SUPPLEMENTARY INFORMATION FILE

for

Short flexible peptides linking non-interacting protein domains appear to resist proteolysis by facilitating domain motions that sterically inhibit protease approach

Snehal Waghmare and Purnananda Guptasarma*

Centre for Protein Science, Design and Engineering (CPSDE) and Hyperthermophile Enzyme Hydrolase Research Centre (HEHRC), Department of Biological Sciences, Indian Institute of Science Education and Research (IISER), Mohali, Knowledge City, Sector-81, SAS Nagar, Punjab 140306

Keywords: linker peptide; protein fusion; proteolytic susceptibility; environmental proteases; non-specific proteases; thermostable domains

* Corresponding author Tel: +91 9815417265 Email: guptasarma@iisermohali.ac.in

CONTENTS

Primers used to create Coh2, CelF and the five fusion constructs incorporating five different linkers flanked by Coh2 and BSX

1. Coh2:

Forward - FP_Coh2_Nde1:

ATATATCATATGTCAGACGGTGTGGTAGTAGAAATTGGCAAAG

Reverse - Coh2XhoR:

ATATTACTCGAGTGTTGCATTGCCAACGTTAACACCACCG

2. BSX:

The codon-optimized gene encoding BSX was gotten synthesized commercially

3. CelF:

Forward - CThCelFNhe1F:

ATATATGCTAGCGCGGATTTCAACTATGGTGAGGC

Reverse - CThCelFXho1R:

TATATTCTCGAGCTGTTCAGCCGGGAATTTTTCAATAAG

4. Common forward and reverse primers for the five fusion constructs.

Forward - FP_Coh2_Nde1:

ATATATCATATGTCAGACGGTGTGGTAGTAGAAATTGGCAAAG

Reverse - BSXXhoIR:

ATATATCTCGAGAGACTGGAAGTACAGGTTCTCGTCGATGATACGCCAGAACGC

5. *Nat-Quarter*:

Forward - Coh2BSXlinkF:

GTTGGCAATGCAACACCGACCAAGGGAGCAACAGTTCAGCCGTTCGCGTGG

Reverse - Coh2BSXlinkR:

TGTTGCTCCCTTGGTCGGTGTTGC

6. *Nat-Half*:

Forward - FP2cohbsxhalf:

AAATACAGCTACGCCGACAAAATCAGCTACGGTTCAGCCCTTCGCGTGG

Reverse - RP2cohlinkhalf:

AGCTGATTTTGTCGGCGTAGCTGTATTTGTTGGTGTTGCTCCCTTGGTCGGTGTTGC

7. *Nat-Full*:

Forward - FP2cohbsxfull:

CCGACAAACACACCGACAAACACACCCGCAAATACACCAGTTCAGCCGTTCGCGTGG

Reverse - RP2cohlinkfull:

GTCGGTGTGTTTGTCGGTACAGATGGCCTGGTGGGGGTAGCCGTAGCTGATTTTGTCGG

8. *Rigid*: (5 amino acids were deleted during cloning)

Forward - FP2cohbsxrigid:

GCGGCAGCGAAAGAGGCGGCAGCAAAAGAGGCAGCAGCTAAAGTTCAGCCGTTCGCGTGG

Reverse - RP2cohlinkrigid:

GCTGCCGCCTCTTTCGCTGCCGCCTCTTTCGCTGCCGCCTCGCCAACGTTAACACCACC

9. *Flexible*: (10 amino acids were deleted during cloning)

Forward - FP2cohbsxflex:

GGTGGTGGCAGCGGTGGTGGCGGTTCTGGTGGTGGTGGAGTTCAGCCGTTCGCGTGG

Reverse - RP2cohlinkflex:

GCCACCACCGCTGCCACCACCACCAGAACCGCCACCACCTGAGCCAACGTTAACACCACC

Amino acid sequences of domains and fusion constructs

Note: All genes were cloned in the pET23a vector, between NdeI and XhoI restriction sites, except CelF which was cloned between NheI and XhoI. All genes were expressed in *E.coli* BL21 Star (DE3) pLysS. Further, in all of the sequences shown below:

- The N-terminal methione and the 6xHis histidine-affinity tag at the C-terminal are shown with green highlighting;
- Linker sequences are shown with grey highlighting;
- Restriction site-derived sequences are shown with red highlighting;
- A Tev protease site that was included in some constructs is shown with yellow highlighting;
- The sequences of Coh2 and BSX are shown without highlighting.
- In the sequence of CelF, the dockerin sequence is shown in cyan.

