

**Supplementary Information**  
**for**  
**Phase variable glycosylation in non-typeable *Haemophilus influenzae***

**Danila Elango<sup>1,2</sup> and Benjamin L. Schulz<sup>1,2,3,\*</sup>**

<sup>1</sup> School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia 4072, Queensland, Australia.

<sup>2</sup> Australian Infectious Diseases Research Centre, The University of Queensland, Brisbane 4072, Australia.

<sup>3</sup> Centre for Biopharmaceutical Innovation, Australian Institute of Bioengineering and Nanotechnology, The University of Queensland, St Lucia 4072, Queensland, Australia.

\* To whom correspondence should be addressed: Benjamin L. Schulz, [b.schulz@uq.edu.au](mailto:b.schulz@uq.edu.au), +61 7 3365 4875.

## Table of Contents

<b>Supplementary Table S1.</b> List of NTHi strains used in investigation of SSRs in the promoter regions of hmw1A, hmw1C, hmw2A, hmw2C.	S-1
<b>Supplementary Table S2.</b> Measurement of site-specific glycan occupancy in HMW1A with HMW1C abundance fixed with 0.2% arabinose and HMW1A abundance altered with varying concentrations of IPTG (0.05-1 mM).	S-2
<b>Supplementary Table S3.</b> Measurement of site-specific glycan occupancy in HMW1A with HMW1A abundance fixed with 0.1 mM IPTG and HMW1C abundance altered with varying concentrations of arabinose (0.00002-2%).	S-2
<b>Supplementary Figure S1.</b> MS/MS peptide identification of TTLTN[+162.053]TTLESILK	S-3
<b>Supplementary Figure S2.</b> MS/MS peptide identification of GTFVN[+162.053]ITANQR	S-4
<b>Supplementary Figure S3.</b> MS/MS peptide identification of IYVNSSIN[+486.159]LSNGSLTLWSEGR.	S-5
<b>Supplementary Figure S4.</b> MS/MS peptide identification of GAN[+162.053]LTIYSGGWVDVHK	S-6
<b>Supplementary Figure S5.</b> MS/MS peptide identification of FEGTLN[+162.053]ISGK	S-7
<b>Supplementary Figure S6.</b> MS/MS peptide identification of YFN[+162.053]VSTGSSLR	S-8
<b>Supplementary Figure S7.</b> MS/MS peptide identification of DLTLN[+162.053]ATGGNITLLQVEGTDGMIGK	S-9
<b>Supplementary Figure S8.</b> MS/MS peptide identification of NITFEGGN[+162.053]ITFGSR	S-10
<b>Supplementary Figure S9.</b> MS/MS peptide identification of AITN[+162.053]FTFNVGGGLFDNK	S-11
<b>Supplementary Figure S10.</b> MS/MS peptide identification of NLSITTNN[+162.053]SSSTYR	S-12
<b>Supplementary Figure S11.</b> MS/MS peptide identification of TIISGN[+162.053]ITNK	S-13
<b>Supplementary Figure S12.</b> MS/MS peptide identification of NGDLN[+162.053]ITNEGSDTEMQIGGDVSQK	S-14
<b>Supplementary Figure S13.</b> MS/MS peptide identification of AGVDGENSDSDATNNAN[+162.053]LTIK	S-15
<b>Supplementary Figure S14.</b> MS/MS peptide identification of LTQDLN[+162.053]ISGFNK	S-16
<b>Supplementary Figure S15.</b> MS/MS peptide identification of GQVNLSAQDGGSVAGSINAAN[+162.053]VTLNTTGTLLTVK	S-17
<b>Supplementary Figure S16.</b> MS/MS peptide identification of GSNIN[+162.053]ATSGTLLVINAK	S-18
<b>Supplementary Figure S17.</b> MS/MS peptide identification of DAELNGAALGNHTVVN[+162.053]ATNANGSGSVIATTSSR	S-19
<b>Supplementary Figure S18.</b> MS/MS peptide identification of DAELNGAALGNHTVVN[+486.159]ATNANGSGSVIATTSSR.	S-20
<b>Supplementary Figure 19.</b> Secondary structure prediction of HMW1A signal peptide and pro-piece domains using JPred.	S-21
<b>Supplementary Figure 20.</b> Secondary structure prediction of mature HMW1A using JPred.	S-22

**Supplementary Table S1.** List of NTHi strains used in investigation of SSRs in the promoter regions of *hmw1A*, *hmw1C*, *hmw2A*, *hmw2C*. Genbank accessed 6 August 2019.

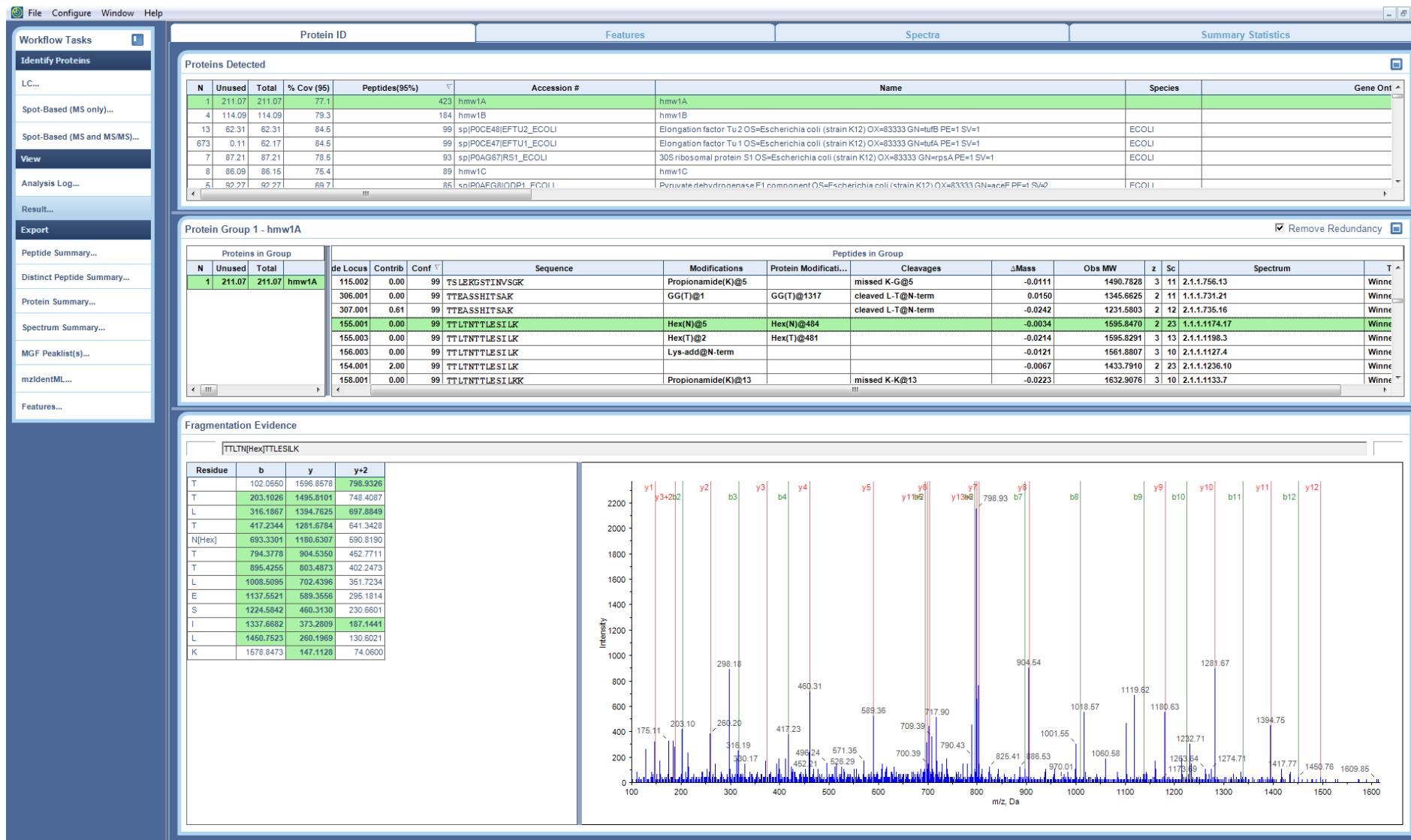
Strain	Genbank accession (Nucleotide)	Gene	GenBank accession (Protein)
<b>R2846/12</b>	CP002276.1	<i>hmw1A</i>	ADO96128.1
		<i>hmw1C</i>	ADO96126.1
		<i>hmw2A</i>	ADO96470.1
		<i>hmw2C</i>	ADO96472.1
<b>86-028NP</b>	CP000057.2	<i>hmw1A</i>	AAX88733.1
		<i>hmw1C</i>	AAX88735.1
		<i>hmw2A</i>	AAX88269.1
		<i>hmw2C</i>	AAX88267.1
<b>2019</b>	CP008740.1	<i>hmw1A</i>	AKA47280.1
		<i>hmw1C</i>	AKA47278.1
		<i>hmw2A</i>	AKA47663.1
		<i>hmw2C</i>	AKA47665.1
<b>NCTC8143</b>	LN831035.1	<i>hmw1A</i>	CKH04147.1
		<i>hmw1C</i>	CKH04095.1
		<i>hmw2A</i>	CKH14369.1
		<i>hmw2C</i>	CKH14421.1
<b>PittEE</b>	CP000671.1	<i>hmw1A</i>	ABQ98149.1
		<i>hmw1C</i>	ABQ98147.1
		<i>hmw2A</i>	ABQ98487.1
		<i>hmw2C</i>	ABQ98489.1

**Supplementary Table S2.** Measurement of site-specific glycan occupancy in HMW1A with HMW1C abundance fixed with 0.2% arabinose and HMW1A abundance altered with varying concentrations of IPTG (0.05-1 mM). Asn residues in sequons are indicated in **bold**.

