



OPEN

## A protein–protein interaction analysis tool for targeted cross-linking mass spectrometry

Jongham Park<sup>1,5</sup>, Ahrum Son<sup>2,5</sup> & Hyunsoo Kim<sup>1,3,4</sup>✉

Protein networking is critical to understanding the biological functions of proteins and the underlying mechanisms of disease. However, identifying physical protein–protein interactions (PPIs) can be challenging. To gain insights into target proteins that interact with a particular disease, we need to profile all the proteins involved in the disease beforehand. Although the cross-linking mass spectrometry (XL-MS) method is a representative approach to identify physical interactions between proteins, calculating theoretical mass values for application to targeted mass spectrometry can be difficult. To address this challenge, our research team developed PPIAT, a web application that integrates information on reviewed human proteins, protein–protein interactions, cross-linkers, enzymes, and modifications. PPIAT leverages publicly accessible databases such as STRING to identify interactomes associated with target proteins. Moreover, it autonomously computes the theoretical mass value, accounting for all potential cross-linking scenarios pertinent to the application of XL-MS in SRM analysis. The outputs generated by PPIAT can be concisely represented in terms of protein interaction probabilities, complemented by findings from alternative analytical tools like Prego. These comprehensive summaries enable researchers to customize the results according to specific experimental conditions. All functions of PPIAT are available for free on the web application, making it a valuable tool for researchers studying protein–protein interactions.

Protein–protein interactions (PPIs) play a crucial role in the formation of protein networks and complex structures, which are essential for understanding the biological functions of proteins<sup>1–4</sup>. Abnormal PPIs involving endogenous proteins, proteins from pathogens, or both can lead to various human diseases, including cancer<sup>4–6</sup>. To understand a target protein's interaction with a specific disease, it is necessary to profile all proteins involved in the disease beforehand. However, the profiling results may not match with an experimental target protein<sup>7–9</sup>. Therefore, it is essential to identify physical protein–protein interactions using various methods<sup>10,11</sup>. Several web-based databases, such as STRING<sup>12</sup>, MINT<sup>13</sup>, BioGRID<sup>14</sup>, and IntAct<sup>15</sup> provide theoretical protein–protein interaction search and contribute significantly to protein interactome profiling analysis.

Various methodologies can identify these interaction sites, including covalent labeling<sup>16</sup> footprinting<sup>17</sup>, hydrogen–deuterium exchange mass spectrometry (HDX-MS)<sup>18</sup>, cross-linking mass spectrometry (XL-MS or CX-MS or CL-MS)<sup>11,19–26</sup>, ion-mobility MS<sup>27</sup>, and native MS<sup>10,28</sup>. Among these methods, XL-MS has become a powerful approach for mapping protein–protein interactions over several decades<sup>23</sup>. However, its effectiveness has long been impeded by three primary obstacles: (1) complex tandem mass spectrometry (MS/MS or MS2) fragmentation of cross-linked peptides; (2) low abundance of cross-linked peptides in complex peptide mixtures; (3) heterogeneity of cross-linked products<sup>10</sup>. Furthermore, calculating all possible cases and their mass values by considering the site, charge, and modification where interactions between proteins occur can be challenging. The first hurdle makes accurate identification of cross-linked peptides and unambiguous assignment of cross-linked sites challenging, while the second and third hurdles hinder effective MS detection of cross-linked peptides<sup>4</sup>. The fourth hurdle becomes significantly hinders the efficiency of research. Selected reaction monitoring (SRM) method is used to identify PPIs perform based on MS/MS values obtained through profiling. However, MS/MS fragmentation of cross-linked peptides is typically convoluted and unpredictable. Therefore, to identify the site of the predicted theoretical actual interaction occurs and calculate their mass values, we need to compare experimental MS/MS spectra against a computed library of theoretical spectra. This process of calculating mass

<sup>1</sup>Department of Bio-AI Convergence, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Republic of Korea. <sup>2</sup>Department of Molecular Medicine, Scripps Research, La Jolla, CA 92037, USA. <sup>3</sup>Department of Convergent Bioscience and Informatics, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Republic of Korea. <sup>4</sup>SCICS, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Republic of Korea. <sup>5</sup>These authors contributed equally: Jongham Park and Ahrum Son. ✉email: kimlab@cnu.ac.kr

