	15	*S. albus*	−
104	Champacyclin	$F-95.1$	D?	802.5	1.9×10^{-20}	20	*S. albus*	−
105	Champacyclin	$V-81.1$	D?	816.6	2.7×10^{-20}	17	*S. albus*	−
106	Champacyclin	$V-71.1$	D	826.5	1.5×10^{-19}	17	*S. albus*	−
107	Champacyclin	$A-28.0$	-	869.6	2.1×10^{-21}	19	*S. albus*	−
108	Champacyclin	$L/I-14.1$	-	883.5	1.4×10^{-24}	16	*S. albus*	−
109	Champacyclin	$L/I-12.0$	-	885.6	3.9×10^{-27}	15	*S. albus*	−
110	Champacyclin	$F+4.0$	-	901.6	5.5×10^{-20}	16	*S. albus*	−
111	Champacyclin	$L/I+14.0$	-	911.6	3.7×10^{-27}	25	*S. albus*	+
112	Champacyclin	$V+15.1$	-	912.7	4.9×10^{-19}	15	*S. albus*	−
113	Champacyclin	$V+30.0$	-	927.6	8.3×10^{-24}	16	*S. albus*	+
114	Champacyclin	$V+36.0$	-	933.6	4.2×10^{-23}	21	*S. albus*	−
115	Champacyclin	$K+78.0$	I?	975.6	1.2×10^{-20}	15	*S. albus*	−
116	Champacyclin	$K+128.0$	I	1025.6	1.1×10^{-18}	14	*S. albus*	−
117	Surugamide A	$L/I-113.1$	D!	798.5	4.3×10^{-21}	18	*S. albus*	+
118	Surugamide A	$L/I-95.1$	D?	816.5	2.0×10^{-21}	19	*S. albus*	−
119	Surugamide A	$L/I-95.1$	D?	816.6	2.6×10^{-22}	14	*S. albus*	−
120	Surugamide A	$L/I-14.0$	-	897.6	6.4×10^{-33}	24	N/A	+
121	Surugamide A	$L/I+14.0$	-	925.6	9.8×10^{-21}	23	*S. albus*	−
122	Surugamide A	$L/I+16.0$	-	927.6	1.2×10^{-26}	27	*S. albus*	−
123	Surugamide A	$L/I+18.0$	-	929.6	4.9×10^{-22}	20	*S. albus*	−
124	Surugamide A	$L/I+22.0$	-	933.6	8.7×10^{-25}	21	*S. albus*	−
125	Surugamide A	$L/I+28.0$	-	939.6	8.3×10^{-26}	18	*S. albus*	−
126	Surugamide A	$L/I+34.0$	-	945.6	1.9×10^{-20}	17	*S. albus*	−
127	Surugamide A	$K+42.0$	-	953.6	4.6×10^{-23}	19	*S. albus*	−
128	Surugamide A	$K+56.0$	I	967.7	1.7×10^{-20}	14	*S. albus*	−
129	Surugamide A	$K+72.0$	I	983.6	1.2×10^{-20}	17	*S. albus*	−
130	Surugamide A	$K+128.0$	I	1039.6	5.0×10^{-24}	21	*S. albus*	−
131	Surugamide A	$F+131.1$	I	1042.7	4.7×10^{-20}	20	*S. albus*	−
132	Surugamide A	$L/I+148.0$	I	1059.6	5.4×10^{-25}	20	*S. albus*	−
133	Surugamide A	$K+157.1$	I	1068.7	2.2×10^{-22}	21	*S. albus*	−
134	Surugamide A	$K+163.0$	I	1074.7	2.3×10^{-19}	14	*S. albus*	−
135	Surugamide A	$K+171.0$	I?	1082.7	1.7×10^{-22}	18	*S. albus*	−
136	Surugamide A	$L/I+186.0$	I	1097.7	2.5×10^{-20}	16	*S. albus*	−
137	Surugamide A	$L/I+197.1$	I?	1108.7	4.0×10^{-22}	16	*S. albus*	−
138	Surugamide A	$L/I+251.1$	I?	1162.7	2.1×10^{-20}	15	*S. albus*	−
139	**Surugamide B**	$K-128.1$	D!	769.5	1.7×10^{-19}	20	*S. albus*	−
140	Surugamide B	$K-85.0$	D?	812.6	4.2×10^{-25}	21	*S. albus*	−
141	Surugamide B	$F-34.0$	-	863.6	9.7×10^{-21}	24	*S. albus*	+
142	Surugamide B	$L/I-28.0$	-	869.6	5.3×10^{-21}	19	*S. albus*	−
143	Surugamide B	$L/I-14.0$	-	883.6	5.0×10^{-26}	22	*S. albus*	+
144	Surugamide B	$L/I-12.0$	-	885.6	1.4×10^{-21}	16	*S. albus*	−
145	Surugamide B	$K+14.0$	-	911.6	4.8×10^{-23}	21	*S. albus*	+
146	Surugamide B	$L/I+16.0$	-	913.6	2.6×10^{-24}	19	*S. albus*	+