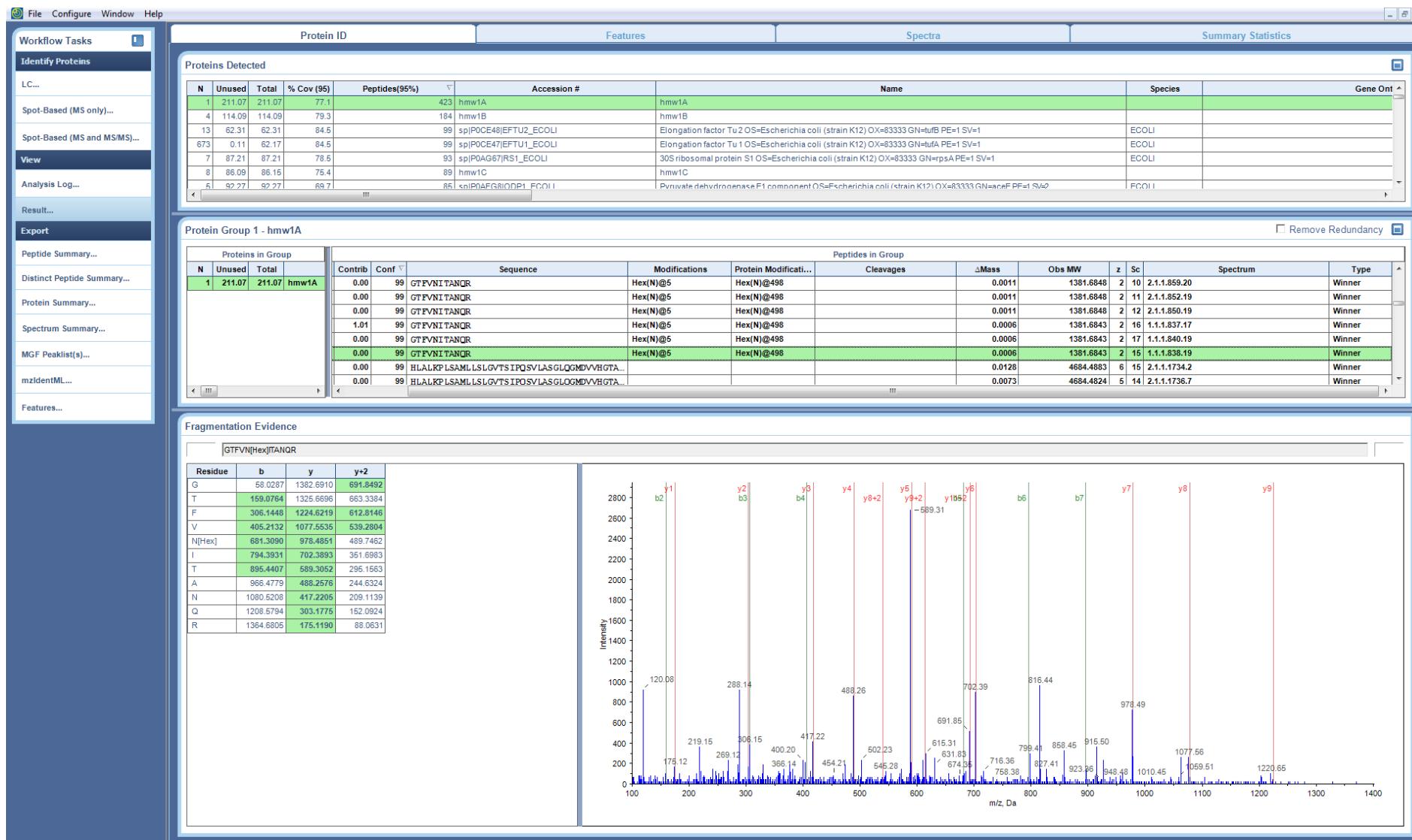
Peptide	1	0.5	0.2	0.1	0.05
QFNIDQNEMVQFLQENN <sub>120</sub> NSAVFNR	0.00%	0.00%	0.00%	0.00%	0.00%
<b>N</b> <sub>183</sub> FTFEQTK	0.00%	0.00%	0.00%	0.00%	0.00%
EWLLDPDN <sub>444</sub> VSINAETAGR	0.00%	0.00%	0.00%	0.00%	0.00%
SN <sub>456</sub> TSEDDEYTGSNSASTPK	0.00%	0.00%	0.00%	0.00%	0.00%
TTLTN <sub>484</sub> TTLESILK	8.38%	10.27%	18.94%	47.81%	78.44%
GTFVN <sub>498</sub> ITANQR	16.98%	19.61%	32.54%	60.19%	93.02%
GAN <sub>546</sub> LTIYSGGWVDVHK	16.87%	21.82%	33.74%	63.67%	61.86%
FEGTLN <sub>636</sub> ISGK	30.77%	31.06%	45.81%	79.32%	60.01%
VN <sub>642</sub> ISMVLPK	0.00%	0.00%	0.00%	0.00%	0.00%
YFN <sub>773</sub> VSTGSSLR	6.66%	8.67%	17.22%	50.07%	68.78%
AITN <sub>912</sub> FTFNVGGLFDNK	5.06%	4.23%	7.71%	14.07%	26.81%
TIISGN <sub>964</sub> ITNK	17.15%	20.74%	32.01%	63.32%	67.31%
AGVDGENSDSDATNNAN <sub>1029</sub> LTIK	17.28%	17.73%	31.02%	63.80%	75.42%
LTQDLN <sub>1044</sub> ISGFNK	12.17%	13.73%	22.90%	50.39%	50.27%
VETSGSNN <sub>1107</sub> NTEDSSDNNAGLTIDAK	0.00%	0.00%	0.00%	0.00%	0.00%
GSNIN <sub>1366</sub> ATSGTLVINAK	1.09%	1.36%	3.54%	9.25%	11.76%
VN <sub>1412</sub> ITGDLITINGLNIISK	0.00%	0.00%	0.00%	0.00%	0.00%
FIEPN <sub>1491</sub> NTITVDTQNEFATR	0.00%	0.00%	0.00%	0.00%	0.00%

**Supplementary Table S3.** Measurement of site-specific glycan occupancy in HMW1A with HMW1A abundance fixed with 0.1 mM IPTG and HMW1C abundance altered with varying concentrations of arabinose (0.00002-2%). Asn residues in sequons are indicated in **bold**.

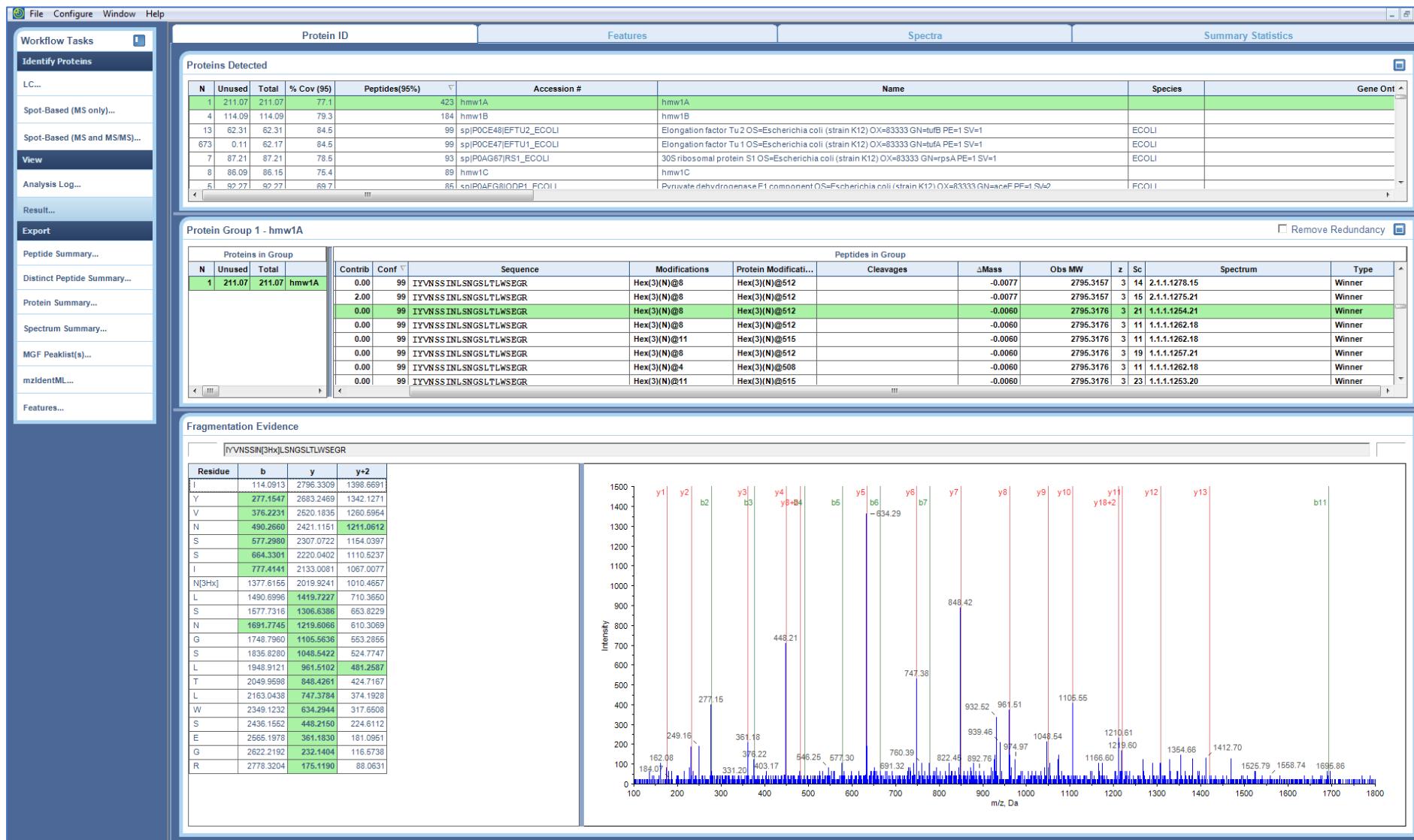
Peptide	2	0.2	0.02	0.002	0.0002	0.00002
QFNIDQNEMVQFLQENN <sub>120</sub> NSAVFNR	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
<b>N</b> <sub>183</sub> FTFEQTK	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
EWLLDPDN <sub>444</sub> VSINAETAGR	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
SN <sub>456</sub> TSEDDEYTGSNSASTPK	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
TTLTN <sub>484</sub> TTLESILK	72.85%	73.99%	47.22%	7.23%	2.60%	2.39%
GTFVN <sub>498</sub> ITANQR	86.67%	80.54%	65.31%	17.61%	9.21%	0.40%
GAN <sub>546</sub> LTIYSGGWVDVHK	88.53%	89.99%	74.04%	25.09%	13.58%	11.13%
FEGTLN <sub>636</sub> ISGK	83.79%	85.73%	70.71%	20.28%	8.34%	7.60%
VN <sub>642</sub> ISMVLPK	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
YFN <sub>773</sub> VSTGSSLR	78.62%	79.21%	51.58%	8.99%	3.31%	2.90%
AITN <sub>912</sub> FTFNVGGLFDNK	17.86%	13.95%	18.47%	7.45%	4.66%	3.23%
TIISGN <sub>964</sub> ITNK	79.22%	80.97%	62.16%	16.32%	6.23%	5.56%
AGVDGENSDSDATNNAN <sub>1029</sub> LTIK	83.08%	84.97%	60.45%	11.96%	5.05%	4.66%
LTQDLN <sub>1044</sub> ISGFNK	68.36%	82.76%	48.45%	10.58%	5.89%	5.06%
VETSGSNN <sub>1107</sub> NTEDSSDNNAGLTIDAK	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
GSNIN <sub>1366</sub> ATSGTLVINAK	16.15%	16.62%	7.10%	0.71%	0.17%	0.11%
VN <sub>1412</sub> ITGDLITINGLNIISK	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
FIEPN <sub>1491</sub> NTITVDTQNEFATR	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%



