



RESEARCH ARTICLE

**REVISED** **Identifying and characterising Thrap3, Bclaf1 and Erh interactions using cross-linking mass spectrometry [version 2; peer review: 2 approved]**

Liudmila Shcherbakova , Mercedes Pardo , Theodoros Roumeliotis, Jyoti Choudhary

Cancer Biology, Institute of Cancer Research, UK, London, UK

**v2** **First published:** 12 Oct 2021, **6:260**  
<https://doi.org/10.12688/wellcomeopenres.17160.1>  
**Latest published:** 06 Jan 2023, **6:260**  
<https://doi.org/10.12688/wellcomeopenres.17160.2>

**Abstract**





**Background:** Cross-linking mass spectrometry (XL-MS) is a powerful technology capable of yielding structural insights across the complex cellular protein interaction network. However, up to date most of the studies utilising XL-MS to characterise individual protein complexes' topology have been carried out on over-expressed or recombinant proteins, which might not accurately represent native cellular conditions.



**Methods:** We performed XL-MS using MS-cleavable crosslinker disuccinimidyl sulfoxide (DSSO) after immunoprecipitation of endogenous BRG/Brahma-associated factors (BAF) complex and co-purifying proteins. Data are available via ProteomeXchange with identifier PXD027611.

**Results:** Although we did not detect the expected enrichment of crosslinks within the BAF complex, we identified numerous crosslinks between three co-purifying proteins, namely Thrap3, Bclaf1 and Erh. Thrap3 and Bclaf1 are mostly disordered proteins for which no 3D structure is available. The XL data allowed us to map interaction surfaces on these proteins, which overlap with the non-disordered portions of both proteins. The identified XLs are in agreement with homology-modelled structures suggesting that the interaction surfaces are globular.

**Conclusions:** Our data shows that MS-cleavable crosslinker DSSO can be used to characterise in detail the topology and interaction surfaces of endogenous protein complexes without the need for overexpression. We demonstrate that Bclaf1, Erh and Thrap3 interact closely with each other, suggesting they might form a novel complex, hereby referred to as TEB complex. This data can be exploited for modelling protein-protein docking to characterise the three-dimensional structure of the complex. Endogenous XL-MS might be challenging due to crosslinker accessibility, protein complex abundance or isolation efficiency, and require further optimisation for

**Open Peer Review****Approval Status**  

|   | 1   | 2   |
|---|---|---|
| <b>version 2</b><br>(revision)<br>06 Jan 2023 | <br><a href="#">view</a>   |   |
|   |                          |   |
| <b>version 1</b><br>12 Oct 2021               | <br><a href="#">view</a> | <br><a href="#">view</a> |

- Alexander Leitner** , ETH Zurich, Zurich, Switzerland
- J. Mark Skehel** , Francis Crick Institute, London, UK

Any reports and responses or comments on the article can be found at the end of the article.

some complexes like the BAF complex to detect a substantial number of crosslinks.

### Keywords

Endogenous complexes, XL-MS, CLMS cross-linking mass spectrometry, DSSO, Thrap3, Bclaf1, Erh

**Corresponding authors:** Liudmila Shcherbakova ([liudmila.shcherbakova@icr.ac.uk](mailto:liudmila.shcherbakova@icr.ac.uk)), Jyoti Choudhary ([jyoti.choudhary@icr.ac.uk](mailto:jyoti.choudhary@icr.ac.uk))

**Author roles:** **Shcherbakova L:** Data Curation, Investigation, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Pardo M:** Methodology, Project Administration, Supervision, Writing – Review & Editing; **Roumeliotis T:** Investigation, Methodology, Writing – Review & Editing; **Choudhary J:** Conceptualization, Funding Acquisition, Project Administration, Supervision, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This work was supported by the Wellcome Trust [208391]; the Institute of Cancer Research (Hensley Nankivell Studentship Fund); and the Cancer Research UK Centre grant (C309/A25144).

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2023 Shcherbakova L *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Shcherbakova L, Pardo M, Roumeliotis T and Choudhary J. **Identifying and characterising Thrap3, Bclaf1 and Erh interactions using cross-linking mass spectrometry [version 2; peer review: 2 approved]** Wellcome Open Research 2023, 6:260 <https://doi.org/10.12688/wellcomeopenres.17160.2>

**First published:** 12 Oct 2021, 6:260 <https://doi.org/10.12688/wellcomeopenres.17160.1>

**REVISED Amendments from Version 1**

Following the review of the first version of this paper, we implemented some changes in the second version. The novel complex we originally proposed to be called BET was renamed as TEB (THRAP3-ERH-BCLAF1) to avoid confusion with the existing nomenclature of the BET domain. The crosslink distance constraint was changed from 37 Å to a more stringent value of 32 Å.

Figure 5 was amended to display a representative structural model where all crosslinks satisfied the 32 Å distance constraint. Further details on sample preparation (amount of protein used in the IP, IPP150 buffer composition, crosslinker concentration) were added to the methods section.

**Any further responses from the reviewers can be found at the end of the article**

**Introduction**

Cross-linking mass spectrometry (XL-MS) is a powerful technique that enables the identification of proximal amino acid residues within a single protein as well as residues in close proximity in interacting proteins. Intramolecular crosslinks provide distance constraint parameters that can support homology-based tertiary structure modelling or even guide *de novo* modelling (Adikaram *et al.*, 2019; Kahraman *et al.*, 2013; Liu *et al.*, 2020; Orbán-Németh *et al.*, 2018; O'Reilly & Rappsilber, 2018). Moreover, intermolecular cross-linking information has been used to determine spatial orientation and elucidate the topology of protein complex subunits (Gaik *et al.*, 2015; Herzog *et al.*, 2012; Leitner *et al.*, 2016; Politis *et al.*, 2014). In the absence of a full tertiary structure model protein-protein interaction (PPI) surface information derived from XL-MS can potentially serve as a guide to direct mutation or small molecule screens to disrupt specific subunit interactions within a given complex.

Most XL-MS studies to date have been performed on over-expressed or recombinant proteins or protein complexes (O'Reilly & Rappsilber, 2018). More rarely, XL-MS has been applied to study protein interactions in a more complex endogenous setting like cell lysates (Götze *et al.*, 2019; Klykov *et al.*, 2018; Liu *et al.*, 2015; Yugandhar *et al.*, 2020), organelles (Bartolec *et al.*, 2020; Makepeace *et al.*, 2020; Ryl *et al.*, 2020) specific cellular compartments (Fasci *et al.*, 2018; Ser *et al.*, 2019) or purified complexes expressed at endogenous levels (Makowski *et al.*, 2016; Spruijt *et al.*, 2021). Recent advances in XL-MS have introduced MS-cleavable crosslinkers, which generate distinct fragment pairs in MS2 and therefore substantially reduce the complexity of data analysis and improve identification accuracy (Matzinger & Mechtler, 2021). Disuccinimidyl sulfoxide (DSSO) (Kao *et al.*, 2011) is the most extensively used MS-cleavable reagent, and has been applied to the characterisation of protein interactions both in isolated proteins as well as complex samples (Adams *et al.*, 2020; Klykov *et al.*, 2018; Nguyen *et al.*, 2020; Ser *et al.*, 2019; Smith *et al.*, 2018; Wang *et al.*, 2017).

Here, we performed XL-MS on affinity-purified (AP) endogenous BRG/Brahma-associated factors (BAF) complex in native conditions to define the interaction surfaces between complex subunits and with associated protein partners. The strategy was based on immunoprecipitation of Arid1a, considered to be a scaffolding/bridging component of the complex (Han *et al.*, 2020; He *et al.*, 2020; Mashtalir *et al.*, 2018), and unlike most published XL-MS studies, involved no artificial increase of the target protein complex by overexpression. In addition to BAF purification, we concomitantly achieved a significant co-enrichment of Thrap3, Bclaf1 and Erh proteins that allowed the identification of a substantial number of crosslinks (XLs) between them, implicating this protein cluster as a native protein assembly, that we hereby refer to as TEB complex. We report the interaction surfaces between Thrap3, Bclaf1 and Erh, determined through chemical crosslinking mass-spectrometry.

**Results**

Bclaf1, Erh and Thrap3 interact directly with each other. To investigate interaction surfaces of Arid1a and the BAF complex, we carried out five experiments where immunoprecipitation of endogenous Arid1a from mouse embryonic stem cells (mESCs) was coupled to crosslinking using MS cleavable crosslinker DSSO and MS3 mass spectrometry. Crosslink peptide identification was performed using the full mouse Uniprot database, with crosslink assignment using XlinkX (Liu *et al.*, 2015) as a part of Proteome Discoverer at a 1% false discovery rate (FDR).

Whilst we were able to detect several XLs between BAF subunits and interacting proteins, these were mostly single crosslink spectra matches (CSMs) with exception of two. We also detected a significant number of XLs between Thrap3, Erh and Bclaf1, due to a strong enrichment of these proteins in Arid1a APs. Specifically, 13% of all XLs identified in the five experiments involved these three proteins. For comparison, 9.6% of total XLs obtained were assigned to BAF, which, in addition to being the bait in the AP, is also a much larger complex with 27 subunits. Given the richness of XL information associated with these proteins, we used this data to gain insight into the structural topology of this protein cluster. Our data suggests that they interact closely and might represent a stable complex. We hereby refer to it as the Bclaf1-Erh-Thrap3 (TEB) complex.

Bclaf1 and Thrap3 are highly homologous proteins (Figure 1). They share the Thrap3-Bclaf1 domain, which defines the THRAP3/BCLAF1 family, that contains these two proteins. There is no available crystal or cryo-electron microscopy structure for Thrap3 or Bclaf1. Interestingly, domain analysis with Pfam database (v33.1) (Mistry *et al.*, 2021) showed that Thrap3 and Bclaf1 are both highly disordered proteins along the whole length (Figure 2A), which limits the application of structure prediction modelling on them without additional information. Hence our XL-MS data could add useful information for structural elucidation of the complex.