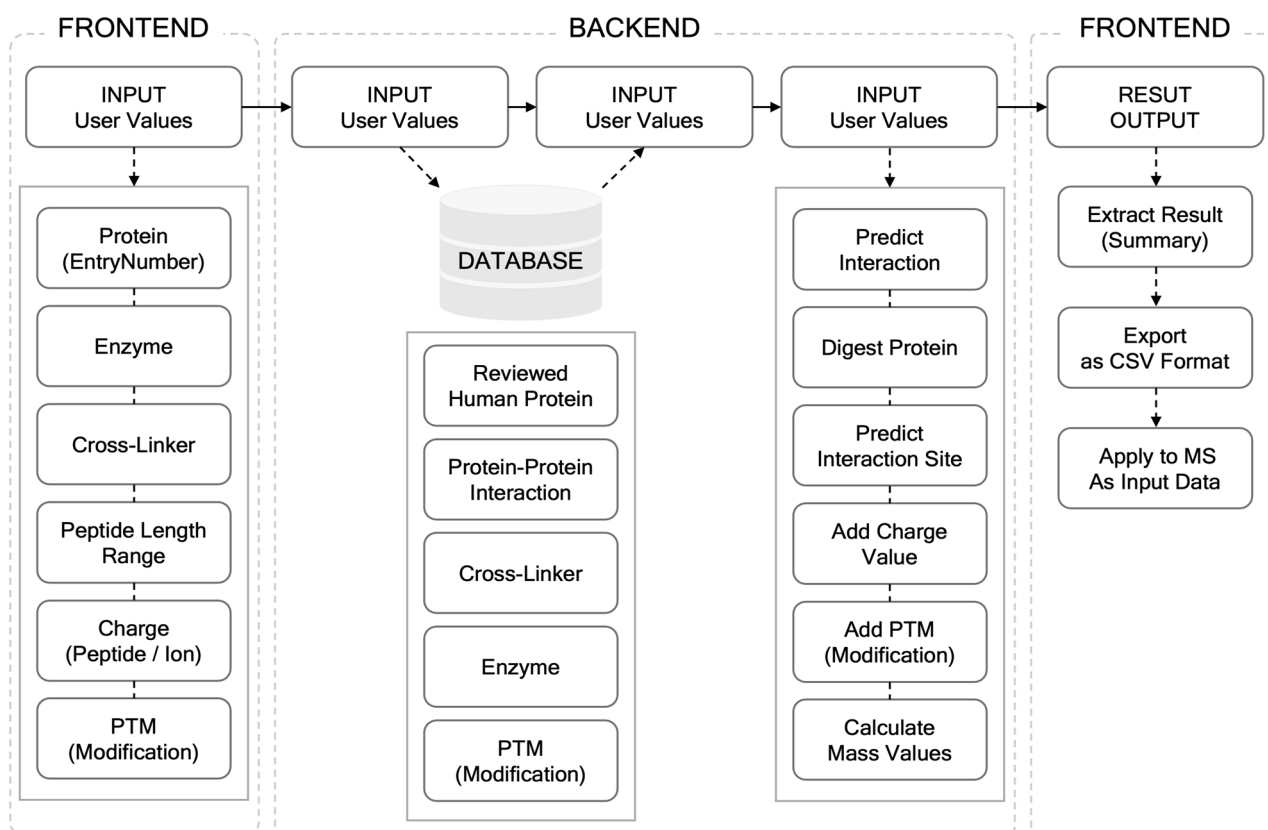
values must not only consider all cases of theoretical interaction but also the properties of cross-linker, charge of peptide/fragmented peptide ion, and modification. Finally, several tools can help analyze XL-MS, such as SRMcollider<sup>29</sup>, Prego<sup>30</sup>, and X-Link Transition Calculator<sup>31</sup> in Skyline<sup>32</sup>. SRMcollider is software that predicts interference probability between target transitions<sup>29</sup>. Prego is software that predicts high-responding peptides for SRM experiments<sup>30</sup>. The X-Link Transition Calculator is software that calculates cross-linked mass values<sup>31</sup>. However, these tools are unable to search theoretical protein interactors for the target protein, and it must be known which sites are actual interact and modified. In this respect, the development of analysis tools for XL-MS is still needed.

We proposed PPIAT, a targeted mass spectrometry-based protein–protein interaction analysis tool. PPIAT is a web-based analysis tool that can search interaction information about human proteins and calculate mass values given the properties of cross-linkers, enzymes, charges, and modifications. We expect PPIAT to improve experimental efficiency and become a significant contributor to theoretical information for XL-MS analysis.

## Materials and methods

### Overview of PPIAT

PPIAT was developed using XAMPP, CI3 (Codeignator3), and CSS (Cascading Style Sheet). XAMPP is a software package that includes the Apache web server, PHP (Hypertext Preprocessor), and MySQL, which means that PPIAT runs on the Apache server and relies on the MySQL database. The software offers three main functions. First, it allows users to search for human proteins by utilizing information from UniProt's reviewed human proteins database<sup>24,33</sup>. Second, it provides users with the ability to search for information on commercially available cross-linkers, including their name, binding site, cleavability<sup>25</sup>, spacer arm length, and mass values, which are differentiated between non-cleavage and cleavage mass values. Third, PPIAT can predict protein–protein interactions for the user's target protein and calculate their mass values based on input parameters such as target protein, cross-linker, enzyme, peptide length range, peptide charge, ion charge, and modifications specified on the search page. Finally, the results obtained from PPIAT can be summarized and extracted based on the probability score of protein interaction in the derived results and other analysis tools such as Prego<sup>30</sup>. With these functions, PPIAT serves as a valuable tool for targeted mass spectrometry-based protein interactor analysis in XL-MS. (see Fig. 1) To aid users in effectively utilizing the tool, a comprehensive tutorial is accessible at: <http://ppiat.cnu.ac.kr/userguide>.



