Dataset S1.

(a) Summary of proteomic evidence supporting novel cyclotide sequences from *C. ternatea*.

The following MS spectra show the ProteinPilot-assigned peptide spectral matches (PSM) for full-length linearized cyclotides generated by endoproteinase Glu-C digestion (spectra named with suffix E), trypsin digestion (suffix T) and chymotrypsin digestion (suffix C). In many instances semi-endoGlu-C, semi-tryptic or semi-chymotryptic peptides were observed and the corresponding PSM have been included as additional lines of evidence supporting the full-length peptide assignments. In one case, Cter N, the assignment was performed using RapiDeNovo software (Bruker) after MS/MS performed on a Bruker UltrafleXtreme mass spectrometer. In all other cases, spectra were acquired on an Applied Biosystems QStar Elite mass spectrometer. As an example, the PSM of Cter M in spectrum E1a (Endo Glu-C digestion) shows the assignment of 25/29 amino acids. The unassigned amino acids were assigned in subsequent spectra (EndoGlu-C: E1b, E1c; tryptic: T1a-T1c). This is the case for each peptide sequenced. In the following dataset outputs from ProteinPilot, amino acid sequences coloured green represent PSMs identified with 95% confidence, yellow represent 50% confidence, red represent <50% confidence, and black are not matched. Sequence ion assignment is shown in each spectrum with y-ions shown in red and b-ions shown in green. The matched ions are also shown in the left panel, where green shading represents a matched ion.

Evidence for Cter M:

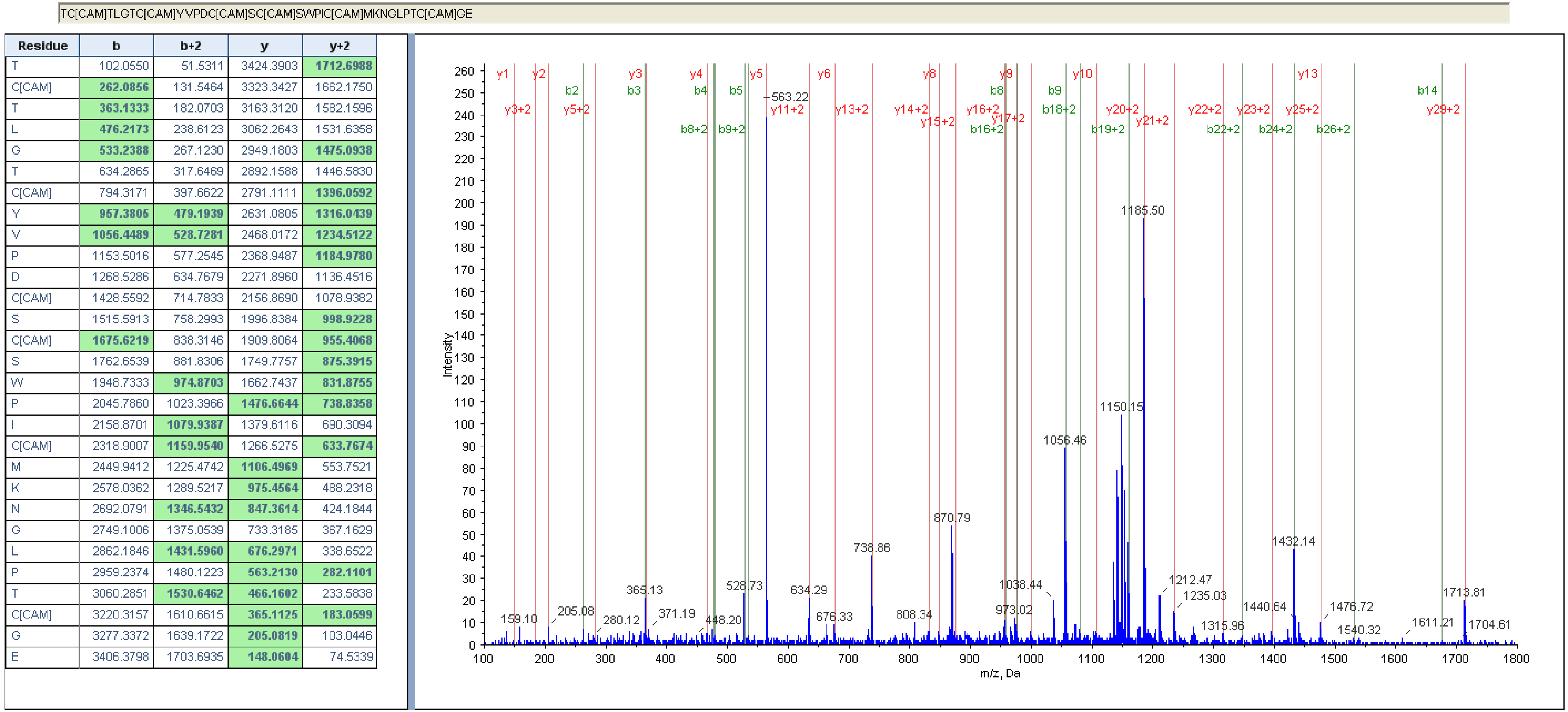
Endoproteinase Glu-C digest:

TCTLGTCYVPDCSCSWPICMKNGLPTCGE

NB: This peptide shows 75% sequence homology to kalata B7 (P58457) from *Oldenlandia affinis*.

```
1    GLPTCGETCTLGTCYVPDCSCSWPICMKN    29  Cter_M
1    GLPVCGETCTLGTCYTQGCTCSWPICKRN    29  kB7
     ***.************.  .*:****** :*
```

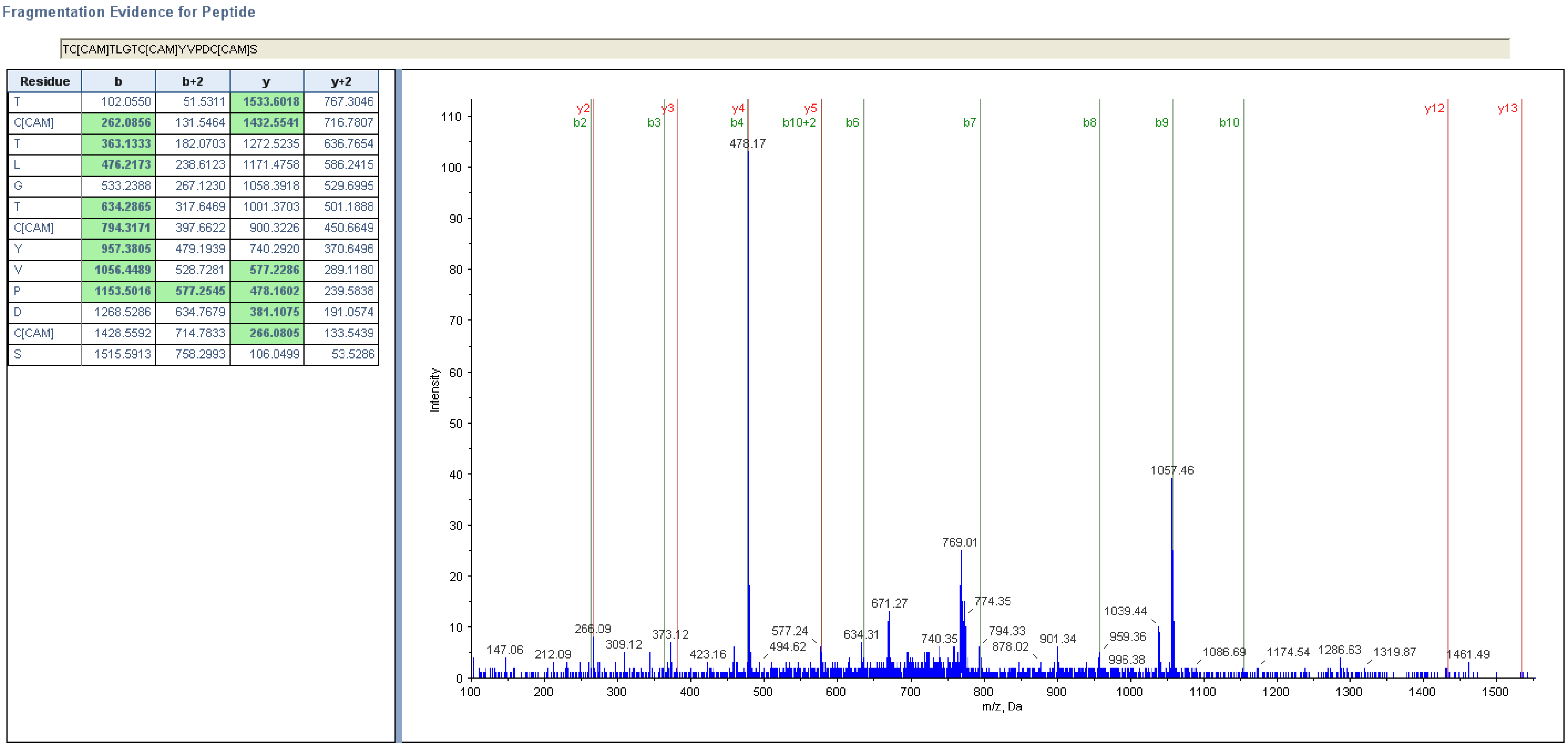
Fragmentation Evidence for Peptide



TC[CAM]TLGTC[CAM]YVPDC[CAM]SC[CAM]SWPIC[CAM]MKNGLPTC[CAM]GE

Residue	b	b+2	y	y+2
T	102.0550	51.5311	3424.3903	**1712.6988**
C[CAM]	**262.0856**	131.5464	3323.3427	1662.1750
T	**363.1333**	182.0703	3163.3120	1582.1596
L	**476.2173**	238.6123	3062.2643	1531.6358
G	**533.2388**	267.1230	2949.1803	**1475.0938**
T	634.2865	317.6469	2892.1588	1446.5830
C[CAM]	794.3171	397.6622	2791.1111	**1396.0592**
Y	**957.3805**	**479.1939**	2631.0805	**1316.0439**
V	**1056.4489**	**528.7281**	2468.0172	**1234.5122**
P	1153.5016	577.2545	2368.9487	**1184.9780**
D	1268.5286	634.7679	2271.8960	1136.4516
C[CAM]	1428.5592	714.7833	2156.8690	1078.9382
S	1515.5913	758.2993	1996.8384	**998.9228**
C[CAM]	**1675.6219**	838.3146	1909.8064	**955.4068**
S	1762.6539	881.8306	1749.7757	**875.3915**
W	1948.7333	**974.8703**	1662.7437	**831.8755**
P	2045.7860	1023.3966	**1476.6644**	**738.8358**
I	2158.8701	**1079.9387**	1379.6116	690.3094
C[CAM]	2318.9007	**1159.9540**	1266.5275	**633.7674**
M	2449.9412	1225.4742	**1106.4969**	553.7521
K	2578.0362	1289.5217	**975.4564**	488.2318
N	2692.0791	**1346.5432**	**847.3614**	424.1844
G	2749.1006	1375.0539	733.3185	367.1629
L	2862.1846	**1431.5960**	**676.2971**	338.6522
P	2959.2374	1480.1223	**563.2130**	**282.1101**
T	3060.2851	**1530.6462**	**466.1602**	233.5838
C[CAM]	3220.3157	1610.6615	**365.1125**	**183.0599**
G	3277.3372	1639.1722	**205.0819**	103.0446
E	3406.3798	1703.6935	**148.0604**	74.5339

Supplementary spectrum E1a: MS/MS of *m/z* 1142.13 (mass 3423.4 Da) showing sequence ion coverage. The derived sequence is TCTLGTCYVPDCSCSWPICMKNGLPTCGE.

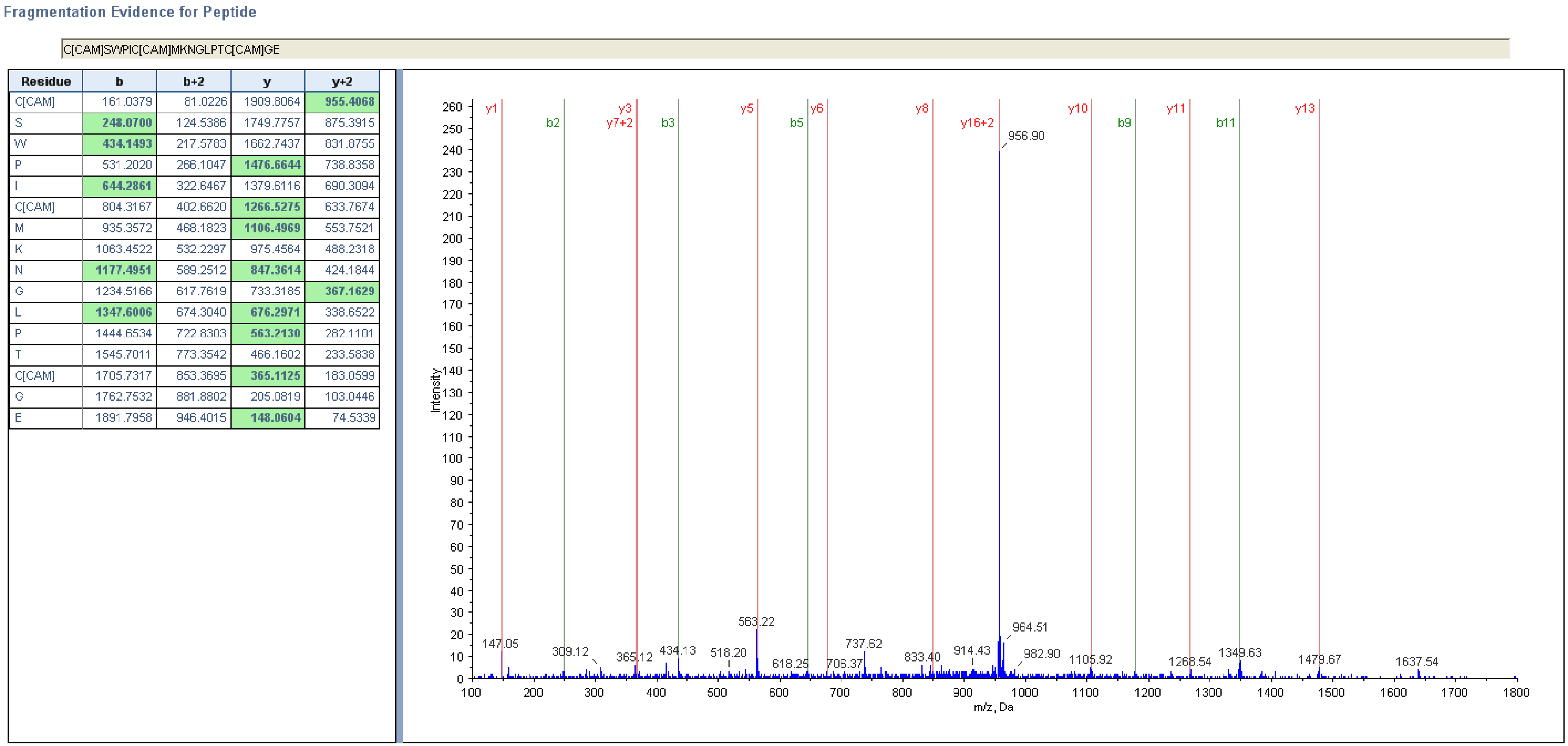


Fragmentation Evidence for Peptide

TC[CAM]TLGTC[CAM]YVPDC[CAM]S

Residue	b	b+2	y	y+2
T	102.0550	51.5311	**1533.6018**	767.3046
C[CAM]	**262.0856**	131.5464	**1432.5541**	716.7807
T	**363.1333**	182.0703	1272.5235	636.7654
L	**476.2173**	238.6123	1171.4758	586.2415
G	533.2388	267.1230	1058.3918	529.6995
T	**634.2865**	317.6469	1001.3703	501.1888
C[CAM]	**794.3171**	397.6622	900.3226	450.6649
Y	**957.3805**	479.1939	740.2920	370.6496
V	**1056.4489**	528.7281	**577.2286**	289.1180
P	**1153.5016**	**577.2545**	**478.1602**	239.5838
D	1268.5286	634.7679	**381.1075**	191.0574
C[CAM]	1428.5592	714.7833	**266.0805**	133.5439
S	1515.5913	758.2993	106.0499	53.5286

Supplementary spectrum E1b: MS/MS of *m/z* 767.32 (mass 1532.62 Da) showing sequence ion coverage. The derived sequence is TCTLGTCYVPDCS.



Fragmentation Evidence for Peptide

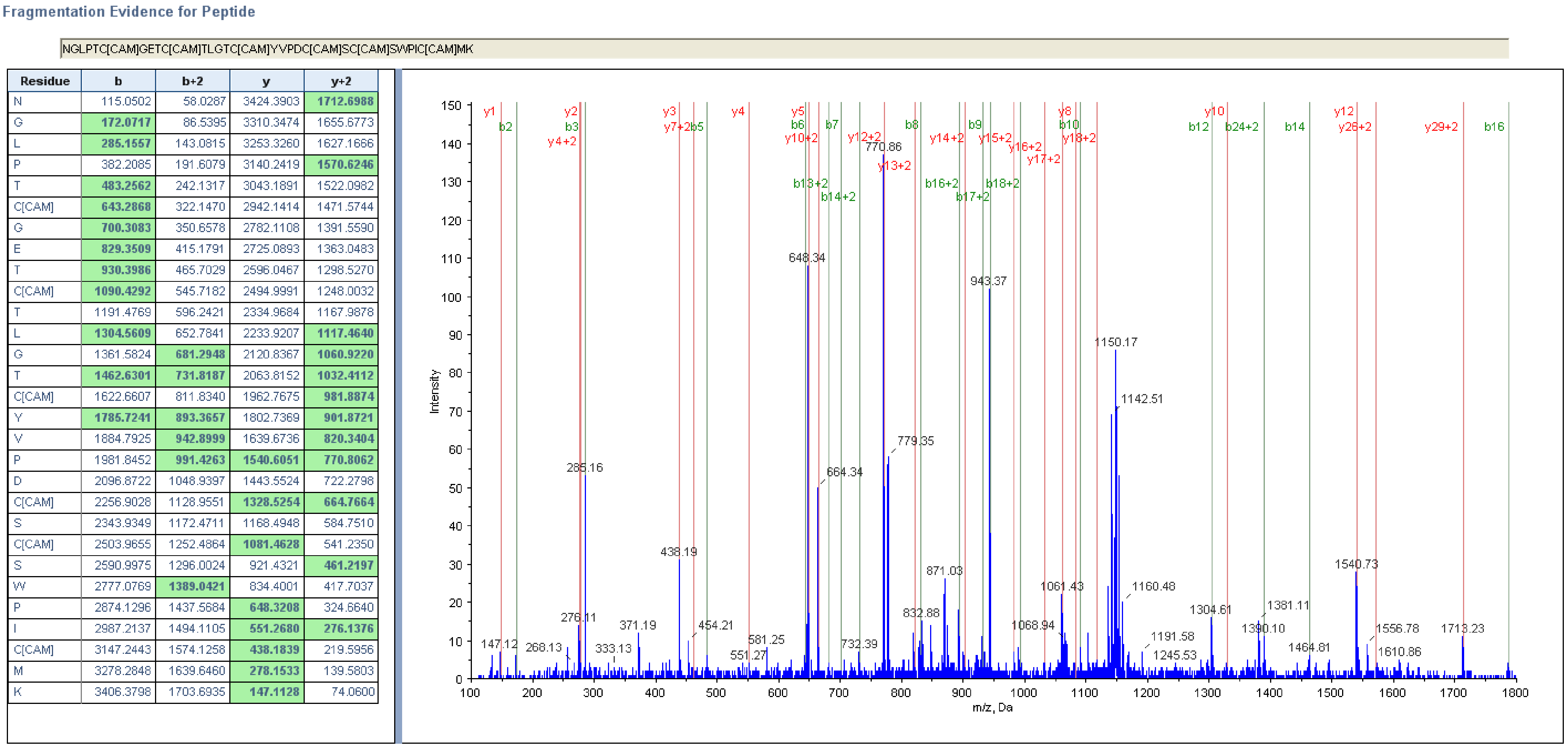
C[CAM]SWPIC[CAM]MKNGLPTC[CAM]GE

Residue	b	b+2	y	y+2
C[CAM]	161.0379	81.0226	1909.8064	**955.4068**
S	**248.0700**	124.5386	1749.7757	875.3915
W	**434.1493**	217.5783	1662.7437	831.8755
P	531.2020	266.1047	**1476.6644**	738.8358
I	**644.2861**	322.6467	1379.6116	690.3094
C[CAM]	804.3167	402.6620	**1266.5275**	633.7674
M	935.3572	468.1823	**1106.4969**	553.7521
K	1063.4522	532.2297	975.4564	488.2318
N	**1177.4951**	589.2512	**847.3614**	424.1844
G	1234.5166	617.7619	733.3185	**367.1629**
L	**1347.6006**	674.3040	**676.2971**	338.6522
P	1444.6534	722.8303	**563.2130**	282.1101
T	1545.7011	773.3542	466.1602	233.5838
C[CAM]	1705.7317	853.3695	**365.1125**	183.0599
G	1762.7532	881.8802	205.0819	103.0446
E	1891.7958	946.4015	**148.0604**	74.5339

Supplementary spectrum E1c: MS/MS of *m/z* 955.41 (mass 1908.81 Da) showing sequence ion coverage. The derived sequence is SWPICMKNGLPTCGE.