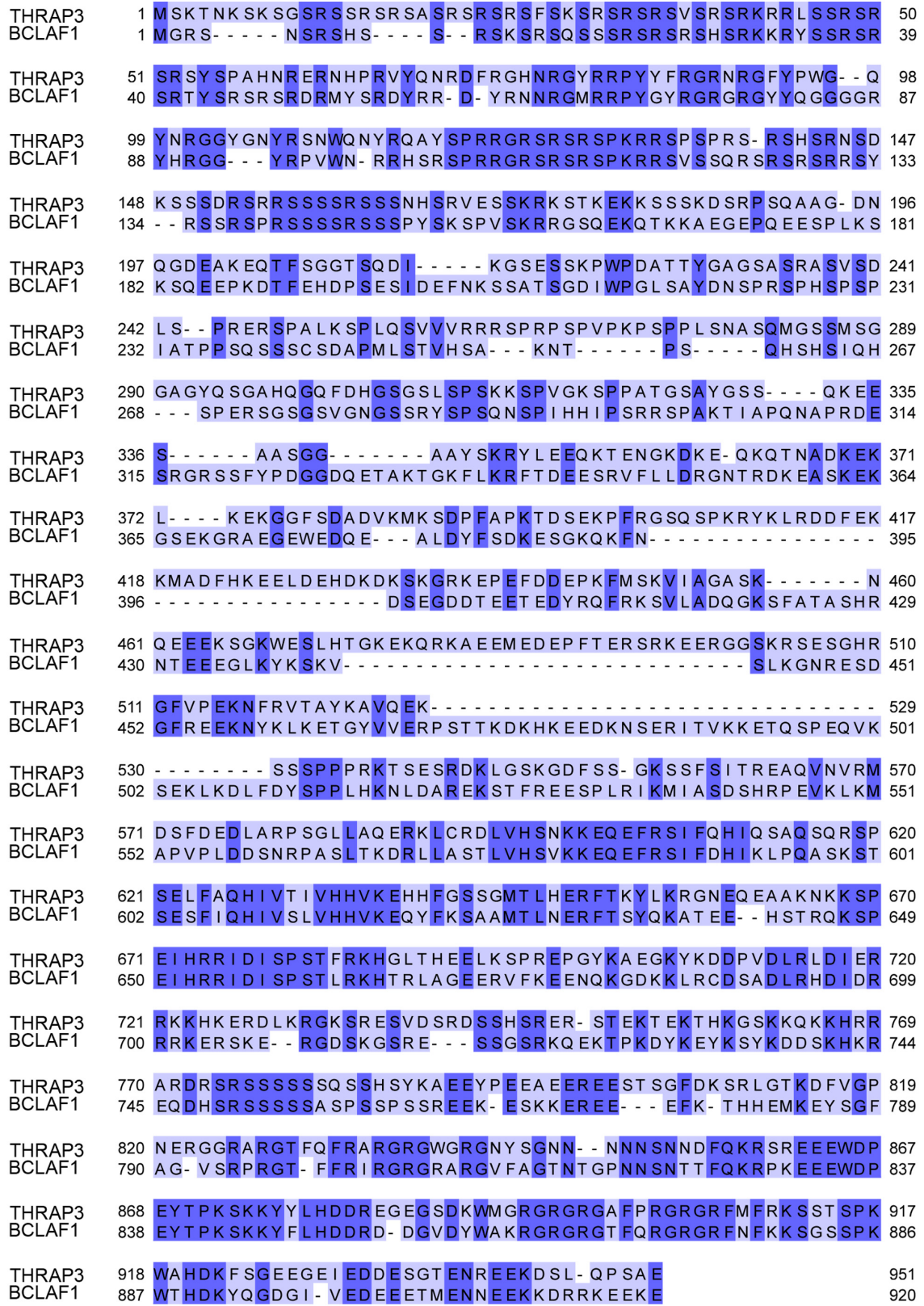
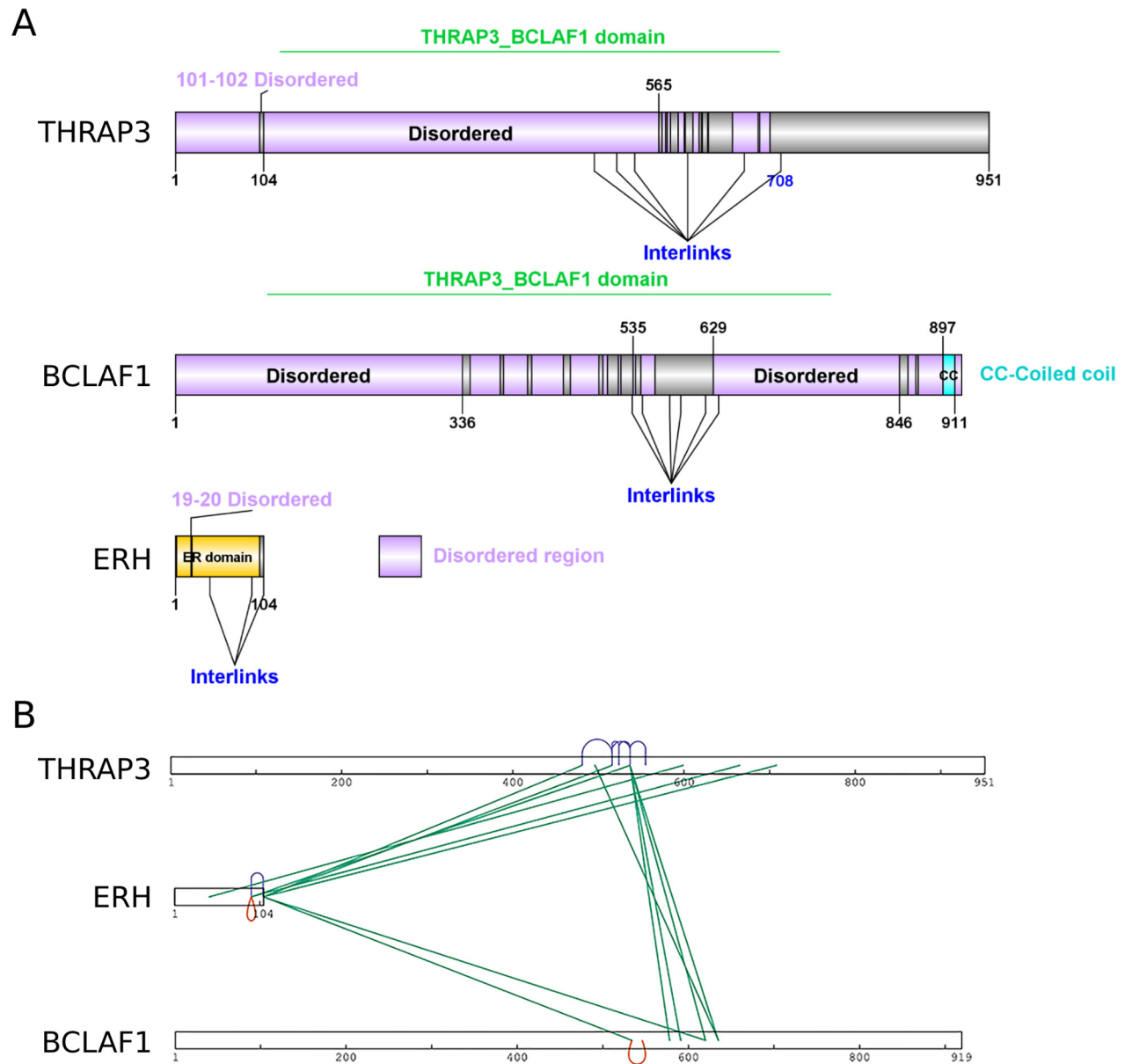


Figure 1. Protein sequence alignment of mouse Thrap3 and mouse Bclaf1. Proteins showed ~43% homology. The alignment was performed using Clustal Omega sequence alignment tool (Madeira et al., 2019) within the Jalview v2.11.1.4 (Waterhouse et al., 2009).



**Figure 2. Structural features of TEB complex members and identified cross-linked sites. (A)** Annotations of domains and disorder regions of Thrap3, Bclaf1 and Erh as reported in Pfam database (Mistry *et al.*, 2021). The illustration was made using DOG 2.0 (Ren *et al.*, 2009). **(B)** A two dimensional visualisation of Thrap3, Bclaf1 and Erh depicting the high confidence cross-links using xiVIEW (Graham *et al.*, 2019).