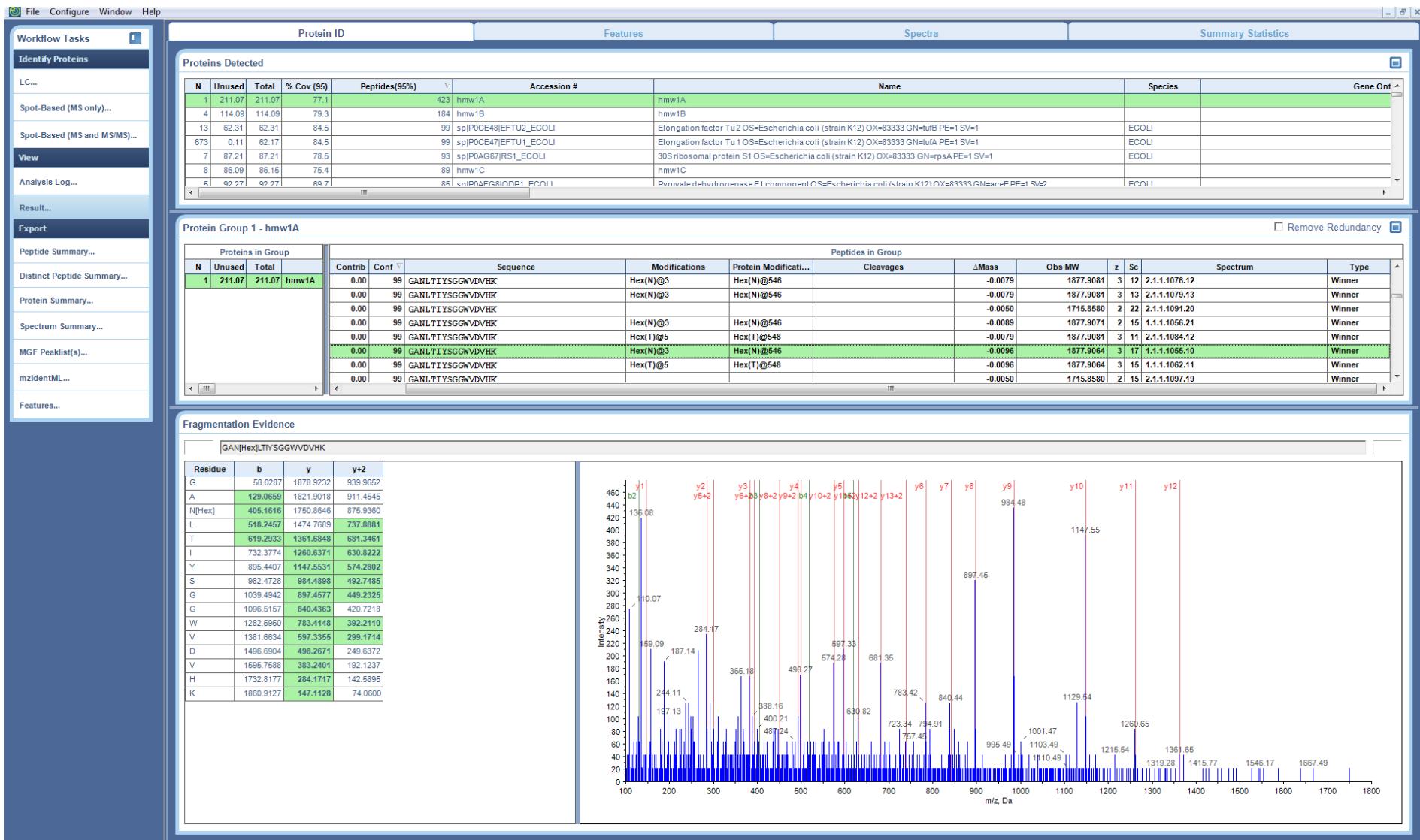
**Supplementary Figure S1. MS/MS peptide identification of TTLTN[+162.053]TTLESILK**



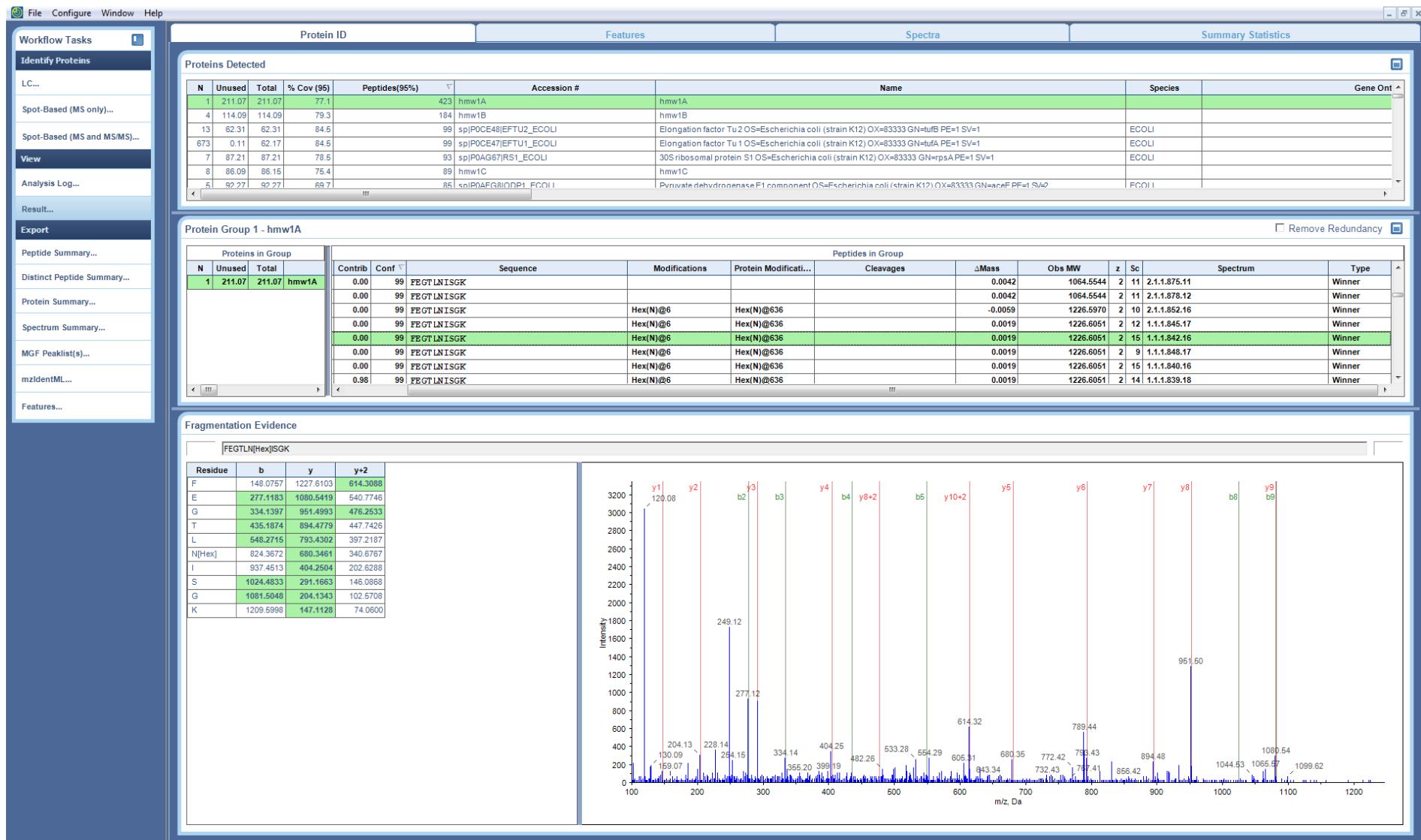
**Supplementary Figure S2.** MS/MS peptide identification of GTFVN[+162.053]ITANQR



