Combined workflow for XL-MS data processing: Application to the FtsH-HflK-HflC membrane protein complex

Hatice Akkulak[1,†], *H. Kerim İnce*[1,†], *Gunce Goc*[2,3], *Burak V. Kabasakal*[2,4,*] *and Sureyya Ozcan*[1,5,*]

[1] Department of Chemistry, Middle East Technical University, Ankara, 06800, Turkiye

[2] Turkish Accelerator and Radiation Laboratory (TARLA), Ankara University, Gölbaşı, Ankara, 06830, Turkiye

[3] Department of Biology, Kocaeli University, Umuttepe, Kocaeli, 41380, Turkiye

[4] School of Biochemistry, University of Bristol, Bristol, BS8 1TD, UK

[5] Cancer Systems Biology Laboratory (CanSyL) Middle East Technical University, 06800 Ankara, Turkey

[†] These authors contributed equally to this manuscript.

* To whom correspondence may be addressed

KEYWORDS: Cross-linking, Mass Spectrometry, Structural Proteomics, Membrane Protein Complexes

ABSTRACT

Recent advances in mass spectrometry (MS) yielding sensitive and accurate measurements along with developments in software tools have enabled the characterization of complex systems routinely. Thus, structural proteomics and cross-linking mass spectrometry (XL-MS) have become a useful method for structural modeling of protein complexes. Here we assessed the utility of commonly used XL-MS software tools, XlinkX, MS Annika, MeroX, and MaxLynx to elucidate the protein interactions within a membrane protein complex containing FtsH, HflK, and HflC, over-expressed in *E.coli*. The MS data were processed using various XL-MS software-tools. The comparison involved both software performance criteria such as processing time, computer usage, and number of identified inter and intra-protein interactions. Each interaction was manually checked using the raw MS/MSMS data to verify inter- and intra-protein interactions. Total 106 interactions of HflC, 96 interactions of FtsH, and 23 interactions of HflK, including the FtsH-HflK-HflC complex, with other proteins were determined. Approximately 93% of the interactions were found in the purified complex sample, which can be combined with molecular docking for single or complex protein structural modeling. In addition to XL-MS, according to protein mapping and network studies, genes encoding proteins were in 60 and 108 different categories for the purified protein complex and the solubilized membrane sample. The top 15 biological processes were determined, 60% of which are involved in protein translation and quality control. Our study will provide a complete workflow for XL-MS studies of protein complexes, an overview of XL-MS software.

INTRODUCTION

The role and function of proteins in the cell could only be revealed by uncovering all aspects of a protein: structure, alterations, localizations, abundances and the protein-protein interactions that are the basis of proteomics research. Since proteins might gain their function through forming complexes with various proteins under various conditions, functional proteomics is one of the most effective approaches to understand the role of the proteins by identifying protein-protein interactions[1].

Recent developments in proteomics with more rapid and sensitive mass spectrometry (MS) analyses have enabled us to determine various features in a single MS run[2]. Current software used in MS allows the analysis of thousands of spectra, leading to deciphering even very complex matrices to unravel thousands of features. Software tools have made MS spectra analysis processes faster and automated, which aided large-scale clinical and multidisciplinary studies[3].

Proteomics has become a fashionable method for structural studies rather than a complementary technique by providing comprehensive information especially in protein-protein interactions[4,5]. X-ray crystallography and electron cryo-microscopy (cryo-EM) often reveal only one form of a protein complex with interactions defined by the organization of the protein components whereas cross-linking mass spectrometry (XL-MS) can give overall information from multiple forms of a protein complex simultaneously. XL-MS software are designed to identify interacting residue pairs and give information about connections between these residues by utilizing spectrum data. Numerous software[6] are available, including Kojak[7], MaxLynx[8], MeroX (StavroX)[9–12], MS

Annika[13], OpenPepXL[14], pLink2[15], XlinkX[16], and XiSEARCH[17], for various cross-linking applications.

Crucial processes of the cell, such as signal transduction, membrane transportation, signal detection, cell-cell interactions are operated by membrane proteins. Therefore, 25% of cell proteome is composed of membrane proteins, making them an important target for developing new therapeutics[18]. Folding and degradation of membrane proteins are controlled by chaperones and proteases[19]. As a member of AAA protease family, FtsH has a cytoplasmic ATPase domain and possesses a proteolytic activity with a zinc-binding site to degrade the improperly folded integral membrane proteins[20,21]. HflK and HflC are known to form a very large membrane-bound complex with FtsH, and modulate FtsH proteolytic activity[19,21,22]. FtsH-HflK-HflC complex plays a chaperone role to stabilize proteins in mitochondria[20], and a modulator role to limit ATPase activity of FtsH[22,23].

Here we used a cross-linking MS approach to investigate the structure and the interactions of a recombinantly produced bacterial membrane protein complex composed of three proteins, FtsH, HflK, and HflC. We compared commonly used XL-MS software, XlinkX, MS Annika, MeroX, MaxLynx, with their efficiency, performance, output for number of peptides and cross-links for both the purified protein complex and total solubilized membrane samples. The relationship of the complex proteins with other proteins were also investigated and analyzed through common protein-protein interaction (PPI) databases, STRING[24], IntAct[25,26], DIP[27], and MINT[28]. Our findings will guide protein interaction and structural proteomics studies using XL-MS.

RESULTS AND DISCUSSION

This study offers a comprehensive workflow to get insights into a membrane protein complex structure and protein interactions using XL-MS (Fig. 1).

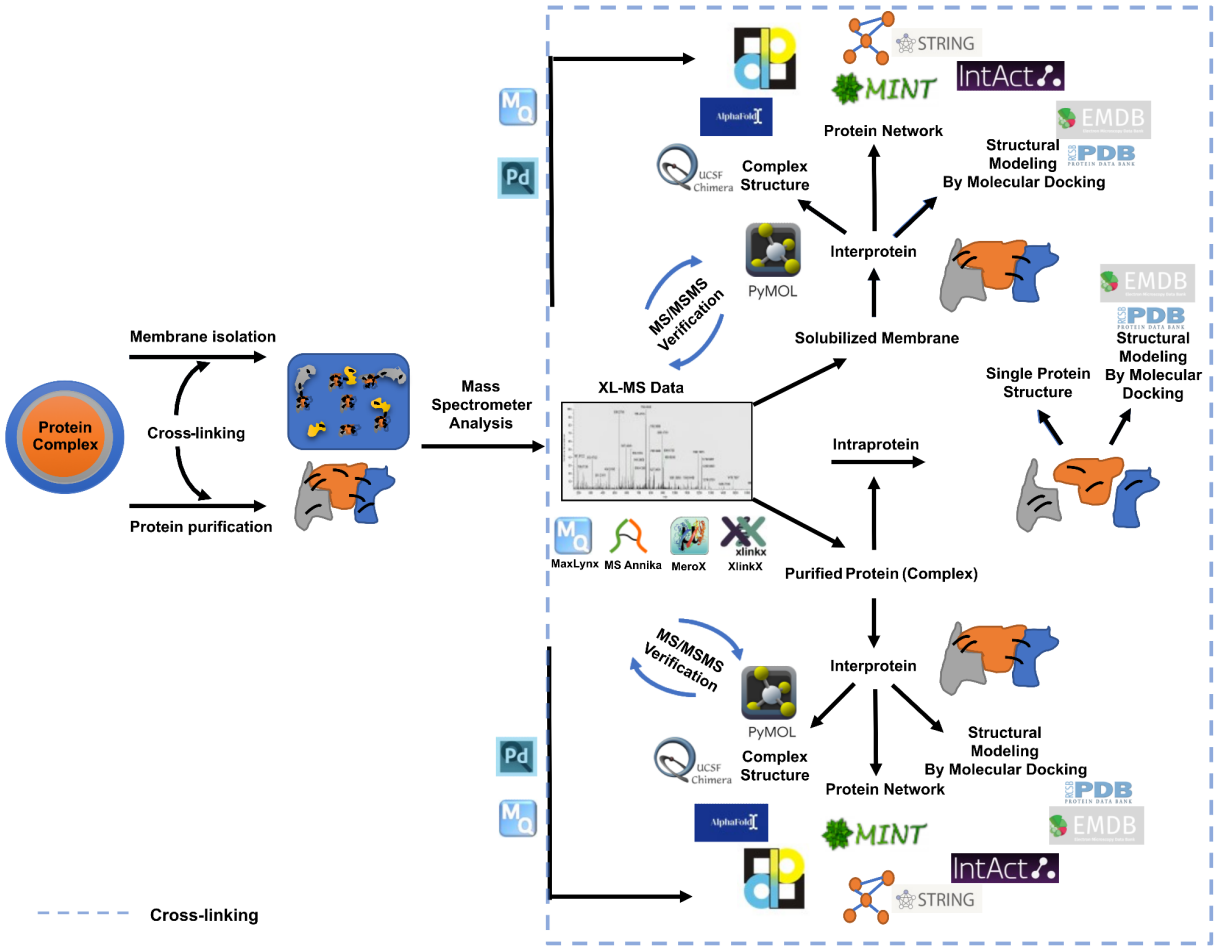


Figure 1. The overall workflow of XL-MS analysis: General steps of XL-MS studies and possible application areas.