Trypsin digest:

NGLPTCGETCTLGTCYVPDCSCSWPICMK

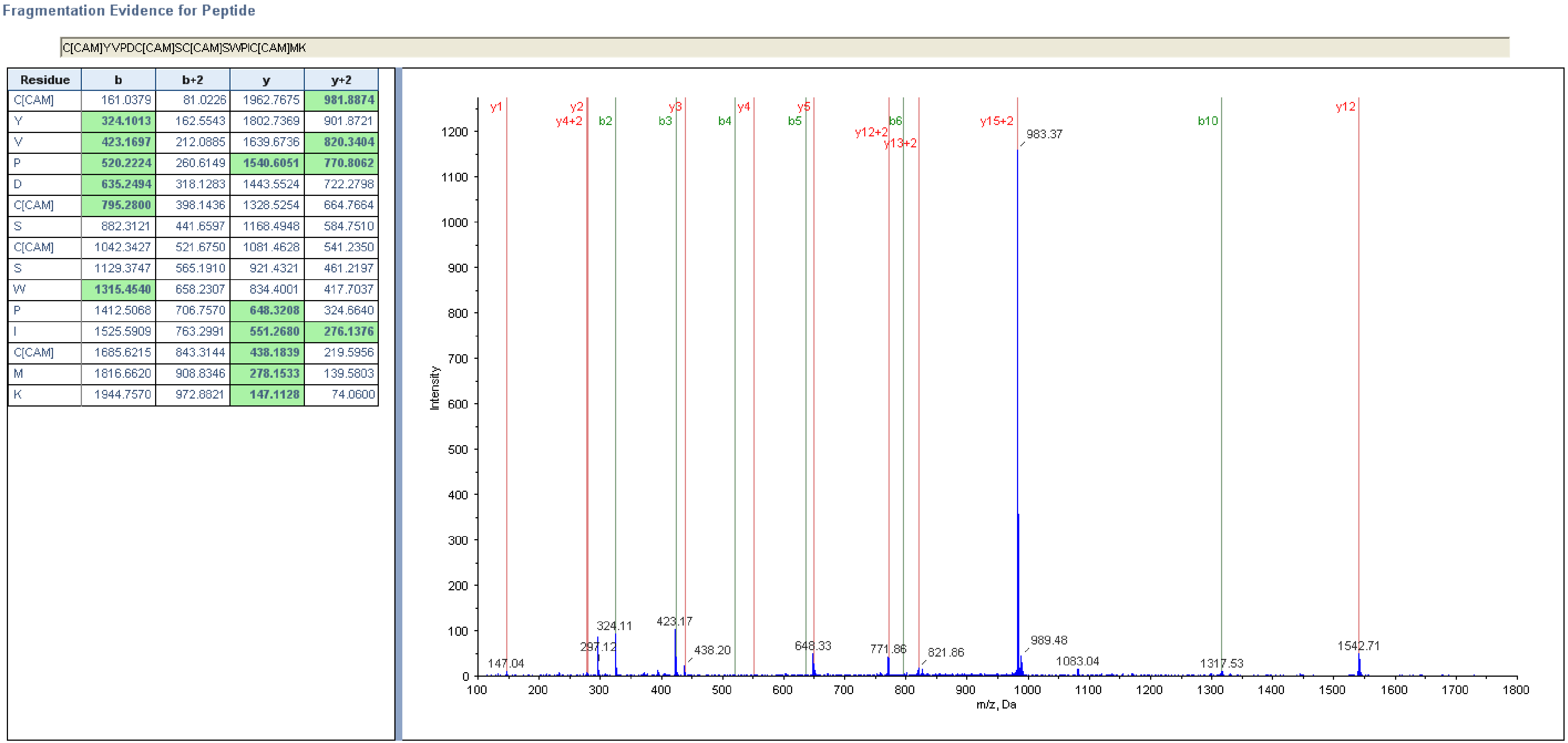


Fragmentation Evidence for Peptide

NGLPTC[CAM]GETC[CAM]TLGTC[CAM]YVPDC[CAM]SC[CAM]SWPIC[CAM]MK

Residue	b	b+2	y	y+2
N	115.0502	58.0287	3424.3903	**1712.6988**
G	**172.0717**	86.5395	3310.3474	1655.6773
L	**285.1557**	143.0815	3253.3260	1627.1666
P	382.2085	191.6079	3140.2419	**1570.6246**
T	**483.2562**	242.1317	3043.1891	1522.0982
C[CAM]	**643.2868**	322.1470	2942.1414	1471.5744
G	**700.3083**	350.6578	2782.1108	1391.5590
E	**829.3509**	415.1791	2725.0893	1363.0483
T	**930.3986**	465.7029	2596.0467	1298.5270
C[CAM]	**1090.4292**	545.7182	2494.9991	1248.0032
T	1191.4769	596.2421	2334.9684	1167.9878
L	**1304.5609**	652.7841	2233.9207	**1117.4640**
G	1361.5824	**681.2948**	2120.8367	**1060.9220**
T	**1462.6301**	**731.8187**	2063.8152	**1032.4112**
C[CAM]	1622.6607	811.8340	1962.7675	**981.8874**
Y	**1785.7241**	**893.3657**	1802.7369	**901.8721**
V	1884.7925	**942.8999**	1639.6736	**820.3404**
P	1981.8452	**991.4263**	**1540.6051**	**770.8062**
D	2096.8722	1048.9397	1443.5524	722.2798
C[CAM]	2256.9028	1128.9551	**1328.5254**	**664.7664**
S	2343.9349	1172.4711	1168.4948	584.7510
C[CAM]	2503.9655	1252.4864	**1081.4628**	541.2350
S	2590.9975	1296.0024	921.4321	**461.2197**
W	2777.0769	**1389.0421**	834.4001	417.7037
P	2874.1296	1437.5684	**648.3208**	324.6640
I	2987.2137	1494.1105	**551.2680**	**276.1376**
C[CAM]	3147.2443	1574.1258	**438.1839**	219.5956
M	3278.2848	1639.6460	**278.1533**	139.5803
K	3406.3798	1703.6935	**147.1128**	74.0600

Supplementary spectrum T1a: MS/MS of *m/z* 1142.13 (mass 3423.38 Da) showing sequence ion coverage. The derived sequence is NGLPTCGETCTLGTCYVPDCSCSWPICMK.

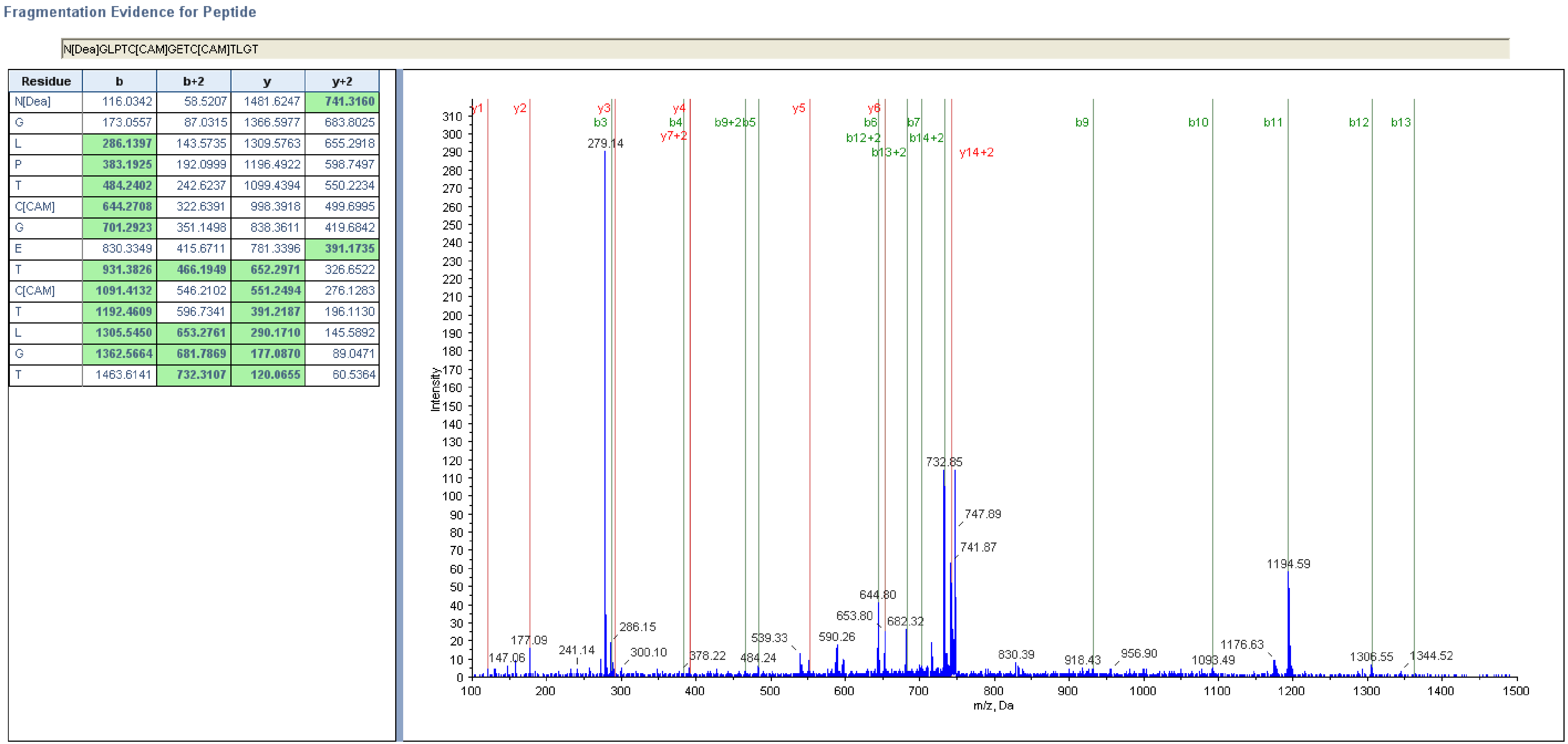


Fragmentation Evidence for Peptide

C[CAM]YVPDC[CAM]SC[CAM]SWPIC[CAM]MK

Residue	b	b+2	y	y+2
C[CAM]	161.0379	81.0226	1962.7675	**981.8874**
Y	**324.1013**	162.5543	1802.7369	901.8721
V	**423.1697**	212.0885	1639.6736	**820.3404**
P	**520.2224**	260.6149	**1540.6051**	**770.8062**
D	**635.2494**	318.1283	1443.5524	722.2798
C[CAM]	**795.2800**	398.1436	1328.5254	664.7664
S	882.3121	441.6597	1168.4948	584.7510
C[CAM]	1042.3427	521.6750	1081.4628	541.2350
S	1129.3747	565.1910	921.4321	461.2197
W	**1315.4540**	658.2307	834.4001	417.7037
P	1412.5068	706.7570	**648.3208**	324.6640
I	1525.5909	763.2991	**551.2680**	**276.1376**
C[CAM]	1685.6215	843.3144	**438.1839**	219.5956
M	1816.6620	908.8346	**278.1533**	139.5803
K	1944.7570	972.8821	**147.1128**	74.0600

Supplementary spectrum T1b: MS/MS of *m/z* 981.90 (mass 1961.78 Da) showing sequence ion coverage. The derived sequence is CYVPDCSCSWPICMK.



Fragmentation Evidence for Peptide

N[Dea]GLPTC[CAM]GETC[CAM]TLGT

Residue	b	b+2	y	y+2
N[Dea]	116.0342	58.5207	1481.6247	**741.3160**
G	173.0557	87.0315	1366.5977	683.8025
L	**286.1397**	143.5735	1309.5763	655.2918
P	**383.1925**	192.0999	1196.4922	598.7497
T	**484.2402**	242.6237	1099.4394	550.2234
C[CAM]	**644.2708**	322.6391	998.3918	499.6995
G	**701.2923**	351.1498	838.3611	419.6842
E	830.3349	415.6711	781.3396	**391.1735**
T	**931.3826**	**466.1949**	**652.2971**	326.6522
C[CAM]	**1091.4132**	546.2102	**551.2494**	276.1283
T	**1192.4609**	596.7341	**391.2187**	196.1130
L	**1305.5450**	**653.2761**	**290.1710**	145.5892
G	**1362.5664**	**681.7869**	**177.0870**	89.0471
T	1463.6141	**732.3107**	**120.0655**	60.5364

Supplementary spectrum T1c: MS/MS of *m/z* 741.33 (mass 1480.65 Da) showing sequence ion coverage. The derived sequence is NGLPTCGETCTLGT.

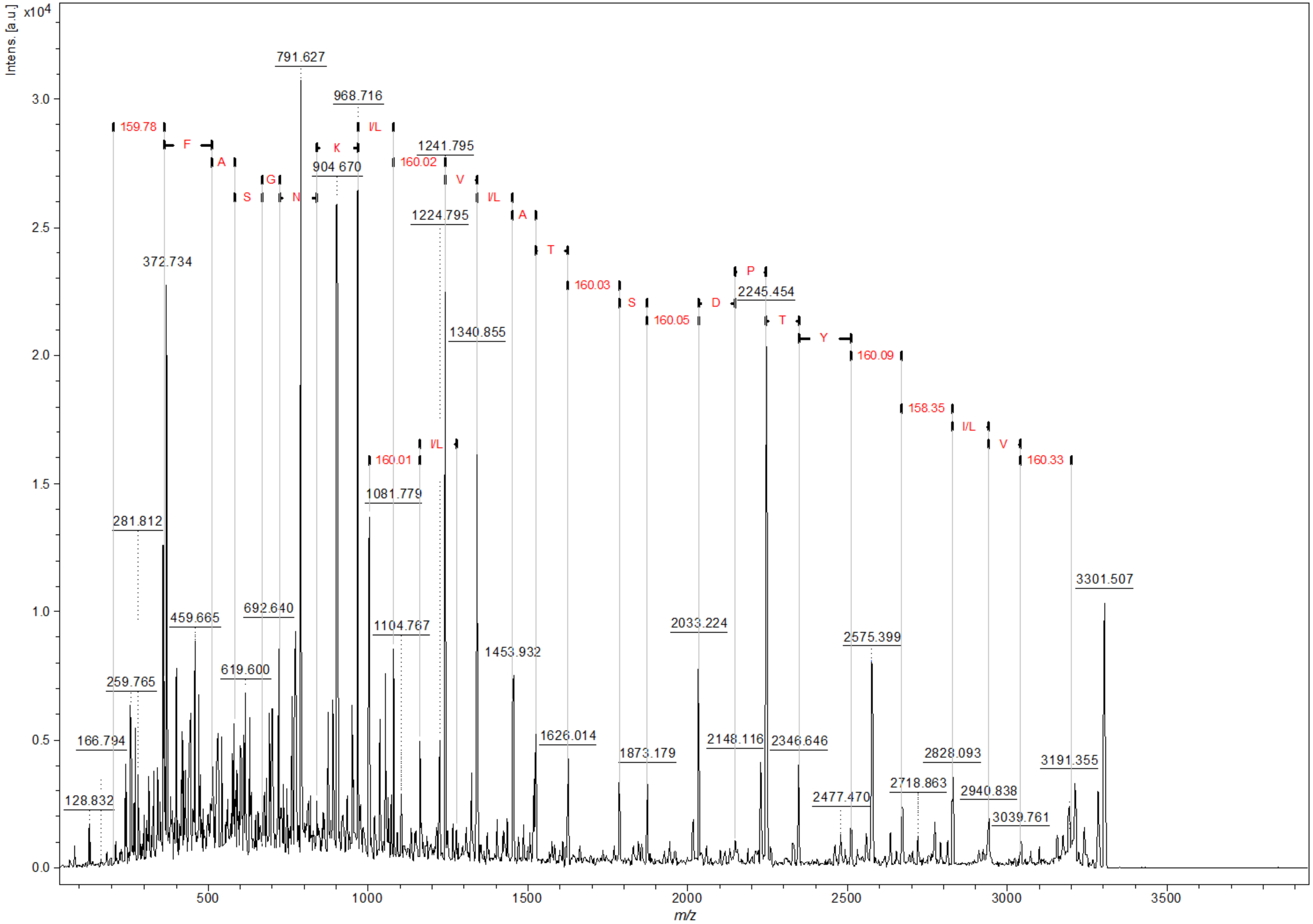
Evidence for Cter N:

Endoproteinase Glu-C digest:

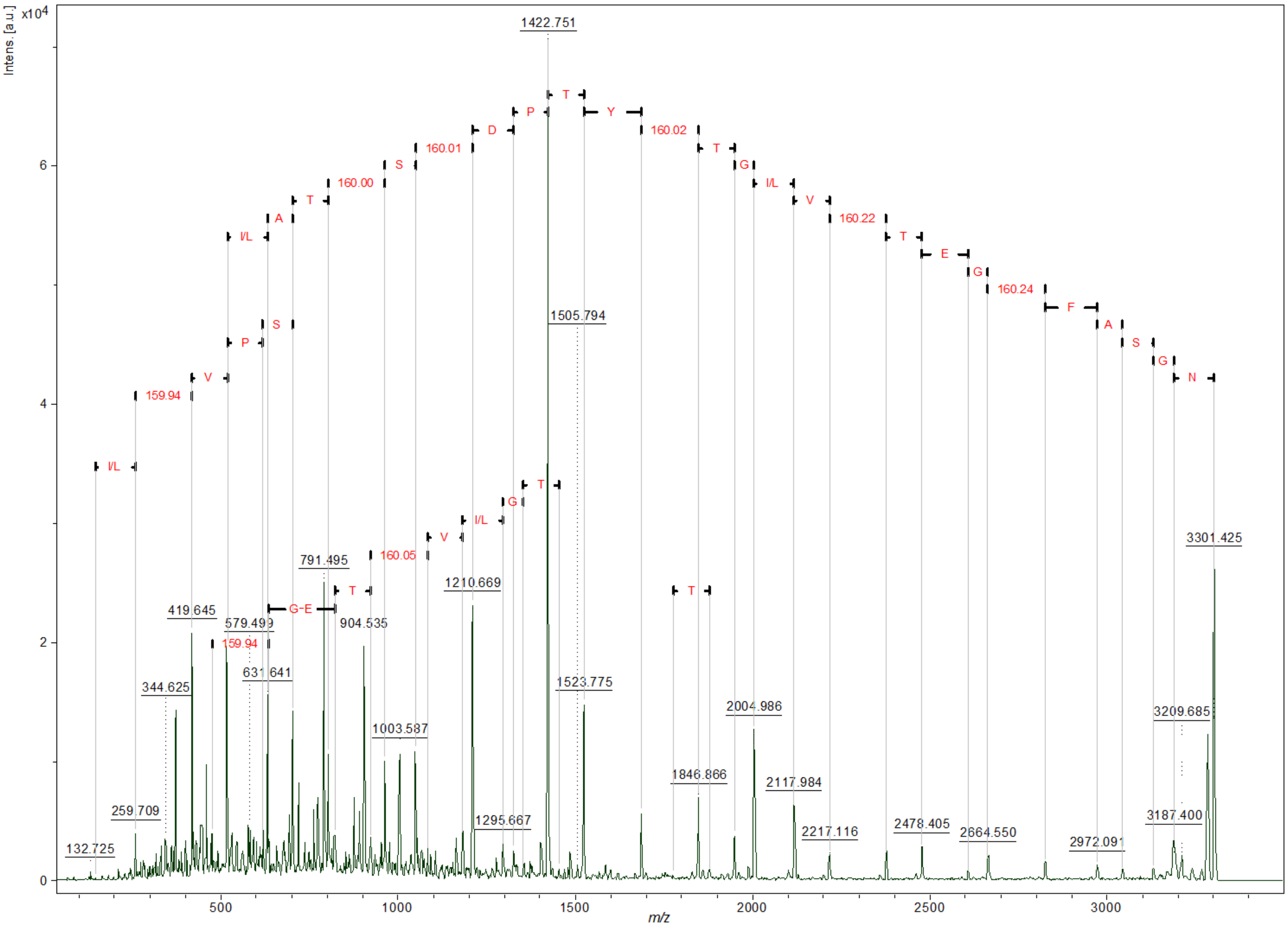
TCVLGTCYTPDCSCTALVCLKNGSAFCGE

NB: This peptide shows homology to kalata B9 (P85127) from *Oldenlandia affinis*.

```
1     GSAF-CGETCVLGTCYTPDCSC-TALVCLKN   29  Cter_N
1     GSVFNCGETCVLGTCYTPGCTCNTYRVCTKD   31  kB9
      **.* *************.*:* *  ** *:
```



Supplementary spectrum E2a: MS/MS of *m/z* 3301.42 (mass 3300.42 Da) showing sequence ion coverage. The derived sequence is TCVLGTCYTPDCSCTALVCLKNGSAFCGE.



Supplementary spectrum T2a: MS/MS of *m/z* 3301.42 (mass 3300.42 Da) showing sequence ion coverage. The derived sequence is NGSAFCGETCVLGTCYTPDCSCTALVCLK.

