

## ***Selective cross-linking of coinciding protein assemblies by in-gel cross-linking mass-spectrometry***

*Johannes F. Hevler<sup>1,2#</sup>, Marie V. Lukassen<sup>1,2#</sup>, Alfredo Cabrera-Orefice<sup>3</sup>, Susanne Arnold<sup>3</sup>,  
Matti F. Pronker<sup>1,2</sup>, Vojtech Franc<sup>1,2</sup> and Albert J.R. Heck<sup>1,2\*</sup>*

# contributed equally

\*Correspondence: Albert Heck, [a.j.r.heck@uu.nl](mailto:a.j.r.heck@uu.nl); phone: +31302536797

<sup>1</sup> Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands

<sup>2</sup> Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, The Netherlands

<sup>3</sup> Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, 6525 GA Nijmegen, The Netherlands

### **Abstract**

Cross-linking mass spectrometry has developed into an important method to study protein structures and interactions. The in-solution cross-linking workflows involve time and sample consuming steps and do not provide sensible solutions for differentiating cross-links obtained from co-occurring protein oligomers, complexes, or conformers. Here we developed a cross-linking workflow combining blue native PAGE with in-gel cross-linking mass spectrometry (IGX-MS). This workflow circumvents steps, such as buffer exchange and cross-linker concentration optimization. Additionally, IGX-MS enables the parallel analysis of co-occurring protein complexes using only small amounts of sample. Another benefit of IGX-MS observed by experiments on GroEL and purified bovine heart mitochondria, is the substantial reduction of artificial over-length cross-links when compared to in-solution cross-linking. We next used IGX-MS to investigate the complement components C5, C6, and their hetero-dimeric C5b6 complex. The obtained cross-links were used to generate a refined structural model of the complement component C6, resembling C6 in its inactivated state. This finding shows that IGX-MS can be used to provide new insights into the initial stages of the terminal complement pathway.

## Introduction

Over the last decades bimolecular mass spectrometry (MS), with its ability to analyze low amounts of samples with high speed and sensitivity, has evolved into a central pillar beneficial for integrative structural biology (de Souza and Picotti, 2020; Kaur et al., 2019; Lossel et al., 2016; Robinson, 2019). The structural MS toolbox contains multiple complementary approaches. Next to native MS and top-down MS, a variety of peptide-centric MS methods, such as thermal proteome profiling (TPP), limited proteolysis (LiP), hydrogen/deuterium exchange (HDX) MS and chemical cross-linking MS (XL-MS or CLMS), have emerged and enabled structural studies of a wide range of biomolecules (Feng et al., 2014; Heck, 2008; Leitner et al., 2010; Savitski et al., 2014; Zheng et al., 2019). With recent advances in instrumentation, sample preparation, and data analysis, especially XL-MS has started to fulfill its potential to complement well established structural methods such as X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR), and cryo-electron microscopy (cryo-EM) (Leitner et al., 2016; Matthew Allen Bullock et al., 2016; Rappsilber, 2011). XL-MS has a particular utility to capture protein-protein interactions in solution by measuring spatial distance restraints, mirroring structural conformations of intact proteins. Concomitantly, a wide range of chemical cross-linkers have been explored so far, often relying on similar chemical principles (Sinz, 2003; Steigenberger et al., 2020). Most used cross-linkers are small, homo-bifunctional reagents, with two reactive moieties capable of covalently binding two nearby amino acids. The reactive groups are separated by a spacer arm of varying lengths, which can be gas-phase cleavable or non-cleavable, thereby determining different MS data acquisition methods (Kao et al., 2011; Leitner et al., 2010; Muller et al., 2010; Staros, 1982). Recent advances in search engines for more efficient identification of cross-linked peptides allowed structural studies of purified proteins or protein complexes, as well as large scale experiments with more complex samples like purified organelles or cell lysates, using buffer systems which aim to meet physiological relevant conditions (Beveridge et al., 2020; Chen and Rappsilber, 2019; Gotze et al., 2019; Klykov et al., 2018). A typical XL-MS workflow begins with the optimization of the cross-linker concentration. Next, a protein mixture is incubated with the cross-linking reagent, and the reaction is subsequently quenched to prevent the generation of unwanted random protein contacts. After (tryptic) digestion, cross-linked peptides are subjected to various pre-fractionation steps or enrichment strategies, to distinguish them from the vast majority of unmodified peptides. Cross-linked residues are eventually identified using dedicated XL-MS search algorithms, providing structural information in the form of distance

restraints, which can be utilized to guide computational homology modeling, refinement of flexible regions within structural models, protein-protein docking and the generation of protein interaction networks (Albanese et al., 2020; Bullock et al., 2018; Iacobucci et al., 2019; Kim et al., 2018; Ryl et al., 2020). Currently, technological developments in XL-MS aim to further improve the cross-linking reaction efficiency and detection. The research is mainly focused on sample preparation techniques, MS fragmentation and enrichment strategies, data acquisition and analysis of cross-linked peptides, as well as the design of novel cross-linkers (Chen et al., 2019; Dau et al., 2019; Iacobucci et al., 2018; Leitner et al., 2012; Liu et al., 2017; Mendes et al., 2019; Steigenberger et al., 2019).

Although the latest advances significantly revised and reformed the field of XL-MS, some challenges remain. A central problem of XL-MS data analysis is the occurrence of both false-positive and false-negative cross-link hits. Especially, the existence of proteins with highly dynamic/flexible conformations and the presence of co-occurring alike protein complexes (e.g., protein oligomers and co-occurring complexes sharing distinct sub-units), significantly complicate the analysis by current in-solution XL-MS approaches. Interaction specific cross-links are relevant as structural changes can be triggered by the presence of a binding partner or the environment, thereby eventually displaying a physiological relevant protein conformation (de Souza and Picotti, 2020; Feng et al., 2014; Mannige, 2014; Uversky, 2011). Additionally, when using too high concentrations of cross-linker or too high concentrations of proteins, undesired artificial interactions are likely being picked up by XL-MS. In-solution XL-MS experiments, therefore, need careful experimental optimization of, in particular, the concentration of the proteins and the cross-linker.

Here we describe an alternative approach, performing in-gel cross-linking mass spectrometry (IGX-MS), re-discovering the great separation power of gel-electrophoresis. Prior to cross-linking, we load the samples and perform blue native polyacrylamide gel electrophoresis (BN-PAGE), allowing the separation of distinct structural states of the proteins or protein complexes. The distinct bands are subsequently excised and cross-linked in the gel, enabling the measurement of conformation- and interaction-specific cross-links and derived distance restraints (Fig. 1). We show that this IGX-MS workflow has certain advantages compared to the in-solution based XL-MS methods. These include, amongst other things, no need for cross-linker concentration optimization, generation of conformation-specific cross-links, and relatively low sample concentration requirements. Moreover, through IGX-MS data obtained for several protein assemblies, we provide evidence that proteins retain not only their

quarternary but also their secondary and tertiary native structural states under BN-PAGE separation. In proof-of-concept studies, comparing IGX-MS with in-solution XL-MS, targeting the 14-mer *E. Coli* GroEL chaperone and the ATP synthase (complex V) from bovine heart mitochondria (BHM), we show that IGX-MS reduces substantially the number of (potentially false) over-length cross-links. Ultimately, we applied the optimized IGX-MS for investigating structures of the terminal complement proteins C5 and C6, which are involved in the initial steps of the assembly of membrane protein attack (MAC). Our cross-linking data lead us to propose a refined alternative conformation of the complement component C6, providing new insights into the terminal complement pathway. In summary, our data show that BN-PAGE based IGX-MS is a powerful tool, allowing the efficient generation of conformation- and interaction-specific distance constraints, with the potential of refining structural models of a large variety of protein assemblies, even when they co-occur in solution.