We summarised confident crosslinks and corresponding number of CSMs for each of five experiments in Table 1. Due to the high sequence homology between Thrap3 and Bclaf1 (Figure 1), the crosslinks were manually checked for unambiguous assignment between the two proteins and undistinguishable pairs were removed from analysis. We detected 24 unique XL sites and a total of 121 CSMs across all five experiments. The high density of crosslinks in the regions from 496-537 amino acids (aa) for Thrap3 and 578-635aa for Bclaf1 (Figure 2B) indicates the interaction surface between the two proteins. This

region of Thrap3 also contains most of XLs to-self, suggesting that it is a potentially globular area and is used for PPI. On the contrary, no intra-links were retained for Bclaf1 upon high confidence filtering (Figure 2B). Even though they share ~43% homology, the middle parts of these proteins are quite different between approximately 400aa and 560aa, with fragments missing in one protein or the other (Figure 1). The Bclaf1 and Thrap3 interaction interface also contains their cross-linked sites to Erh, which extends slightly further (up to 481-708aa in Thrap3 and 534-620aa in Bclaf1). Moreover, the region of

**Table 1. TEB complex confident crosslinks.** This table reports the confident crosslinks, which are defined as being detected in more than one experiment, and its corresponding CSMs number for each of five experiments. XL-pairs.ID indicate peptide that contains crosslink to the unique site pair. For each unique crosslinked peptide pair we report the crosslinked spectra match (CSMs) count for each individual experiment. All CSMs fulfill the 1% FDR cut-off.

| XL-pair_ID | Crosslink Type | Sequence A        | Accession A | Protein Descriptions A           | Position A | Sequence B       | Accession B | Protein Descriptions B           | Position B | Exp.1 #CSMs | Exp.2 #CSMs | Exp.3 #CSMs | Exp.4 #CSMs | Exp.5 #CSMs |
|------------|----------------|-------------------|-------------|----------------------------------|------------|------------------|-------------|----------------------------------|------------|-------------|-------------|-------------|-------------|-------------|
| 1          | Inter          | EESTSGFD[K]SR     | Q569Z6      | Thrap3                           | 808        | LLIY[K]VSNR      | P01631      | Ig kappa chain V-II region 26-10 | 55         | 1           | 0           | 4           | 5           | 0           |
| 2          | Inter          | Y[K]DDPVDLR       | Q569Z6      | Thrap3                           | 708        | QAQQAG[K]        | P84089      | Erh                              | 104        | 0           | 0           | 1           | 1           | 0           |
| 3          | Inter          | GNEQEA[K]NKK      | Q569Z6      | Thrap3                           | 665        | E[K]YVLLR        | P84089      | Erh                              | 90         | 0           | 0           | 1           | 1           | 0           |
| 4          | Inter          | FTSQ[K]ATEEHSTR   | Q8K019      | Bclaf1                           | 635        | SR[K]EER         | Q569Z6      | Thrap3                           | 496        | 2           | 0           | 3           | 0           | 0           |
| 5          | Inter          | FTSQ[K]ATEEHSTR   | Q8K019      | Bclaf1                           | 635        | [K]TSESR         | Q569Z6      | Thrap3                           | 537        | 1           | 0           | 2           | 0           | 1           |
| 6          | Inter          | EQYF[K]SPAVTLNER  | Q8K019      | Bclaf1                           | 620        | QAQQAG[K]        | P84089      | Erh                              | 104        | 0           | 1           | 0           | 1           | 3           |
| 7          | Inter          | EQYF[K]SPAVTLNER  | Q8K019      | Bclaf1                           | 620        | [K]TSESR         | Q569Z6      | Thrap3                           | 537        | 1           | 0           | 2           | 0           | 3           |
| 8          | Inter          | SIFDHI[K]LPQANK   | Q8K019      | Bclaf1                           | 591        | [K]TSESR         | Q569Z6      | Thrap3                           | 537        | 0           | 0           | 1           | 1           | 0           |
| 9          | Inter          | LLASTLVHSV[K]K    | Q8K019      | Bclaf1                           | 578        | [K]TSESR         | Q569Z6      | Thrap3                           | 537        | 0           | 0           | 1           | 1           | 0           |
| 10         | Intra          | GFSSG[K]SSEFSTR   | Q569Z6      | Thrap3                           | 555        | [K]TSESR         | Q569Z6      | Thrap3                           | 537        | 0           | 0           | 4           | 1           | 0           |
| 11         | Inter          | IK[M]IASDSHRPEVK  | Q8K019      | Bclaf1                           | 534        | QAQQAG[K]        | P84089      | Erh                              | 104        | 1           | 0           | 1           | 0           | 0           |
| 12         | Intra          | IK[M]IASDSHRPEVK  | Q8K019      | Bclaf1                           | 534        | MIASDSHRPEV[K]LK | Q8K019      | Bclaf1a                          | 546        | 0           | 0           | 0           | 0           | 6           |
| 13         | Intra          | VTAV[K]AVQEK      | Q569Z6      | Thrap3                           | 524        | GFVPE[K]NFR      | Q569Z6      | Thrap3                           | 516        | 0           | 0           | 1           | 1           | 0           |
| 14         | Intra          | VTAV[K]AVQEK      | Q569Z6      | Thrap3                           | 524        | [K]TSESR         | Q569Z6      | Thrap3                           | 537        | 1           | 0           | 4           | 0           | 0           |
| 15         | Inter          | GFVPE[K]NFR       | Q569Z6      | Thrap3                           | 516        | QAQQAG[K]        | P84089      | Erh                              | 104        | 2           | 0           | 4           | 0           | 3           |
| 16         | Intra          | GFVPE[K]NFR       | Q569Z6      | Thrap3                           | 516        | QAQQAG[K]        | P84089      | Erh                              | 104        | 0           | 0           | 0           | 5           | 0           |
| 17         | Inter          | [K]AEEMEDPEPFTER  | Q569Z6      | Thrap3                           | 481        | [K]TSESR         | Q569Z6      | Thrap3                           | 537        | 0           | 0           | 1           | 2           | 0           |
| 18         | Inter          | [K]AEEMEDPEPFTER  | Q569Z6      | Thrap3                           | 481        | QAQQAG[K]        | P84089      | Erh                              | 104        | 0           | 1           | 1           | 1           | 0           |
| 19         | Intra          | [K]AEEMEDPEPFTER  | Q569Z6      | Thrap3                           | 481        | QAQQAG[K]        | P84089      | Erh                              | 104        | 0           | 0           | 0           | 0           | 3           |
| 20         | Intra          | [K]AEEMEDPEPFTER  | Q569Z6      | Thrap3                           | 481        | GFVPE[K]NFR      | Q569Z6      | Thrap3                           | 516        | 0           | 0           | 0           | 0           | 3           |
| 21         | Intra          | E[K]YVLLR         | P84089      | Erh                              | 90         | E[K]YVLLR        | P84089      | Erh                              | 90         | 0           | 0           | 1           | 2           | 5           |
| 22         | Intra          | E[K]YVLLR         | P84089      | Erh                              | 90         | QAQQAG[K]        | P84089      | Erh                              | 104        | 0           | 0           | 2           | 2           | 1           |
| 23         | Inter          | FSGSGGTDFTL[K]ISR | P01631      | Ig kappa chain V-II region 26-10 | 79         | [K]TSESR         | Q569Z6      | Thrap3                           | 537        | 0           | 0           | 4           | 3           | 1           |
| 24         | Inter          | MYEEHL[K]R        | P84089      | Erh                              | 41         | EESTSGFD[K]ISR   | Q569Z6      | Thrap3                           | 808        | 0           | 0           | 3           | 4           | 4           |

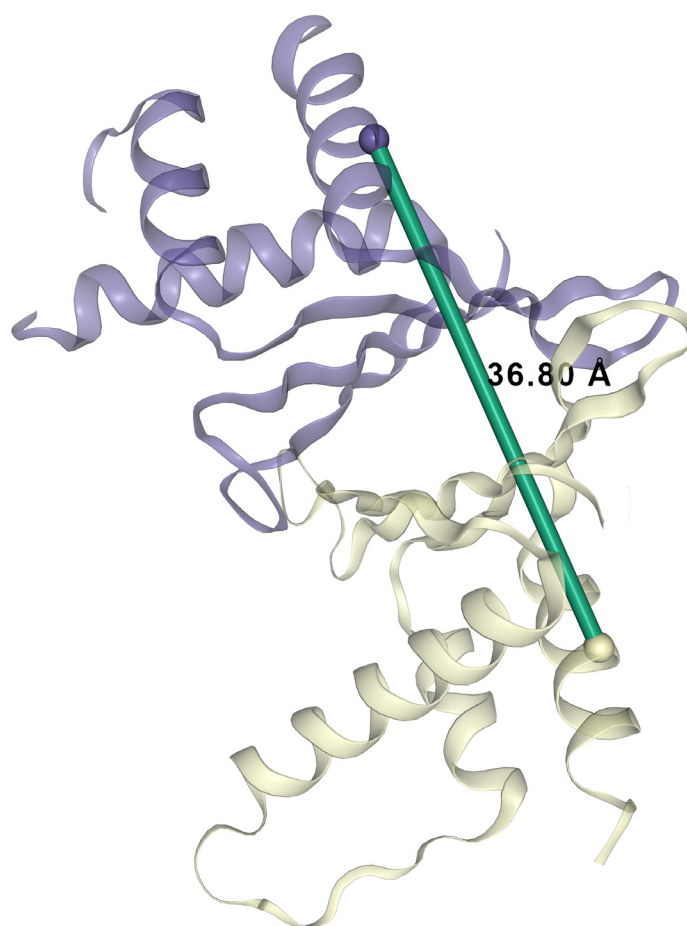
dense crosslinks in Bclaf1 lies outside of the disordered region as reported in Pfam database (v33.1) (Mistry *et al.*, 2021) (561–629aa; Figure 2A), suggesting that the Bclaf1 interaction surface, like that of Thrap3, may also be a globular structure. Our data demonstrates that Bclaf1, Erh and Thrap3 interaction is direct and suggests that they might form a ternary complex which we have named TEB (Thrap3-Erh-Bclaf1a).

Previous studies identified that both the N-terminus (1–190aa) and the C-terminus (359–951aa) of Thrap3 are important for its function in DNA repair of post double stranded breaks and stalled replication forks, whilst the residues 190–359aa appear to be dispensable for this function (Vohhodina *et al.*, 2017). The C-terminal fragment of Thrap3 has been shown to partially rescue the Thrap3 knockout phenotype in terms of ability to respond to ionising radiation (IR) induced DNA damage (Vohhodina *et al.*, 2017). Since this Thrap3 C-terminal fragment covers the XL-rich region, the rescue may have been mediated by restoration of interactions with Erh and Bclaf1. Moreover, there is a number of highly frequent phosphorylation sites (>100 reports per site) reported in

PhosphoSitePlus (v6.6.0.1) (Hornbeck *et al.*, 2015) within the Thrap3 C-terminus in mice (S379, S572, S679, S924).