1. **Coh2:**

MSDGVVVEIGKVTGSVGTTVEIPVYFRGVPSKGIANCDFVFRYDPNVLEIIGIDPGDIIVDPNPTKSFDTA
IYPDRKIIVFLFAEDSGTGAYAITKDGVFAKIRATVKSSAPGYITFDEVGGFADNDLVEQKVSFIDGGVN
VGNATLEHHHHHH

2. **BSX:**

MVQPFAWQVASLADRYEESFDIGAAVEPHQLNGRQGKVLKHHYNSIVAENAMKPISLQPEEGVFTWD
GADAIVEFARKNNMNLRFHTLVWHNQVPDWFFLDEEGNPMVEETNEAKRQANKELLLERLETHIKTV
VERYKDDVTAWDVVNEVVDDGTPNERGLRESVWYQITGDEYIRVAFETARKYAGEDAKLFINDYNTE
VTPKRDHLYNLVQDLLADGVPIDGVGHQAHIQIDWPTIDEIRTSMEMFAGLGLDNQVTELDVSLYGWP
PRPAFPTYDAIPQERFQAQADRYNQLFELYEELDADLSSVTFWGIADNHTWLDDRAREYNDGVGKDA
PFVFDPNYRVKPAFWRIIDGSEFELRRQACGRTRLEHHHHHH

3. **CelF:**

MASADFNYGEALQKAIMFYEFQRSGKLPENKRNNWRGDSALNDGADNGLDLTGGWYDAGDHVKFN
LPMAYAVTMLAWSVYESRDAYVQSGQLPYILDNIKWATDYFIKCHPSPNVYYYQVGDGALDHSWWG
PAEVMQMPRPSFKVDLTNPGSTVVAETAAAMAASSIVFKPTDPEYAATLLRHAKELFTFADTTRSDAG
YRAAEGYYSSHSGFYDELTWASIWLYLATGDQSYLDKAESYEPHWERERGTTLISYSWAHCWDNKLY
GSLLLLAKITGKSYYKQCIENHLDYWTVGFNGSRVQYTPKGLAYLDRWGSLRYATTQAFLASVYADW
SGCDPAKAAVYKEFAKKQVDYALGSTGRSFVVGFGKNPPRNPHHRTAHSSWSALMTEPAECRHILVG
ALVGGPDGSDSYVDRLDDYQCNEVANDYNAGFVGALAKMYEKYGGEPIPNFVAFETPGEEFYVEAA
VNAAGPGFVNIKASIINKSGWPARGSDKLSAKYFVDISEAVAKGITLDQITVQSTTNGGAKVSQLLPWD
PDNHIYYVNIDFTGINIFPGGINEYKRDVYFTITAPYGEGNWDNTNDFSFQGLEQGFTSKKTEYIPLYDG
NVRVWGKVPDGGSEPDPTPTITVGPTPSVTPTSVPGIMLGDVNFDGRINSTDYSRLKRYVIKSLEFTDPE
EHQKFIAAADVDGNGRINSTDLYVLNRYILKLIEKFPAEQLEHHHHHH

4. ***Nat-Quarter* or Coh2-*Nat-Quarter*-BSX:**

MSDGVVVEIGKVTGSVGTTVEIPVYFRGVPSKGIANCDFVFRYDPNVLEIIGIDPGDIIVDPNPTKSFDTA
IYPDRKIIVFLFAEDSGTGAYAITKDGVFAKIRATVKSSAPGYITFDEVGGFADNDLVEQKVSFIDGGVN
VGNATPTKGATVQPFAWQVASLADRYEESFDIGAAVEPHQLNGRQGKVLKHHYNSIVAENAMKPISL
QPEEGVFTWDGADAIVEFARKNNMNLRFHTLVWHNQVPDWFFLDEEGNPMVEETNEAKRQANKELL
LERLETHIKTVVERYKDDVTAWDVVNEVVDDGTPNERGLRESVWYQITGDEYIRVAFETARKYAGED
AKLFINDYNTEVTPKRDHLYNLVQDLLADGVPIDGVGHQAHIQIDWPTIDEIRTSMEMFAGLGLDNQV
TELDVSLYGWPPRPAFPTYDAIPQERFQAQADRYNQLFELYEELDADLSSVTFWGIADNHTWLDDRAR
EYNDGVGKDAPFVFDPNYRVKPAFWRIIDENLYFQSLEHHHHHH

