## DXMSMS Match User Guide 20141124.

Below is the example how to process mass spectrometric crosslinking data with ICC-CLASS DXMSMS Match.

Example is Orbitrap LC-MS-MS/MS run of test peptide TP1 crosslinked with CBDPS-H8/D8, digested with proteinase K and affinity purified with immobilized avidin. MS data were acquired with Xcalibur (ver. 2.1.0.1140) with Mass Tags and Dynamic Exclusion precursor selection methods enabled in global data dependent settings. For CBDPS-H8/D8 mass difference between light and heavy isotopic forms of 8.05824 Da was used in Mass Tags setting. Mass Tag run used a Top 3 method. MS scans (m/z range from 200 to 2000) and MSMS scans were acquired in the Orbitrap mass analyzer at 60000 and 30000 resolution, respectively. Fragment ions for MSMS acquisition were produced by collision-induced dissociation (CID) at normalized collision energy of 35% for 10 ms and activation q = 0.25.

1. Detecting doublets of signals corresponding to light (H8) and heavy (D8) forms of crosslinks.

When Mass Tags method was used for acquisition, doublets of signals can be obtained in two ways: searching MS spectra (A) or using precursors lists (B). Here we will present both ways.

A. Doublets list from MS spectra.





XtractAll Settings	
Raw Files         Xtract File:       ICC-CLASS-ESI\20121026_TP1_CBDPS_PKSol_MT_mz_top3_Xtract.raw         Scan Filter:       FTMS + p NSI Full ms [200.00-2000.00]	Xtract Cancel
Generate Masses Mode Extract What: C M  MH+ C Full Pattern C Input Data after Filtering	Help
Time Range Begin Time: 15.00 End Time: 45.00	
Mass Range     Resolution@400     S/N Threshold       Low Mass:     200     High Mass:     2000.00     30	Advanced >>

This will create new xtracted .raw file with \_xtract addition to the file name.

A.2. Convert this file to the text file using XConvert program of XCalibur package (usually it is located in Xcalibur>system>programs> folder). You can copy it to the data folder together with

Thermofisher.Foundation.FConvert.dll file and run it in there. In newer version of XCalibur package this is substituted by FileConverter.exe program which does not require copy of .dll file.

Thermo File Con	verter					
Conversion source Source data type:	Xcalibur Files *.raw	Folder:	C:\		Bro	<u>w</u> se
File Name			Туре	Size	Date	
0121026_TP1_0	CBDPS_PKSol_MT_mz_top	o3.raw	Xcalibur Raw File	1406024 kb	10/28/2012 4:23:05 /	٨M
<u>20121026_TP1_</u>	CBDPS_PKSol_MT_mz_top	o3_xtract.raw	Xcalibur Raw File	11069 kb	5/14/2013 6:46:58 P!	И
Select <u>All</u> Conversion destination Destination data type:	Clear Selection	Add Job(s)	C:\		▼ Bro	ws <u>e</u>
Jobs Status						
C:\\20121026_TP1_	CBDPS_PKSol_MT_mz_to	p3_xtract.raw	> C:\\20121026_TP1_CBDP5	S_PKSol_MT_mz_top3_xtra	act.bd	
<						>
Remove Job(s)						
	<u>_</u>	nvert	Close	<u>H</u> elp		

This will create text file containing xtracted MS data \_xtract.txt.

A.3. Extract mass list of the run from \_xtract.txt file using Mass List From Xtract Text File program.

🔜 Mass list from xtract text file	
Run	

This will create \_xtract\_MassList.txt file.

A.4. Find list of doublets from mass list using DX ESI LCMS program.

🔜 DX LCMS 20	111006	
DX	8.05021 💌	Da
Tolerance +-	0.01	Da
	Run	

This will create \_xtract\_MassList\_DX.txt file.

A.5. Remove repeats from doublets list with DX ESI LCMS Mass List Filter program.

🔜 DX ESI LCMS Massi	list Filter 20111006 🔳 🗖 🔀
ESI LCMS_DX mass list file	Load
Mass tolerance +- 10	ppm
Retention time tolerance +- 1	🖌 min
	Run

This will create \_xtract\_MassList\_DX\_Filtered.txt file.