**Figure 1.** Workflow of PPIAT. The figure illustrates the flow of PPIAT, with input Information for searching protein–protein interaction (PPIs) at the front-end. Data is queried from the database based on the input information, and PPIs and mass values are calculated according to the input conditions. All calculated data is presented at the front-end, and the results can be exported in CSV format. The exported data can be used as input for MS/MS analysis.

## Database construction

PPIAT is built on a comprehensive database that integrates data from various sources, including reviewed human proteins, protein–protein interactions, cross-linkers, enzymes, and modifications. The database contains 20,386 reviewed human protein entries (43,277 including isoforms), approximately 20 billion interaction data, and 83 cross-linker entries. These data were collected from reputable sources such as UniProt<sup>33</sup>, STRING<sup>12</sup>, and publicly published white papers. PPIAT employs the STRING database to identify interacting proteomes associated with target proteins. The databases for PPIs and reviewed human proteins, provided respectively by STRING and UniProt, are initially distinct in format. PPIAT enhances its predictive capability by integrating these two databases into a cohesive format for PPI search. The database forms the backbone of the PPIAT, enabling the efficient and accurate prediction of protein–protein interactions and the calculation of their respective mass values.

## User guide about PPIAT

PPIAT offers a user-friendly interface, allowing for the input of various parameters, such as (1) the entry number of the desired target protein (e.g., P04439, Q6ZWK4), (2) the enzyme used for protein digestion, (3) the cross-linker employed in XL-MS<sup>11,19–26</sup> for identifying protein–protein interactions, (4) the preferred peptide length range for MS/MS analysis, (5) the peptide and fragmented peptide ion charges, and (6) any anticipated modifications in the protein–protein interactions (refer to Fig. 2). The information of reviewed human proteins available at UniProt<sup>33</sup> comprises several columns, including ID, name, organism, entry number, entry name, and sequence ID. The entry number and entry name information can be utilized to search for protein–protein interactions in resources such as STRING<sup>12</sup>.

Proteins are differentiated between canonical and isoform. Isoform refers to a member of a set of highly similar proteins that originate from a single gene or gene family and are the result of genetic differences. Although isoforms have the same or similar function, they do not have an identical sequence and therefore have different mass values<sup>34</sup>. To address this issue, PPIAT divides the same entry name and protein name between canonical and isoform using a dash (“-”) in the entry number. However, while the values of the string column at UniProt<sup>33</sup> and values of protein A and protein B at STRING<sup>12</sup> are linked to each other (e.g., format of 9606.ENSPP00000379873), they are not completely matched, and the canonical and isoform use the same value. It should be noted that PPIAT does not differentiate between canonical forms and isoforms of a target protein based solely on the protein A and B values from STRING<sup>12</sup>. For distinguishing between these forms, PPIAT utilizes the entry number when searching for a target protein (e.g., P02649, P06727).

Users can select either trypsin or chymotrypsin as the enzyme for protein digestion, depending on the properties of the digestion site and exceptions. In addition, users can choose from a list of 83 commercialized cross-linkers to identify interactions between the target protein and other proteins based on the properties of the cross-linker binding site and cleavability. (see Data S1) Finally, by inputting the peptide length range, the number of interaction rankings, the charge of the peptide and fragmented peptide ion, and any modifications predicted to occur in the protein–protein interactions. The peptide length range refers to the maximum and minimum peptide lengths that the user wants to identify at the interaction level, and the number of interaction rankings refers to the number of interactable proteins that the user wants to identify. The input values for peptide charge

**Figure 2.** PPIAT search page. This search page of PPIAT requires users to input specific information for searching protein–protein interactions (PPIs) and their mass values. Required information includes the entry number of the target protein, the enzyme used for protein digestion, the cross-linker utilized for crosslinking, the range of peptide length, the ranking of protein–protein interactions, the charges of the peptide and ion, and any predicted protein modifications.

and fragmented peptide ion charge considered for MS/MS analysis allow for overlapping selection. There are two types of modifications, static and variable, which are used to calculate mass values, and they also allow for overlapping selection. PPIAT is designed to identify interactions with target proteins and automatically compute the totality of actual interaction cases, along with their corresponding masses resulting from cross-linking at both peptide and ionic levels. This is done while taking into consideration the charge of the peptide and fragmented peptide ion, properties of the cross-linker and enzyme, as well as any modifications.