## Results

**BN-PAGE forms the basis for IGX-MS.** Blue native polyacrylamide gel electrophoresis, BN-PAGE, has proven to be a robust and sensitive method for the separation of protein complexes from various sample types. It requires only minimal amounts of sample to sensitively estimate native protein molecular weights ( $M_w$ ), respective conformational states, as well as protein-protein interactions. Further, proteins and protein complexes are thought to maintain not only their overall quarternary structural organization in the gel but also their secondary and tertiary structural organization, as they can still after that be subjected to further structural and functional analysis (2D crystallization, cryo-EM, in-gel activity assay) (Poetsch et al., 2000; Schafer et al., 2006; Wittig et al., 2007).

Here, we combine BN-PAGE with XL-MS, to efficiently isolate co-occurring protein oligomeric states and sub-complexes, and subsequently structurally characterize each of them by XL-MS. We first determined whether proteins and protein complexes can be cross-linked in the BN gel. For this, 10  $\mu$ g of purified *E. Coli* GroEL diluted in a cross-linking incompatible Tris buffer was subjected to BN-PAGE as described previously (Wittig et al., 2006). Bands corresponding to the native 14-mer GroEL ( $M_w = 800$  kDa) were excised from the BN-PAGE (SI Fig. 1A), and further, cut into small pieces and incubated with or without the cross-linker reagent DSS. Next, GroEL was extracted from the gel pieces and subsequently loaded onto a reducing SDS-PAGE (SI Fig. 1B). The control lane (no cross-linker) revealed only one distinct band at 57 kDa representing the GroEL monomers. In contrast, the BN-PAGE band that was



incubated with DSS, showed several additional high-molecular-weight bands above 100 kDa, indicating the successful cross-linking of GroEL subunits. The results also indicate that a successful in-gel cross-linking reaction can be performed even though the initial buffer in which GroEL was diluted, is incompatible with the cross-linking reaction.

**IGX-MS optimization is straightforward.** In order to prevent protein precipitation caused by over cross-linking, standard in-solution XL-MS crucially depends on a careful evaluation of the optimal concentration of the cross-linker. For this, a subset of sample needs to be incubated with varying cross-linker concentrations prior to the experiment and subsequently analyzed by gel electrophoresis. The time and sample consuming optimization step led us to investigate the effect of varying cross-linker concentrations for IGX-MS experiments. GroEL was subjected to BN-PAGE, and relevant bands were cross-linked using the two different cross-linking reagent DSS and DSSO in concentrations varying in five steps from 0.5 to 5 mM. After quenching of the cross-linking reaction with Tris, the protein-containing bands were prepared for MS analysis following a standard in-gel digestion procedure. Cross-links obtained for DSS and DSSO, and each concentration were validated by mapping them onto the GroEL structure (PDB ID: 1KP8), and lysine C<sub>α</sub>-C<sub>α</sub> distances were obtained (SI Table 1). The distance distribution for both cross-linkers was highly similar at all used concentrations, and almost no cross-link distances over 30 Å were observed across the varying concentrations, indicating that IGX-MS is highly resistant against over cross-linking of proteins (Fig. 2A). Our data suggest that IGX-MS is also less hampered by unspecific cross-links. We also observed that the total number of unique cross-links was not affected by the concentration of DSS, and only marginally effected for DSSO as the obtained cross-links are slightly lower for concentrations below 2 mM (Fig. 2A). However, no significant difference was detected between 2 mM and 5 mM (Fig. 2A). The cross-linked sites onto the GroEL structure showed good consistency in cross-linked regions for both DSS and DSSO experiments (Fig. 2B, SI Table 1). Finally, we compared the cross-linking results for each cross-linker and concentration across the three replicates, which demonstrated excellent reproducibility of the IGX-MS experiments (SI Fig. 2). Based on these results, a DSS concentration of 1.5 mM (SI Fig. 2A) and a DSSO concentration of 2 mM (SI Fig. 2B) were used for the subsequent experiments.

**Direct comparison of IGX-MS and in-solution XL-MS.** BN-PAGE facilitates the distinction of oligomeric states, but it can also be particularly useful when protein complexes are reconstituted. In such experiments, one or more of the subunits may be (unwillingly) in excess. These preparations can then lead to false cross-link interpretations, as especially intra cross-

links can originate from the free monomer subunit or from the subunit in the complex (which may exhibit another conformation). By comparing IGX-MS and in-solution XL-MS of GroEL, we aimed to assess the relevance of this additional separation aspect and confirm that proteins maintain their native structural integrity during in-gel separation. First, the optimal cross-linker concentration for in-solution experiments was determined by SDS-PAGE to avoid over cross-linking (SI Fig. 3A). Next, the sample for XL-MS was cross-linked with DSS, while IGX-MS was cross-linked with both DSS and DSSO. Subsequent comparison of DSS- and DSSO-in-gel cross-linked samples showed a 53% overlap of identified cross-linked sites. Next, we directly compared XL-MS and IGX-MS. Although a significant number (60 %) of DSS-in-gel cross-links was also detected by XL-MS (Fig. 3A), XL-MS resulted in a seemingly higher total number of unique cross-links compared to IGX-MS. Nevertheless, our detailed evaluation of the identified cross-links confirmed that GroEL maintains an in-solution like structure in the BN-PAGE gel. First, we ruled out that the higher number of unique cross-links in-solution could be explained by insufficient extraction of long peptides from the gel, based on the observation that the detected cross-linked peptides displayed a similar length distribution (SI Fig. 3B). Plotting all the cross-links onto the GroEL structure revealed that a large portion of the "exclusive" in-solution cross-links originated from paired lysine residues separated by more than 30 Å, our distance cut-off. In contrast, virtually all IGX-MS cross-links remained below this cut-off (Fig. 3B, SI Table 1), reflecting a Gaussian distance distribution expected for the used structural model. We are convinced that the BN-PAGE gel separation removes the co-analysis of higher-order protein aggregates. The latter often leads to the observation of over-length cross-links in solution. Further, mapping and directly comparing the overlapped cross-links between the IGX-MS and XL-MS revealed high consistency of cross-linked sites, which further supports that GroEL preserved its native conformation in the gel.

**IGX-MS facilitates the analysis of distinct co-occurring assemblies.** Proteins in cells or extracted from various biological sources can be part of multiple different complexes. Whether a protein is a free monomer or part of one or more protein complexes can have a substantial effect on its structure. The identification of distinctive structural states of a protein by XL-MS in solution is often hampered, especially when the "free" monomer co-exist with the same protein being part of one or more complexes. For this reason, we investigated whether IGX-MS can be used to exclusively obtain cross-links for proteins in a single configuration of the complex. As a first test-sample, we incubated GroEL with one of its known natural unfolded substrates, namely the bacteriophage T4 capsid protein (gp23, 56 kDa). Following incubation

of GroEL with unfolded gp23, we analyzed this sample by BN-PAGE and observed three distinctive bands, corresponding to free GroEL and GroEL with one or two copies of gp23 bound (SI Fig. 4A). That GroEL can bind two substrate molecules (in the *cis* and *trans* ring) agrees with previously reported data (van Duijn et al., 2006) and could be additionally confirmed by relative quantification of the subunits in the respective bands (SI Fig. 4B). In parallel cross-linking of each band, i.e., GroEL, GroEL:gp23 and GroEL:(gp23)<sub>2</sub>, with DSS, revealed interlinks between GroEL and gp23 exclusively in the middle and upper band whereby the primary inter-linked residues in GroEL were identified as K42, K122, and K272 (SI Table 2). The major site of interaction was found to be K272, which is located at the outer edge of the cavity (SI Fig. 4C-D). IGX-MS of each BN-PAGE bands enabled the identification of protein assembly specific distance restraints, which would have been impossible by in-solution XL-MS.