This file is ready to be used as input file for DXMSMS Match program.

The file will contain sorted by mass list of light mass, intensity, retention time in minutes, heavy mass, intensity, retention time in minutes and residual difference between heavy and light masses of the doublets minus isotopic label mass difference:

Copy of 2012	1012_TP1_CB	DPS_PKSol_I	nclList_M1_xtract_MassLis	t_DX_Filter	ed - Notepad		
<u>F</u> ile <u>E</u> dit F <u>o</u> rmat	<u>V</u> iew <u>H</u> elp						
402.148163	348798	26.537193	410.192108	799589	26.537193	0063	~
402.148193	663119	38.990860	410.192139	46572	38.990860	0063	_
402.148163	16455	45.375027	410.192169	24858	45.375027	0062	
424.211914	14206	17.411693	432.256439	66010	17.411693	0057	
451.113800	405039	33.537858	459.164154	114629	33.537858	.0001	
452.188080	363882	26.304358	460.238342	362882	26.304358	.0001	
479.177185	1263432	27.176530	487.229889	198017	27.176530	.0025	
481.228882	123000	19.081027	489.285492	131703	19.081027	.0064	
481.229034	4132908	26.265367	489.285065	120924	26.265367	.0058	
481.229309	22578	31.323360	489.285889	38761	31.323360	.0064	
495.067322	49767	45.694530	503.107574	65415	45.694530	0100	
519.138794	166797	15.156860	527.198120	6449	15.156860	.0091	
521.135864	37089	54.853192	529.189758	16116	54.853192	.0037	
528.115540	519829	41.023860	536.166016	328378	41.023860	.0003	
529.189819	4710	54.717192	537.239807	23692	54.717192	0002	
531.191162	117557	21.501525	539.241455	18312	21.501525	.0001	
533.159607	136683	31.647692	541.209778	136342	31.647692	.0000	
534.259033	25088	21.252200	542.309143	19100	21.252200	0001	
535.242798	395640	23.302358	543.292908	279777	23.302358	0001	
536.165222	72122	17.687192	544.207092	20685	17.687192	0083	~

A.6. Create .mgf file of the run using Thermo's Proteome Discoverer program.

This file is ready to be used as input file for DXMSMS Match program and will look like this:

📕 20121012_TP1_CBDPS_PKSol_InclList_M1 - Notepad 🔳 🗖 🔀
<u>File E</u> dit F <u>o</u> rmat <u>Vi</u> ew <u>H</u> elp
MASS=Monoisotopic
BEGIN IONS
TITLE=File4651 Spectrum1 scans: 890
PEPMASS=437.24222 15274.88379
CHARGE=2+
RTINSECONDS=841
SCANS=890
158.09280 32.3536
174.08868 310.584
175.11909 761.718
176.12210 39.779
176.12566 25.3981
191.11519 42.8924
205.14857 53.2761
210.95042 234.749
212.94749 206.815
236.11450 71.974

A.7. Perform search for crosslinks masses assignments and verifications using DXMSMS Match ESI DXH program.

B. Doublets list from .mgf file.

When Mass Tags method is used for the precursor selection, Xcalibur acquisition software detects doublets of signals corresponding to isotopic coding mass difference during the run and sequentially acquires MS/MS spectra of light and heavy isotopic forms. Thus, doublets masses can be deducted from the precursors masses stored in .mgf file of the run. For this option there is version of the program DX MSMS Match ESI MGF DXH, which uses only .mgf file and protein sequences as an input. This version can be used also for data acquired using TopN or TopSpeed methods, but if heavy precursors MS/MS spectra were not acquired, isotopic coding information will be lost for those crosslink assignments.

B.1. Create .mgf file of the run using Thermo's Proteome Discoverer program. This file is ready to be used as input file for DXMSMS Match program. B.2. Perform search for crosslinks masses assignments and verifications using DXMSMS Match ESI MGF DXH program.