## Result

The XL-MS workflow can be divided into several steps. Firstly, a target protein complex is cross-linked in solution and digested with trypsin into peptides. Secondly, the peptides are analyzed by liquid chromatography coupled with mass spectrometry (LC-MS/MS) to obtain precursor masses and fragment masses for cross-linked peptides. Thirdly, the fragmentation spectra of all peptides are subjected to database searching to identify cross-linked peptides<sup>20</sup>. This procedure involves searching for the binding site of the cross-linker in each sequence of the precursor ion digested by the enzyme and the fragmented ion generated during SRM analysis. Considering the varying mass of the cross-linked interactor depending on the cross-linker's characteristics, PPIAT consults a database encompassing information about the cross-linker's properties, such as its cleavability. If there are amino acids that can bind to the cross-linker, PPIAT calculates the mass, accounting for the cross-linker's properties and all potential cross-linking types. Finally, the output of PPIAT can be summarized and extracted based on the probability score of protein interaction in the derived results and other analysis tools such as Prego<sup>30</sup>. This output can be used as input data for targeted mass spectrometry analysis. The combined score is divided into four groups based on the confidence level: highest confidence (> 0.9), high confidence (> 0.7), medium confidence (> 0.4), and low confidence (> 0.15). The analysis program Prego provides a list of predicted high-responding peptides for SRM experiments<sup>30</sup>. PPIAT stands out for its advanced mass calculation capabilities, vital for integrating XL-MS studies with targeted mass spectrometry techniques such as SRM (Selected Reaction Monitoring) and PRM (Parallel Reaction Monitoring). It enables the generation of comprehensive mass spectral libraries from results obtained through DDA (Data-Dependent Acquisition) or DIA (Data-Independent Acquisition) analyses, effectively targeting specific proteins. It also facilitates conducting experiments with crosslinkers on peptide sequences predicted by Prego<sup>30</sup>.

The PPIAT software outputs results that can be utilized as input for SRM analysis. This software searches for PPIs and calculates their mass values. It generates separate tables for identified PPIs and predicts interaction information, indicating which proteins are interacting with the target protein based on a combined score. Additionally, another table provides printed information on the theoretical actual interaction, including combined score, the protein A' peptide, cross linker, protein B' peptide, precursor charge, precursor  $m/z$ , protein A' ion, protein A' ion type, protein B' ion, protein B' ion type, product charge, and product  $m/z$ .

After searching for the target protein using PPIAT, the software generates a list of theoretical PPIs for the target protein, along with the number of actual interactions that occur in the interactome and their corresponding mass values. The theoretical PPIs list predicts which proteins interact with the target protein and assigns scores based on the STRING database<sup>12</sup>. In each column of the tables, "Protein A" refers to the target protein that was inputted, while "Protein B" refers to the proteins that interact with the target protein in both peptide and ion levels. The precursor  $m/z$  and product  $m/z$  columns are divided into the mass value of protein A linked with one side of the cross linker, the cross-linker mass after cleavage, and protein B linked with the other side of cross linker. If a modification is expected to occur statically, the fluctuating mass values are added. However, if the modification is expected to occur variably, the results are printed as two lines—one with the change in mass values due to the modification and one without. Figure 3 provides an example of the output format.

