Supplementary material for Cycloquest:

Identification of cyclopeptides via database search

of their mass spectra against genome databases

Hosein Mohimani,[†] Wei-Ting Liu,[‡] Joshua S. Mylne,[¶] Aaron G. Poth,^{¶,§} Michelle L. Colgrave,[§] Dat Tran,^{||,⊥,#} Michael E. Selsted,^{||,⊥,#} Pieter C. Dorrestein,^{‡,@} and Pavel A. Pevzner^{*,△}

Department of Electrical and Computer Engineering, UC San Diego, Department of Chemistry and Biochemistry, UC San Diego, Institute for Molecular Bioscience, The University of Queensland, Brisbane, Division of Livestock Industries, CSIRO, Brisbane, Department of Pathology and Laboratory Medicine, School of Medicine, UC Irvine, Center for Immunology, UC Irvine, Department of Pathology and Laboratory Medicine, Keck School of Medicine, USC, Skaggs School of Pharmacy and Pharmaceutical Sciences, UC San Diego, and Department of Computer Science and Engineering, UC San Diego

E-mail: ppevzner@ucsd.edu

*To whom correspondence should be addressed

[†]Department of Electrical and Computer Engineering, UC San Diego

[‡]Department of Chemistry and Biochemistry, UC San Diego

[¶]Institute for Molecular Bioscience, The University of Queensland, Brisbane

[§]Division of Livestock Industries, CSIRO, Brisbane

^{||}Department of Pathology and Laboratory Medicine, School of Medicine, UC Irvine

[⊥]Center for Immunology, UC Irvine

[#]Department of Pathology and Laboratory Medicine, Keck School of Medicine, USC

[@]Skaggs School of Pharmacy and Pharmaceutical Sciences, UC San Diego

[△]Department of Computer Science and Engineering, UC San Diego

Fig. S1: The number of cyclopeptides reported in different species.

Fig. S2: (A) Distribution of cyclic scores of all 14 amino acid long candidate peptides in the database (within the 0.5 Da mass accuracy), matched to spectra of cyclic SFTI-1. The red arrow shows the maximum score. (B) The respective results for linear score and cyclic SFTI-1. (C) The respective results for cyclic score and linear SFTI-1[K,S]. (D) The respective results for linear score and linear SFTI-1[K,S].

Table S1: Cyclopeptides found between 1965-1997.

Peptide	Discovery year	Reference
Circulin A (Bacteria)	1965	5
Alpha Amanitin	1968	10
Circulin B (Bacteria)	1968	6
Kalata B1	1973	15,16
Virotoxin	1980	11
Subtilosin A	1985	2,3
Microcin J25	1992	4
Cyclosochotride A	1994	17
Circulins A and B (Plant)	1994	18

Text S1. Cyclic Score

Given a cyclopeptide *Peptide*, a singly-charged spectrum S (represented as a set of masses), and an accuracy threshold δ , we define $CyclicScore_{\delta}(Peptide, S)$ as the number of elements (masses) shared between $\Delta(Peptide)$ and S within an accuracy δ . We say that a mass m in the theoretical spectrum $\Delta(peptide)$ is shared with S within an accuracy δ if there exists a mass s in the experimental spectrum S such that $|s - (m + ChargeMass)| < \delta$, where $ChargeMass = 1.007276Da$. Similarly, for a multiply-charged spectrum S with maximum charge of $+n$, we define $CyclicScore_{\delta}(Peptide, S)$ as the number of elements (masses) shared between $\Delta(Peptide)$ and S within an accuracy δ , where a mass m in the theoretical spectrum $\Delta(peptide)$ is said to be shared with S within an accuracy δ with charge state i , if there exists a mass s in the experimental spectrum S such that $|s - \frac{m+i \cdot ChargeMass}{i}| < \delta$. Most cyclopeptides that we encountered had charges 1, 2 or 3.

