A blind test of computational technique for predicting peptide sequences which can be cyclized by the cyanobactin macrocyclase $PatG_{mac}$.

Materials, Experimental Methods, and Experimental Data.

Peptide Synthesis

Fmoc amino acid derivatives, 2-(1H-benzotriazol-1-yl)-1,1,3,3- tetramethyluronium hexafluorophosphate (HBTU) and 2-chlorotrityl chloride resin were purchased from Novabiochem, Merck Biosciences, UK. Trifluoroacetic acid (TFA), N,N-diisopropylethylamine (DIEA), N,N-dimethylformamide (DMF) and piperidine were obtained from Alfa Aesar, UK, and used without further purification.

The linear peptide substrates were synthesized manually in-house using the standard Fmoc-based solid phase peptide synthesis (SPPS) strategy Ref: (Chan, W.; White, P.D (1999). *Fmoc Solid Phase Peptide Synthesis: A Practical Approach*. Clarendon Press, Oxford: Oxford University Press). In brief, the attachment of the first amino acid was conducted using a molar ratio of amino acid: DIEA: resin of 1.2: 4.8: 1 in DCM. Subsequent amino acids were sequentially coupled following removal of the Fmoc protecting group at each cycle. Fmoc deprotection steps were carried out with 20% piperidine in DMF (v/v) for 6 min (3x: 2min each); coupling reactions were performed in DMF using a molar ratio of amino acid: HBTU: DIEA: resin of 5: 5: 10: 1. Reactions were monitored using the Kaiser test.

The peptides were cleaved from the support and side chain protecting groups removed by treatment with a mixture consisting of 95% TFA, 2.5% triisopropylsilane (TIPS), 2.5% H_2O (20 mL of mixture per g of peptide resin, 5 hours at room temperature). The resin was then filtered and washed twice with TFA. The combined filtrates were concentrated under reduced pressure. The peptide was precipitated with cold diethyl ether and recovered by centrifugation. The peptide was submitted to LC-MS analysis and the sequence was verified by MS-MS analysis.

Protein Purification

PatGmac and LynGmac macrocyclases were cloned from genomic DNA into pHISTEV vectors and expressed in *Escherichia coli* BL21 (DE3) grown on auto-induction medium (Formedium Terrific broth base containing trace elements) for 48 h at 20 °C, with shaking at 200 rpm. Cells were harvested by centrifugation at 4,000 x g, 4 °C, for 15 min, and re-suspended in lysis buffer (500 mM NaCl, 20 mM Tris, pH 8.0, 20 mM Imidazole, 3mM BME, DTA-free protease inhibitor tablets (Roche) and 0.4 mg DNase (Sigma) g−1 wet cells). Cells were lysed by passage through a cell disruptor at 30 kPSI and the lysate was cleared by centrifugation at 20,000 x g, 4 °C for 45 min. It was then filtered through a 0.45 µm membrane filter and then to a Ni-sepharose FF column (GE Healthcare) prewashed with lysis buffer. Protein was eluted with 250 mM Imidazole and passed through a HiPrep 26/10 Desalting column (GE Healthcare) into buffer containing 350mM NaCl and 10mM Bicine. Desalted protein was concentrated using Vivaspin 20 MWCO 30 000 concentrators (GE Healthcare) down to required concentrations for future enzymatic assays.

Enzyme assays

Macrocyclisation reactions were prepared for 100µM peptide substrate and 20µM protein in buffer containing 350mM NaCl, 10mM Bicine pH 7.5 and 5% DMSO. Reaction mixtures were incubated at 37°C with shaking at 200 rpm for up to 4 days and monitored by ESI LCMS.

List of peptides tested

Peptide original order number	Peptide Sequence and reference to experimental data			P(PCC)	BXD prediction	Experimental result	BXD correct?
2	VPAPIPFP	C	(Fig. S1, S2)	8.70E-01	Yes	Yes	**Yes**
20	VTR(ThH)VTM(ThH)	C	(Fig. S3, S4)	8.50E-01	Yes	Yes	**Yes**
18	RTV(ThH)MTV(ThH)	C	(Fig. S5, S6)	8.40E-01	Yes	Yes	**Yes**
21	VTM(ThH)VTR(ThH)	C	(Fig. S7, S8)	5.80E-01	Yes	Yes	**Yes**
8	ZSKLQIDP	C	(Fig. S9, S10)	5.50E-01	Yes	Yes	**Yes**
19	MTV(ThH)RTV(ThH)	C	(Fig. S11, S12)	3.40E-01	Yes	Yes	**Yes**
3	QENHVFIQFP	B	(Fig. S13, S14)	1.80E-01	Yes	Yes	**Yes**
5	VGAGIGF(Pip)	C	(Fig. S15)	1.30E-01	Yes	No	**No**
4	EDWYFDHP	B	(Fig. S16, S17)	9.20E-02	Yes	Yes	**Yes**
22	NEFMQTGSYSGP	A	(Fig. S18)	9.00E-02	Yes	No	**No**
14	Z(Ac)SKLQIDP	C	(Fig. S19, S20)	8.30E-02	Yes	Yes	**Yes**
6	VGAGIGF(ΨP)	C	(Fig. S21)	7.40E-02	No	No	**Yes**
23	DCSPAKCSLLCSNP	C	(Fig. S22, S23)	2.40E-02	No	Yes	**No**
24	LTPGQWHMKWVP	B	(Fig. S24)	1.90E-02	No	No	**Yes**
10	Z(Fmoc)SKLQIDP	C	(Fig. S25, S26)	1.3E-02	No	Yes	**No**
1	CITJC	C	(Fig. S27)	7.80E-03	No	No	**Yes**
7	GSKLQIDP	C	(Fig. S28)	5.50E-03	No	No	**Yes**
25	VALKLALKLALPRGPRP	C	(Fig. S29)	1.50E-03	No	No	**Yes**
12	Z(TFA)SKLQIDP	C	(Fig. S30)	3.90E-04	No	No	**Yes**
13	Z(TFA)SKLQIDP	C	(Fig. S31)	8.00E-05	No	No	**Yes**
16	S(Ac)SKLQIDP	C	(Fig. S32)	4.70E-05	No	No	**Yes**
17	T(Ac)SKLQIDP	C	(Fig. S33)	1.10E-05	No	No	**Yes**
15	Z(Ac)SKLQIDP	C	(Fig. S34)	9.40E-07	No	No	**Yes**
9	ZSKLQIDP	C	(Fig. S35)	5.70E-10	No	No	**Yes**
11	Z(Fmoc)SKLQIDP	C	(Fig. S36)	2.50E-10	No	No	**Yes**

Table 1: List of peptides tested; J = Fmoc-L-propargylglycine; Pip = Piperidine; ΨP = pseudo Proline; Z = amino Alanine; Fmoc = Fmoc protected; TFA = TFA protected; Ac = Acetylated; Amino acids in red are in D conformation.

The table also contains the reference to the origin of the peptides.

A: From the reference: (Leikoski, David P. Fewer, Jouni Jokela, Matti Wahlsten, Leo Rouhiainen, and Kaarina Sivonen, Highly Diverse Cyanobactins in Strains of the Genus *Anabaena*_Niina APPLIED AND ENVIRONMENTAL MICROBIOLOGY, 2010 76 (Feb.1) pp. 701–709)

B: Randomly generated sequences from RandSeq (Expasy)

C: sequences designed to generate analogues of homophymine A. See ref: (Angela Zampella, Valentina Sepe, Paolo Luciano, Filomena Bellotta, Maria Chiara Monti, Maria Valeria D'Auria, Trine Jepsen, Sylvain

Petek, Marie-Thérèse Adeline, Olivier Laprévôte, Anne-Marie Aubertin, Cécile Debitus, Christiane Poupat and Alain Ahond, Homophymine A, an Anti-HIV Cyclodepsipeptide from the Sponge *Homophymia* sp. J. Org Chem 2008 73 (14) pp 5319-5327)
The reference to the experimental data figures is also given

LC-MS data

Positive-mode LC-MS data show the mass-over-charge ratio, the charge, the chemical formula, the theoretical mass (derived from the chemical formula) and the deviation between the theoretical and the actual masses observed for each substrate. A low delta/deviation (-5 to 5) value indicates the mass observed corresponds to the chemical formula expected for each compound. Single and double charged ions were detected for both linear and cyclic forms for the peptides tested. The data is arranged in the order of descending predicted P(PCC). For all substrates with high P(PCC) the presence of cyclic form was identified and for all substrates with low P(PCC) the cyclic form was not detected.

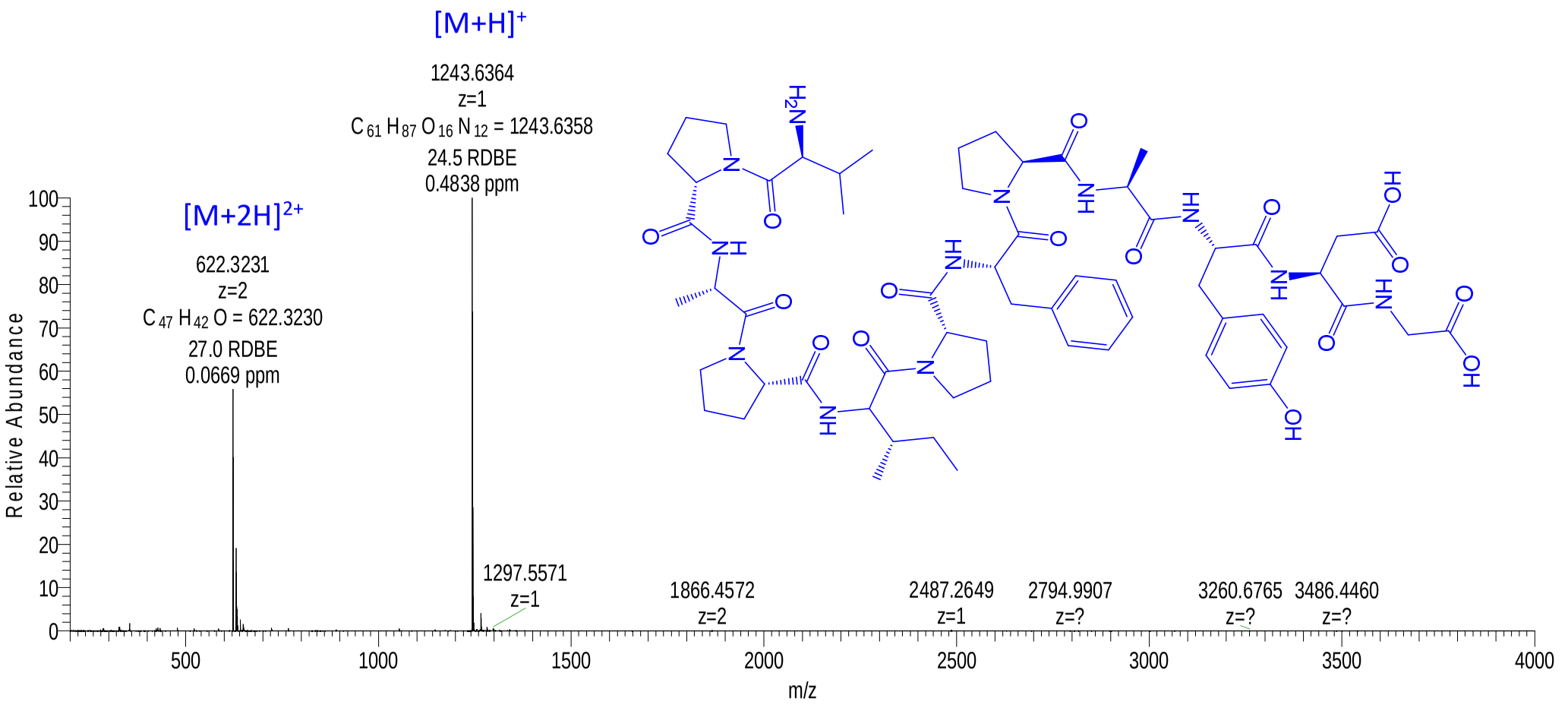


Fig. S1: ESI LC-MS for Peptide 2: Linear VPAPIPFP - AYDG

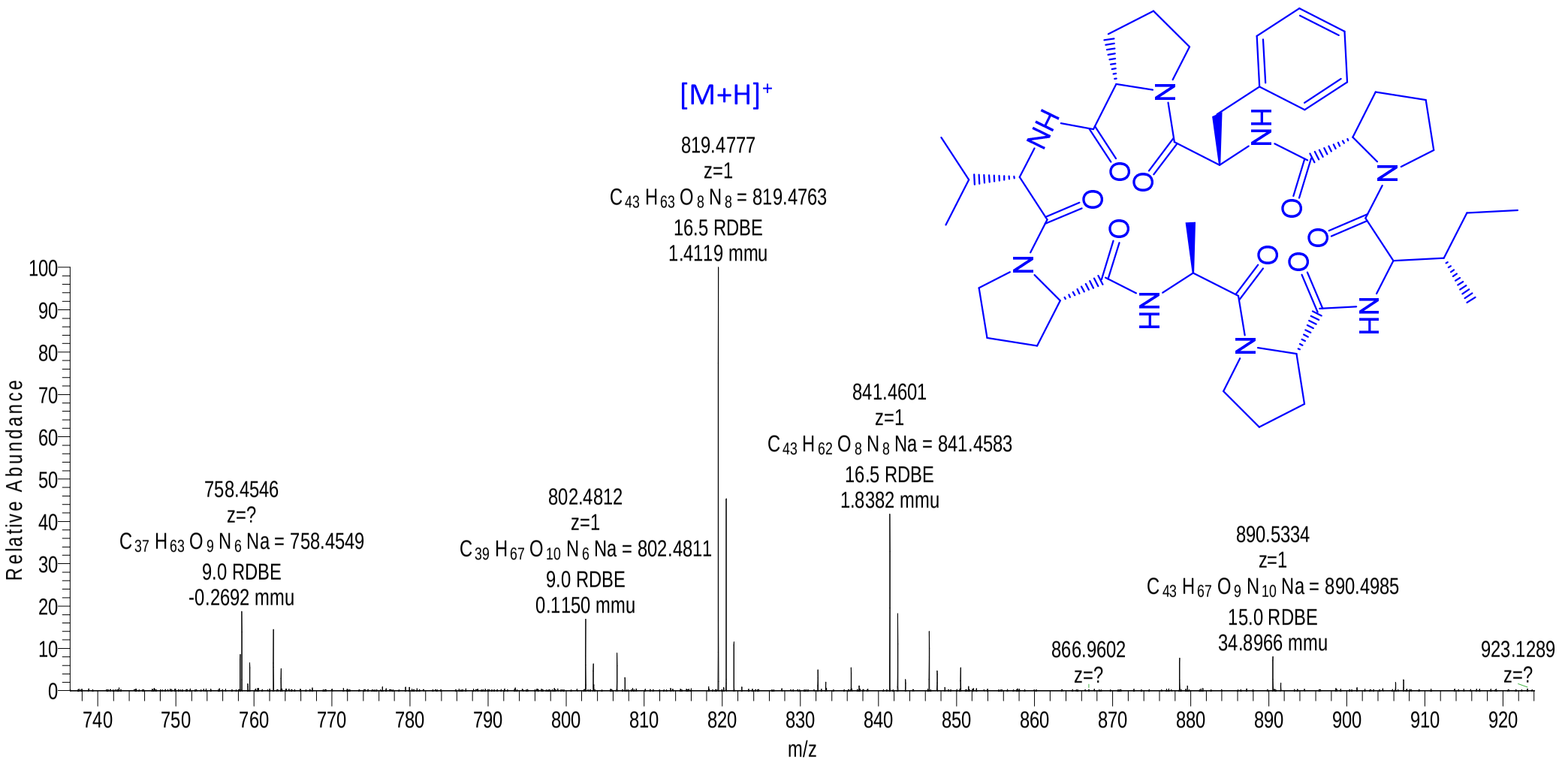


Fig. S2: ESI LC-MS for cyclic Peptide 2: Cyclo [VPAPIPFP]

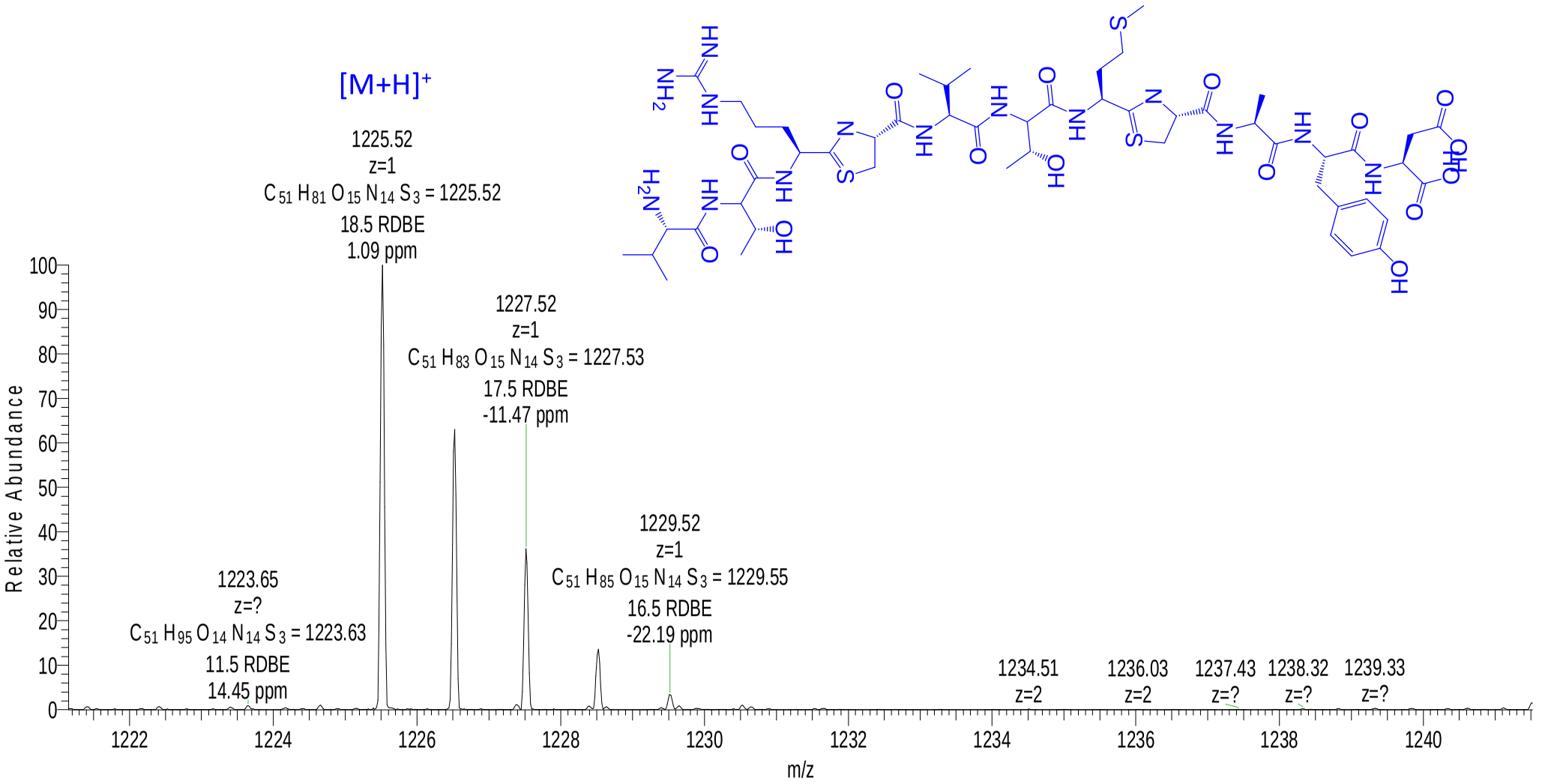


Fig. S3: ESI LC-MS for Peptide 20: Linear VTR(ThH)VTM(ThH) - AYDG

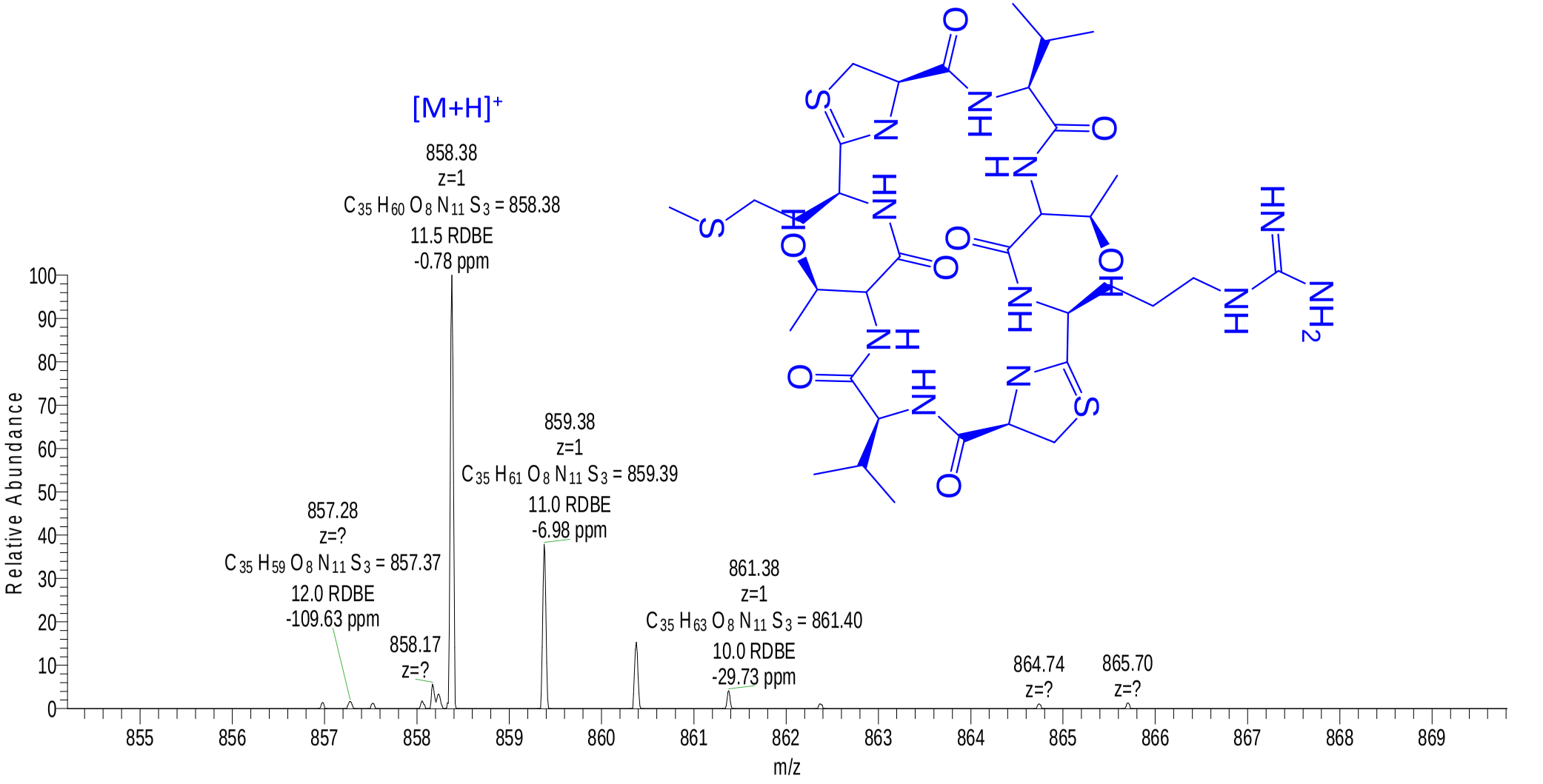


Fig. S4: ESI LC-MS for cyclic Peptide 20: Cyclo [VTR(ThH)VTM(ThH)]