#	Peptide	Offset	M	Mass	P-value	Sc	Strain	SN
147	Surugamide B	$F+18.0$	-	915.6	4.5×10^{-23}	15	*S. albus*	+
148	Surugamide B	$K+22.0$	-	919.6	4.0×10^{-21}	22	*S. albus*	−
149	Surugamide B	$K+28.0$	-	925.6	2.5×10^{-25}	26	*S. albus*	+
150	Surugamide B	$L/I+34.0$	-	931.6	2.8×10^{-21}	17	*S. albus*	−
151	Surugamide B	$K+42.0$	-	939.6	3.1×10^{-20}	15	*S. albus*	−
152	Surugamide B	$K+58.0$	I	955.6	2.8×10^{-19}	17	*S. albus*	−
153	Surugamide B	$K+70.0$	I?	967.6	3.0×10^{-27}	27	*S. albus*	−
154	Surugamide B	$K+72.0$	I	969.6	2.7×10^{-22}	19	*S. albus*	−
155	Surugamide B	$K+100.0$	I	997.6	2.8×10^{-21}	14	*S. albus*	−
156	Surugamide B	$K+148.0$	I	1045.6	2.0×10^{-25}	17	*S. albus*	−
157	Surugamide B	$K+169.1$	I?	1066.7	6.5×10^{-21}	23	*S. albus*	−
158	Surugamide B	$K+177.0$	I?	1074.6	3.9×10^{-21}	27	*S. albus*	−
159	Surugamide C	$K-99.1$	D	798.5	1.6×10^{-19}	18	*S. albus*	+
160	Surugamide C	$V-81.1$	D?	816.5	7.3×10^{-22}	19	*S. albus*	−
161	Surugamide C	$L/I-42.0$	-	855.6	5.9×10^{-23}	16	*S. albus*	−
162	Surugamide C	$L/I-28.0$	-	869.6	6.7×10^{-27}	31	*S. albus*	−
163	Surugamide C	$L/I-14.0$	-	883.6	1.8×10^{-27}	23	*S. albus*	+
164	Surugamide C	$L/I+16.0$	-	913.6	4.3×10^{-19}	13	*S. albus*	+
165	Surugamide C	$K+28.0$	-	925.6	3.5×10^{-21}	21	*S. albus*	+
166	Surugamide C	$K+42.0$	-	939.6	1.8×10^{-19}	17	*S. albus*	−
167	Surugamide C	$L/I+72.0$	I	969.6	2.1×10^{-19}	13	*S. albus*	−
168	Surugamide C	$L/I+146.1$	I?	1043.7	9.2×10^{-19}	13	*S. albus*	−
169	Surugamide C	$K+180.0$	I?	1077.6	5.2×10^{-21}	14	*S. albus*	−
170	Surugamide D	$F-113.1$	D	784.5	7.2×10^{-19}	15	*S. albus*	−
171	Surugamide D	$L/I-110.1$	D?	787.5	4.8×10^{-20}	14	*S. albus*	−
172	Surugamide D	$L/I-99.1$	D	798.5	1.3×10^{-22}	22	*S. albus*	−
173	Surugamide D	$F-95.1$	D?	802.5	3.9×10^{-22}	22	*S. albus*	−
174	Surugamide D	$L/I-94.1$	D?	803.5	1.9×10^{-23}	21	*S. albus*	−
175	Surugamide D	$F-34.0$	-	863.6	1.6×10^{-19}	23	*S. albus*	+
176	Surugamide D	$L/I-28.0$	-	869.6	5.4×10^{-26}	24	*S. albus*	−
177	Surugamide D	$L/I-14.0$	-	883.6	7.6×10^{-27}	26	*S. albus*	+
178	Surugamide D	$L/I+8.0$	-	905.6	2.5×10^{-19}	19	*S. albus*	−
179	Surugamide D	$K+14.0$	-	911.6	9.2×10^{-27}	31	*S. albus*	+
180	Surugamide D	$F+16.0$	-	913.6	1.5×10^{-23}	18	*S. albus*	+
181	Surugamide D	$K+18.0$	-	915.6	3.3×10^{-22}	18	*S. albus*	+
182	Surugamide D	$K+22.0$	-	919.6	5.6×10^{-22}	24	*S. albus*	−
183	Surugamide D	$L/I+28.0$	-	925.6	5.9×10^{-21}	17	*S. albus*	−
184	Surugamide D	$L/I+34.0$	-	931.6	6.1×10^{-20}	13	*S. albus*	−
185	Surugamide D	$K+70.0$	I?	967.6	4.7×10^{-19}	16	*S. albus*	−
186	Surugamide D	$K+100.0$	I	997.6	1.5×10^{-21}	14	*S. albus*	−
187	Surugamide D	$L/I+116.1$	I?	1013.7	2.8×10^{-27}	15	*S. albus*	−
188	Surugamide D	$L/I+148.0$	I	1045.6	3.2×10^{-21}	16	*S. albus*	−
189	Surugamide D	$K+149.0$	I?	1046.6	2.1×10^{-25}	30	*S. albus*	−
PNP family: Phepropeptins and similar; Producers: Streptomyces [40], Eurypon [33]								
190	Phepropeptin A	$L/I-95.1$	D?	587.4	1.5×10^{-19}	16	*S. albus*	−
191	Rolloamide A	$F-113.0$	D	650.4	1.4×10^{-18}	12	*S. e14*	−
Singletons; Producers: Streptomyces [41–43]								
192	Desotamide	$G+32.0$	-	542.3	1.0×10^{-19}	12	N/A	−
193	Stenothricin	$A+18.0$	-	1149.6	8.4×10^{-23}	16	*S. roseosporus*	−
194	Virginiamycin S1	$m.F-78.0$	D?	745.3	4.5×10^{-19}	11	*S. pristinaesp.*	−
PNP family: Isariins; Producer: fungus Isaria [44, 45]								
195	Isariin A	$V+179.1$	I?	816.6	8.3×10^{-20}	16	*S. albus*	−
196	Isoisariin B	$G+207.1$	I?	802.5	1.9×10^{-18}	15	*S. albus*	−
Singletons; Producers: NOT Streptomyces [46–49]								
197	Aurilide B	$lt+54.0$	-	887.6	1.1×10^{-18}	12	*S. albus*	−
198	Barangamide B	$L/I+193.0$	I?	1270.7	5.8×10^{-24}	26	*S. sviceus*	−
199	KMM1364C	$L/I-14.0$	-	1035.7	2.1×10^{-18}	11	*S. lividans*	+
200	Unguisin D	$W-101.0$	D	637.5	1.6×10^{-21}	20	*S. albus*	+
$Spectra_{STREP_2}$								
PNP family: Surugamides and similar; Producer: Streptomyces [37–39]								
201	Champacyclin	$V+14.0$	-	911.6	1.2×10^{-19}	20	*S. CNQ329*	+
202	Champacyclin	$F+139.1$	I?	1036.7	2.5×10^{-12}	13	*S. hygroscop.*	−
203	Surugamide A	$L/I-14.0$	-	897.6	2.5×10^{-16}	15	*S. CNQ329*	+

#	Peptide	Offset	M	Mass	P-value	Sc	Strain	SN
204	Surugamide A	$F+18.0$	-	929.6	3.4×10^{-13}	13	*S. CNQ329*	+
205	Surugamide A	$L/I+20.0$	-	929.6	4.9×10^{-13}	11	*S. CNQ329*	+
206	Surugamide A	$K+22.0$	-	933.6	7.8×10^{-16}	14	*S. albus*	−
207	Surugamide A	$K+42.0$	-	953.6	4.3×10^{-19}	14	*S. CNQ329*	−
208	Surugamide B	$A-14.0$	-	883.6	1.9×10^{-16}	14	*S. albus*	−
209	Surugamide B	$L/I+14.0$	-	911.6	1.1×10^{-19}	14	*S. CNT302*	+
210	Surugamide B	$F+18.0$	-	915.6	2.2×10^{-12}	11	*S. albus*	−
211	Surugamide B	$L/I+36.0$	-	933.6	3.6×10^{-15}	14	*S. CNQ329*	−
212	Surugamide B	$K+56.0$	I	953.6	3.9×10^{-15}	13	*S. albus*	−
213	Surugamide C	$L/I+14.0$	-	911.6	9.3×10^{-21}	16	*S. CNY228*	+
214	Surugamide C	$L/I+56.0$	-	953.6	5.6×10^{-16}	18	*S. CNQ329*	−
215	Surugamide D	$L/I+14.0$	-	911.6	1.3×10^{-16}	16	*S. albus*	+
PNP family: Calcium-Dependent Antibiotics (CDA); Producer: Streptomyces [50]								
216	CDA*[1]	$G+14.0$	-	1494.5	6.2×10^{-16}	18	*S. coelicolor*	+
217	CDA*[1]	$ns1+36.0$	-	1516.5	2.3×10^{-12}	13	*S. coelicolor*	−
218	CDA 4a	$T+42.0$	-	1494.5	1.8×10^{-14}	16	*S. coelicolor*	+
Singletons; Producers: Streptomyces [51–57]								
219	Actinomycin G3	$Achr-168.0$	D?	1104.6	2.5×10^{-12}	12	*S. TAA040*	−
220	Glycinocin D	$P+44.0$	-	1276.6	1.0×10^{-11}	14	*S. CNX435*	−
221	Grisemycin	$A+16.0$	-	1848.0	3.2×10^{-30}	17	*S. griseus*	−
222	Grisemycin	$Q+22.0$	-	1854.0	5.2×10^{-29}	19	*S. griseus*	−
223	Nocardamine	$Suc-40.0$	-	544.4	7.2×10^{-13}	8	*S. griseoflavus*	−
224	Nocardamine	$Suc+18.0$	-	602.4	2.3×10^{-12}	8	*S. CNT318*	−
225	Streptofactin	$K+126.1$	I?	953.6	5.0×10^{-17}	10	*S. albus*	+
226	Venepeptide	$L/I+15.0$	-	2138.1	3.3×10^{-13}	10	*S. lividans*	−
227	**Venepeptide**	$N+31.0$	-	2154.1	3.2×10^{-15}	13	*S. lividans*	−
228	WS9326C	$ns2+14.0$	-	1036.5	1.8×10^{-12}	9	*S. griseoflavus*	+
PNP family: Surfactins and similar; Producer: Bacillus [14–16]								
229	Esperin	$E-14.0$	-	1021.7	3.2×10^{-13}	8	N/A	+
230	Surfactin*[1]	$L/I+30.0$	-	1065.7	2.6×10^{-12}	13	*S. CNS580*	−
231	Surfactin D	$V+14.0$	-	1063.7	8.0×10^{-12}	9	N/A	+
232	Surfactin*[6]	$L/I+14.0$	-	1063.7	6.0×10^{-14}	10	N/A	+
Singletons; Producers: NOT Streptomyces [21, 58–66]								
233	Clavariopsin B	$G-7.0$	-	1132.7	1.3×10^{-12}	13	*S. CNQ865*	−
234	Keenamide A	$Thz-62.0$	D?	556.3	8.5×10^{-12}	8	*S. coelicolor*	−
235	Lariatin A	$S+88.1$	I?	2138.1	2.1×10^{-12}	12	*S. coelicolor*	−
236	Myxochromide A1	$Q-65.1$	D?	754.4	8.2×10^{-12}	7	*S. CNS654*	−
237	Nostopeptolide A*[1]	$But+12.0$	-	1092.6	8.9×10^{-14}	12	*S. CNQ329*	−
238	Petriellin A	$A-10.0$	-	1420.9	2.3×10^{-13}	17	*S. CNQ865*	−
239	Pleofungin C	$m.L/I-28.0$	-	1052.7	9.7×10^{-12}	13	*S. hygroscop.*	−
240	Pyoverdin Pf1547	$A-30.0$	-	1517.7	4.5×10^{-14}	10	*S. griseus*	−
241	Sch 378167*[1]	$Leuc-114.0$	D!	1021.6	8.4×10^{-12}	14	*S. CNS580*	−
242	Sch 378167*[1]	$F-50.0$	-	1085.7	9.6×10^{-13}	14	*S. CNS580*	−
243	Taxillaid	$L/I+146.1$	I	953.6	1.0×10^{-14}	12	*S. albus*	+