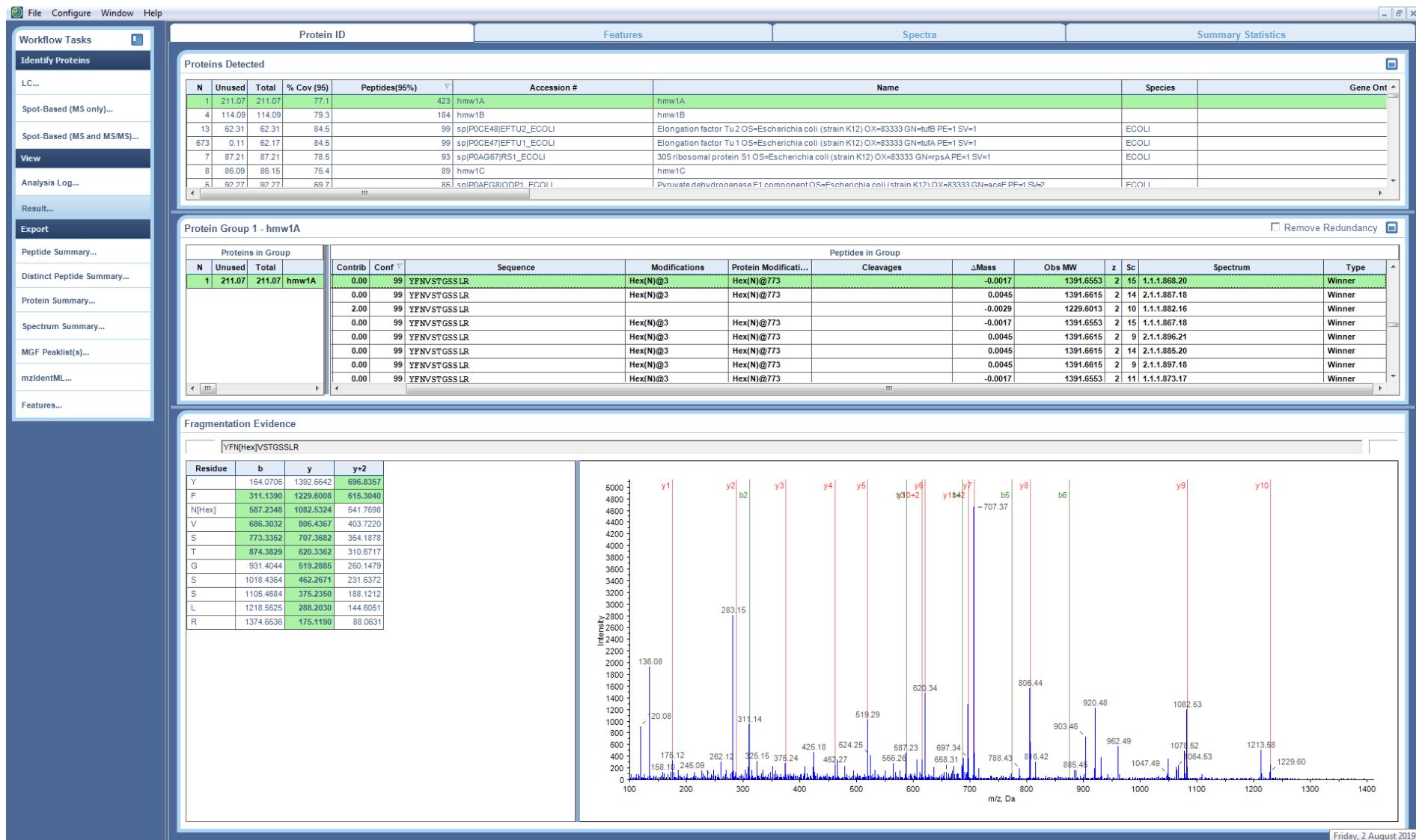
**Supplementary Figure S3. MS/MS peptide identification of IYVNSSIN[+486.159]LSNGSLTLWSEGR.** Site assignment of hexose-Asn is ambiguous, and assigned in Table 2 based on the presence of glycosylation sequons.



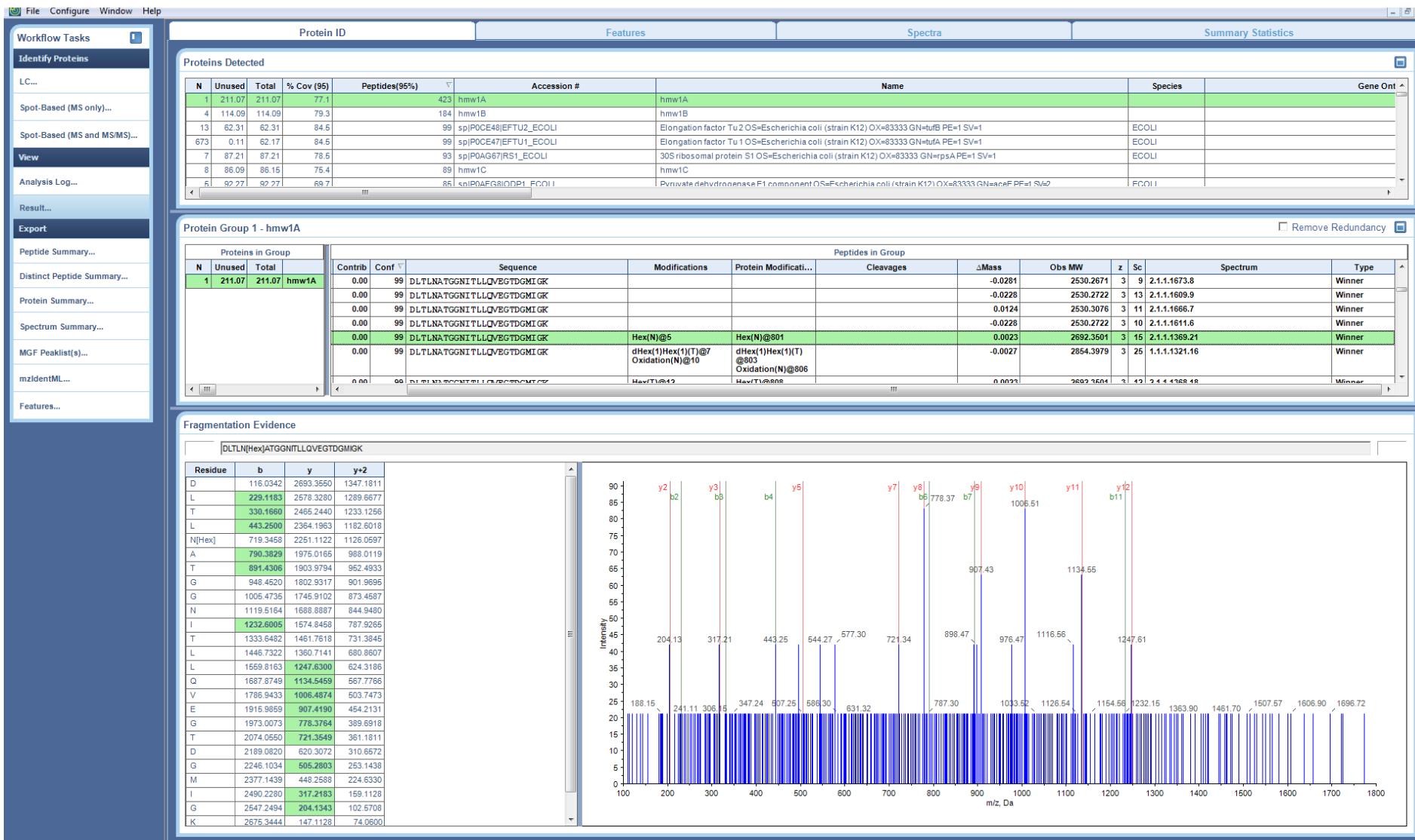
Supplementary Figure S4. MS/MS peptide identification of GAN[+162.053]LTIYSGGWVDVHK



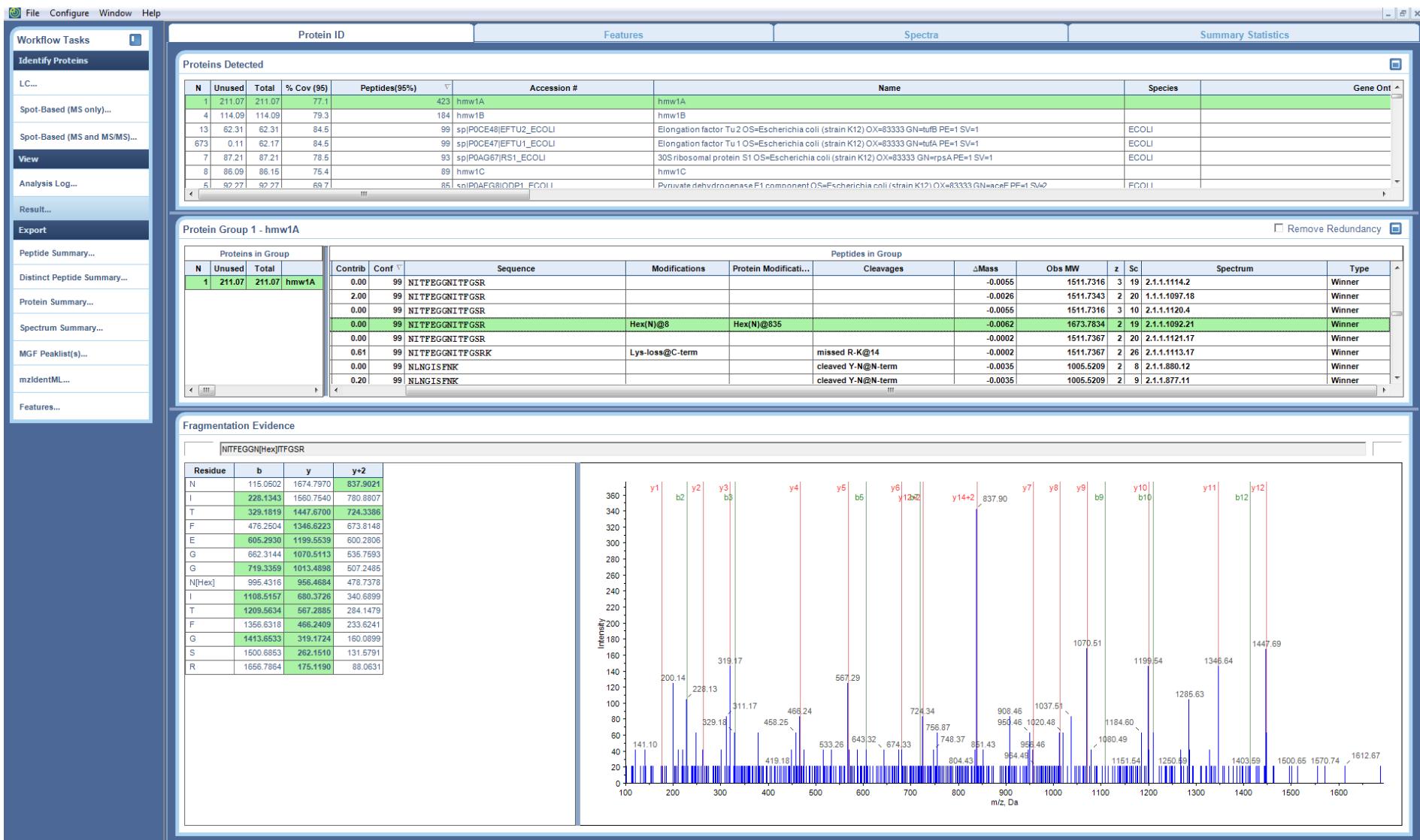
**Supplementary Figure S5. MS/MS peptide identification of FEGTLN[+162.053]ISGK**