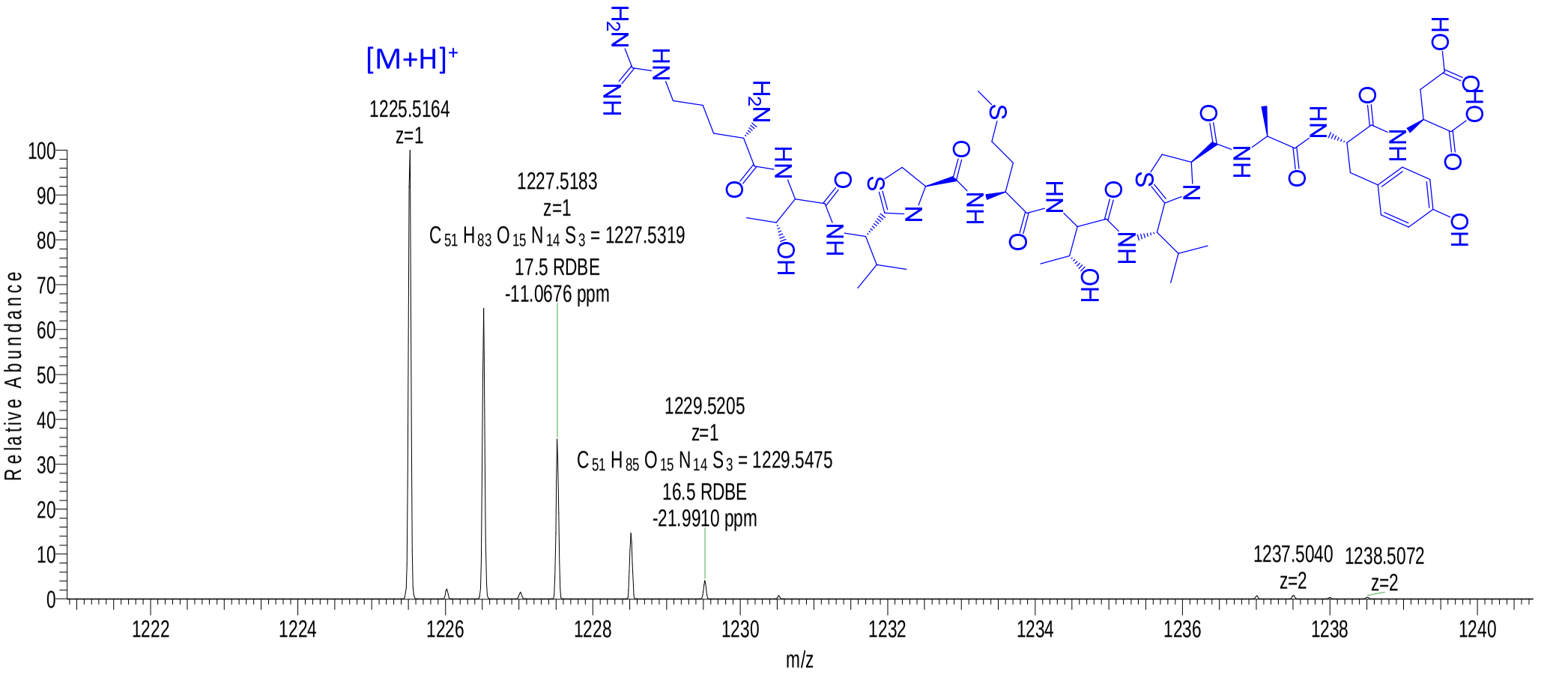


Fig. S5: ESI LC-MS for Peptide 18: Linear RTV(ThH)MTV(ThH) - AYDG

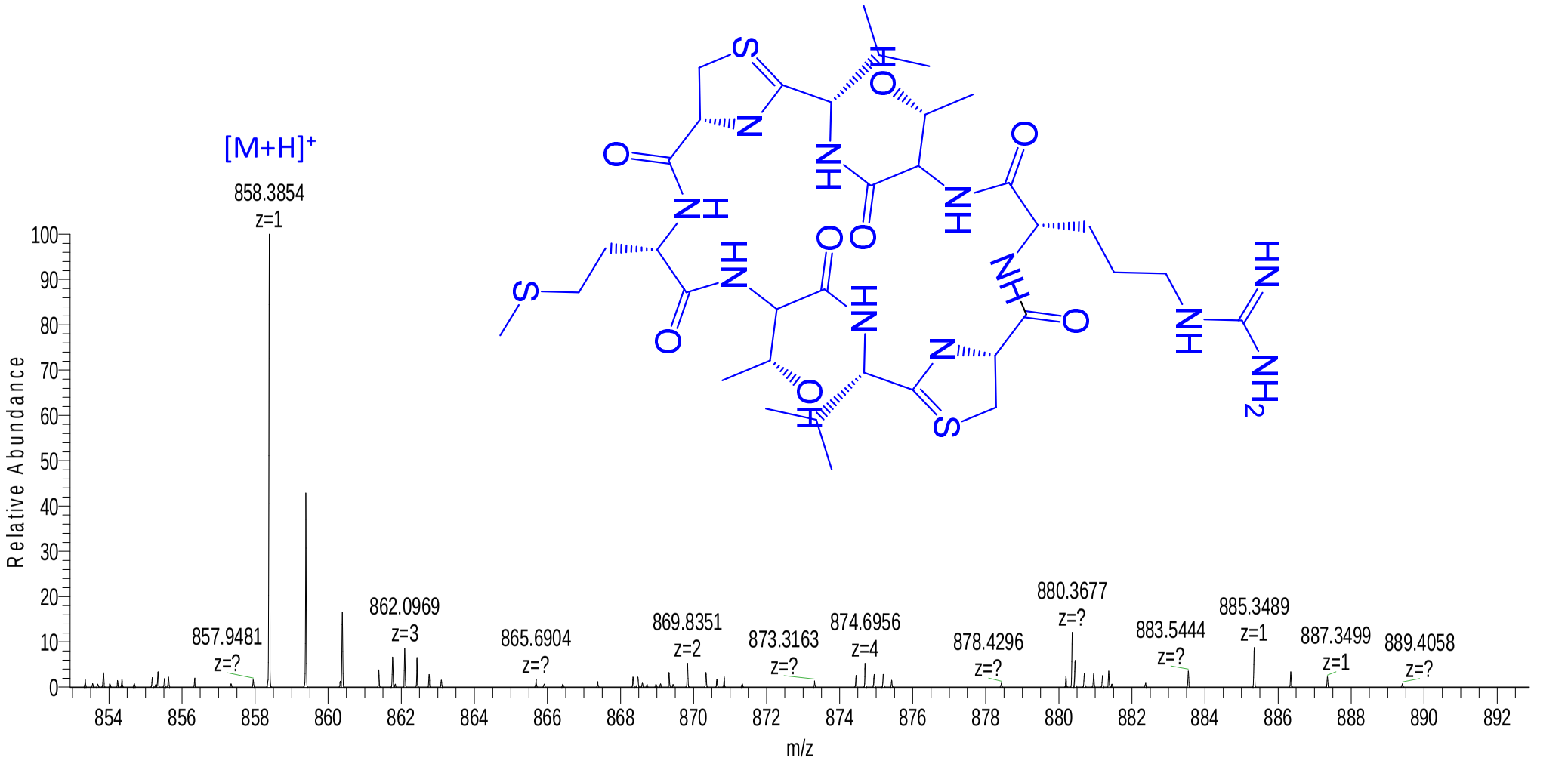


Fig. S6: ESI LC-MS for cyclic Peptide 18: Cyclo [RTV(ThH)MTV(ThH)]

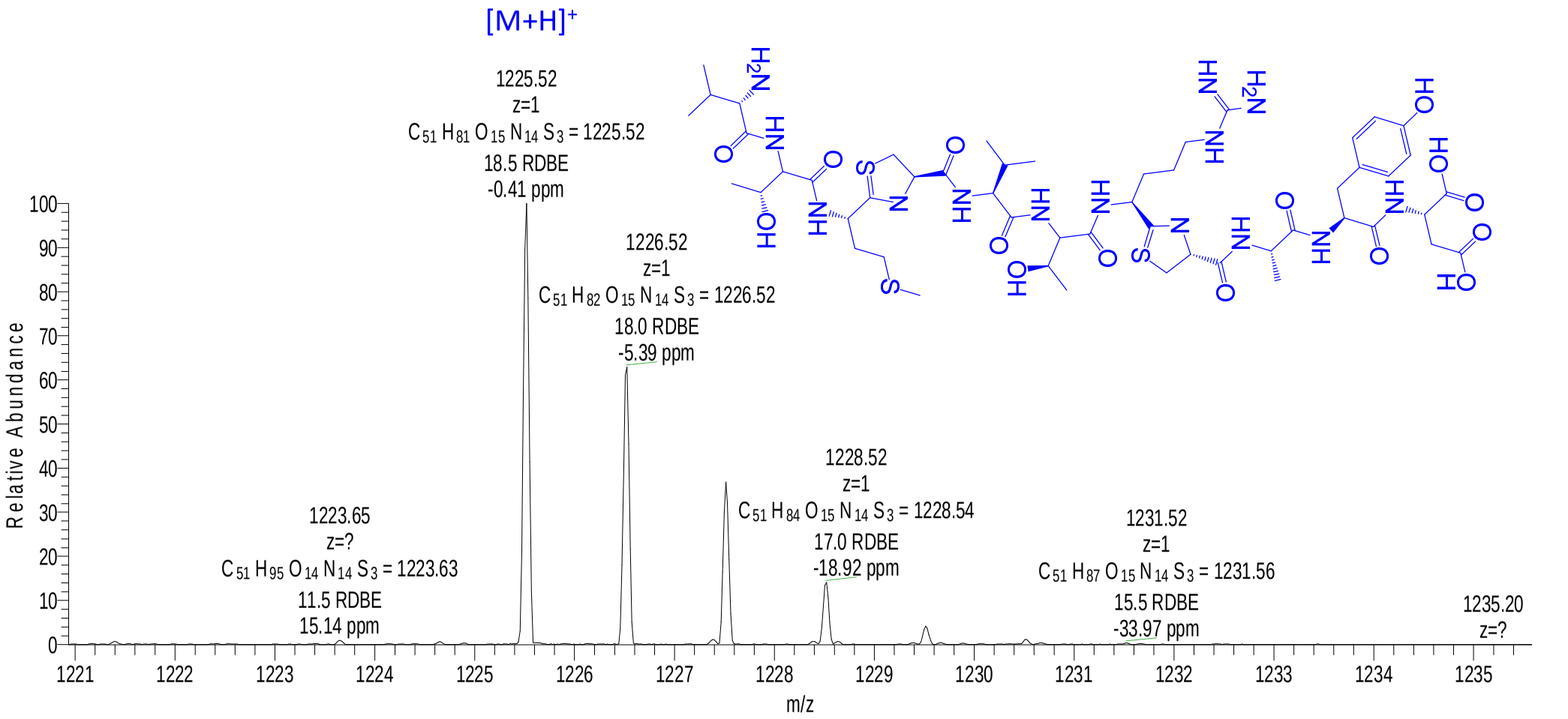


Fig. S7: ESI LC-MS for Peptide 21: Linear VTM((ThH)VTR(ThH) - AYDG

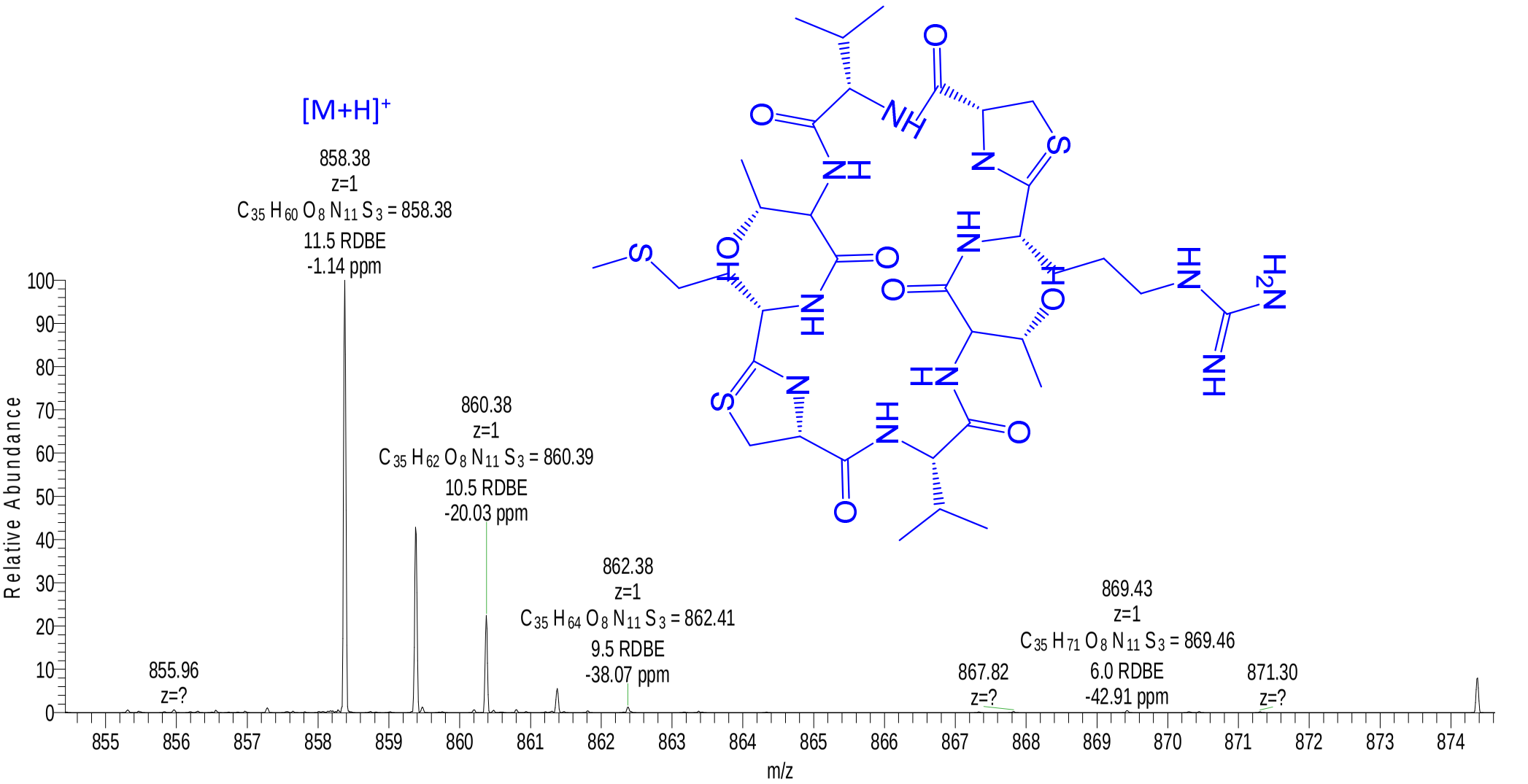


Fig. S8: ESI LC-MS for cyclic Peptide 21: Cyclo [VTM((ThH)VTR(ThH)]

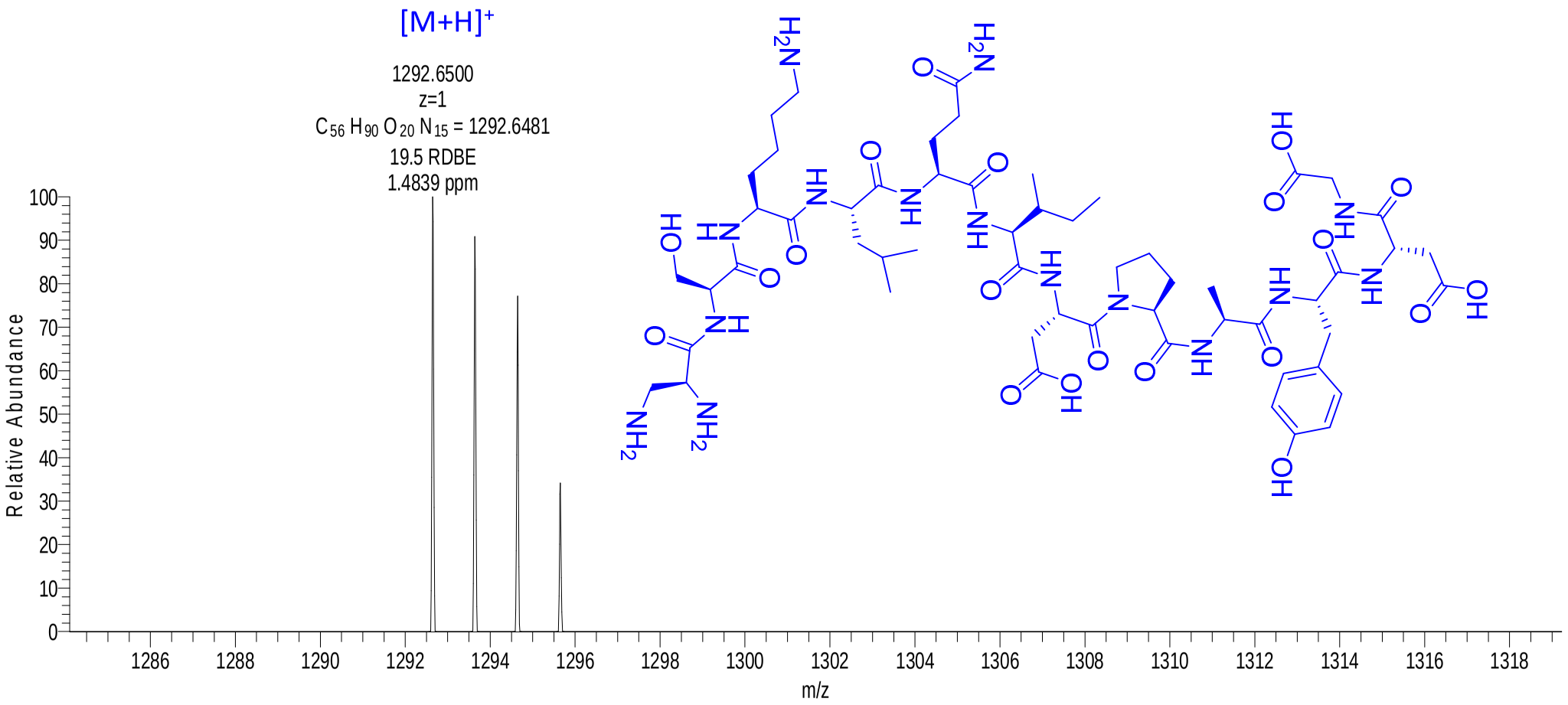


Fig. S9: ESI LC-MS for Peptide 8: Linear ZSKLQIDP - AYDG

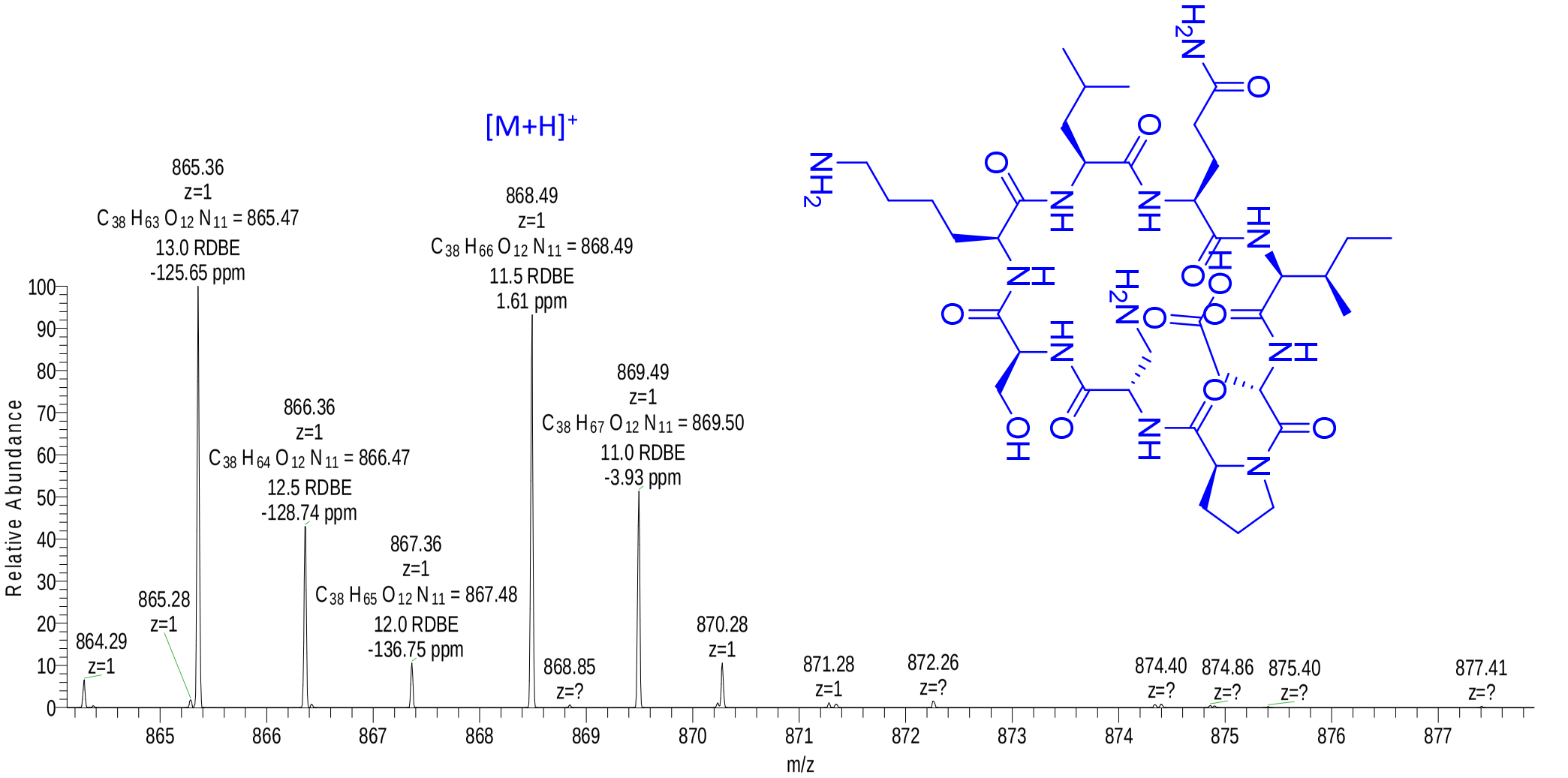


Fig. S10: ESI LC-MS for Peptide 8: Cyclo [ZSKLQIDP]