Supplementary Table 16. Iterative VarQuest run on the $Spectra_{CYANO}$ dataset. *First iteration* is the run of VarQuest on the entire $Spectra_{CYANO}$ dataset against $PNPdatabase$ (5021 PNPs). *Second iteration* is the run of VarQuest on the $Spectra_{CYANO}$ dataset without already identified spectra against the most reliable PNP variants (target and decoy) identified on the first iteration ($FirstIterationDB$ with 81 PNP variants). All PSMs with P-values above 10^{-10} were removed beforehand. *# PSMs* and *# variants* stand for the number of identified PSMs and PNP variants, respectively. *# PNPs (best)* stands for the number of unique PNP identifications in target and decoy databases independently (together, i.e. if a PNP was identified in both target and decoy databases, only the best identification is counted). *Target* and *Decoy* stand for identifications in the target PNP database ($PNPdatabase$ or $FirstIterationDB$) and its decoy version, respectively. *Decoy'* is for the second iteration identifications in $FirstIterationDB$ matching with PNPs originated from decoy PNP variants of the first iteration. *FDR* stands for the False Discovery Rate, the ratio of the total number of decoy hits to the number of target hits.

	First iteration			Second iteration		
	# PSMs	# variants	# PNPs (best)	# PSMs	# variants	# PNPs (best)
Target	3573	2083	702 (688)	626	353	41 (41)
Decoy	101	95	84 (40)	5	5	5 (0)
Decoy'	-	-	-	2	2	1 (1)
FDR (%)	2.8	4.6	12.0 (5.8)	1.1	2.0	14.6 (2.4)

Supplementary Table 17. List of common mass offsets identified by VarQuest in $Spectra_{CYANO}$ at the 2nd iterative run. Only offsets identified in at least 5 PNP variants are shown. The total offset range $-300 \ldots 300$ Da was divided into non-overlapping intervals of size 0.1 Da. All PNP variants (385) were grouped according to the second iteration offset. The groups were sorted by the size. *Offset* is the mass difference between a PNP from $FirstIterationDB$ and its identified variant (in Da). *Possible explanation* is a name of common modification/mutation matching the mass offset (if known). *# PNP variants* stands for the number of PNP variants identified with specified mass offset.

Offset	Possible explanation	# PNP variants
±14.0	methylation	17
±32.0	dioxidation	10
±18.0	hydration	10
±38.0	acrolein [67]	6
±6.0	—	5
±16.0	hydroxylation	5
±26.0	—	5
±28.0	dimethylation	5
±40.0	—	5
±44.0	—	5

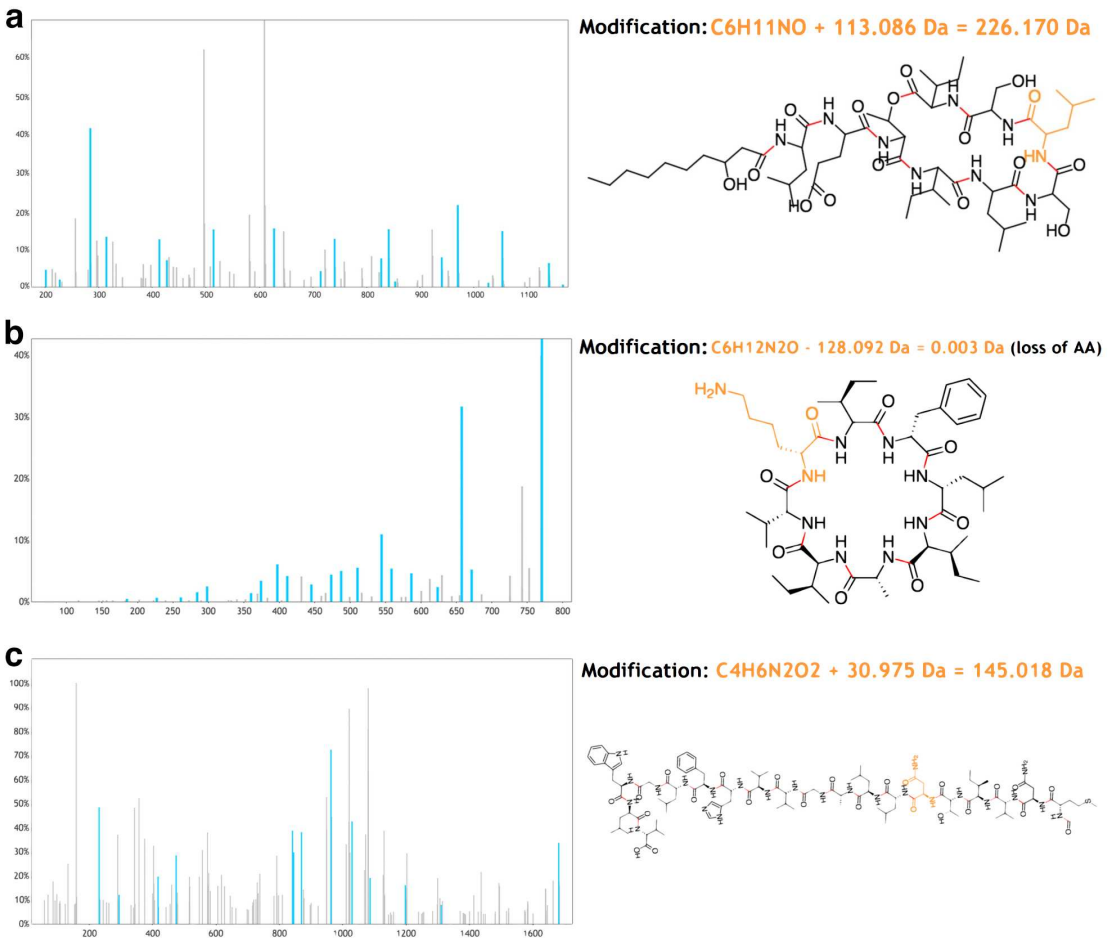
Supplementary Table 18. Abbreviations of PNPs with complex names in the PNPdatabase. *Abbreviation* is a short name used in Supplementary Table 15. *Full name* stands for the full name present in the PNPdatabase. *Mass* is PNP monoisotopic mass (in Da). *Origin* is original database of the compound (AntiMarin [3], DNP [2], MIBiG [4], or StreptomeDB [5]).

Abbreviation	Full name	Mass	Origin
Bacillomycin*1	Bacillomycin D methyl ester	1058.6	AntiMarin
CDA*1	Antibiotic CDA 2.1,3.1-Didehydro(Z-)	1480.5	DNP
Massetolide*1	Massetolides Diastereoisomer	1167.7	DNP
Nostopeptolide A*1	Nostopeptolide A 4-Epimer	1080.6	DNP
Orfamide A*1	Orfamide A 7-Valine analogue	1280.8	DNP
Putisolvin I*1	Putisolvin I 2-Isoleucine or Leucine analogue	1393.8	DNP
SNA 60-367*1	SNA 60-367 2-Deoxy 1	1446.8	DNP
SNA 60-367*2	SNA 60-367 2-Deoxy 2	1474.8	DNP
SNA 60-367*3	SNA 60-367 2-Deoxy 3	1488.8	DNP
Sch 378167*1	Sch 378167 5,5-Diamide	1135.7	DNP
Surfactin*1	Surfactin 1-Me ester 1	1035.7	DNP
Surfactin*2	Surfactin 1-Me ester 2	1049.7	DNP
Surfactin*3	(Ile2,Ile7)-Surfactin C15 monomethyl est	1049.7	AntiMarin
Surfactin*4	(Ile7)-Surfactin C13ai dimethyl ester	1035.7	AntiMarin
Surfactin*5	(Ile7)-Surfactin C14 monomethyl ester	1035.7	AntiMarin
Surfactin*6	(Ile7)-Surfactin C14i dimethyl ester	1049.7	AntiMarin
Tolaasin*1	Tolaasin Ring-opened form	2004.2	DNP
Xantholysin A*1	Xantholysin A 14-Valine analogue	1761.1	DNP
Xantholysin A*2	Xantholysin A N1-Deacyl, N1-(3-hydroxy-5-dodecenoyl)	1801.1	DNP