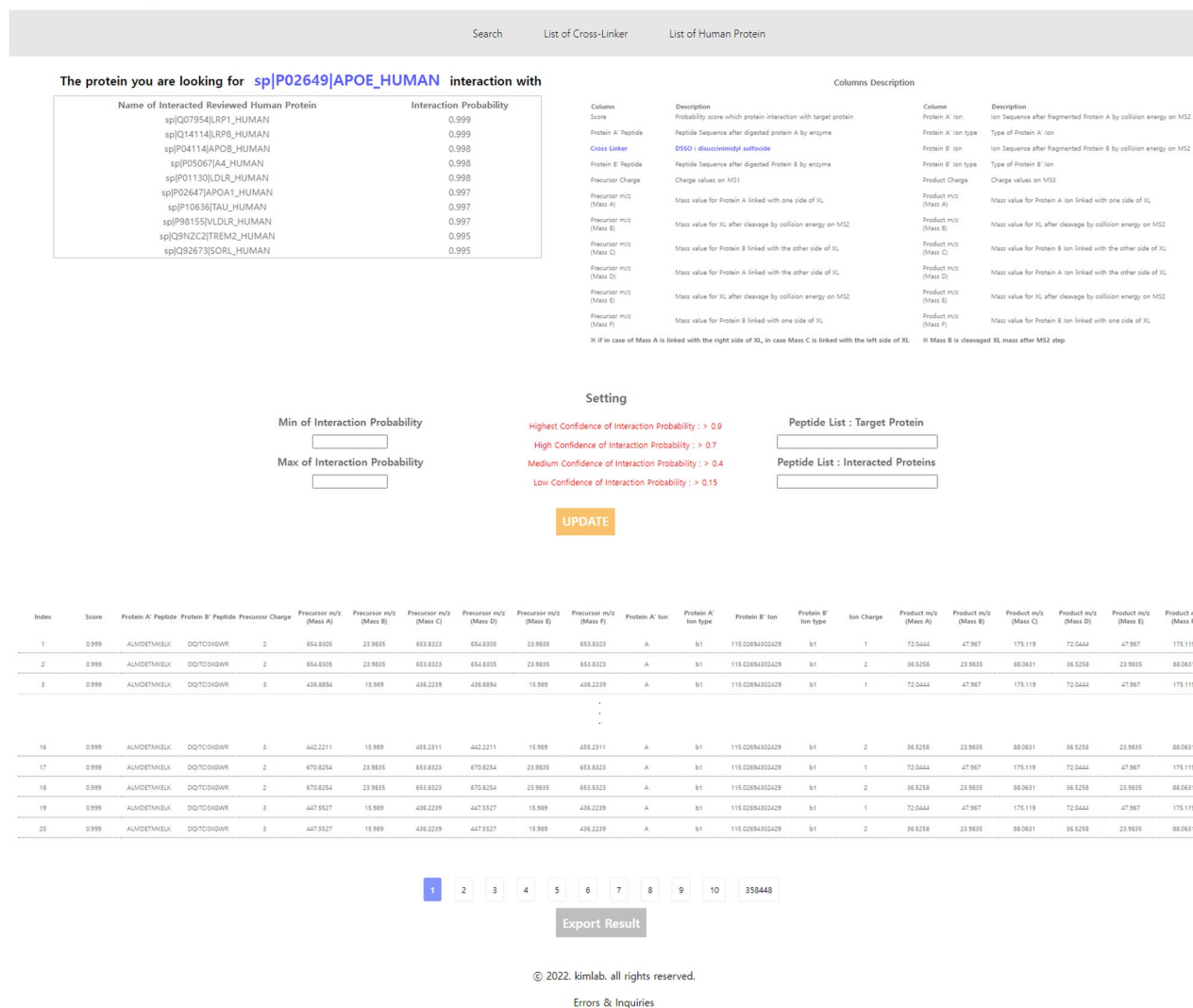
Cross-linking mass spectrometry (XL-MS) involves performing a mass spectrometry-based analysis on cross-linked complexes that have been digested by enzymes. When using PPIAT, the software calculates the mass values of the digested cross-linked protein complexes at the peptide and fragmented peptide ion levels. However, it does not consider the total mass of the cross-linked complex. To take into account all other products of the cross-linking reaction between Protein A and Protein B, it is necessary to consider four possibilities at the peptide and fragmented peptide ion levels: (1) cross-linked between Protein A and Protein B (cross-linking or interpeptide); (2) only cross-linked with Protein A, not cross-linked to Protein B (loop-link or intrapeptide); (3) only cross-linked with Protein B, not cross-linked to Protein A (loop-link or intrapeptide); and (4) non-cross-linked between Protein A and Protein B. (mono-link or dead-end)<sup>20,21</sup> In addition, when conducting XL-MS experiments, it is possible to encounter situations where interactions occur on one side, and the other side of the cross-linker may interact with Protein A and Protein B, respectively. To obtain more accurate data of the protein–protein interactions, these possibilities must be considered.

Cross-linkers are typically classified into two categories: cleavable and non-cleavable cross-linkers<sup>35,36</sup>. When using a cleavable cross-linker, it is important to consider its properties, such as the cleavage site and the type of cleavage that occurs (e.g., symmetric or asymmetric)<sup>35,36</sup>. If a cleavable cross-linker is used and is cleaved symmetrically, the same mass values should be added to Protein A and Protein B, respectively. However, if the cross-linker is cleaved asymmetrically, different mass values must be added. It is important to consider these factors when using a cleavable cross-linker to obtain accurate results. All calculated mass values can be exported to CSV format and used as input data for MS/MS analysis, which can provide further insights into the protein–protein interactions.

We utilized the PPIAT tool to explore potential PPIs involving APOE, a major biomarker for Alzheimer's disease<sup>37–45</sup>. Our search was performed under the following conditions: APOE(UniProt ID: P02649), trypsin digestion, cross-linker DSSO, top 10 interactions, peptide length of 9, peptide charge of 2 and 3, ion charge of 1 and 2, and modifications of carbamidomethylation(static) and oxidation(variable). Ten proteins were

## PPIAT

Protein-Protein Interaction Analytics Tool



**Figure 3.** Result page of PPIAT, which consists of two main sections. The first section identifies the protein–protein interactions (PPIs) and predicts the interacting proteins for the target protein, based on a combined score from the STRING database. The second section calculates the mass values for the PPIs, taking into consideration the conditions for cross-linking mass spectrometry (XL-MS).

identified by PPIAT, including LRP8(UniProt ID: Q14114, score:0.999), LRP1(UniProt ID: Q07954, score:0.999), A4(APP, UniProt ID: P05067, score:0.998), APOB(UniProt ID: P04114, score:0.998), LDLR(UniProt ID: P01130, score:0.998), APOA1(UniProt ID: P02647, score:0.997), VLDLR(UniProt ID: P98155, score:0.997), TAU(MAPT, UniProt ID: P10636, score:0.997), SORL(UniProt ID: Q92673, score:0.995), and TREM2(UniProt ID: Q9NZC2, score: 0.995) (Table 1), along with their theoretical information in XL (see Data S2 and S3). We verified that the PPIs output information matched the search results from STRING. Additionally, we confirmed the mass shift was consistent with the cross-linker DSSO by comparing the mass values using Skyline software (see Data S2 and S3).