The intra- and inter-protein interactions within and between the large multimeric membrane protein complex, FtsH-HflK-HflC were analyzed in the purified form and over-expressed in the solubilized membrane to obtain both structural and protein network analyses. The proteins were cross-linked with DSBU and then digested by using trypsin. High

resolution MS spectra were analyzed using common XL-MS software MaxLynx, MeroX, MS Annika, XlinkX. The intra-protein and inter-protein interactions were further confirmed using MS and MS/MS data.

XL-MS Analysis

To identify inter- and intra-protein interactions, signature fragment ions representing cross-linkers and peptides were searched through four software.

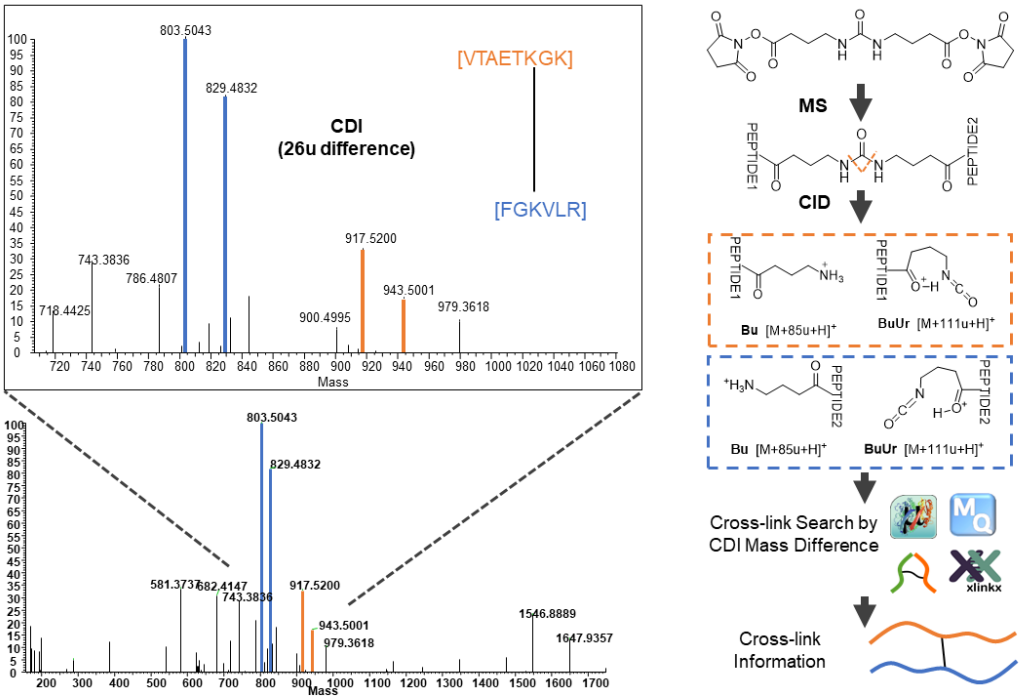


Figure 2. Fingerprint fragments confirming cross-linked peptides.

A representative MS/MS spectrum of a cross-linked peptide pair is presented in Fig. 2. The MS cleavable cross-linker, DSBU, provided a significant advantage, as software were able to confirm specific interactions using the mass difference between two cleavages of cross-links. The DSBU was covalently connected to residues of proteins using NHS ester reactive sides. The Collision-Induced Dissociation (CID) cleaves DSBU between carbonyl and amine parts and

results in two characteristic diagnostic ions. The diagnostic ion corresponds to the peptide and cross-linker with protonated amine group at the cleavage point (PEPTIDE+Bu), which provides 85u mass adduct on the peptide[11,12]. In Fig. 2, the FGKVLR peptide with Bu part of the cross-linker on the lysine gives the peak of 803.5043. Also, the VTAETKGK peptide with Bu part of the cross-linker on the lysine peak appears at 917.5200. The other diagnostic ion is the BuUr part of the cross-linker bound to the peptide, where the nitrogen on amine group loses one proton due to the cross-linker cleavage, resulting in a 111u mass adduct to the connecting peptide (PEPTIDE+BuUr). Fig. 2 shows the FGKVLR peptide with BuUr part of the cross-linker on the lysine giving the peak of 829.4832, and the VTAETKGK peptide with BuUr part of the cross-linker on the lysine peak appearing at 943.5001. The cross-link search was conducted using specific software designed to search 26 u mass difference owing to the cleavage of the cross-linker described above[8,12,13,16].

The XL-MS software predominantly uses lysine specific cross-linking sites to identify interactions. Literature also suggests that serine, threonine, and tyrosine could be cross-linked to DSBU[29]. However, corresponding connections have not been validated in tandem MS level up-to-date. Among four software, only MeroX enables simultaneous multi/variable amino acid-crosslinker search. A representative fragmentation pattern confirming threonine-DSBU interactions is shown in Fig. 3.

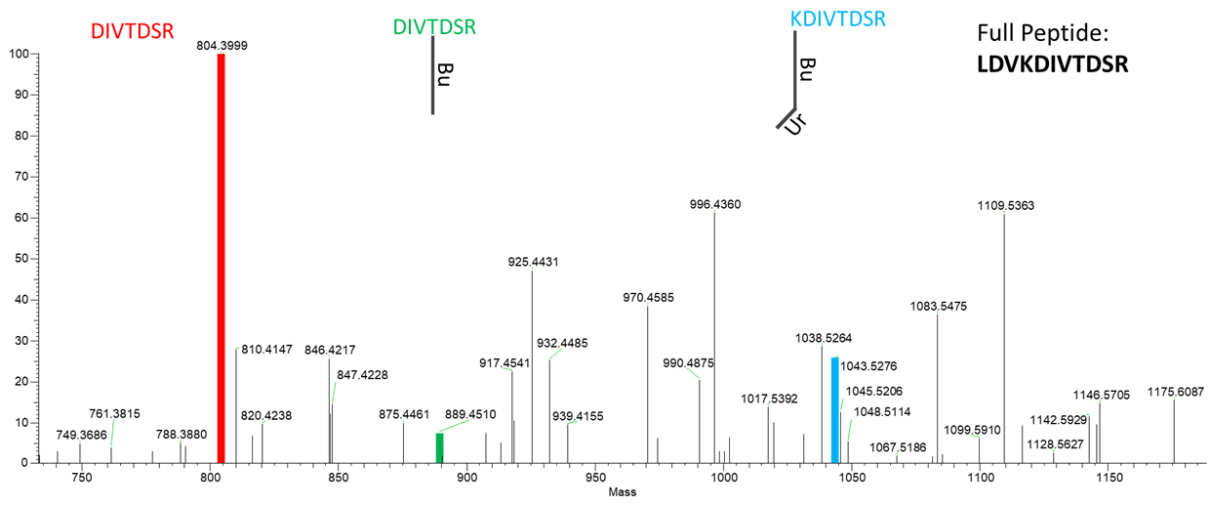


Figure 3. Different cleavage patterns of the cross-linker and backbone revealing the cross-link between threonine and lysine.

The deconvoluted MS/MS spectrum of LDVKDIVTDSR-GVIGKYTmDR peptide pair was given in Fig. 3. In addition to two diagnostic ions specific to peptide-DSBU (85 and 111), several ions were observed to validate the threonine-DSBU connection. The peak at the 1043.5276 mass unit belongs to Y8 cleavage of the peptide LDVKDIVTDSR and BuUr part of the cross-linker which is KDIVTDSR+BuUr. In this case, the cross-linker is bound to lysine. The peak at 889.4510 mass unit belongs to Y7 cleavage of the peptide LDVKDIVTDSR and Bu part of the cross-linker which is DIVTDSR+Bu. In this example, there is no lysine in the sequence, but threonine. Furthermore, the peak at 804.3999 mass unit belongs to Y7 cleavage of the LDVKDIVTDSR yielding the DIVTDSR without the cross-linker. Data suggested that the DSBU was linked to both lysine and threonine residues and corresponding peptides were co-eluted.