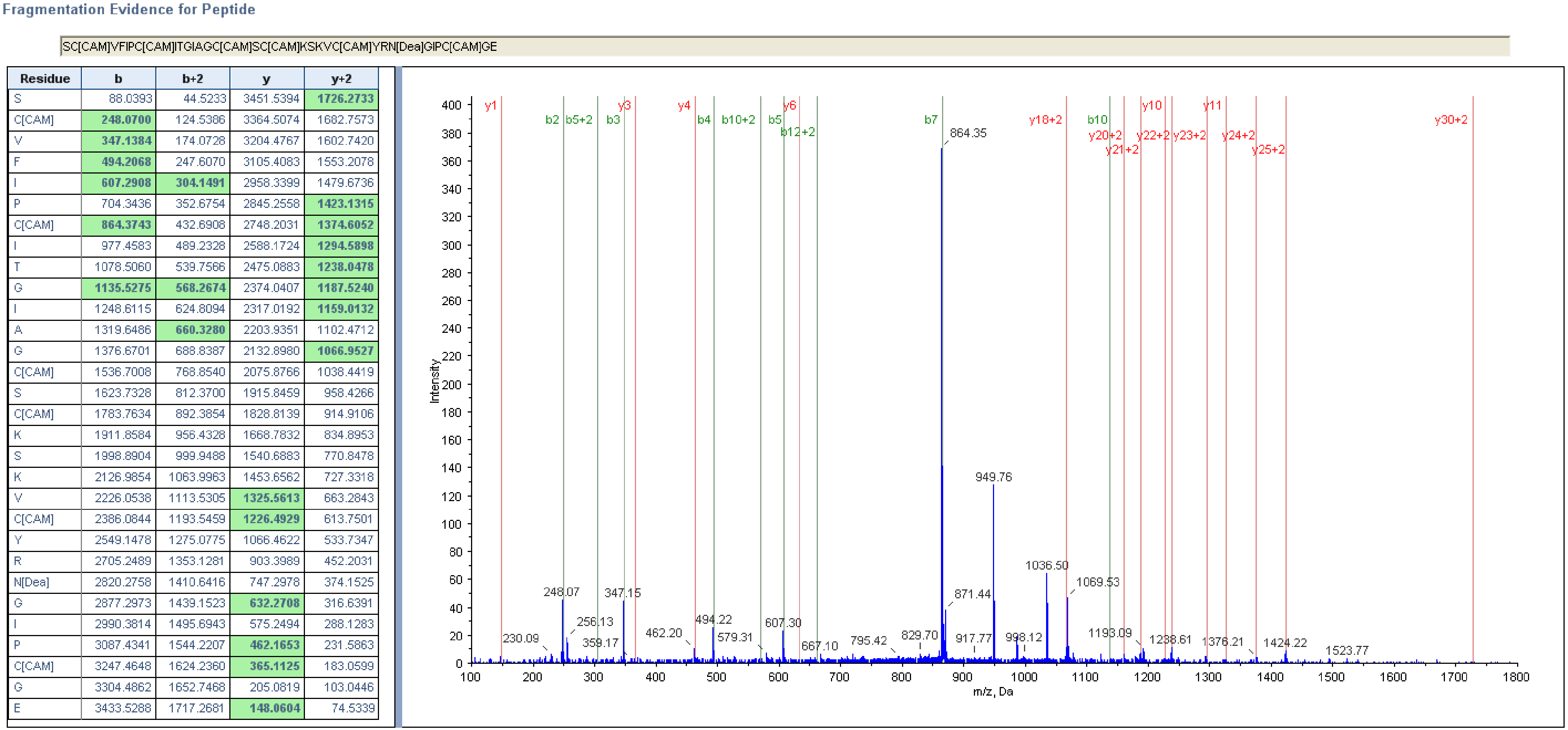
Evidence for Cter O:

Endoproteinase Glu-C digest:

SCVFIPCITGIAGCSCKSKVCYRNGIPCGE

NB: This peptide shows 93% sequence homology to circulin-C (P84641) from *Chassalia parvifolia*.

```
1     GIPCGESCVFIPCITGIAGCSCKSKVCYRN    30  Cter_O
1     GIPCGESCVFIPCITSVAGCSCKSKVCYRN    30  Circulin-C
      ***************.:*************
```

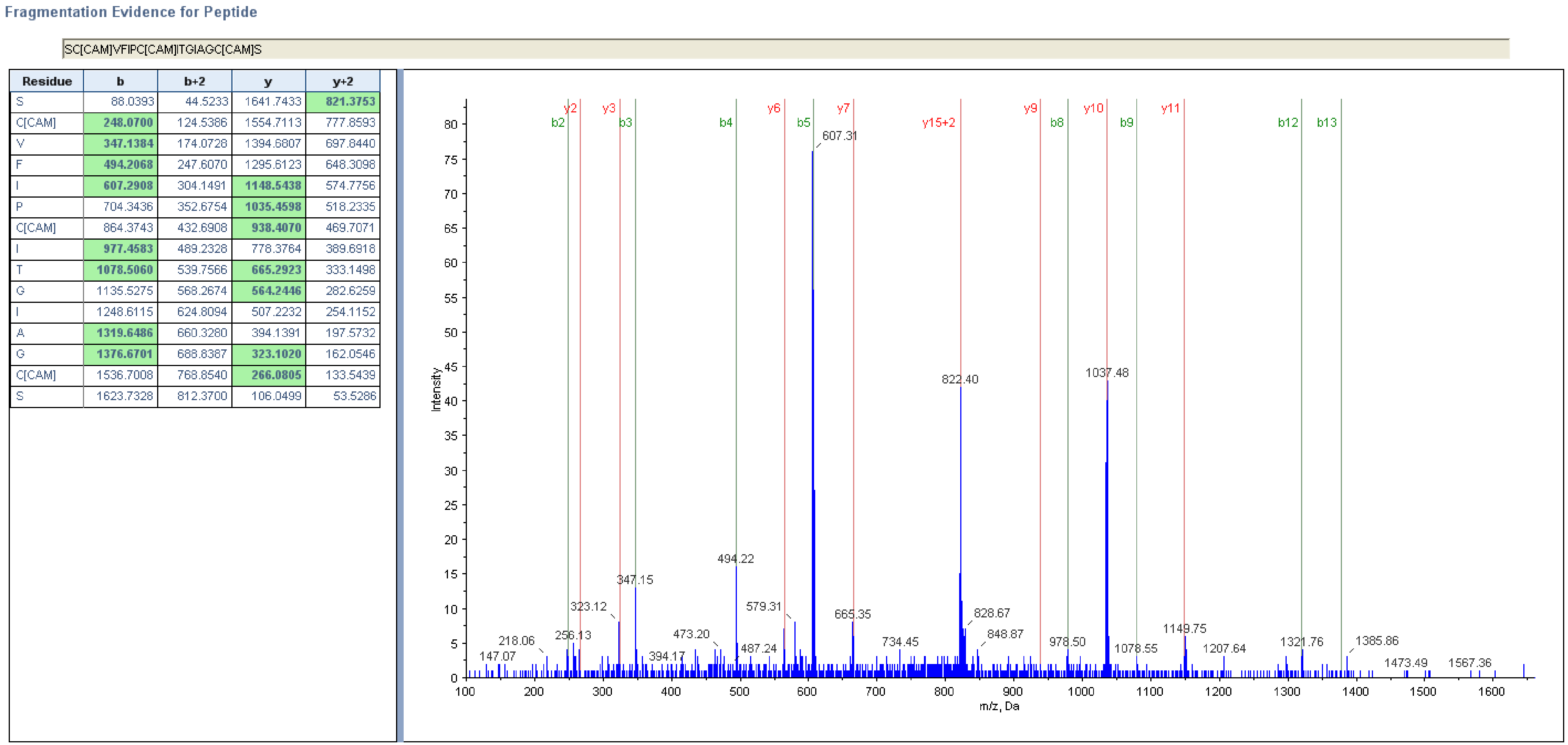


Fragmentation Evidence for Peptide

SC[CAM]VFIPC[CAM]ITGIAGC[CAM]SC[CAM]KSKVC[CAM]YRN[Dea]GIPC[CAM]GE

Residue	b	b+2	y	y+2
S	88.0393	44.5233	3451.5394	**1726.2733**
C[CAM]	**248.0700**	124.5386	3364.5074	1682.7573
V	**347.1384**	174.0728	3204.4767	1602.7420
F	**494.2068**	247.6070	3105.4083	1553.2078
I	**607.2908**	**304.1491**	2958.3399	1479.6736
P	704.3436	352.6754	2845.2558	**1423.1315**
C[CAM]	**864.3743**	432.6908	2748.2031	**1374.6052**
I	977.4583	489.2328	2588.1724	**1294.5898**
T	1078.5060	539.7566	2475.0883	**1238.0478**
G	**1135.5275**	**568.2674**	2374.0407	**1187.5240**
I	1248.6115	624.8094	2317.0192	**1159.0132**
A	1319.6486	**660.3280**	2203.9351	1102.4712
G	1376.6701	688.8387	2132.8980	**1066.9527**
C[CAM]	1536.7008	768.8540	2075.8766	1038.4419
S	1623.7328	812.3700	1915.8459	958.4266
C[CAM]	1783.7634	892.3854	1828.8139	914.9106
K	1911.8584	956.4328	1668.7832	834.8953
S	1998.8904	999.9488	1540.6883	770.8478
K	2126.9854	1063.9963	1453.6562	727.3318
V	2226.0538	1113.5305	**1325.5613**	663.2843
C[CAM]	2386.0844	1193.5459	**1226.4929**	613.7501
Y	2549.1478	1275.0775	1066.4622	533.7347
R	2705.2489	1353.1281	903.3989	452.2031
N[Dea]	2820.2758	1410.6416	747.2978	374.1525
G	2877.2973	1439.1523	**632.2708**	316.6391
I	2990.3814	1495.6943	575.2494	288.1283
P	3087.4341	1544.2207	**462.1653**	231.5863
C[CAM]	3247.4648	1624.2360	**365.1125**	183.0599
G	3304.4862	1652.7468	205.0819	103.0446
E	3433.5288	1717.2681	**148.0604**	74.5339

Supplementary spectrum E3a: MS/MS of *m/z* 863.41 (mass 3449.62 Da) showing sequence ion coverage. The derived sequence is SCVFIPCITGIAGCSCKSKVCYRNGIPCGE.

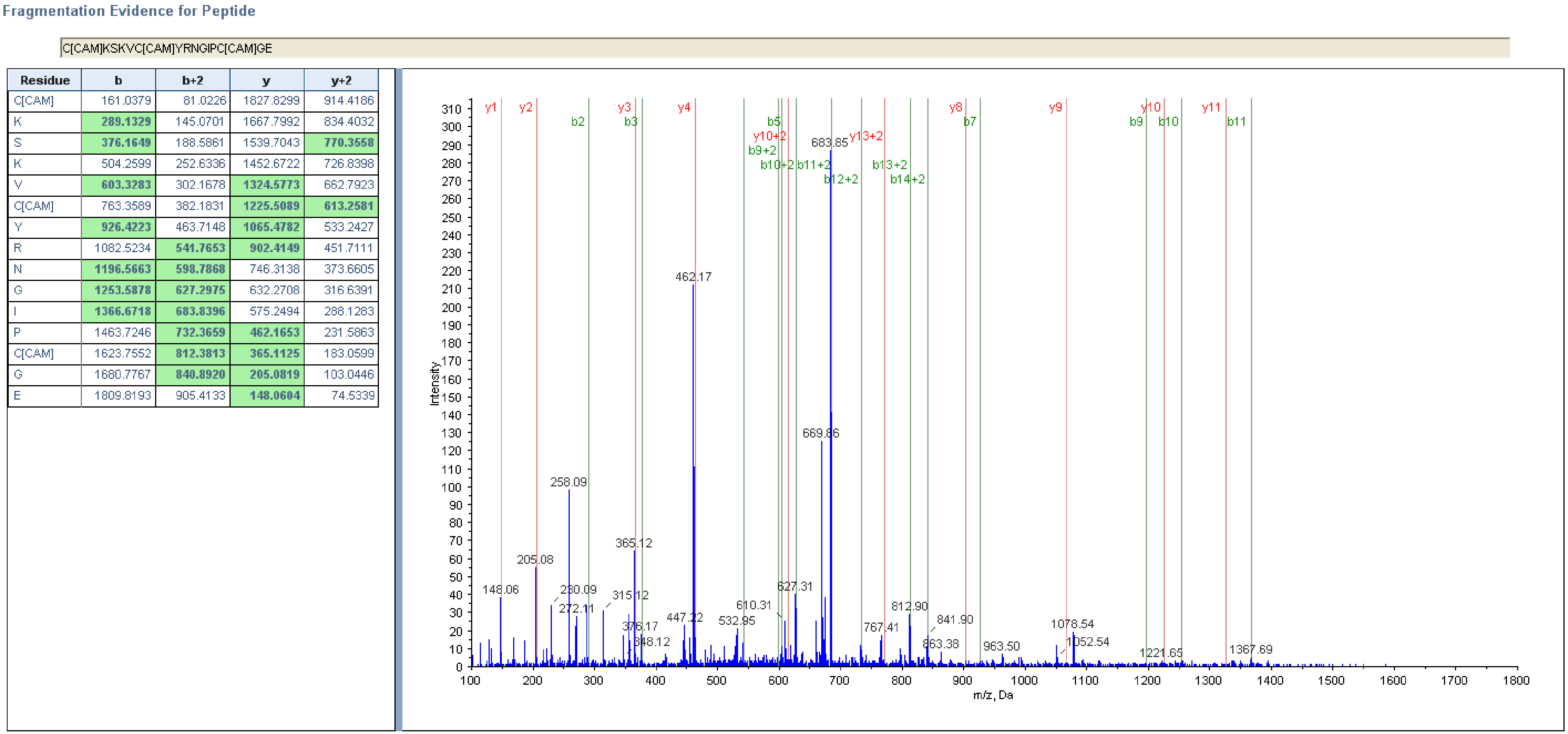


Fragmentation Evidence for Peptide

SC[CAM]VFIPC[CAM]ITGIAGC[CAM]S

Residue	b	b+2	y	y+2
S	88.0393	44.5233	1641.7433	**821.3753**
C[CAM]	**248.0700**	124.5386	1554.7113	777.8593
V	**347.1384**	174.0728	1394.6807	697.8440
F	**494.2068**	247.6070	1295.6123	648.3098
I	**607.2908**	304.1491	**1148.5438**	574.7756
P	704.3436	352.6754	**1035.4598**	518.2335
C[CAM]	864.3743	432.6908	**938.4070**	469.7071
I	**977.4583**	489.2328	778.3764	389.6918
T	**1078.5060**	539.7566	**665.2923**	333.1498
G	1135.5275	568.2674	**564.2446**	282.6259
I	1248.6115	624.8094	507.2232	254.1152
A	**1319.6486**	660.3280	394.1391	197.5732
G	**1376.6701**	688.8387	**323.1020**	162.0546
C[CAM]	1536.7008	768.8540	**266.0805**	133.5439
S	1623.7328	812.3700	106.0499	53.5286

Supplementary spectrum E3b: MS/MS of *m/z* 821.39 (mass 1640.77 Da) showing sequence ion coverage. The derived sequence is CVFIPCITGIAGCS.



Fragmentation Evidence for Peptide

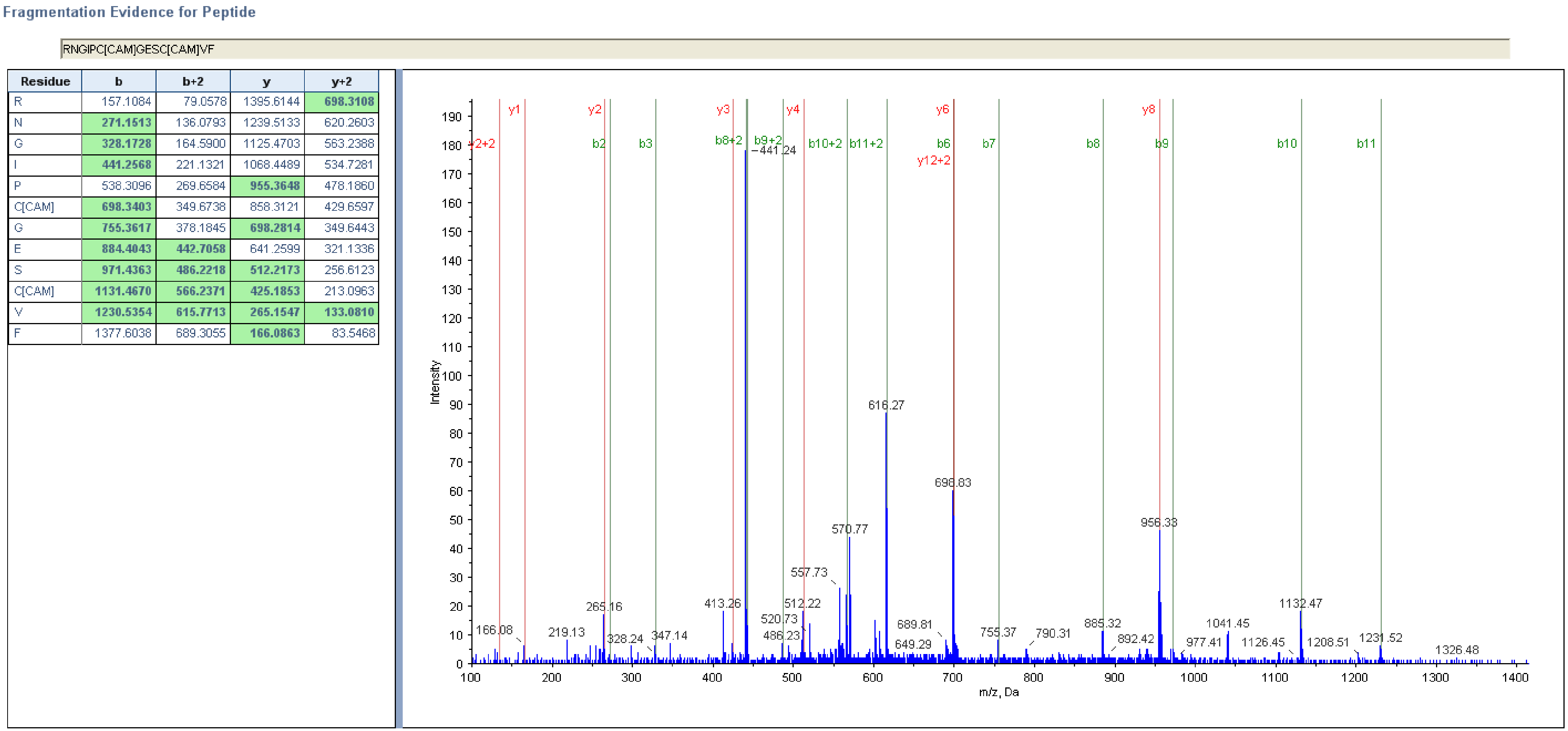
C[CAM]KSKVC[CAM]YRNGIPC[CAM]GE

Residue	b	b+2	y	y+2
C[CAM]	161.0379	81.0226	1827.8299	914.4186
K	**289.1329**	145.0701	1667.7992	834.4032
S	**376.1649**	188.5861	1539.7043	**770.3558**
K	504.2599	252.6336	1452.6722	726.8398
V	**603.3283**	302.1678	**1324.5773**	662.7923
C[CAM]	763.3589	382.1831	**1225.5089**	**613.2581**
Y	**926.4223**	463.7148	**1065.4782**	533.2427
R	1082.5234	**541.7653**	**902.4149**	451.7111
N	**1196.5663**	**598.7868**	746.3138	373.6605
G	**1253.5878**	**627.2975**	632.2708	316.6391
I	**1366.6718**	**683.8396**	575.2494	288.1283
P	1463.7246	**732.3659**	**462.1653**	231.5863
C[CAM]	1623.7552	**812.3813**	**365.1125**	183.0599
G	1680.7767	**840.8920**	**205.0819**	103.0446
E	1809.8193	905.4133	**148.0604**	74.5339

Supplementary spectrum E3c: MS/MS of *m/z* 609.94 (mass 1826.81 Da) showing sequence ion coverage. The derived sequence is CKSKVCYRNGIPCGE.

Chymotrypsin digest:

RNGIPCGESCVFIPCITGIAGCSCKSKVCY

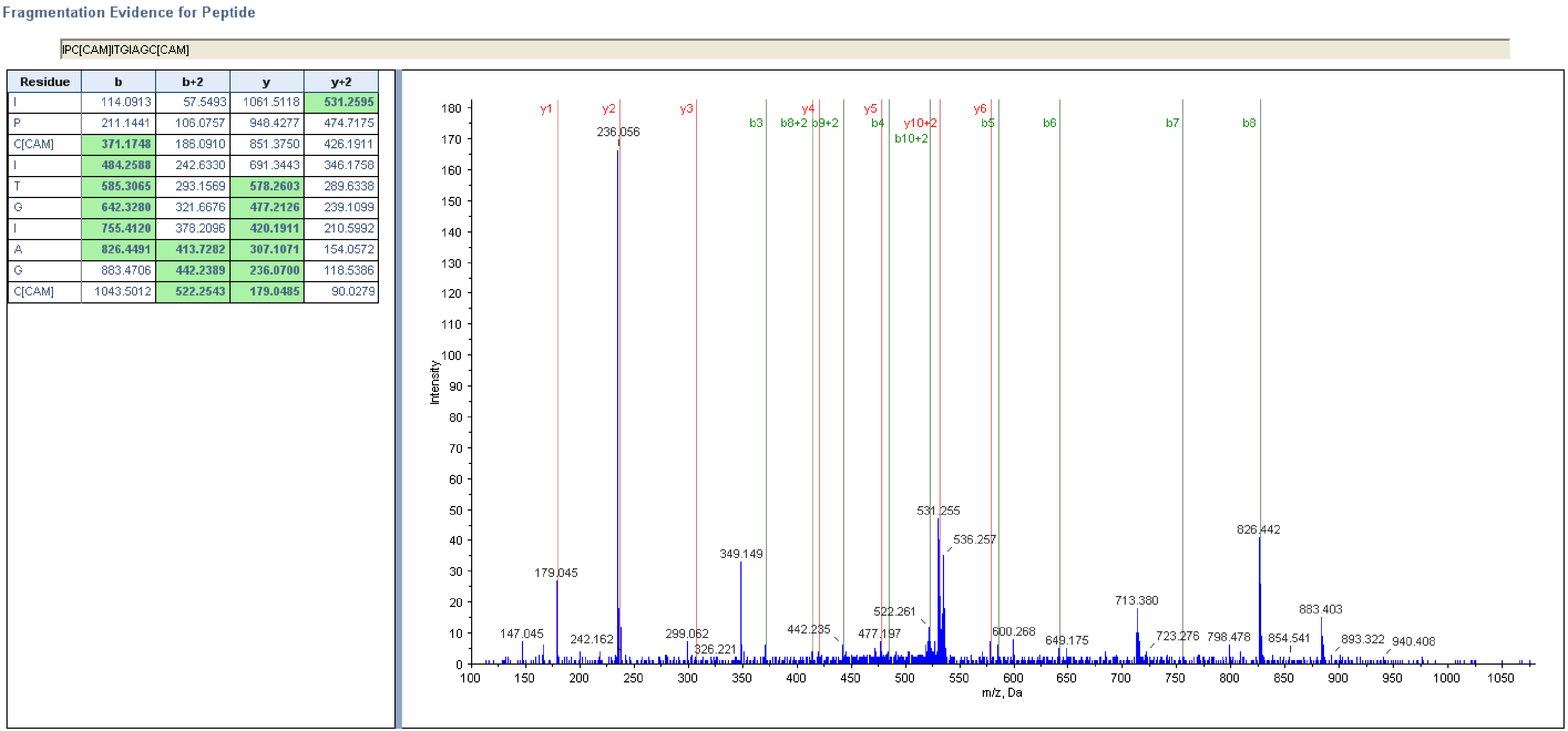


Fragmentation Evidence for Peptide

RNGIPC[CAM]GESC[CAM]VF

Residue	b	b+2	y	y+2
R	157.1084	79.0578	1395.6144	**698.3108**
N	**271.1513**	136.0793	1239.5133	620.2603
G	**328.1728**	164.5900	1125.4703	563.2388
I	**441.2568**	221.1321	1068.4489	534.7281
P	538.3096	269.6584	**955.3648**	478.1860
C[CAM]	**698.3403**	349.6738	858.3121	429.6597
G	**755.3617**	378.1845	**698.2814**	349.6443
E	**884.4043**	**442.7058**	641.2599	321.1336
S	**971.4363**	**486.2218**	**512.2173**	256.6123
C[CAM]	**1131.4670**	**566.2371**	**425.1853**	213.0963
V	**1230.5354**	**615.7713**	**265.1547**	**133.0810**
F	1377.6038	689.3055	**166.0863**	83.5468

Supplementary spectrum C3a: MS/MS of *m/z* 698.31 (mass 1394.60 Da) showing sequence ion coverage. The derived sequence is RNGIPCGESCVF.