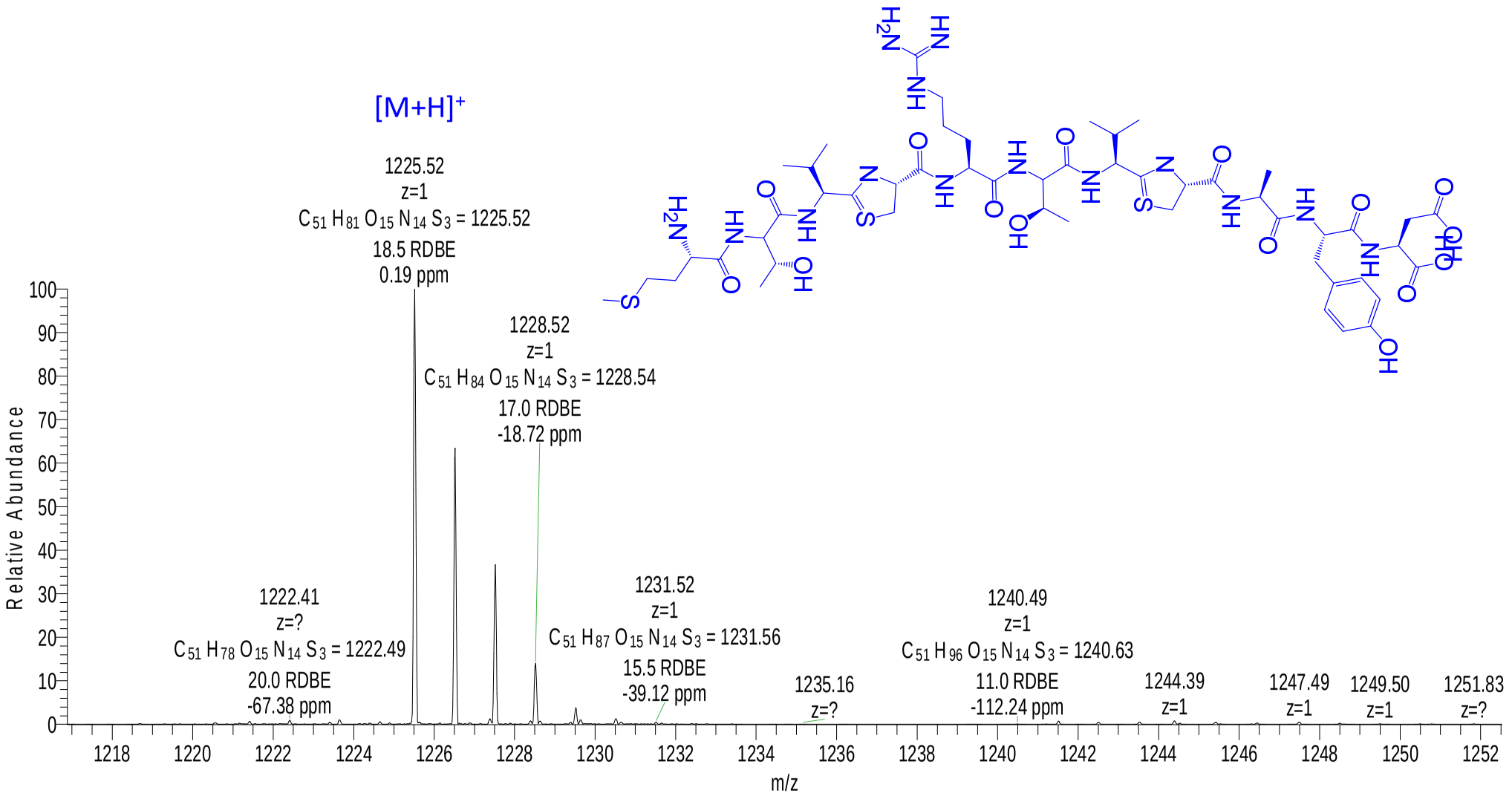


Fig. S11: ESI LC-MS for Peptide 19: Linear MTV(ThH)RTV(ThH)- AYDG

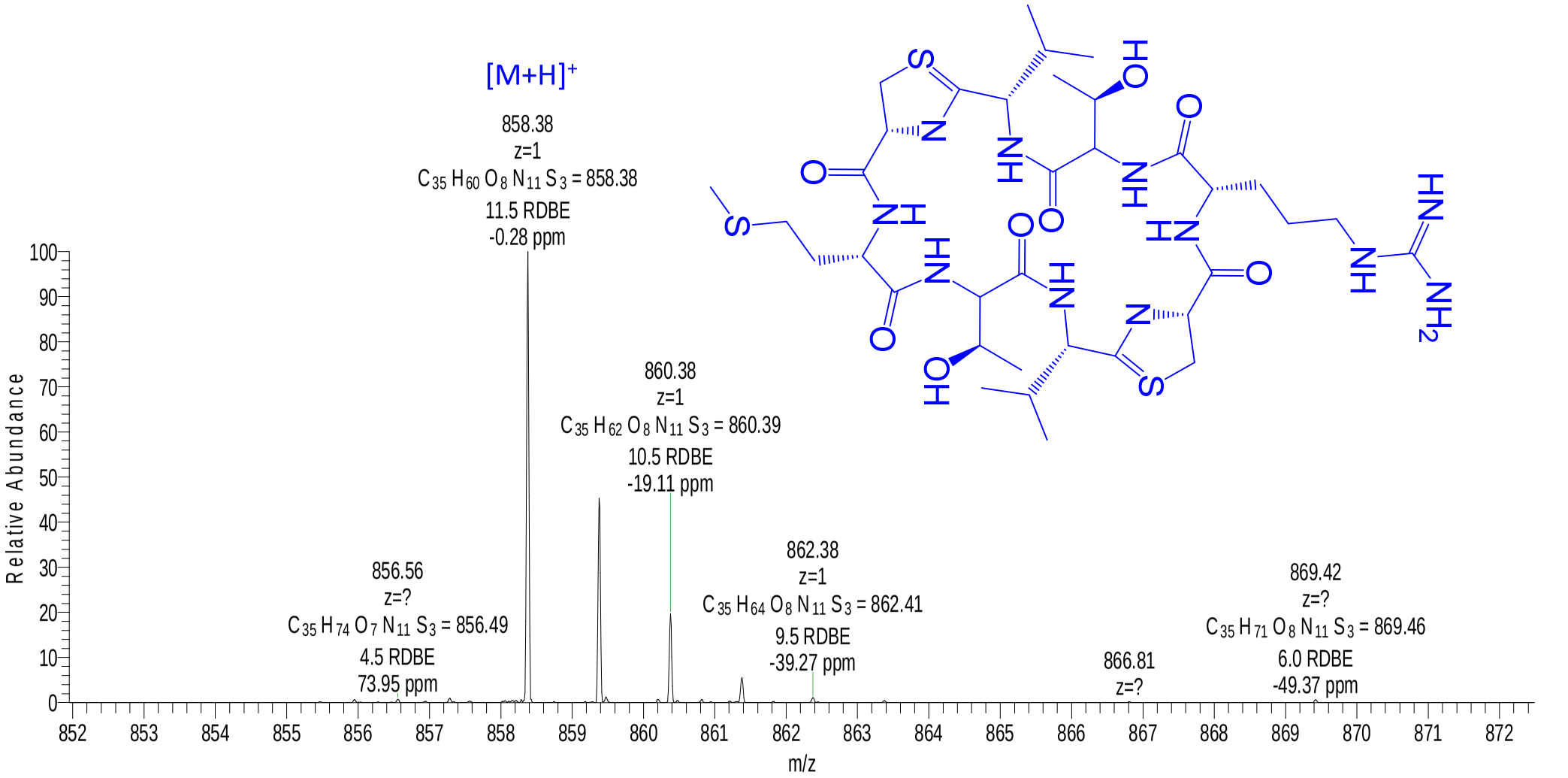


Fig. S12: ESI LC-MS for cyclic Peptide 19: Cyclo [MTV(ThH)RTV(ThH)]

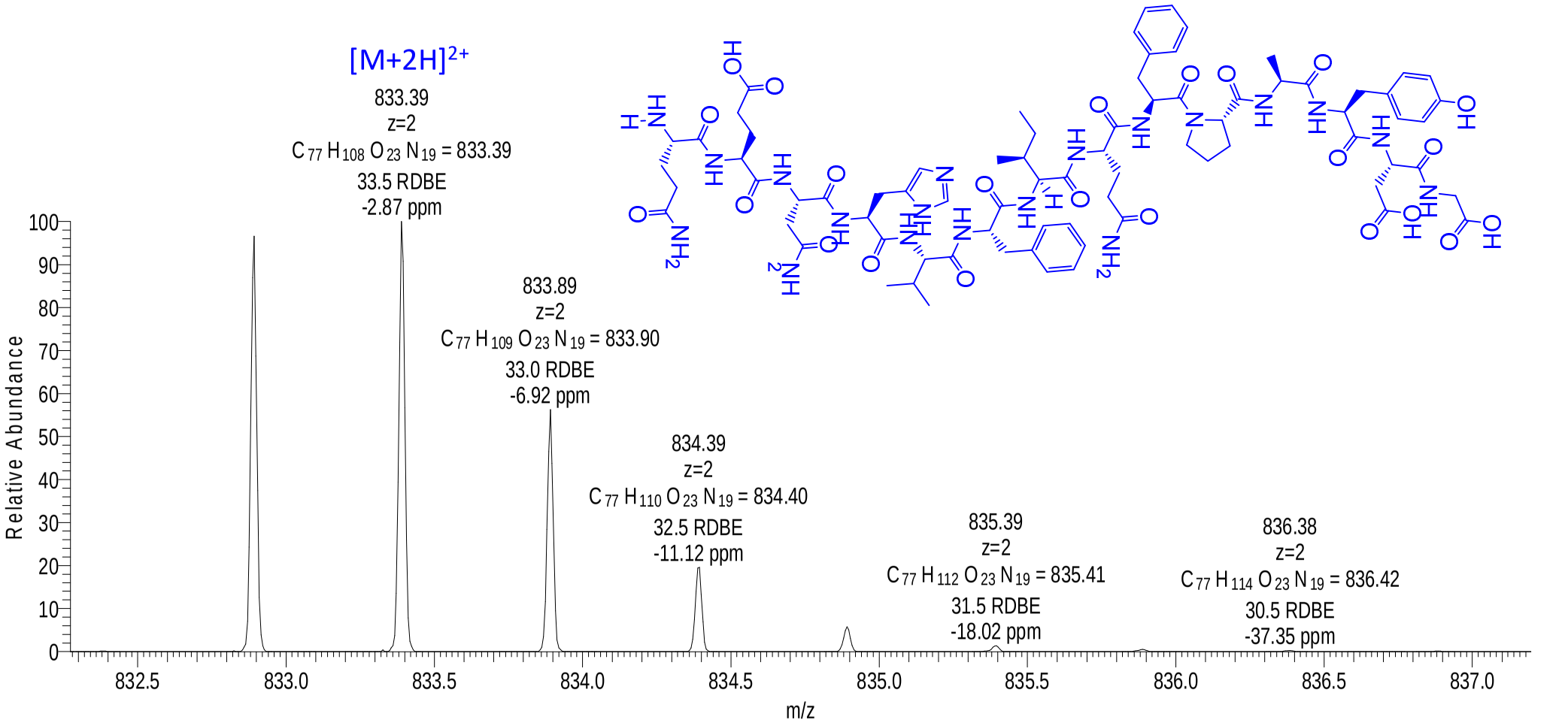


Fig. S13: ESI LC-MS for Peptide 3: Linear QENHVFIQFP - AYDG

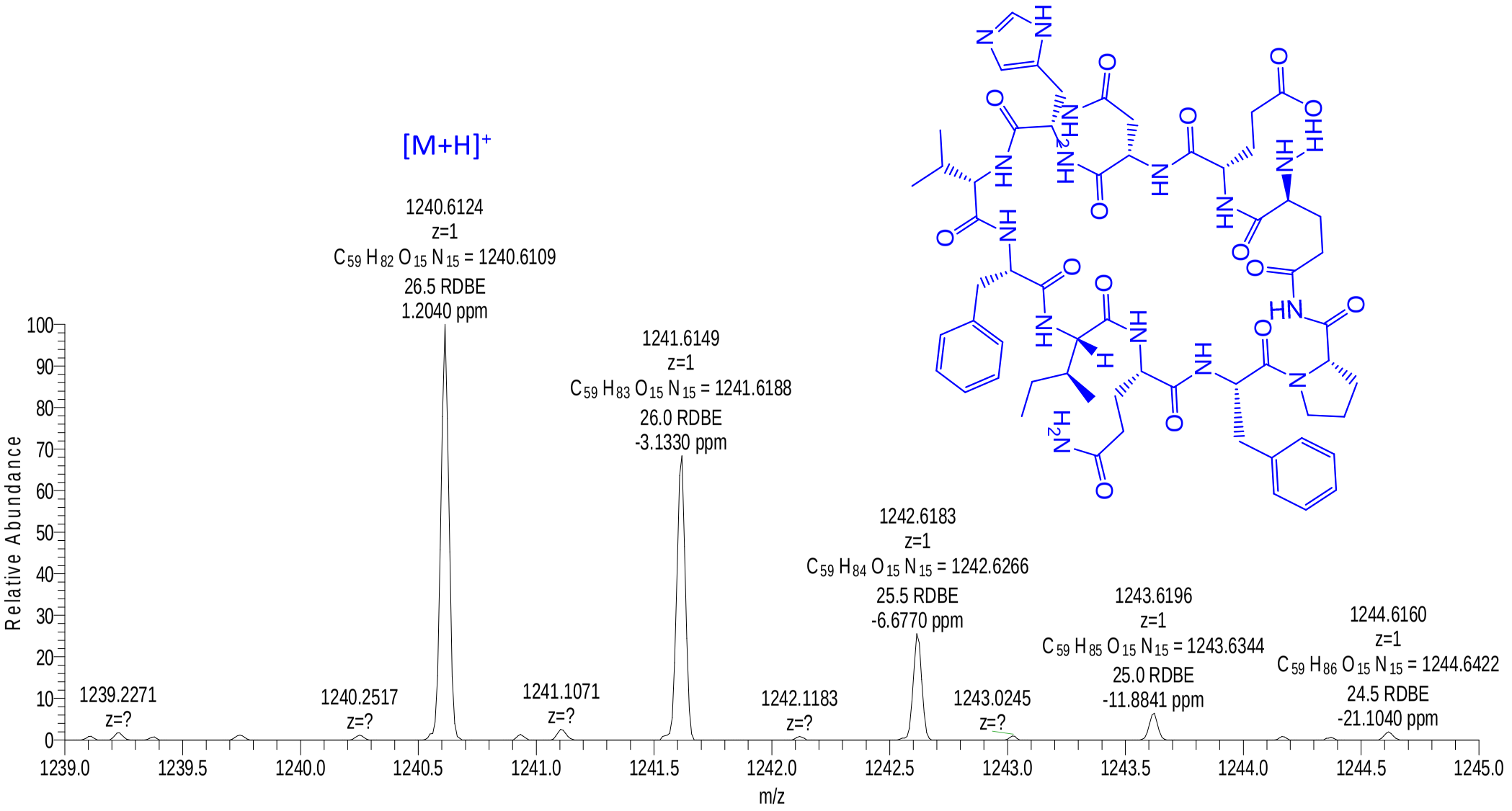


Fig. S14: ESI LC-MS for cyclic Peptide 3: Cyclo [QENHVFIQFP]

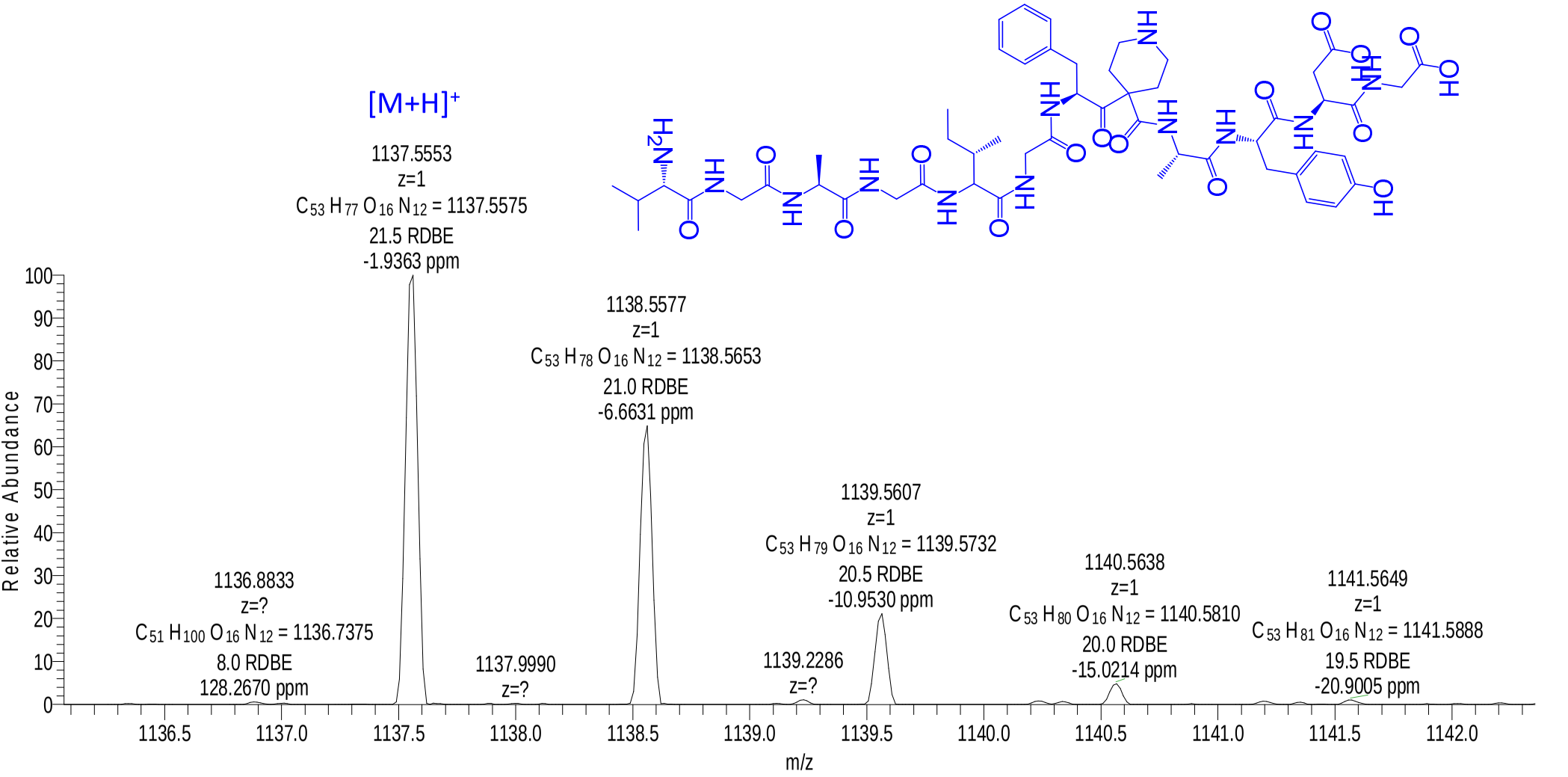


Fig. S15: ESI LC-MS for Peptide 5: Linear VGAGIGF(Pip) – AYDG

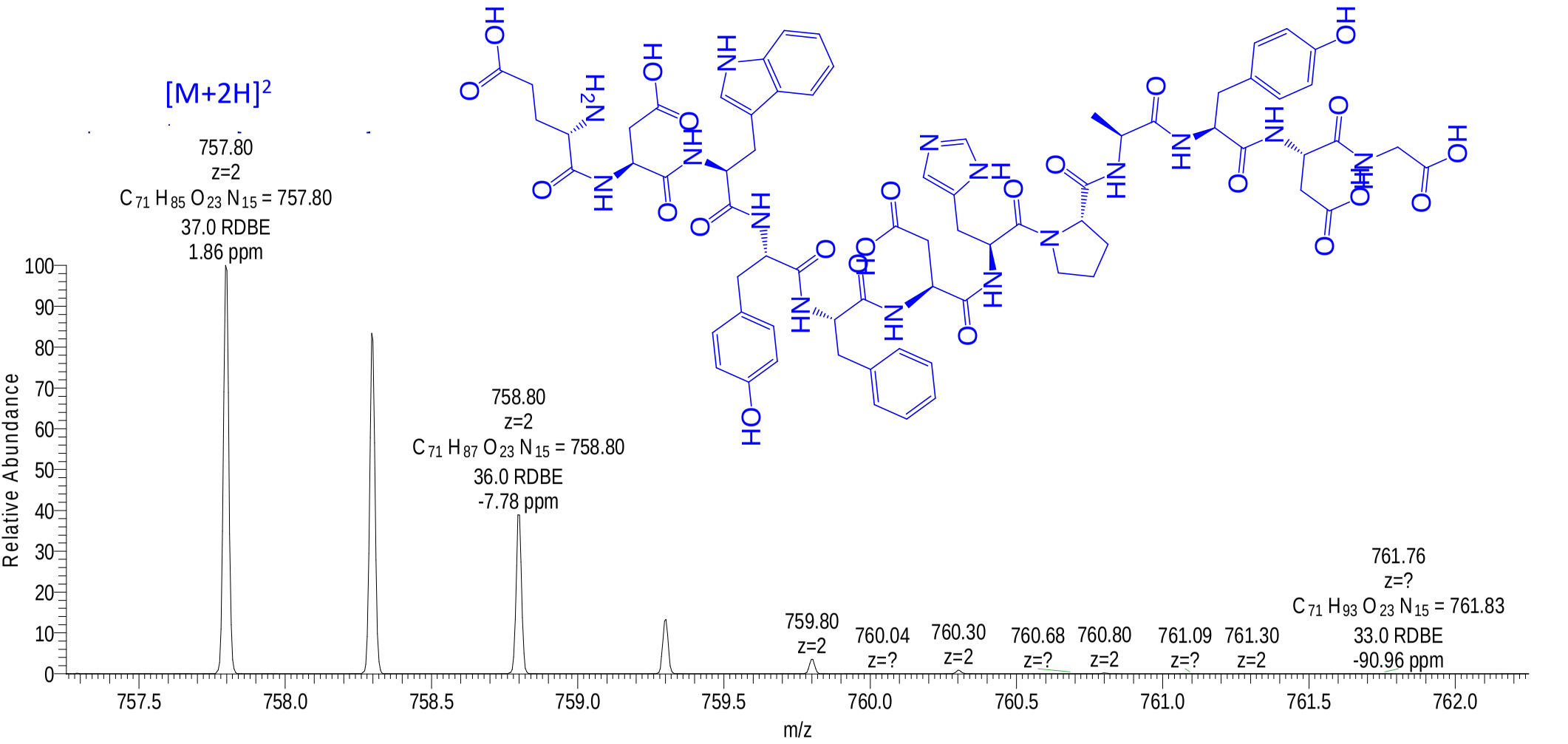


Fig. S16: ESI LC-MS for Peptide 4: Linear EDWYFDHP – AYDG

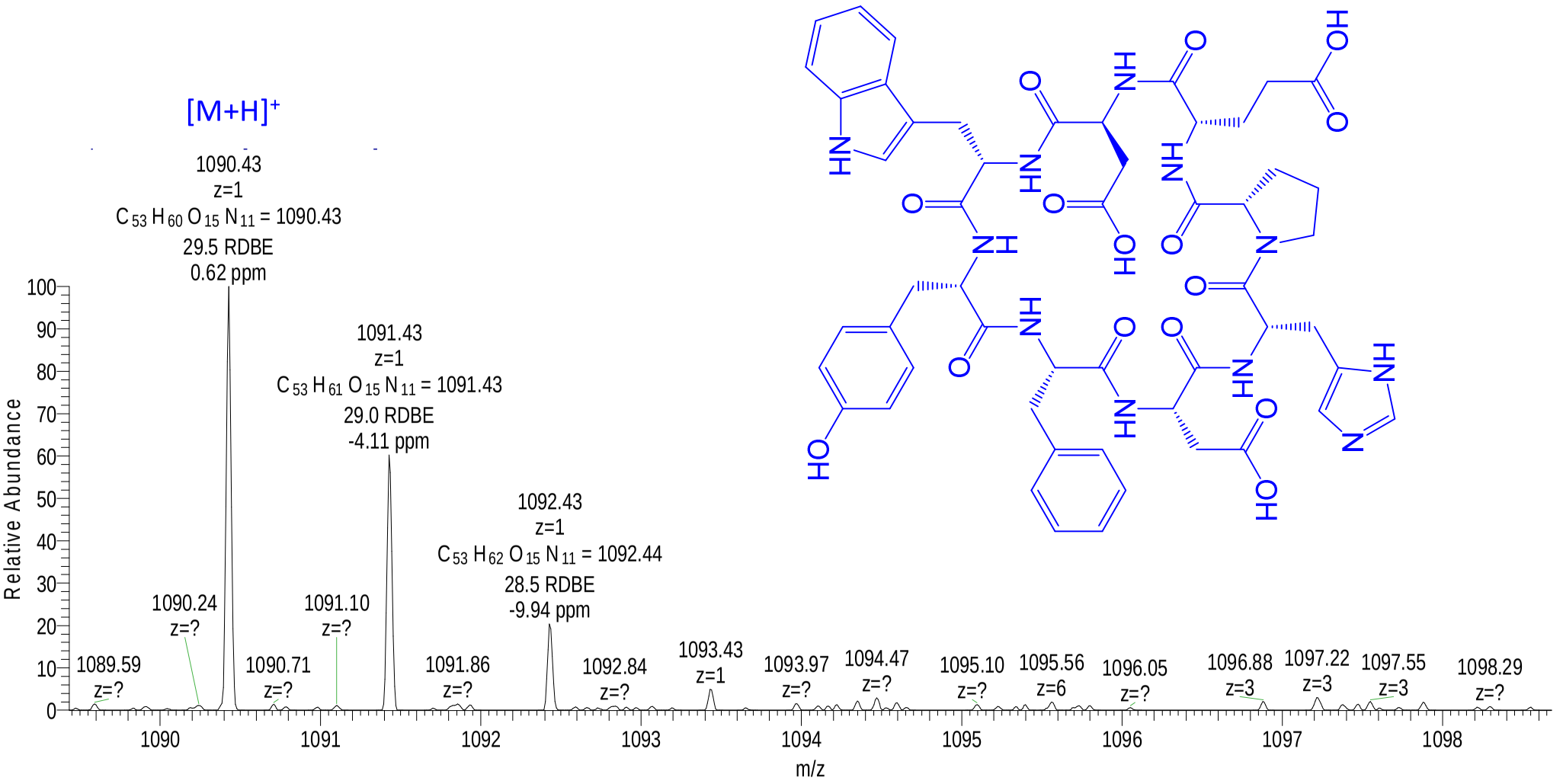


Fig. S17: ESI LC-MS for cyclic Peptide 4: Cyclo [EDWYFDHP]