5. ***Nat-Half or Coh2-Nat-Half-BSX*:**

MSDGVVVEIGKVTGSVGTTVEIPVYFRGVPSKGIANCDFVFRYDPNVLEIIGIDPGDIIVDPNPTKSFDTA
IYPDRKIIVFLFAEDSGTGAYAITKDGVFAKIRATVKSSAPGYITFDEVGGFADNDLVEQKVSFIDGGVN
VGNATPTKGATPTNTATPTKSATVQPFAWQVASLADRYEESFDIGAAVEPHQLNGRQGKVLKHHYNSI
VAENAMKPISLQPEEGVFTWDGADAIVEFARKNNMNLRFHTLVWHNQVPDWFFLDEEGNPMVEETN
EAKRQANKELLLERLETHIKTVVERYKDDVTAWDVVNEVVDDGTPNERGLRESVWYQITGDEYIRVA
FETARKYAGEDAKLFINDYNTEVTPKRDHLYNLVQDLLADGVPIDGVGHQAHIQIDWPTIDEIRTSME
MFAGLGLDNQVTELDVSLYGWPPRPAFPTYDAIPQERFQAQADRYNQLFELYEELDADLSSVTFWGIA
DNHTWLDDRAREYNDGVGKDAPFVFDPNYRVKPAFWRIIDENLYFQSLEHHHHHH

6. ***Nat-Full or Coh2-Nat-Full-BSX*:**

MSDGVVVEIGKVTGSVGTTVEIPVYFRGVPSKGIANCDFVFRYDPNVLEIIGIDPGDIIVDPNPTKSFDTA
IYPDRKIIVFLFAEDSGTGAYAITKDGVFAKIRATVKSSAPGYITFDEVGGFADNDLVEQKVSFIDGGVN
VGNATPTKGATPTNTATPTKSATATPTRPSVPTNTPTNTPANTPVQPFAWQVASLADRYEESFDIGAAV
EPHQLNGRQGKVLKHHYNSIVAENAMKPISLQPEEGVFTWDGADAIVEFARKNNMNLRFHTLVWHN
QVPDWFFLDEEGNPMVEETNEAKRQANKELLLERLETHIKTVVERYKDDVTAWDVVNEVVDDGTPN
ERGLRESVWYQITGDEYIRVAFETARKYAGEDAKLFINDYNTEVTPKRDHLYNLVQDLLADGVPIDGV
GHQAHIQIDWPTIDEIRTSMEMFAGLGLDNQVTELDVSLYGWPPRPAFPTYDAIPQERFQAQADRYNQ
LFELYEELDADLSSVTFWGIADNHTWLDDRAREYNDGVGKDAPFVFDPNYRVKPAFWRIIDENLYFQS
LEHHHHHH

7. ***Flexible or Coh2-Flexible-BSX*:**

MSDGVVVEIGKVTGSVGTTVEIPVYFRGVPSKGIANCDFVFRYDPNVLEIIGIDPGDIIVDPNPTKSFDTA
IYPDRKIIVFLFAEDSGTGAYAITKDGVFAKIRATVKSSAPGYITFDEVGGFADNDLVEQKVSFIDGGVN
VGSGGGGSGGGGVQPFAWQVASLADRYEESFDIGAAVEPHQLNGRQGKVLKHHYNSIVAENAMKPIS
LQPEEGVFTWDGADAIVEFARKNNMNLRFHTLVWHNQVPDWFFLDEEGNPMVEETNEAKRQANKEL
LLERLETHIKTVVERYKDDVTAWDVVNEVVDDGTPNERGLRESVWYQITGDEYIRVAFETARKYAGE
DAKLFINDYNTEVTPKRDHLYNLVQDLLADGVPIDGVGHQAHIQIDWPTIDEIRTSMEMFAGLGLDNQ
VTELDVSLYGWPPRPAFPTYDAIPQERFQAQADRYNQLFELYEELDADLSSVTFWGIADNHTWLDDRA
REYNDGVGKDAPFVFDPNYRVKPAFWRIIDENLYFQSLEHHHHHH