Fragmentation Evidence for Peptide

IPC[CAM]ITGIAGC[CAM]

Residue	b	b+2	y	y+2
I	114.0913	57.5493	1061.5118	**531.2595**
P	211.1441	106.0757	948.4277	474.7175
C[CAM]	**371.1748**	186.0910	851.3750	426.1911
I	**484.2588**	242.6330	691.3443	346.1758
T	**585.3065**	293.1569	**578.2603**	289.6338
G	**642.3280**	321.6676	**477.2126**	239.1099
I	**755.4120**	378.2096	**420.1911**	210.5992
A	**826.4491**	**413.7282**	**307.1071**	154.0572
G	883.4706	**442.2389**	**236.0700**	118.5386
C[CAM]	1043.5012	**522.2543**	**179.0485**	90.0279

Supplementary spectrum C3b: MS/MS of *m/z* 531.25 (mass 1060.49 Da) showing sequence ion coverage. The derived sequence is IPCITGIAGC.

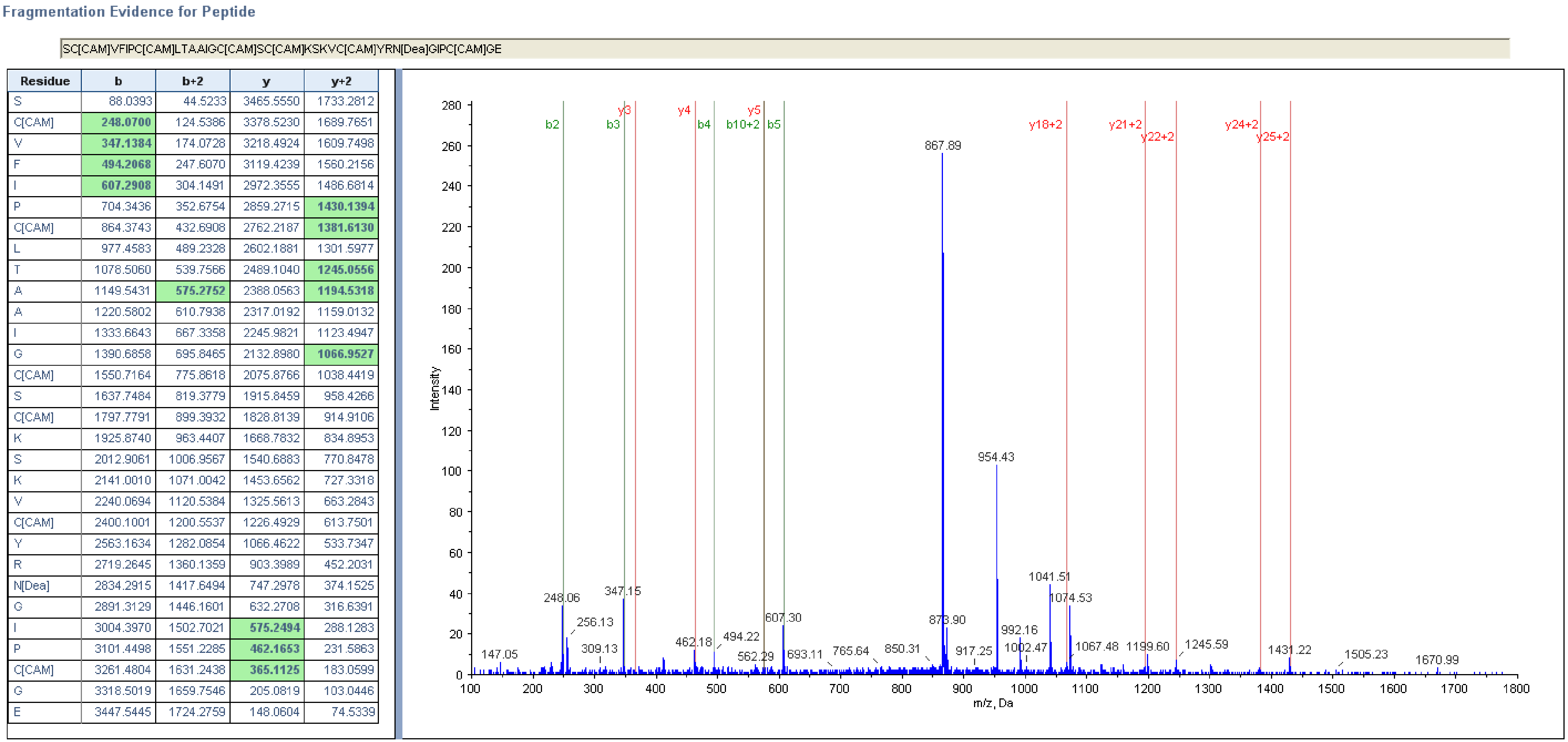
Evidence for Cter P:

Endoproteinase Glu-C digest:

SCVFIPCITAAIGCSCKSKVCYRNGIPCGE

NB: This peptide shows 96% sequence homology to cyclotide A (D2WPK4) from *Gloeospermum blakeanum*.

```
1     GIPCGESCVFIPCITAAIGCSCKSKVCYRN   30  Cter_P
1     GIPCGESCVFIPCITAAIGCSCKTKVCYRN   30  Cyclotide-A
      ***********************:******
```

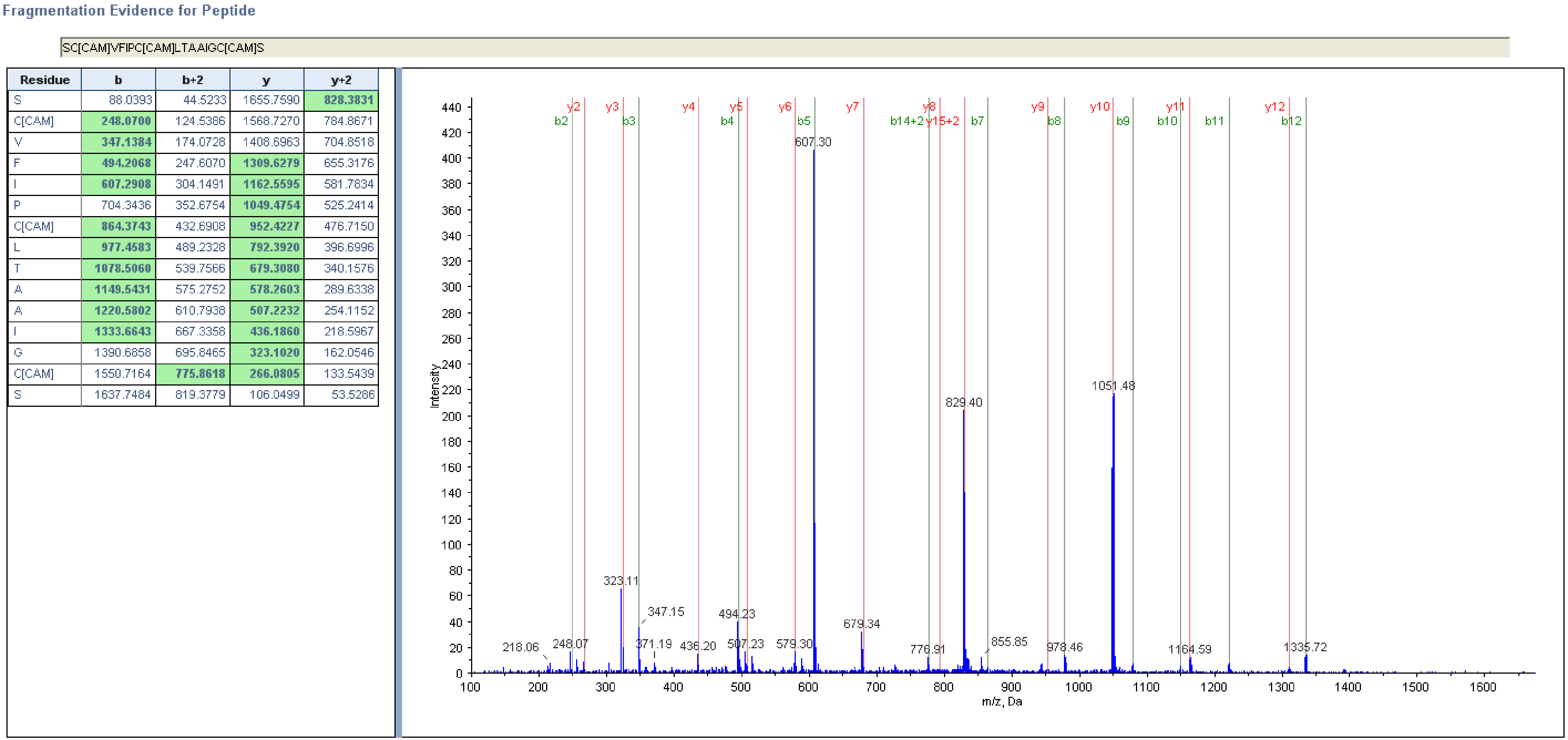


Fragmentation Evidence for Peptide

SC[CAM]VFIPC[CAM]LTAAIGC[CAM]SC[CAM]KSKVC[CAM]YRN[Dea]GIPC[CAM]GE

Residue	b	b+2	y	y+2
S	88.0393	44.5233	3465.5550	1733.2812
C[CAM]	**248.0700**	124.5386	3378.5230	1689.7651
V	**347.1384**	174.0728	3218.4924	1609.7498
F	**494.2068**	247.6070	3119.4239	1560.2156
I	**607.2908**	304.1491	2972.3555	1486.6814
P	704.3436	352.6754	2859.2715	**1430.1394**
C[CAM]	864.3743	432.6908	2762.2187	**1381.6130**
L	977.4583	489.2328	2602.1881	1301.5977
T	1078.5060	539.7566	2489.1040	**1245.0556**
A	1149.5431	**575.2752**	2388.0563	**1194.5318**
A	1220.5802	610.7938	2317.0192	1159.0132
I	1333.6643	667.3358	2245.9821	1123.4947
G	1390.6858	695.8465	2132.8980	**1066.9527**
C[CAM]	1550.7164	775.8618	2075.8766	1038.4419
S	1637.7484	819.3779	1915.8459	958.4266
C[CAM]	1797.7791	899.3932	1828.8139	914.9106
K	1925.8740	963.4407	1668.7832	834.8953
S	2012.9061	1006.9567	1540.6883	770.8478
K	2141.0010	1071.0042	1453.6562	727.3318
V	2240.0694	1120.5384	1325.5613	663.2843
C[CAM]	2400.1001	1200.5537	1226.4929	613.7501
Y	2563.1634	1282.0854	1066.4622	533.7347
R	2719.2645	1360.1359	903.3989	452.2031
N[Dea]	2834.2915	1417.6494	747.2978	374.1525
G	2891.3129	1446.1601	632.2708	316.6391
I	3004.3970	1502.7021	**575.2494**	288.1283
P	3101.4498	1551.2285	**462.1653**	231.5863
C[CAM]	3261.4804	1631.2438	**365.1125**	183.0599
G	3318.5019	1659.7546	205.0819	103.0446
E	3447.5445	1724.2759	148.0604	74.5339

Supplementary spectrum E4a: MS/MS of *m/z* 866.90 (mass 3463.58 Da) showing sequence ion coverage. The derived sequence is SCVFIPCITAAIGCSCKSKVCYRNGIPCGE.

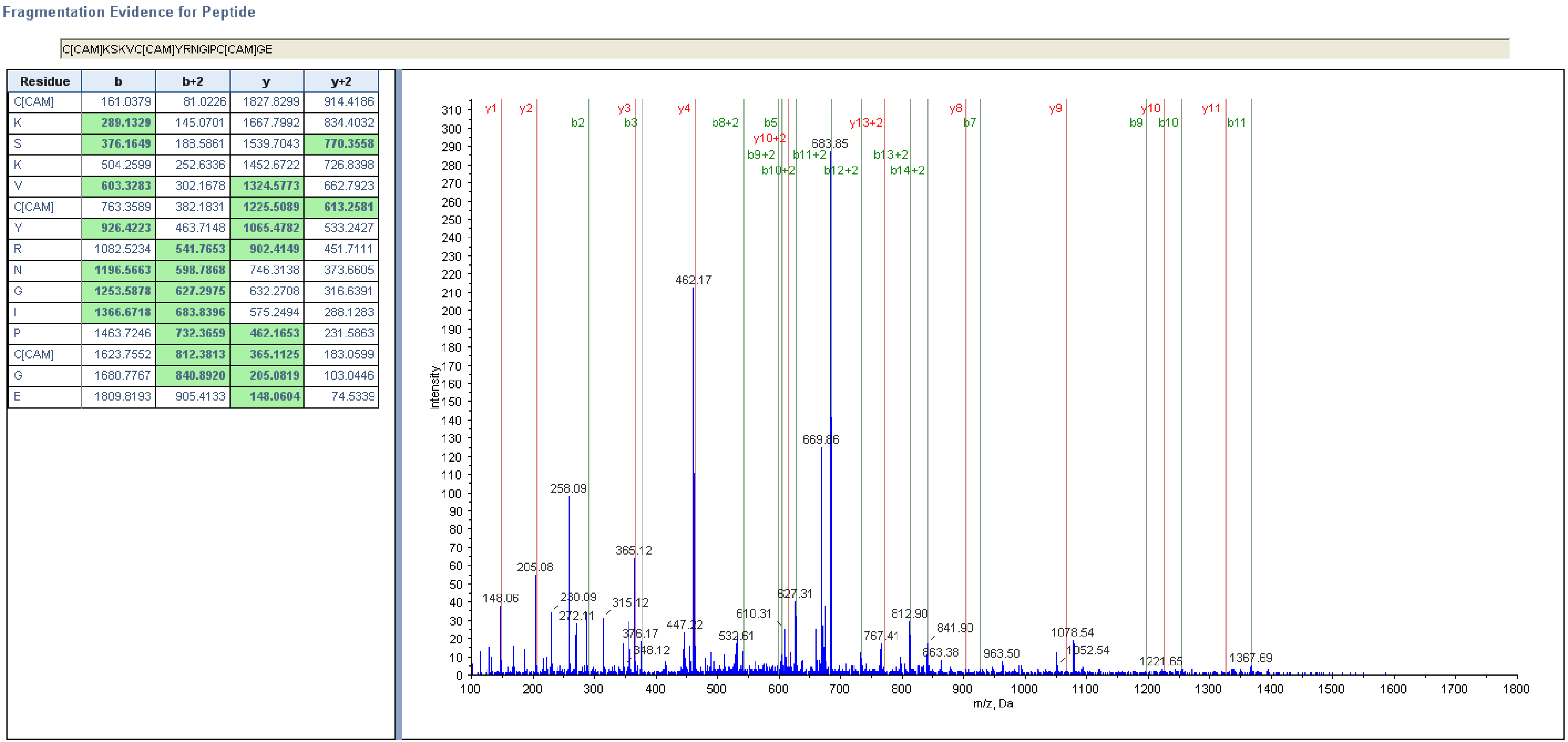


Fragmentation Evidence for Peptide

SC[CAM]VFIPC[CAM]LTAAIGC[CAM]S

Residue	b	b+2	y	y+2
S	88.0393	44.5233	1655.7590	**828.3831**
C[CAM]	**248.0700**	124.5386	1568.7270	784.8671
V	**347.1384**	174.0728	1408.6963	704.8518
F	**494.2068**	247.6070	**1309.6279**	655.3176
I	**607.2908**	304.1491	**1162.5595**	581.7834
P	704.3436	352.6754	**1049.4754**	525.2414
C[CAM]	**864.3743**	432.6908	**952.4227**	476.7150
L	**977.4583**	489.2328	**792.3920**	396.6996
T	**1078.5060**	539.7566	**679.3080**	340.1576
A	**1149.5431**	575.2752	**578.2603**	289.6338
A	**1220.5802**	610.7938	**507.2232**	254.1152
I	**1333.6643**	667.3358	**436.1860**	218.5967
G	1390.6858	695.8465	**323.1020**	162.0546
C[CAM]	1550.7164	**775.8618**	**266.0805**	133.5439
S	1637.7484	819.3779	106.0499	53.5286

Supplementary spectrum E4b: MS/MS of *m/z* 828.39 (mass 1654.78 Da) showing sequence ion coverage. The derived sequence is SCVFIPCITAAIGCS.



Fragmentation Evidence for Peptide

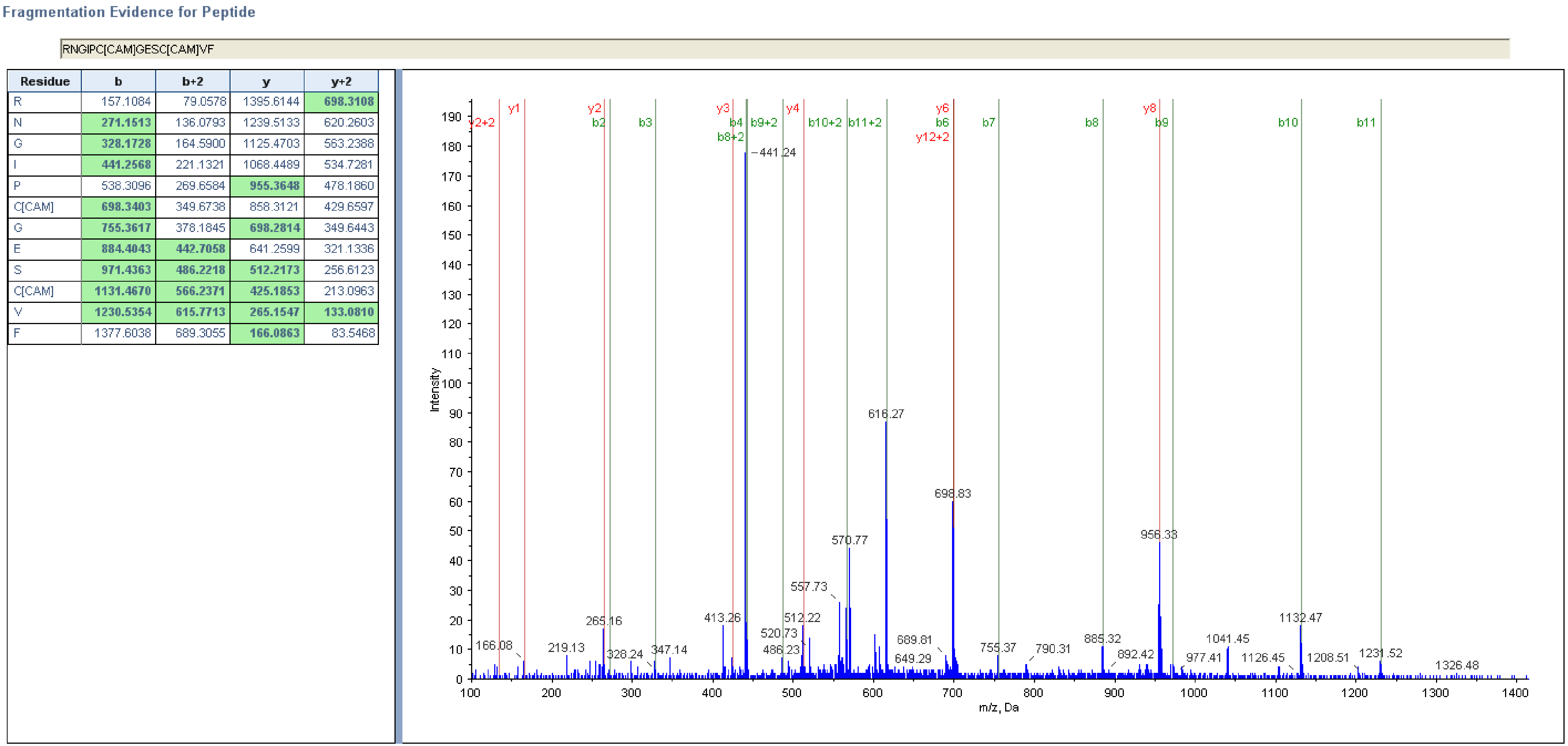
C[CAM]KSKVC[CAM]YRNGIPC[CAM]GE

Residue	b	b+2	y	y+2
C[CAM]	161.0379	81.0226	1827.8299	914.4186
K	**289.1329**	145.0701	1667.7992	834.4032
S	**376.1649**	188.5861	1539.7043	**770.3558**
K	504.2599	252.6336	1452.6722	726.8398
V	**603.3283**	302.1678	**1324.5773**	662.7923
C[CAM]	763.3589	382.1831	**1225.5089**	**613.2581**
Y	**926.4223**	463.7148	**1065.4782**	533.2427
R	1082.5234	**541.7653**	**902.4149**	451.7111
N	**1196.5663**	**598.7868**	746.3138	373.6605
G	**1253.5878**	**627.2975**	632.2708	316.6391
I	**1366.6718**	**683.8396**	575.2494	288.1283
P	1463.7246	**732.3659**	**462.1653**	231.5863
C[CAM]	1623.7552	**812.3813**	**365.1125**	183.0599
G	1680.7767	**840.8920**	**205.0819**	103.0446
E	1809.8193	905.4133	**148.0604**	74.5339

Supplementary spectrum E4c: MS/MS of *m/z* 609.94 (mass 1826.81 Da) showing sequence ion coverage. The derived sequence is CKSKVCYRNGIPCGE.