#### Mapping TEB crosslinks on available structures

Erh is the only component of the TEB complex with an available crystal structure (Arai *et al.*, 2005; Hazra *et al.*, 2020; Jin *et al.*, 2007; Kwon *et al.*, 2020; Li *et al.*, 2005; Wan *et al.*, 2005; Xie *et al.*, 2019). We detected a XL for Erh that involves two connected peptides with overlapping sequences (Table 1; XL-pair ID 19). Although XL-MS is not able to distinguish between intra- or inter-links in the case of homomultimeric proteins, the overlap in the crosslinked sequences conclusively identifies an inter-molecular interaction, that is, a crosslink involving two molecules of the same protein (or alternatively a false positive peptide identification). In agreement with this data, Erh is known to form a homodimer (Arai *et al.*, 2005; Hazra *et al.*, 2020; Xie *et al.*, 2019) and therefore we concluded that the self-link at the position 90 involves two Erh molecules. We utilised available Erh homodimer protein structure PDB:1WZ7 (Arai *et al.*, 2005) to model this crosslink within a structural context (Figure 3). The length of



**Figure 3. Erh dimer inter-crosslink.** ERH inter-crosslink at position K90-K90 was mapped onto available Erh (Uniprot accession: P84089) structure PDB:1WZ7 (Arai *et al.*, 2005).

the mapped crosslink is in agreement with the DSSO maximum distance constraint of 37Å threshold used in other modelling studies (Liu *et al.*, 2020).

#### TEB crosslinks satisfy the model based on sequence homology

Since there is no available crystal or cryo-EM structure for Thrap3 or Bclaf1, we used online modelling tool Robetta for *de novo* modelling of putative structures (Raman *et al.*, 2009; Song *et al.*, 2013), and five predicted structures were rendered for each protein. In all models Thrap3 appeared to be largely unfolded and flexible (Figure 4), with roughly the same area stretching from 330-725aa folding into a number of helices that resemble a globular structure. This region overlapped with the crosslink-dense area of Thrap3 (Figure 5). When XLs were mapped onto the proposed models, 5 out of the 5 crosslinks we detected satisfied the more restrictive distance threshold of 32Å (Armony *et al.*, 2021) in models 2, 3 and 5, whilst 4 out of 5 did in models 1 and 4.

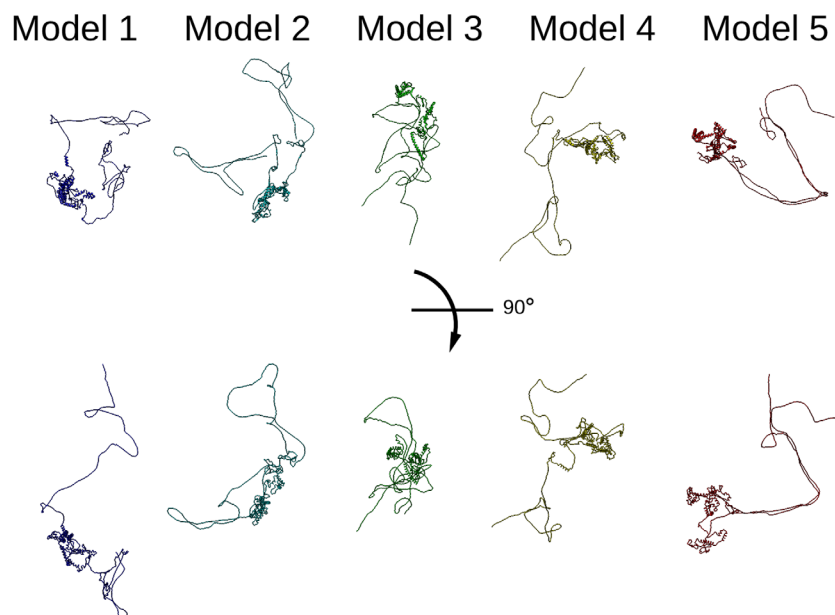
Surprisingly, despite the homology between Bclaf1 and Thrap3, the predicted structures for Bclaf1 appeared generally much more folded (Figure 6) than those of Thrap3 and overall looked very different.

#### Discussion

Thrap3 and Bclaf1 have been shown to form part of a high molecular weight complex (the SNARP complex) involved in the regulation of cyclin D1 stability together with SNIP1, SKIP and Pinin (Bracken *et al.*, 2008). Interactions between Bclaf1 and Thrap3 with Erh have been identified through

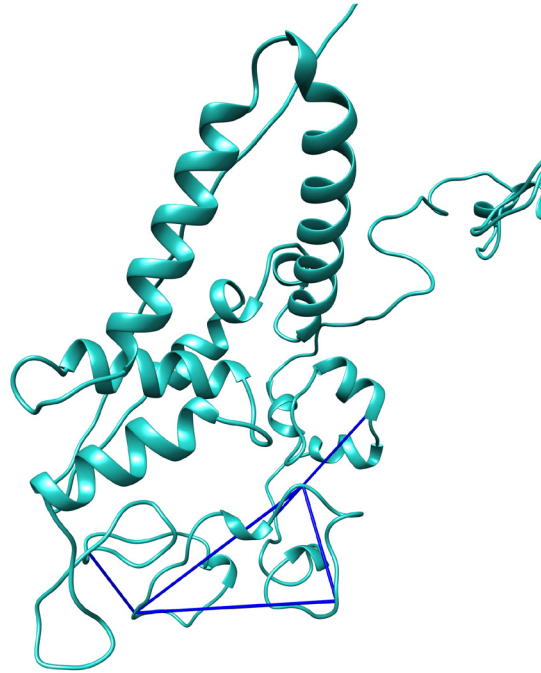
immunoprecipitation of Erh (Kavanaugh *et al.*, 2015), but these interactions were not further validated. Our XL-MS data confirms that Bclaf1, Thrap3 and Erh are in close proximity to each other and provides evidence that they can interact directly. Furthermore, the experimental conditions of cell lysis (high salt buffer) and the number of XLs detected suggest that they form a tight complex that we refer to as the TEB complex. Bclaf1, Erh and Thrap3 are functionally related and are all involved in splicing regulation. Thrap3 and Bclaf1 have also been shown to play an important role in DNA damage repair (DDR) through regulation of splicing of ataxia telangiectasia mutated (*ATM*) serine/threonine kinase and export of its mRNA, as well as regulation of other transcripts involved in DDR (Vohhodina *et al.*, 2017), while Erh has been reported to regulate splicing of ataxia telangiectasia and Rad3-related (*ATR*) protein (Kavanaugh *et al.*, 2015). Thrap3 and Bclaf1 are able to compensate for the loss of each other (Vohhodina *et al.*, 2017) indicating partial functional redundancy.

There is a discrepancy in the literature as to whether any components of TEB localise at the DNA damage (DD) sites with some data suggesting that Thrap3 and Bclaf1a do not localise (Beli *et al.*, 2012), and neither does Erh (Kavanaugh *et al.*, 2015). In fact, Thrap3 is suggested to be excluded from double strand breaks (Beli *et al.*, 2012). However, another study showed that Bclaf1 interacts and co-localises with H2AX after IR in damage foci (Lee *et al.*, 2012). Bclaf1 has also been reported to form a complex with phosphorylated BRCA1, downstream of its function as *ATM* splicing regulator, promoting splicing of DDR proteins such as ATRIP, BACH1 and EXO1 (Savage *et al.*, 2014). The diverse function of Bclaf1 can be partially explained

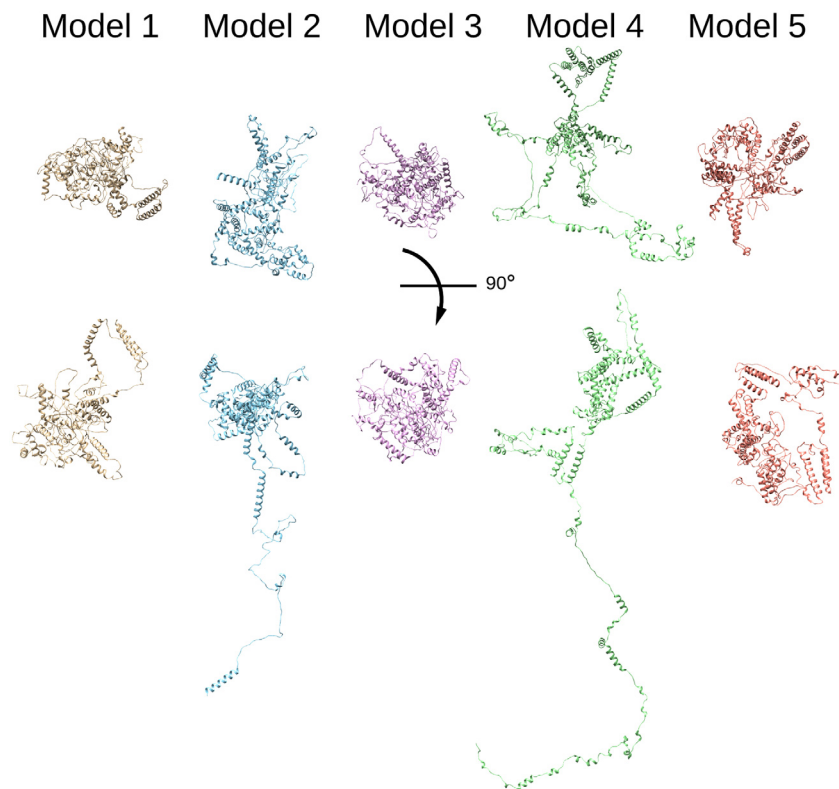


**Figure 4. Thrap3 sequence-based modelling.** Thrap3 (mouse) protein models generated using Robetta protein modelling tool (Raman *et al.*, 2009; Song *et al.*, 2013) based solely on the protein sequence.