8. ***Rigid or Coh2-Rigid-Quarter-BSX*:**

MSDGVVVEIGKVTGSVGTTVEIPVYFRGVPSKGIANCDFVFRYDPNVLEIIGIDPGDIIVDPNPTKSFDTA
IYPDRKIIVFLFAEDSGTGAYAITKDGVFAKIRATVKSSAPGYITFDEVGGFADNDLVEQKVSFIDGGVN
VGEAAAKEAAAKEAAAKVQPFAWQVASLADRYEESFDIGAAVEPHQLNGRQGKVLKHHYNSIVAEN
AMKPISLQPEEGVFTWDGADAIVEFARKNNMNLRFHTLVWHNQVPDWFFLDEEGNPMVEETNEAKR
QANKELLLERLETHIKTVVERYKDDVTAWDVVNEVVDDGTPNERGLRESVWYQITGDEYIRVAFETA
RKYAGEDAKLFINDYNTEVTPKRDHLYNLVQDLLADGVPIDGVGHQAHIQIDWPTIDEIRTSMEMFAG
LGLDNQVTELDVSLYGWPPRPAFPTYDAIPQERFQAQADRYNQLFELYEELDADLSSVTFWGIADNHT
WLDDRAREYNDGVGKDAPFVFDPNYRVKPAFWRIIDENLYFQSLEHHHHHH

Mass spectrometric confirmation of the sequence of Coh2 by MALDI-MS

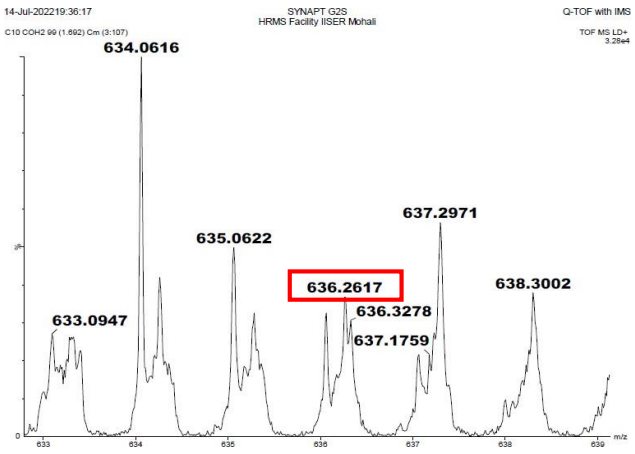
Coh2 is 16.54 kDa in size (including its C-terminal 6xHis tag).

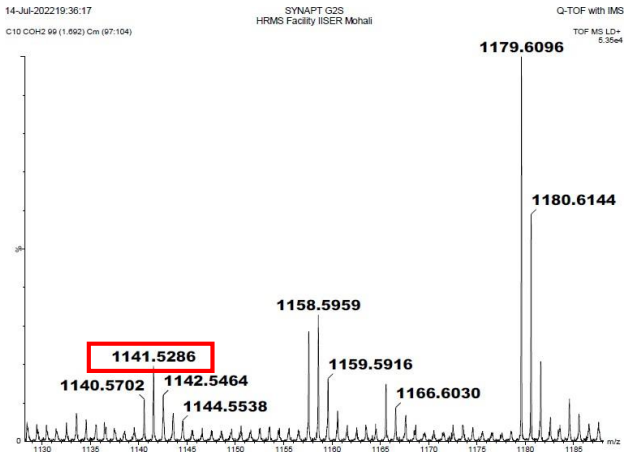
Upon digestion by trypsin, it is expected to generate 9 peptides (ignoring missed cleavages).

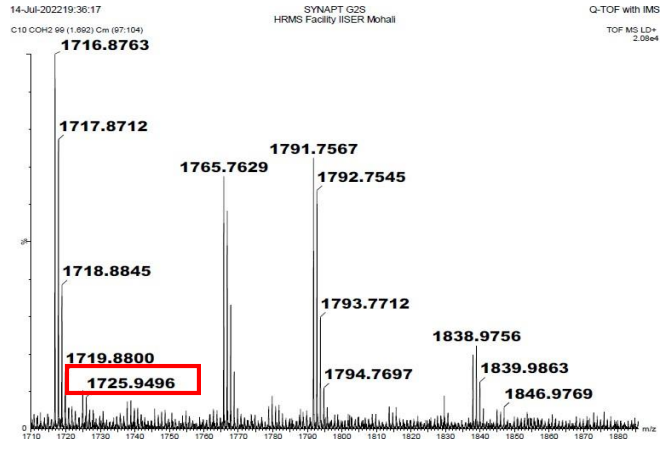
Of these 9, 4 peptides were observed. These are mentioned in the table along with their sequences.