The cross-link search through the whole proteome yielded 1093 unique interactions (Supplementary Table 2-1). The total number of inter-protein and intra-protein connections were 620 and 473, respectively.

FtsH-HflK-HflC Complex Inter-protein Interactions

There were 30 inter-protein interactions obtained from four software and two different samples. Excluding repeating and unspecified connections from different software, 20 interactions between the complex proteins were found (Table 1). The unspecific connections confirm the link not on amino acid level, but in consecutive peptide levels allowing uncertainty up to 4 AAs. Four of the interactions were found through all four software considering both samples. Among them, one key interaction between FtsH and HflK was only found by MaxLynx from the purified protein sample. The number of interactions between FtsH and HflC was 7, and the number of interactions between HflK and HflC was 12 from all the software and samples. The highest number of unique connections was obtained from the purified protein complex sample by MaxLynx. MeroX resulted in 5 unspecific connections from the purified protein complex and 2 from the solubilized membrane.

Table 1. Inter-protein interactions of three proteins in the complex obtained in two samples (PP: purified protein, SM: solubilized membrane) analyzed through four software, MaxLynx, MeroX, MS Annika, XlinkX.

		MaxLynx		MeroX		MS Annika		XlinkX	
	Connection Points	**PP**	**SM**	**PP**	**SM**	**PP**	**SM**	**PP**	**SM**
FtsH-HflK	31-133	+							
FtsH-HflC	84-120	+				+		+	
	61-211	+							
	60-211	+				+			
	31-63	+		+*		+		+	
	31-36					+			
	60-120					+			
	61 - 120		+						
HflK-HflC	172-36	+				+			
	346-327	+	+		+	+		+	
	172-122	+				+			
	294-327	+							
	198-137	+	+	+*	+	+	+	+	+
	294 -281	+						+	+
	198 - 141/143			+*	+*				
	294 - 248		+	+*	+*	+		+	+
	294 - 328/330/331/333			+*					
	198 - 82							+	
	198 - 84							+	
	310 - 281						+		+

*Unspecific connection

FtsH, HflK, HflC Intra-protein Interactions

Total of 119 intra-protein interactions were obtained from four software and two different samples, and 86 of them were non-repeating and unique interactions (Supplementary Table 2-2). There were 70, 14, and 2 intra-protein interactions of FtsH, HflC, and HflK, respectively from all the software and samples. In the purified protein sample, XlinkX gave the highest result with 39

interactions. The lowest number of unique intra-protein interactions was obtained by MS Annika in the solubilized membrane sample with 7 interactions. Fifteen of the interactions were found in all the software considering both samples.

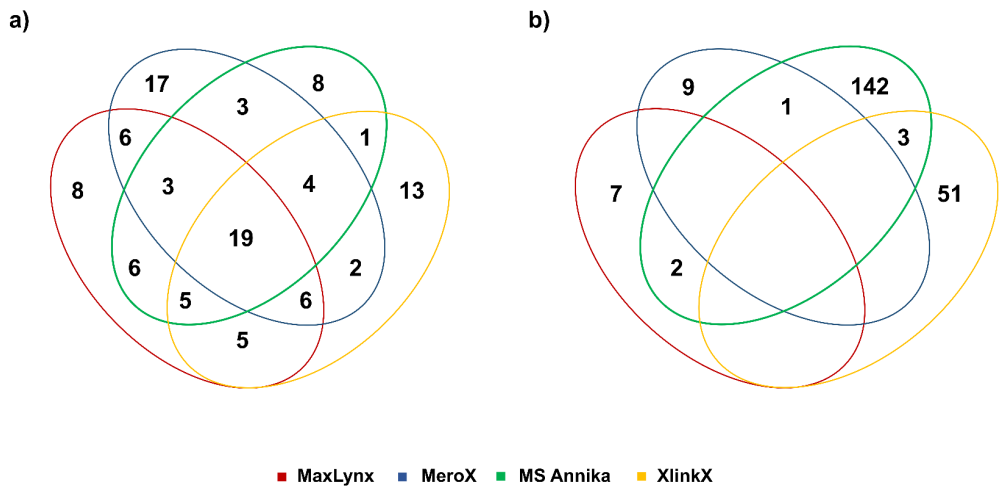


Figure 4. Venn diagram of interactions a) within and between complex proteins and b) interactions of complex proteins with other proteins.

The non-repeating and specific inter- and intra-protein interactions found in both samples were compared through all software based on the interactions within and between the complex proteins and the interactions of proteins in the complex with other proteins (Fig. 4). Nineteen of the connections within and between the complex proteins were obtained in all four software. MeroX gave the highest result with a total of 60 connections, whereas MS Annika gave the least with 49 connections. MeroX, in addition, gave the highest number of interactions (17 interactions) not found in other software (Fig. 4a). On the other hand, most of the interactions between the complex proteins and other proteins were obtained through MS Annika (Fig. 4b).

Compared with the previous results, MS Annika surprisingly caught a successful number of protein-protein interactions with 148 interactions including both samples. On the contrary, MaxLynx and MeroX gave the least number of interactions (9 and 10, respectively). Furthermore, there are not any common interactions found by all software. MeroX and MS Annika found 1 common interaction, MaxLynx and MS Annika found 2, MS Annika and XlinkX found 3.

The FtsH-HflK-HflC complex was analyzed with the FASTA files of only these three proteins in all software for close-up examination of the interactions and for the comparison with proteome-wide search. The most abundant 10 inter- and intra-protein interactions from both samples were listed in Supplementary Table 3. Additional miscleavages and modified peptides were indicated in the table with different symbols. Since serine, threonine and tyrosine modifications were set as modification sides, we clearly observed binding of the cross-linker to a residue other than lysine (Fig. 3).

The interactions were further manually verified using raw data (MS and MS/MS) and 120 out of 160 interactions were successfully validated. The verification of results obtained from MeroX, MS Annika and XlinkX was possible, whereas in MaxLynx the verification was difficult since the MS/MS spectrum could not be visualized in the software.

Protein Interactions in the Complex

Intra-protein and inter-protein interactions determined by various XL-MS software were visualized in previously published experimental structures and AlphaFold2-predicted models of the components of the membrane protein complex, FtsH-HflK-HflC in Fig. 5. The cross-links between HflK and HflC, and their interactions between FtsH were confirmed in the recently

published cryo-EM structure of the FtsH-HflK-HflC complex[21]. In the cryo-EM structure (PDB: 7WI3), HflK and HflC form a heterodimer with a close dimer interface, and the HflK-HflC dimer forms a dodecamer, comprising 12 HflK and HflC monomers. Thus, the cross-links between HflK and HflC may represent the connections between either the HflK-HflC heterodimer or neighboring HflK and HflC monomers. Moreover, cross-links between FtsH and HflK-HflC were determined only between the periplasmic region of FtsH, and stomatin/prohibitin/flotillin/HflK/C (SPFH) domain and N-terminal of HflK-HflC proteins, aligning with interactions observed in the cryo-EM structure. Both intra- and inter-protein cross-links obtained by MaxLynx and MeroX were well distributed on the experimental structure of the membrane protein complex; however, less number of interactions were determined between FtsH and HflK-HflC with MaxLynx and MeroX, compared to those obtained by MS Annika and XlinkX.