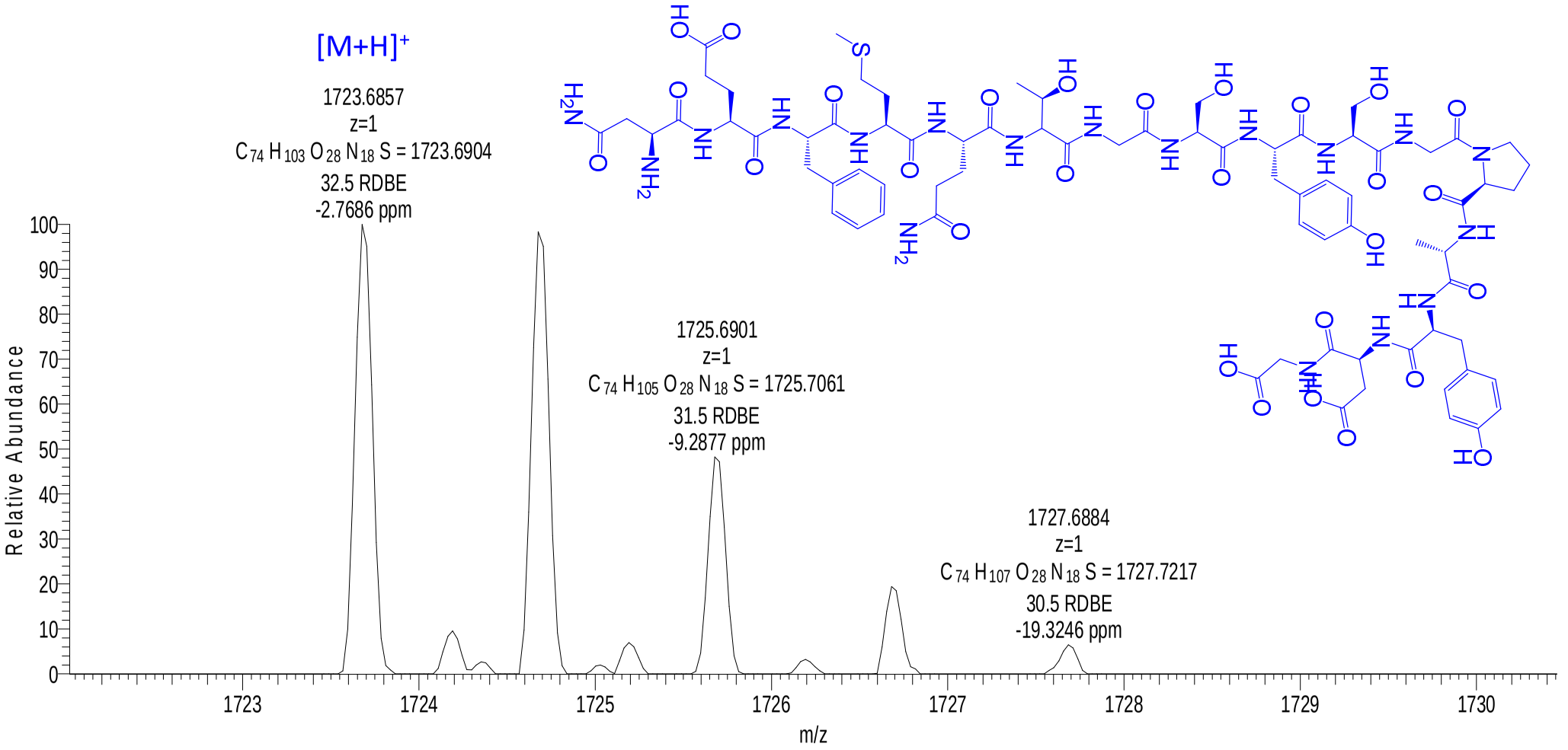


Fig. S18: ESI LC-MS for Peptide 22: Linear NEFMQTGSYSGP - AYDG

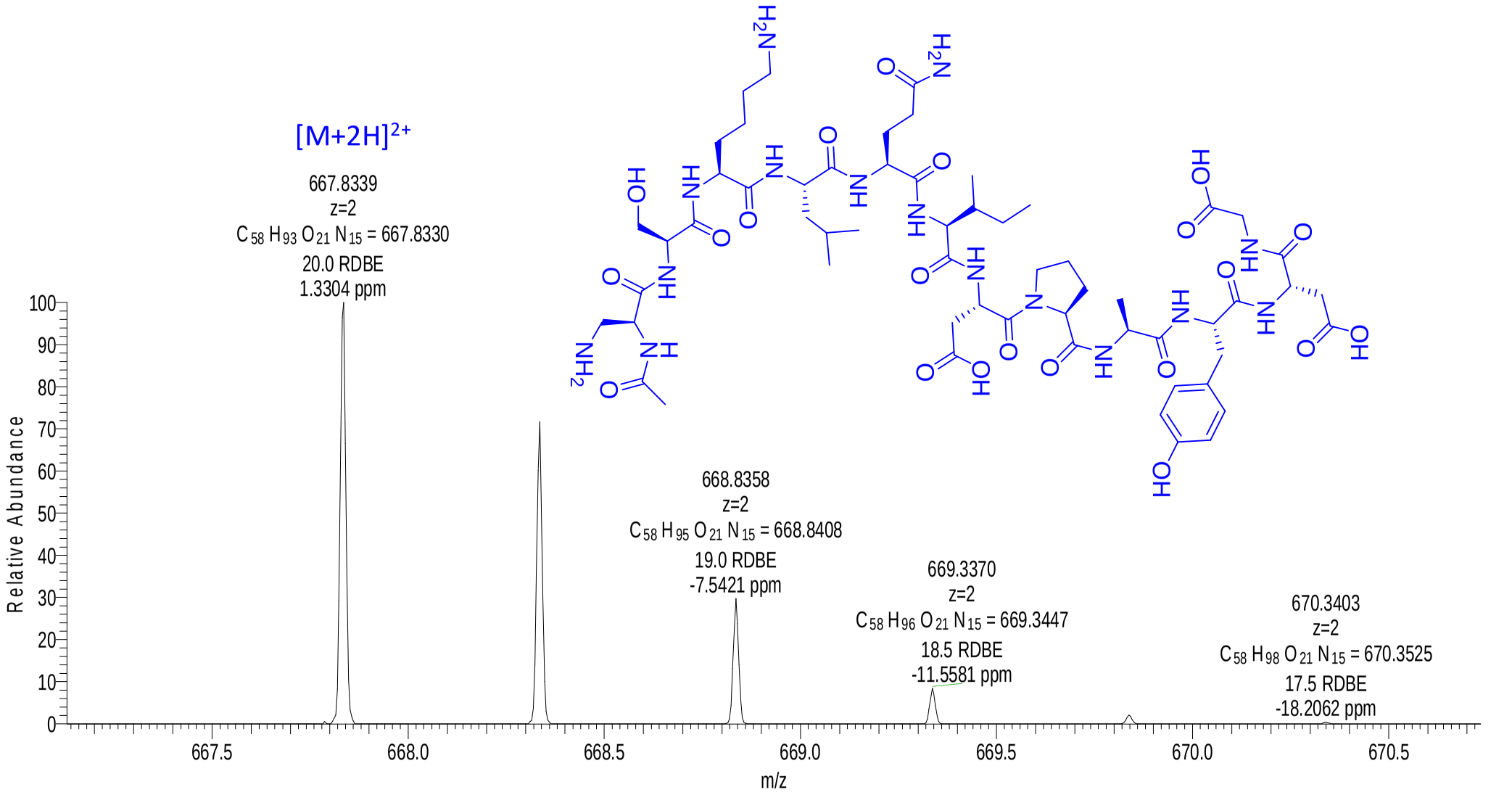


Fig. S19: ESI LC-MS for Peptide 14: Linear Z(Ac)SKLQIDP - AYDG

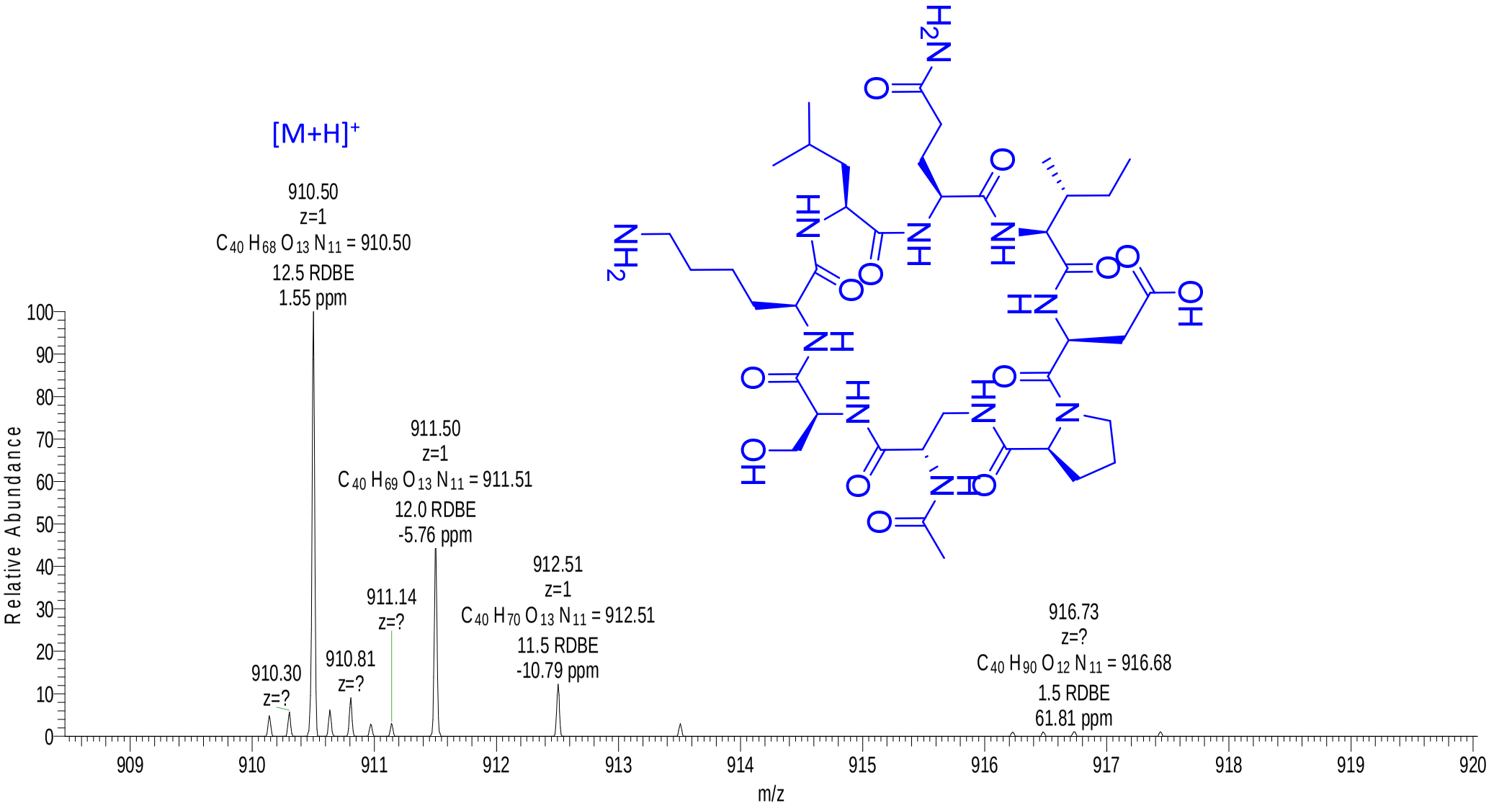


Fig. S20: ESI LC-MS for cyclic Peptide 14: Cyclo [Z(Ac)SKLQIDP]

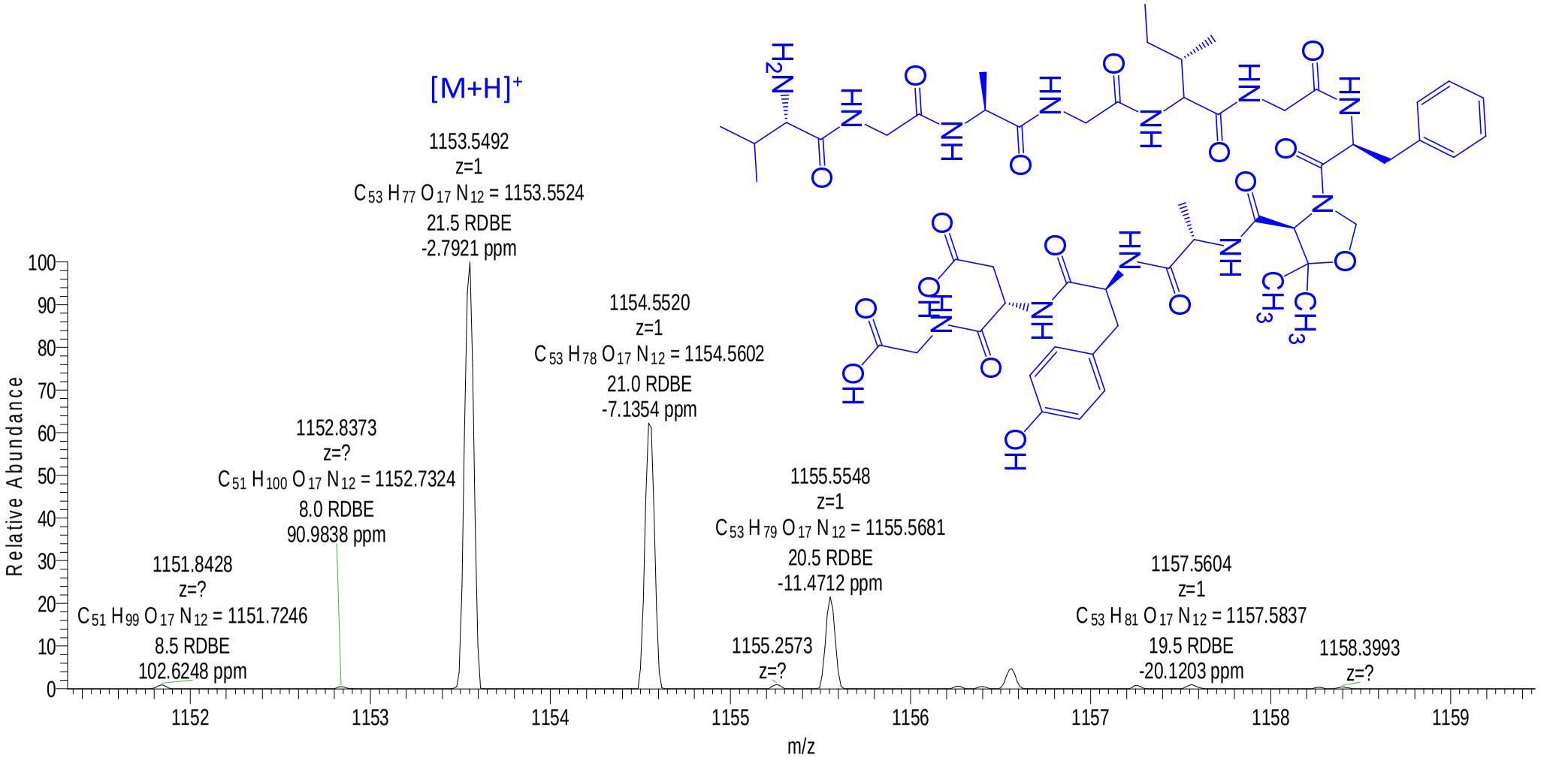


Fig. S21: ESI LC-MS for Peptide 6: Linear VGAGIGF(ΨP) - AYDG

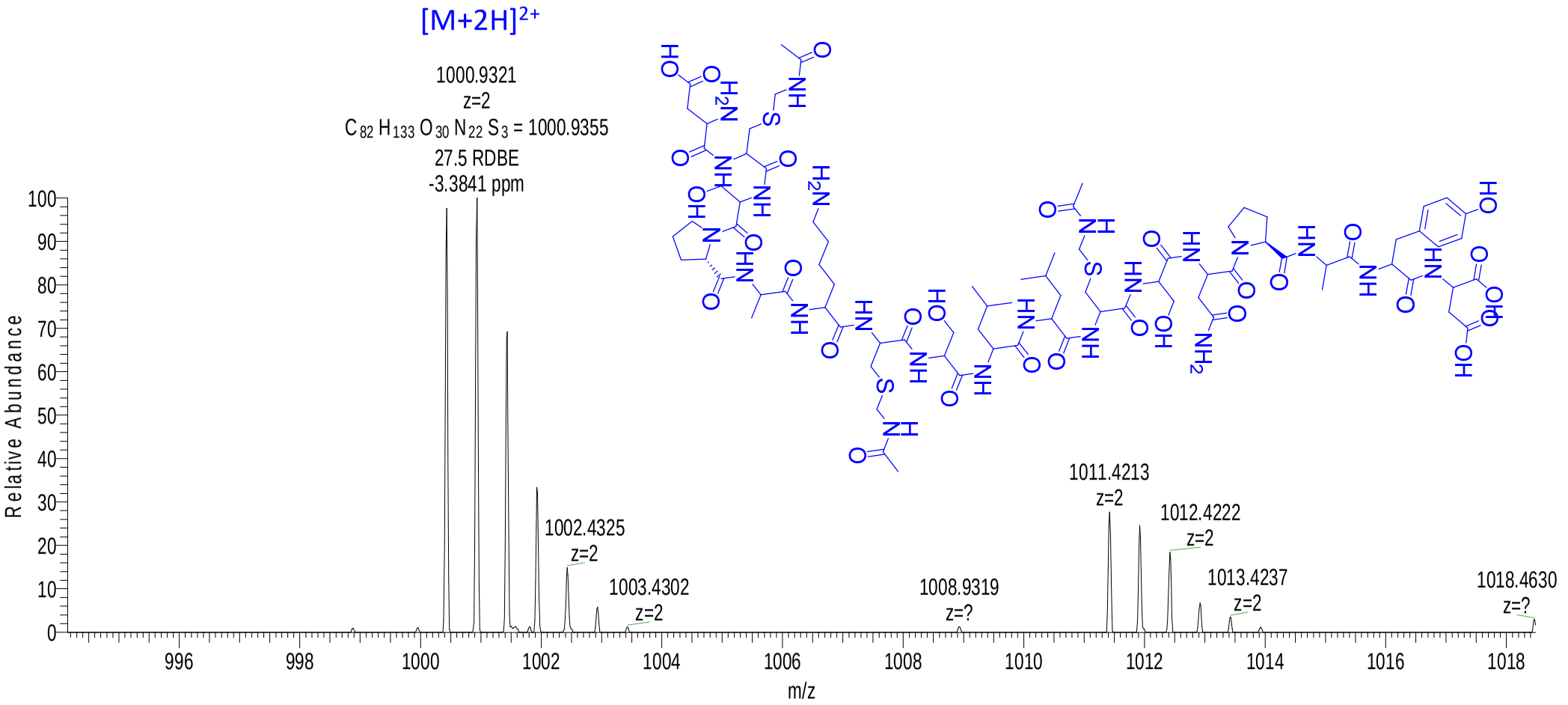


Fig. S22: ESI LC-MS for Peptide 23: Linear DCSPAKCSLLCSNP - AYDG

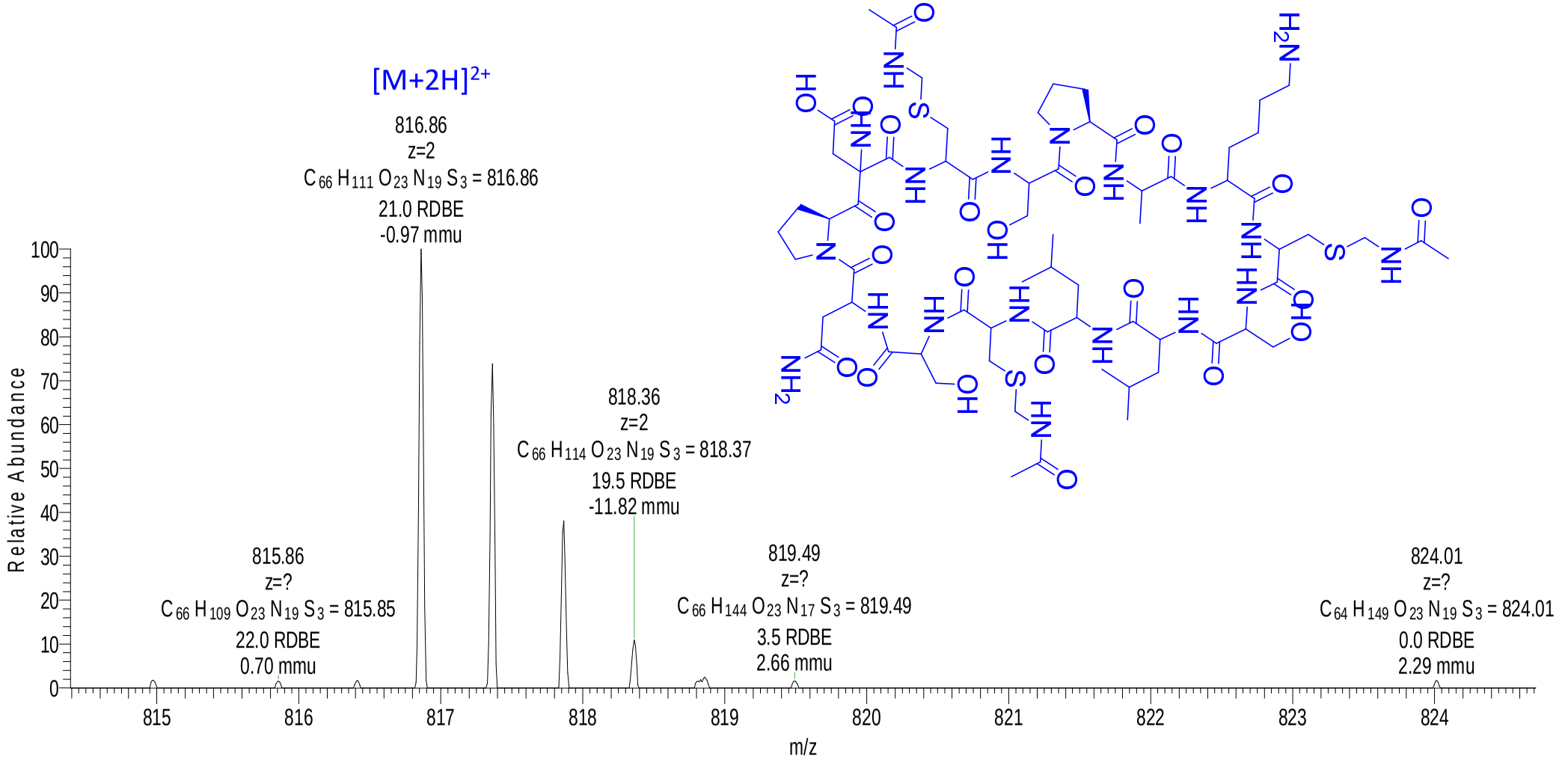


Fig. S23: ESI LC-MS for cyclic Peptide 23: Cyclo [DCSPAKCSLLCSNP]