Next, we assessed the capabilities of IGX on a more complex sample and subjected 20 µg of purified bovine heart mitochondria, solubilized with digitonin, to BN-PAGE (Fig. 4A). It is well-known that BN-PAGE can separate and visualize the different complexes of the mitochondrial respiratory chain, including many of the co-occurring super-complexes (Schagger and Pfeiffer, 2000). The band corresponding to the monomeric form of complex V (the well-studied ATP synthase, which can also be abundantly present in a V<sub>2</sub> dimeric form), was excised and subjected to IGX-MS using DSS (Fig. 4A). The detected cross-links were plotted onto the 3D structure (PDB ID: 5ARA) and compared to cross-links detected in a previously published study from our lab, by in-solution XL-MS (Liu et al., 2018) (Fig. 4B). The cross-linked regions detected by IGX-MS and in-solution XL-MS were found to be virtually identical, indicating that also these membrane protein complexes largely retain their -quarternary-tertiary and secondary structures in the BN-PAGE gel. Similar to the previous in-solution XL-MS experiments, only solvent-accessible regions of complex V subunits (which in intact mitochondria are facing the matrix) were detected in the IGX data (Fig. 4C). In-solution XL-MS resulted in a higher number of unique cross-links (248 vs. 53 for IGX). However, like for GroEL, the C<sub>α</sub>-C<sub>α</sub> distance distribution revealed that many in-solution XL-MS cross-links are well above the 30 Å cut-off (149 unique cross-links). The IGX-MS cross-links are predominantly below this set cut-off (only two unique cross-links above), highlighting the accordance of the IGX-MS generated restrains with the previously published structure of monomeric ATPase (Fig. 4D, SI Table 3). We argue that some of these over-length cross-links detected by in-solution XL-MS may originate from crosslinks from in solution co-occurring

dimeric complex V or other ATPase conformations induced upon binding of one (or several) of its many previously identified interactors (Liu et al., 2018; Ryl et al., 2020; Schweppe et al., 2017). In summary, comparing the IGX-MS and in-solution XL-MS cross-links for mitochondrial complex V highlights the capability of IGX-MS to generate sufficient and reliable distance restraints, also for complex samples requiring still only minimal amounts of sample.

### **Structural features of the complement proteins C5 and C6 and how these adapt when complexed into C5b6.**

The terminal pathway of the complement system is mediated by sequentially interacting proteins (a.o. C5 to C9) that undergo various conformational changes in response to interactions with each other and the membrane environment (Bajic et al., 2015; Bayly-Jones et al., 2017; Hadders et al., 2012; Schatz-Jakobsen et al., 2016). In this process, membrane attack complex (MAC) is typically formed on the membrane of bacteria or pathogens, which leads to their elimination. Briefly, C6 binds to C5b, which originates from C5 by proteolytic cleavage. The resulting C5b6 complex binds C7, C8, and C9 sequentially, forming the C5b-9 complex. This complex is assembled onto the bacterial membrane and combines with polymerizing C9 molecules to create a lytic pore termed MAC (Esser, 1994). Several structures of the different components of the terminal pathway have been explored by X-ray crystallography and electron microscopy (EM) (Aleshin et al., 2012; DiScipio et al., 1988; DiScipio and Hugli, 1989; Fredslund et al., 2008; Hadders et al., 2012; Lovelace et al., 2011; Menny et al., 2018). Especially well-resolved structures of monomeric C5, C6, and C5b6 contributed to the understanding of the conformational changes these proteins undergo in forming the C5b6 complex (Aleshin et al., 2012; Fredslund et al., 2008; Hadders et al., 2012). Here, we set out to further investigate these different conformations by applying IGX-MS on monomeric C5, monomeric C6, and the hetero-dimeric C5b6 complex. Therefore, the BN-PAGE bands representing pure monomeric C5, monomeric C6, and C5b6 were subjected to IGX-MS (SI Fig. 5A). Both C5 and complexed C5b showed a Gaussian distribution of lysine  $C_{\alpha}$ - $C_{\alpha}$  distances when identified cross-links were plotted on the respective available structural models, corroborating the existing structural models (PDB ID: 3CU7 and 4A5W) (Fig. 5A-B, SI Table 4). In the C-terminal region of free C5, we detected some cross-links that exceed the distance restraint. This region is known to undergo significant structural rearrangements, as it adopts a more open conformation after conversion to C5b (Fig. 5A, SI Fig. 5B). Likewise, cross-links obtained for monomeric C6 and complexed C6 were plotted onto previously reported structural models (PDB ID: 3T5O and 4A5W) (Fig. 5C, SI Table 4). Cross-links

obtained for C6 when present in C5b6 were in good agreement with the previously published C5b6 structure (PDB ID: 4A5W). In contrast, cross-links obtained for monomeric C6 (PDB ID: 3T5O) did not substantiate with the existing structural model and showed a noticeable bimodal distance distribution (Fig. 5D). Full-length C6 is composed of three thrombospondin (TSP) domains, a membrane attack complex/perforin (MACPF) domain, an LDL-receptor class A (LDL) domain, and an epidermal growth factor-like (EGF) domain. The EGF domain is followed by the C5b-binding domain composed of two complement control protein domains (CCP1 and CCP2) and two C-terminal factor I modules (FIM1 and FIM2), which are connected to the main body through a partially unresolved flexible linker (Fig. 5C). Interestingly, a high number of over-length cross-links ( $> 30 \text{ \AA}$ ) in monomeric C6 were observed between the LDL domain and the MACPF, as well as within the MACPF domain itself. Cross-links exceeding the distance constraint within the MACPF are connecting the previously described autoinhibitory region (AI, residue 480-522) to two  $\sim 50$ -residue helical clusters (CH1, residue 236-288; CH2 residue 363-416) (Fig. 5C, SI Fig. 5C) (Hadders et al., 2012). Secondly, several over-length cross-links were observed located between the FIM2 domain and residues of the LDL and MACPF domain. These over-length cross-links are nearly exclusively detected for monomeric C6 and not for C6 when present in C5b6 (Fig. 5C, SI Fig. 5C).

**Cross-link guided structural refinement of monomeric free C6.** Monomeric C6 displayed an intolerably high number of over-length cross-links suggesting that an alternative conformation of monomeric free C6 may (co-)exist. Based on the identification of the lysine residues involved in these over-length cross-links, such an alternative structure would include re-positioning of the MACPF-, LDL- and the C5b-binding domains. We sought to define a structural model for monomeric C6 in two consecutive modeling steps. Our final refined model (Fig. 6A) retained the characteristic sequence-specific secondary structure elements and is only missing the flexible linker domain (residues 591-619), for which no confident distance restraints were available, likely due to the lack of lysine residues in this region. When comparing the inter-domain rotation angles and domain centroid displacements between our model and the monomeric C6 X-ray structure (PDB ID: 3T5O), it becomes apparent that the main body (TSP1-1-TSP1-2-LDL-MACPF-EGF-TSP1-3) and the C5b-binding region (CCP1-CCP2-FIM1-FIM2) undergo a significant motion to each other (angle of  $32^\circ$  and displacement of  $110.0 \text{ \AA}$  between TSP1-3 and FIM1 – Fig. 6B, SI Table 5). On the other hand, the C5b-binding region moves almost like a single body with small inter-domain angles and displacements (largest angle of  $2^\circ$  and largest displacement of  $1.1 \text{ \AA}$ ). Within the main body,