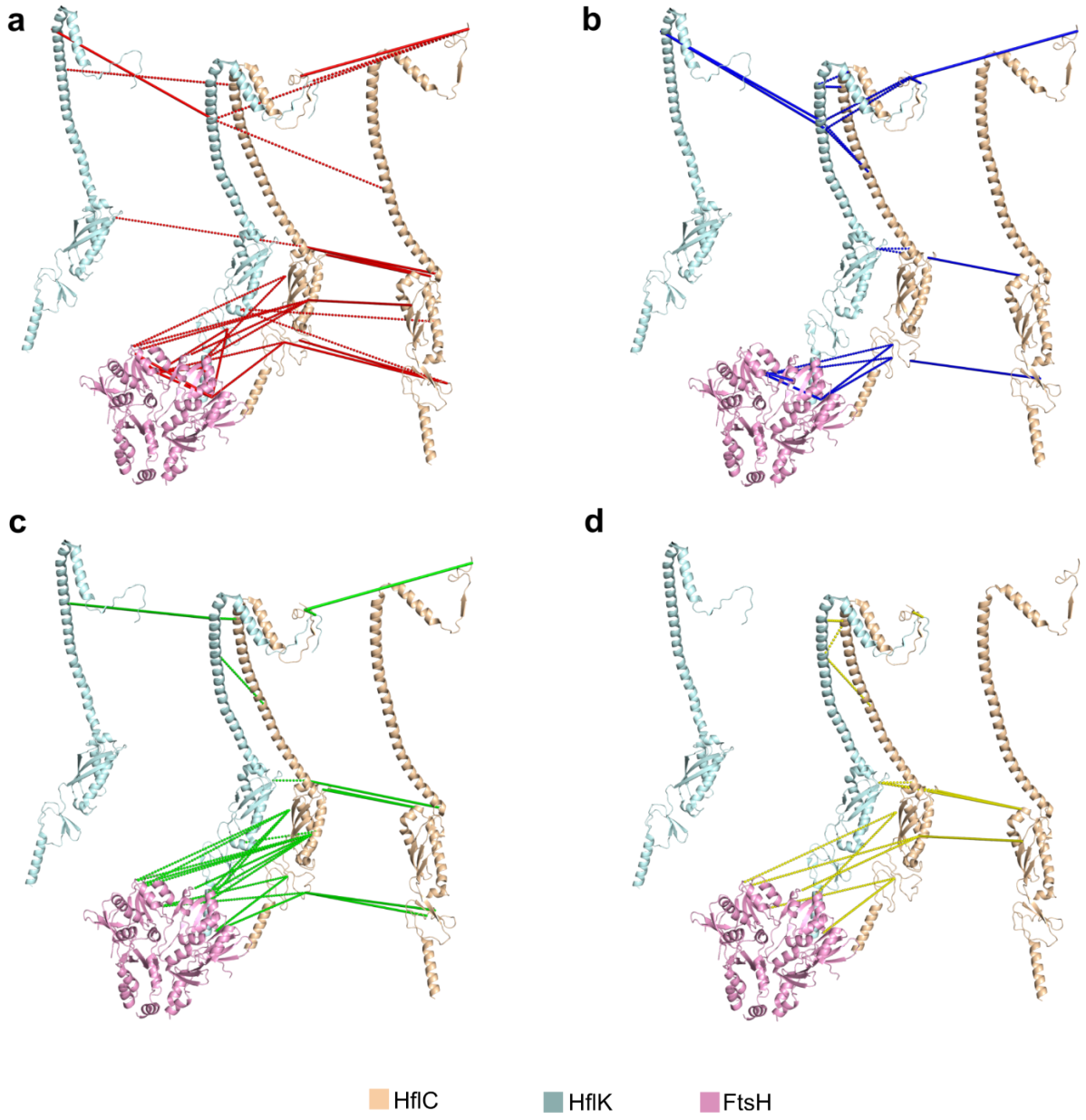


Figure 5. Top ten intra-protein (lines) and inter-protein (dashed lines) interactions of FtsH-HflK-HflC (pink - light blue - wheat) membrane protein complex (PDB ID: 7WI3) obtained from XL-MS software and colored red MaxLynx (a), blue MeroX (b), green MS Annika (c) and yellow XlinkX (d). Periplasmic and transmembrane regions of hexameric FtsH are shown only. The possible interactions are also shown between neighboring HflK and HflC

that are positioned away from the HflK-HflC heterodimer.

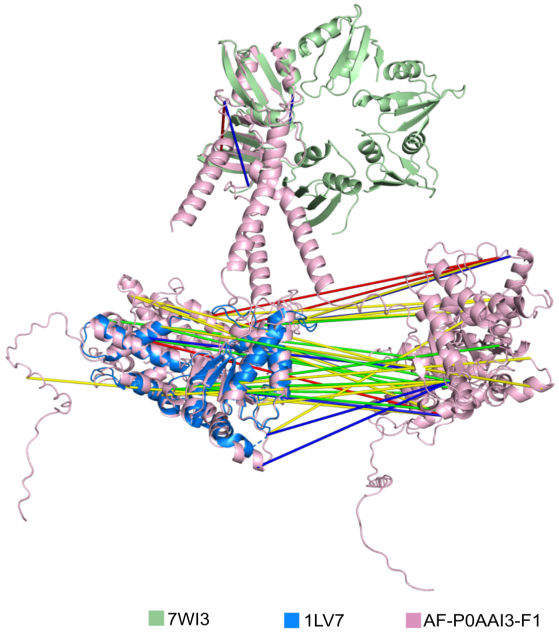


Figure 6. The detailed intra-protein connections of FtsH, represented between two monomers, obtained from XL-MS software MaxLynx (red), MeroX (blue), MS Annika (green) and XlinkX (yellow). Experimental structures are superposed with the predicted AlphaFold2 model (pink) (AF-P0AAI3-F1). The periplasmic region of hexameric FtsH (PDB ID: 7WI3) is shown in pale green, the partial cytoplasmic structure (PDB ID: 1LV7) is shown in blue. 91 residues of FtsH, in which no linkage was obtained, in N-terminus were removed in the AlphaFold2 model for clarity.

The intra-protein interactions in FtsH were confirmed on both experimental structures (PDB: 1LV7, 7WI3)[21,30] and the predicted AlphaFold2 model (AF-P0AAI3-F1) due to the lack of full-length FtsH structure deposited in the PDB although experimental structures align well with the predicted model (Fig. 6). There are cross-links determined between not only the periplasmic

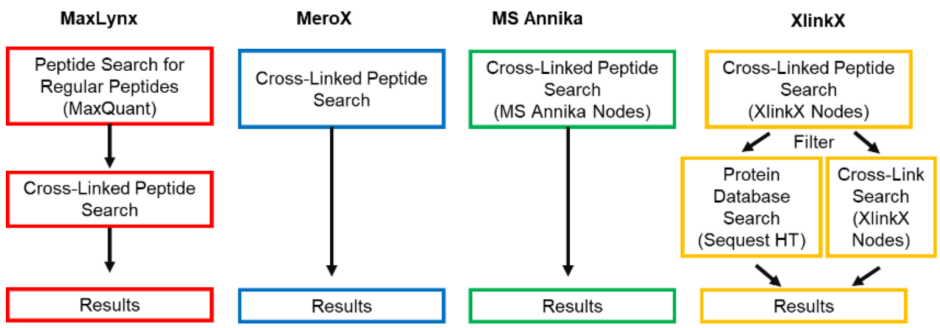
region (N-domain) but also the cytoplasmic region of FtsH. These connections may reflect both intra-protein interactions within the FtsH monomer and between FtsH monomers at the dimer interface as FtsH exists in a hexameric conformation. MaxLynx and MeroX determined cross-links within the N-domain of the FtsH whereas the number of interactions within the cytoplasmic region were low compared to other software. On the other hand, there were many connections determined by MS Annika and XlinkX within the cytoplasmic region; however, there was no cross-link determined in the N-domain with these software.

Overall, there are sufficient intra- and inter-protein interactions determined by each XL-MS software. Furthermore, there are interactions between residues which are not seen in experimental structures. For instance, HflK-Lys360 interacts with FtsH-Lys408, HflC-Lys211 interacts with FtsH-Lys60, HflC-Thr330, Ser331, and Thr333 interact with HflK-Lys294 and Lys346. All these interactions may be used to model the protein complex using experimental and predicted structures, and molecular docking programs such as HADDOCK[31]. Multimeric structures of proteins, for instance hexameric FtsH or dodecameric HflK-HflC, can also be modeled by molecular docking and cross-linking data. Likewise, the multimer structures predicted by structure prediction programs such as AlphaFold[32,33] and RoseTTAFold[34] can be confirmed by XL-MS data. The predicted structures can also be fitted into low resolution cryo-EM maps or SAXS (Small Angle X-ray Scattering) models.

In addition to structural proteomics studies, the interaction between proteins can be used for protein mapping and protein-protein networking, especially for complex samples such as a cell membrane as in our study.

XL-MS Software Performance Characteristics

Four software were compared according to performance parameters; the number of cross-links identified, and processing time for two different datasets of solubilized membrane and purified protein complex (Fig. 7). According to overall time spent during data processing by the software, MS Annika was the fastest.