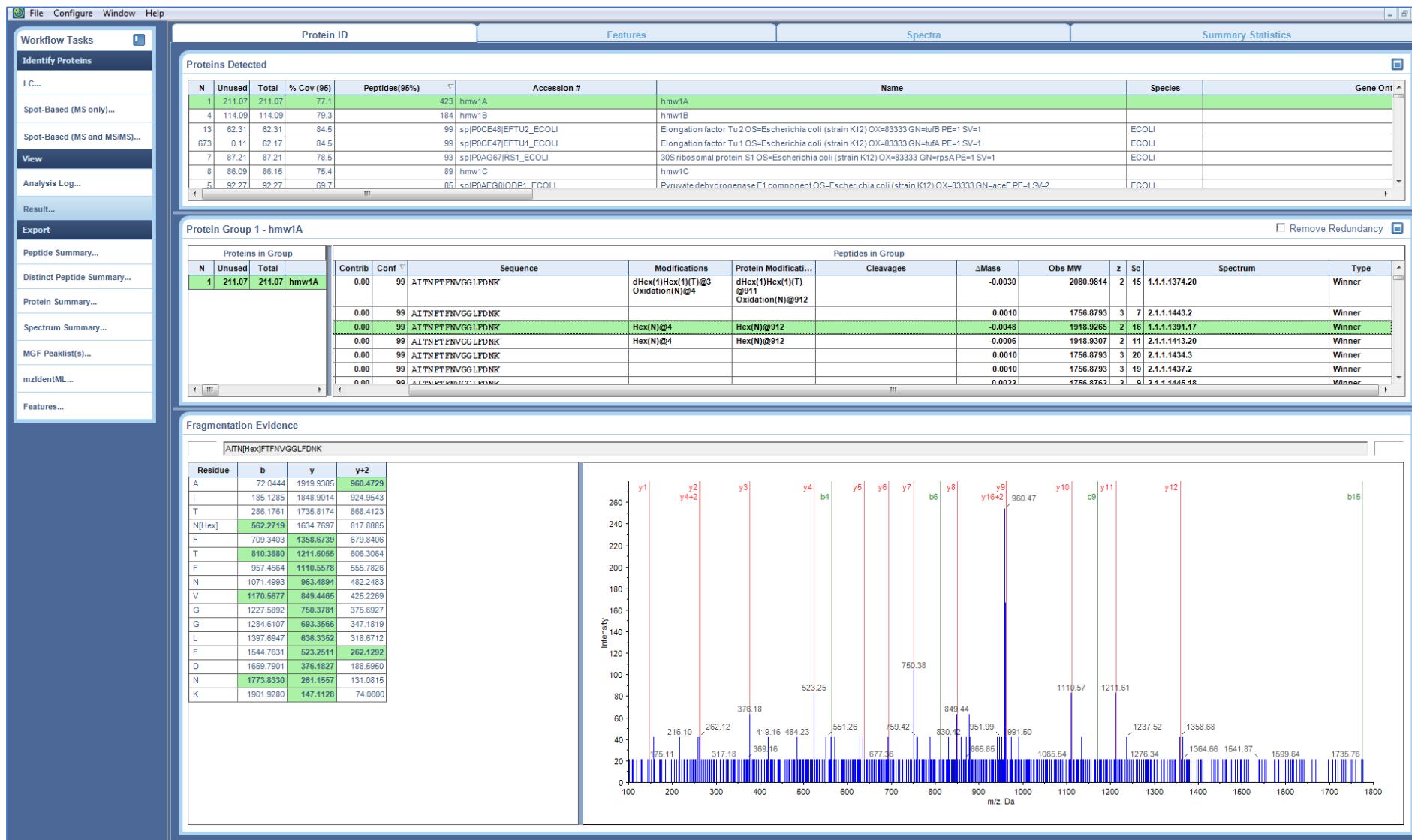
**Supplementary Figure S6.** MS/MS peptide identification of YFN[+162.053]VSTGSSLR



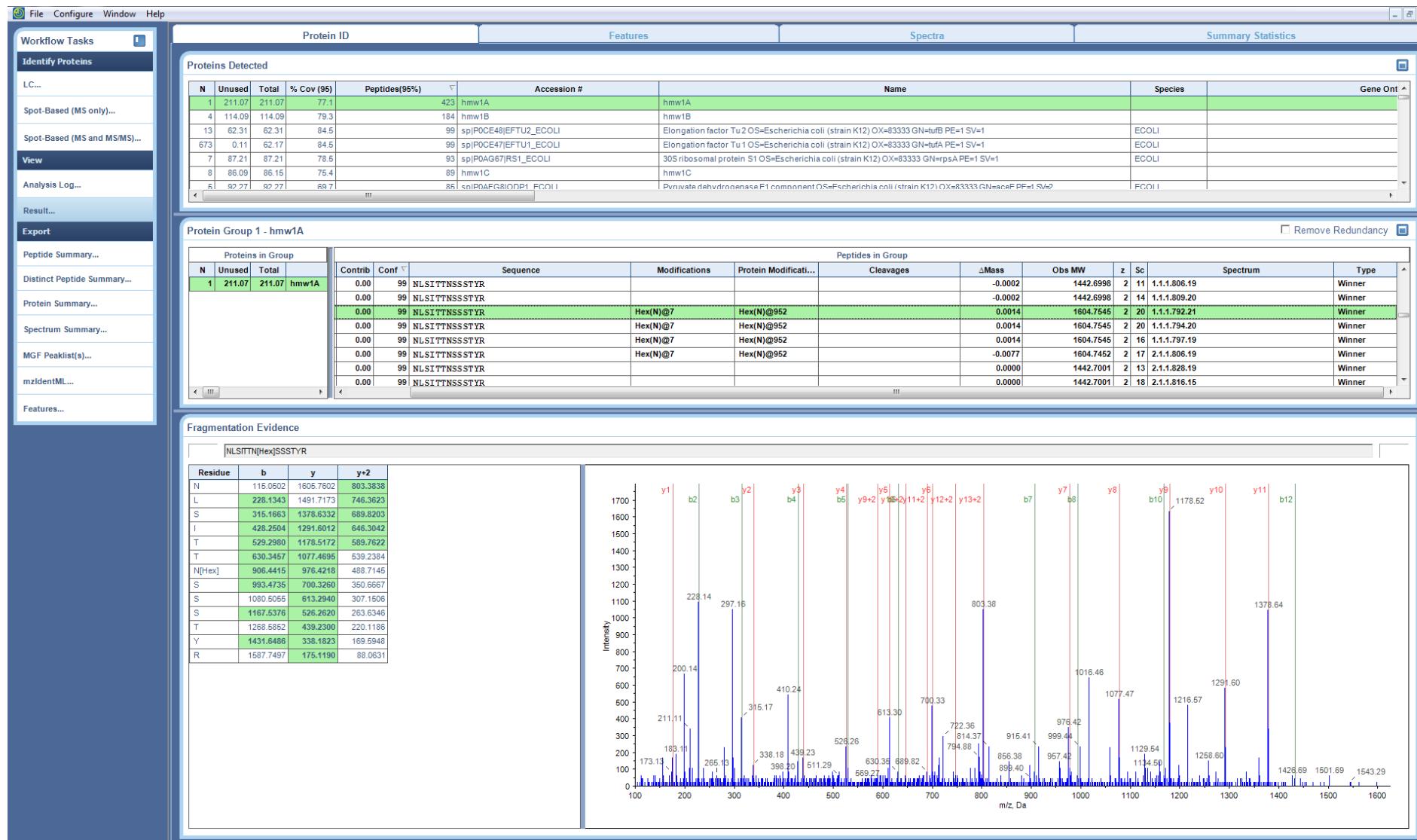
**Supplementary Figure S7. MS/MS peptide identification of DLTLN[+162.053]ATGGNITLLQVEGTDGMIGK**



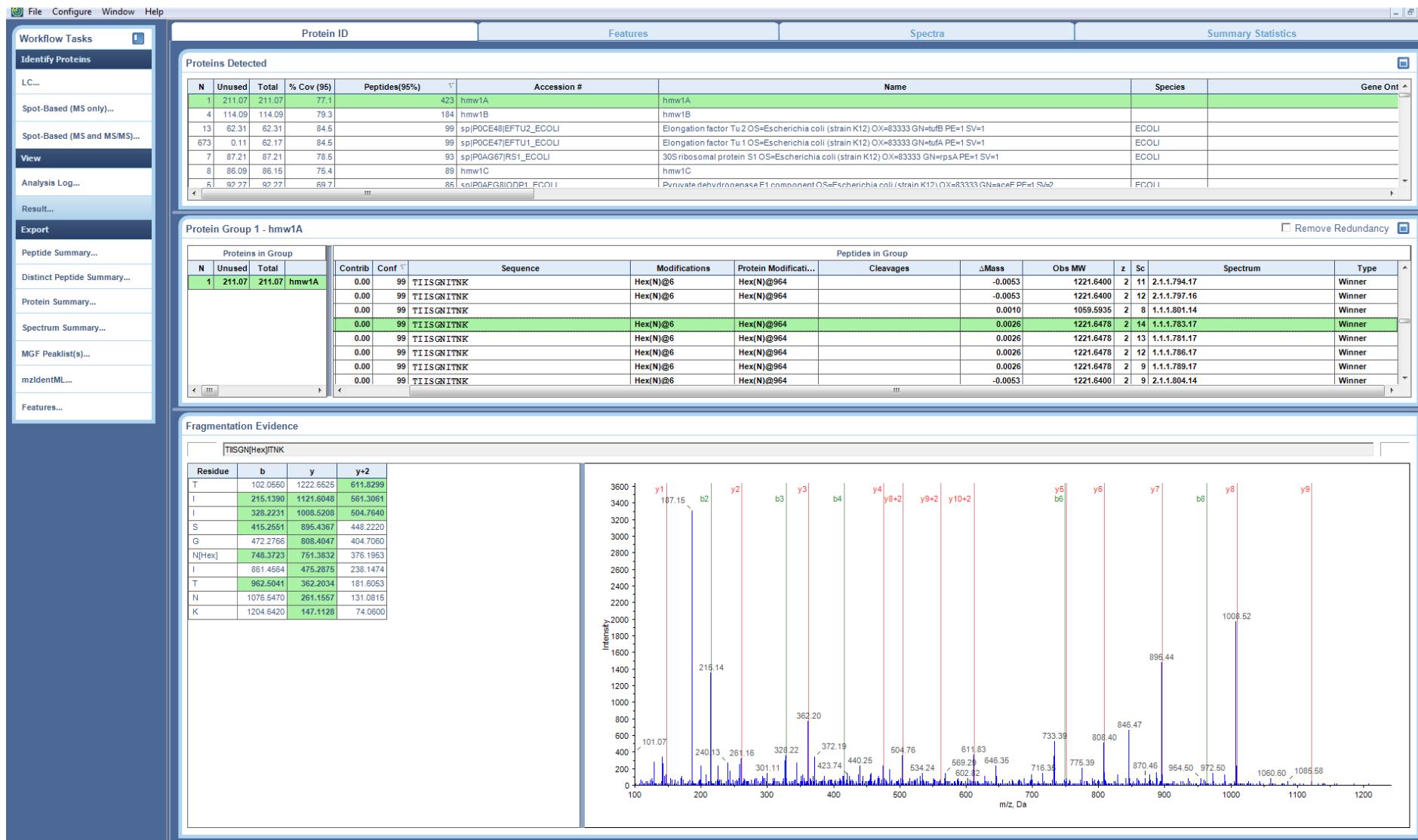
**Supplementary Figure S8.** MS/MS peptide identification of NITFEGGN[+162.053]ITFGSR