Chymotrypsin digest:

RNGIPCGESCVFIPCLTAAIGCSCKSKVCY

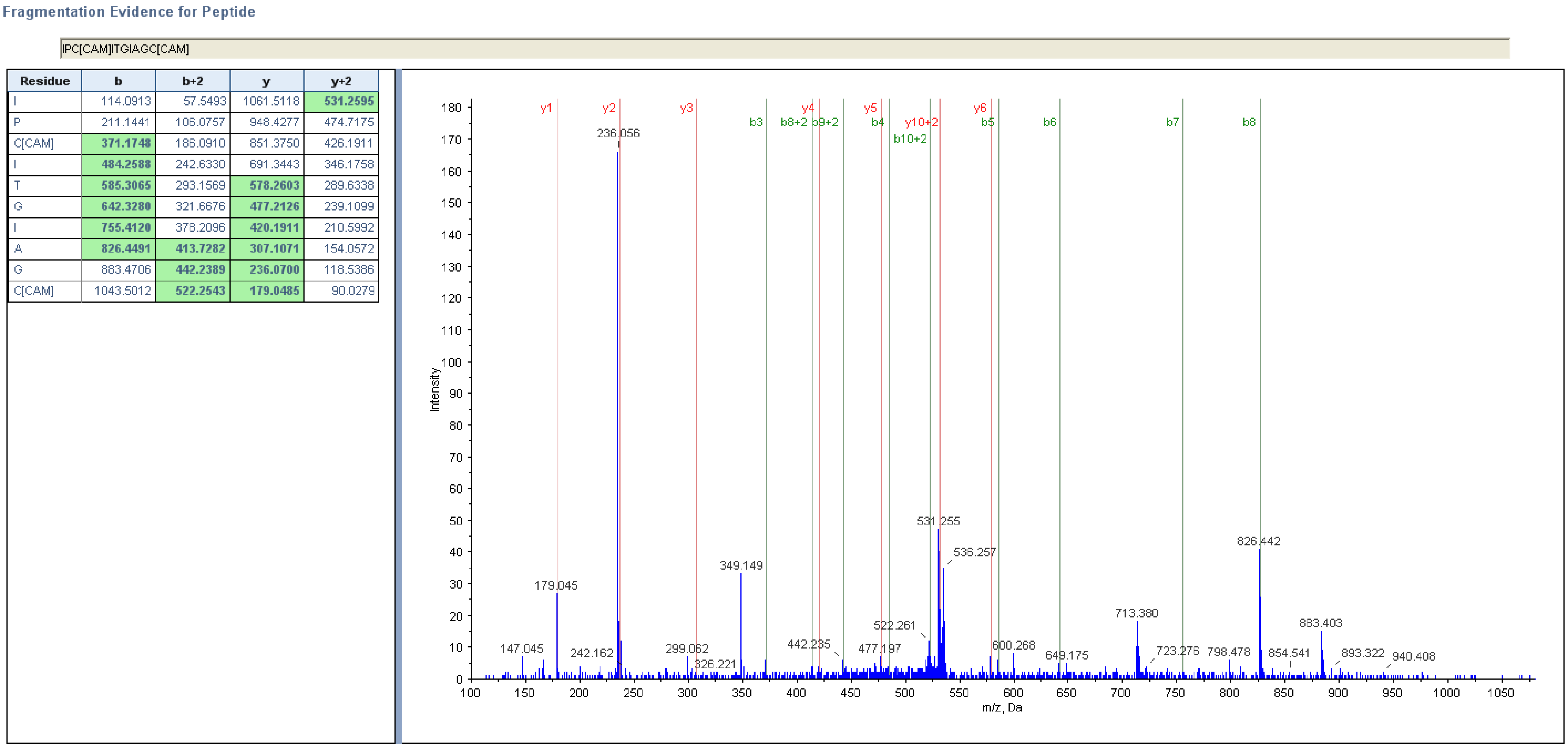


Fragmentation Evidence for Peptide

RNGIPC[CAM]GESC[CAM]VF

Residue	b	b+2	y	y+2
R	157.1084	79.0578	1395.6144	**698.3108**
N	**271.1513**	136.0793	1239.5133	620.2603
G	**328.1728**	164.5900	1125.4703	563.2388
I	**441.2568**	221.1321	1068.4489	534.7281
P	538.3096	269.6584	**955.3648**	478.1860
C[CAM]	**698.3403**	349.6738	858.3121	429.6597
G	**755.3617**	378.1845	**698.2814**	349.6443
E	**884.4043**	**442.7058**	641.2599	321.1336
S	**971.4363**	**486.2218**	**512.2173**	256.6123
C[CAM]	**1131.4670**	**566.2371**	**425.1853**	213.0963
V	**1230.5354**	**615.7713**	**265.1547**	**133.0810**
F	1377.6038	689.3055	**166.0863**	83.5468

Supplementary spectrum C4a: MS/MS of *m/z* 698.31 (mass 1394.60 Da) showing sequence ion coverage. The derived sequence is RNGIPCGESCVF.



Fragmentation Evidence for Peptide

IPC[CAM]ITGIAGC[CAM]

Residue	b	b+2	y	y+2
I	114.0913	57.5493	1061.5118	**531.2595**
P	211.1441	106.0757	948.4277	474.7175
C[CAM]	**371.1748**	186.0910	851.3750	426.1911
I	**484.2588**	242.6330	691.3443	346.1758
T	**585.3065**	293.1569	**578.2603**	289.6338
G	**642.3280**	321.6676	**477.2126**	239.1099
I	**755.4120**	378.2096	**420.1911**	210.5992
A	**826.4491**	**413.7282**	**307.1071**	154.0572
G	883.4706	**442.2389**	**236.0700**	118.5386
C[CAM]	1043.5012	**522.2543**	**179.0485**	90.0279

Supplementary spectrum C4b: MS/MS of *m/z* 531.25 (mass 1060.49 Da) showing sequence ion coverage. The derived sequence is IPCITGIAGC.

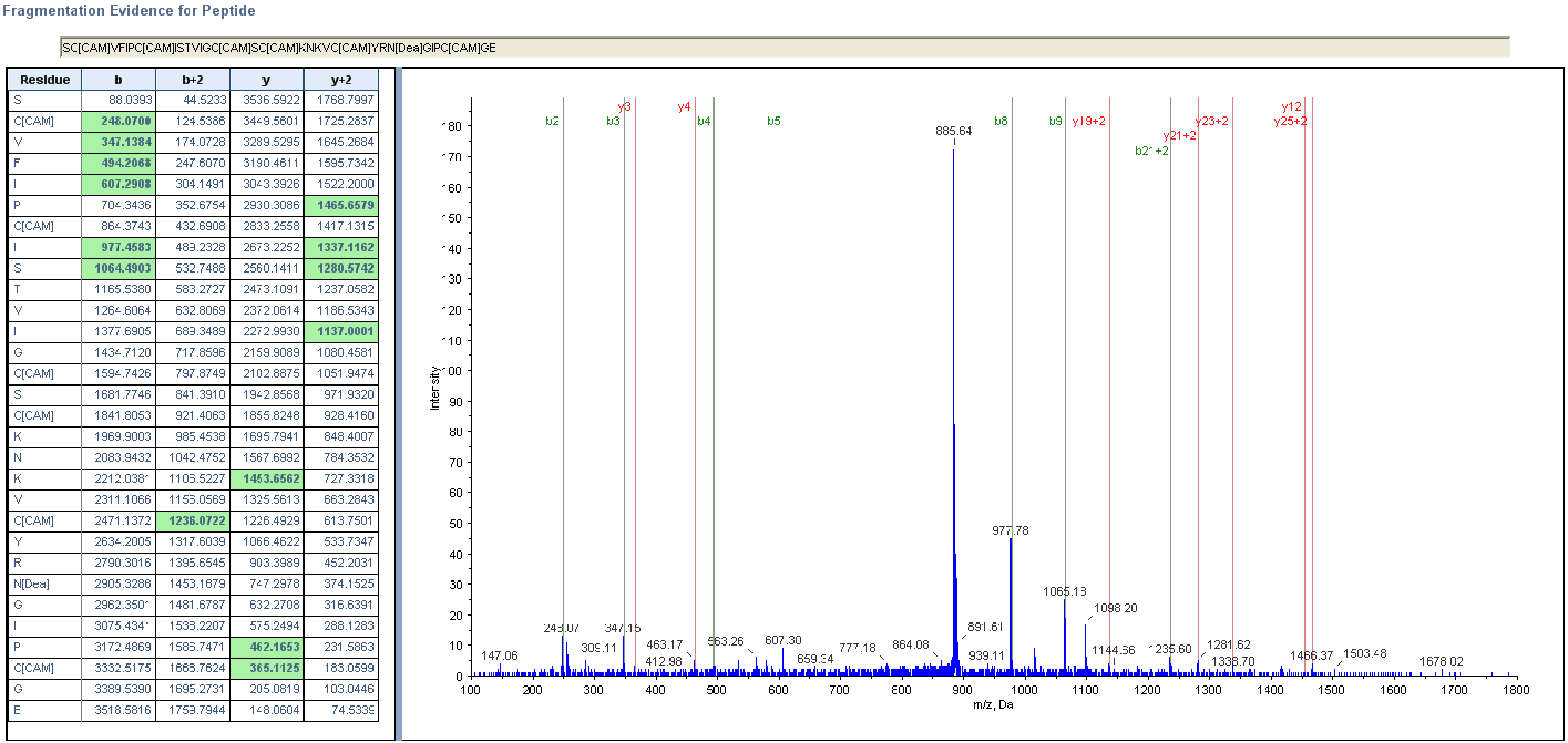
Evidence for Cter Q:

Endoproteinase Glu-C digest:

SCVFIPCISTVIGCSCKNKVCYRNGIPCGE

NB: This peptide shows 93% sequence homology to cyclotide I (Q30CB5) from *Hybanthus floribundus.*

```
1     GIPCGESCVFIPCISTVIGCSCKNKVCYRN   30  Cter_Q
1     GIPCGESCVFIPCISGVIGCSCKSKVCYRN   30  Cyclotide-I
      *************** *******.******
```

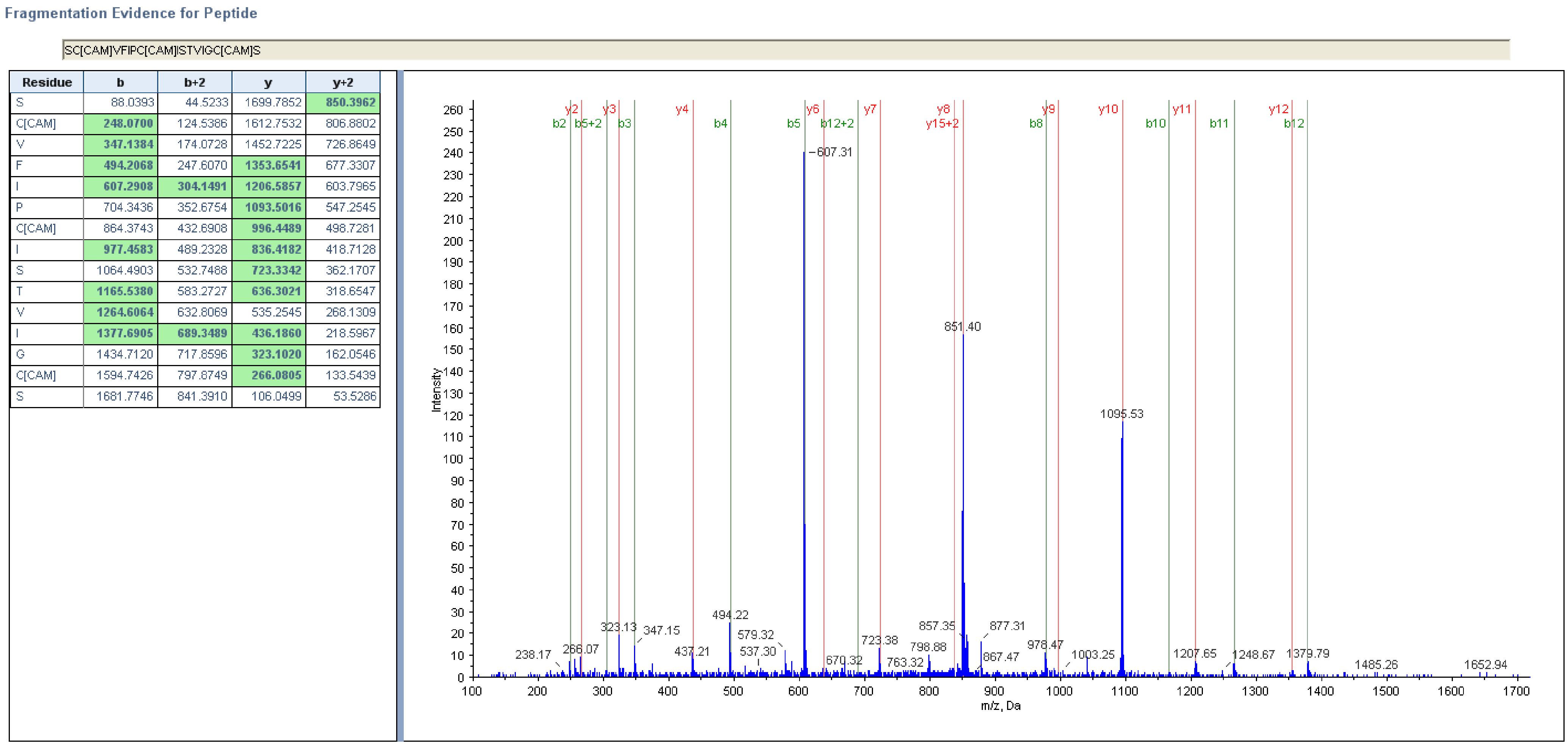


Fragmentation Evidence for Peptide

SC[CAM]VFIPC[CAM]ISTVIGC[CAM]SC[CAM]KNKVC[CAM]YRN[Dea]GIPC[CAM]GE

Residue	b	b+2	y	y+2
S	88.0393	44.5233	3536.5922	1768.7997
C[CAM]	**248.0700**	124.5386	3449.5601	1725.2837
V	**347.1384**	174.0728	3289.5295	1645.2684
F	**494.2068**	247.6070	3190.4611	1595.7342
I	**607.2908**	304.1491	3043.3926	1522.2000
P	704.3436	352.6754	2930.3086	**1465.6579**
C[CAM]	864.3743	432.6908	2833.2558	1417.1315
I	**977.4583**	489.2328	2673.2252	**1337.1162**
S	**1064.4903**	532.7488	2560.1411	**1280.5742**
T	1165.5380	583.2727	2473.1091	1237.0582
V	1264.6064	632.8069	2372.0614	1186.5343
I	1377.6905	689.3489	2272.9930	**1137.0001**
G	1434.7120	717.8596	2159.9089	1080.4581
C[CAM]	1594.7426	797.8749	2102.8875	1051.9474
S	1681.7746	841.3910	1942.8568	971.9320
C[CAM]	1841.8053	921.4063	1855.8248	928.4160
K	1969.9003	985.4538	1695.7941	848.4007
N	2083.9432	1042.4752	1567.6992	784.3532
K	2212.0381	1106.5227	**1453.6562**	727.3318
V	2311.1066	1156.0569	1325.5613	663.2843
C[CAM]	2471.1372	**1236.0722**	1226.4929	613.7501
Y	2634.2005	1317.6039	1066.4622	533.7347
R	2790.3016	1395.6545	903.3989	452.2031
N[Dea]	2905.3286	1453.1679	747.2978	374.1525
G	2962.3501	1481.6787	632.2708	316.6391
I	3075.4341	1538.2207	575.2494	288.1283
P	3172.4869	1586.7471	**462.1653**	231.5863
C[CAM]	3332.5175	1666.7624	**365.1125**	183.0599
G	3389.5390	1695.2731	205.0819	103.0446
E	3518.5816	1759.7944	148.0604	74.5339

Supplementary spectrum E5a: MS/MS of *m/z* 884.60 (mass 3534.61 Da) showing sequence ion coverage. The derived sequence is SCVFIPCISTVIGCSCKNKVCYRNGIPCGE.

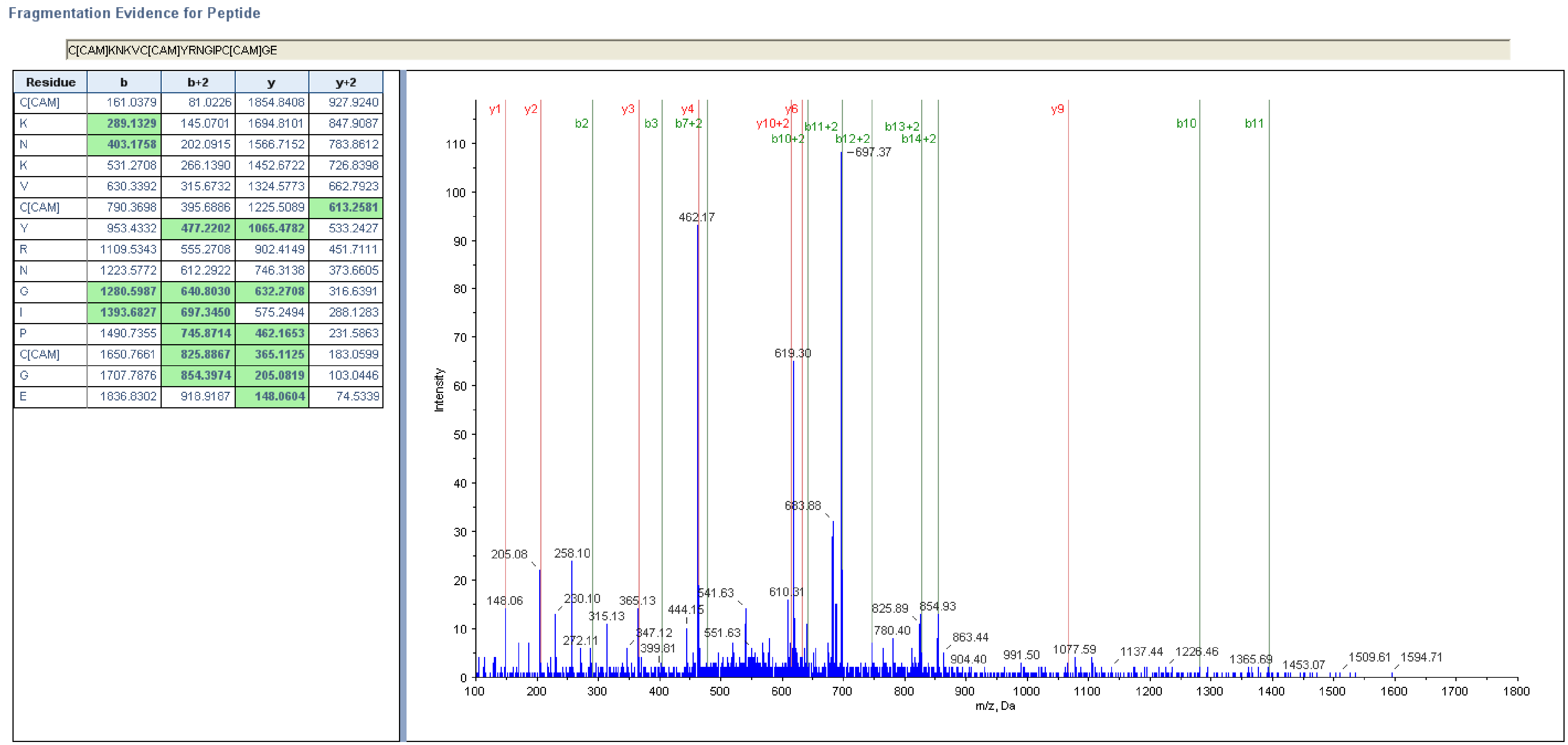


Fragmentation Evidence for Peptide

SC[CAM]VFIPC[CAM]ISTVIGC[CAM]S

Residue	b	b+2	y	y+2
S	88.0393	44.5233	1699.7852	**850.3962**
C[CAM]	**248.0700**	124.5386	1612.7532	806.8802
V	**347.1384**	174.0728	1452.7225	726.8649
F	**494.2068**	247.6070	**1353.6541**	677.3307
I	**607.2908**	**304.1491**	**1206.5857**	603.7965
P	704.3436	352.6754	**1093.5016**	547.2545
C[CAM]	864.3743	432.6908	**996.4489**	498.7281
I	**977.4583**	489.2328	**836.4182**	418.7128
S	1064.4903	532.7488	**723.3342**	362.1707
T	**1165.5380**	583.2727	**636.3021**	318.6547
V	**1264.6064**	632.8069	535.2545	268.1309
I	**1377.6905**	**689.3489**	**436.1860**	218.5967
G	1434.7120	717.8596	**323.1020**	162.0546
C[CAM]	1594.7426	797.8749	**266.0805**	133.5439
S	1681.7746	841.3910	106.0499	53.5286

Supplementary spectrum E5b: MS/MS of *m/z* 850.41 (mass 1698.81 Da) showing sequence ion coverage. The derived sequence is SCVFIPCISTVIGCS.