	MaxLynx	MeroX	MS Annika	XlinkX
Connecting Residues	**K**	**K,S,T,Y**	**K(S,T,Y)**	**K**
Processing Time (Mins)*	106 (SM) / 146 (PP)	32 (SM) / 23 (PP)	21 (SM) / 5 (PP)	62 (SM) / 96 (PP)
Complex Inter-Protein Interactions	13	11	11	10
Complex Intra-Protein Interaction	45	49	38	45
Complex Proteome-wide Interactions	9	11	148	54

Figure 7. Comparison of four XL-MS software; MaxLynx (Red), MeroX (Blue), MS Annika (Green), XlinkX (Yellow). Overall search process, including connection residues, processing time for solubilized membrane (SM) and purified protein (PP), complex inter-protein interactions, complex intra-protein interactions and complex proteome-wide interactions are given.

Proteins were analyzed based on unique peptides. The number of interactions are based on unique peptides connected via DSBU. The DSBU cross-linker was already integrated into all

software. The software outcome file lists inter- and intra-protein interactions by indicating specific AA positions and corresponding tryptic peptide sequence for each protein pair. The matching score is also listed along with standard MS search parameters such as number of matching spectra, precursor, mass error, retention time and intensities.

MaxLynx is a software embedded into MaxQuant, a protein database search software, and thus, the cross-link search is performed only after the protein search. Although it is a useful process to identify proteins and their interactions simultaneously, the overall processing time is longer. This could be a limitation considering the size and spectra count of a high-resolution MS data searched against a large database.

The cross-linker connection side could be set to only lysine residue by default in MaxLynx, contrary to MeroX software which conducts only cross-linked peptide search at a time. Unlike MaxLynx and other software, MeroX enables selecting multiple residues such as serine, threonine, and tyrosine in addition to lysine simultaneously. Searching various cross-linking possibilities does not make MeroX's processing time longer, compared to MaxLynx and XlinkX. In addition, the graphical user interface (GUI) is user-friendly and fluent, and the size of data is not a limitation. However, in MeroX, the protein search is not an option, and only cross-linking information can be obtained.

MS Annika is a Proteome Discoverer 2.5 node, and the workflow used for MS Annika contains a spectrum selector mode. MS Annika allows conduct searches with multiple residues such as serine, threonine, and tyrosine in addition to lysine, but cannot achieve this simultaneously. Also, it contains both protein IDs and cross-link information. The processing time is significantly shorter than any other software. This might be due to the less number of

nodes in the overall search workflow, and doing only cross-linked peptide search by MS Annika nodes. Additionally, MS Annika nodes provide different search modes and these modes result in fast processing times.

XlinkX is another Proteome Discoverer 2.5 node with a workflow consisting of different nodes, such as Sequest HT. Thus, protein database search is performed with the cross-link search. This is a useful outcome for identifying proteins in samples, and their interactions, like MaxLynx. This may cause the longer processing time of XlinkX than MS Annika in a way. However, the cross-link search is performed with only lysine residue, and this is one of the limitations of XlinkX.

All the software provided relatively the same number of inter-protein connections between complex proteins. MaxLynx provided 13, MeroX provided 11, MS Annika Provided 11, and XlinkX provided 10 inter-protein connections between complex proteins. Likewise, the number of intra-protein connections are similar. MaxLynx provided 45, MeroX provided 49, MS Annika Provided 38, and XlinkX provided 45 intra-protein connections within complex proteins. However, MS Annika yielded the highest number of cross-links (148 interactions) between complex proteins and other proteins, compared to all other software (Fig.7).

All the software have different source usage during data analysis (Supplementary Figure 1). MaxLynx is the most source-friendly software in terms of CPU and RAM usage. While the RAM usage of MeroX was comparable to that of MaxLynx, its CPU usage was exceptionally high, reaching 99% utilization during analysis. On the other hand, MS Annika utilized all 5 GB of RAM assigned to it; however, had a CPU utilization of around 25%, lower than that of MeroX and XlinkX, and similar to that of MaxLynx. XlinkX also had a high CPU usage of 99% like

MeroX, and utilized more RAM than MeroX, with an average of 3.8 GB, but still less than that of MS Annika. These differences can be due to the different search and working algorithms of each software[6,8,13,16,29,35]. Additionally, it was observed that computer source usage was the same for both purified protein and solubilized membrane MS data. However, the processing time changes depending on the file size.

Protein Network

Protein Search

The identification of proteins in the purified protein complex and solubilized membrane fractions allowed us to compare differences in interaction maps generated by cross-links. We aimed to determine interactions not only within the FtsH-HflK-HflC membrane protein complex, but also with its surroundings. The network map generated by the STRING database revealed numerous connections, indicating that FtsH, HflK, and HflC have multiple biological relations with other proteins involved in different pathways. Upon comparing the two samples, two factors stood out: genomic neighborhood and functional similarity. The solubilized membrane sample contained 42% membrane protein, whereas the purified FtsH-HflK-HflC sample had an overall membrane protein content of 96% (Supplementary Figure 2).

A total of 20 connections from the findings obtained with protein search of both samples were curated in the STRING, IntAct and DIP databases as functionally associated with FtsH, HflK, and HflC. Approximately half of the interacting partners were found in the purified sample (Fig. 8a), while the solubilized membrane fraction had 17 interacting partners (Fig. 8b). STRING, IntAct and DIP databases were used for the demonstration of functional association.

RpIL, RpID, YajC, GpmA and RlpIU proteins were found to be associated only on DIP and IntAct, whereas all remaining interacting partners were found by all three databases.

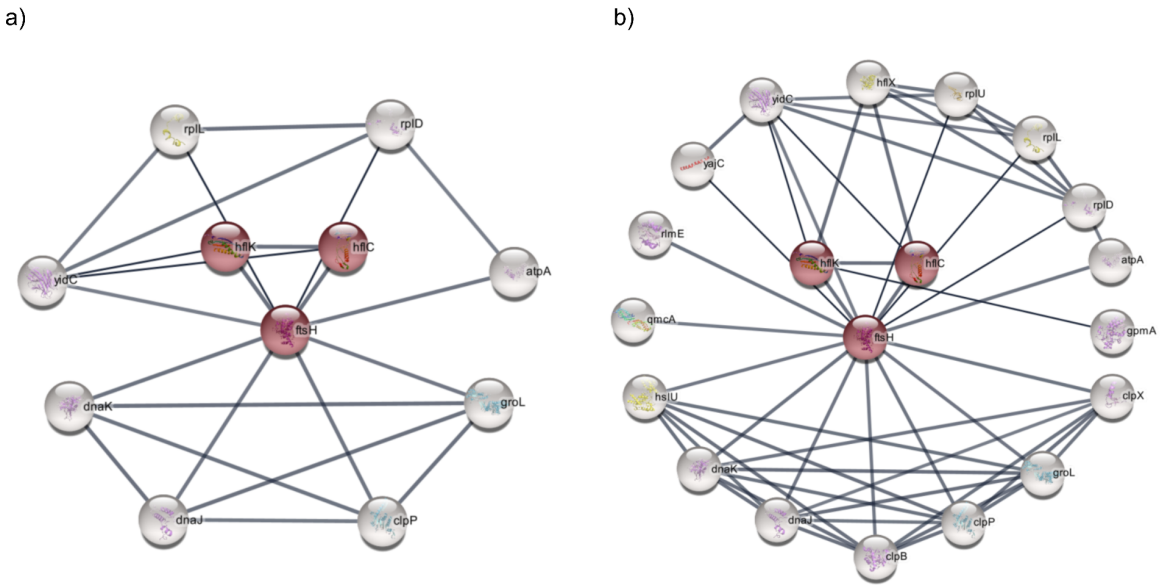
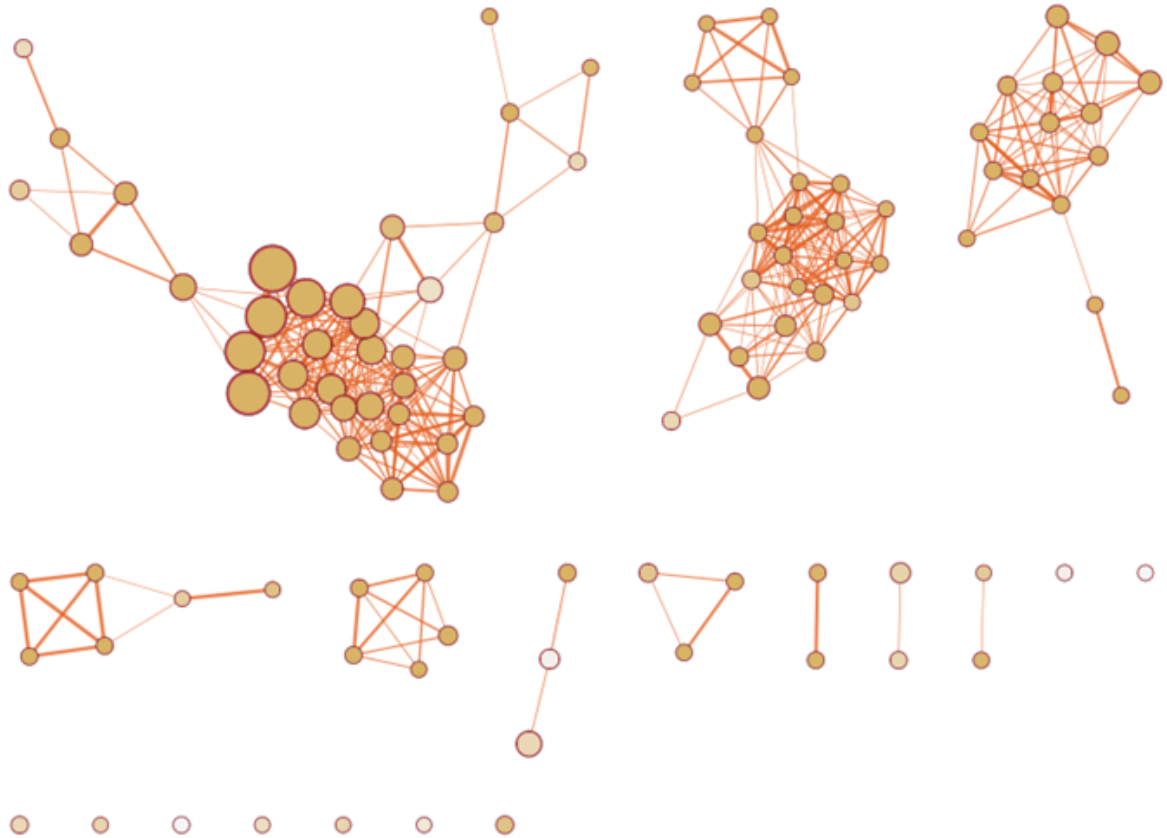