**Figure 5. Thrap3 crosslinks satisfy the *de novo* homology-based model.** Predicted zoomed in structure of Thrap3 model 2 using Robetta protein modelling tool (Raman *et al.*, 2009; Song *et al.*, 2013). Cross-linking data was mapped using Chimera with distance constraint set at 32Å (Pettersen *et al.*, 2004).



**Figure 6. Bclaf1 sequence-based modelling.** Bclaf1 (mouse) protein model generated using Robetta protein modelling tool (Raman *et al.*, 2009; Song *et al.*, 2013) based solely on the protein sequence.

by itself being a subject of splicing regulation, with one specific splicing isoform implicated in regulation of tumour growth (Zhou *et al.*, 2014).

Crosslinks can not only be used to refine low or medium-resolution structures, but also aid the generation of protein models from their amino acid sequences (Liu *et al.*, 2020; Orbán-Németh *et al.*, 2018). The crosslinking data presented here could be incorporated as  $C_{\alpha}$ - $C_{\alpha}$  distance restraints using I-TASSER Roy *et al.*, 2010; Yang *et al.*, 2015; Zhang, 2008) to refine the preliminary protein models we generated and improve their reliability. If satisfactory model structures are obtained, XL data can again be exploited through the HADDOCK platform for inter-molecular docking of the TEB complex subunits (van Zundert *et al.*, 2016). However, due to lack of confident intramolecular XLs for Bclaf1 or available resolved structure of its homologues, its modelling may not be accurate enough to perform this task successfully.

In summary, we have performed XL-MS on endogenous protein complexes to derive useful topological information. We have shown that Thrap3, Bclaf1 and Erh interact directly with each other to form a tight protein assembly and we have identified their interaction interfaces through XL-MS. Abnormal splicing events are often observed in cancer cells and have been involved in many types of cancer. Hence, characterisation of the interactions between these proteins will be useful for better understanding their role in oncogenesis

## Methods

### Cell culture

Mouse embryonic stem cells (mESCs) were cultured by StemCell Technologies Inc. (Cambridge, UK).

### Immunoprecipitation and crosslinking

Protein G-Dynabeads (#10004D, Invitrogen) were prepared by coupling to antibodies against specific target protein in Table 2 for 15 minutes at room temperature (RT). Nuclear extraction was performed with isotonic buffer containing 10mM Tris-HCl, 10mM NaCl, 1.5mM MgCl<sub>2</sub>, 0.34M Sucrose, 1mM dithiothreitol (DTT), Halt Protease inhibitors (Thermo Scientific) and 0.05% NP-40. Nuclei were lysed for 10 minutes on ice with lysis buffer composed of 50 mM TrisHCl pH 8.0, 450 mM NaCl, 1 mM EDTA, 1 mM DTT, Halt Protease inhibitors (Thermo Scientific). NaCl was diluted to 150mM for immunoprecipitation (IP) for experiments 1-3 and 5, but kept

at 450mM for experiment 4. Protein concentration was measured by Quick Start™ Bradford 1x Dye Reagent (#5000205, Biorad) following the manufactures guidelines. IP was performed from 10–60 mg protein depending on the experiment for 1–2 hours at 4°C. Crosslinking reagent disuccinimidyl sulfoxide (DSSO) (#A33545, ThermoFisher) was dissolved in DMSO at 50mM. The crosslinking was performed at 1mM (experiments 1-4) or 5mM (experiment 5) on the IP sample while still coupled to the beads, with DMSO final concentration being 2% at RT for 1 hour. The reaction was quenched with 125mM Tris-HCl pH 8.0 at RT for 15 minutes. The beads were washed three times with IPP150 (150mM NaCl, 10mM Tris-HCl, 0.1% NP40). For samples that were destined for MS analysis, the beads were washed and digested as previously described in Hillier *et al.* (2019) with a modified digestion schedule: Lys C digestion overnight, followed by Trypsin for approximately 4 hours, two times and a Trypsin overnight digestion.

### Mass spectrometry

Peptides generated by trypsin digestion were fractionated with the Pierce High pH Reversed-Phase Peptide Fractionation Kit (#84868, ThermoFisher) according to manufacturer instructions, and eight fractions were collected and dried. Liquid chromatography mass spectrometry (LC-MS) analysis was performed on the Dionex UltiMate 3000 UHPLC system coupled with the Orbitrap Lumos Mass Spectrometer (Thermo Scientific). Each peptide fraction was reconstituted in 15  $\mu$ L 0.1% formic acid and loaded to the Acclaim PepMap 100, 100  $\mu$ m  $\times$  2 cm C18, 5  $\mu$ m trapping column at 10  $\mu$ L/min flow rate of 0.1% formic acid loading buffer. The sample was then subjected to a gradient elution on the EASY-Spray C18 capillary column (75  $\mu$ m  $\times$  50 cm, 2  $\mu$ m) at 50°C. Mobile phase A was 0.1% formic acid and mobile phase B was 80% acetonitrile, 0.1% formic acid. The gradient separation method at flow rate 300 nL/min was as follows: for 90 minutes gradient from 5%–38% B, for 10 minutes up to 95% B, for 5 minutes isocratic at 95% B, re-equilibration to 5% B in 5 minutes, for 10 minutes isocratic at 5% B. MS scans were acquired at a mass resolution of 120,000 and precursors between 375–1,600 m/z and charge equal or higher than +3 were isolated for collision-induced dissociation (CID) fragmentation with quadrupole isolation width 1.6 Th in the top speed mode in cycles of 5 seconds. Collision energy was set at 25%. Fragments with targeted mass difference of 31.9721 (DSSO crosslinker) were further subjected to CID fragmentation at the MS3 level with collision energy 35%, iontrap detection and MS2 isolation window 2 Th. Two precursor groups were selected with both ions in the

**Table 2.** The antibodies used in immunoprecipitation (IP).

| Target | Name      | Catalogue number | RRID      | Type       | Species in which the antibody was raised | Company                  |
|--------|-----------|------------------|-----------|------------|--|--------------------------|
| Arid1a | PSG3      | sc-32761         | AB_673396 | Monoclonal | Mouse                                    | Santa Cruz Biotechnology |
| N/A    | Mouse-IgG | 12-371           | AB_145840 | Polyclonal | Mouse                                    | Millipore                |

pair. Targeted MS precursors were dynamically excluded for further isolation and activation for 30 seconds with 10 ppm mass tolerance.

### Mass spectrometry and crosslinking data analysis

MS raw data from mESCs IPs was analysed using Proteome Discoverer 2.4 (#OPTON-30945; Thermo Scientific). XLinkX was used to search crosslinked peptides pairs (Liu *et al.*, 2015) for tryptic peptides with a minimum peptide length of 5 aa and maximum of 2 miss cleavages with dynamic modifications set to as oxidation of methionine (+15.995 Da) at 1% FDR. Precursor mass tolerance set to 20ppm, FTMS fragment mass tolerance to 30ppm, and ITMS fragment tolerance to 0.5 Da. Sequest was used for general MS2 search for tryptic peptides with a minimum peptide length of 6 aa and maximum of 2 missed cleavages with DSSO modifications followed by Target Decoy PSM Validator set at 1% FDR. Dynamic modifications were set as oxidation of methionine (+15.995 Da), DSSO hydrolysed for lysine (+176.014 Da) and DSSO quenched with Tris/dead end for lysine (+279.078 Da). The searches were performed against full UniProt Mouse database (August 2019) together with cRAP contaminant database. As an alternative to Proteome Discoverer 2.4 with XLinkX and Sequest, open-access software such as MaxQuant (Cox & Mann, 2008) pLink (Chen *et al.*, 2019), XQuest/xProphet (Leitner *et al.*, 2014; Rinner *et al.*, 2008; Walzthoeni *et al.*, 2012), StravoX (Götze *et al.*, 2012), MeroX (Götze *et al.*, 2015), Kojak (Hoopmann *et al.*, 2015), XiSEARCH (Mendes *et al.*, 2019) or MaxLinker (Yugandhar *et al.*, 2020) could be used for XL identification.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE

(Perez-Riverol *et al.*, 2019) partner repository with the dataset identifier PXD027611 (see underlying data).

### Crosslinking mapping, protein modelling and visualisation

Annotations of domains and disorder regions of Thrap3, Bclaf1 and Erh as reported in Pfam database (v33.1) (Mistry *et al.*, 2021). The illustration of the domains was made using DOG 2.0 (Ren *et al.*, 2009). Two dimensional visualisation of crosslinks was performed using xiVIEW online tool (Graham *et al.*, 2019). The crosslinks visualised were filtered so at least one of the proteins from the cross-linked pair is a member of the TEB complex and the site of the cross-linked pair was observed in more than one experiment. Structural models of Bclaf1 and Thrap3 were predicted by Robetta online tool with TrRefineRosetta modelling method for Bclaf1, and comparative modelling and *ab initio* modelling for Thrap3 (Raman *et al.*, 2009; Song *et al.*, 2013). PDB files of the model and the available structures were visualised using UCSF Chimera package (Pettersen *et al.*, 2004) and crosslinks were mapped using a Chimera plug-in Xlink Analyzer (Kosinski *et al.*, 2015). Chimera was developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311).

### Data availability

#### Underlying data

The underlying data has been deposited in the ProteomeXchange Consortium via the PRIDE partner repository, accession number PXD027611: <https://identifiers.org/pride.project:PXD027611>.