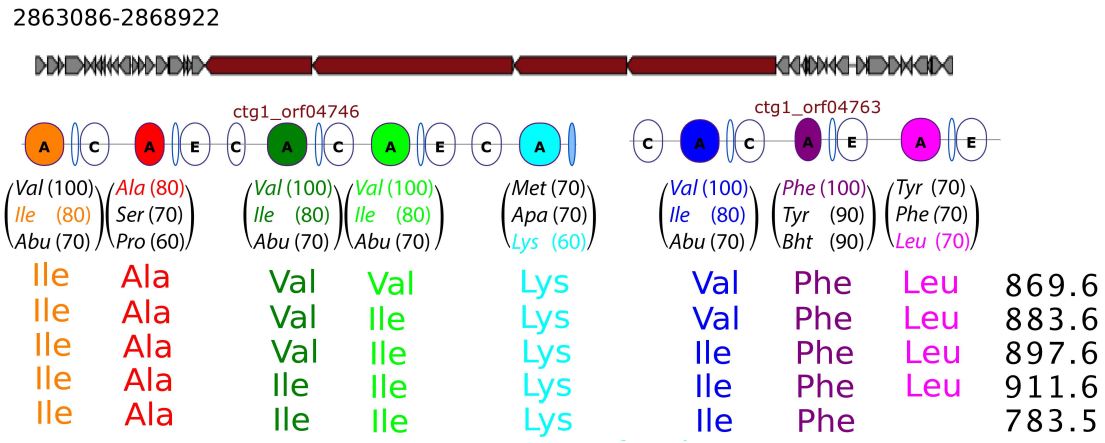
Supplementary Table 19. Abbreviations of non-standard amino acids (residues) in PNPs in the PNPdatabase. *Abbreviation* is a short name used in Supplementary Table 4 and S15. *Full name* stands for the proposed full name of the residue (note: the same chemical formula can represent different residues but mass-spectrometry and VarQuest are blind to residue stereochemistry). *Formula* is for chemical formula. *Mass* is monoisotopic mass (in Da).

Abbreviation	Full name	Formula	Mass
Achr	Actinoyl chromophore	$C_{16}H_{10}O_5N_2$	310.1
But	Butyrate	C_4H_6O	86.0
Cap	Capreomycidine	$C_6H_{10}ON_4$	154.1
Dia	Diaminoacrylic acid	$C_8H_{12}O_3N_2$	184.1
Hiv	Hydroxyisovaleric acid	$C_5H_8O_2$	100.1
Iva	Isovalerate	C_5H_8O	84.1
Lac	Lactic Acid	$C_3H_6O_3$	90.0
Leuc	Leucic acid	$C_6H_{10}O_2$	114.1
NMA	N-methylol acrylamide	C_4H_5ON	131.0
Orn	Ornithine	$C_5H_{10}ON_2$	114.1
Suc	Succinyl amide	$C_9H_{16}O_3N_2$	200.1
Thz	Thiazole-4-carboxylic acid	C_4HONS	111.0
Δ But	Dehydroaminobutyric acid	C_4H_5ON	83.0
ns1	non-standard 1	$C_8H_7O_2N$	149.0
ns2	non-standard 2	$C_{14}H_{14}O$	198.1

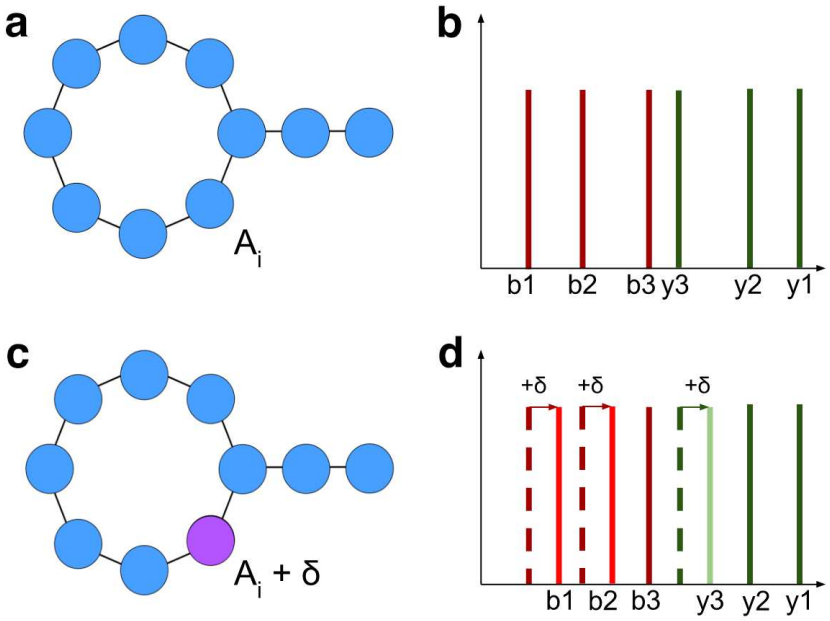
Supplementary Figures



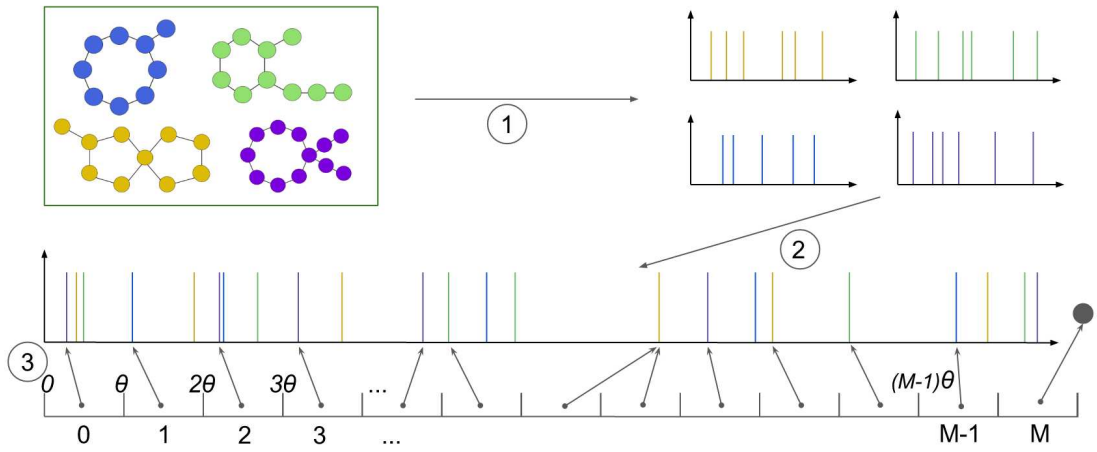
Supplementary Figure 1. Fragments of VarQuest interactive visualization of PSMs for Massetolide-1252 (a), Surugamide-769 (b), and Venepeptide-2154 (c). Tandem mass spectra are shown on the left (matched peaks are highlighted) and known PNP structures are shown on the right, peptide bonds are colored red. Predicted modifications are colored orange and information on the mass offsets and the modified amino acid are given above the chemical structures. User may select any matched peak to see the corresponding fragmentation of the chemical structure. Interactive visualizations are available at http://cab.spbu.ru/software/varquest.



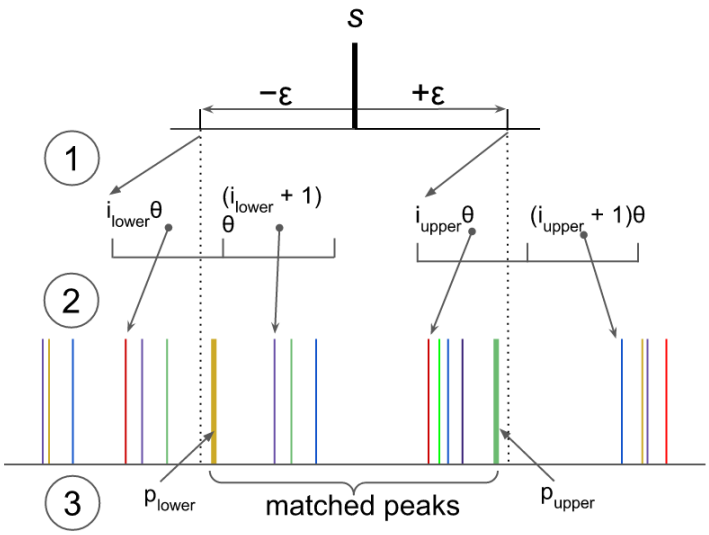
Supplementary Figure 2. The proposed surugamide biosynthetic gene cluster and various surugamides (along with their masses in Da) explained by it. The first five PNPs were reported in [39], the last one is Surugamide-769 found by VarQuest as a modification of surugamide B (mass is 897.6 Da). The ORFs 4751 (4 adenylation domains) and 4755 (6 adenylation domains), separating ORFs 4748 and 4759 are not shown. Three most likely amino acids for each adenylation domain are shown along with their NRPS2predictor [68] scores. The residues in these NRPs correlate with the residues predicted by NRPS2predictor (shown by the same colors). The biosynthetic gene cluster was identified by antiSMASH [69]. Although the previous studies of spectral networks [70–72] demonstrated that each connected component usually corresponds to related peptides (e.g., to a PNP family), the opposite statement is not necessarily true, i.e., related peptides sometimes form multiple connected components. E.g., the SpecNets approach missed Surugamide-769 because its connected component does not contain known surugamides.