Figure 8. Interacting partners (colored in grey) of FtsH-HflK-HflC (colored in red) a) in the purified protein complex sample b) in the solubilized membrane fraction (STRING database combined with IntAct and DIP, score ≥ 0.9, FDR Stringency 1%).

Fig. 9 and Fig. 10 show biological processes in which the proteins in the solubilized membrane and purified protein samples are involved, respectively. We obtained 108 nodes in the solubilized membrane sample (Fig. 9) and 60 nodes in the purified protein sample (Fig. 10). The top 15 biological processes with the lowest p-values are shown. Proteins involved in translation, membrane protein insertion and quality control processes constitute 60% of all interacting partners identified in both samples (Supplementary Table 4). Several chaperones, including ClpB, DnaK, DnaJ, GroEL, and proteases such as ClpP, ClpX, HslU, as well as YidC, and YajC,

were identified in both samples. These chaperones are functionally associated with the FtsH-HflK-HflC membrane protein complex[36,37]**.** Specifically, YidC and YajC are proteins involved in preprotein translocation, and a part of the membrane-bound holo-translocon system comprising seven proteins including SecB (chaperone), SecA (ATPase), SecYEG (protein secretion complex), and two membrane proteins, SecD and SecF, playing roles in the release of the mature peptide into the periplasm[38–40]. The proximity of YidC to FtsH, HflK, and HflC in the bacterial cell was previously shown via *in vitro* photo cross-linking[41]. The confirmed interaction of YidC with the FtsH-HflK-HflC complex, and coexistence of YajC with it suggest that FtsH-HflK-HflC complex interacts and cooperates with the holo-translocon system in membrane protein secretion and quality control.



Top 15 Biological Processes with the Lowest p-values

Translation
Amide biosynthetic process
Peptide biosynthetic process
Peptide metabolic process
Cellular amide metabolic process
Organonitrogen compound biosynthetic process
Cellular metabolic process
Primary metabolic process
Metabolic process
Organonitrogen compound metabolic process
Biosynthetic process
Organic substance metabolic process
Organic substance biosynthetic process
Cellular protein metabolic process
Protein metabolic process

Figure 9. Top 15 biological processes of proteins found in the solubilized membrane sample with the lowest p-values.

Top 15 Biological Processes with the Lowest p-values

Translation
Peptide biosynthetic process
Amide biosynthetic process
Ribosome assembly
Cellular amide metabolic process
Ribonucleoprotein complex assembly
Protein-containing complex subunit organization
Ribosome biogenesis
Protein-containing complex assembly
Cellular protein-containing complex assembly
Cellular component assembly
Cellular protein metabolic process
Organonitrogen compound biosynthetic process
Organelle organization
Protein metabolic process

Figure 10. Top 15 biological processes of proteins found in the purified protein sample with the lowest p-values.

Cross-linking

We obtained 106 interactions of HflC, 96 interactions of FtsH, and 23 interactions of HflK, including FtsH-HflK-HflC complex, with other proteins separately from the cross-linking analysis of both samples through four software. Approximately 93% of the interactions were found in the purified complex sample and 223 interactions (out of 225) were verified through the raw data by fragment (MS/MS) and precursor ions (MS). These interactions were compared with literature through online databases, STRING (https://string-db.org), IntAct (https://www.ebi.ac.uk/intact/home), DIP (http://dip.doe-mbi.ucla.edu), and MINT (https://mint.bio.uniroma2.it/). The connections of FtsH (Supplementary Table 5), HflK (Supplementary Table 6), and HflC (Supplementary Table 7) were presented.

Although the literature suggests that GroEL, ClpB, and HflX are functionally associated with HflC, we have identified direct interactions between these proteins and HflC. Chaperonins such as ClpB and GroEL play an essential role in protein folding, particularly under stress conditions where unfolded proteins are generated. Therefore, the novel direct interactions of ClpB and GroEL with HflC may be responsible for the exposure to stress during the expression of the recombinant FtsH-HflK-HflC membrane protein complex. Another novel interaction of HflX to HflC indicates a possible cooperation between these proteins according to their regulating role for protein synthesis and quality control of membrane proteins.

Conclusion

In this study, we assessed the utility of commonly used XL-MS software tools, MaxLynx, MeroX, MS Annika, and XlinkX to elucidate the protein interactions within a membrane protein complex containing FtsH, HflK, and HflC, over-expressed in *E. coli*. Among four software, MS Annika gave the highest number of inter- and intra-protein connections. A higher number of the FtsH-HflK-HflC inter- and intra-protein interactions were obtained from the purified protein complex sample than the solubilized membrane sample. Given that the affinity-based purified complex yielded more interactions with other proteins, FtsH-HflK-HflC complex may be interacting with many associated proteins in the cell. The outcomes of the study were compared with the complex structure obtained from Protein Data Bank (PDB) to validate intra- and inter-protein interactions. Interactions between residues not seen in experimental structures were also determined. These interactions may be used to model monomeric and multimeric protein complexes using experimental and predicted structures, and molecular docking programs, assisted with protein structure/assembly prediction programs. Outcomes were not only the cross-link information within the overexpressed FtsH-HflK-HflC complex but also interactions between other proteins in close proximity with the protein complex. Thus, XL-MS can be used for protein mapping and protein-protein networking studies in various matrices. Moreover, novel physical interaction partners of the target membrane protein complex have been determined. Further studies are ongoing to elucidate the biological phenomena behind these protein-protein interactions.

MATERIALS AND METHOD

Expression and Purification of FtsH-HflK-HflC Membrane Protein Complex

The plasmid containing FtsH-HflK-HflC membrane protein complex genes was transformed into *E. coli* Lemo21 (DE3) (NEB, GB) for protein expression. Overexpression was carried out by using 0.4 mM IPTG (Sigma, USA) and 0.2% (w/v) arabinose (Carl Roth, DEU) in 2xYT media (Sigma, USA). Cells were harvested by centrifugation and were disrupted by sonication (Sonics VCX 130, USA). Afterwards, the membrane was separated by ultracentrifugation, and solubilized using n-Dodecyl-β-D-maltoside (DDM) (Carbosynth, USA) and purified samples of FtsH-HflK-HflC were collected using affinity chromatography.

Chemical Cross-Linking Reactions with DSBU and Tryptic Digestion

Cross-linking reactions with disuccinimidyl dibutyric urea (DSBU) (Thermo Fisher Sigma, USA) were carried out using purified FtsH-HflK-HflC membrane protein complex and solubilized membrane with overexpressed protein complex according to manufacturer's instructions[11,42–44]. A 5 µM of purified FtsH-HflK-HflC sample was treated with 2.5 mM DSBU for 60 min at room temperature. The reaction was stopped using 20-fold Tris-HCl (1 M, pH 8.0) and samples were concentrated using pre-equilibrated 100 kDa MWCO concentrators up to ~20 µL.