## References

- Adams MK, Banks CAS, Thornton JL, *et al.*: **Differential Complex Formation via Paralogs in the Human Sin3 Protein Interaction Network.** *Mol Cell Proteomics.* 2020; **19**(9): 1468–1484.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Adikaram PR, Zhang JH, Kitzcock CM, *et al.*: **Development of R7BP inhibitors through cross-linking coupled mass spectrometry and integrated modeling.** *Commun Biol.* 2019; **2**: 338.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Arai R, Kukimoto-Niino M, Uda-Tochio H, *et al.*: **Crystal structure of an enhancer of rudimentary homolog (ERH) at 2.1 Angstroms resolution.** *Protein Sci.* 2005; **14**(7): 1888–93.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Armony G, Heck AJR, Wu W: **Extracellular crosslinking mass spectrometry reveals HLA class I – HLA class II interactions on the cell surface.** *Mol Immunol.* 2021; **136**: 16–25.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bartolec TK, Smith DL, Pang CNI, *et al.*: **Cross-linking Mass Spectrometry Analysis of the Yeast Nucleus Reveals Extensive Protein–Protein Interactions Not Detected by Systematic Two-Hybrid or Affinity Purification-Mass Spectrometry.** *Anal Chem.* 2020; **92**(2): 1874–1882.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Beli P, Lukashchuk N, Wagner SA, *et al.*: **Proteomic investigations reveal a role for RNA processing factor THRAP3 in the DNA damage response.** *Mol Cell.* 2012; **46**(2): 212–25.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bracken CP, Wall SJ, Barré B, *et al.*: **Regulation of cyclin D1 RNA stability by SNIP1.** *Cancer Res.* 2008; **68**(18): 7621–8.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chen ZL, Meng JM, Cao Y, *et al.*: **A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides.** *Nat Commun.* 2019; **10**(1): 3404.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cox J, Mann M: **MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.** *Nat Biotechnol.* 2008; **26**(12): 1367–1372.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Fasci D, van Ingen H, Scheltema RA, *et al.*: **Histone Interaction Landscapes Visualized by Crosslinking Mass Spectrometry in Intact Cell Nuclei.** *Mol Cell Proteomics.* 2018; **17**(10): 2018–2033.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gaïk M, Flemming D, von Appen A, *et al.*: **Structural basis for assembly and function of the Nup82 complex in the nuclear pore scaffold.** *J Cell Biol.* 2015; **208**(3): 283–297.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Götze M, Iacobucci C, Ihling CH, *et al.*: **A Simple Cross-Linking/Mass Spectrometry Workflow for Studying System-wide Protein Interactions.**

- Anal Chem.* 2019; **91**(15): 10236–10244.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Götze M, Pettelkau J, Fritzsche R, et al.: **Automated Assignment of MS/MS Cleavable Cross-Links in Protein 3D-Structure Analysis.** *J Am Soc Mass Spectrom.* 2015; **26**(1): 83–97.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Götze M, Pettelkau J, Schaks S, et al.: **StavroX—a software for analyzing crosslinked products in protein interaction studies.** *J Am Soc Mass Spectrom.* 2012; **23**(1): 76–87.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Graham M, Combe C, Kolbowski L, et al.: **xiView: A common platform for the downstream analysis of Crosslinking Mass Spectrometry data.** *bioRxiv.* 2019.  
[Publisher Full Text](#)
- Han Y, Reyes AA, Malik S, et al.: **Cryo-EM structure of SWI/SNF complex bound to a nucleosome.** *Nature.* 2020; **579**(7799): 452–455.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hazra D, Andrić V, Palancade B, et al.: **Formation of *S. pombe* Erh1 homodimer mediates gametogenic gene silencing and meiosis progression.** *Sci Rep.* 2020; **10**(1): 1034.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- He S, Wu Z, Tian Y, et al.: **Structure of nucleosome-bound human BAF complex.** *Science.* 2020; **367**(6480): 875–881.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Herzog F, Kahraman A, Boehringer D, et al.: **Structural Probing of a Protein Phosphatase 2A Network by Chemical Cross-Linking and Mass Spectrometry.** *Science.* 2012; **337**(6100): 1348–52.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hillier C, Pardo M, Yu L, et al.: **Landscape of the *Plasmodium* Interactome Reveals Both Conserved and Species-Specific Functionality.** *Cell Rep.* 2019; **28**(6): 1635–1647.e5.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hoopmann MR, Zelter A, Johnson RS, et al.: **Kojak: Efficient Analysis of Chemically Cross-Linked Protein Complexes.** *J Proteome Res.* 2015; **14**(5): 2190–2198.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hornbeck PV, Zhang B, Murray B, et al.: **PhosphoSitePlus, 2014: mutations, PTMs and recalibrations.** *Nucleic Acids Res.* 2015; **43**(Database issue): D512–20.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jin T, Guo F, Serebriiskii IG, et al.: **A 1.55 Å resolution X-ray crystal structure of HEF2/ERH and insights into its transcriptional and cell-cycle interaction networks.** *Proteins.* 2007; **68**(2): 427–37.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kahraman A, Herzog F, Leitner A, et al.: **Cross-link guided molecular modeling with ROSETTA.** *PLoS One.* 2013; **8**(9): e73411.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kao A, Chiu CL, Vellucci D, et al.: **Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes.** *Mol Cell Proteomics.* 2011; **10**(1): M110.002212.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kavanaugh G, Zhao R, Guo Y, et al.: **Enhancer of Rudimentary Homolog Affects the Replication Stress Response through Regulation of RNA Processing.** *Mol Cell Biol.* 2015; **35**(17): 2979–90.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Klykov O, Steigenberger B, Pektaş S, et al.: **Efficient and robust proteome-wide approaches for cross-linking mass spectrometry.** *Nat Protoc.* 2018; **13**(12): 2964–2990.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kosinski J, von Appen A, Ori A, et al.: **Xlink Analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures.** *J Struct Biol.* 2015; **189**(3): 177–83.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kwon SC, Jang H, Shen S, et al.: **ERH facilitates microRNA maturation through the interaction with the N-terminus of DGCR8.** *Nucleic Acids Res.* 2020; **48**(19): 11097–11112.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lee YY, Yu YB, Gunawardena HP, et al.: **BCLAF1 is a radiation-induced H2AX-interacting partner involved in γH2AX-mediated regulation of apoptosis and DNA repair.** *Cell Death Dis.* 2012; **3**(7): e359.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Leitner A, Faini M, Stengel F, et al.: **Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines.** *Trends Biochem Sci.* 2016; **41**(1): 20–32.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Leitner A, Walzthoeni T, Aebersold R.: **Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline.** *Nat Protoc.* 2014; **9**(1): 120–37.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Li H, Inoue M, Yabuki T, et al.: **Solution structure of the mouse enhancer of rudimentary protein reveals a novel fold.** *J Biomol NMR.* 2005; **32**(4): 329–34.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Liu F, Rijkers DT, Post H, et al.: **Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry.** *Nat Methods.* 2015; **12**(12): 1179–1184.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Liu X, Zhang Y, Wen Z, et al.: **Driving integrative structural modeling with serial capture affinity purification.** *Proc Natl Acad Sci U S A.* 2020; **117**(50): 31861–31870.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Madeira F, Park YM, Lee J, et al.: **The EMBL-EBI search and sequence analysis tools APIs in 2019.** *Nucleic Acids Res.* 2019; **47**(W1): W636–W641.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Makepeace KAT, Mohammed Y, Rudashevskaya EL, et al.: **Improving Identification of *In-organello* Protein-Protein Interactions Using an Affinity-enrichable, Isotopically Coded, and Mass Spectrometry-cleavable Chemical Crosslinker.** *Mol Cell Proteomics.* 2020; **19**(4): 624–639.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Makowski MM, Willems E, Jansen PW, et al.: **Cross-linking immunoprecipitation-MS (xIP-MS): Topological Analysis of Chromatin-associated Protein Complexes Using Single Affinity Purification.** *Mol Cell Proteomics.* 2016; **15**(3): 854–65.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mashtalir N, D'Avino AR, Michel BC, et al.: **Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes.** *Cell.* 2018; **175**(5): 1272–1288.e20.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Matzinger M, Mechtler K.: **Cleavable Cross-Linkers and Mass Spectrometry for the Ultimate Task of Profiling Protein-Protein Interaction Networks *in Vivo*.** *J Proteome Res.* 2021; **20**(1): 78–93.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mendes ML, Fischer L, Chen ZA, et al.: **An integrated workflow for crosslinking mass spectrometry.** *Mol Syst Biol.* 2019; **15**(9): e8994.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mistry J, Chuguransky S, Williams L, et al.: **Pfam: the protein families database in 2021.** *Nucleic Acids Res.* 2021; **49**(D1): D412–D419.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nguyen TT, Blackburn MR, Sussman MR.: **Intermolecular and Intramolecular Interactions of the *Arabidopsis* Plasma Membrane Proton Pump Revealed Using a Mass Spectrometry Cleavable Cross-Linker.** *Biochemistry.* 2020; **59**(24): 2210–2225.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- O'Reilly FJ, Rappsilber J.: **Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology.** *Nat Struct Mol Biol.* 2018; **25**(11): 1000–1008.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Orbán-Németh Z, Beveridge R, Hollenstein DM, et al.: **Structural prediction of protein models using distance restraints derived from cross-linking mass spectrometry data.** *Nat Protoc.* 2018; **13**(3): 478–494.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Perez-Riverol Y, Csordas A, Bai J, et al.: **The PRIDE database and related tools and resources in 2019: improving support for quantification data.** *Nucleic Acids Res.* 2019; **47**(D1): D442–D450.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Pettersen EF, Goddard TD, Huang CC, et al.: **UCSF Chimera—a visualization system for exploratory research and analysis.** *J Comput Chem.* 2004; **25**(13): 1605–12.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Politis A, Stengel F, Hall Z, et al.: **A mass spectrometry-based hybrid method for structural modeling of protein complexes.** *Nat Methods.* 2014; **11**(4): 403–406.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Raman S, Vernon R, Thompson J, et al.: **Structure prediction for CASP8 with all-atom refinement using Rosetta.** *Proteins.* 2009; **77** Suppl 9(0 9): 89–99.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ren J, Wen L, Gao X, et al.: **DOG 1.0: illustrator of protein domain structures.** *Cell Res.* 2009; **19**(2): 271–3.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Rinner O, Seebacher J, Walzthoeni T, et al.: **Identification of cross-linked peptides from large sequence databases.** *Nat Methods.* 2008; **5**(4): 315–8.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Roy A, Kucukural A, Zhang Y.: **I-TASSER: a unified platform for automated protein structure and function prediction.** *Nat Protoc.* 2010; **5**(4): 725–38.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ryl PS, Bohlke-Schneider JM, Lenz S, et al.: ***In Situ* Structural Restraints from Cross-Linking Mass Spectrometry in Human Mitochondria.** *J Proteome Res.* 2020; **19**(1): 327–336.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Savage KI, Gorski JJ, Barros EM, et al.: **Identification of a BRCA1-mRNA splicing complex required for efficient DNA repair and maintenance of genomic stability.** *Mol Cell.* 2014; **54**(3): 445–59.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ser Z, Cifani P, Kentsis A.: **Optimized Cross-Linking Mass Spectrometry for *In Situ* Interaction Proteomics.** *J Proteome Res.* 2019; **18**(6): 2545–2558.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Smith DL, Götze M, Bartolec TK, et al.: **Characterization of the Interaction**

**between Arginine Methyltransferase Hmt1 and Its Substrate Npl3: Use of Multiple Cross-Linkers, Mass Spectrometric Approaches, and Software Platforms.** *Anal Chem.* 2018; **90**(15): 9101–9108.