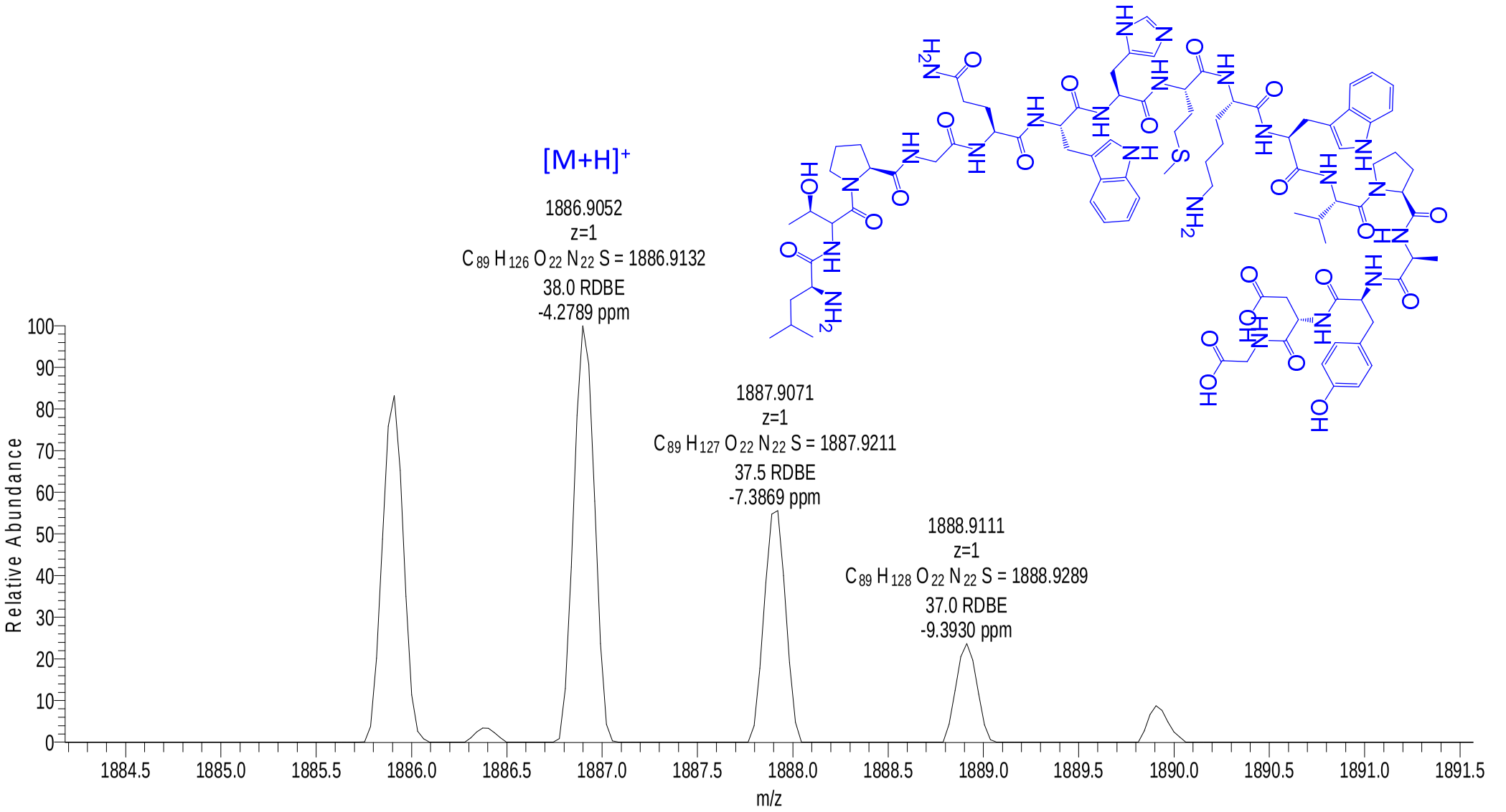


Fig. S24: ESI LC-MS for Peptide 24: Linear LTPGQWHMKWVP - AYDG

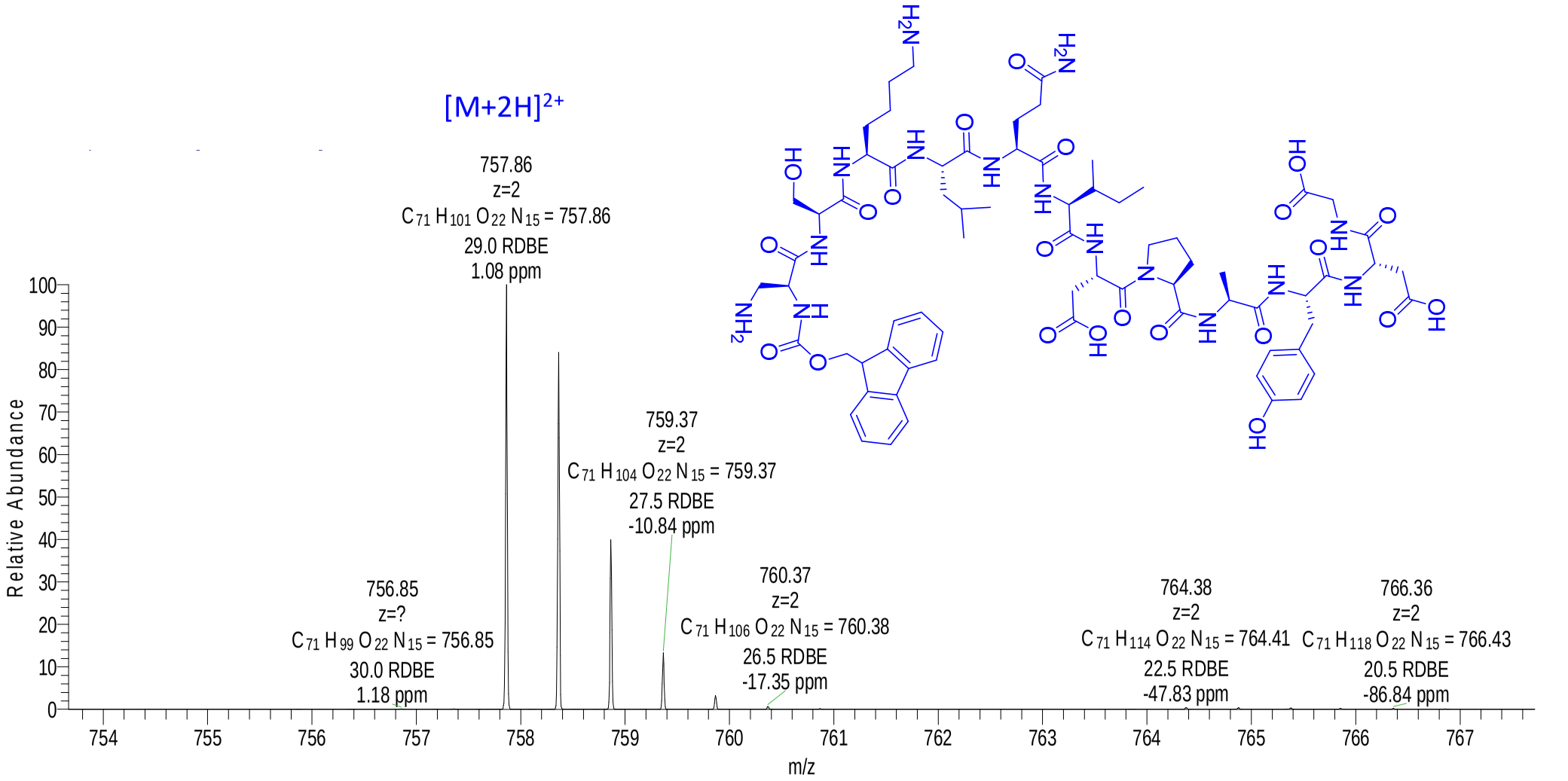


Fig. S25: ESI LC-MS for Peptide 10: Linear Z(Fmoc)SKLQIDP - AYDG

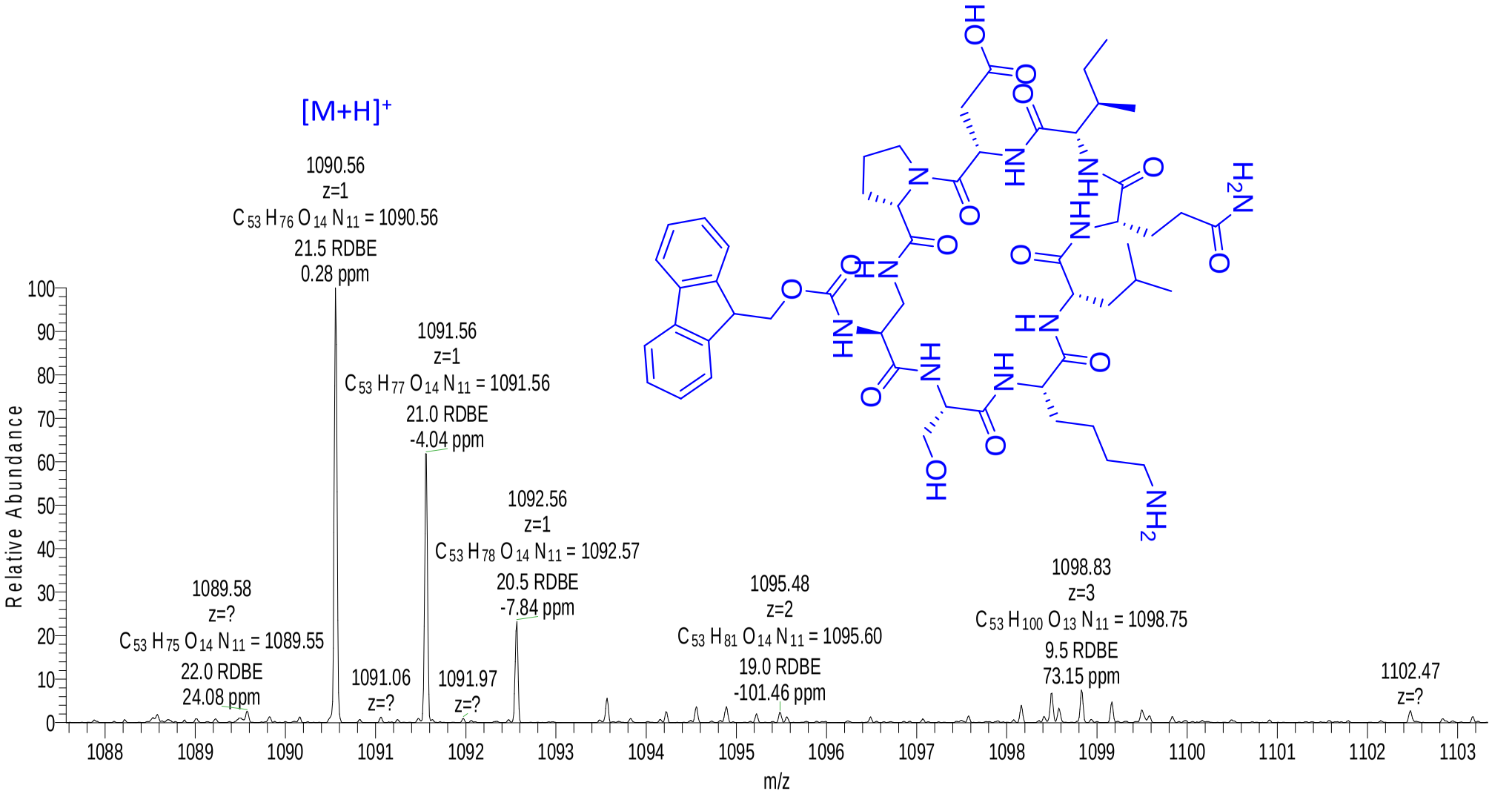


Fig. S26: ESI LC-MS for cyclic Peptide 10: Cyclo [Z(Fmoc)SKLQIDP]