substantial domain reorientations can also be observed; in particular between EGF and TSP1-3 (angle of 83 ° and displacement of 16.7 Å), between TSP1-1 and TSP1-2 (angle of 42 ° and displacement of 19.7 Å), and between EGF and TSP1-3 (angle of 29 ° and displacement of 16.3 Å) (Fig. 6B, SI Table 5). Further, when comparing the MACPF domain of the IGX-MS driven model to the previously reported structures of monomeric, complexed, and activated C6, noticeable intra-domain differences are obtained (SI Fig. 6A-D). Overall, the MACPF domain is comprised of a central four-stranded  $\beta$ -sheet, an AI region which is dominated by a linchpin helix, and three helical clusters (CH1-3) of which CH1 and CH2 unfold upon C6 activation (SI Fig. 6A). Closer examining the MACPF domain of the monomeric C6 X-ray structure (SI Fig. 6B) revealed a remarkable conformational resemblance with the MACPF domains of complexed C6 (SI Fig. 6C) and activated C6 (SI Fig. 6D). Interestingly, in our cross-linked driven structural model, we were able to obtain a different conformational orientation of the regions within the MACPF domain (SI Fig. 6A). A clear re-positioning of the linchpin helix (part of the AI region), as well as the CH1 and CH2 cluster, can be delineated, when compared to the MACPF of the X-ray structure (SI Fig. 6E). Here, the linchpin helix (red helix) of the AI is tilted towards the central  $\beta$ -sheets and the helical clusters (CH1-3) (SI Fig. 6A). Additionally, the CH1 and CH2 domains are shown to be re-located, with the CH1 domain moved upwards, and the CH2 domain tilted towards the linchpin helix when compared to the MACPF domains of the previously reported monomeric, complexed and activated C6 structures (SI Fig. 6F). Conclusively, the structural rearrangements guided by our IGX-MS data, result in a more closed conformation of the MACPF domain for free monomeric C6. Besides over-length cross-links within the MACPF domain, we detected eight cross-link restraints between the C5b-binding domain (specifically CCP2, FIM1, and FIM2) and LDL- and the MACPF-domain (specifically CH2) (SI Fig. 7A). The domains that are usually involved in the binding interface of C6 and C5b (CCP1 and CCP2 – see Fig. 5C) are in our model predicted to wrap around the MACPF domain, sharing an interaction interface with its CH2 and CH3 cluster (Fig. 6A, SI Fig. 7B). The FIM2 domain forms an interaction interface with residues of the TSP2, LDL, and MACPF domain, thereby locking the C5b-binding domain to the main body of C6 (Fig. 6A-B, SI Fig. 7C, SI Table 6). This observation is in sharp contrast to the reported X-ray structure, in which the C5b-binding domain shows an "elongated" conformation, with no interaction interface between the mentioned domains (Fig. 6B). Further, we generated contact-maps to assess the overlap of cross-link data with the C6 X-ray structure and the IGX-MS driven model. The IGX-MS driven model provides new contact possibilities between the C5b-binding domain and the LDL-, and MACPF domain as well as within the



MACPF domain (Fig. 6C), thereby significantly improving the overlap between the IGX and reported structural data (Fig. 6D, SI Table 4 and 7).

## Discussion

In-solution XL-MS has become a useful tool to study protein structures and protein-protein interactions (Henry et al., 2018; Koukos and Bonvin, 2020; Liu et al., 2014). Even though the technology of XL-MS advanced substantially over the last decade, there are still quite a few challenges left. The current in-solution XL-MS method is time and labor demanding. Further, XL-MS requires a considerable amount of sample, which is necessary for the optimization of the cross-linker concentration and optionally also enrichment of cross-linked peptides. Also, the presence of co-occurring protein oligomers and distinct protein complexes within a sample in solution can complicate the structural analysis. Unfortunately, there is currently no separation method with sufficient sensitivity and resolution power, which could be easily implemented into the standard XL-MS workflows and enable convenient separation of protein complexes in their native state.

Here we introduce IGX-MS, an alternative approach that aims at tackling some of the abovementioned challenges of in-solution XL-MS. Firstly, we demonstrate the feasibility of cross-linking proteins in a BN-PAGE gel environment. We show that the efficiency of the cross-link reaction in-gel is independent of the cross-linker used (DSS and DSSO work equally well) and also concentration-independent. The latter makes the time- and sample consuming cross-linker concentration optimization obsolete. BN-PAGE combined with IGX-MS is very sensitive, requiring only a few micrograms of a protein sample to dissect and study different co-occurring protein complexes. These features do not only reduce sample preparation time as no purification of a protein/ protein complex of interest is needed but also enables the analysis of samples that are only available in minimal quantities (and therefore not susceptible to conventionally used purification methods in XL-MS). Another step, which can be omitted by using IGX-MS, is the buffer exchange of the sample into a cross-link compatible buffer since this exchange occurs already in the gel. The IGX-MS data presented here is hallmarked by high reproducibility, and to a large extent, the observed cross-links agree with those found by parallel in-solution cross-links. We find this an important conclusion, as it supports that proteins in the BN-PAGE gel maintain their native quaternary, tertiary, and secondary structures, as previously suggested (Poetsch et al., 2000; Schafer et al., 2006; Wittig et al., 2007). Compared to in-solution XL-MS, IGX-MS shows a desired significant reduction of

over-length cross-links when plotted on reported structures. A closer investigation of in-solution and in-gel cross-links revealed that IGX-MS generates protein state-specific cross-links with a lower number of undesired cross-linking products compared to in-solution XL-MS. This specificity is mainly achieved due to the option to precisely cross-link a specific oligomeric state or individual protein complex of interest. Compared to in-solution XL-MS, IGX-MS also has some caveats. Not all proteins and proteins assemblies do enter a BN-PAGE gel easily, and some of them do not migrate as well defined bands. Further, the resolving power of commercially available gels can provide a challenge in resolving distinct protein assemblies close in size. However, gels can easily be cast in-house, using gradients optimized for specific mass regions, potentially allowing a fast and easy adaption and modification suitable for specific sample/protein types.

Following proof of concept studies on samples, ranging from GroEL to mitochondrial respiratory chain complexes, we ultimately demonstrate that the distance restraints generated by IGX-MS can also be used to guide computational homology modeling and refine structural models. The reported structure of monomeric free C6 obtained by X-ray crystallography has been suggested to be in a partially activated and extended state (Aleshin et al., 2012). This partial activation was attributed to crystal lattice contacts with neighboring C6 molecules mimicking C5b in the C5b6 hetero-dimer (SI Fig. 8). Our IGX-MS driven refined structural model suggests a more compact structural arrangement for monomeric C6. The linchpin helix of the AI region becomes tilted with an imaginary rotational center around one of the previously identified hinge regions (Aleshin et al., 2012) and more closely located to the central  $\beta$ -sheet as well as the helical clusters (CH1, CH2), which are known to unfold in the activated MACPF complex (Menny et al., 2018). This arrangement indicates that the linchpin helix interacts with the CH1 domain as previously suggested (Aleshin et al., 2012) and also with the central four-stranded  $\beta$ -sheet and the CH2 domain. A closer position of the linchpin helix might hint at an auto-inhibitory function, to tightly control the unfolding of respective regions. Next, the C5b-binding domain can be found wrapped around the LDL- and MACPF domain, stabilizing the more compacted MACPF domain. The presented IGX-MS-driven model points towards a new structural state of the C6 domains that is characteristic for its unbound state. We compared our data to previously published structures of complexed and activated C6 to obtain new insights into the dynamic activation process of C6. Upon binding of C6 to C5b, the C5b-binding domain is no longer wrapped around the LDL-and MACPF domain of C6, offering a binding interface for C5b. Simultaneously, the linchpin helix rotates around the hinge region to the left, and the

CH1 domain moves downwards, bringing the EGF domain of the AI and the CH1 cluster in close proximity. Further, the CH2 cluster opens up by bending slightly to the right, resulting in a more open conformation of the MACPF domain, which culminates in the complete unfolding of CH1 and CH2 to elongated  $\beta$ -sheets in the MAC complex (SI Fig. 6A-E, SI Video 1). This initial unfolding of C6 provides an alternative conformational step in the terminal complement pathway and further suggests that a similar mechanism could be true for C7, which shares a similar domain structure to C6.