🖶 DX MSMS Mat	ch ESI single	TN N	IGF DXH :	20140327	
			Protein	1	ıth
Crosslinker	CBDPS 🗸		sequences		
Мір	508.08899	Da	Load	TRTESTDIKRASSREADYLINKER	
Mde	0	Da	Update		
Mi	-18.01057	Da			
McIrest 1	54.01002	Da	Crosslink	720         1503.64897         +3         1936         4         1503.65012         .8 ppm         (N)KE(R)         1         22-23         (E)STDIKR(A)         1         5-10         25         75         31         1           720         1503.64897         +3         1936         5         1503.65012         .8 ppm         (N)KER(-)         1         22-24         (E)STDIKR(A)         1         5-9         44         75         34         111.	-
McIrest2	455.08625	Da		720 1503.64897 +3 1936 6 1503.65012 .8 ppm (E)STDIK(R) 1 5-9 (N)KER(-) 1 22-24 44 75 34 1.11 720 1503.64897 +3 1936 7 1503.65012 .8 ppm (T)ESTDIK(R) 1 4-9 (I)KR(A) 1 9-10 14 NeM 1 NeM	
DX	8.05021 🗸	Da		720 1503.64897 +3 1936 8 1503.65012 .8 ppm (E)STDIKR(A) 1 5-10 (N)KE(R) 1 22-23 25 75 311. 720 1503.64897 +3 1936 9 1503.65012 .8 ppm (T)ESTDIKR(A) 1 4-10 (I)K(R) 1 9-9 19 NaN 211 720 1503.64897 +3 1936 9 1503.65012 .8 ppm (T)ESTDIKR(A) 1 4-10 (N)K(R) 1 9-9 19 NaN 211	
DX mass + tolerance	.01 🗸	Da		721 1503.64906 +3 1725 1 1503.65012 .5 ppm (1)E(R) 1 9-9 (T)ESTDIKR(A) 1 4-10 8 NaN 1 11.	
DX retention time + tolerance	30 🗸	s		721 1503.64906 +3 1725 3 1503.65012 .7 ppm (N/K(E) 1 22-22 (N)ESIDIAK(E) 1 4-0 8 NaN 11 721 1503.64906 +3 1725 3 1503.65012 .7 ppm (I)KR(E) 1 9-10 (T)ESIDIK(E) 1 4-9 3 NaN NaN 721 1505.64906 +3 1725 4 1503 55012 .7 ppm (N/K(E) 1 32-29 (F)ETDIK(E) 1 5-10 20 67 25 NaN	
Filter DX repeats				721 1503.64906 +3 1725 5 1503.65012 .7 ppm (N)KER(-) 1 22-23 (E)STER(R) 1 5-9 42 67 27 111. 25	
Filter DX	10 🗸	ppm		721 1503.64906 +3 1725 6 1503.65012 .7 ppm (E)SIDIK(R) 1 5-9 (N)KER(-) 1 22-24 42 67 27 1.11 25 721 1503.64906 +3 1725 7 1503.65012 .7 ppm (T)ESIDIK(R) 1 4-9 (I)KR(A) 1 9-10 3 NaN NaN	
mass tolerance Filter DX	60			721 1503.64906 +3 1725 8 1503.65012 .7 ppm (E)STDIKR(A) 1 5-10 (N)KE(R) 1 22-23 28 67 25 NAM 721 1503.64906 +3 1725 9 1503.65012 .7 ppm (E)STDIKR(A) 1 4-10 (I)K(R) 1 9-9 8 NaN111	
time window	00	3		721 1503.64906 +3 1725 10 1503.65012 .7 ppm (T)ESTDIKR(A) 1 4-10 (N)K(E) 1 22-22 8 NAN 1 .11 723 1503.6501 4 .1 100 1 1 100 5011 1 100 1 100 1 1 100 1	
MGF file	Load			722 1503.65001 +2 1682 2 1503.65012 .1 ppm (N)K(E) 1 22-22 (T)ESTDIK(K) A 1 4-10 8 100 1.1. 100	