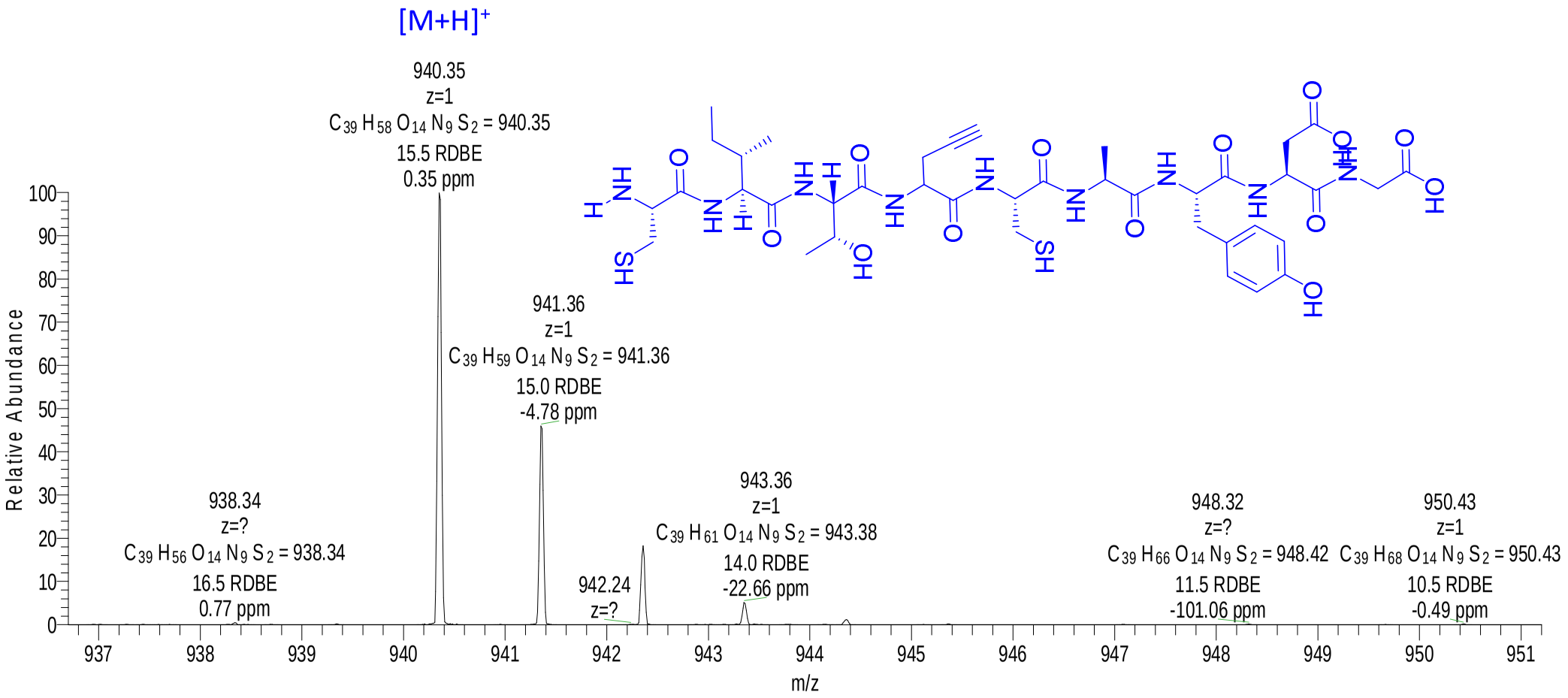


Fig. S27: ESI LC-MS for Peptide 1: Linear CITJC - AYDG

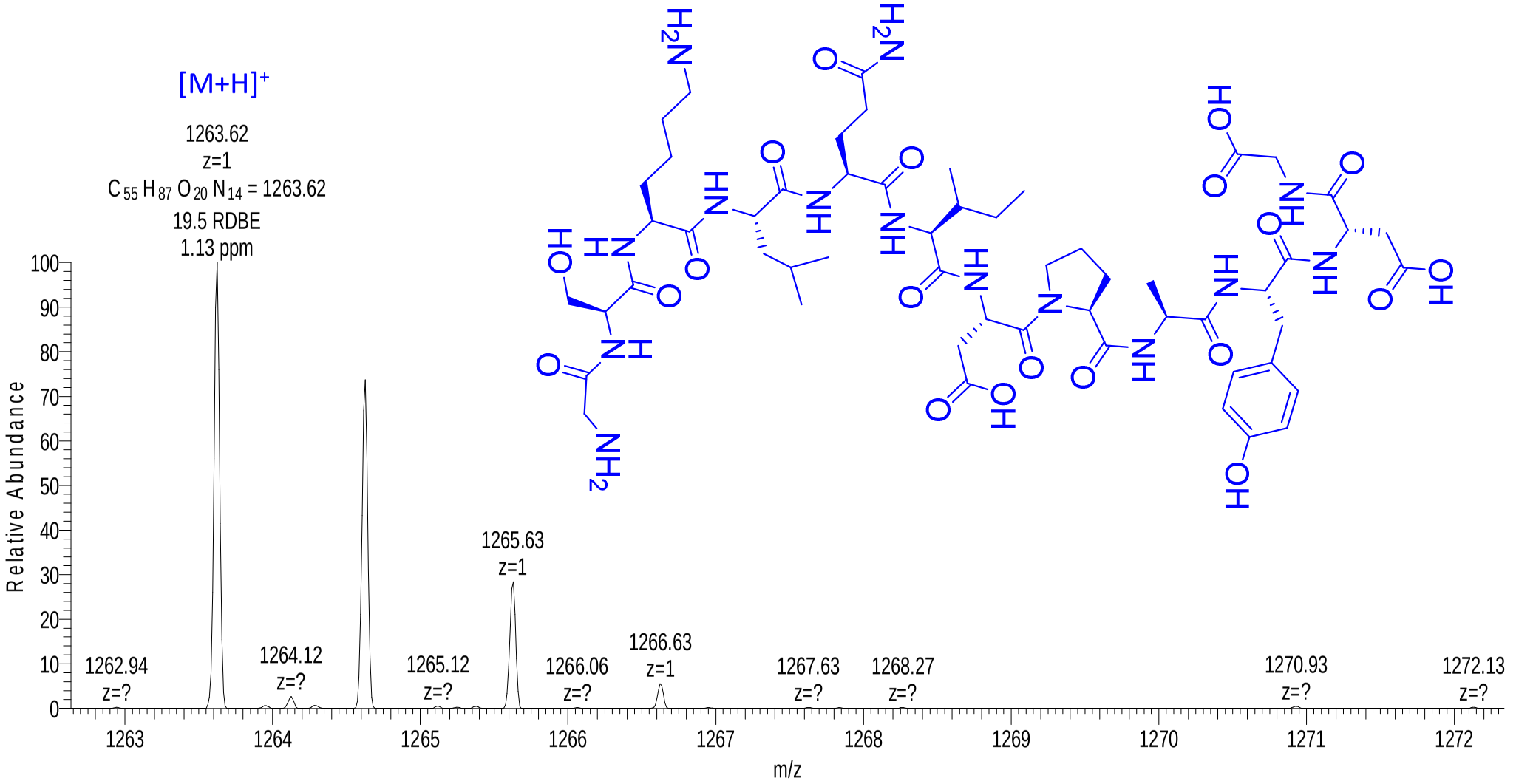


Fig. S28: ESI LC-MS for Peptide 7: Linear GSKLQIDP - AYDG

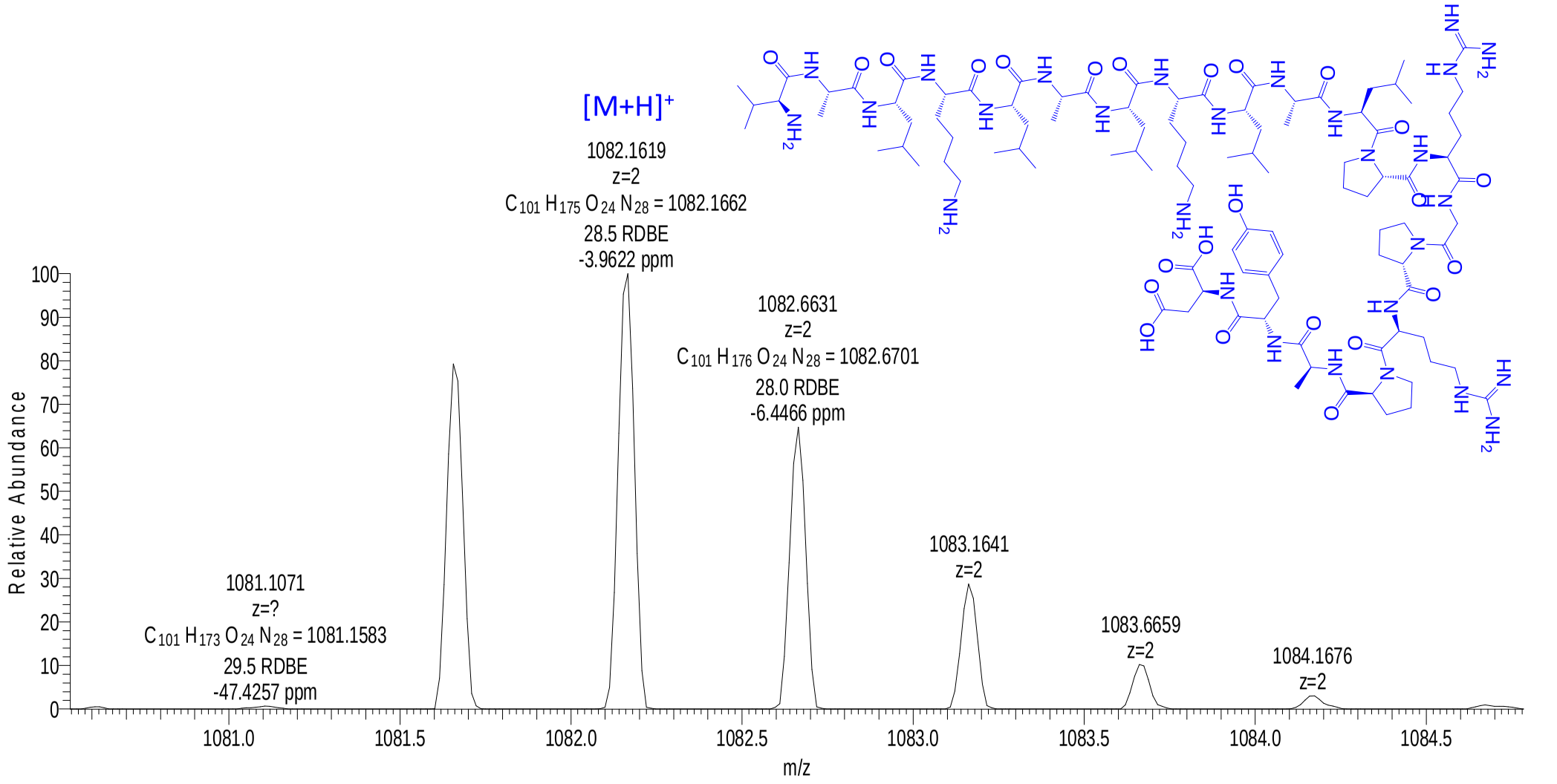


Fig. S29: ESI LC-MS for Peptide 25: Linear VALKLALKLALPRGPRP – AYDG

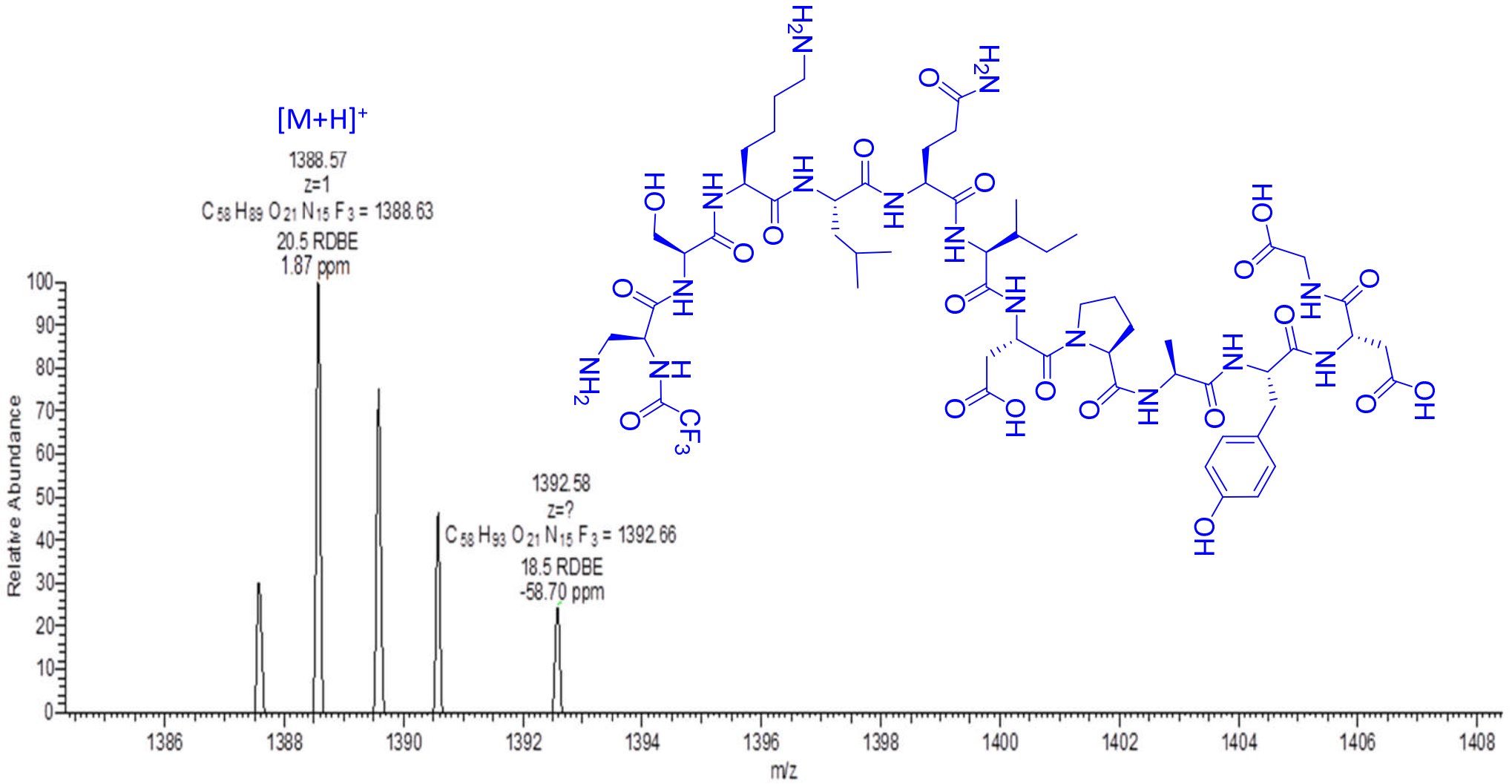


Fig. S30: ESI LC-MS for Peptide 12: Linear Z(TFA)SKLQIDP - AYDG

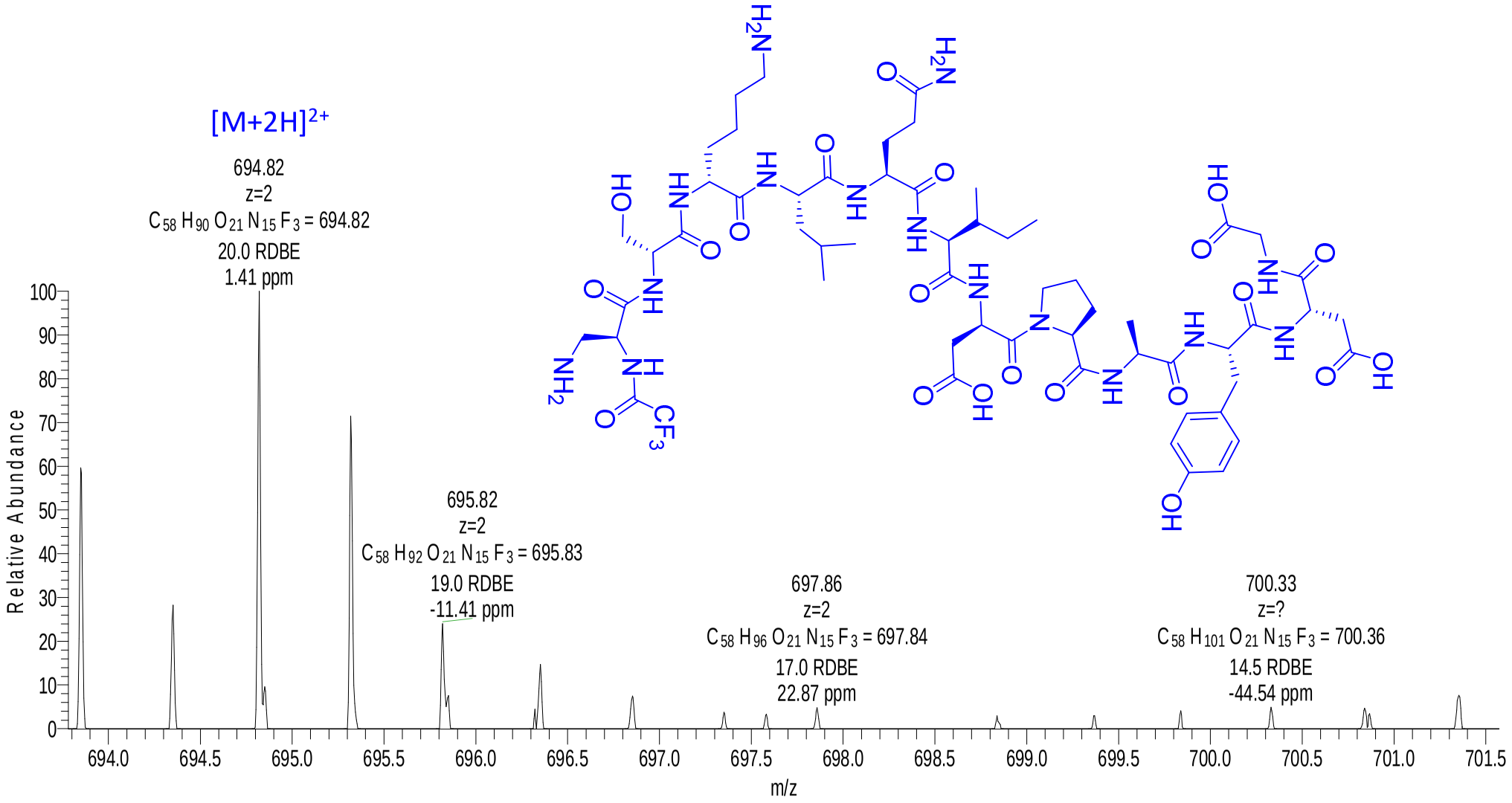


Fig. S31: ESI LC-MS for Peptide 13: Linear Z(TFA)SKLQIDP - AYDG

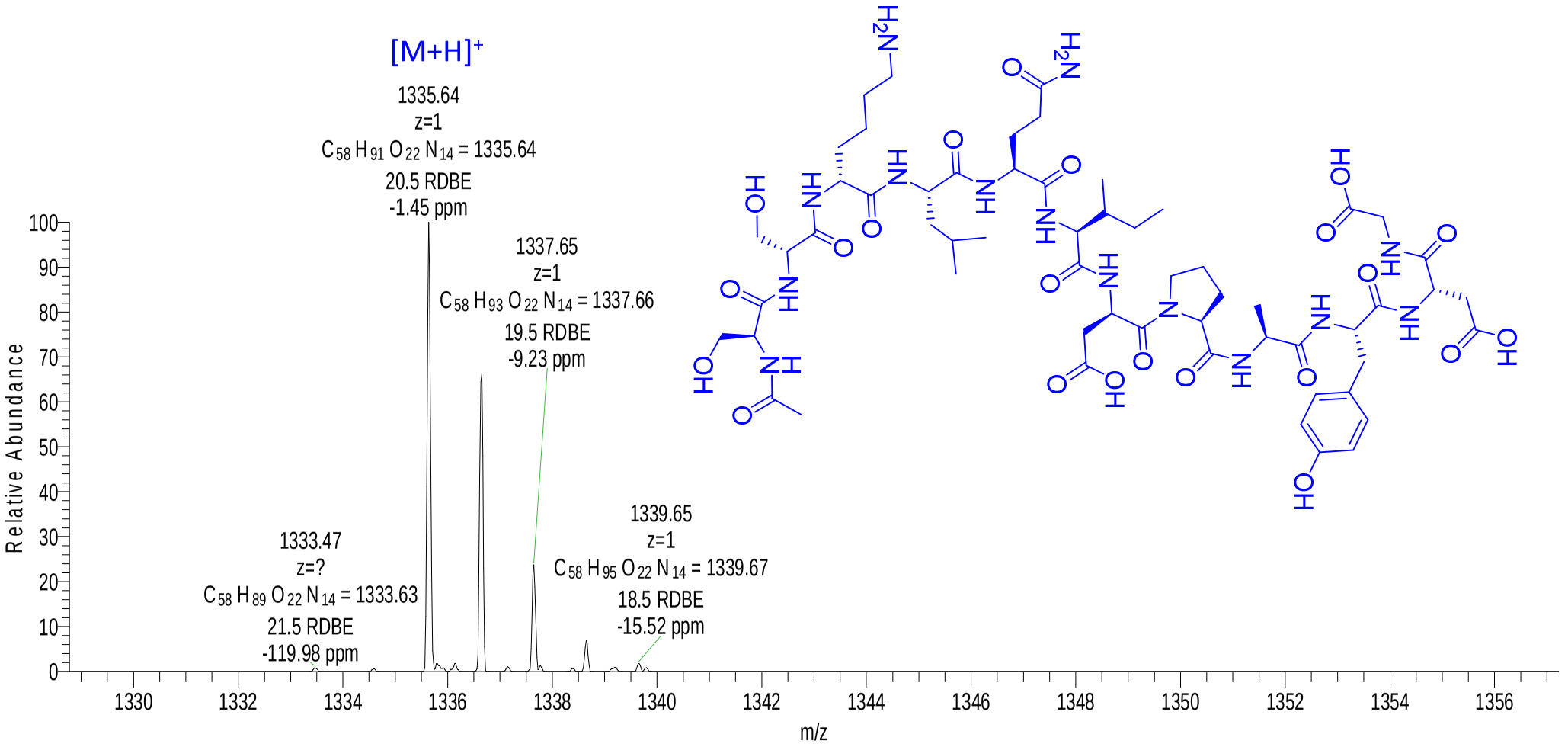


Fig. S32: ESI LC-MS for Peptide 16: Linear S(Ac)SKLQIDP - AYDG

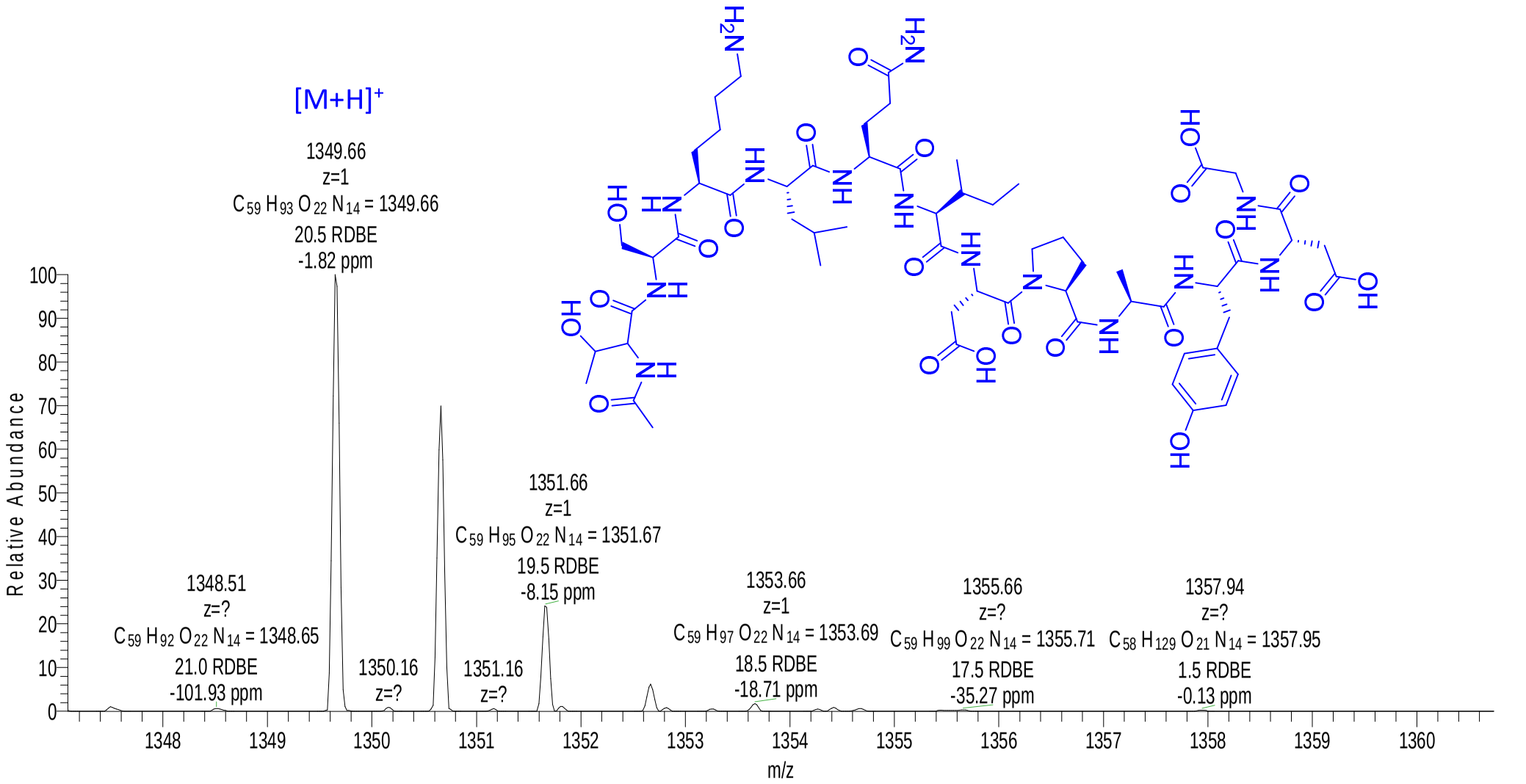


Fig. S33: ESI LC-MS for Peptide 17: Linear T(Ac)SKLQIDP - AYDG

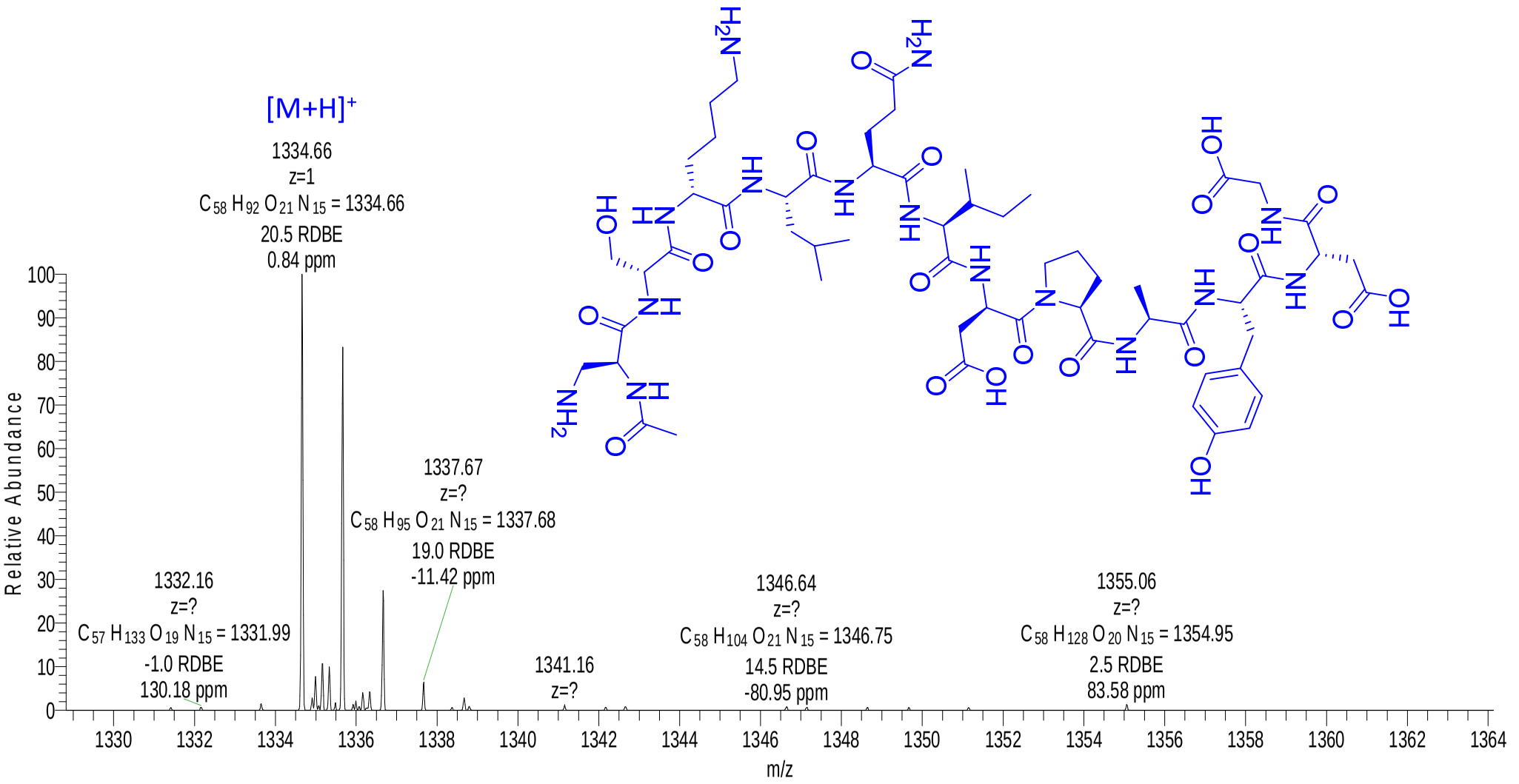


Fig. S34: ESI LC-MS for Peptide 15: Linear Z(Ac)SKLQIDP - AYDG

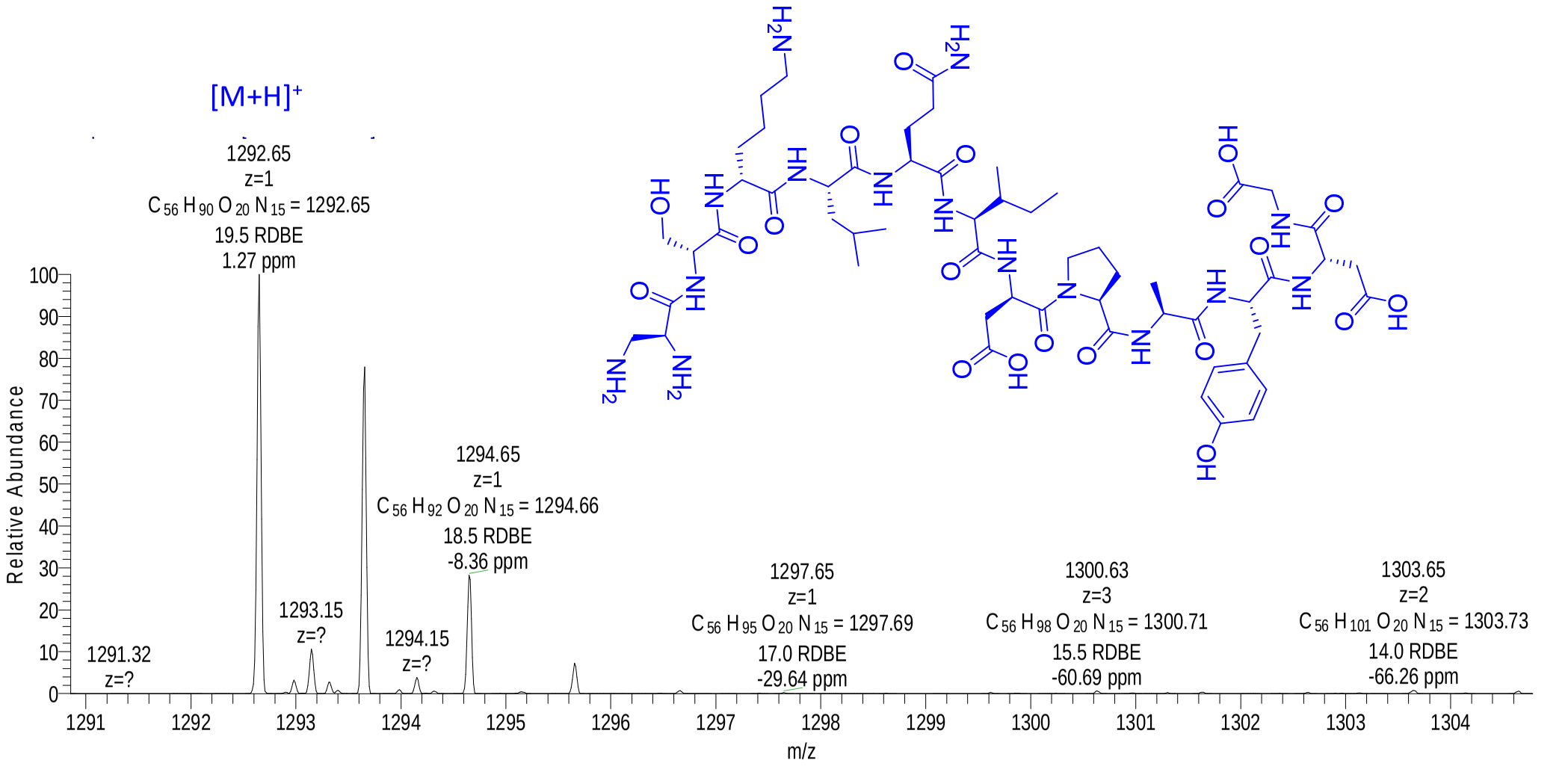


Fig. S35: ESI LC-MS for Peptide 9: Linear ZSKLQIDP - AYDG

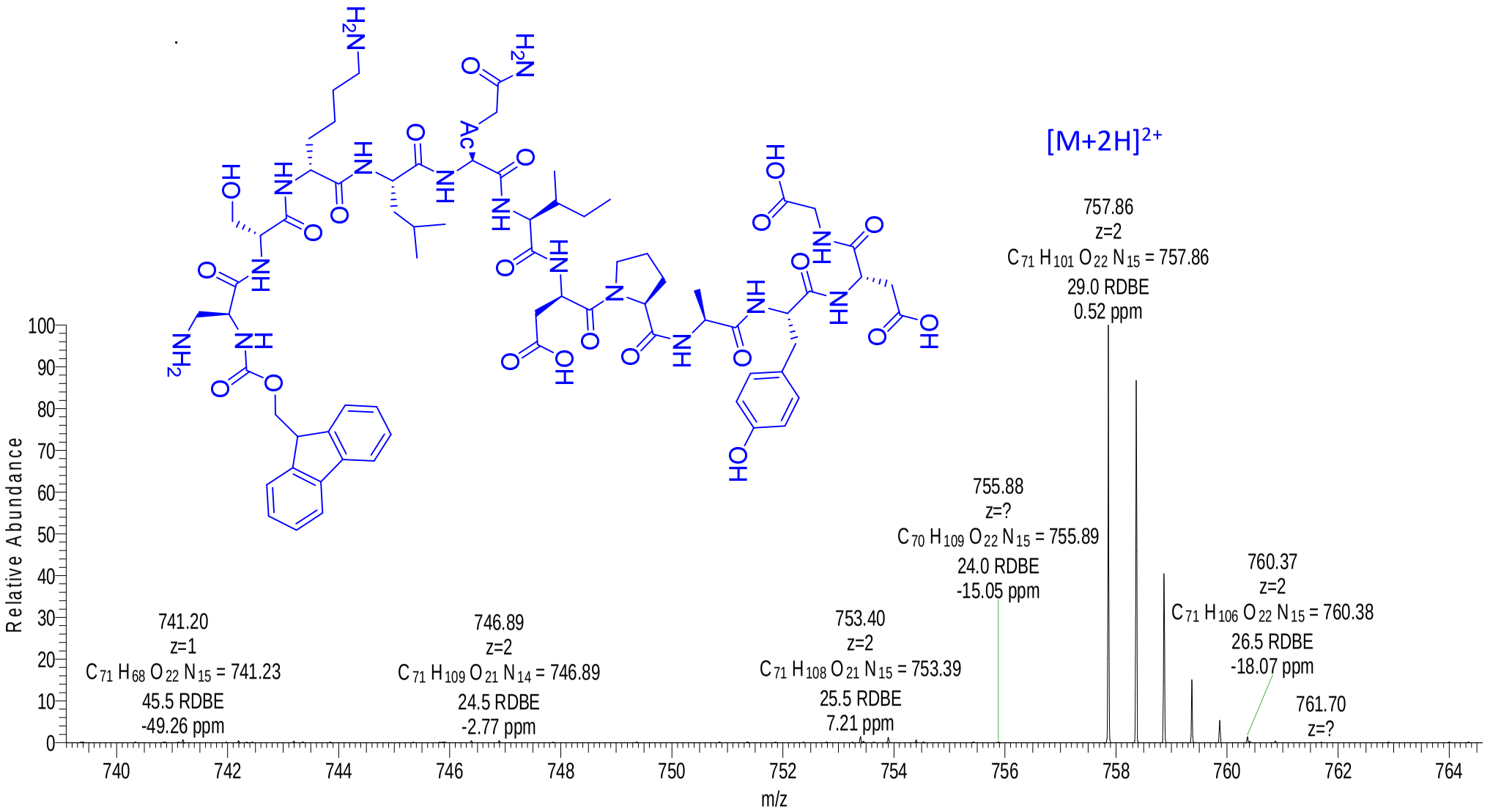
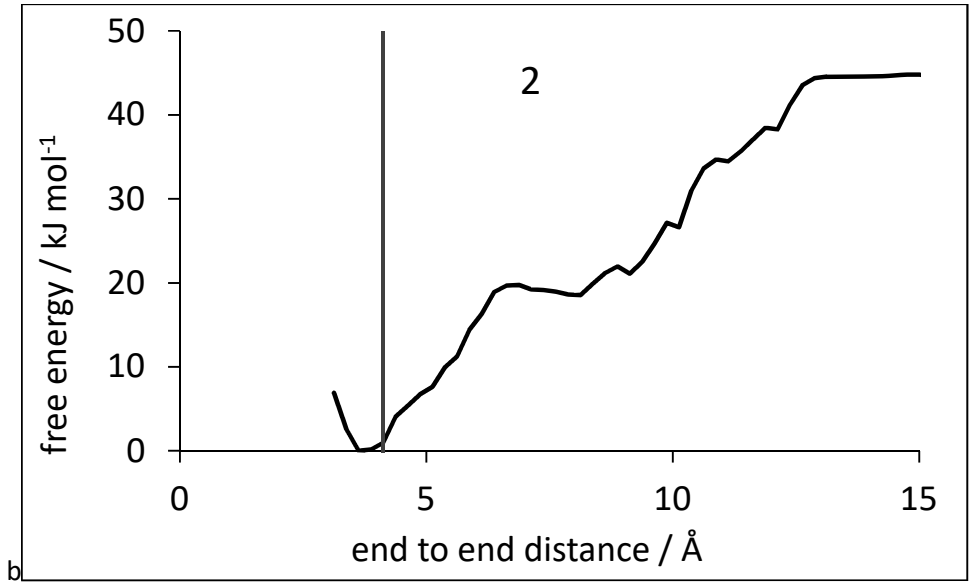