Collectively, the work presented here describes a novel methodology termed IGX-MS, which allows the effortless, sample-frugal, and reproducible generation of specific distance restraints by cross-linking, enabling to distinctively analyze co-occurring protein oligomers and protein assemblies.

## Material and Methods

**Materials.** Chemicals and reagents were purchased from Sigma Aldrich (Steinheim, Germany) unless otherwise stated. Acetonitrile (ACN) was purchased from Biosolve (Valkenswaard, The Netherlands). Sequencing grade trypsin was obtained from Promega (Madison, WI). NativePAGE 3-12% Bis-Tris protein gels, NativePAGE Sample Buffer, NativePAGE Running Buffer, NativePAGE Cathode Additive, and NativeMark were purchased from Invitrogen (California, USA). Criterion XT Bis-Tris Precast Gels (4-12%), XT MOPS Running Buffer, and sample buffer were purchased from BioRad (California, USA). DSSO was produced in-house according to a previous protocol (Kao et al., 2011). Oasis HLB 96-well  $\mu$ Elution Plates were purchased from Waters (Massachusetts, USA). GroEL and the major capsid protein gp23 were expressed and purified as previously described. (Quaite-Randall and Joachimiak, 2000; van Duijn et al., 2005; van Duijn et al., 2006) Complement components C5, C6, and C5b6 were purchased from CompTech (Texas, USA). The bovine heart was freshly obtained from a slaughterhouse.

**Separation of proteins using Blue native PAGE (BN-PAGE).** Blue native page analysis was performed according to the manufacturer's protocol and recently published protocols (Wittig et al., 2006). Briefly, proteins were mixed with NativePAGE sample buffer (1x final concentration), and subsequently, 5-20  $\mu$ g of protein sample was loaded onto a Bis-Tris gel (3-12 %). Electrophoresis was started with a dark blue cathode buffer (1x NativePAGE cathode additive) for 30 min at 80 V before the dark blue buffer was changed to light blue cathode buffer (0.1x NativePAGE cathode additive). After change of buffer, the voltage was increased to 120-140 V for additional 2-4 hours. Readily run gels were briefly rinsed with ddH<sub>2</sub>O before gel bands of interest were excised and further cut into smaller pieces under a laminar flow hood. Excised protein bands were stored in Eppendorf tubes for subsequent cross-linking experiments.

**Verification of in-gel cross-linking (IGX) by SDS-PAGE.** Purified GroEL (10  $\mu$ g) in Tris buffer (50 mM Tris-HCL, pH 7.7, 1 mM EDTA, 1 mM DTT, 10% glycerol) was analyzed using BN-PAGE as described above. Next, excised gel bands were incubated in 50  $\mu$ L PBS with or without 1.5 mM DSS for 30 min at room temperature (RT). Then the cross-linking reaction was quenched by addition of Tris to a final concentration of 50 mM for 30 min at RT. Next, the supernatant was removed from the gel pieces, and proteins were extracted in 200  $\mu$ L extraction buffer (50 mM Tris, pH 7.9, 1 mM dithiothreitol (DTT), 150 mM NaCl, 0.1% SDS)

overnight at RT. The gel pieces were separated by centrifugation at 14,000 x g for 2 min, and the supernatant dried to 20  $\mu$ L. The samples were heated to 95 °C for 5 min with 50 mM DTT and sample buffer before loaded onto SDS-PAGE (4-12%). The gel was prepared according to the manufacturer's protocol using MOPS running buffer. After the SDS-PAGE separation finished, the gel was briefly washed with ddH<sub>2</sub>O and subjected to Coomassie brilliant blue staining solution for approximately one hour. Stained SDS gels were de-stained in ddH<sub>2</sub>O, overnight and shaking at RT.

**DSS and DSSO concentration range experiments with GroEL.** BN-PAGE followed by IGX of purified GroEL was performed as described before. Excised gel pieces were incubated with an increasing concentration of DSS and DSSO (0.5, 1, 1.5, 2, and 5 mM) to determine the effect of different cross-linker concentrations. Cross-linking experiments were performed in triplicates. After quenching of cross-linking reactions, the supernatant was removed, and gel pieces were briefly washed using ddH<sub>2</sub>O and subsequently subjected to standard in-gel digestion (Shevchenko et al., 2006). Briefly, gel pieces containing cross-linked proteins were washed, reduced by incubation in reduction buffer (50 mM ammonium bicarbonate (AmBic), 6.5 mM DTT, pH 8.5) for one hour at RT. Reduction buffer was removed, and gel pieces were dehydrated using 100 % ACN. Next, dehydrated gel pieces were subjected to alkylation buffer (50 mM AmBic, 54 mM iodoacetamide (IAA), pH 8.5) for 30 min at RT in the dark. Alkylation buffer was removed, and gel pieces were dehydrated using 100 % ACN. For digestion of cross-linked proteins, dehydrated gel pieces were covered with digestion buffer (3ng/ $\mu$ L Trypsin in 50 mM AmBic, pH 8.5) and pre-incubated for minimum 30 min on ice. Next, excess of digestion buffer was removed, and an equivalent volume of AmBic buffer (50 mM AmBic, pH 8.5) was added to cover the gel pieces, before incubating at 37 °C overnight. Next, the supernatant containing digested peptides was collected, and gel pieces were once again dehydrated using 100 % ACN for 15 min at RT. Resulted supernatant was collected and combined with the previous supernatant. Finally, the samples were completely dried and stored at -80 °C until MS analysis. For MS analysis, cross-linked peptides were resuspended in MS buffer (10 % FA in water) and analyzed as described below.

**LC-MS analysis.** Data for IGX-MS samples was acquired using an UHPLC 1290 system (Agilent Technologies, Santa Clara, CA) coupled on-line to an Orbitrap Fusion or Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific, San Jose, CA). Firstly, peptides were trapped using a 100- $\mu$ m inner diameter 2-cm trap column (packed in-house with ReproSil-Pur C18-AQ, 3 $\mu$ m) prior to separation on an analytical column (50 cm of length, 75  $\mu$ M inner

diameter; packed in-house with Poroshell 120 EC-C18, 2.7  $\mu\text{m}$ ). Trapping of peptides was performed for 5 min in solvent A (0.1 % FA in water) at a flow rate of 0.005 mL/min. DSS cross-linked peptides were subsequently separated as follows: 0-13 % solvent B (0.1 % FA in 80 % v/v ACN) in 10 sec, 13-44 % in 40 min, 44-100 % in 3 min and finally 100 % for 2 min. DSSO cross-linked samples were separated using the following gradient: 0-10 % solvent B in 10 sec, 10-40 % in 40 min, 40-100 % in 3 min, and finally 100 % for 2 min. For each of the gradients, the flow was passively split to approximately 200 nL/min. Mass spectrometers were operated in a data-dependent mode (DDA). For DSS cross-linked peptides, full scan MS spectra from 350-1500 Th were acquired in the Orbitrap at a resolution of 60,000 with the AGC target set to  $1 \times 10^6$  and maximum injection time of 20 ms. For measurements on the Orbitrap Fusion, in-source fragmentation was turned on and set to 15 eV. Cycle time for MS<sup>2</sup> fragmentation scans was set to 3 s. Only peptides with charged states 3-8 were fragmented, and dynamic exclusion properties were set to  $n = 1$ , for a duration of 20 ms. Fragmentation was performed using in a stepped HCD collision energy mode (31.5, 35, 38.5 %) in the ion trap and acquired in the Orbitrap at a resolution of 30,000 after accumulating a target value of  $1 \times 10^5$  with an isolation window of 1.4 Th and maximum injection time of 120 ms. For the acquisition of DSSO cross-linked peptides, full scan MS spectra from 310-1600 Th were acquired in the Orbitrap at a resolution of 120,000 with the AGC target set to  $5 \times 10^5$  and maximum injection time of 50 ms. For the identification of DSSO signature peaks, peptides were fragmented using a fixed CID collision energy (30 %) and MS<sup>2</sup> scan was performed at in the Orbitrap at a resolution of 30,000 after accumulating a target value of  $5 \times 10^4$  ions using an isolation window of 1.6 Th and maximum injection time of 54 ms. For sequencing selected signature peaks, selected ions were fragmented using a fixed HCD collision energy (30 %) in the ion trap MS<sup>3</sup>, with the AGC target set to  $1 \times 10^4$  and maximum injection time of 120 ms.