Digest sites	all 🗸	1		722 1503.65001 +2 1682 3 1503.65012 .1 ppm (1/KKA) 1 9-10 (1/SSIDIK(K) 1 4-9 1/ 6/1 NAN 722 1503.65001 +2 1682 4 1503.65012 .1 ppm (N)KE(R) 1 22-23 (E)STDIKR(A) 1 5-10 42 100 42 NAN	
including CL site				722 1503.65001 +2 1682 5 1503.65012 .1 ppm (N)KER(-) 1 22-24 (E)STDIK(R) 1 5-9 63 100 48 1111 100 722 1503.65001 +2 1682 6 1503.65012 .1 ppm (E)STDIK(R) 1 5-9 (N)KER(-) 1 22-24 63 100 48 1111 100	
Missed up to	al 🗸			722 1503.65001 +2 1682 7 1503.65012 .1 ppm (T)ESTDIK(R) 1 4-9 (I)KR(A) 1 9-10 17 67 1 NeN 722 1503.65001 +2 1682 8 1503.65012 1 ppm (E)STDIK(A) 1 5-10 (N)KF(B) 1 22-23 42 100 42 NeN	
CL sites	К	1		722 1503 65001 +2 1682 9 1503 65012 1 ppm (T)ESTDIR(A) 1 4-10 (I)K(R) 1 9-9 8 10011 100	
Dead-ends only				723 1503.65017 +3 1612 1 1503.65012 ppm (1/2510/LR(A) 1 = +10 (1/(K)) 1 22-2 5 1001 1 100 723 1503.65017 +3 1612 1 1503.65012 ppm (1/(K)(K) 1 = -9 (T)(K)(K) 1 4-10 28 50 211 100	
Intra-peptide only				723 1503.65017 +3 1512 2 1503.65012 .0 ppm (N)K(R) 1 22-22 (T)ESTDIK(A) 1 4-10 28 50 2 11 100 723 1503.65017 +3 1512 3 1503.65012 .0 ppm (N)K(A) 1 9-10 (T)ESTDIK(A) 1 4-9 28 801 NaN	
Precursor +-	2 🗸	000		723 1503.65017 +3 1612 4 1503.65012 .0 ppm (N)KE(R) 1 22-24 (E)STDIR(A) 1 5-10 47 90 37 11 723 1503.65017 +3 1612 5 1503.65012 .0 ppm (N)KER(-) 1 22-24 (E)STDIR(A) 1 5-9 56 100 41 111 83	
tolerance Fragments +-	10			723 1503.65017 +3 1612 6 1503.65012 .0 ppm (E)STDIK(R) 1 5-9 (N)KER(-) 1 22-24 56 100 41 1111 83	
tolerance		ppm		723 1503.65017 +3 1612 / 1503.65012 .0 ppm (1/ESDER(R) 1 4-5 (1/RR(R) 1 5-10 20 80 1 NEW 723 1503.65017 +3 1612 8 1503.65012 .0 ppm (E)STDIKR(R) 1 5-10 (N)KE(R) 1 22-23 47 90 37 .11.	
DXclrest1	4.02511 💌	Da		723 1503.65017 +3 1612 9 1503.65012 .0 ppm (T)ESTDIKR(A) 1 4-10 (I)K(R) 1 9-9 28 50211 100 723 1503.65017 +3 1612 10 1503.65012 .0 ppm (T)ESTDIKR(A) 1 4-10 (N)K(E) 1 22-22 8 50211 100	
DXclrest2	4.02511 🗸	Da		724 1503.65111 +2 1610 1 1503.65012 -7 ppm (I)K(R) 1 9-9 (I)ESTDIKR(A) 1 4-10 21 501 1 100 724 1503.65111 +2 1610 2 1503.65012 -7 ppm (I)K(R) 1 22-22 (I)ESTDIKR(A) 1 4-10 21 501 1 100	
Write to file				724 1503.65111 +2 1610 3 1503.650127 ppm (I)KR(A) 1 9-10 (T)ESTDIK(R) 1 4-9 33 67 11	
	Run			724 1503.65111 +2 1610 4 1503.650127 ppm (N)KE(R) 1 22-23 (E)STDIKR(A) 1 5-10 63 100 43 1111 724 1503.65111 +2 1610 5 1503.650127 ppm (N)KER(-) 1 22-24 (E)STDIK(R) 1 5-9 67 100 49 1111 60	
	Spectrum			724 1503.65111 +2 1610 6 1503.650127 ppm (E)STDIK(R) 1 5-9 (N)KER(-) 1 22-24 67 100 49 1111 60	~