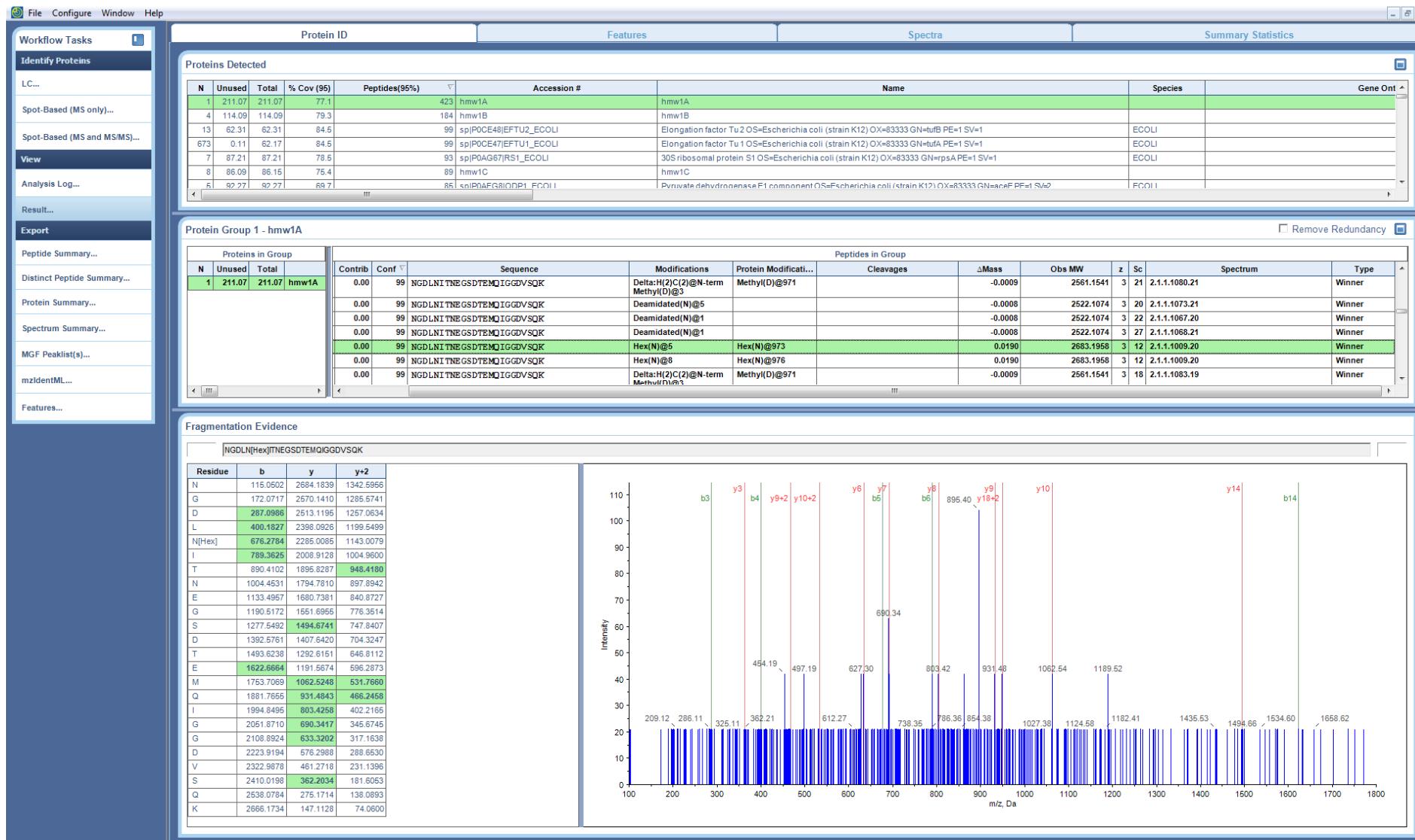
**Supplementary Figure S9.** MS/MS peptide identification of AITN[+162.053]FTFNVGGLFDNK



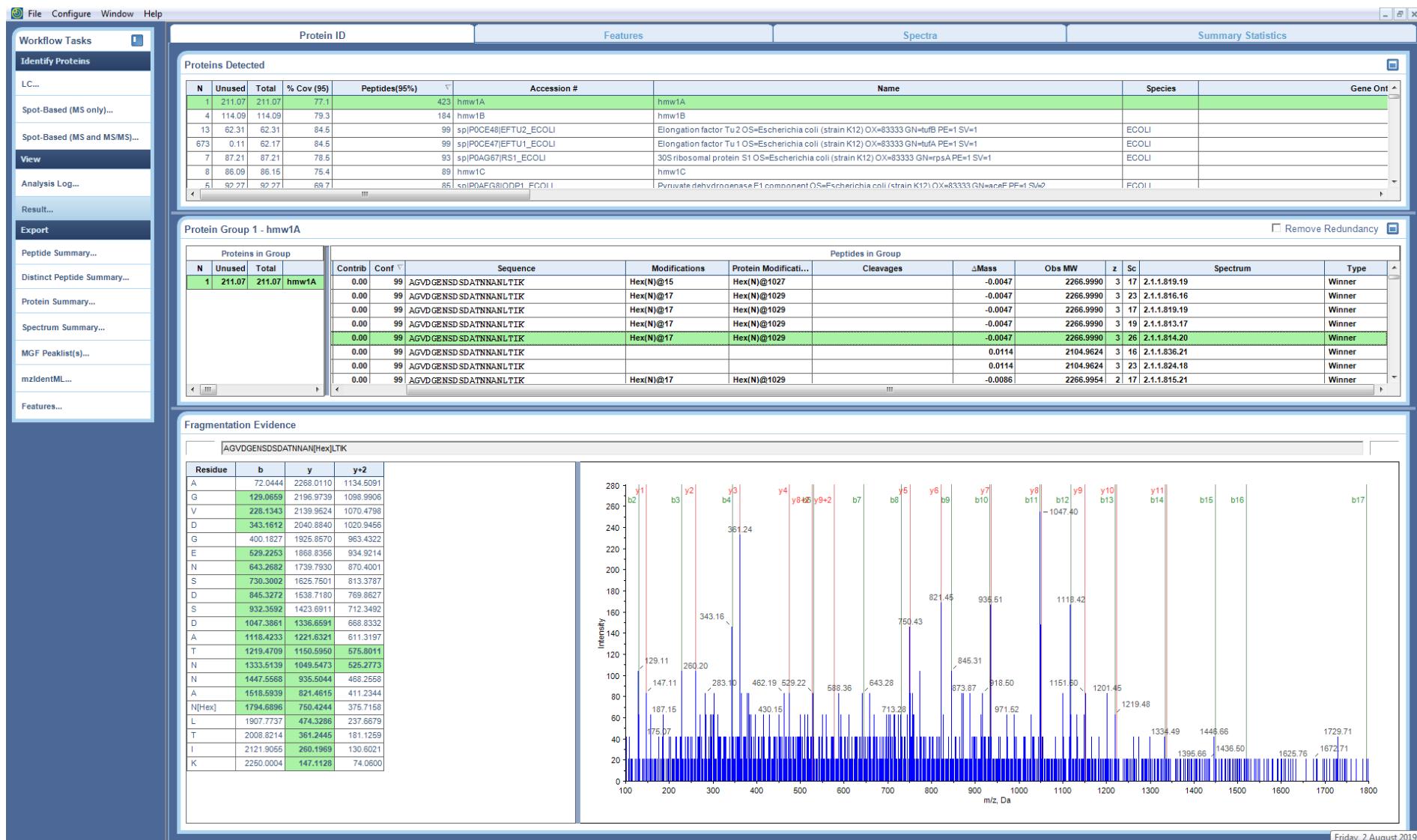
**Supplementary Figure S10. MS/MS peptide identification of NLSITTN[+162.053]SSSTYR**



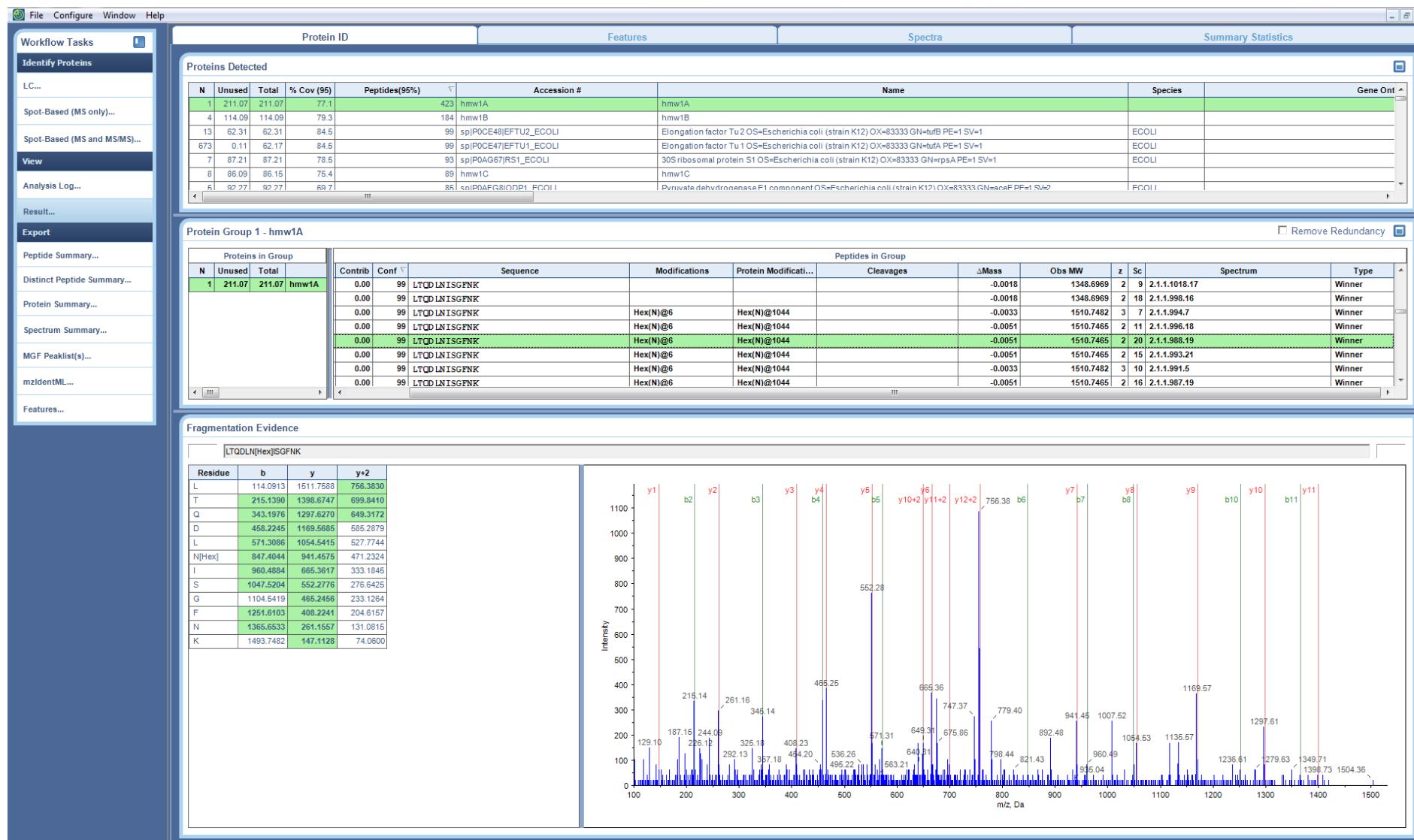
**Supplementary Figure S11. MS/MS peptide identification of TIISGN[+162.053]ITNK**