Fragmentation Evidence for Peptide

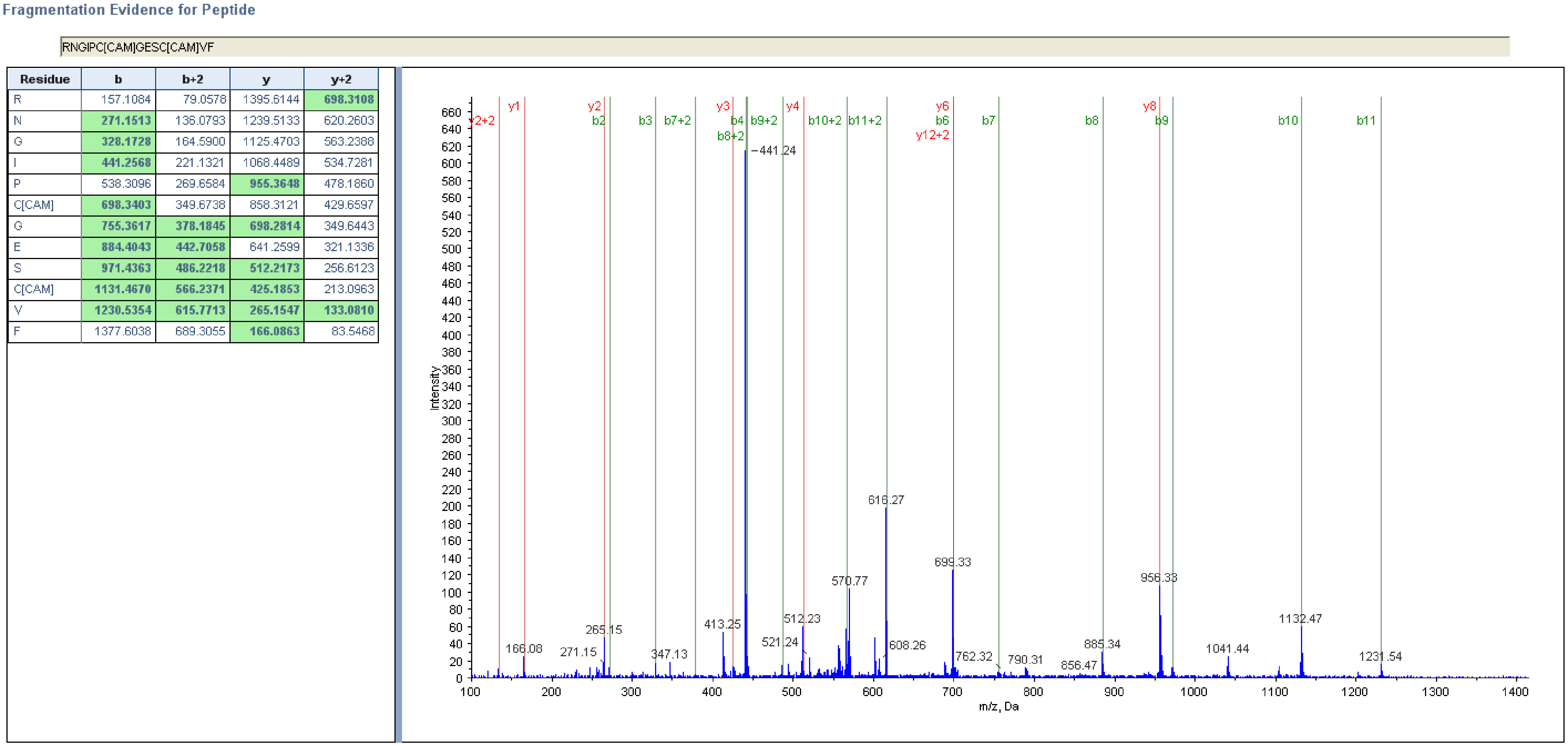
C[CAM]KNKVC[CAM]YRNGIPC[CAM]GE

Residue	b	b+2	y	y+2
C[CAM]	161.0379	81.0226	1854.8408	927.9240
K	**289.1329**	145.0701	1694.8101	847.9087
N	**403.1758**	202.0915	1566.7152	783.8612
K	531.2708	266.1390	1452.6722	726.8398
V	630.3392	315.6732	1324.5773	662.7923
C[CAM]	790.3698	395.6886	1225.5089	**613.2581**
Y	953.4332	**477.2202**	**1065.4782**	533.2427
R	1109.5343	555.2708	902.4149	451.7111
N	1223.5772	612.2922	746.3138	373.6605
G	**1280.5987**	**640.8030**	**632.2708**	316.6391
I	**1393.6827**	**697.3450**	575.2494	288.1283
P	1490.7355	**745.8714**	**462.1653**	231.5863
C[CAM]	1650.7661	**825.8867**	**365.1125**	183.0599
G	1707.7876	**854.3974**	**205.0819**	103.0446
E	1836.8302	918.9187	**148.0604**	74.5339

Supplementary spectrum E5c: MS/MS of *m/z* 618.95 (mass 1853.84 Da) showing sequence ion coverage. The derived sequence is CKNKVCYRNGIPCGE.

Chymotrypsin digest:

RNGIPCGESCVFIPCISTVIGCSCKNKVCY

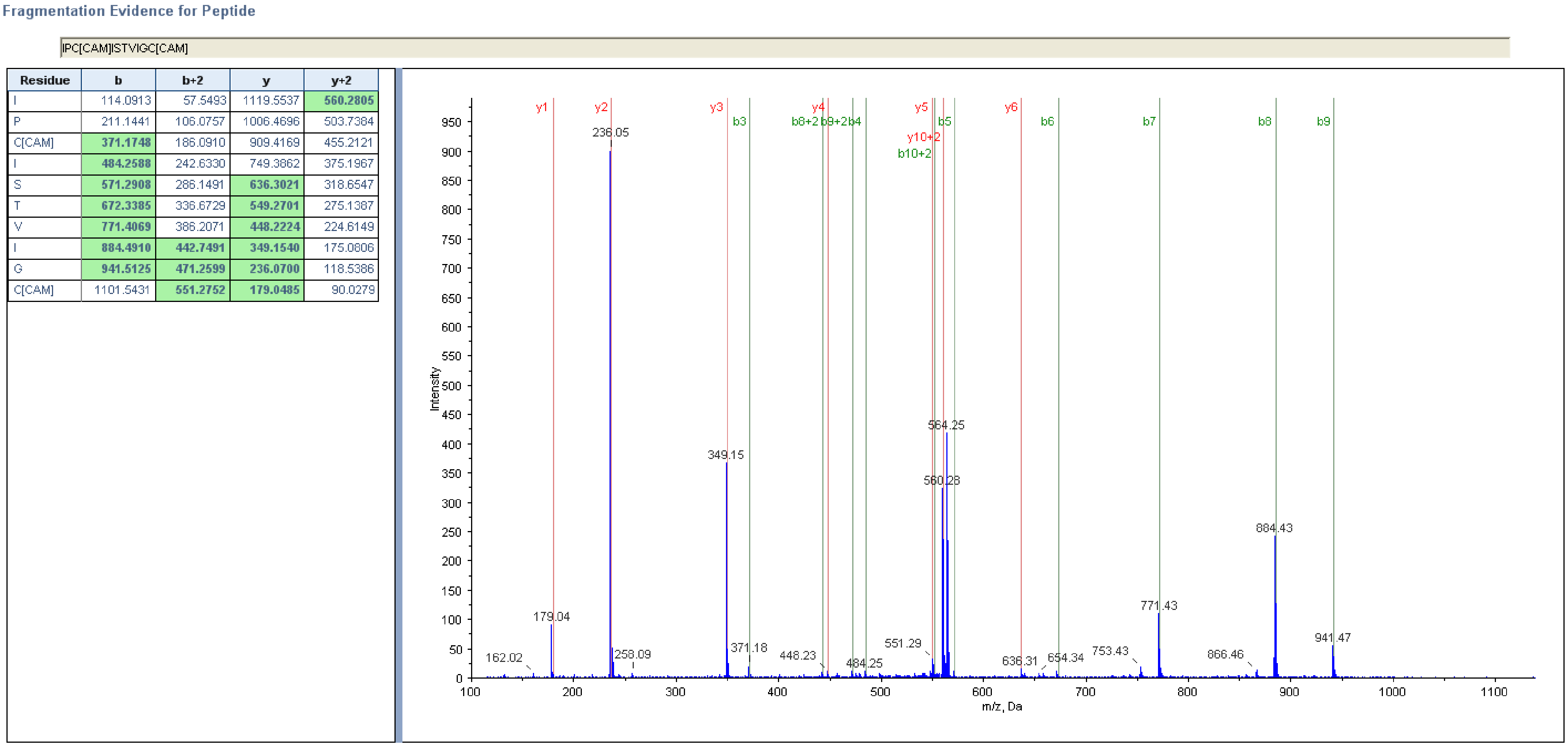


Fragmentation Evidence for Peptide

RNGIPC[CAM]GESC[CAM]VF

Residue	b	b+2	y	y+2
R	157.1084	79.0578	1395.6144	**698.3108**
N	**271.1513**	136.0793	1239.5133	620.2603
G	**328.1728**	164.5900	1125.4703	563.2388
I	**441.2568**	221.1321	1068.4489	534.7281
P	538.3096	269.6584	**955.3648**	478.1860
C[CAM]	**698.3403**	349.6738	858.3121	429.6597
G	**755.3617**	**378.1845**	**698.2814**	349.6443
E	**884.4043**	**442.7058**	641.2599	321.1336
S	**971.4363**	**486.2218**	**512.2173**	256.6123
C[CAM]	**1131.4670**	**566.2371**	**425.1853**	213.0963
V	**1230.5354**	**615.7713**	**265.1547**	**133.0810**
F	1377.6038	689.3055	**166.0863**	83.5468

Supplementary spectrum C5a: MS/MS of *m/z* 698.31 (mass 1394.60 Da) showing sequence ion coverage. The derived sequence is RNGIPCGESCVF.

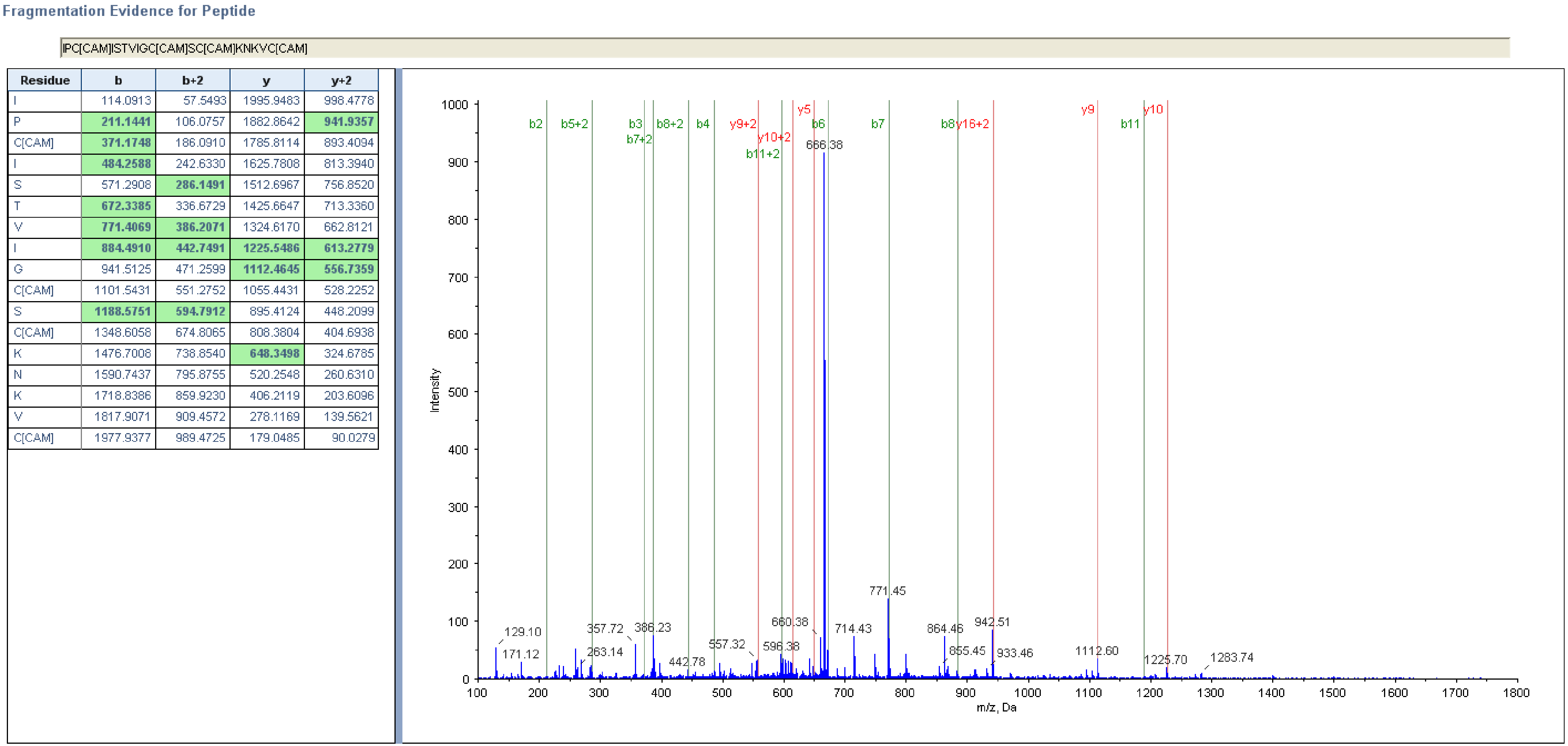


Fragmentation Evidence for Peptide

IPC[CAM]ISTVIGC[CAM]

Residue	b	b+2	y	y+2
I	114.0913	57.5493	1119.5537	**560.2805**
P	211.1441	106.0757	1006.4696	503.7384
C[CAM]	**371.1748**	186.0910	909.4169	455.2121
I	**484.2588**	242.6330	749.3862	375.1967
S	**571.2908**	286.1491	**636.3021**	318.6547
T	**672.3385**	336.6729	**549.2701**	275.1387
V	**771.4069**	386.2071	**448.2224**	224.6149
I	**884.4910**	**442.7491**	**349.1540**	175.0806
G	**941.5125**	**471.2599**	**236.0700**	118.5386
C[CAM]	1101.5431	**551.2752**	**179.0485**	90.0279

Supplementary spectrum C5b: MS/MS of *m/z* 560.28 (mass 1118.54 Da) showing sequence ion coverage. The derived sequence is IPCISTVIGC.



Fragmentation Evidence for Peptide

IPC[CAM]ISTVIGC[CAM]SC[CAM]KNKVC[CAM]

Residue	b	b+2	y	y+2
I	114.0913	57.5493	1995.9483	998.4778
P	**211.1441**	106.0757	1882.8642	**941.9357**
C[CAM]	**371.1748**	186.0910	1785.8114	893.4094
I	**484.2588**	242.6330	1625.7808	813.3940
S	571.2908	**286.1491**	1512.6967	756.8520
T	**672.3385**	336.6729	1425.6647	713.3360
V	**771.4069**	**386.2071**	1324.6170	662.8121
I	**884.4910**	**442.7491**	**1225.5486**	**613.2779**
G	941.5125	471.2599	**1112.4645**	**556.7359**
C[CAM]	1101.5431	551.2752	1055.4431	528.2252
S	**1188.5751**	**594.7912**	895.4124	448.2099
C[CAM]	1348.6058	674.8065	808.3804	404.6938
K	1476.7008	738.8540	**648.3498**	324.6785
N	1590.7437	795.8755	520.2548	260.6310
K	1718.8386	859.9230	406.2119	203.6096
V	1817.9071	909.4572	278.1169	139.5621
C[CAM]	1977.9377	989.4725	179.0485	90.0279

Supplementary spectrum C5c: MS/MS of *m/z* 666.03 (mass 1995.09 Da) showing sequence ion coverage. The derived sequence is IPCISTVIGCSCKNKVC.

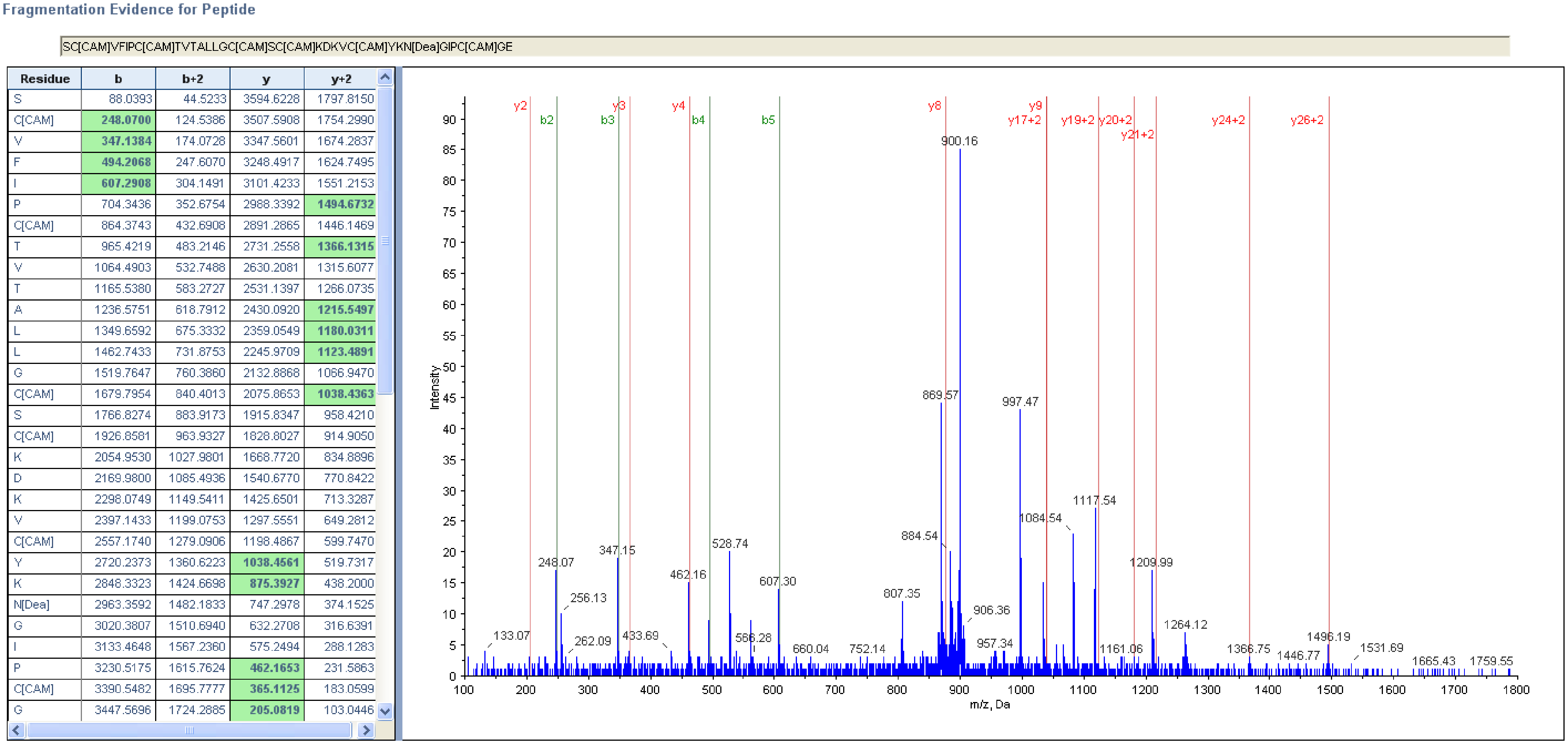
Evidence for Cter R:

Endoproteinase Glu-C digest:

SCVFIPCTVTALLGCSCKDKVCYKNGIPCGE

NB: This peptide shows 93% sequence homology to Cyclopsychotride-A (P56872) from *Psychotria longipes*.

```
1    GIPCGESCVFIPCTVTALLGCSCKDKVCYKN   31  Cter_R
1    SIPCGESCVFIPCTVTALLGCSCKSKVCYKN   31  Cyclopsychotride-A
     .***********************.******
```