Table S2: Cycloquest results for different reduction forms and different collision energies of SFTI-1 and SFT-L2. The table shows cycloquest usually works much better for reduced peptides. Moreover, the performance is much better in a fixed range of collision energies (between 4500eV to 6000eV).

Peptide	Modification	Mass	Collision Energy (eV)	Rank	Score	NScore
SFTI-1	Reduced	1515.90	4000	57	35	3.96
SFTI-1	Reduced	1515.90	4500	3	44	4.97
SFTI-1	Reduced	1515.90	5000	1	52	5.48
SFTI-1	Reduced	1515.90	5500	3	47	4.64
SFTI-1	Reduced	1515.90	6000	4	48	4.81
SFTI-1	Reduced	1515.90	6500	617	33	2.71
SFTI-1	Reduced	1515.90	7000	20	37	3.82
SFTI-1	Oxidized	1513.90	4000	1558	25	2.12
SFTI-1	Oxidized	1513.90	4500	584	33	2.73
SFTI-1	Oxidized	1513.90	5000	>10000		
SFTI-1	Oxidized	1513.90	5500	1398	32	2.23
SFTI-1	Oxidized	1513.90	6000	2090	34	2.07
SFTI-1	Oxidized	1513.90	6500	2391	31	1.87
SFTI-1	Oxidized	1513.90	7000	349	37	2.95
SFT-L1	Reduced	1205.38	4000	2	40	4.99
SFT-L1	Reduced	1205.38	4500	1	49	5.49
SFT-L1	Reduced	1205.38	5000	8	45	5.10
SFT-L1	Reduced	1205.38	5500	2	44	5.20
SFT-L1	Reduced	1205.38	6000	1	42	6.99
SFT-L1	Reduced	1205.38	6500	3	42	5.60
SFT-L1	Reduced	1205.38	7000	1	36	5.56
SFT-L1	Oxidized	1203.46	4000	13	34	4.44
SFT-L1	Oxidized	1203.46	4500	29	31	3.83
SFT-L1	Oxidized	1203.46	5000	458	25	2.77
SFT-L1	Oxidized	1203.46	5500	112	29	3.43
SFT-L1	Oxidized	1203.46	6000	209	28	3.19
SFT-L1	Oxidized	1203.46	6500	126	30	3.45
SFT-L1	Oxidized	1203.46	7000	87	31	3.52
SFTI-1[K,S]	Reduced	1533.67	4000	1	17	9.19
SFTI-1[K,S]	Reduced	1533.67	4500	1	21	10.81
SFTI-1[K,S]	Reduced	1533.67	5000	1	16	8.09
SFTI-1[K,S]	Reduced	1533.67	5500	1	12	6.48
SFTI-1[K,S]	Reduced	1533.67	6000	3	11	5.36
SFTI-1[K,S]	Reduced	1533.67	6500	2	12	6.17
SFTI-1[K,S]	Reduced	1533.67	7000	1	15	8.13
SFTI-1[K,S]	Oxidized	1531.65	4000	785	6	2.66
SFTI-1[K,S]	Oxidized	1531.65	4500	870	6	2.61
SFTI-1[K,S]	Oxidized	1531.65	5000	786	6	2.57
SFTI-1[K,S]	Oxidized	1531.65	5500	>10000		
SFTI-1[K,S]	Oxidized	1531.65	6000	6701	4	1.42
SFTI-1[K,S]	Oxidized	1531.65	6500	2618	5	2.01
SFTI-1[K,S]	Oxidized	1531.65	7000	2708	5	1.94
SFTI-1[D,G]	Oxidized	1531.70	4000	193	8	3.26
SFTI-1[D,G]	Oxidized	1531.70	4500	418	7	2.87
SFTI-1[D,G]	Oxidized	1531.70	5000	175	8	3.40
SFTI-1	Iodoacetamide	1629.77		14	40	4.13
SFTI-1	Maleimide	1709.70		266	33	3.17
SFT-L1	Iodoacetamide	1319.50		1	45	6.44
SFT-L1	Maleimide	1399.45		2	55	6.08