Cross-linking reactions were performed at 100 mg/ml solubilized membranes of overexpressed FtsH-HflK-HflC with a final concentration of 10 mM DSBU in 12.5 µL. Reactions were incubated for 60 min at room temperature and terminated using 20-fold Tris-HCl (1 M, pH 8.0).

Samples were extracted in 100 µL of 50 mM Ammonium bicarbonate and incubated with 11 µL of 100 mM Dithiothreitol (DTT) for 50 min at 60 °C. Free cysteines were alkylated with 22 µL of 100 mM Iodoacetamide (IAA) for 30 min at room temperature in the dark. Then 50 µL 0.1 µg/µL trypsin was added and samples were incubated at 37 °C for 18 h. All samples were dried in miVac at room temperature. Dry peptide samples were dissolved in 0.1% Formic acid and peptide concentrations were determined with Pierce™ Quantitative Peptide Assay (Thermo Fisher Scientific, USA) before the MS analysis.

Mass Spectrometry

The 1 µg digested samples were analyzed by the UltiMate™ WPS-3000RS nanoLC system coupled with Orbitrap Fusion Lumos (Thermo Fisher Scientific, USA). The peptides were separated on Acclaim™ PepMap™ 100 C18 HPLC Columns (3 µm, 0.075 mm × 500 mm, Thermo Fisher Scientific, USA). The mobile phase A containing 0.1% aqueous formic acid and mobile phase B comprising 0.1% formic acid in 80% acetonitrile were set to a gradient; 0–5 min, 4–4% (B); 5–130 min, 4–35% (B); 130–150 min, 35–50% (B); 150–153 min, 50–100% (B); 153–168 min, 100–100% (B); 168–170 min, 100–4% (B); and 170–180 min, 4–4% (B). The MS and MSMS spectra were collected with a mass range of m/z 300-1800 in positive ionization ion mode. The high-energy C-trap dissociation (HCD) fragmentation was performed with nitrogen gas, and collision energies of 25%, 30%, and 35%. The precursor and the product ions were detected at 120K resolution and 15K resolution respectively.

Data Analysis

Raw MS data corresponding to the purified protein complex and the solubilized membrane were analyzed in MaxLynx, MeroX 2.0, MS Annika and XlinkX to identify cross-linked peptides as described below.

MaxLynx software is an add-on to the MaxQuant. MS Annika and XlinkX are two separate nodes for the Proteome Discoverer 2.5 (Thermo Fisher Scientific). Corresponding workflows were obtained from the website. MeroX is a JAVA based software that can be operated independently.

E.coli proteome (UniProt ID: UP000000625) was used for the proteome-wide search. While the raw MS data could be directly used in MaxLynx, MS Annika and XlinkX, it needs to be converted to a mzml format using an external software such as ProteoWizard[45] for MeroX. The same settings were used in all four software to ensure fair comparison (Supplementary Table 1). C-terminals of lysine (K) and arginine (R) residues were set as specific cleavage sites, and maximum missed cleavages were set to 3. Precursor and fragmentation mass tolerance were limited to 10 ppm and 20 ppm, respectively. The carbamidomethylation at cysteine was assigned as the fixed modification, and the oxidation at methionine was selected as the variable modification. Signal to noise ratio was set to 2.0. The minimum precursor mass limit was 1000 Da, whereas the maximum precursor mass limit was 20,000 Da. In addition, the minimum and maximum peptide lengths were set as 5 and 40 AAs. The cross-link modification sites were specified as Lysine for all software. In addition, Serine (S) , Threonine (T), and Tyrosine (Y) were also added in MeroX search as the software enables variable amino acid search[29]. The FDR (False Discovery Rate) was limited to 1%. The upper limit to Cα-Cα distance was used as 30Å as suggested by previous studies[12,46,47].

The specifications of the computer used were Intel i5 9300-H Processor with 4 Cores (8 Logical Cores), and 8GB DDR4 2666MHz RAM. The maximum RAM usage was limited to 5GB of RAM (3GB RAM was assigned for the operating system). The capacity of the processor and the amount of free RAM Memory amount were two limiting factors.

For the protein network database search and visualization, STRING (v.11.5)[24], IntAct[25,26], DIP[27], MINT[28] and Cytoscape (v.3.9.1)[48] were used. The search on STRING was done by applying advanced settings of required score as highest score (0.900) and FDR Stringency as 1%. The EnrichmentMap, which is a Cytoscape App, was used for the visualization of the results of gene-set enrichment. Generic enrichment results were used for operation and p-values were operated for visualization as radial heat-maps. Gene sets are represented by nodes and the sizes show the number of genes.

Supplementary Material Description

Supplementary Figure 1. Cross-linking software performance characteristics

Supplementary Figure 2. Relative abundances of membrane and other proteins

Supplementary Table 1. Cross-link search parameters

Supplementary Table 2. Total interaction numbers and connections represented

Supplementary Table 3. The most abundant 10 inter- and intra-protein interactions

Supplementary Table 4. Proteins identified in the samples

Supplementary Table 5. The connections of FtsH with other proteins

Supplementary Table 6. The connections of HflK with other proteins

Supplementary Table 7. The connections of HflC with other proteins

ACKNOWLEDGMENT

H.A., H.K.I., G.G., and B.V.K. are supported by TUBITAK (Turkish Scientific and Technological Research Council) BIDEB 2232 Program (Project No: 118C225). We thank Prof. Carlito B. Lebrilla and Lebrilla Group from University of California, Davis, USA for supporting us in cross-linking mass spectrometry analyses. Also we thank Bergers Lab - Prof. Christiane Schaffitzel-Berger and Prof. Imre Berger from University of Bristol, UK for providing the plasmid expressing the membrane protein complex.

REFERENCES

1. Wheat A, Yu C, Wang X, Burke AM, Chemmama IE, Kaake RM, Baker P, Rychnovsky SD, Yang J, Huang L (2021) Protein interaction landscapes revealed by advanced in vivo cross-linking-mass spectrometry. Proc. Natl. Acad. Sci. U. S. A. 118:e2023360118.
2. Smith BJ, Martins-de-Souza D (2021) Biological Applications for LC-MS-Based Proteomics. Adv. Exp. Med. Biol. 1336:17–29.
3. Chen C, Hou J, Tanner JJ, Cheng J (2020) Bioinformatics Methods for Mass Spectrometry-Based Proteomics Data Analysis. Int. J. Mol. Sci. 21:2873.
4. Tang X, Wippel HH, Chavez JD, Bruce JE (2021) Crosslinking mass spectrometry: A link between structural biology and systems biology. Protein Sci. 30:773–784.
5. Graziadei A, Rappsilber J (2022) Leveraging crosslinking mass spectrometry in structural and cell biology. Structure 30:37–54.
6. Piersimoni L, Kastritis PL, Arlt C, Sinz A (2022) Cross-Linking Mass Spectrometry for Investigating Protein Conformations and Protein–Protein Interactions─A Method for All Seasons. Chem. Rev. 122:7500–7531.
7. Hoopmann MR, Zelter A, Johnson RS, Riffle M, MacCoss MJ, Davis TN, Moritz RL (2015) Kojak: Efficient analysis of chemically cross-linked protein complexes. J. Proteome Res. 14:2190–2198.
8. Yılmaz Ş, Busch F, Nagaraj N, Cox J (2021) Accurate and automated high-coverage identification of chemically cross-linked peptides with MaxLynx. :2021.08.26.457759. Available from: https://www.biorxiv.org/content/10.1101/2021.08.26.457759v1
9. Götze M, Pettelkau J, Fritzsche R, Ihling CH, Schäfer M, Sinz A (2015) Automated assignment of MS/MS cleavable cross-links in protein 3D-structure analysis. J. Am. Soc. Mass Spectrom. 26:83–97.
10. Götze M, Pettelkau J, Schaks S, Bosse K, Ihling CH, Krauth F, Fritzsche R, Kühn U, Sinz A (2012) StavroX—A Software for Analyzing Crosslinked Products in Protein Interaction Studies. J. Am. Soc. Mass Spectrom. 23:76–87.
11. Müller MQ, Dreiocker F, Ihling CH, Schäfer M, Sinz A (2010) Cleavable cross-linker for