## Discussion

XL-MS is a widely used method for identifying protein interactions and elucidating protein structures in three dimensions. In recent years, significant technological advancements in XL-MS studies have propelled the field of proteomics forward. However, there are still several limitations that need to be overcome, such as the need for better strategies for cross-linking. Unfortunately, there is currently no available solution or analysis tool that can accurately calculate the mass values of theoretical interactions required for SRM analysis. To address this challenge, our research team has developed PPIAT, a tool that can search for protein–protein interactions with target proteins, identify theoretical interactions that occur, and calculate their mass values at the peptide and fragmented peptide ion levels. With PPIAT as input data, mass values can be accurately calculated. This new tool provides an important advancement in the field of proteomics, helping researchers to overcome the challenges of XL-MS and improve their ability to identify protein–protein interactions.

Rank	Protein name	Interaction probability
1	sp Q07954 LRP1_HUMAN	0.999
2	sp Q14114 LRP8_HUMAN	0.999
3	sp P04114 APOB_HUMAN	0.998
4	sp P05067 A4_HUMAN	0.998
5	sp P01130 LDLR_HUMAN	0.998
6	sp P02647 APOA1_HUMAN	0.997
7	sp P10636 TAU_HUMAN	0.997
8	sp P98155 VLDLR_HUMAN	0.997
9	sp Q9NZC2 TREM2_HUMAN	0.995
10	sp Q92673 SORL_HUMAN	0.995

**Table 1.** APOE E4 PPIs search from PPIAT. Top 10 list of proteins interaction with APOE E4 protein. The results are matched with results on STRING.

Overall, the development of PPIAT represents a significant advancement in XL-MS studies and has the potential to enhance our understanding of protein–protein interactions. Although strategies such as the development of various cross-linkers have been employed in XL-MS, a lack of solutions or analysis tools for accurately calculating the mass values of theoretical interactions required for MS analysis still exists. To address this issue, several software tools have been developed, including SRMcollider<sup>29</sup>, Prego<sup>30</sup>, and X-Link Transition Calculator<sup>31</sup> in Skyline. However, these tools are not integrated with the platform used to search for theoretical PPIs, and the protein modification, binding site, and mass value of the cross-linker must be calculated and entered directly for XL-MS.

While PPIAT is a valuable tool for XL-MS analysis, it has some limitations that need to be addressed. Firstly, it does not cover all the products of the cross-linker, nor does it provide information about the enzyme or modification. Additionally, the output generated by PPIAT is only theoretical and requires experimental verification to confirm the results. To address these limitations, our team has built a database that includes information on 83 different cross-linkers, two types of enzymes (trypsin and chymotrypsin)<sup>20</sup>, and the most commonly occurring modifications (oxidation and carbamidomethylation). While these updates are a step forward, we will continue to improve and update the database regularly to ensure that it remains a useful resource for the scientific community.

The field of cross-linkers has seen a wide range of types with unique properties, including chemical cross-linkers and Photo/MS-cleavable cross-linkers<sup>35,36,46</sup>, which has been studied and commercialized. Additionally, researchers often synthesize their own cross-linkers with novel characteristics, such as BMSO, DBB, DHSO, and SDASO, for their experiments<sup>47–51</sup>. Although these synthetic cross-linkers are not currently included in the PPIAT database, an editing function will be added to the software to allow users to input information on these cross-linkers when searching for target proteins.

The PPIAT database is continuously updated with reviewed information on human proteins and protein–protein interactions. While the current version of the database has been built using PPI information from STRING<sup>12</sup> only, our plan is to expand it by integrating with other databases such as MINT<sup>13</sup>, BioGRID<sup>14</sup>, and IntAct<sup>15</sup> for PPI searching. The output generated by PPIAT can be downloaded in CSV format and used as input data for SRM analysis.

This analysis tool was developed on i9-10940X CPU, 128 GB RAM, Window server system. Researchers can identify PPIs up to the following conditions: (1) protein: AP02649 (APOE protein), (2) enzyme: trypsin, (3) cross-linker: DSSO, (4) peptide length range (min): 6, (5) peptide length range (max):10, number of interaction ranking: 10, (6) precursor charge: 2, 3, (7) product charge: 1, 2, (8) modifications: carbamidomethyl (C): static, oxidation (M): variable. The current analysis conditions are limited. We plan to make a server system that allows search conditions with at least 10 interacting proteins for the target protein and a peptide length ranging from 6 to 20.

## Conclusions

In conclusion, this paper introduces PPIAT, a web-based open platform designed for the analysis of XL-MS data. PPIAT provides a solution for the search of protein–protein interactions involving target proteins and calculation of the mass values of all theoretical interaction cases between proteins at both peptide and fragmented peptide ion levels, overcoming a significant obstacle in XL-MS analysis. We expect that this tool will be widely adopted by researchers in proteomics and bioinformatics, and we welcome contributions to its development. By continuously updating and refining the database, we aim to address the current limitations and improve the accuracy and efficiency of XL-MS analysis.