Make necessary selections of the crosslinker, digestion, crosslinking sites and crosslink type restrictions. If user-specific crosslinker is selected, fill in Mip mass value  $([M_{12}+H]^+ = [M_1+H]^+ + [M_2+H]^+ + Mip$ , where  $M_1, M_2$  - masses of free peptides;  $M_{12}$  - mass of inter-peptide crosslink), McIrest1 and McIrest2 mass values  $([M_{12}+H]^+ = [M_1cl+H]^+ + [M_2cl+H]^+ + McIiploss, [M_1cl+H]^+ = [M_1+H]^+ + McIrest, where M_1cl and M_2cl - cleaved peptide products from the cleavage of the inter-peptide crosslink, McIrest, - mass of the cleaved portion of the crosslinking reagent, attached to the cleaved peptide product of the inter-peptide crosslink).$ 

Load protein sequences using Load button. Sequences can be manually modified in the protein sequences window. Hit Update button to convert all the sequences to uniform text. Number of proteins and their lengths will be displayed in protein window.

Load MGF file. Change, if necessary, precursor mass and retention time tolerance values (precision for finding of MS/MS spectra in .mgf file) and fragments tolerance value (precision for crosslink MS/MS fragments matches to theoretical values).

Hit Run button. Message box with a number of the peptides considered for the search will be displayed. Hit OK button to run the search. Precursors mass list, doublets mass list and filtered doublets mass list (if filtering option is selected) files will be written to the same as program's directory. Found matches will be displayed in the Crosslink text box. If necessary, modify restrictions and run search again. To store output of the search in the file, check Write to file button.

Output will contain peptide and protein data for the found matches, scores and "." or "1" indicators of found crosslinks cleavage products. You can select the line of the match you want to inspect and hit Spectrum button. This will open Spectrum View window, in which you can see the selected MS/MS spectrum and inspect fragment masses matches made.



Found matches for light isotopic form will be highlighted in red. In the window on the right corresponding portion of the light and heavy .mgf files will be displayed. In the window on the bottom theoretical pairs (without and with counterpart crosslinked peptide, which correspondently should manifest in the spectrum as single and doublet signals) of cleavage products, b- and y- fragment masses of the crosslink will be listed. If crosslinker containing fragments were detected as doublets in a merged light and heavy spectrum, they are marked with "D" for peptide backbone fragments and "d" for cleavage products.

Spectrum view can be zoomed in by selecting rectangular range within the spectrum field with left mouse button. Zoomed in mass range can be shifted by selecting direction and length of the shift by left mouse button in the space below x-axis of the plot. Zoom out to the original mass range can be done by single left mouse button click within the spectrum field.

Following the spectrum inspection this window can be closed and next spectrum can be selected for the verification of the assignment from the Crosslink text box of the search DXMSMS Match window.

Output file can be opened in Excel, where it can be sorted, filtered and formatted according to the user preferences. Example of the processed output of the CBDPS-H8/D8 crosslinked test peptide TP1, digested with proteinase K, enriched with monomeric avidin beads and analyzed by LC-ESI-MS and MS/MS on Orbitrap instrument using Mass Tags acquisition method, is presented below. Unique inter-peptide Lys-Lys crosslinks were selected.