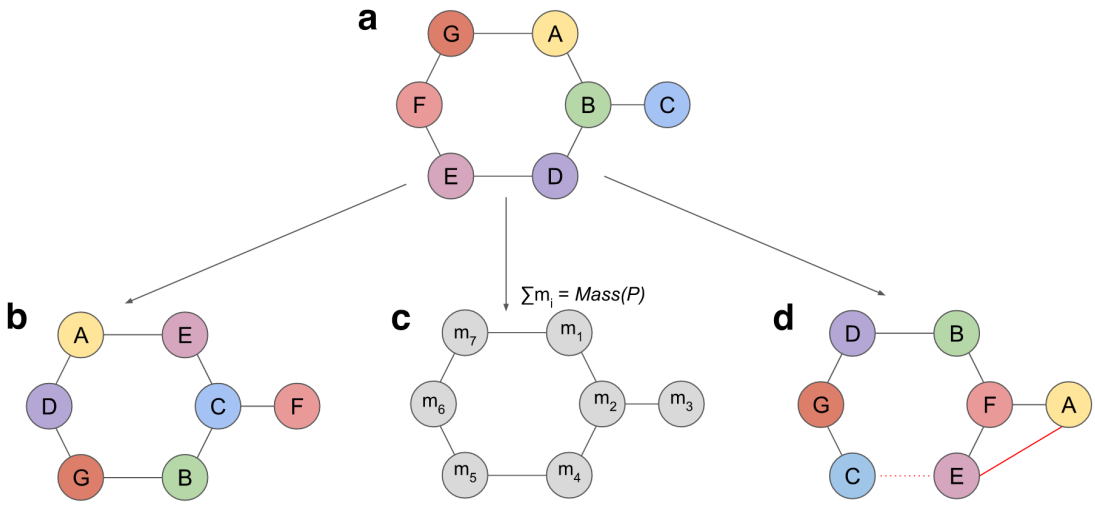
Supplementary Figure 3. Correspondence between theoretical spectra of a PNP and its variant. A PNP P (a) along with its theoretical spectrum (b) and a variant of P with a modification δ on one of its amino acids (c) along with its theoretical spectrum (d). Complementary peaks are arbitrarily labeled as $b-$ and $y-$ peaks with the same indexes. Corresponding peaks in both spectra have identical labels. For simplicity, only 6 peaks are shown.



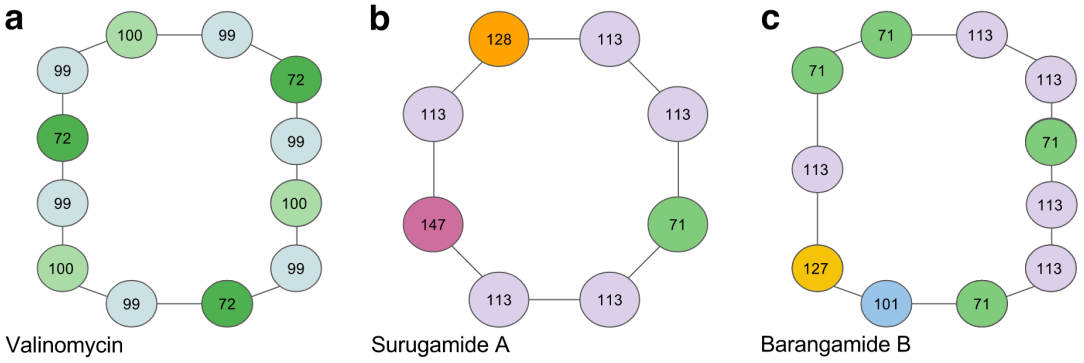
Supplementary Figure 4. **Preprocessing a given PNP database** *Peptides*. Stage 1: generating theoretical spectra for each peptide in the database. Stage 2: combining and sorting peaks from each spectrum altogether to create a list of sorted theoretical peaks $SortedPeaks(Peptides)$. Each peak is associated with the related peptide (shown by color). Stage 3: indexing the list of peaks to create an indexing table $Index(Peptides, M, \theta)$. The i-th cell contains a pointer to the minimal peak larger than or equal to $i\theta$.



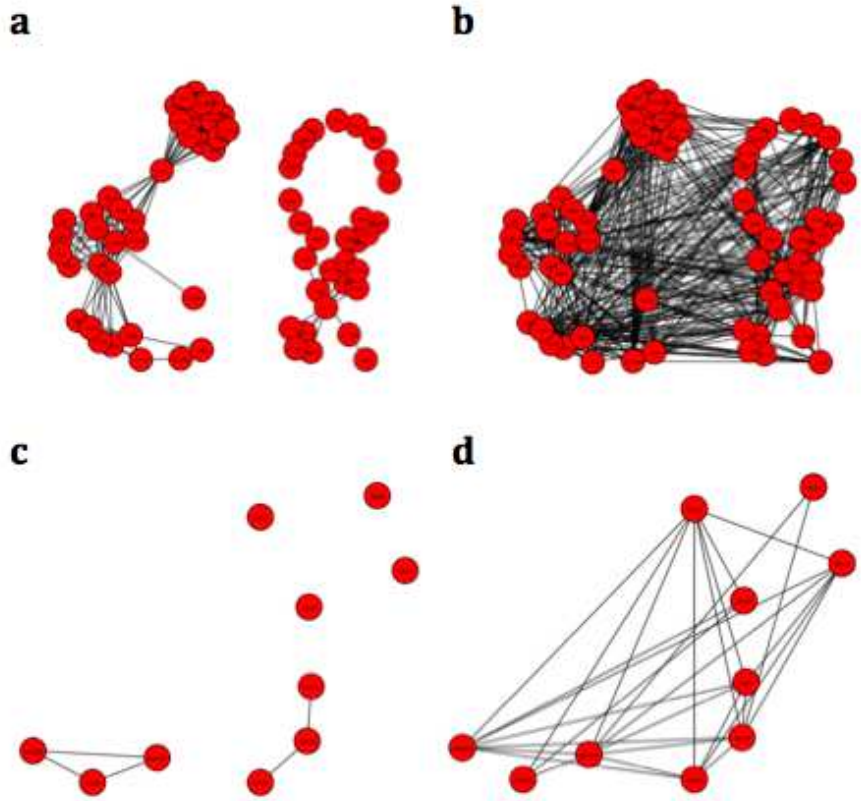
Supplementary Figure 5. **Scoring an experimental peak** s. Theoretical peak p matches s if $s - \varepsilon < p < s + \varepsilon$. We refer to the smallest (largest) matched peak as p_{lower} (p_{upper}). Stage 1: determine indexes of the intervals containing p_{lower} and p_{upper}. Stage 2: use binary search to find exact values of p_{lower} and p_{upper}. Stage 3: iterate from p_{lower} to p_{upper} to find all matching peaks.