[PubMed Abstract](#) | [Publisher Full Text](#)

Song Y, DiMaio F, Wang RY, *et al.*: **High-resolution comparative modeling with RosettaCM.** *Structure.* 2013; **21**(10): 1735–42.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Spruijt CG, Gräwe C, Kleinendorst SC, *et al.*: **Cross-linking mass spectrometry reveals the structural topology of peripheral NuRD subunits relative to the core complex.** *FEBS J.* 2021; **288**(10): 3231–3245.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

van Zundert GCP, Rodrigues JPLM, Trellet M, *et al.*: **The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes.** *J Mol Biol.* 2016; **428**(4): 720–725.

[PubMed Abstract](#) | [Publisher Full Text](#)

Vohhodina J, Barros EM, Savage AL, *et al.*: **The RNA processing factors THRAP3 and BCLAF1 promote the DNA damage response through selective mRNA splicing and nuclear export.** *Nucleic Acids Res.* 2017; **45**(22): 12816–12833.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Walzthoeni T, Claassen M, Leitner A, *et al.*: **False discovery rate estimation for cross-linked peptides identified by mass spectrometry.** *Nat Methods.* 2012; **9**(9): 901–903.

[PubMed Abstract](#) | [Publisher Full Text](#)

Wan C, Tempel W, Liu ZJ, *et al.*: **Structure of the conserved transcriptional repressor enhancer of rudimentary homolog.** *Biochemistry.* 2005; **44**(13): 5017–23.

[PubMed Abstract](#) | [Publisher Full Text](#)

Wang X, Cimermancic P, Yu C, *et al.*: **Molecular Details Underlying Dynamic Structures and Regulation of the Human 26S Proteasome.** *Mol Cell Proteomics.* 2017; **16**(5): 840–854.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Waterhouse AM, Procter JB, Martin DMA, *et al.*: **Jalview Version 2—a multiple sequence alignment editor and analysis workbench.** *Bioinformatics.* 2009; **25**(9): 1189–1191.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Xie G, Vo TV, Thillainadesan G, *et al.*: **A conserved dimer interface connects ERH and YTH family proteins to promote gene silencing.** *Nat Commun.* 2019; **10**(1): 251.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Yang J, Yan R, Roy A, *et al.*: **The I-TASSER Suite: protein structure and function prediction.** *Nat Methods.* 2015; **12**(1): 7–8.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Yugandhar K, Wang TY, Leung AKY, *et al.*: **MaXLinker: Proteome-wide Cross-link Identifications with High Specificity and Sensitivity\*.** *Mol Cell Proteomics.* 2020; **19**(3): 554–568.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Zhang Y: **I-TASSER server for protein 3D structure prediction.** *BMC Bioinformatics.* 2008; **9**: 40.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Zhou X, Li X, Cheng Y, *et al.*: **BCLAF1 and its splicing regulator SRSF10 regulate the tumorigenic potential of colon cancer cells.** *Nat Commun.* 2014; **5**: 4581.

[PubMed Abstract](#) | [Publisher Full Text](#)

# Open Peer Review

Current Peer Review Status:  

---

## Version 2

Reviewer Report 11 January 2023

<https://doi.org/10.21956/wellcomeopenres.20372.r53966>

© 2023 Leitner A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Alexander Leitner** 

Institute of Molecular Systems Biology, Department of Biology, ETH Zurich, Zurich, Switzerland

The changes to the text are appreciated. I have no further comments.

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** cross-linking mass spectrometry; proteomics; structural biology; protein-protein interactions, protein-RNA interactions

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---

## Version 1

Reviewer Report 08 July 2022

<https://doi.org/10.21956/wellcomeopenres.18954.r51240>

© 2022 Skehel J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**J. Mark Skehel** 

Francis Crick Institute, London, UK

The manuscript by Shcherbakova *et al.*, describes the use of chemical cross-linking to study the protein interactions of proteins and protein complexes at endogenous levels following affinity purification (IP-XL-MS). The manuscript is well written and straightforward to follow. The study

sets out to investigate the interactions of the BAF complex, by immunoprecipitating Arid1a, a scaffolding protein of the BAF complex. Unfortunately, the experiment yielded too few cross-links of the BAF complex to enable detailed topological investigation, however, three proteins, Thrap3, Bclaf1, and Erh were identified with numerous cross-links, suggesting that they may form a previously unidentified complex. In all, five experiments were carried out yielding a total of 121 cross-links. Of these, 24 CSMs (cross-link spectral matches) were attributed to subunits of this putative complex.

The study highlights an important aspect of structural mass spectrometry, namely that chemical cross-linking enables the structural analysis of intrinsically disordered proteins and disordered regions within a protein. Thrap3 and Bclaf1 are predicted to have extensively disordered regions, which makes it very difficult if not impossible to study the structures by crystallography or cryo-electron microscopy. While the data here provide limited intra and inter cross-links, they are starting point for more extensive investigation of this putative new complex.

Raw data underlying this work has been deposited in the ProteomeXchange Consortium via the PRIDE repository.

Minor comments:

1. The authors suggest naming the putative Bclaf1a-Erh-Thrap3 complex, the BET complex. This is potentially confusing, as there are already the Bromodomain and Extra-Terminal domain (BET) proteins (BRD2, BRD3, BRD4, and BRDT). Perhaps an alternative nomenclature would be better.
2. Table 1 should additionally show the confidence score attributed to a CSM by the software used for data analysis (XLinkX). The legend also refers to 4 experiments where five are shown.
3. The authors have employed the Robetta protein modelling tool to predict possible structures for these intrinsically disordered proteins. It would be useful to compare these structures with those predicted by AlphaFold.
4. Searches were performed against the full UniProt Mouse database. Potentially an alternative approach would have been to characterise the IP obtained using Arid1a and build a more targeted database against which to search the data.
5. The authors specify how protein concentrations were measured, but don't say how much protein was used for the IP. Additionally, a value of 1 mM is given for the cross-linking reagent, how does this relate to a protein: cross-linker ratio? Were several concentration ratios investigated?
6. On page 11, the authors mention that alternative software is available for data analysis. This list is not extensive and should at least include other applications such as Xi, StavroX, etc.
7. The authors get slightly fixated on the distance restraint of 37Å. This is a lenient value and

may not have any particular importance for these disordered proteins.

8. For the cross-linking data analysis, it may prove useful to include serine and threonine in the reaction specificity for DSSO.

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Yes

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Proteomics, structural mass spectrometry

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 25 Nov 2022

**Liudmila Shcherbakova**, Institute of Cancer Research, UK, London, UK

- The authors suggest naming the putative Bclaf1a-Erh-Thrap3 complex, the BET complex. This is potentially confusing, as there are already the Bromodomain and Extra-Terminal domain (BET) proteins (BRD2, BRD3, BRD4, and BRDT). Perhaps an alternative nomenclature would be better.

*We thank the reviewer for pointing this out. We have now changed the name to TEB complex throughout the text.*

- Table 1 should additionally show the confidence score attributed to a CSM by the software used for data analysis (XLinkX). The legend also refers to 4 experiments where five are shown.

*The XlinkX Target Decoy PSM Validator was set at 1% FDR as reported in the methods. All crosslinks reported in table 1 fulfilled the 1% FDR cut-off. This has now been added to the legend for clarity and the number of experiments (5) has also been corrected.*

- The authors have employed the Robetta protein modelling tool to predict possible



structures for these intrinsically disordered proteins. It would be useful to compare these structures with those predicted by AlphaFold.

*AlphaFold was released after the manuscript had been submitted, but we agree with the reviewer that it would be interesting to compare the modelled structures with the AlphaFold predicted structures. AlphaFold structures for human THRAP3 and BCLAF1 look more disordered than those presented in the paper for the mouse counterparts but retained a few alpha helices in the central region of both THRAP3 and BCLAF1, similarly to the structures generated by Robetta.*

- Searches were performed against the full UniProt Mouse database. Potentially an alternative approach would have been to characterise the IP obtained using Arid1a and build a more targeted database against which to search the data.

*Originally, since the experiments were designed around the BAF complex, we performed searches using a database restricted to BAF subunits and compared them with the full database search. The restricted database search did not result in a substantial improvement in crosslink assignment for BAF complex subunits and therefore the full mouse database was used to search for additional crosslinks.*

- The authors specify how protein concentrations were measured, but don't say how much protein was used for the IP. Additionally, a value of 1mM is given for the cross-linking reagent, how does this relate to a protein: cross-linker ratio? Were several concentration ratios investigated?