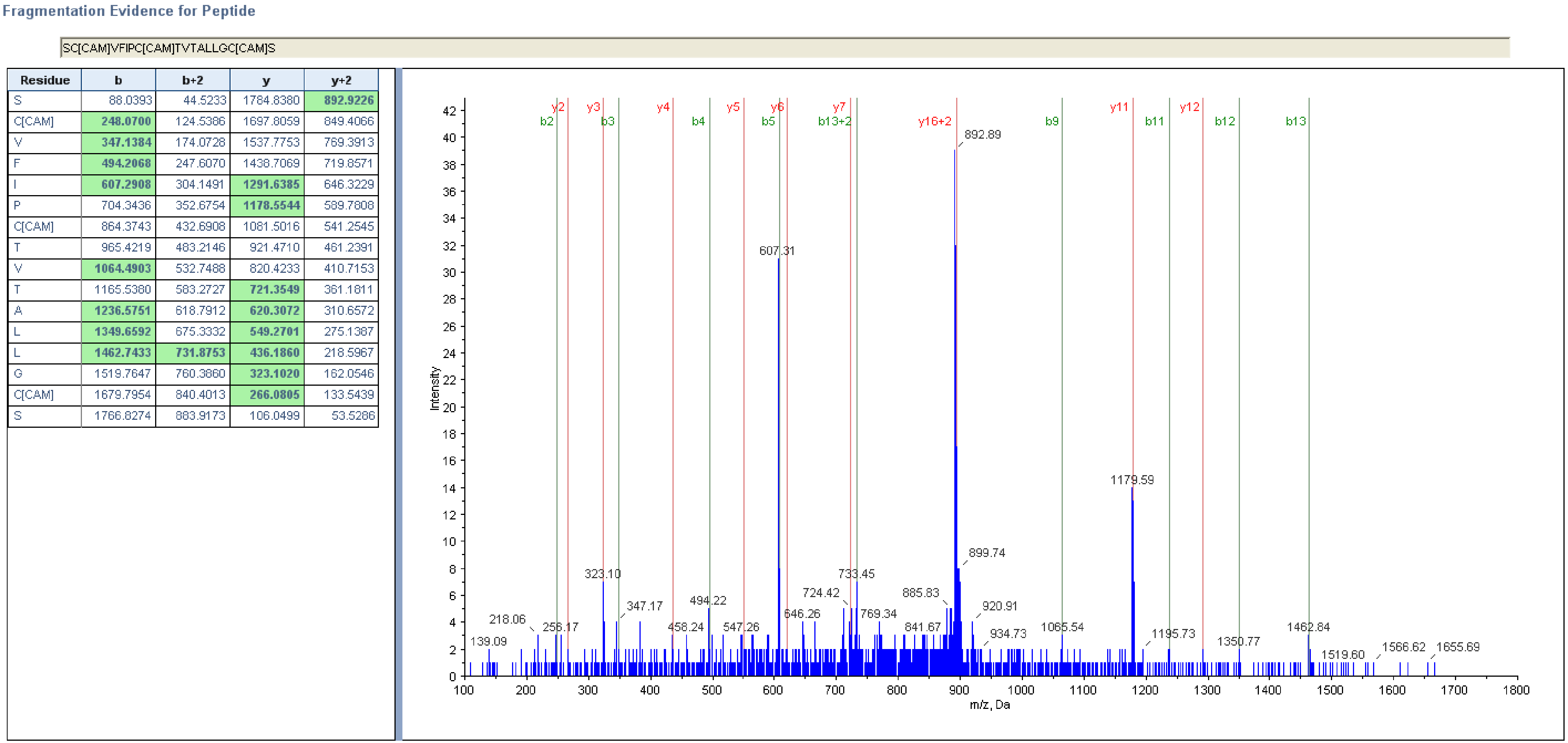


Fragmentation Evidence for Peptide

SC[CAM]VFIPC[CAM]TVTALLGC[CAM]SC[CAM]KDKVC[CAM]YKN[Dea]GIPC[CAM]GE

Residue	b	b+2	y	y+2
S	88.0393	44.5233	3594.6228	1797.8150
C[CAM]	**248.0700**	124.5386	3507.5908	1754.2990
V	**347.1384**	174.0728	3347.5601	1674.2837
F	**494.2068**	247.6070	3248.4917	1624.7495
I	**607.2908**	304.1491	3101.4233	1551.2153
P	704.3436	352.6754	2988.3392	**1494.6732**
C[CAM]	864.3743	432.6908	2891.2865	1446.1469
T	965.4219	483.2146	2731.2558	**1366.1315**
V	1064.4903	532.7488	2630.2081	1315.6077
T	1165.5380	583.2727	2531.1397	1266.0735
A	1236.5751	618.7912	2430.0920	**1215.5497**
L	1349.6592	675.3332	2359.0549	**1180.0311**
L	1462.7433	731.8753	2245.9709	**1123.4891**
G	1519.7647	760.3860	2132.8868	1066.9470
C[CAM]	1679.7954	840.4013	2075.8653	**1038.4363**
S	1766.8274	883.9173	1915.8347	958.4210
C[CAM]	1926.8581	963.9327	1828.8027	914.9050
K	2054.9530	1027.9801	1668.7720	834.8896
D	2169.9800	1085.4936	1540.6770	770.8422
K	2298.0749	1149.5411	1425.6501	713.3287
V	2397.1433	1199.0753	1297.5551	649.2812
C[CAM]	2557.1740	1279.0906	1198.4867	599.7470
Y	2720.2373	1360.6223	**1038.4561**	519.7317
K	2848.3323	1424.6698	**875.3927**	438.2000
N[Dea]	2963.3592	1482.1833	747.2978	374.1525
G	3020.3807	1510.6940	632.2708	316.6391
I	3133.4648	1567.2360	575.2494	288.1283
P	3230.5175	1615.7624	**462.1653**	231.5863
C[CAM]	3390.5482	1695.7777	**365.1125**	183.0599
G	3447.5696	1724.2885	**205.0819**	103.0446

Supplementary spectrum E6a: MS/MS of *m/z* 899.17 (mass 3592.66 Da) showing sequence ion coverage. The derived sequence is SCVFIPCTVTALLGCSCKDKVCYKNGIPCGE.

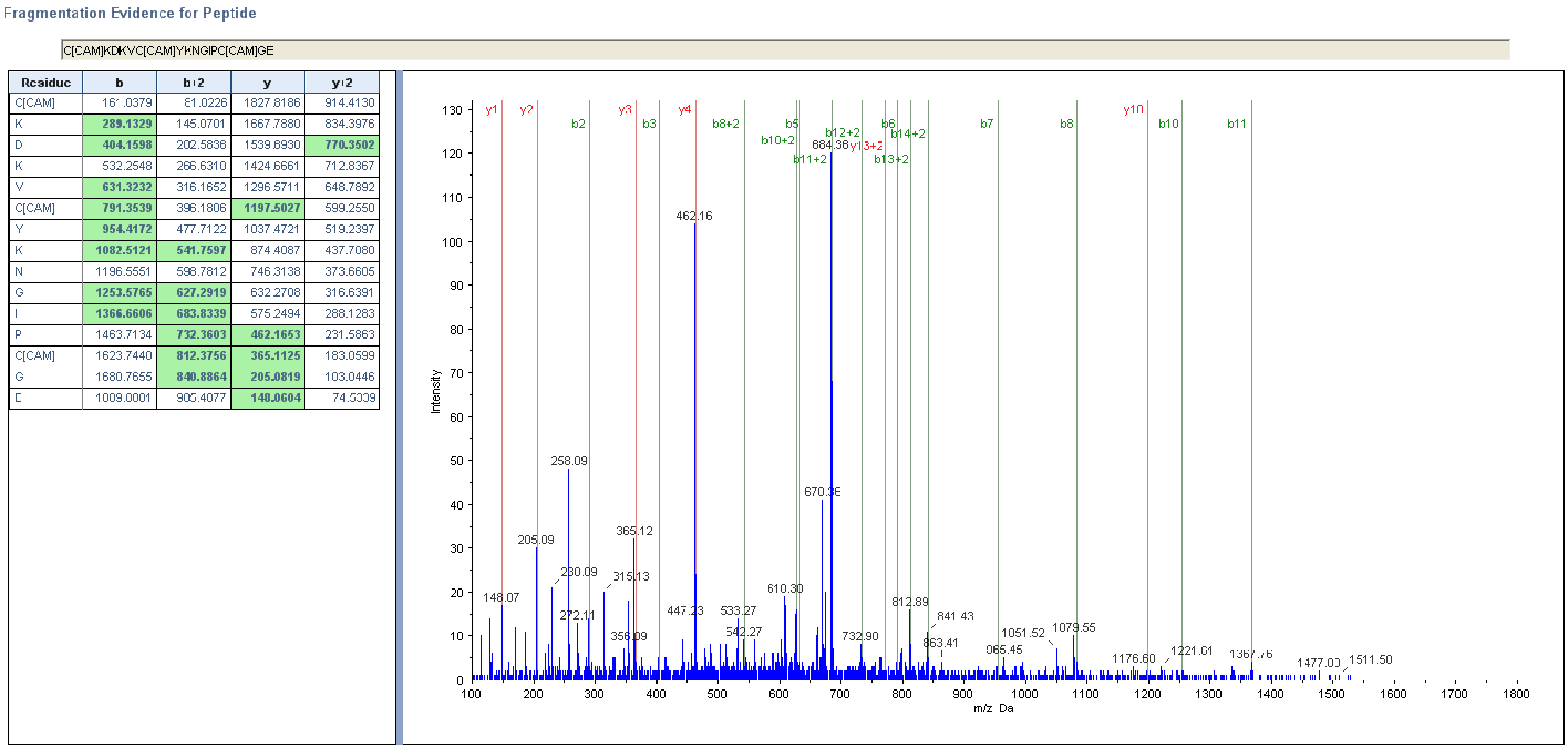


Fragmentation Evidence for Peptide

SC[CAM]VFIPC[CAM]TVTALLGC[CAM]S

Residue	b	b+2	y	y+2
S	88.0393	44.5233	1784.8380	**892.9226**
C[CAM]	**248.0700**	124.5386	1697.8059	849.4066
V	**347.1384**	174.0728	1537.7753	769.3913
F	**494.2068**	247.6070	1438.7069	719.8571
I	**607.2908**	304.1491	**1291.6385**	646.3229
P	704.3436	352.6754	**1178.5544**	589.7808
C[CAM]	864.3743	432.6908	1081.5016	541.2545
T	965.4219	483.2146	921.4710	461.2391
V	**1064.4903**	532.7488	820.4233	410.7153
T	1165.5380	583.2727	**721.3549**	361.1811
A	**1236.5751**	618.7912	**620.3072**	310.6572
L	**1349.6592**	675.3332	**549.2701**	275.1387
L	**1462.7433**	**731.8753**	**436.1860**	218.5967
G	1519.7647	760.3860	**323.1020**	162.0546
C[CAM]	1679.7954	840.4013	**266.0805**	133.5439
S	1766.8274	883.9173	106.0499	53.5286

Supplementary spectrum E6b: MS/MS of *m/z* 892.95 (mass 1783.88 Da) showing sequence ion coverage. The derived sequence is SCVFIPCTVTALLGCS.



Fragmentation Evidence for Peptide

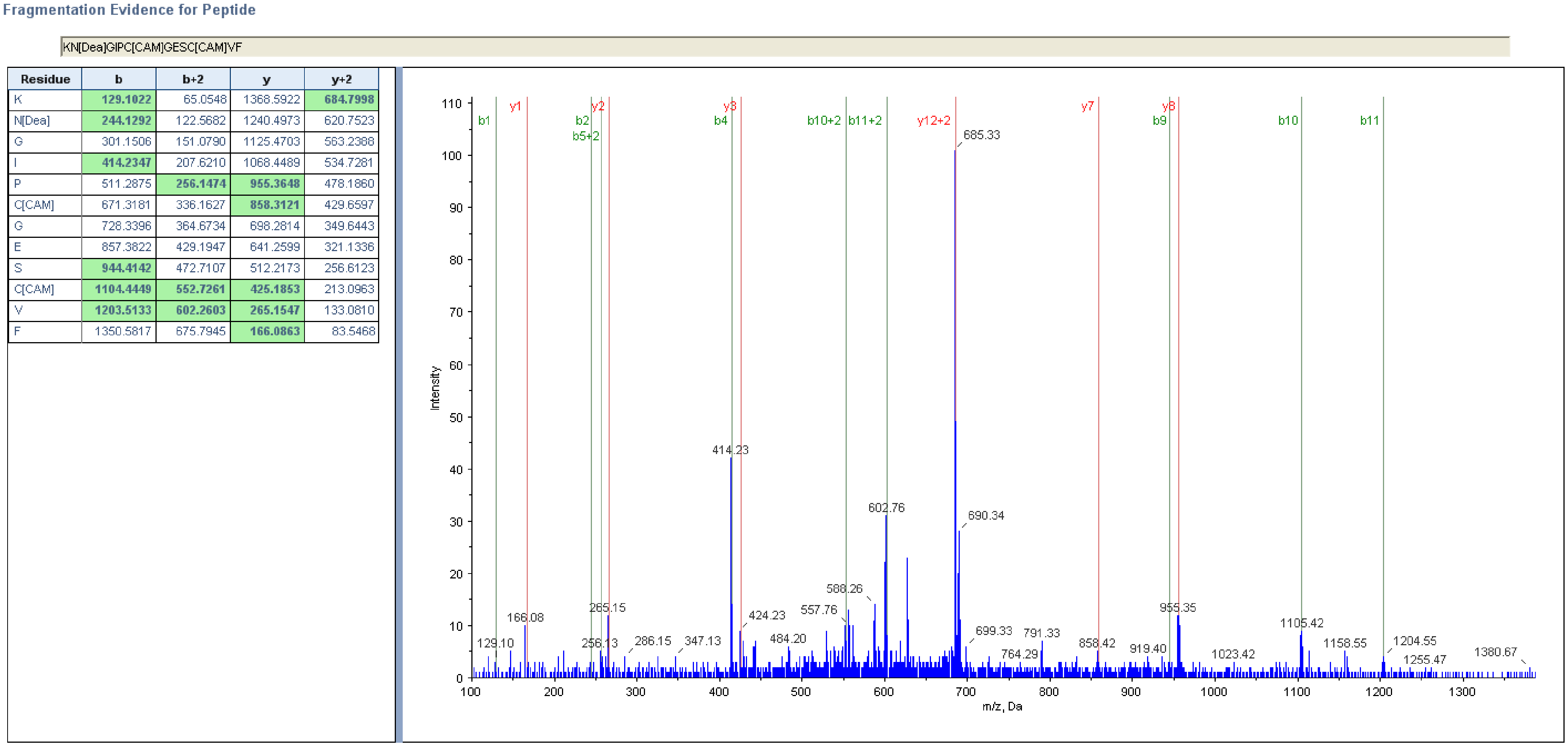
C[CAM]KDKVC[CAM]YKNGIPC[CAM]GE

Residue	b	b+2	y	y+2
C[CAM]	161.0379	81.0226	1827.8186	914.4130
K	**289.1329**	145.0701	1667.7880	834.3976
D	**404.1598**	202.5836	1539.6930	**770.3502**
K	532.2548	266.6310	1424.6661	712.8367
V	**631.3232**	316.1652	1296.5711	648.7892
C[CAM]	**791.3539**	396.1806	**1197.5027**	599.2550
Y	**954.4172**	477.7122	1037.4721	519.2397
K	**1082.5121**	**541.7597**	874.4087	437.7080
N	1196.5551	598.7812	746.3138	373.6605
G	**1253.5765**	**627.2919**	632.2708	316.6391
I	**1366.6606**	**683.8339**	575.2494	288.1283
P	1463.7134	**732.3603**	**462.1653**	231.5863
C[CAM]	1623.7440	**812.3756**	**365.1125**	183.0599
G	1680.7655	**840.8864**	**205.0819**	103.0446
E	1809.8081	905.4077	**148.0604**	74.5339

Supplementary spectrum E6c: MS/MS of *m/z* 609.95 (mass 1826.83 Da) showing sequence ion coverage. The derived sequence is CKDKVCYKNGIPCGE.

Chymotrypsin digest:

KNGIPCGESCVFIPCTVTALLGCSCKDKVCY



Fragmentation Evidence for Peptide

KN[Dea]GIPC[CAM]GESC[CAM]VF

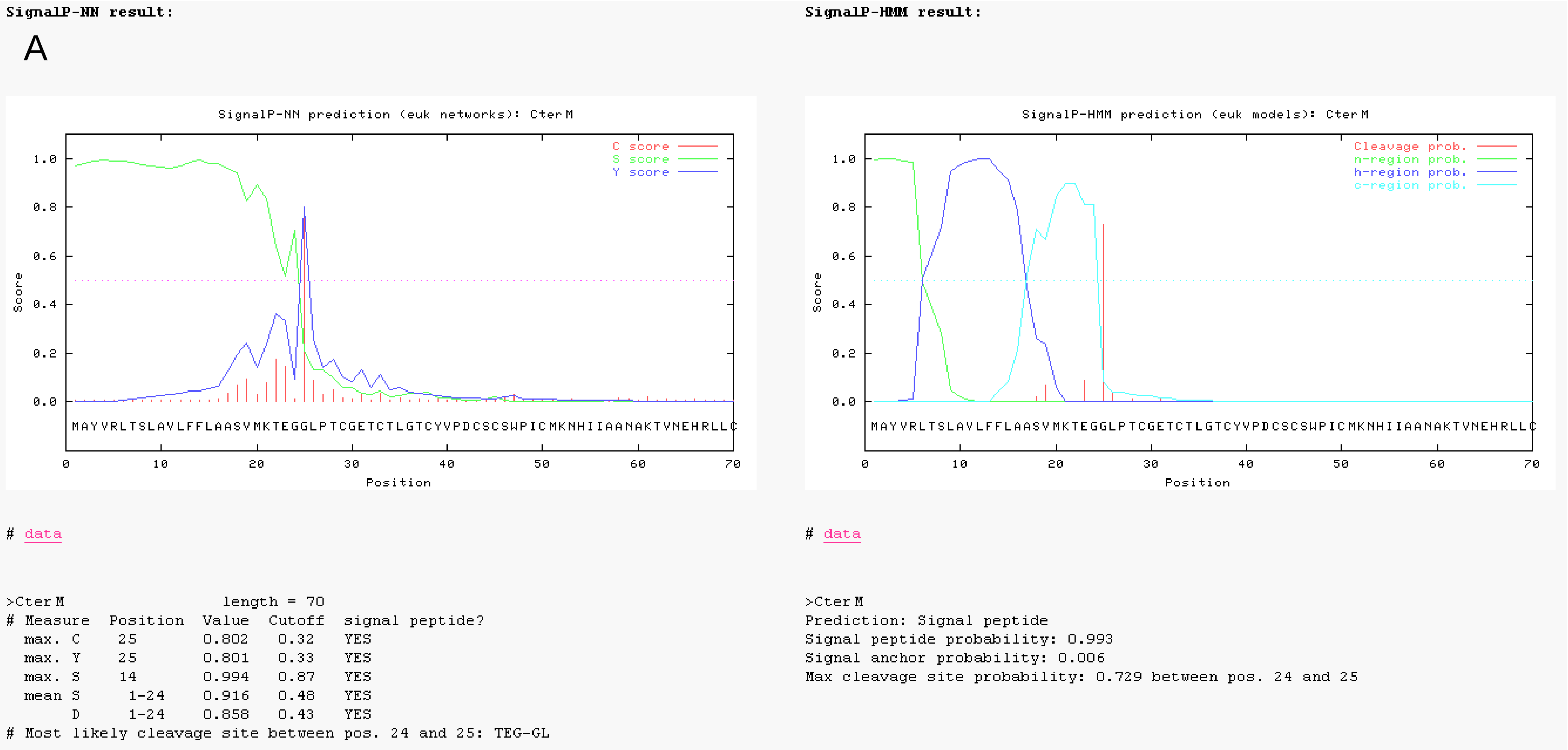
Residue	b	b+2	y	y+2
K	**129.1022**	65.0548	1368.5922	**684.7998**
N[Dea]	**244.1292**	122.5682	1240.4973	620.7523
G	301.1506	151.0790	1125.4703	563.2388
I	**414.2347**	207.6210	1068.4489	534.7281
P	511.2875	**256.1474**	**955.3648**	478.1860
C[CAM]	671.3181	336.1627	**858.3121**	429.6597
G	728.3396	364.6734	698.2814	349.6443
E	857.3822	429.1947	641.2599	321.1336
S	**944.4142**	472.7107	512.2173	256.6123
C[CAM]	**1104.4449**	**552.7261**	**425.1853**	213.0963
V	**1203.5133**	**602.2603**	**265.1547**	133.0810
F	1350.5817	675.7945	**166.0863**	83.5468

Supplementary spectrum C6a: MS/MS of *m/z* 684.81 (mass 1367.61 Da) showing sequence ion coverage. The derived sequence is KNGIPCGESCVF.

(b) SignalP analyses of Cter M and kalata B1 alongside selected albumin-1 precursors from Fabaceae.

Panel A- SignalP (1) analysis of Cter M precursor protein predicts signal peptidase cleavage at the proto-N-terminus of the mature cyclotide sequence, between precursor protein residues 24 and 25 (72.9% probability). Panel B- SignalP analysis of kalata B1 precursor protein predicts signal peptidase cleavage between precursor protein residues 22 and 23 (82.5% probability). As in all previously characterized cyclotide genes, a pro-region and an N-terminal repeat region are encoded prior to the start of the first cyclotide domain. Panels C through F- Respective SignalP analyses of albumin-1 precursor proteins from *Pisum sativum*, *Medicago truncatula*, *Phaseolus vulgaris*, and *Glycine max* predict signal peptidase cleavage at the proto-N-termini of mature PA1b peptide sequences. Cleavages are predicted between residues 26 and 27 (53.0%), 22 and 23 (51.1%), 27 and 28 (69.7%), and 19 and 20 (98.6%) respectively.