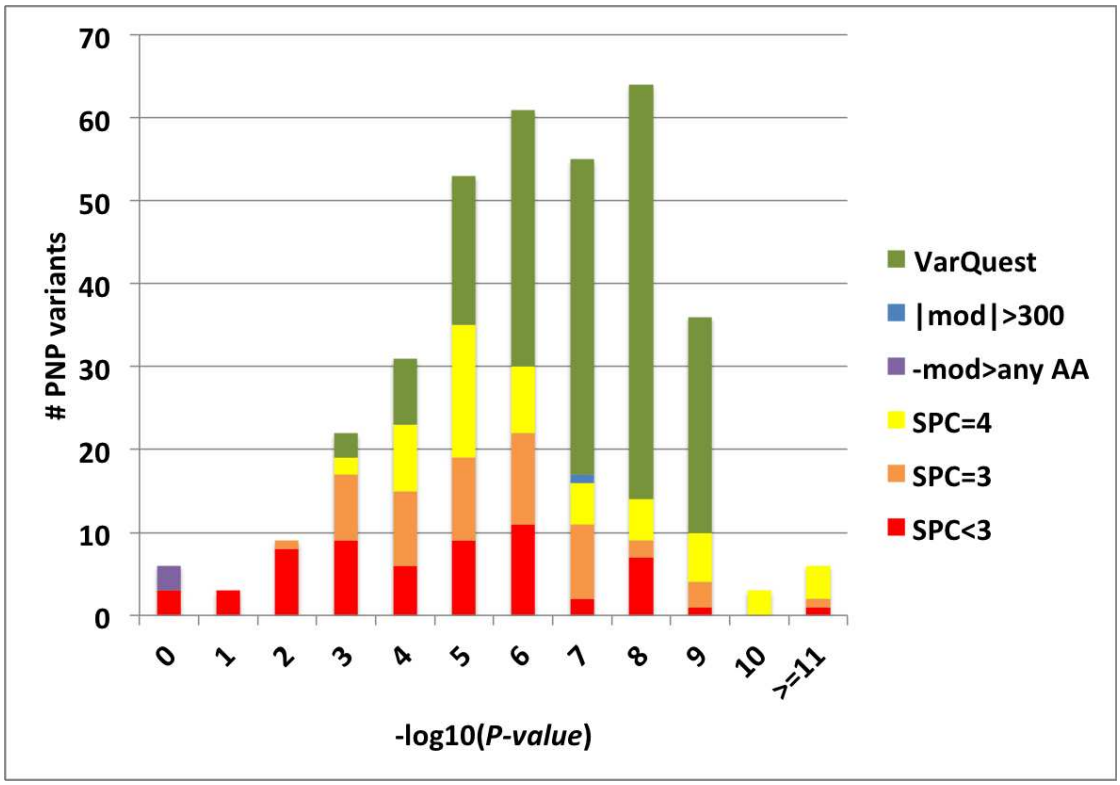
Supplementary Figure 6. Various decoy generation strategies. A target peptide (a) and its decoy versions generated by amino acid shuffling (b), mass redistribution (c), and amino acid shuffling plus displacement of a random bond (d).



Supplementary Figure 7. Examples of cyclic PNPs with high number of duplicated amino acids. Numbers in nodes refer to amino acid masses in Da. The nodes with equal mass are colored the same.



Supplementary Figure 8. Strain graphs for $Spectra_{CYANO}$ (top) and $Spectra_{STREP1}$ (bottom) datasets. Each node represents a strain and two nodes are connected by an edge if they share a common known PNP (DEREPLICATOR identifications: left) or PNP variants of the same known PNP (VarQuest identifications: right). Note that, two leftmost nodes are connected by an edge in the bottom left (DEREPLICATOR) but not the bottom right (VarQuest) figure. The corresponding strains share a single PNP identified by DEREPLICATOR and when we switch to VarQuest, one of these identifications changed to a variant of another PNP with a better P-value. As the result, while these two strains are in the same connected component, they are not connected by an edge in the bottom right Figure.



Supplementary Figure 9. Distribution of the first iteration P-values for PNP variants identified in the second iterative run of VarQuest on $Spectra_{CYANO}$. *VarQuest* stands for PNP variants identified by VarQuest on the first iteration but not reported due to P-values above the standard 10^{-10} threshold. $|mod| > 300$ stands for PNP variants with the absolute value of the total mass offset larger than default value of $MaxMod$=300 Da. $-mod > any\ AA$ is for PNP variants with negative mass offsets larger than mass of any amino acid in the corresponding known PNP (in this case $VariableScore$ is not defined and VarQuest reports P-value=1). SPC stands for PNP variants with specified $SPCScore(P, S)$ value. Note that a known PNP P is included in the VarQuest $CandidatePeptides(S)$ lists for a spectrum S if $SPCScore(P, S) \geq \eta$, where default value for η is 5.

Supplementary References

[1] M. Wang *et al.*, "Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking," *Nat. Biotechnol.*, vol. 34, pp. 828–837, Aug 2016.

[2] R. Gozalbes and A. Pineda-Lucena, "Small molecule databases and chemical descriptors useful in chemoinformatics: an overview," *Comb. Chem. High Throughput Screen.*, vol. 14, pp. 548–458, Jul 2011.

[3] J. Blunt, M. Munro, and H. Laatsch, "AntiMarin database," *University of Canterbury; Christchurch, New Zealand: University of Gottingen; Gottingen, Germany*, 2007.

[4] M. H. Medema *et al.*, "Minimum Information about a Biosynthetic Gene cluster," *Nat. Chemical Biology*, vol. 11, pp. 625–631, Sep 2015.

[5] X. Lucas, C. Senger, A. Erxleben, B. A. Gruning, K. Doring, J. Mosch, S. Flemming, and S. Gunther, "StreptomeDB: a resource for natural compounds isolated from Streptomyces species," *Nucleic Acids Res.*, vol. 41, pp. D1130–1136, Jan 2013.

[6] Y. Djoumbou Feunang, R. Eisner, C. Knox, L. Chepelev, J. Hastings, G. Owen, E. Fahy, C. Steinbeck, S. Subramanian, E. Bolton, R. Greiner, and D. S. Wishart, "ClassyFire: automated chemical classification with a comprehensive, computable taxonomy," *J Cheminform*, vol. 8, p. 61, 2016.

[7] Z. Ma, N. Geudens, N. P. Kieu, D. Sinnaeve, M. Ongena, J. C. Martins, and M. Hofte, "Biosynthesis, Chemical Structure, and Structure-Activity Relationship of Orfamide Lipopeptides Produced by Pseudomonas protegens and Related Species," *Front Microbiol*, vol. 7, p. 382, 2016.

[8] W. Li, H. Rokni-Zadeh, M. De Vleeschouwer, M. G. Ghequire, D. Sinnaeve, G. L. Xie, J. Rozenski, A. Madder, J. C. Martins, and R. De Mot, "The antimicrobial compound xantholysin defines a new group of Pseudomonas cyclic lipopeptides," *PLoS ONE*, vol. 8, no. 5, p. e62946, 2013.

[9] C. Bassarello, S. Lazzaroni, G. Bifulco, P. Lo Cantore, N. S. Iacobellis, R. Riccio, L. Gomez-Paloma, and A. Evidente, "Tolaasins A–E, five new lipodepsipeptides produced by Pseudomonas tolaasii," *J. Nat. Prod.*, vol. 67, pp. 811–816, May 2004.

[10] I. Kuiper, E. L. Lagendijk, R. Pickford, J. P. Derrick, G. E. Lamers, J. E. Thomas-Oates, B. J. Lugtenberg, and G. V. Bloemberg, "Characterization of two Pseudomonas putida lipopeptide biosurfactants, putisolvin I and II, which inhibit biofilm formation and break down existing biofilms," *Mol. Microbiol.*, vol. 51, pp. 97–113, Jan 2004.

[11] J. Gerard, R. Lloyd, T. Barsby, P. Haden, M. T. Kelly, and R. J. Andersen, "Massetolides A-H, antimycobacterial cyclic depsipeptides produced by two pseudomonads isolated from marine habitats," *J. Nat. Prod.*, vol. 60, pp. 223–229, Mar 1997.

[12] D. Sinnaeve, P. M. Hendrickx, J. Van Hemel, E. Peys, B. Kieffer, and J. C. Martins, "The solution structure and self-association properties of the cyclic lipodepsipeptide pseudodesmin A support its pore-forming potential," *Chemistry*, vol. 15, pp. 12653–12662, Nov 2009.

[13] D. Sorensen, T. H. Nielsen, C. Christophersen, J. S?rensen, and M. Gajhede, "Cyclic lipoundecapeptide amphisin from Pseudomonas sp. strain DSS73," *Acta Crystallogr C*, vol. 57, pp. 1123–1124, Sep 2001.

[14] K. Arima *et al.*, "Surfactin, a crystalline peptide lipid surfactant produced by Bacillus subtilis: Isolation, characterization and its inhibition of fibrin clot formation," *Biochem. Biophys. Res. Commun*, vol. 31, pp. 488–494, 1968.