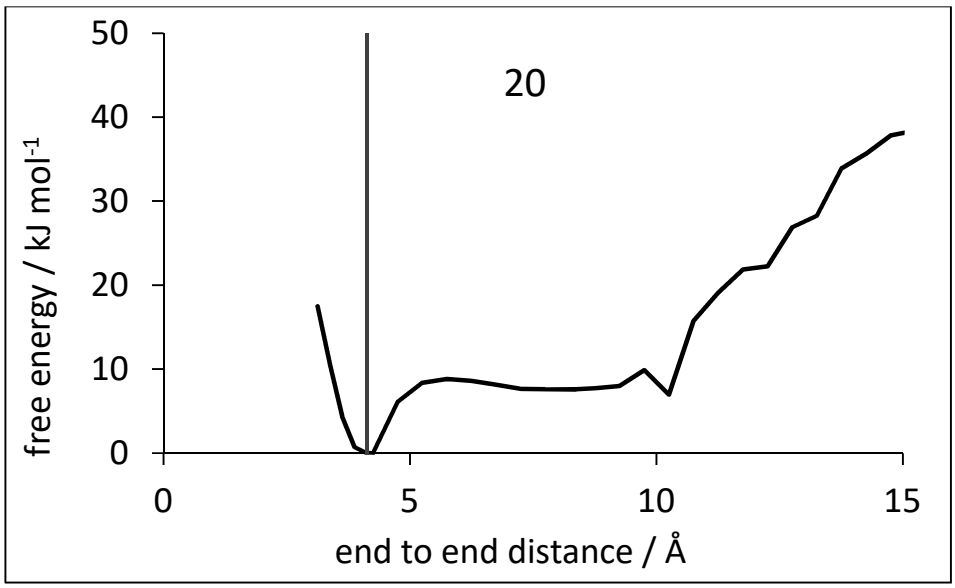
Fig. S36: ESI LC-MS for Peptide 11: Linear Z(Fmoc)SKLQIDP - AYDG

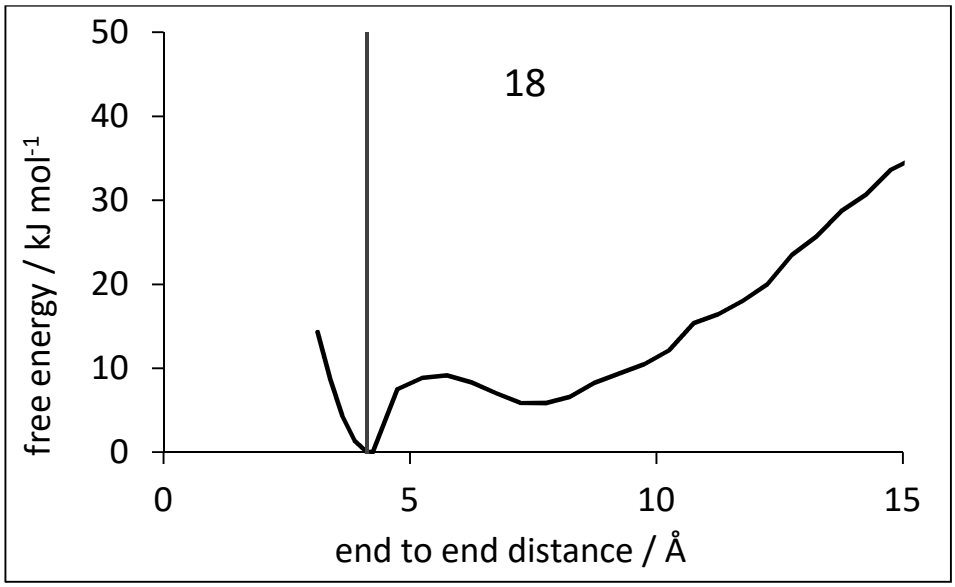
Computational Data. Calculated Free Energies as a Function the C-N Distance.

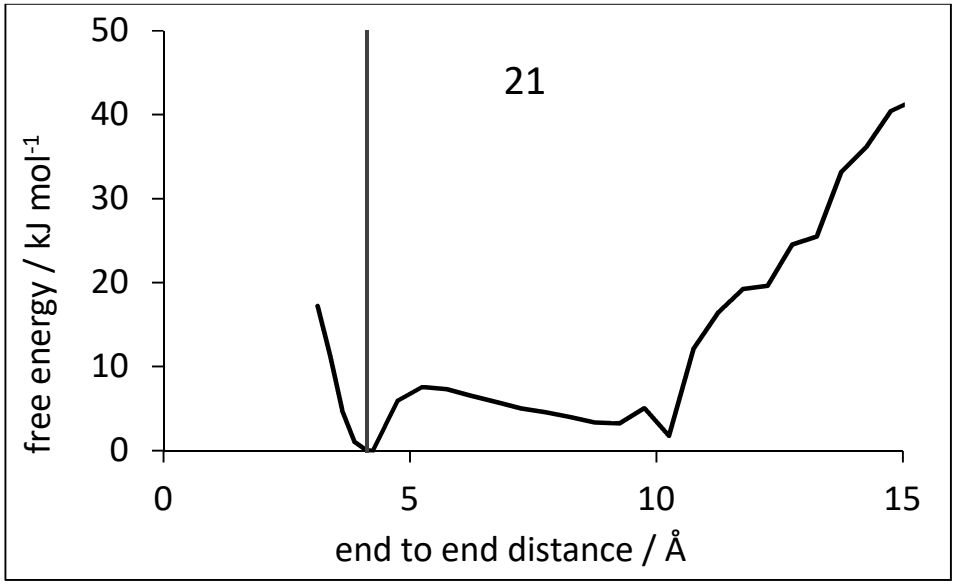
Fig. S37. The 25 frames below show the cyclisation free energies vs C-N distance for each of the 25 peptides arranged in the order of descending P(PCC), the probability of PCC formation, same as in the Table 1. The PCC end-to-end distance, below which the peptide is said to be in the PCC, is shown by the vertical line at 4 Å.