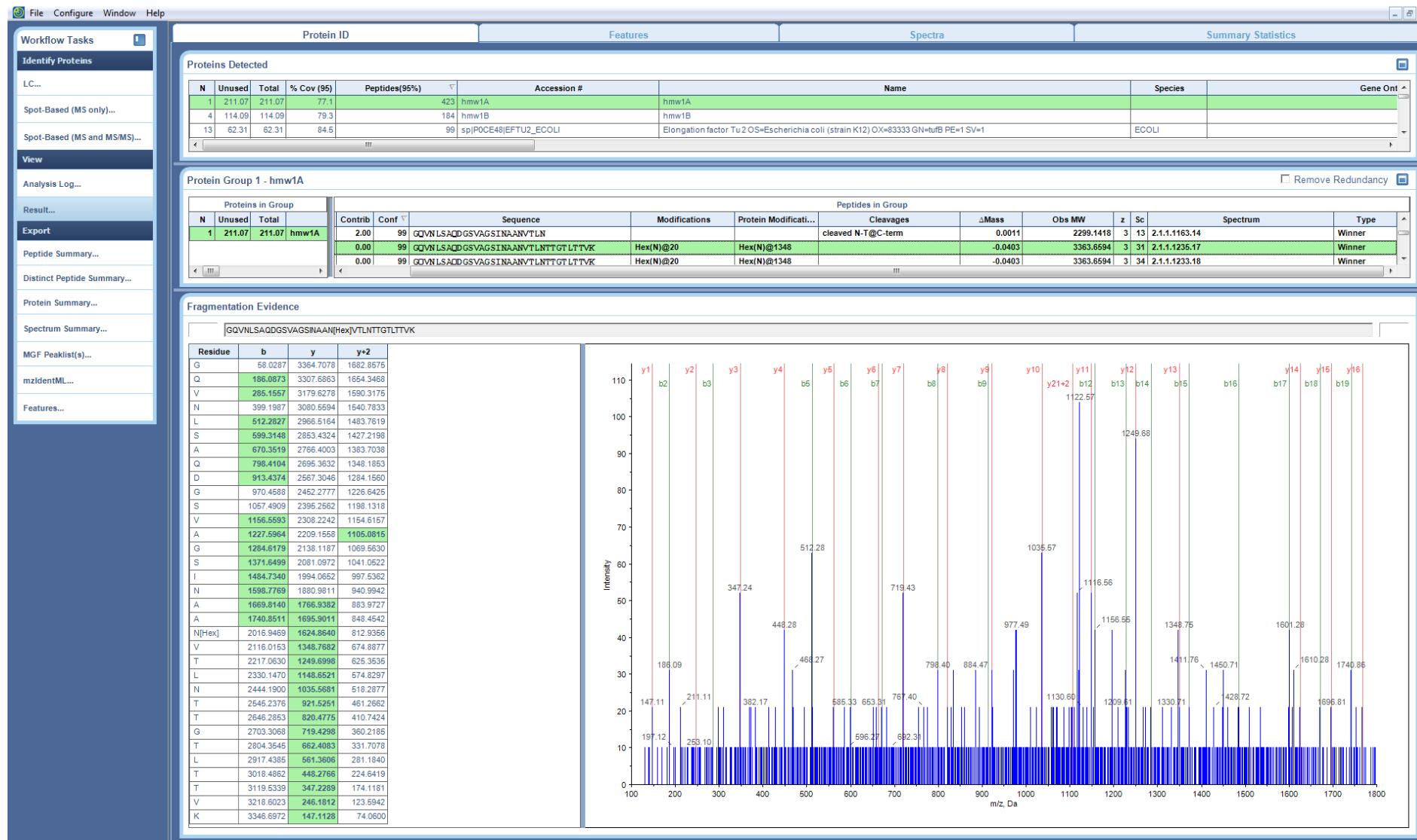
**Supplementary Figure S12. MS/MS peptide identification of NGDLN[+162.053]ITNEGSDETEMQIGGDVSQK**



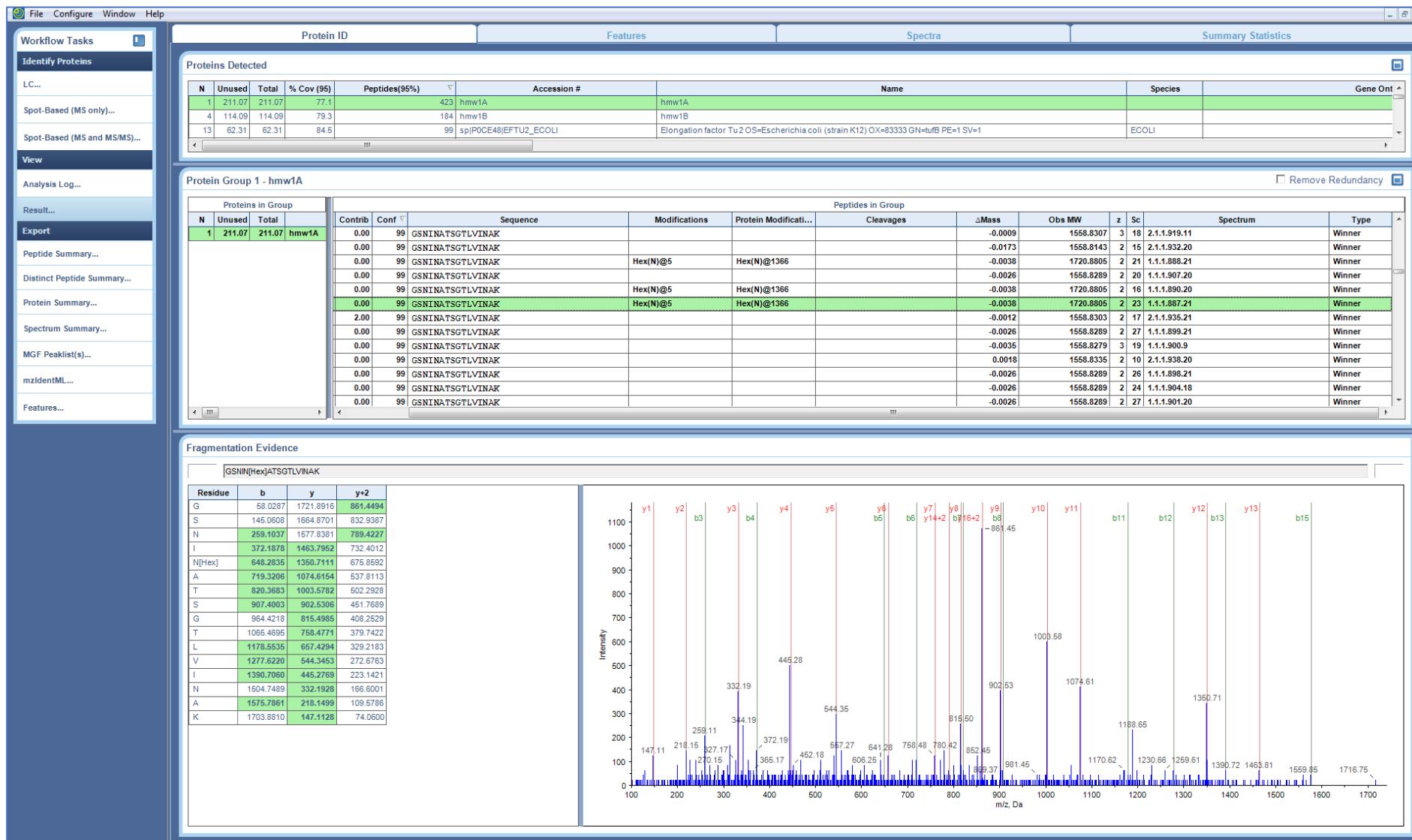
**Supplementary Figure S13. MS/MS peptide identification of AGVDGENSDATNNAN[+162.053]LTIK**