*We have now specified that 10-60 mg of nuclear lysate were used for the IP depending on the yield of the protein extraction on the day of the experiment.*

*Protein concentration was not measured after the IP as proteins were crosslinked and digested on the beads. Therefore, we are unable to provide an accurate protein to cross-linker ratio. We also tried 5 mM DSSO concentration (for experiment 5, now clarified in methods). An increase of ~1.8 times in crosslinks was observed. Although it is a substantial improvement, it was not proportional to the 5-fold increase in amount of crosslinker and was therefore deemed to not be cost-effective.*

- On page 11, the authors mention that alternative software is available for data analysis. This list is not extensive and should at least include other applications such as Xi, StavroX, etc.

*We have now added other software for XL data analysis: "As an alternative to Proteome Discoverer 2.4 with XLinkX and Sequest, open-access software such as MaxQuant (Cox and Mann, 2008), pLink (Chen et al., 2019), XQuest/xProphet (Rinner et al., 2008, Walzthoeni et al., 2012, Leitner et al., 2014), StravoX (Götze et al., 2012), MeroX (Götze et al., 2015), Kojak (Hoopmann et al., 2015), XiSEARCH (Mendes et al., 2019) or MaxLinker (Yugandhar et al., 2020) could be used for XL identification."*

- The authors get slightly fixated on the distance restraint of 37Å. This is a lenient value and may not have any particular importance for these disordered proteins.

*We agree that 37 Å is a lax value. We have now relaxed the use of the 37 Å threshold throughout the manuscript. Further examination of our results has revealed that most of Thrap3 XLs satisfy a distance constraint of 32 Å in all models, and we have modified the manuscript accordingly.*

- For the cross-linking data analysis, it may prove useful to include serine and threonine in the reaction specificity for DSSO.

*We thank the reviewer for this suggestion. We did not check for serine and threonine sites as we generally see less than 1% reactivity with these residues when using other reagents with similar reaction chemistry to DSSO such as TMT.*

**References:**

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., WU, L., ZHOU, W.-J., CHI, H., SUN, R.-X., DONG, M.-Q. & HE, S.-M. 2019. A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nature Communications*, 10, 3404.

COX, J. & MANN, M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, 26, 1367-1372.

GÖTZE, M., PETTELKAU, J., FRITZSCHE, R., IHLING, C. H., SCHÄFER, M. & SINZ, A. 2015. Automated Assignment of MS/MS Cleavable Cross-Links in Protein 3D-Structure Analysis. *Journal of the American Society for Mass Spectrometry*, 26, 83-97.

GÖTZE, M., PETTELKAU, J., SCHAKS, S., BOSSE, K., IHLING, C. H., 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.

HOOPMANN, M. R., ZELTER, A., JOHNSON, R. S., RIFFLE, M., MACCOSS, M. J., DAVIS, T. N. & MORITZ, R. L. 2015. Kojak: Efficient Analysis of Chemically Cross-Linked Protein Complexes. *Journal of Proteome Research*, 14, 2190-2198.

LEITNER, A., WALZTHOENI, T. & AEBERSOLD, R. 2014. Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. *Nat Protoc*, 9, 120-37.

MENDES, M. L., FISCHER, L., CHEN, Z. A., BARBON, M., O'REILLY, F. J., GIESE, S. H., BOHLKE-SCHNEIDER, M., BELSOM, A., DAU, T., COMBE, C. W., GRAHAM, M., EISELE, M. R., BAUMEISTER, W., SPECK, C. & RAPPSILBER, J. 2019. An integrated workflow for crosslinking mass spectrometry. *Mol Syst Biol*, 15, e8994.

RINNER, O., SEEBACHER, J., WALZTHOENI, T., MUELLER, L. N., BECK, M., SCHMIDT, A., MUELLER, M. & AEBERSOLD, R. 2008. Identification of cross-linked peptides from large sequence databases. *Nat Methods*, 5, 315-8.

WALZTHOENI, T., CLAASSEN, M., LEITNER, A., HERZOG, F., BOHN, S., FÖRSTER, F., BECK, M. & AEBERSOLD, R. 2012. False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nature Methods*, 9, 901-903.

YUGANDHAR, K., WANG, T.-Y., LEUNG, A. K.-Y., LANZ, M. C., MOTORYKIN, I., LIANG, J., SHAYHIDIN, E. E., SMOLKA, M. B., ZHANG, S. & YU, H. 2020. MaXLinker: Proteome-wide Cross-link Identifications with High Specificity and Sensitivity\*. *Molecular & Cellular Proteomics*, 19, 554-568.

**Competing Interests:** No competing interests were disclosed.

Reviewer Report 27 June 2022

<https://doi.org/10.21956/wellcomeopenres.18954.r50940>

© 2022 Leitner A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Alexander Leitner** 

Institute of Molecular Systems Biology, Department of Biology, ETH Zurich, Zurich, Switzerland

In this contribution, Shcherbakova *et al.* report the characterization of the ternary interaction between mouse Thrap3, Bclaf1, and Erh by chemical cross-linking (XL) and mass spectrometry. The cross-linking data was obtained from an experiment that initially targeted the BAF complex, which did not yield sufficient cross-links for a detailed structural interpretation. Thrap3, Bclaf1, and Erh, however, were found to be connected through a mutually supporting set of cross-links, suggesting that the observed connections specify the binding interfaces in a ternary complex. The manuscript describes data generation from the IP-XL-MS experiment, for which results from multiple, independent experiments were combined. In total, 24 non-redundant peptide pairs involving at least one of the three proteins were identified repeatedly.

Thrap3 and Bclaf1 are predicted to have large unstructured/intrinsically disordered regions, which makes it difficult to study their interactions by established structural biology methods. The XL data presented here serve as a starting point for further experiments, including integrative modeling approaches to define the organization of the complex. The role of Thrap3 and Bclaf1 in some splicing events may make the complex relevant for diseases such as cancer.

The manuscript is clearly written and experiments are described in sufficient detail to enable replication. XL-MS results have been deposited to the PRIDE repository, facilitating reanalysis or reuse of the data.

Minor comments that could be addressed in a revised version:

- On page 7, the authors discuss the potential ambiguity of XL data when it comes to site pairs on the same protein. While this is true, the case described here - a cross-link that connects partially overlapping sequences - needs to be inter-molecular (or a false positive identification). I would emphasize this more in the text.
- Several times throughout the manuscript, the authors mention that they used an upper distance of 37 Å to assess "compatibility" of a cross-link with available PDB structures or structure predictions, citing Liu et al., 2020, as a reference. I would like to point out that 37 Å is a quite lenient threshold, and upper distances of 30-32 Å are much more commonly used. Nevertheless, an interaction between largely unstructured proteins may sample a larger conformational space. Somewhat related to this, the authors discuss that intra-molecular links on Thrap3 were compatible with one of the proposed models (page 7). Since more than one model was predicted, how did the links fit to other models?
- In the methods section (page 10), an "IPP150" buffer is mentioned that needs to be defined.

On the same page, "precursor selection ... with a mass resolution of 120,000" should be rephrased. Precursors were detected at a resolution of 120,000 but their selection for fragmentation is only dependent on the isolation width. "miss cleavages" should read "missed cleavages".

- The legend in Table 1 mentions four experiments, although the results from five are displayed.

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Not applicable

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** cross-linking mass spectrometry; proteomics; structural biology; protein-protein interactions, protein-RNA interactions

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 25 Nov 2022

**Liudmila Shcherbakova**, Institute of Cancer Research, UK, London, UK

- On page 7, the authors discuss the potential ambiguity of XL data when it comes to site pairs on the same protein. While this is true, the case described here - a cross-link that connects partially overlapping sequences - needs to be inter-molecular (or a false positive identification). I would emphasize this more in the text.

*We have now emphasised that the cross-link that connects partially overlapping sequences corresponds to an inter-molecular connection (or a false positive identification).*

- Several times throughout the manuscript, the authors mention that they used an upper distance of 37 Å to assess "compatibility" of a cross-link with available PDB structures or structure predictions, citing Liu et al., 2020, as a reference. I would like

to point out that 37 Å is a quite lenient threshold, and upper distances of 30-32 Å are much more commonly used. Nevertheless, an interaction between largely unstructured proteins may sample a larger conformational space. Somewhat related to this, the authors discuss that intra-molecular links on Thrap3 were compatible with one of the proposed models (page 7). Since more than one model was predicted, how did the links fit to other models?

*We thank the reviewer for his expert comment on the cross-link distance upper threshold.*

*The cross-links fit to the other 4 models was similar, with models 2,3 and 5 satisfying all 5/5 crosslinks at a distance of 32 Å and models 1 and 4 satisfying 4/5 crosslinks at 32 Å. We have now modified the text to include this information.*

- In the methods section (page 10), an "IPP150" buffer is mentioned that needs to be defined. On the same page, "precursor selection ... with a mass resolution of 120,000" should be rephrased. Precursors were detected at a resolution of 120,000 but their selection for fragmentation is only dependent on the isolation width. "miss cleavages" should read "missed cleavages".

*The IPP150 buffer composition (150 NaCl, 10mM Tris-HCl, 0.1% NP40) is now defined in the text.*

*We have rephrased the MS mass resolution and precursor selection parameters to: "MS scans were acquired at a mass resolution of 120,000 and precursors between 375-1,600 m/z and charge equal or higher than +3 were isolated for collision-induced dissociation (CID) fragmentation with quadrupole isolation width 1.6 Th in the top speed mode in cycles of 5 seconds."*

*We have also replaced "miss cleavages" with "missed cleavages".*

- The legend in Table 1 mentions four experiments, although the results from five are displayed

*The legend has been corrected.*

**Competing Interests:** No competing interests were disclosed.