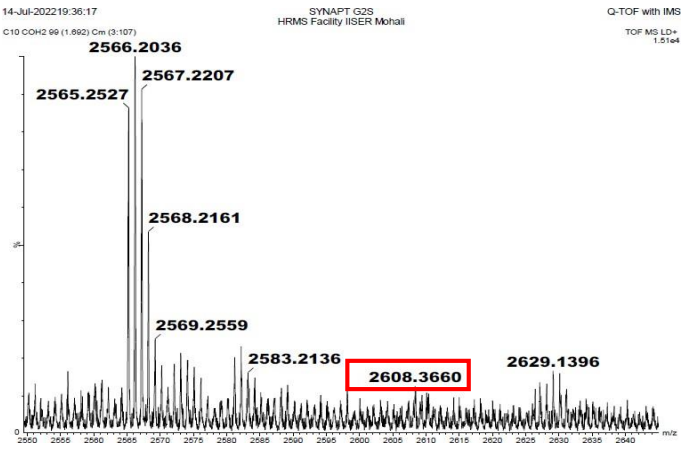
Mass spectral data for the 4 observed peptides is shown below the table.

Theoretical mass	Observed mass	Sequences of observed peptide masses
636.3351	636.3278	DGVFAK
1141.546	1141.529	GIANCDFVFR
1724.922	1725.93	VTGSVGTTVEIPVYFR
2607.355	2608.366	YDPNVLEIIGIDPGDIIVDPNPTK









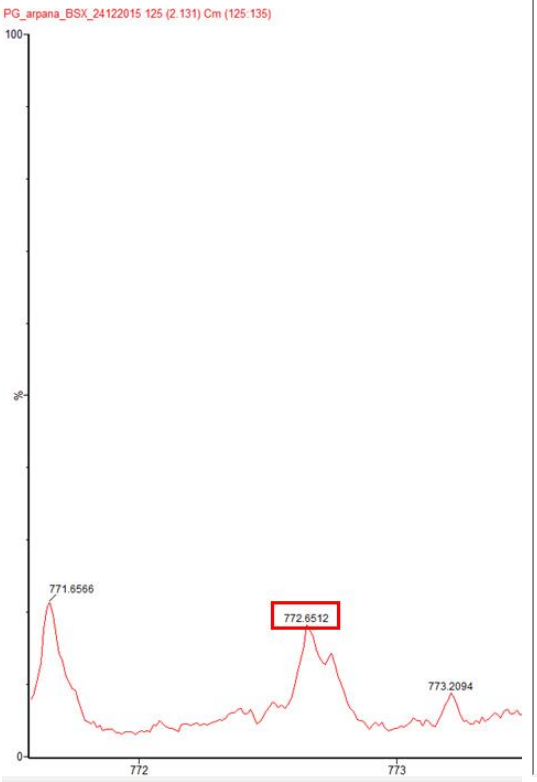
Mass spectrometric confirmation of the sequence of BSX by ESI-Q-TOF MS and MALDI MS

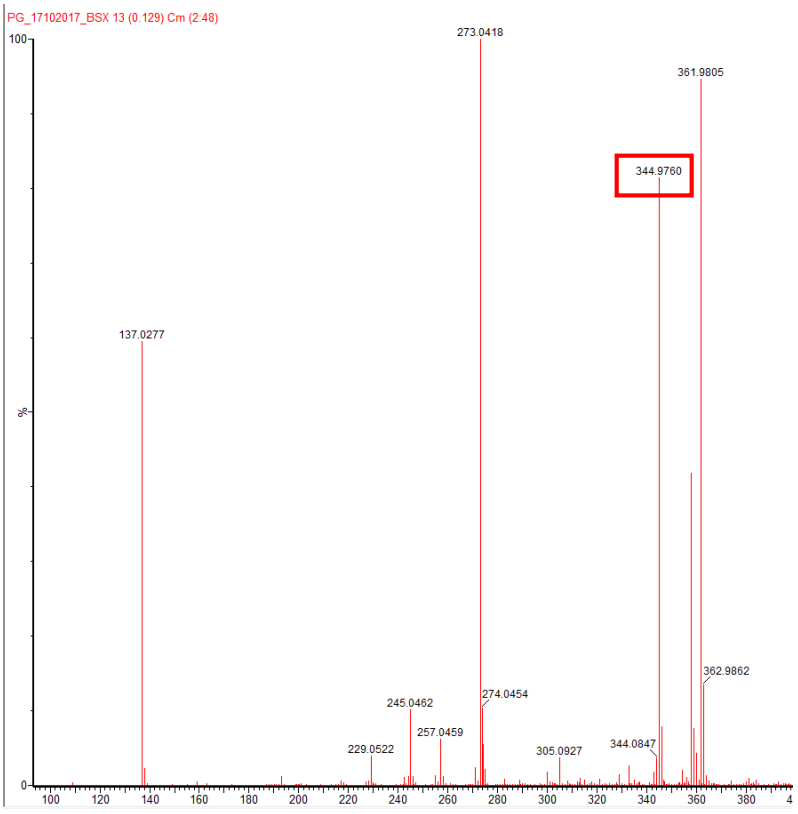
BSX is 43.88 kDa in size (including its C-terminal 6xHis tag).