[15] T. Ito and H. Ogawa, "Chemical studies on the antibiotic esperin," *Bulletin of the Agricultural Chemical Society of Japan*, vol. 23, no. 6, pp. 536–547, 1959.

[16] N. Naruse, O. Tenmyo, S. Kobaru, H. Kamei, T. Miyaki, M. Konishi, and T. Oki, "Pumilacidin, a complex of new antiviral antibiotics. Production, isolation, chemical properties, structure and biological activity," *J. Antibiot.*, vol. 43, pp. 267–280, Mar 1990.

[17] Q. Zhou *et al.*, "Xentrivalpeptides AQ: Depsipeptide Diversification in Xenorhabdus," *J. Nat. Prod.*, vol. 75, no. 10, pp. 1717–1722, 2012.

[18] F. Peypoux, F. Besson, G. Michel, and L. Delcambe, "Structure of bacillomycin D, a new antibiotic of the iturin group," *Eur. J. Biochem.*, vol. 118, pp. 323–327, Aug 1981.

[19] Y. Kajimura, M. Sugiyama, and M. Kaneda, "Bacillopeptins, new cyclic lipopeptide antibiotics from Bacillus subtilis FR-2," *J. Antibiot.*, vol. 48, pp. 1095–1103, Oct 1995.

[20] Y. Esumi, Y. Suzuki, Y. Itoh, M. Chijimatsu, M. Uramoto, K. Kimura, S. Nakayama, M. Yoshihama, T. Ichikawa, T. Haramo, and J. Fujishige, "SNA-60-367 components, new peptide enzyme inhibitors of aromatase: structure of the fatty acid side chain and amino acid sequence by mass spectrometry," *J. Antibiot.*, vol. 56, pp. 716–720, Aug 2003.

[21] K. Kaida *et al.*, "New cyclic depsipeptide antibiotics, clavariopsins a and b, produced by an aquatic hyphomycetes, clavariopsis aquatica. 1. taxonomy, fermentation, isolation, and biological properties," *J Antibiot (Tokyo)*, vol. 54, no. 1, pp. 17–21, 2001.

[22] R. B. Bates, S. Caldera, and M. D. Ruane, "Synthesis and Stereochemistry of Axinastatin 4," *J. Nat. Prod.*, vol. 61, p. 405, Mar 1998.

[23] A. Sakurai *et al.*, "Isolation and structure elucidation of substance-ib, a hexapeptide inducing sexual agglutination in saccharomyces cerevisiae," *Agricultural and Biological Chemistry*, vol. 40, no. 7, pp. 1451–1452, 1976.

[24] S. R. Ibrahim, C. C. Min, F. Teuscher, R. Ebel, C. Kakoschke, W. Lin, V. Wray, R. Edrada-Ebel, and P. Proksch, "Callyaerins A-F and H, new cytotoxic cyclic peptides from the Indonesian marine sponge Callyspongia aerizusa," *Bioorg. Med. Chem.*, vol. 18, pp. 4947–4956, Jul 2010.

[25] J. Vicente, B. Vera, A. D. Rodriguez, I. Rodriguez-Escudero, and R. G. Raptis, "Euryjanicin A: a new cycloheptapeptide from the Caribbean marine sponge Prosuberites laughlini," *Tetrahedron Lett.*, vol. 50, pp. 4571–4574, Aug 2009.

[26] B. F. Erlanger and L. Goode, "Gramicidin S; relationship of cyclic structure to antibiotic activity," *Nature*, vol. 174, pp. 840–841, Oct 1954.

[27] J. Kobayashi *et al.*, "Hymenamide F, new cyclic heptapeptide from marine sponge Hymeniacidon sp," *Tetrahedron*, vol. 52, no. 18, pp. 6355–6360, 1996.

[28] N. Sitachitta *et al.*, "Yanucamides A and B, two new depsipeptides from an assemblage of the marine cyanobacteria Lyngbya majuscula and Schizothrix species," *J Nat Prod*, vol. 63, no. 2, pp. 197–200, 2000.

[29] J. B. MacMillan, M. A. Ernst-Russell, J. S. de Ropp, and T. F. Molinski, "Lobocyclamides A-C, lipopeptides from a cryptic cyanobacterial mat containing Lyngbya confervoides," *J. Org. Chem.*, vol. 67, pp. 8210–8215, Nov 2002.

[30] R. Fernandez *et al.*, "Malaysiatin, the first cyclic heptapeptide from a marine sponge," *Tetrahedron Letters*, vol. 33, no. 40, pp. 6017–6020, 1992.

[31] O. Sterner, W. Etzel, A. Mayer, and H. Anke, "Omphalotin, a new cyclic peptide with potent nematicidal activity from omphalotus olearius ii. isolation and structure determination," *Natural Product Letters*, vol. 10, no. 1, pp. 33–38, 1997.

[32] G. R. Pettit, J. W. Lippert, S. R. Taylor, R. Tan, and M. D. Williams, "Synthesis of phakellistatin 11: a micronesia (Chuuk) marine sponge cyclooctapeptide," *J. Nat. Prod.*, vol. 64, pp. 883–891, Jul 2001.

[33] D. Williams, K. Yu, H. Behrisch, R. Van Soest, and R. Andersen, "Rolloamides A and B, cytotoxic cyclic heptapeptides isolated from the Caribbean marine sponge Eurypon laughlini," *J Nat Prod*, vol. 72, no. 7, pp. 1253–7, 2009.

[34] M. Arai, Y. Yamano, M. Fujita, A. Setiawan, and M. Kobayashi, "Stylissamide X, a new proline-rich cyclic octapeptide as an inhibitor of cell migration, from an Indonesian marine sponge of Stylissa sp," *Bioorg. Med. Chem. Lett.*, vol. 22, pp. 1818–1821, Feb 2012.

[35] X. Zou, S. Niu, J. Ren, E. Li, X. Liu, and Y. Che, "Verrucamides A-D, antibacterial cyclopeptides from Myrothecium verrucaria," *J. Nat. Prod.*, vol. 74, pp. 1111–1116, May 2011.

[36] J. Tabudravu *et al.*, "Wainunuamide, a histidine-containing proline-rich cyclic heptapeptide isolated from the fijian marine sponge Stylotella aurantium," *Tetrahedron Letters*, vol. 42, no. 52, pp. 9273–9276, 2001.

[37] K. Takada, A. Ninomiya, M. Naruse, Y. Sun, M. Miyazaki, Y. Nogi, S. Okada, and S. Matsunaga, "Surugamides A-E, cyclic octapeptides with four D-amino acid residues, from a marine Streptomyces sp.: LC-MS-aided inspection of partial hydrolysates for the distinction of D- and L-amino acid residues in the sequence," *J. Org. Chem.*, vol. 78, pp. 6746–6750, Jul 2013.

[38] A. Pesic, H. I. Baumann, K. Kleinschmidt, P. Ensle, J. Wiese, R. D. Sussmuth, and J. F. Imhoff, "Champacyclin, a new cyclic octapeptide from Streptomyces strain C42 isolated from the Baltic Sea," *Mar Drugs*, vol. 11, pp. 4834–4857, Dec 2013.

[39] H. Mohimani, A. Gurevich, A. Mikheenko, N. Garg, L. F. Nothias, A. Ninomiya, K. Takada, P. C. Dorrestein, and P. A. Pevzner, "Dereplication of peptidic natural products through database search of mass spectra," *Nat. Chem. Biol.*, vol. 13, pp. 30–37, Jan 2017.

[40] R. Sekizawa, I. Momose, N. Kinoshita, H. Naganawa, M. Hamada, Y. Muraoka, H. Iinuma, and T. Takeuchi, "Isolation and structural determination of phepropeptins A, B, C, and D, new proteasome inhibitors, produced by Streptomyces sp," *J Antibiot (Tokyo)*, vol. 54, no. 11, pp. 874–881, 2001.

[41] Q. Li, Y. Song, X. Qin, X. Zhang, A. Sun, and J. Ju, "Identification of the biosynthetic gene cluster for the anti-infective desotamides and production of a new analogue in a heterologous host," *J Nat Prod*, vol. 78, no. 4, pp. 944–8, 2015.

[42] W. T. Liu, A. Lamsa, W. R. Wong, P. D. Boudreau, R. Kersten, Y. Peng, W. J. Moree, B. M. Duggan, B. S. Moore, W. H. Gerwick, R. G. Linington, K. Pogliano, and P. C. Dorrestein, "MS/MS-based networking and peptidogenomics guided genome mining revealed the stenothricin gene cluster in Streptomyces roseosporus," *J. Antibiot.*, vol. 67, pp. 99–104, Jan 2014.