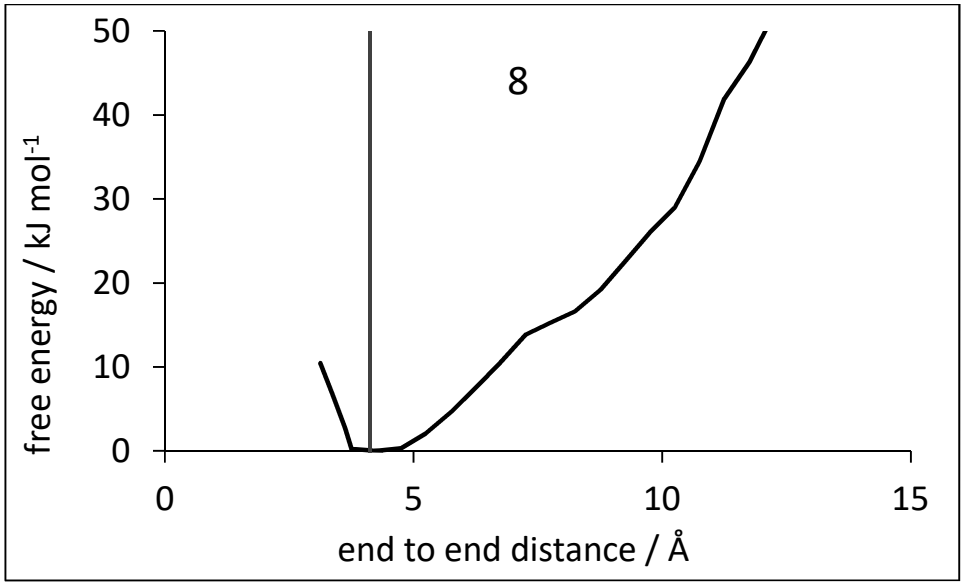


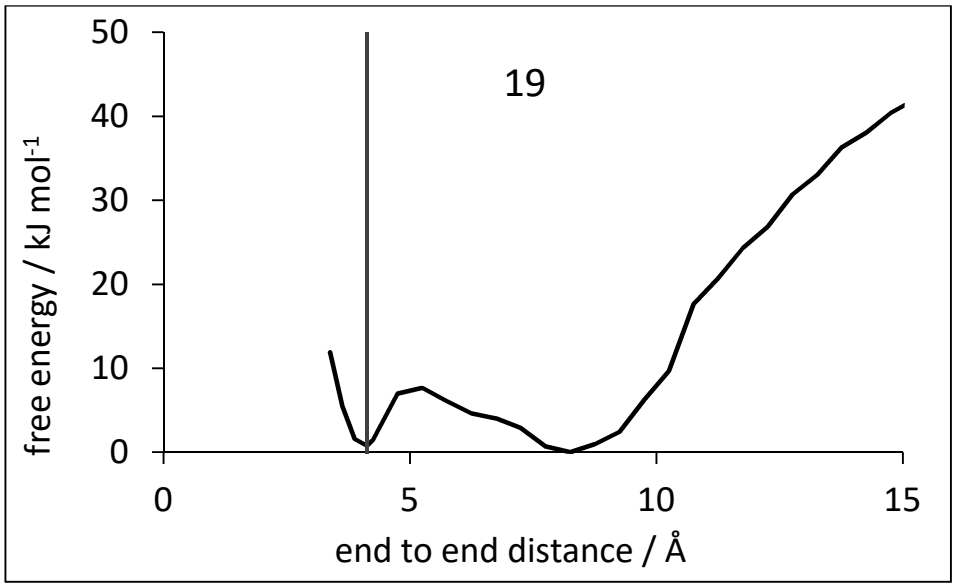
b

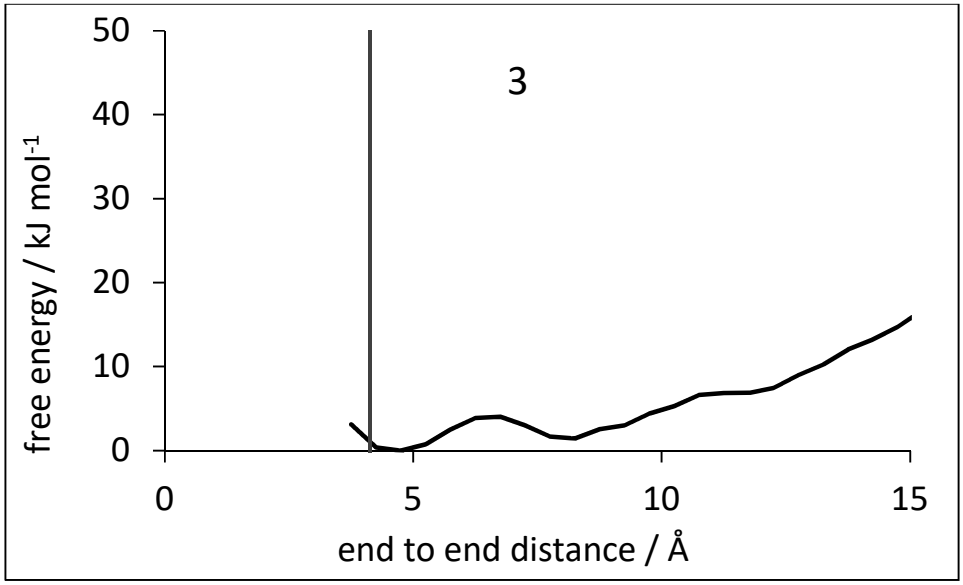


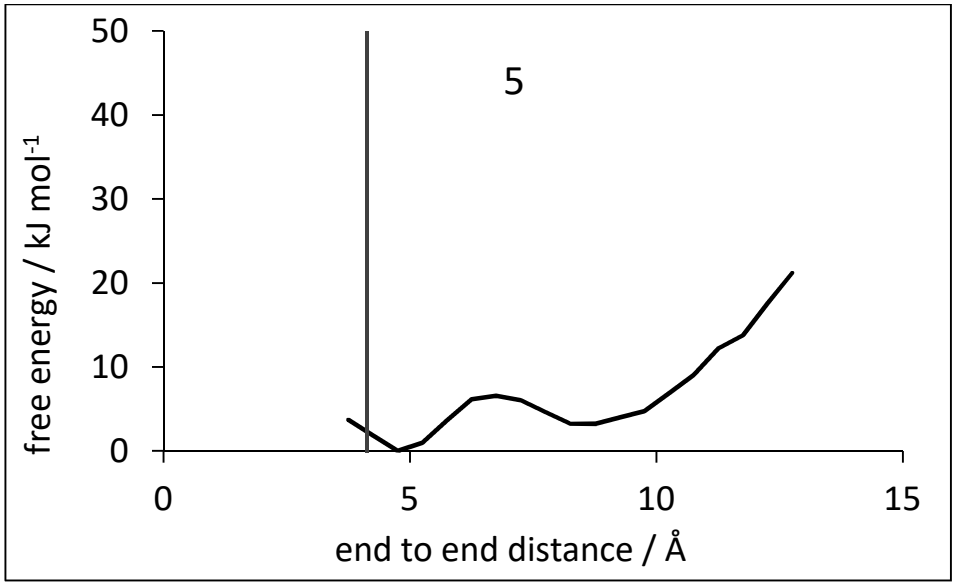

18
free energy / kJ mol-1
50
40
30
20
10
0
0
5
10
15
end to end distance / Å

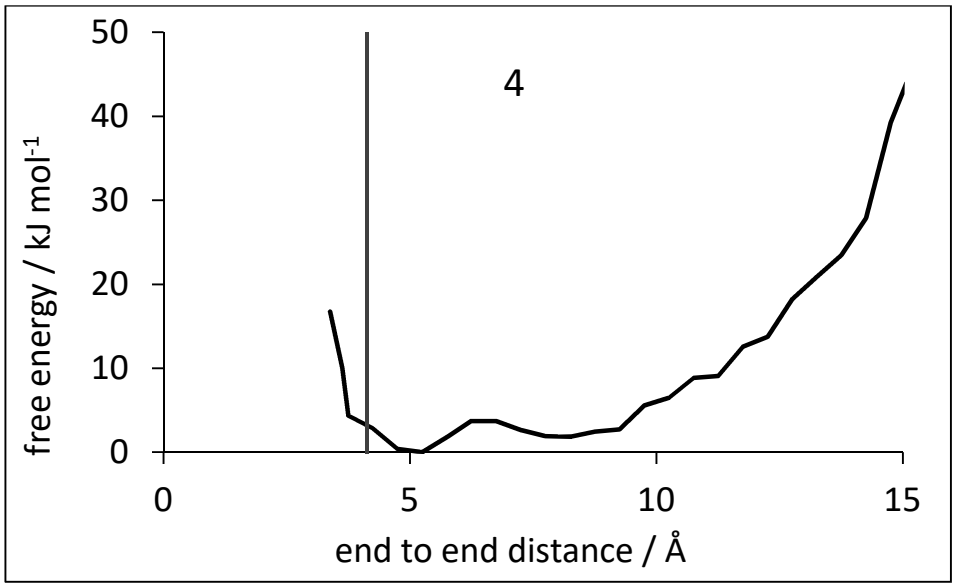

21
free energy / kJ mol-1
50
40
30
20
10
0
0
5
10
15
end to end distance / Å

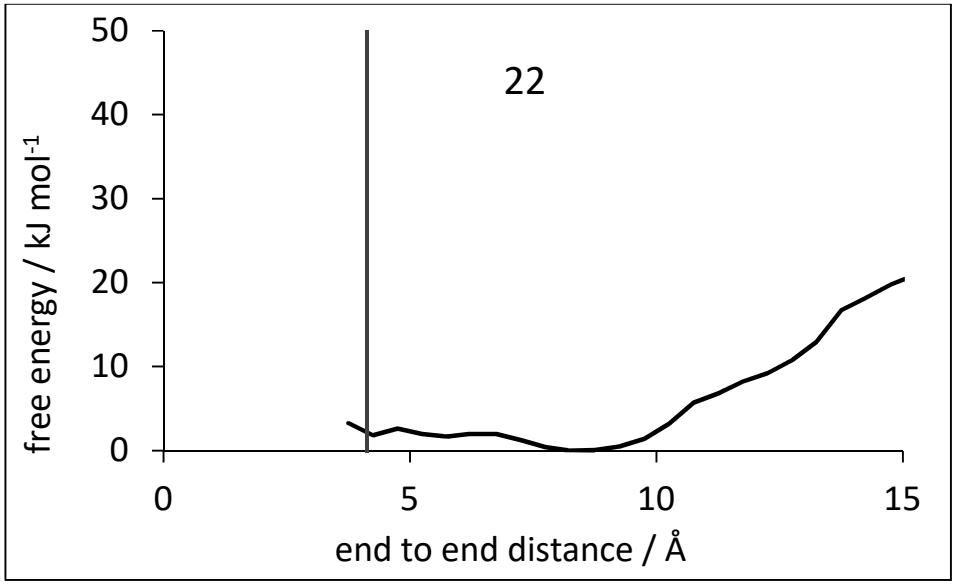

8
free energy / kJ mol-1
50
40
30
20
10
0
0
5
10
15
end to end distance / Å

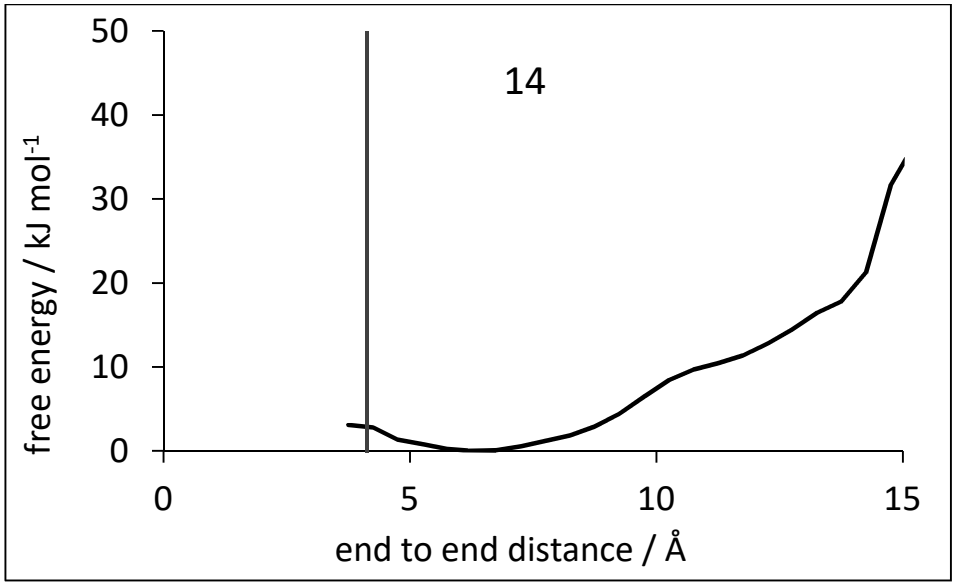

19
free energy / kJ mol-1
end to end distance / Å
0
10
20
30
40
50
0
5
10
15

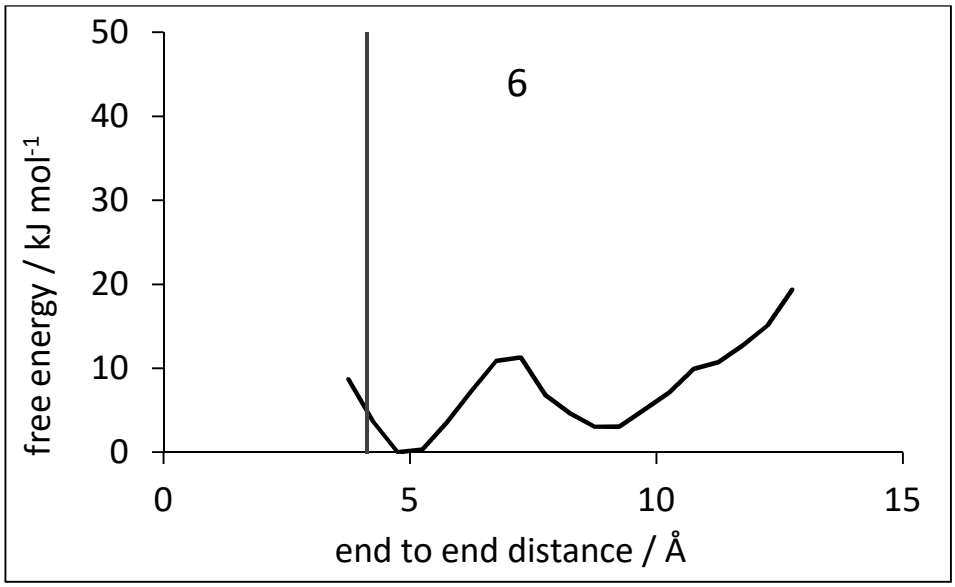

3
free energy / kJ mol-1
end to end distance / Å
0
10
20
30
40
50
0
5
10
15

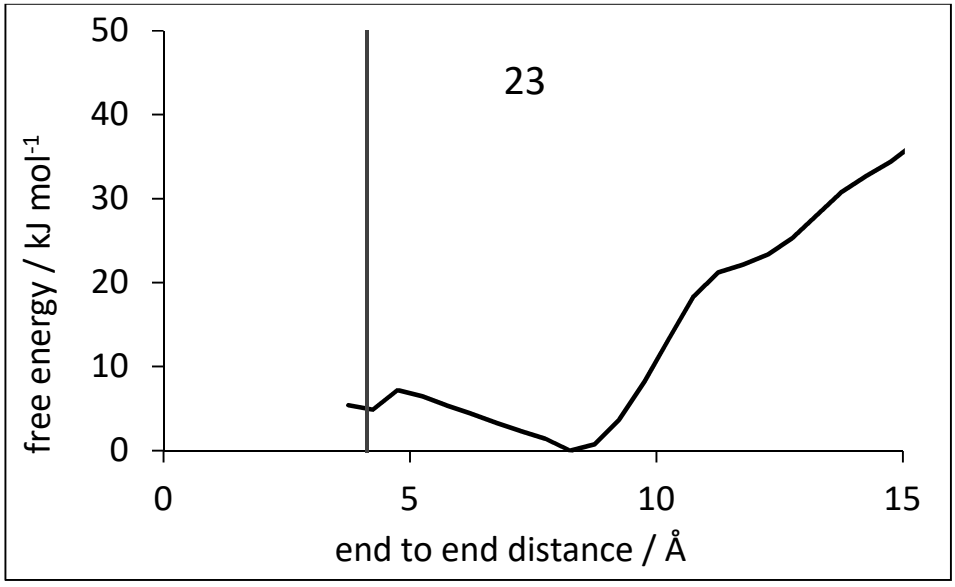

5
free energy / kJ mol-1
end to end distance / Å
0
10
20
30
40
50
0
5
10
15

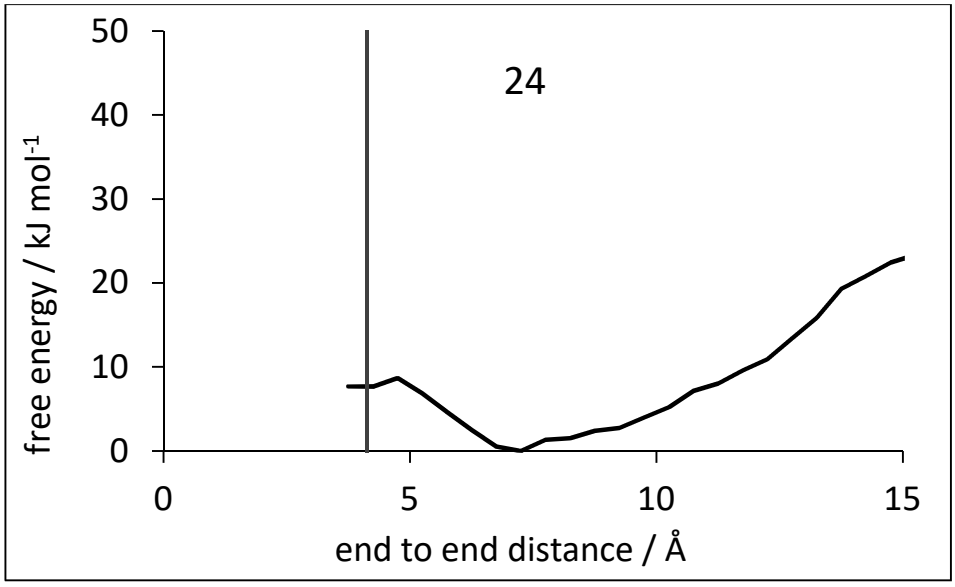

4
free energy / kJ mol-1
50
40
30
20
10
0
0
5
10
15
end to end distance / Å

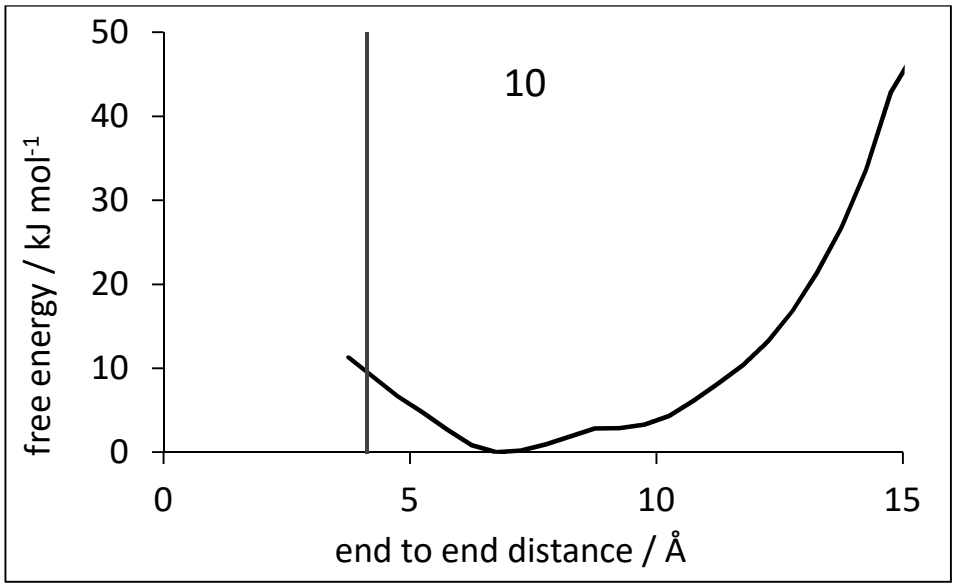

22
free energy / kJ mol-1
50
40
30
20
10
0
0
5
10
15
end to end distance / Å

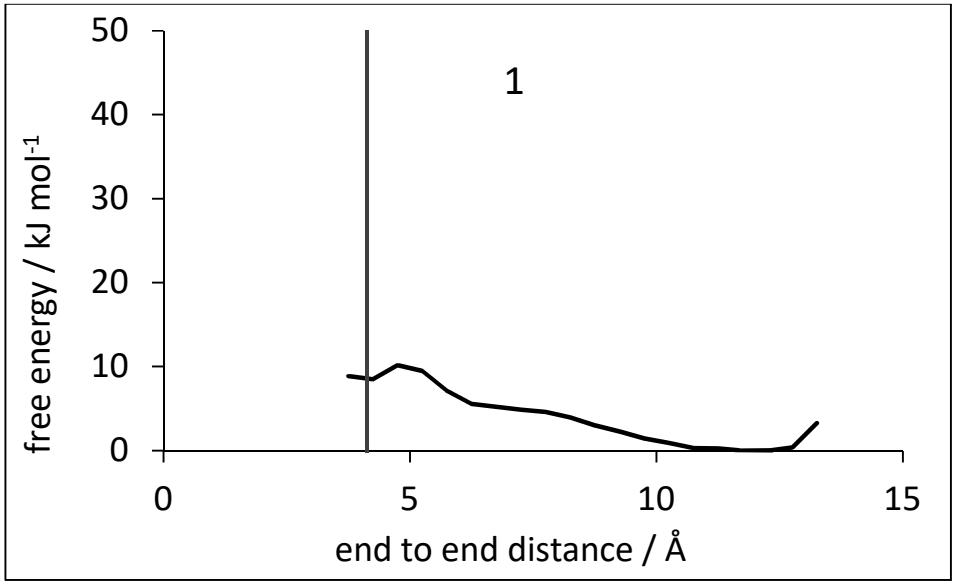

14
free energy / kJ mol-1
50
40
30
20
10
0
0
5
10
15
end to end distance / Å

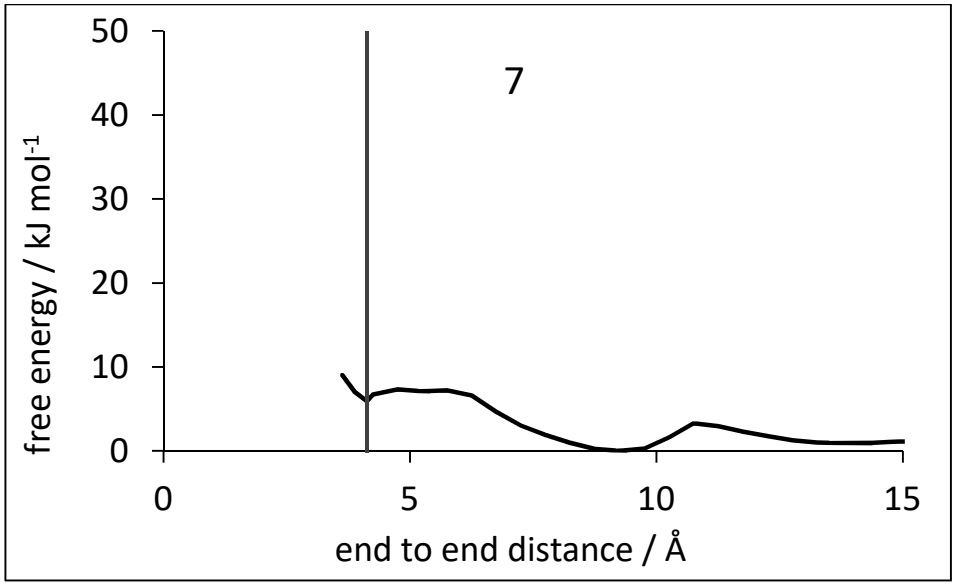

6
free energy / kJ mol-1
50
40
30
20
10
0
0
5
10
15
end to end distance / Å

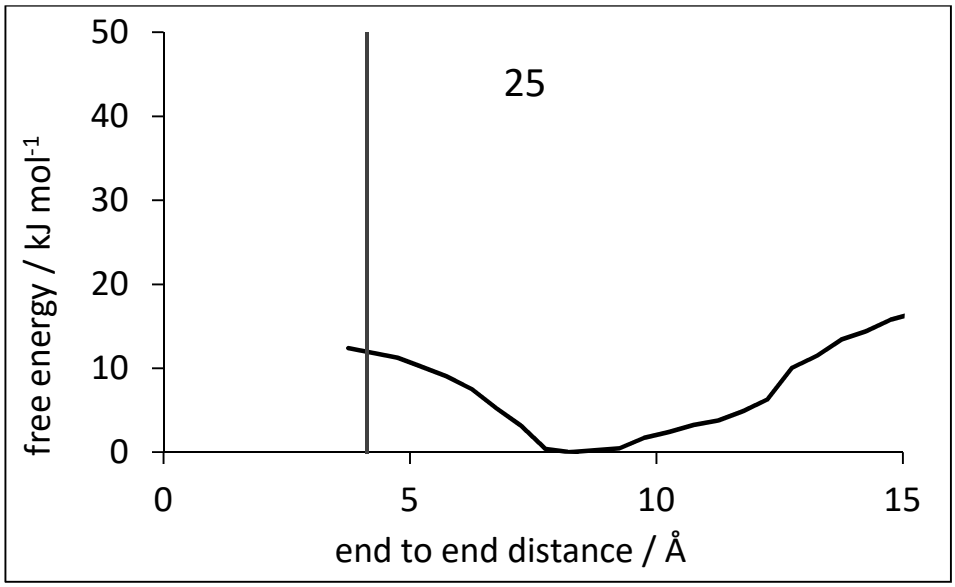

23
free energy / kJ mol-1
50
40
30
20
10
0
0
5
10
15
end to end distance / Å

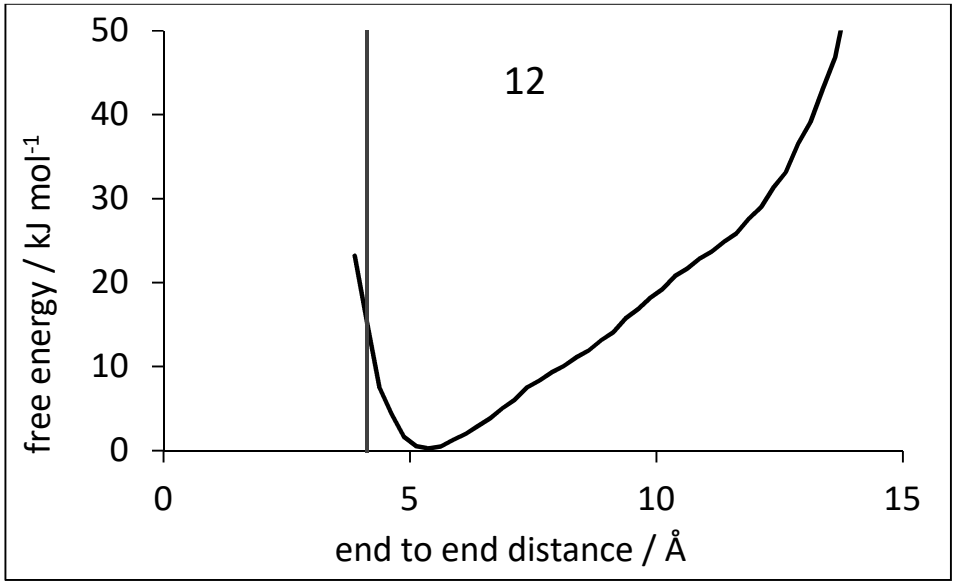

24
free energy / kJ mol-1
50
40
30
20
10
0
0
5
10
15
end to end distance / Å

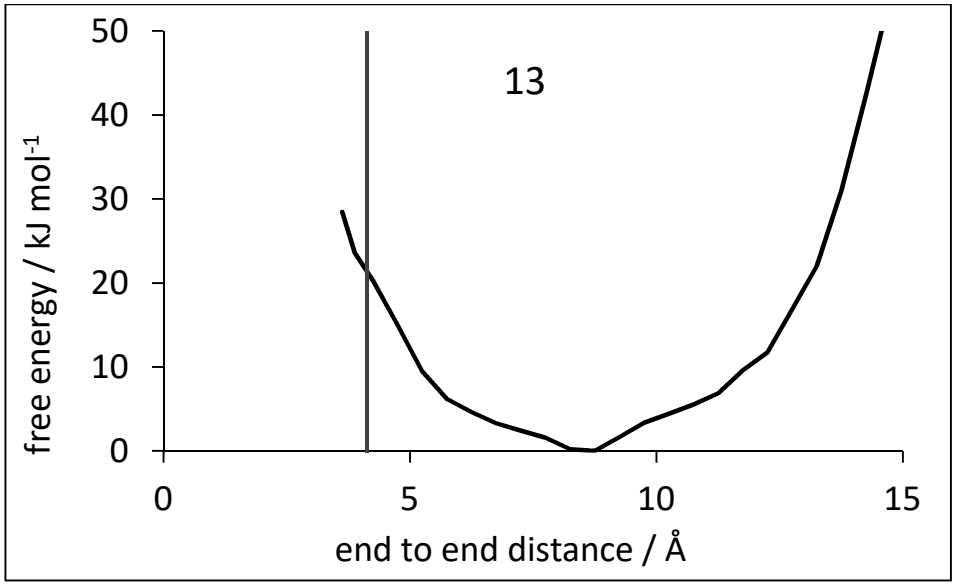

10
free energy / kJ mol-1
50
40
30
20
10
0
0
5
10
15
end to end distance / Å

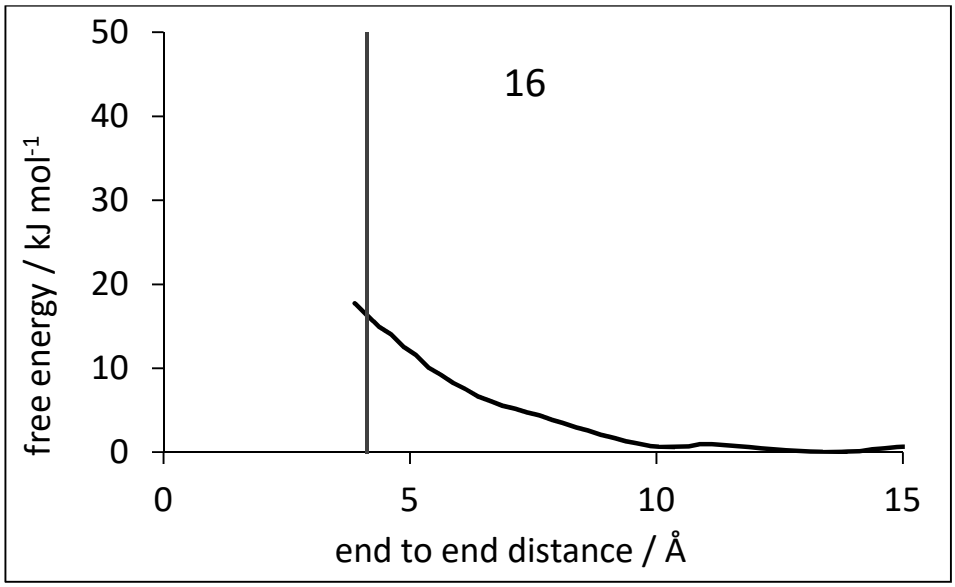

1
free energy / kJ mol-1
50
40
30
20
10
0
0
5
10
15
end to end distance / Å

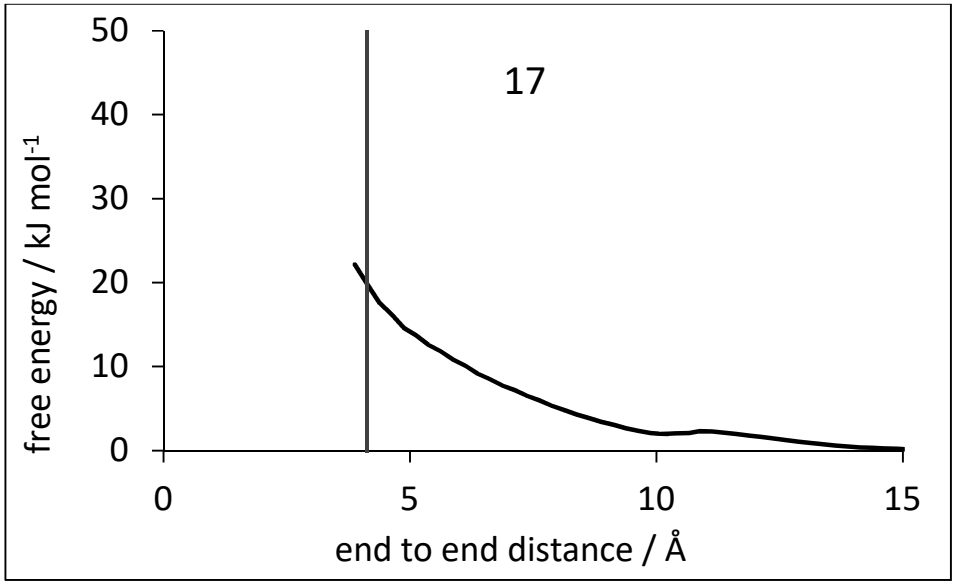

7
free energy / kJ mol-1
50
40
30
20
10
0
0
5
10
15
end to end distance / Å

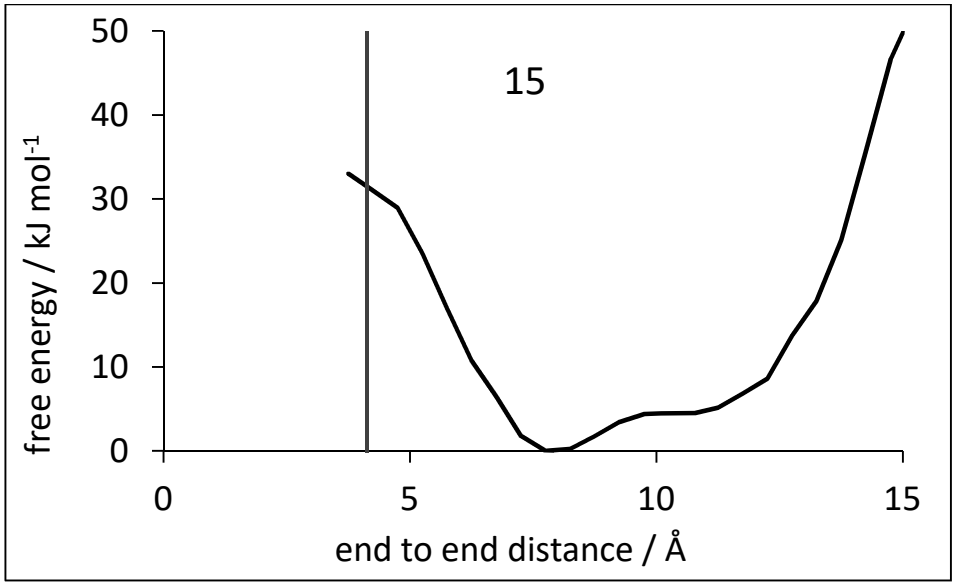

25
free energy / kJ mol-1
50
40
30
20
10
0
0
5
10
15
end to end distance / Å

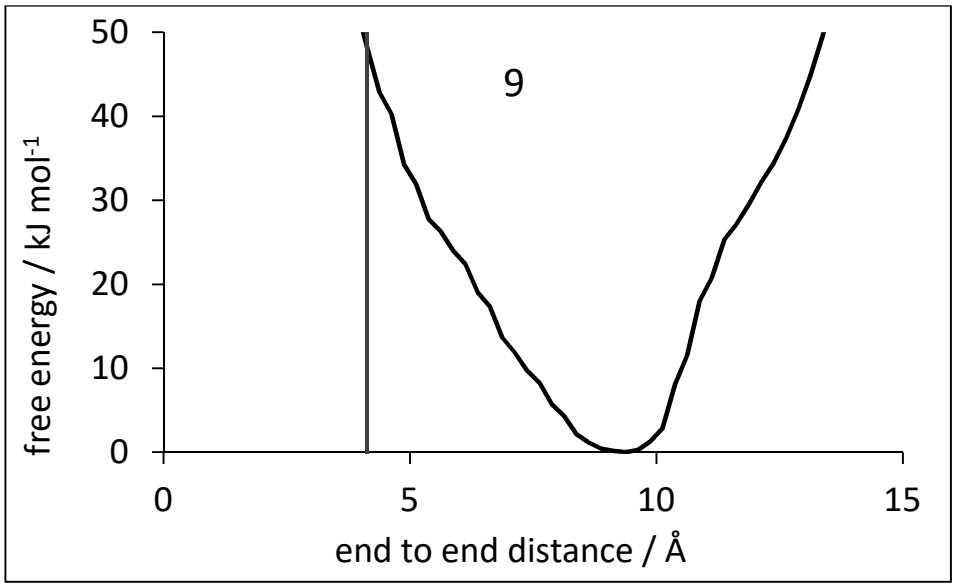

12
free energy / kJ mol-1
50
40
30
20
10
0
0
5
10
15
end to end distance / Å

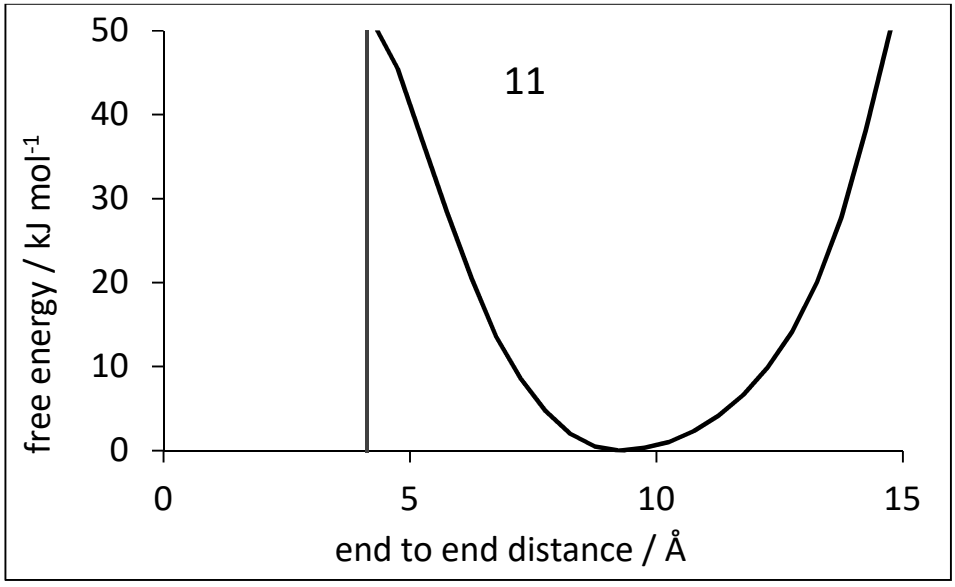

13
free energy / kJ mol-1
50
40
30
20
10
0
0
5
10
15
end to end distance / Å


16
free energy / kJ mol-1
50
40
30
20
10
0
0
5
10
15
end to end distance / Å


17
free energy / kJ mol-1
50
40
30
20
10
0
0
5
10
15
end to end distance / Å


15
free energy / kJ mol-1
50
40
30
20
10
0
0
5
10
15
end to end distance / Å


9
free energy / kJ mol-1
50
40
30
20
10
0
0
5
10
15
end to end distance / Å


11
free energy / kJ mol-1
50
40
30
20
10
0
0
5
10
15
end to end distance / Å