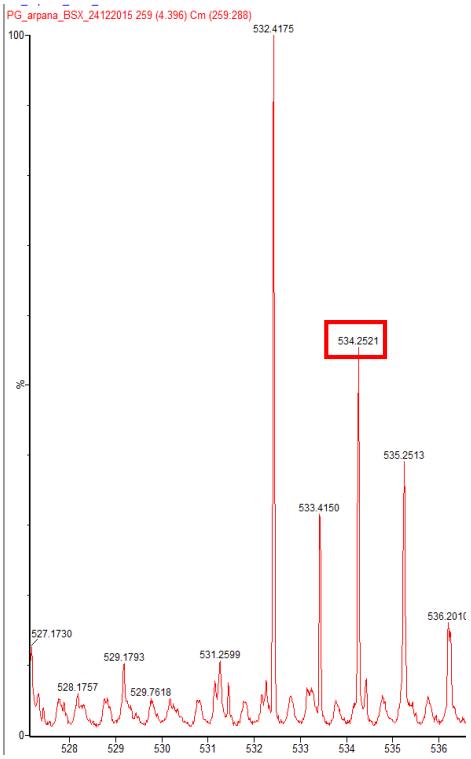
Upon digestion by trypsin, it is expected to generate 69 peptides (with 0 or 1 missed cleavages).

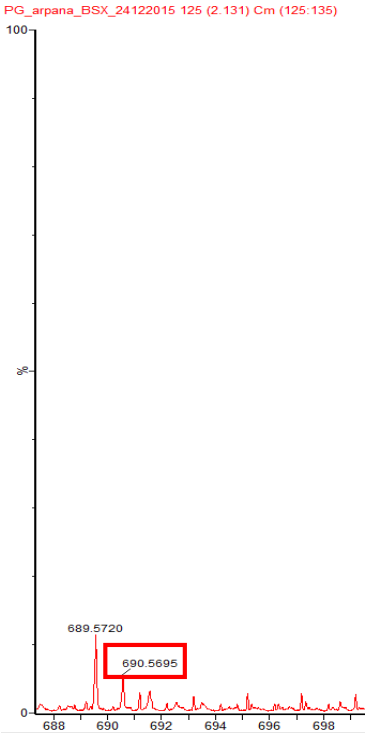
Of these 69 masses, the table (and spectra) below show the masses of 7 representative small peptides using ESI-TOF MS data and 5 representative large peptides using MALDI-TOF data. The sequences are mentioned.

Theoretical mass	Observed mass	Amino acid sequence	Missed cleavages required	MS technique
175.1189	175.0202	R	0	ESI-TOF
246.156	245.0462	AR	0	ESI-TOF
345.2245	344.9760	GLR	0	ESI-TOF
460.2514	460.7669	QANK	0	ESI-TOF
534.2453	534.2521	QACG	0	ESI-TOF
690.3464	690.5695	RQACGR	1	ESI-TOF
772.4563	772.6512	ELLLER	0	ESI-TOF
835.4057	836.5156	FQAQADR	0	MALDI-TOF
2085.0399	2088.7859	GLRESVWY QITGDEYIR	1	MALDI-TOF
2288.1081	2289.3584	YAGEDAKLFIN DYNTEVTPK	1	MALDI-TOF
2345.0527	2344.2581	DTVTAWDVVNE VVDDGTPNER	0	MALDI-TOF
3914.8231	3911.0579	FHTLVWHNQVPDWFFL DEEGNPMVEETNEAKR	1	MALDI-TOF