[43] K. Matsuno, Y. Yamada, C. Lee, and T. Nihira, "Identification by gene deletion analysis of barB as a negative regulator controlling an early process of virginiamycin biosynthesis in Streptomyces virginiae," *Arch Microbiol.*, vol. 181, pp. 52–9, Jan 2004.

[44] L. C. Vining and W. A. Taber, "Isariin, a new depsipeptide from isaria cretacea," *Canadian Journal of Chemistry*, vol. 40, no. 8, pp. 1579–1584, 1962.

[45] R. Baute, G. Deffieux, D. Merlet, M. Baute, and A. Neveu, "New insecticidal cyclodepsipeptides from the fungus Isaria felina. I. Production, isolation and insecticidal properties of isariins B, C and D," *J Antibiot (Tokyo)*, vol. 34, no. 10, pp. 1261–1265, 1981.

[46] S. Kiyotake, M. Tsuyoshi, S. Takunobu, I. Takashi, F. Tatsuya, T. Noboru, H. Kozue, T. Masaki, I. Taiji, K. Hideo, and Y. Kiyoyuki, "Aurilide, a cytotoxic depsipeptide from the sea hare Dolabella auricularia: isolation, structure determination, synthesis, and biological activity," *Tetrahedron*, vol. 60, pp. 8509–8527, Sep 2004.

[47] M. Roy *et al.*, "Barangamide A, a new cyclic peptide from the Indonesian sponge Theonella swinhoei," *Tetrahedron Letters*, vol. 40, pp. 5373–5376, Jul 1999.

[48] L. A. Romanenko, M. Uchino, N. I. Kalinovskaya, and V. V. Mikhailov, "Isolation, phylogenetic analysis and screening of marine mollusc-associated bacteria for antimicrobial, hemolytic and surface activities," *Microbiol. Res.*, vol. 163, no. 6, pp. 633–644, 2008.

[49] J. Malmstrom, A. Ryager, U. Anthoni, and P. H. Nielsen, "Unguisin C, a GABA-containing cyclic peptide from the fungus Emericella unguis," *Phytochemistry*, vol. 60, pp. 869–872, Aug 2002.

[50] D. Hopwood and H. Wright, "CDA is a new chromosomally-determined antibiotic from Streptomyces coelicolor A3(2)," *J Gen Microbiol*, vol. 192, no. 12, pp. 3575–9, 1983.

[51] J. Bitzer, V. Gesheva, and A. Zeeck, "Actinomycins with altered threonine units in the beta-peptidolactone," *J. Nat. Prod.*, vol. 69, pp. 1153–1157, Aug 2006.

[52] F. Kong and G. Carter, "Structure determination of glycinocins a to d, further evidence for the cyclic structure of the amphomycin antibiotics," *J Antibiot (Tokyo)*, vol. 56, no. 6, pp. 557–564, 2003.

[53] Z. Xie, L. Zhou, L. Guo, X. Yang, G. Qu, C. Wu, and S. Zhang, "Grisemycin, a Bridged Angucyclinone with a Methylsulfinyl Moiety from a Marine-Derived Streptomyces sp," *Org. Lett.*, vol. 18, pp. 1402–1405, Mar 2016.

[54] H. S. Lee, H. J. Shin, K. H. Jang, T. S. Kim, K. B. Oh, and J. Shin, "Cyclic peptides of the nocardamine class from a marine-derived bacterium of the genus Streptomyces," *J. Nat. Prod.*, vol. 68, pp. 623–625, Apr 2005.

[55] M. Richter *et al.*, "Streptofactin, a novel biosurfactant with aerial mycelium inducing activity from Streptomyces tendae Tu 901/8c," *FEMS Microbiology Letters*, vol. 163, no. 2, pp. 165–171, 1998.

[56] S. Kodani, K. Sato, H. Hemmi, and M. Ohnish-Kameyama, "Isolation and structural determination of a new hydrophobic peptide venepeptide from Streptomyces venezuelae," *J. Antibiot.*, vol. 67, pp. 839–842, Dec 2014.

[57] Z. Yu, S. Vodanovic-Jankovic, M. Kron, and B. Shen, "New WS9326A congeners from Streptomyces sp. 9078 inhibiting Brugia malayi asparaginyl-tRNA synthetase," *Org. Lett.*, vol. 14, pp. 4946–4949, Sep 2012.

[58] K. J. Wesson and M. T. Hamann, "Keenamide A, a bioactive cyclic peptide from the marine mollusk Pleurobranchus forskalii," *J. Nat. Prod.*, vol. 59, pp. 629–631, Jun 1996.

[59] M. Iwatsuki *et al.*, "Lariatins, antimycobacterial peptides produced by Rhodococcus sp. K01-B0171, have a lasso structure," *J Am Chem Soc*, vol. 128, no. 23, pp. 7486–91, 2006.

[60] O. Perlova, K. Gerth, S. Kuhlmann, Y. Zhang, and R. Muller, "Novel expression hosts for complex secondary metabolite megasynthetases: Production of myxochromide in the thermopilic isolate Corallococcus macrosporus GT-2," *Microb. Cell Fact.*, vol. 8, p. 1, Jan 2009.

[61] D. Hoffmann *et al.*, "Sequence analysis and biochemical characterization of the nostopeptolide A biosynthetic gene cluster from Nostoc sp. GSV224," *Gene*, vol. 311, pp. 171–80, 2003.

[62] K. K. Lee, J. B. Gloer, J. A. Scott, and D. Malloch, "Petriellin a: A novel antifungal depsipeptide from the coprophilous fungus petriella sordida," *The Journal of Organic Chemistry*, vol. 60, no. 17, pp. 5384–5385, 1995.

[63] A. Aoyagi *et al.*, "Pleofungins, novel inositol phosphorylceramide synthase inhibitors, from Phoma sp. SANK 13899," *J. Antibiot*, vol. 60, no. 2, pp. 143–52, 2007.

[64] C. Ruangviryachai *et al.*, "An exceptionally large pyoverdin from a pseudomonas strain collected in Thailand," *Z. Naturforsch*, vol. 55c, pp. 323–327, 2000.

[65] V. Hedge *et al.*, "A family of depsi-peptide fungal metabolites, as selective and competitive human tachykinin receptor (NK2) antagonists: Fermentation, isolation, physico-chemical properties, and biological activity," *J Antibiot (Tokyo)*, vol. 54, no. 2, pp. 125–135, 2001.

[66] M. Kronenwerth *et al.*, "Characterisation of taxlllaids A-G; natural products from Xenorhabdus indica," *Chemistry*, vol. 20, no. 52, pp. 17478–87, 2014.

[67] J. Cai, A. Bhatnagar, and W. M. Pierce, "Protein modification by acrolein: formation and stability of cysteine adducts," *Chem. Res. Toxicol.*, vol. 22, pp. 708–716, Apr 2009.

[68] M. Rottig, M. H. Medema, K. Blin, T. Weber, C. Rausch, and O. Kohlbacher, "NRPSpredictor2–a web server for predicting NRPS adenylation domain specificity," *Nucleic Acids Res.*, vol. 39, pp. W362–367, Jul 2011.

[69] M. H. Medema, K. Blin, P. Cimermancic, V. de Jager, P. Zakrzewski, M. A. Fischbach, T. Weber, E. Takano, and R. Breitling, "antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences," *Nucleic Acids Res.*, vol. 39, pp. W339–346, Jul 2011.

[70] N. Bandeira, "Spectral networks: a new approach to de novo discovery of protein sequences and posttranslational modifications," *BioTechniques*, vol. 42, pp. 687–695, Jun 2007.

[71] J. Watrous, P. Roach, T. Alexandrov, B. S. Heath, J. Y. Yang, R. D. Kersten, M. van der Voort, K. Pogliano, H. Gross, J. M. Raaijmakers, B. S. Moore, J. Laskin, N. Bandeira, and P. C. Dorrestein, "Mass spectral molecular networking of living microbial colonies," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 109, pp. E1743–1752, Jun 2012.

[72] H. Mohimani, W. T. Liu, Y. L. Yang, S. P. Gaudencio, W. Fenical, P. C. Dorrestein, and P. A. Pevzner, "Multiplex de novo sequencing of peptide antibiotics," *J. Comput. Biol.*, vol. 18, pp. 1371–1381, Nov 2011.