**In-solution XL-MS of GroEL.** Purified GroEL (10 $\mu\text{g}$ ) was cross-linked using 0-2 mM DSS for 30 min at RT, followed by quenching using a final concentration of 50 mM Tris. Cross-linked samples were analyzed by SDS-PAGE to determine an optimal cross-linker to protein ratio. The optimal DSS concentration (0.75 mM, SI Fig. 2A) was used for cross-linking of 20  $\mu\text{g}$  GroEL (1mg/mL) in triplicates. After quenching of the reactions, protein precipitation was performed by adding three times 50  $\mu\text{L}$  cold acetone and subsequent incubation at -20 °C overnight. Precipitated samples were centrifuged at 12,000 x g for 20 min. After careful removal of the supernatant, the remaining pellet was air-dried until no acetone solution was visible anymore. Pellets were resuspended in 50  $\mu\text{l}$  ABC with 0.33  $\mu\text{g}$  trypsin (1:60) and



incubated with shaking for 4 h at 37 °C. The solubilized pellets were reduced by 5 mM TCEP for 5 min at 95 °C followed by alkylation with 30 mM CAA for 30 min at 37 °C. Digestion was performed overnight by 0.4 µg trypsin (1:50) at 37 °C. The samples were acidified with TFA before desalting using Oasis HLB plate. Finally, the eluent was dried completely and solubilized in 10% FA before MS-analysis.

**Analysis of GroEL bound to unfolded gp23.** The major capsid protein gp23 was unfolded in 8 M urea for one hour at RT. Unfolded gp23 (11.4 µM) was incubated with GroEL (0.8 µM) in Tris buffer with ADP (50 mM Tris, pH 7.5, 50 mM KCl, MgCl<sub>2</sub>, 1 mM ADP) 10 min at RT. The samples were then subjected to BN-PAGE, as previously described. IGX-MS were performed on the three occurring bands, as described earlier, using 1.5 mM DSS.

**Data analysis of GroEL cross-links.** Raw files obtained from IGX-MS of GroEL were analyzed with the Proteome Discoverer (PD) software suite version 2.3 (Thermo Fisher Scientific) with the incorporated XLinkX node for analysis of cross-linked peptides. For DSS data, the non-cleavable cross-link search option was used, while DSSO data was searched by the MS<sup>2</sup>/MS<sup>3</sup> option. A FASTA file containing the GroEL sequence was used for the XlinkX search. For the samples of GroEL complexed with gp23, the FASTA file was supplemented with the sequences of GroEL and gp23. Carbamidomethyl was set as fixed modification and oxidation (M) and acetylation (protein N-term) as variable modifications. Further, standard settings were used. The obtained cross-links were plotted onto the GroEL structure (PDB ID:1KP8) to extract the C $\alpha$ -C $\alpha$  distances using a python script for PyMol. For further cross-link analysis, both intra-chain and inter-chain combinations to neighboring subunits were considered. In the distance histograms, only the shortest combination is represented if several possible combinations existed. When plotted on the structure, several combinations are shown if they are below 30 Å. Cross-link sequence overviews were generated using xiNET.(Combe et al., 2015). Data obtained from IGX-MS of GroEL bound to gp23 was searched in MaxQuant (version 1.6.10.0) to obtain iBAQ (intensity-Based Absolute Quantification) values. Trypsin was set as a digestion enzyme with two allowed missed cleavages. Carbamidomethyl was set as fixed modification and oxidation (M) and acetylation (protein N-term) as variable modifications. The FASTA file used for the search contained sequences of GroEL and gp23.

**Isolation and purification of bovine heart mitochondria (BHM).** The bovine heart was freshly obtained from a slaughterhouse, kept on ice for 1 h and immediately used for mitochondria isolation. All procedures were performed within a cold room and/or maintaining

the material and solutions on ice. The heart (ca. 600 g) was cut into smaller pieces while removing excess of fat and connective tissue. The pieces of cardiac muscle tissue were homogenized in 4 ml/g tissue of ice-cold isolation buffer (250 mM sucrose, 10 mM Tris/HCl pH 7.4, 0.5 mM EDTA and 2 mM phenyl-methane-sulfonyl fluoride) using a blender at low speed for 5 s and at high speed for 1 min. The pH of the homogenate was measured and corrected to 7.4 with 2 M Tris (unadjusted). After a 15 min stirring, the homogenate was centrifuged at 400 x g (20 min; 4°C). The supernatants were filtered through 8 layers of gauze and centrifuged at 7,000 x g (30 min; 4°C). Resulting mitochondria-enriched pellets were resuspended in isolation buffer and again homogenized this time by applying 10 strokes using a Potter-Elvehjem homogenizer. The mitochondrial homogenates were centrifuged at 10,000 x g (20 min; 4°C) and the resulting pellets (crude mitochondria) were resuspended in isolation buffer supplemented with protease inhibitor cocktail (SIGMAFAST™). Protein concentration was determined by the DC protein assay (Bio-Rad). Aliquots were shock-frozen in liquid nitrogen and stored at -80°C. In order to increase the purity of the preparation, crude mitochondria (4 x 15 ml aliquots; ca. 60 mg prot/ml) were thawed on ice, diluted (1:4) with ice-cold washing buffer (250 mM sucrose, 20 mM Tris/HCl pH 7.4, 1 mM EDTA) and centrifuged at 1,000 x g (10 min; 4°C). The supernatants were recovered and centrifuged at 40,000 x g (20 min; 4°C) and each resulting pellet (clean mitochondria) was resuspended in 2 ml washing buffer. Afterwards, mitochondria were loaded onto a two-layer sucrose gradient (1 M/1.5 M) and centrifuged at 60,000 x g (20 min; 4°C). The fractions accumulated at the interphase (pure mitochondria) were carefully recovered and pooled into one tube. After resuspension in 20 ml ice-cold washing buffer, pure mitochondria were centrifuged at 10,000 x g (20 min; 4°C) and finally resuspended in 5 ml ice-cold washing buffer supplemented with protease inhibitor cocktail (SIGMAFAST™). Protein concentration was determined as above described and the aliquots of pure mitochondria were shock-frozen in liquid nitrogen and stored at -80 °C.

**IGX-MS and data analysis of ATP synthase isolated from purified BHM.** Purified bovine heart mitochondria were solubilized with digitonin (9 g/g protein) on ice for 30 min. Subsequently, 20 µg of solubilized mitochondria was analyzed using BN-PAGE. Afterward, a band corresponding to the ATP synthase was excised, and IGX-MS was applied as described above using 1.5 mM DSS. Triplicates were measured, and individual raw files were searched in MaxQuant against the *Bos Taurus* proteome (2019\_08, downloaded from Uniprot) using previously described settings to generate a subtracted library for the cross-link search in PD

using the XlinkX node. Identified cross-links corresponding to the ATP synthase were subsequently extracted and plotted onto the previously published structure (PDB ID: 5ARA). Resulting C $\alpha$ -C $\alpha$  distances were compared to previously published in-solution data for cross-linked mouse heart mitochondria (Liu et al., 2018). An overview of cross-linked subunits for the bovine ATP synthase was generated using the "circlize" package for R (Gu et al., 2014).