protein structure analysis: reliable identification of cross-linking products by tandem MS. Anal. Chem. 82:6958–6968.
12. Iacobucci C, Götze M, Ihling CH, Piotrowski C, Arlt C, Schäfer M, Hage C, Schmidt R, Sinz A (2018) A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein-protein interactions. Nat. Protoc. 13:2864–2889.
13. Pirklbauer GJ, Stieger CE, Matzinger M, Winkler S, Mechtler K, Dorfer V (2021) MS Annika: A New Cross-Linking Search Engine. J. Proteome Res. 20:2560–2569.
14. Netz E, Dijkstra TMH, Sachsenberg T, Zimmermann L, Walzer M, Monecke T, Ficner R, Dybkov O, Urlaub H, Kohlbacher O (2020) OpenPepXL: An Open-Source Tool for Sensitive Identification of Cross-Linked Peptides in XL-MS. Mol. Cell. Proteomics MCP 19:2157–2168.
15. Chen Z-L, Meng J-M, Cao Y, Yin J-L, Fang R-Q, Fan S-B, Liu C, Zeng W-F, Ding Y-H, Tan D, et al. (2019) A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. Nat. Commun. 10:3404.
16. Liu F, Lössl P, Scheltema R, Viner R, Heck AJR (2017) Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. Nat. Commun. 8:15473.
17. Mendes ML, Fischer L, Chen ZA, Barbon M, O'Reilly FJ, Giese SH, Bohlke-Schneider M, Belsom A, Dau T, Combe CW, et al. (2019) An integrated workflow for crosslinking mass spectrometry. Mol. Syst. Biol. 15:e8994.
18. Perez C, Maier T eds Expression, Purification, and Structural Biology of Membrane Proteins. New York, NY: Springer US; 2020. Available from: http://link.springer.com/10.1007/978-1-0716-0373-4
19. Akiyama Y (2002) Proton-motive force stimulates the proteolytic activity of FtsH, a membrane-bound ATP- dependent protease in Escherichia coli. Proc. Natl. Acad. Sci. 99:8066–8071.
20. Li W, Rao DK, Kaur P (2013) Dual role of the metalloprotease FtsH in biogenesis of the DrrAB drug transporter. J. Biol. Chem. 288:11854–11864.
21. Qiao Z, Yokoyama T, Yan X-F, Beh IT, Shi J, Basak S, Akiyama Y, Gao Y-G (2022) Cryo-EM structure of the entire FtsH-HflKC AAA protease complex. Cell Rep. 39:110890.
22. Ma C, Wang C, Luo D, Yan L, Yang W, Li N, Gao N (2022) Structural insights into the membrane microdomain organization by SPFH family proteins. Cell Res. 32:176–189.
23. Kihara A, Akiyama Y, Ito K (1996) A protease complex in the Escherichia coli plasma membrane: HflKC (HflA) forms a complex with FtsH (HflB), regulating its proteolytic activity against SecY. EMBO J. 15:6122–6131.
24. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, et al. (2021) The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. 49:D605–D612.
25. del Toro N, Shrivastava A, Ragueneau E, Meldal B, Combe C, Barrera E, Perfetto L, How K, Ratan P, Shirodkar G, et al. (2022) The IntAct database: efficient access to fine-grained molecular interaction data. Nucleic Acids Res. 50:D648–D653.
26. Hermjakob H, Montecchi-Palazzi L, Lewington C, Mudali S, Kerrien S, Orchard S, Vingron M, Roechert B, Roepstorff P, Valencia A, et al. (2004) IntAct: an open source molecular interaction database. Nucleic Acids Res. 32:D452–D455.
27. Xenarios I, Rice DW, Salwinski L, Baron MK, Marcotte EM, Eisenberg D (2000) DIP: the Database of Interacting Proteins. Nucleic Acids Res. 28:289–291.

28. Chatr-aryamontri A, Ceol A, Palazzi LM, Nardelli G, Schneider MV, Castagnoli L, Cesareni G (2007) MINT: the Molecular INTeraction database. Nucleic Acids Res. 35:D572–D574.
29. Iacobucci C, Piotrowski C, Aebersold R, Amaral BC, Andrews P, Bernfur K, Borchers C, Brodie NI, Bruce JE, Cao Y, et al. (2019) First Community-Wide, Comparative Cross-Linking Mass Spectrometry Study. Anal. Chem. 91:6953–6961.
30. Krzywda S, Brzozowski AM, Verma C, Karata K, Ogura T, Wilkinson AJ (2002) The crystal structure of the AAA domain of the ATP-dependent protease FtsH of Escherichia coli at 1.5 A resolution. Struct. Lond. Engl. 1993 10:1073–1083.
31. van Zundert GCP, Rodrigues JPGLM, Trellet M, Schmitz C, Kastritis PL, Karaca E, Melquiond ASJ, van Dijk M, de Vries SJ, Bonvin AMJJ (2016) The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J. Mol. Biol. 428:720–725.
32. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589.
33. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Žídek A, Bridgland A, Cowie A, Meyer C, Laydon A, et al. (2021) Highly accurate protein structure prediction for the human proteome. Nature 596:590–596.
34. Yang J, Anishchenko I, Park H, Peng Z, Ovchinnikov S, Baker D (2020) Improved protein structure prediction using predicted interresidue orientations. Proc. Natl. Acad. Sci. U. S. A. 117:1496–1503.
35. Yu C, Huang L (2018) Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. Anal. Chem. 90:144–165.
36. Dill KA, Ozkan SB, Shell MS, Weikl TR (2008) The protein folding problem. Annu. Rev. Biophys. 37:289–316.
37. Schramm FD, Schroeder K, Jonas K (2020) Protein aggregation in bacteria. FEMS Microbiol. Rev. 44:54–72.
38. Botte M, Zaccai NR, Nijeholt JL à, Martin R, Knoops K, Papai G, Zou J, Deniaud A, Karuppasamy M, Jiang Q, et al. (2016) A central cavity within the holo-translocon suggests a mechanism for membrane protein insertion. Sci. Rep. 6:38399.
39. Kabasakal BV, Jiang Q, Schaffitzel C Production of Multi-subunit Membrane Protein Complexes. In: Poterszman A, editor. Multiprotein Complexes: Methods and Protocols. Methods in Molecular Biology. New York, NY: Springer US; 2021. pp. 3–16. Available from: https://doi.org/10.1007/978-1-0716-1126-5_1
40. Schulze RJ, Komar J, Botte M, Allen WJ, Whitehouse S, Gold VAM, Lycklama a Nijeholt JA, Huard K, Berger I, Schaffitzel C, et al. (2014) Membrane protein insertion and proton-motive-force-dependent secretion through the bacterial holo-translocon SecYEG–SecDF–YajC–YidC. Proc. Natl. Acad. Sci. 111:4844–4849.
41. van Bloois E, Dekker HL, Fröderberg L, Houben ENG, Urbanus ML, de Koster CG, de Gier J-W, Luirink J (2008) Detection of cross-links between FtsH, YidC, HflK/C suggests a linked role for these proteins in quality control upon insertion of bacterial inner membrane proteins. FEBS Lett. 582:1419–1424.
42. Arlt C, Götze M, Ihling CH, Hage C, Schäfer M, Sinz A (2016) Integrated Workflow for Structural Proteomics Studies Based on Cross-Linking/Mass Spectrometry with an MS/MS Cleavable Cross-Linker. Anal. Chem. 88:7930–7937.
43. Kao A, Chiu C, Vellucci D, Yang Y, Patel VR, Guan S, Randall A, Baldi P, Rychnovsky SD, Huang L (2011) Development of a novel cross-linking strategy for fast and accurate

identification of cross-linked peptides of protein complexes. Mol. Cell. Proteomics MCP 10:M110.002212.
44. Liu F, Rijkers DTS, Post H, Heck AJR (2015) Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. Nat. Methods 12:1179–1184.
45. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, Gatto L, Fischer B, Pratt B, Egertson J, et al. (2012) A cross-platform toolkit for mass spectrometry and proteomics. Nat. Biotechnol. 30:918–920.
46. Hage C, Iacobucci C, Götze M, Sinz A (2020) A biuret-derived, MS-cleavable cross-linking reagent for protein structural analysis: A proof-of-principle study. J. Mass Spectrom. JMS 55:e4449.
47. Merkley ED, Rysavy S, Kahraman A, Hafen RP, Daggett V, Adkins JN (2014) Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine-lysine distances. Protein Sci. Publ. Protein Soc. 23:747–759.
48. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13:2498–2504.