## Data availability

PPIAT is freely available at [ppiat.cnu.ac.kr/](http://ppiat.cnu.ac.kr/). All data utilized in this study were sourced from esteemed databases and publications. Specifically, the human protein data employed for targeted protein searches and the protein–protein interaction data for interaction analyses were acquired from the UniProt and STRING databases, respectively. These repositories are renowned for their rigorous maintenance and reliability. Additionally, the cross-linker data, utilized by the Protein–Protein Interaction Analysis Tool (PPIAT), was sourced from peer-reviewed white papers and is accessible in Supplementary Data 1. The novel data generated in this study can be

found in Supplementary Data 3. We ensure full accessibility of these data; researchers can freely utilize all data processed through PPIAT, which are exportable in CSV format directly from the results page.

## Code availability

The source code for PPIAT is available on GitHub at <https://github.com/kimlab-cnu>.

Received: 19 September 2023; Accepted: 11 December 2023

Published online: 13 December 2023

## References

- Stelzl, U. *et al.* A human protein-protein interaction network: A resource for annotating the proteome. *Cell* **122**(6), 957–968 (2005).
- Berg, T. Modulation of protein-protein interactions with small organic molecules. *Angew. Chem.-Int. Ed.* **42**(22), 2462 (2003).
- Richards, A. L. *et al.* Mass spectrometry-based protein-protein interaction networks for the study of human diseases. *Mol. Syst. Biol.* **17**(1), e8792 (2021).
- Ewing, R. M. *et al.* Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol. Syst. Biol.* **3**, 89 (2007).
- Ryan, D. P. M. J. M. Protein-protein interactions in human disease. *Curr. Opin. Struct. Biol.* **15**(4), 441 (2005).
- Wheat, A. *et al.* Protein interaction landscapes revealed by advanced in vivo cross-linking-mass spectrometry. *Proc. Natl. Acad. Sci.* **118**(32), e2023360118 (2021).
- Hanash, S. Disease proteomics. *Nature* **422**, 226 (2003).
- Hwang, S. *et al.* HumanNet v2: Human gene networks for disease research. *Nucleic Acids Res.* **47**(D1), D573–D580 (2019).
- Rabbani, G. *et al.* Protein-protein interactions and their role in various diseases and their prediction techniques. *Curr. Protein Pept. Sci.* **19**(10), 948–957 (2018).
- Yu, C. & Huang, L. Cross-linking mass spectrometry: An emerging technology for interactomics and structural biology. *Anal. Chem.* **90**(1), 144–165 (2018).
- Back, J. W. *et al.* Chemical cross-linking and mass spectrometry for protein structural modeling. *J. Mol. Biol.* **331**(2), 303–313 (2003).
- Szklarczyk, D. *et al.* The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**(D1), D605–D612 (2021).
- Chatri-aryamontri, A. *et al.* MINT: The Molecular INTeraction database. *Nucleic Acids Res.* **35**(Database issue), D572–574 (2007).
- Oughtred, R. *et al.* The BioGRID interaction database: 2019 update. *Nucleic Acids Res.* **47**(D1), D529–D541 (2019).
- del Toro, N. *et al.* The IntAct database: Efficient access to fine-grained molecular interaction data. *Nucleic Acids Res.* **50**(D1), D648–D653 (2022).
- Mendoza, V. L. & Vachet, R. W. Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass Spectrom. Rev.* **28**(5), 785–815 (2009).
- Kiselar, J. G. & Chance, M. R. Future directions of structural mass spectrometry using hydroxyl radical footprinting. *J. Mass Spectrom.* **45**(12), 1373–1382 (2010).
- Koneremann, L. *et al.* Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.* **40**(3), 1224–1234 (2011).
- Olegklykov, B. S., Pektaş, S., Fasci, D., Heck, A. J. R. & Scheltema, R. A. Efficient and robust proteome-wide approaches for crosslinking mass spectrometry. *Nat. Protoc.* **6**, 2964–2990 (2018).
- Rappsilber, J. The beginning of a beautiful friendship: Cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J. Struct. Biol.* **173**(3), 530–540 (2011).
- Leitner, A. *et al.* Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol. Cell Proteom.* **9**(8), 1634–1649 (2010).
- Sinz, A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom. Rev.* **25**(4), 663–682 (2006).
- Young, M. M. *et al.* High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc. Natl. Acad. Sci.* **97**(11), 5802–5806 (2000).
- Rappsilber, J. *et al.* A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry. *Anal. Chem.* **72**(2), 267–275 (2000).
- O'Reilly, F. J. & Rappsilber, J. Cross-linking mass spectrometry: Methods and applications in structural, molecular and systems biology. *Nat. Struct. Mol. Biol.* **25**(11), 1000–1008 (2018).
- Maiolica, A. *et al.* Structural analysis of multiprotein complexes by cross-linking, mass spectrometry, and database searching. *Mol. Cell Proteom.* **6**(12), 2200–2211 (2007).
- Uetrecht, C. *et al.* Ion mobility mass spectrometry of proteins and protein assemblies. *Chem. Soc. Rev.* **39**(5), 1633–1655 (2010).
- Lorenzen, K. & van Duijn, E. Native mass spectrometry as a tool in structural biology. *Curr. Protoc. Protein Sci.* **17**(Unit 17), 12 (2010).
- Röst, H. *et al.* A computational tool to detect and avoid redundancy in selected reaction monitoring. *Mol. Cell. Proteom.* **11**(8), 540–549 (2012).
- Searle, B. C. *et al.* Using data independent acquisition (DIA) to model high-responding peptides for targeted proteomics experiments. *Mol. Cell. Proteom.* **14**(9), 2331–2340 (2015).
- Chavez, J. D. *et al.* A general method for targeted quantitative cross-linking mass spectrometry. *PLOS ONE* **11**(12), e0167547 (2016).
- MacLean, B. T. *et al.* Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**(7), 966 (2010).
- Bateman, A. *et al.* UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**(D1), D480–D489 (2021).
- Schluter, H. *et al.* Finding one's way in proteomics: A protein species nomenclature. *Chem. Center J.* **3**, 11 (2009).
- Sinz, A. Divide and conquer: Cleavable cross-linkers to study protein conformation and protein-protein interactions. *Anal. Bioanal. Chem.* **409**(1), 33–44 (2017).
- Schilling, B. *et al.* MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. *J. Am. Soc. Mass Spectrom.* **14**(8), 834–850 (2003).
- Maurer, K. *et al.* Auguste D and Alzheimer's disease. *Lancet* **349**(9064), 1546–1549 (1997).
- Alloul, K. *et al.* Alzheimer's disease: A review of the disease, its epidemiology and economic impact. *Arch. Gerontol. Geriatr.* **27**(3), 189–221 (1998).
- Terry, R. D. & Davies, P. Dementia of the Alzheimer type. *Annu. Rev. Neurosci.* **3**(1), 77–95 (1980).
- Mendez, M. F. Early-onset Alzheimer disease. *Neurol. Clin.* **35**(2), 263–281 (2017).
- Corder, E. H. *et al.* Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**(5123), 921–923 (1993).
- Blennow, K. *et al.* Alzheimer's disease. *Lancet* **368**(9533), 387–403 (2006).