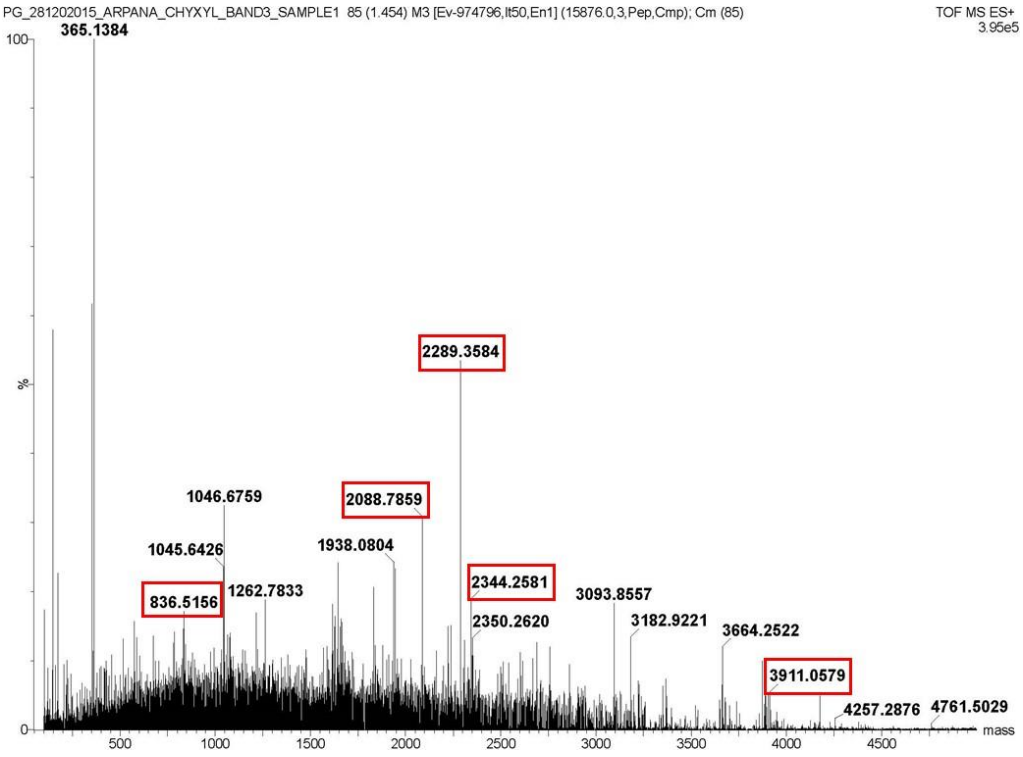

PG_arpana_BSX_24122015 125 (2.131) Cm (125:135)
100
%
0
771.6566
772.6512
773.2094
772
773


PG_17102017_BSX 13 (0.129) Cm (2:48)
100
%
0
273.0418
361.9805
344.9760
137.0277
362.9862
245.0462
274.0454
257.0459
229.0522
305.0927
344.0847
100
120
140
160
180
200
220
240
260
280
300
320
340
360
380


PG_arpana_BSX_24122015 259 (4.396) Cm (259:288)
100
%
0
532.4175
534.2521
535.2513
533.4150
536.2010
527.1730
529.1793
531.2599
528.1757
529.7618
528
529
530
531
532
533
534
535
536


PG_arpana_BSX_24122015 125 (2.131) Cm (125:135)
100
%
0
689.5720
690.5695
688
690
692
694
696
698


PG_08012016_ARPANA_BSX 210 (3.567) Cm (200:217)
100
%
0
164.9596
177.0355
177.9893
175.0202
167.0065
171.1783
182.0124
184.0024
164 166 168 170 172 174 176 178 180 182 184 186
PG_17102017_BSX1 56 (0.498) Cm (4:90)
100
%
0
245.0462
229.0522
257.0459
220 230 240 250 260


PG_281202015_ARPANA_CHYXYL_BAND3_SAMPLE1 85 (1.454) M3 [Ev-974796,It50,En1] (15876.0,3,Pep,Cmp); Cm (85)
TOF MS ES+
3.95e5
100
%
0
365.1384
2289.3584
1046.6759
2088.7859
1045.6426
1938.0804
836.5156
1262.7833
2344.2581
2350.2620
3093.8557
3182.9221
3664.2522
3911.0579
4257.2876
4761.5029
500 1000 1500 2000 2500 3000 3500 4000 4500
mass