**IGX-MS and data analysis of complement proteins.** Complement components C5 (5  $\mu$ g), C6 (5  $\mu$ g), and C5b6 (10  $\mu$ g) were subjected to BN-PAGE followed by IGX-MS as previously described using 1.5 mM DSS. Experiments were done in triplicates. The resulting raw files from the MS-analysis were searched in MaxQuant (version 1.6.10.0) to generate libraries for the C5, C6, and C5b6 bands. The data was searched against the reviewed *Homo Sapiens* Uniprot database (2019\_08, downloaded from UniProt). Trypsin was set as a digestion enzyme with two allowed missed cleavages. Carbamidomethyl was set as fixed modification and oxidation (M) and acetylation (protein N-term) as variable modifications. The data was then searched using PD, as described earlier. Mannosylation of tryptophan residues was added as a variable modification, and MaxQuant generated libraries used in the XLinkX search. Only cross-links observed in two out of the three replicates were included for further analysis. The cross-links were plotted onto the respective structures using PyMol to obtain C $\alpha$ -C $\alpha$  distances. Cross-link sequence overviews were generated using xiNET.(Combe et al., 2015)

**Modeling of an alternative structure of free complement C6.** Cross-links derived for free C6, together with additional structural constraints derived from Uniprot (disulfide-bond information, secondary structure elements; Uniprot Accession: P13671) were used to predict an alternative structural model. Briefly, the modeling process was divided into two consecutive steps. First, an I-Tasser homology model of C6 based on the previously published structure (PDB ID: 3T5O) was generated to resolve missing residues (Yang et al., 2015). Next, a flexible linker region (residue 591-619) and the C5b-binding domain (residues 620-913) were removed from the generated C6 model. Subsequently, regions with a high density of cross-linked residues were removed from the shortened C6 structure, producing a "core-template" for comparative modeling using Modeller 9.24 (Webb and Sali, 2016). Excised regions were provided as additional templates (SI Table 8) to support the modeling process together with the cross-linking restraints (mean = 17 Å, stdev =2) obtained for individual residues (1-590) as well as the secondary structure information which was obtained from Uniprot (Uniprot ID: P13671). In total, 20 cross-linked guided models for free C6 were generated, each first optimized with the variable target function method (VTFM) and afterwards refined using

molecular dynamics (MD) optimization (Sali and Blundell, 1993). For each model, a DOPE score and a GA341 score was calculated to further validate the quality of produced models (John and Sali, 2003; Melo et al., 2002; Shen and Sali, 2006). Additionally, contact maps for each one of the 20 models were generated, and a CM score (Schweppe et al., 2016) was calculated, indicating the overlap of the cross-linking data with the respective contact maps using the XLmap package in R (Schweppe et al., 2016) (SI Table 9). The model satisfying both scores the best (DOPE and CM score) was chosen for the second modeling process, to generate a full-length model of C6 using detected cross-links between C6 (residues 1-590) and the C5b-binding domain (residues 620-913). The structural assembly of both was achieved by predicting an interaction interface by DisVis (van Zundert and Bonvin, 2015) using respective cross-links and solvent-accessible residues as input parameters. Solvent accessible residues were identified using the standalone program Naccess (© S. Hubbard and J. Thornton 1992-6). Residues with relative solvent accessibility  $\geq 40\%$  were used as solvent-accessible residues. Finally, information-driven docking with HADDOCK (Karaca and Bonvin, 2011; van Zundert et al., 2016) with the validated cross-links and the identified active residues was performed, resulting in four distinct clusters. The structure showing the best agreement with the distance restraints used for the docking process and with the best Haddock score was chosen as final model (SI Table 10). Additionally, residues participating in a binding interface between the C6 (residues 1-590) and the C5b-binding domain were predicted using the Prodigy webserver (SI Table 8). For final model validation, all IGX-MS derived cross-links obtained for C6 were plotted onto the model, and distances for respective links were compared to the Xray structure of C6 (PDB ID 3T5O) (SI Table 7).

**Characterization of inter-domain rotation angles.** For the comparison of our IGX-MS driven model with the crystal structure of C6 in isolation (PDB 3T5O), inter-domain rotation angles and centroid displacements were determined by sequentially superposing the domains (with indicated boundaries) of the crystal structure of C6 onto the corresponding domains of our model, using the program Superpose (Krissinel and Henrick, 2004), part of the CCP4 suite (Winn et al., 2011). Superposition was based on  $C_{\alpha}$  atoms of indicated domains. To further compare the protein conformations, a distance map was generated using the Bio3D package for R (Grant et al., 2006). For this, coordinates of superimposed MACPF domains (residue 155-501) extracted from the crystal structure, and our model were provided as input.

**Data availability.** The mass spectrometry data from this publication have been deposited to the ProteomeXchange partner PRIDE database) and assigned the identifier PXD020014 (Vizcaino et al., 2016) (**Reviewer account details:** Username: [reviewer47172@ebi.ac.uk](mailto:reviewer47172@ebi.ac.uk); Password: YXSjuIrx)

### **Acknowledgments**

All authors acknowledge support from the Netherlands Organization for Scientific Research (NWO) funding the Netherlands Proteomics Centre through the X-omics Road Map program (project 184.034.019) and the EU Horizon 2020 program INFRAIA project Epic-XS (Project 823839). MVL thanks Independent Research Fund Denmark (project 9036-00007B).

### **Author contributions**

JFH and AJRH conceptualized the study. JFH and MVL designed the methodology, performed experiments, and analyzed the data. ACO and SA provided the mitochondrial samples. MFP performed the characterization of inter-domain rotation angles. JFH and MVL wrote the original draft. JFH, MVL, MFB, ACO, SA, VF, AJRH carefully revised and edited the manuscript before submission. MVL and AJRH acquired funding and resources. AJRH supervised the project.