43. Ballard, C. *et al.* Alzheimer's disease. *Lancet* **377**(9770), 1019–1031 (2011).
44. Scheltens, P. *et al.* Alzheimer's disease. *Lancet* **388**(10043), 505–517 (2016).
45. Lane, C. A. *et al.* Alzheimer's disease. *Eur. J. Neurol.* **25**(1), 59–70 (2018).
46. Yang, L. *et al.* A photocleavable and mass spectrometry identifiable cross-linker for protein interaction studies. *Anal. Chem.* **82**(9), 3556–3566 (2010).
47. Gutierrez, C. B. *et al.* Development of a novel sulfoxide-containing MS-cleavable homobifunctional cysteine-reactive cross-linker for probing protein–protein interactions. *Anal. Chem.* **90**(12), 7600–7607 (2018).
48. Cui, L. *et al.* Tyrosine-reactive cross-linker for probing protein three-dimensional structures. *Anal. Chem.* **93**(10), 4434–4440 (2021).
49. Gutierrez, C. B. *et al.* Developing an acidic residue reactive and sulfoxide-containing MS-cleavable homobifunctional cross-linker for probing protein–protein interactions. *Anal. Chem.* **88**(16), 8315–8322 (2016).
50. Huang, R. *et al.* A novel mass spectrometry-cleavable, phosphate-based enrichable and multi-targeting protein cross-linker. *Chem. Sci.* **10**(26), 6443–6447 (2019).
51. Gutierrez, C. *et al.* Enabling photoactivated cross-linking mass spectrometric analysis of protein complexes by novel MS-cleavable cross-linkers. *Mol. Cell Proteom.* **20**, 100084 (2021).

## Author contributions

A.R.S., and H.S.K. conceived the supervised the project. J.H.P. developed the PPIAT analysis tool under the guidance of A.R.S. and H.S.K. J.H.P., A.R.S. and H.S.K. wrote the manuscript. J.H.P. taking the lead on writing the methods and results sections and A.R.S. and H.S.K. providing critical revisions and feedback throughout the writing process.

## Funding

The Institute of Information & Communications Technology Planning & Evaluation (IITP) grant, funded by the Korea government (MSIT) (No.RS-2022-00155857), supported the development of this project as part of the Artificial Intelligence Convergence Innovation Human Resources Development program at Chungnam National University. Additionally, the National Research Foundation of Korea (NRF) grant, funded by the Korea government (MSIT), provided support for this research (No. 2022R1F1A1074235 & No. 2021R1G1A1092933). The authors declare that there are no conflicts of interest related to this research.

## Competing interests

The authors declare that they have no conflicts of interest. The Figures included in this manuscript were created using the Biorender software.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-49663-4>.

**Correspondence** and requests for materials should be addressed to H.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023