Deconvoluted MS Peak (MH+) (L)	MS Intensity (L)	MS Retention Time (min) (L)	MGF Precursor # (L)	Original Precursor (m/z) (L)	z (t)	MS Retention Time (sec) (L)	Scan # (L)	Precursor Prediction #	Theorhetical (MH+) (L)	mqq	Protein (1)	AA# Start (1)	AA# End (1)	Modified AA# (1)	AA -1 (1)	Peptide Sequence (1)	AA +1 (1)	Protein (2)	AA# Start (2)	AA# End (2)	Modified AA# (2)	AA -1 (2)	Peptide Sequence (2)	AA +1 (2)	CID Ions Score	Intensity Share Score	401-1	401-2	401-3	401-4	401 Total
1143.51055	467485	32.59	1151.559984	34087	32.59369	-0.0008		2	1143.5108	0.2	1	7	9	6	Т	DIK	R	1	8	9	7	D	IK	R	233	33	1	1	1	1	4
1186.52715	129789	26.20	1194.577318	23574	26.197012	0		4	1186.5278	0.6	1	7	9	6	Т	DIK	R	1	9	10	8	1	KR	A	200	35	1	1	1	1	4
1200.54468	<b>630</b> 1076	22.26	1208.592847	4321319	22.256845	-0.002		1	1200.5435	-1	1	8	9	7	D	IK	R	1	22	24	21	N	KER	-	267	14	1	1	1	1	4
1257.55364	575784	32.81	1265.603563	51460	32.811512	-0.0003		1	1257.5537	0.1	1	7	9	6	Т	DIK	R	1	20	22	19	L	INK	E	163	24	1	1	1	1	4
1258.53826	<b>2</b> 195936	35.08	1266.586962	222441	35.076325	-0.0015		1	1258.5377	-0.4	1	7	9	6	Т	DIK	R	1	7	9	6	Т	DIK	R	100	79	1	1	1	1	4
1314.59763	42559	16.96	1322.64777	13391	16.961503	-0.0001		1	1314.5976	0	1	9	11	8	I	KRA	S	1	22	24	21	N	KER	-	213	32	1	1	1	1	4
1315.56975	193790	26.73	1323.619799	92058	26.728512	-0.0002		2	1315.5704	0.5	1	7	9	6	Т	DIK	R	1	22	24	21	Ν	KER	-	200	35	1	1	1	1	4
1331.59062	50507	31.88	1339.640428	27098	31.881012	-0.0004		2	1331.5905	-0.1	1	5	9	4	E	STDIK	R	1	8	9	7	D	IK	R	170	37	1	1	1	1	4
1332.54973	62909	29.12	1340.600756	6576	29.120012	0.0008		2	1332.5493	-0.3	1	5	9	4	E	STDIK	R	1	21	22	20	1	NK	E	160	40	1	1	1	1	4
1347.54961	360805	28.69	1355.599413	139277	28.69368	-0.0004		4	1347.549	-0.4	1	5	9	4	E	STDIK	ĸ	1	22	23	21	N	KE	ĸ	160	29	1	1	1	1	4
1359.58562	138/18	34.53	1367.635301	231/0	34.53301	-0.0005		2	1359.5854	-0.2	1	6	9	- 5	5	IDIK	к	1	/	9	6	1	DIK	к	190	25	1	1	1	1	4
1372.00301	4559107	10.15	1380.650247	2450172	25 614012	-0.003		/	1372.0031	0.1	1	22	24	21	IN F		-	1	22	24	21	IN I		-	100	22	1	1	1	1	4
1416 61961	05505	23.01	1382.038129	476025	23.014012	-0.0002		4	1416 6191	-0.4	1	5	9	4	C C		n D	1	22	24	21	I NI		A	100	37	1	1	1	1	4
1410.01801	2003222	24.70	1424.008430	470933	24.704423	-0.0004		2	1410.0101	-0.4	1	5	9	2	S F		n D	1	- 22	24	6	т		- P	192	26	1	1	1	1	4
1471 67156	227687	21 51	1479 720859	71367	21 51368	0		2	1440.0174	0.5	1	7	10	4	т		Λ	1	22	24	21	I N	KED		260	20	1	1	1	1	4
1486 64575	91402	16.68	1494 695499	37152	16 677505	-0.0005		2	1486 646	0.2	1	21	24	20		NKER	-	1	22	24	21	N	KER	_	260	20	1	1	1	1	4
1503 64972	638881	25.83	1511 699317	182154	25 834845	-0.0005		6	1503 6501	0.2	1	5	9	4	r F	STDIK	R	1	22	24	21	N	KER	-	192	26	1	0	1	1	3
1632,69389	3335085	25.61	1640.742561	705904	25.614012	-0.0015		7	1632.6927	-0.7	1	4	9	3	т	ESTDIK	R	1	22	24	21	N	KFR	-	157	25	1	1	1	1	4
1659,75107	217941	21.51	1667.800388	47987	21,51368	-0.0009		3	1659.7512	0.1	1	5	10	4	F	STDIKR	A	1	22	24	21	N	KFR	-	221	28	1	1	1	1	4
1730.78794	60303	22.70	1738.83801	20399	22.700515	-0.0001		7	1730.7883	0.2	1	5	11	4	E	STDIKRA	S	1	22	24	21	N	KER	-	188	28	1	1	1	1	4