1. Bendtsen JD, Nielsen H, von Heijne G, & Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340:783-795.



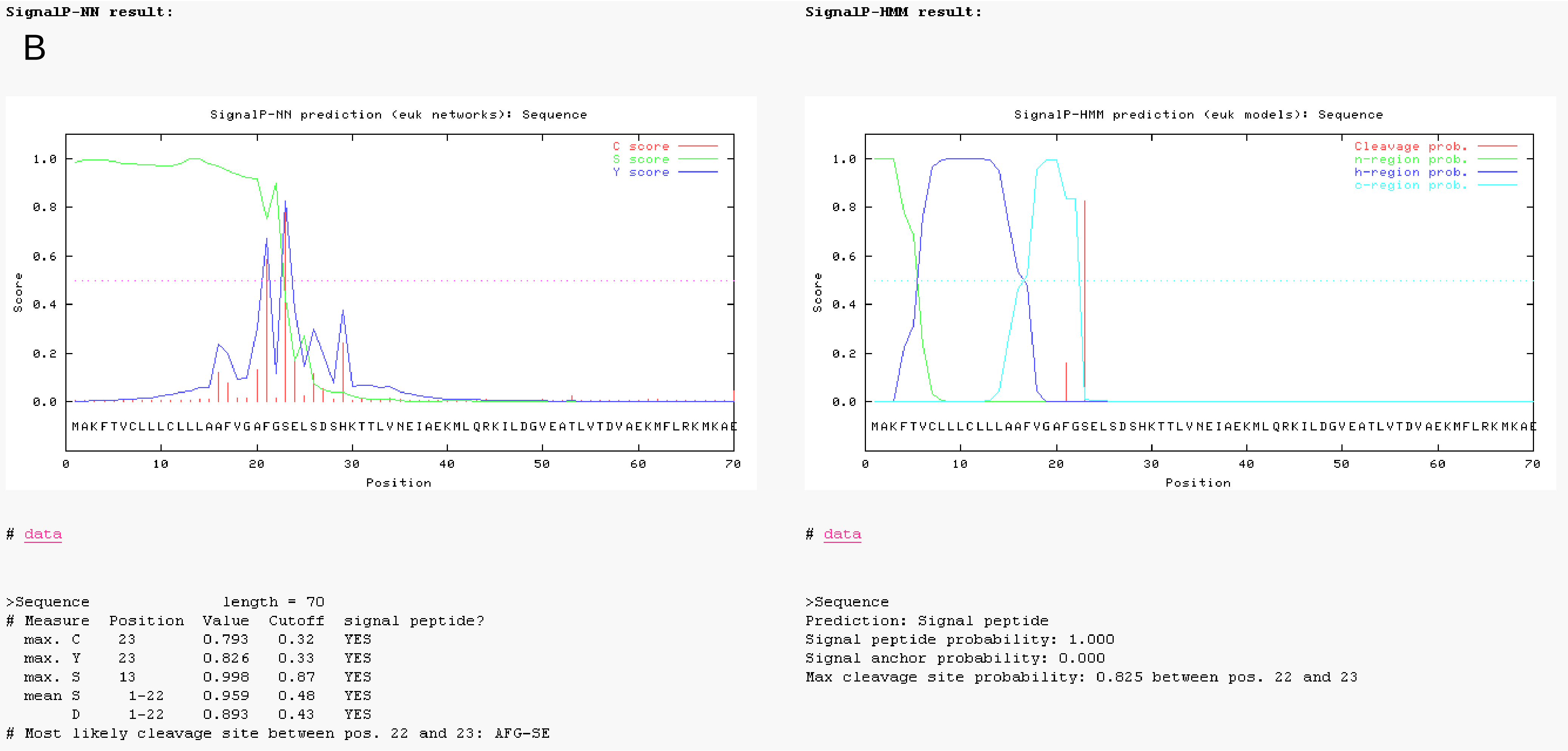
SignalP-NN result:

```
>Cter M                    length = 70
# Measure  Position  Value   Cutoff   signal peptide?
  max. C     25      0.802   0.32    YES
  max. Y     25      0.801   0.33    YES
  max. S     14      0.994   0.87    YES
  mean S      1-24   0.916   0.48    YES
       D      1-24   0.858   0.43    YES
# Most likely cleavage site between pos. 24 and 25: TEG-GL
```

SignalP-HMM result:

```
>Cter M
Prediction: Signal peptide
Signal peptide probability: 0.993
Signal anchor probability: 0.006
Max cleavage site probability: 0.729 between pos. 24 and 25
```

Clitoria ternatea Cter M SignalP analysis.



SignalP-NN result:

data

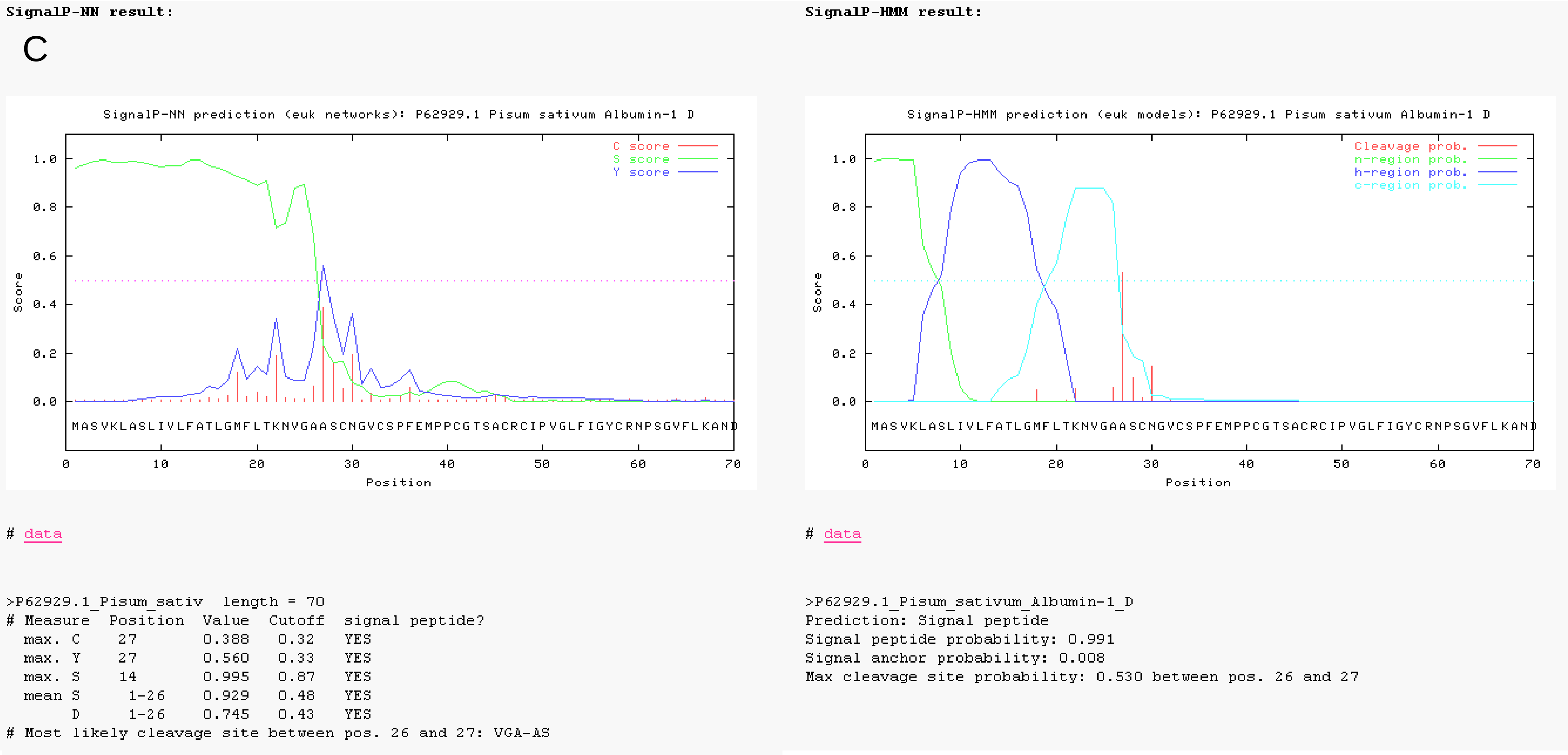
```
>Sequence               length = 70
# Measure  Position  Value   Cutoff  signal peptide?
  max. C     23       0.793   0.32    YES
  max. Y     23       0.826   0.33    YES
  max. S     13       0.998   0.87    YES
  mean S      1-22    0.959   0.48    YES
       D      1-22    0.893   0.43    YES
# Most likely cleavage site between pos. 22 and 23: AFG-SE
```

SignalP-HMM result:

data

```
>Sequence
Prediction: Signal peptide
Signal peptide probability: 1.000
Signal anchor probability: 0.000
Max cleavage site probability: 0.825 between pos. 22 and 23
```

Oldenlandia affinis kalata B1 SignalP analysis.



SignalP-NN result:

data

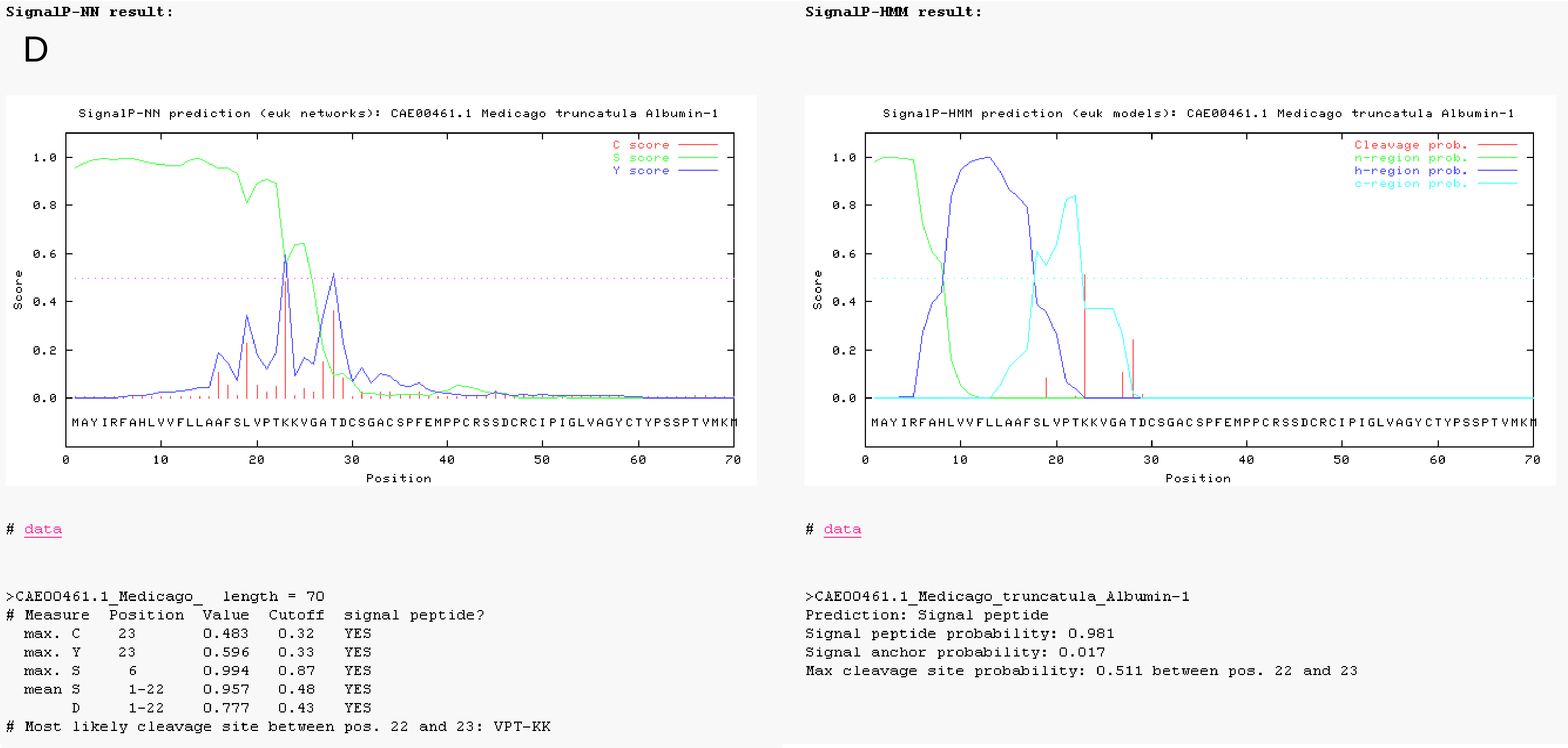
```
>P62929.1_Pisum_sativ  length = 70
# Measure  Position  Value  Cutoff  signal peptide?
  max. C     27      0.388   0.32   YES
  max. Y     27      0.560   0.33   YES
  max. S     14      0.995   0.87   YES
  mean S      1-26   0.929   0.48   YES
       D      1-26   0.745   0.43   YES
# Most likely cleavage site between pos. 26 and 27: VGA-AS
```

SignalP-HMM result:

data

```
>P62929.1_Pisum_sativum_Albumin-1_D
Prediction: Signal peptide
Signal peptide probability: 0.991
Signal anchor probability: 0.008
Max cleavage site probability: 0.530 between pos. 26 and 27
```

Pisum sativum Albumin-1 SignalP analysis.



SignalP-NN result:

data

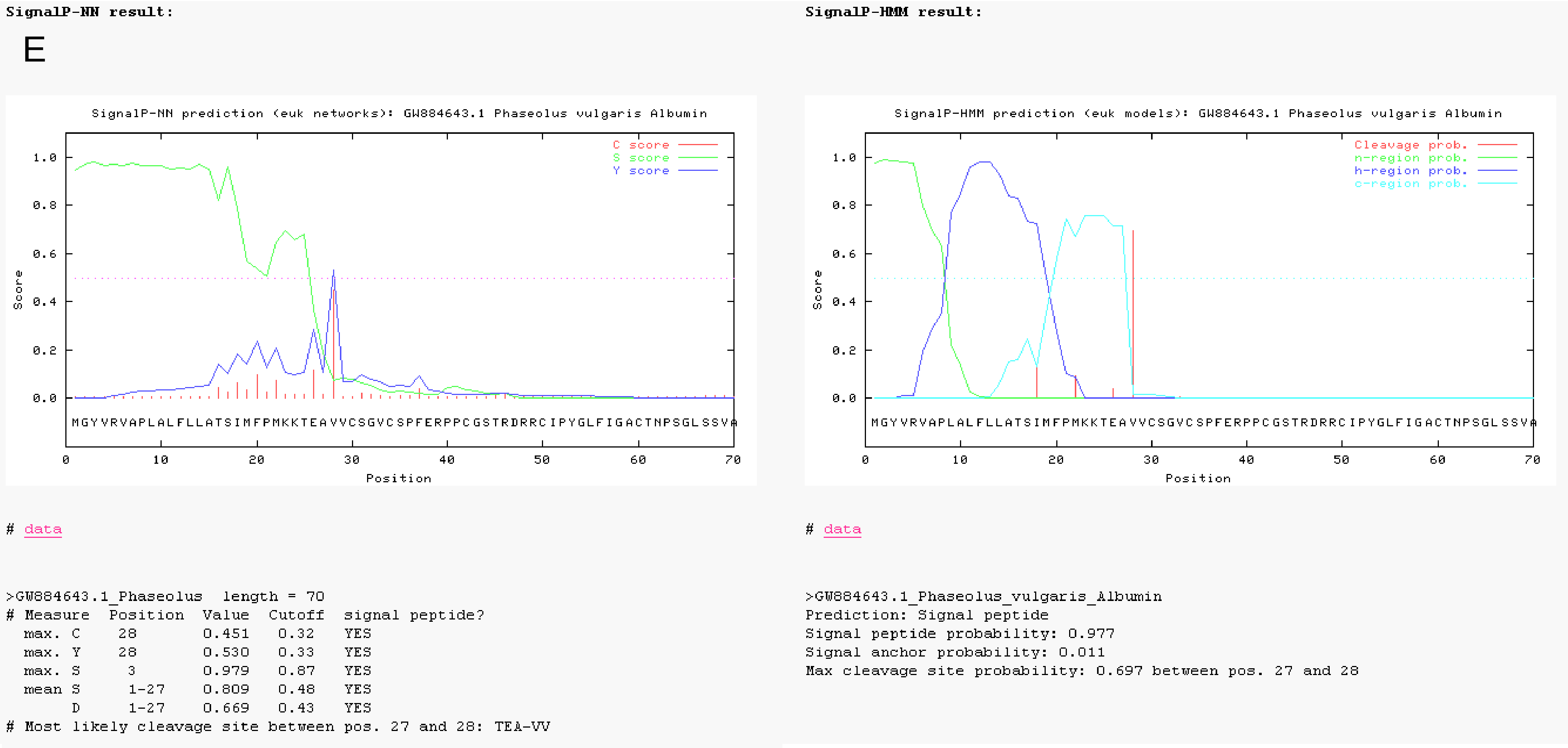
```
>CAE00461.1_Medicago_  length = 70
# Measure  Position  Value  Cutoff  signal peptide?
  max. C    23        0.483   0.32   YES
  max. Y    23        0.596   0.33   YES
  max. S     6        0.994   0.87   YES
  mean S     1-22     0.957   0.48   YES
       D     1-22     0.777   0.43   YES
# Most likely cleavage site between pos. 22 and 23: VPT-KK
```

SignalP-HMM result:

data

```
>CAE00461.1_Medicago_truncatula_Albumin-1
Prediction: Signal peptide
Signal peptide probability: 0.981
Signal anchor probability: 0.017
Max cleavage site probability: 0.511 between pos. 22 and 23
```

Medicago truncatula Albumin-1 SignalP analysis.



SignalP-NN result:

data

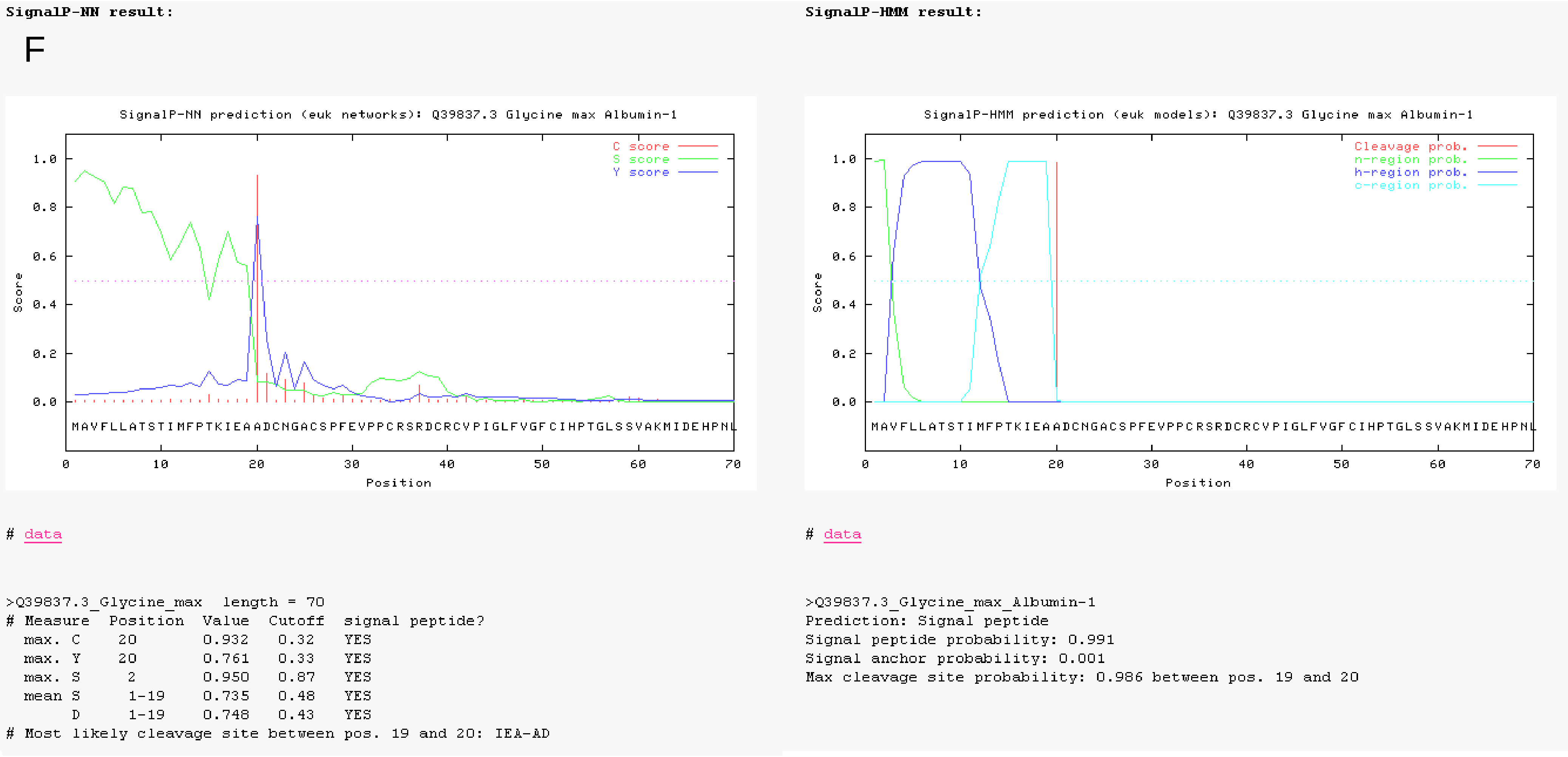
```
>GW884643.1_Phaseolus  length = 70
# Measure  Position  Value   Cutoff   signal peptide?
  max. C     28      0.451    0.32     YES
  max. Y     28      0.530    0.33     YES
  max. S      3      0.979    0.87     YES
  mean S     1-27    0.809    0.48     YES
       D     1-27    0.669    0.43     YES
# Most likely cleavage site between pos. 27 and 28: TEA-VV
```

SignalP-HMM result:

data

```
>GW884643.1_Phaseolus_vulgaris_Albumin
Prediction: Signal peptide
Signal peptide probability: 0.977
Signal anchor probability: 0.011
Max cleavage site probability: 0.697 between pos. 27 and 28
```

Phaseolus vulgaris Albumin-1 SignalP analysis.



SignalP-NN result:

```
>Q39837.3_Glycine_max   length = 70
# Measure  Position  Value   Cutoff   signal peptide?
  max. C     20       0.932   0.32    YES
  max. Y     20       0.761   0.33    YES
  max. S      2       0.950   0.87    YES
  mean S      1-19    0.735   0.48    YES
       D      1-19    0.748   0.43    YES
# Most likely cleavage site between pos. 19 and 20: IEA-AD
```

SignalP-HMM result:

```
>Q39837.3_Glycine_max_Albumin-1
Prediction: Signal peptide
Signal peptide probability: 0.991
Signal anchor probability: 0.001
Max cleavage site probability: 0.986 between pos. 19 and 20
```

Glycine max Albumin-1 SignalP analysis.