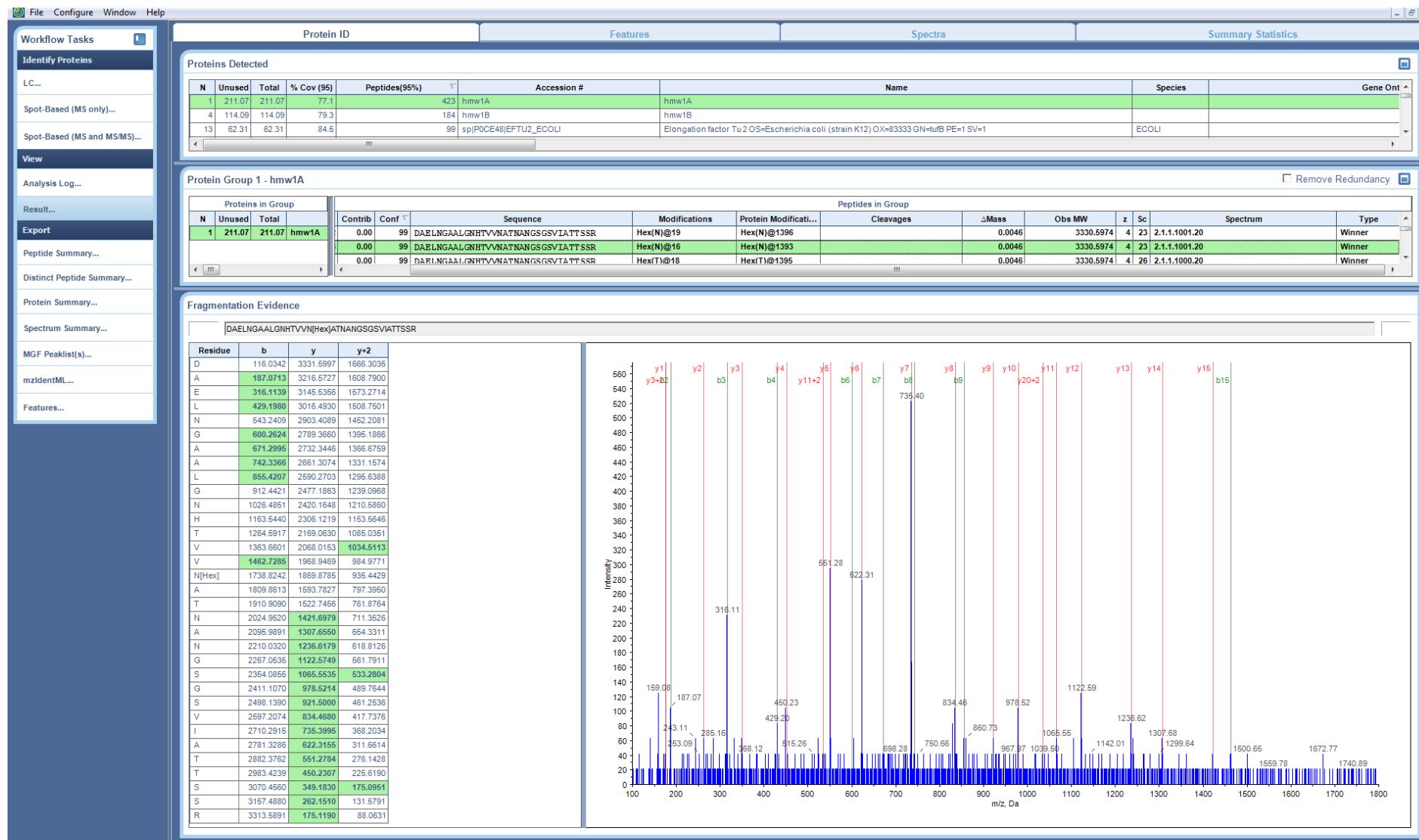
**Supplementary Figure S14. MS/MS peptide identification of LTQDLN[+162.053]ISGFNK**



**Supplementary Figure S15. MS/MS peptide identification of GQVNLSAQDGSVAGSINAAN[+162.053]VTLNNTTGTLLTVK**



**Supplementary Figure S16. MS/MS peptide identification of GSNIN[+162.053]ATSGTLVINA**



**Supplementary Figure S17. MS/MS peptide identification of DAELNGAALGNHTVVN[+162.053]ATNANGSGSVIATTSSR**



**Supplementary Figure S18. MS/MS peptide identification of DAELNGAALGNHTVVN[+486.159]ATNANGSGSVIATTSSR.** Site assignment of hexose-Asn is ambiguous, and assigned in Table 2 based on the presence of glycosylation sequons.

	1	50
Sequence	MNKIYRLKFSKRLNALVAVSELARGCDHSTEKGSEKPARMKVRHLALKPL	
JPred Prediction	--EEEEEE----EEEEEEE-----HHHHHHHHH	
	51	100
Sequence	SAMLLSLGVTSIPQSVLASGLQGMDVVHGTTATMQVDGNKTIIRNSVDAII	
JPred Prediction	HHHHHHHHH-----EEE	
	101	150
Sequence	NWKQFNIDQNEMVQFLQENNNSAVFNRVTSNQISQLKGILDSNGQVFLIN	
JPred Prediction	E-----EEEEEE-----EEEEEEE-----EEEEEE-----EEE-	
	151	200
Sequence	PNGITIGKDAIINTNGFTASTLDISNENIKARNFTFEQTKDKALAEIVNH	
JPred Prediction	--EEE-----E-----EEE-----EE-----EEE-----EEE--	
	201	250
Sequence	GLITVGKDGSVNLIGGKVKNEGVISVNGGSISLLAGQKITISDIINPTIT	
JPred Prediction	-----EEEE-----EEE-----EEE-----EEE-----EEE-----	
	251	300
Sequence	YSIAAPENEAVNLGDIVAKGGNINVRAATIRNQGKLSADSVSKDKSGNIV	
JPred Prediction	EEE-----E-----EEE-----EEE-----EE-----EE	
	301	350
Sequence	LSAKEGEAEIGGVISAQNQQAKGGKLIMITGDVKTLKTGAVIDLSGKEGGE	
JPred Prediction	EE-----EEEEEE-----EEEEEE-----EEE-----EEE-----E	
	351	400
Sequence	TYLGGDERGEKGKNGIQLAKKTSLEKGSTINVSGKEKGGRAIWGGDIALID	
JPred Prediction	EEEE-----EEE-----EEE-----EEE-----EEE-----EEE-	
	401	441
Sequence	GNINAQGSGDIAKTGGFVETSGHDLFIKDNAIVDAKEWLLD	
JPred Prediction	--EEEE-----EEEEEE-----EEE-----E-----	

**Supplementary Figure S19. Secondary structure prediction of HMW1A signal peptide and pro-piece domains using JPred (61).**

	442	491
Sequence	PDNVSINAFTAGRSNTSEDDEYTGSGNSASTPKRNKEKTTLNTTLESIL	
JPred Prediction	---EEE-----	-----EEE-----
	492	541
Sequence	KKGTFVNITANQRIYVNSSINLSNGSLTLWSEGRSGGGVEINNDITTGDD	
JPred Prediction	-----EEEE----EEE-----EEEEEE-----EE-----	
	542	591
Sequence	TRGANLTYSGGWVDVHKNISLGAQGNINITAKQDIAFEKGNSNQVITGQG	
JPred Prediction	-----EEE-----EEE-----EEE-----EEE-----	
	592	641
Sequence	TITSGNQKGFRFNNVSLNGTGSGLQFTTKRTNKYAITNKFEGTLNISGKV	
JPred Prediction	EEEE-----EE-----EEEE-----EEE-----EE-----	
	642	691
Sequence	NISMVLPKNESGYDKFKGRTYWNLTSNVSESGEFNLTIDSRGSDSAGTL	
JPred Prediction	EE-----EEE-----EEE-----	
	692	741
Sequence	TQPYNLNGISFNKDFFNVERNARVNFDIKAPIGINKYSSLNYASFNGNI	
JPred Prediction	-----E-----EEE-----EEE-----	
	742	791
Sequence	SVSGGGSVDFTLASSSNVQTPGVVINSKYFNVSTGSSLRFKTSGSTKTG	
JPred Prediction	EE-----EEE-----EEE-----EE-----EEE-----	
	792	841
Sequence	FSIEKDLTLNATGGNITLLQVEGTDMIGKGIVAKKNITFEGGNITFGSR	
JPred Prediction	EE-----E-----EEE-----EEE-----EEE-----	
	842	891
Sequence	KAVTEIEGNVTINNNANVTLIGSDFDNHQKPLTIKKDVIINSGNLTAGGN	
JPred Prediction	-----EEE-----EEE-----	
	892	941
Sequence	IVNIAGNLTVESNANFKAITNFTFNVGLFDNKGNNSNISIAGGARFKDI	
JPred Prediction	-----	
	942	991
Sequence	DNSKNLSITTNSSSTYRTIISGNITNKGDLNITNEGSDTEMQIGGDVSQ	
JPred Prediction	-----EE-----EE-----EE-----	
	992	1041
Sequence	KEGNLTISSDKINITKQITIKAGVDGENSDSDATNNANLTIKTKELKLQ	
JPred Prediction	-----EEE-----EE-----	
	1042	1091
Sequence	DLNISGFNKAETAKDGSDLTIGNNTNSADGTNAKKVTFNQVKDSKISADG	
JPred Prediction	-----EEE-----E-----	
	1092	1141
Sequence	HKVTLHSKVETSGSNNNTEDSSDNNAGLTIDAKNVTVNNNITSHKAVSIS	
JPred Prediction	--EEE-----EEE-----	
	1142	1191
Sequence	ATSGEITTKTGTINATTGNVEITAQTGSILGGIESSSGSVLTATEGAL	
JPred Prediction	-----	
	1192	1241
Sequence	AVSNISGNTVTVTANGALTTLAGSTIKGTESVTTSSQSGDIGGTISGGT	
JPred Prediction	-----	
	1242	1291
Sequence	VEVKATESLTTQSNSKIKATTGEANVTSATGTIGGTISGNTVNTANAGD	
JPred Prediction	EEE-----	
	1292	1341
Sequence	LTVNGNAGEINATEGAATLTTSSGKLTEAASSHITSAKGQVNLSAQDGSVA	
JPred Prediction	-----EEE-----EE-----	
	1342	1391
Sequence	GSINAANVTLNTTGTLLTVKGSNINATSGTLVINAKDAELNGAALGNHTV	
JPred Prediction	-----	
	1392	1441
Sequence	VNATNANGSGSVIATTSSRVNITGDLITINGLNIISKNGINTVLLKGVKI	
JPred Prediction	-----EEE-----EEEE-----E-----	
	1442	1491
Sequence	DVKYIQPGIASVDEVIEAKRILEKVKDLSDEEREALAKLGVSARFIEPN	
JPred Prediction	E-----EEE-----EEE-----	
	1492	1536
Sequence	NTITVDTQNEFATRPLSRRIVISEGRACFSNSDGATVCVNIADNGR	
JPred Prediction	-----EEEEEE-----EEE-----EEEEEE-----	

**Supplementary Figure S20. Secondary structure prediction of mature HMW1A using JPred (61).**