### **Conflict of interest**

The authors do not declare any conflict of interest

## References

- Albanese, P., Tamara, S., Saracco, G., Scheltema, R.A., and Pagliano, C. (2020). How paired PSII-LHCII supercomplexes mediate the stacking of plant thylakoid membranes unveiled by structural mass-spectrometry. *Nat Commun* *11*, 1361.
- Aleshin, A.E., Schraufstatter, I.U., Stec, B., Bankston, L.A., Liddington, R.C., and DiScipio, R.G. (2012). Structure of complement C6 suggests a mechanism for initiation and unidirectional, sequential assembly of membrane attack complex (MAC). *J Biol Chem* *287*, 10210-10222.
- Bajic, G., Degen, S.E., Thiel, S., and Andersen, G.R. (2015). Complement activation, regulation, and molecular basis for complement-related diseases. *EMBO J* *34*, 2735-2757.
- Bayly-Jones, C., Bubeck, D., and Dunstone, M.A. (2017). The mystery behind membrane insertion: a review of the complement membrane attack complex. *Philos Trans R Soc Lond B Biol Sci* *372*.
- Beveridge, R., Stadlmann, J., Penninger, J.M., and Mechtler, K. (2020). A synthetic peptide library for benchmarking crosslinking-mass spectrometry search engines for proteins and protein complexes. *Nat Commun* *11*, 742.
- Bullock, J.M.A., Sen, N., Thalassinou, K., and Topf, M. (2018). Modeling Protein Complexes Using Restraints from Crosslinking Mass Spectrometry. *Structure* *26*, 1015-1024 e1012.
- Chen, Z.A., and Rappsilber, J. (2019). Quantitative cross-linking/mass spectrometry to elucidate structural changes in proteins and their complexes. *Nat Protoc* *14*, 171-201.
- Chen, Z.L., Meng, J.M., Cao, Y., Yin, J.L., Fang, R.Q., Fan, S.B., Liu, C., Zeng, W.F., Ding, Y.H., Tan, D., *et al.* (2019). A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nat Commun* *10*, 3404.
- Combe, C.W., Fischer, L., and Rappsilber, J. (2015). xiNET: Cross-link Network Maps With Residue Resolution. *Molecular & Cellular Proteomics* *14*, 1137-1147.
- Dau, T., Gupta, K., Berger, I., and Rappsilber, J. (2019). Sequential Digestion with Trypsin and Elastase in Cross-Linking Mass Spectrometry. *Anal Chem* *91*, 4472-4478.
- de Souza, N., and Picotti, P. (2020). Mass spectrometry analysis of the structural proteome. *Curr Opin Struct Biol* *60*, 57-65.
- DiScipio, R.G., Chakravarti, D.N., Muller-Eberhard, H.J., and Fey, G.H. (1988). The structure of human complement component C7 and the C5b-7 complex. *J Biol Chem* *263*, 549-560.
- DiScipio, R.G., and Hugli, T.E. (1989). The molecular architecture of human complement component C6. *J Biol Chem* *264*, 16197-16206.
- Esser, A.F. (1994). The membrane attack complex of complement. Assembly, structure and cytotoxic activity. *Toxicology* *87*, 229-247.
- Feng, Y., De Franceschi, G., Kahraman, A., Soste, M., Melnik, A., Boersema, P.J., de Laureto, P.P., Nikolaev, Y., Oliveira, A.P., and Picotti, P. (2014). Global analysis of protein structural changes in complex proteomes. *Nat Biotechnol* *32*, 1036-1044.
- Fredslund, F., Laursen, N.S., Roversi, P., Jenner, L., Oliveira, C.L., Pedersen, J.S., Nunn, M.A., Lea, S.M., DiScipio, R., Sottrup-Jensen, L., *et al.* (2008). Structure of and influence of a tick complement inhibitor on human complement component 5. *Nat Immunol* *9*, 753-760.
- Gotze, M., Iacobucci, C., Ihling, C.H., and Sinz, A. (2019). A Simple Cross-Linking/Mass Spectrometry Workflow for Studying System-wide Protein Interactions. *Anal Chem* *91*, 10236-10244.
- Grant, B.J., Rodrigues, A.P., ElSawy, K.M., McCammon, J.A., and Caves, L.S. (2006). Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* *22*, 2695-2696.



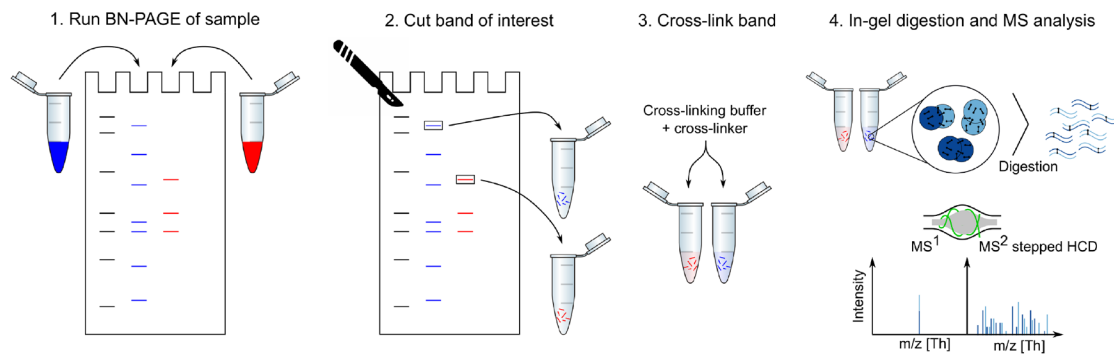
- Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize Implements and enhances circular visualization in R. *Bioinformatics* *30*, 2811-2812.
- Hadders, M.A., Bubeck, D., Roversi, P., Hakobyan, S., Forneris, F., Morgan, B.P., Pangburn, M.K., Llorca, O., Lea, S.M., and Gros, P. (2012). Assembly and regulation of the membrane attack complex based on structures of C5b6 and sC5b9. *Cell Rep* *1*, 200-207.
- Heck, A.J. (2008). Native mass spectrometry: a bridge between interactomics and structural biology. *Nat Methods* *5*, 927-933.
- Henry, N., Krammer, E.M., Stengel, F., Adams, Q., Van Liefferinge, F., Hubin, E., Chaves, R., Efremov, R., Aebersold, R., Vandenbussche, G., *et al.* (2018). Lipidated apolipoprotein E4 structure and its receptor binding mechanism determined by a combined cross-linking coupled to mass spectrometry and molecular dynamics approach. *PLoS Comput Biol* *14*, e1006165.
- Iacobucci, C., Gotze, M., Ihling, C.H., Piotrowski, C., Arlt, C., Schafer, M., Hage, C., Schmidt, R., and Sinz, A. (2018). A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein-protein interactions. *Nat Protoc* *13*, 2864-2889.
- Iacobucci, C., Gotze, M., and Sinz, A. (2019). Cross-linking/mass spectrometry to get a closer view on protein interaction networks. *Curr Opin Biotechnol* *63*, 48-53.
- John, B., and Sali, A. (2003). Comparative protein structure modeling by iterative alignment, model building and model assessment. *Nucleic Acids Res* *31*, 3982-3992.
- Kao, A., Chiu, C.L., Vellucci, D., Yang, Y., Patel, V.R., Guan, S., Randall, A., Baldi, P., Rychnovsky, S.D., and Huang, L. (2011). Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes. *Mol Cell Proteomics* *10*, M110 002212.
- Karaca, E., and Bonvin, A.M. (2011). A multidomain flexible docking approach to deal with large conformational changes in the modeling of biomolecular complexes. *Structure* *19*, 555-565.
- Kaur, U., Johnson, D.T., Chea, E.E., Deredge, D.J., Espino, J.A., and Jones, L.M. (2019). Evolution of Structural Biology through the Lens of Mass Spectrometry. *Anal Chem* *91*, 142-155.
- Kim, S.J., Fernandez-Martinez, J., Nudelman, I., Shi, Y., Zhang, W., Raveh, B., Herricks, T., Slaughter, B.D., Hogan, J.A., Upla, P., *et al.* (2018). Integrative structure and functional anatomy of a nuclear pore complex. *Nature* *555*, 475-482.
- Klykov, O., Steigenberger, B., Pektas, S., Fasci, D., Heck, A.J.R., and Scheltema, R.A. (2018). Efficient and robust proteome-wide approaches for cross-linking mass spectrometry. *Nat Protoc* *13*, 2964-2990.
- Koukos, P.I., and Bonvin, A. (2020). Integrative Modelling of Biomolecular Complexes. *J Mol Biol* *432*, 2861-2881.
- Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* *60*, 2256-2268.
- Leitner, A., Faini, M., Stengel, F., and Aebersold, R. (2016). Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. *Trends Biochem Sci* *41*, 20-32.
- Leitner, A., Reischl, R., Walzthoeni, T., Herzog, F., Bohn, S., Forster, F., and Aebersold, R. (2012). Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Mol Cell Proteomics* *11*, M111 014126.
- Leitner, A., Walzthoeni, T., Kahraman, A., Herzog, F., Rinner, O., Beck, M., and Aebersold, R. (2010). Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol Cell Proteomics* *9*, 1634-1649.

- Liu, F., Lossl, P., Rabbitts, B.M., Balaban, R.S., and Heck, A.J.R. (2018). The interactome of intact mitochondria by cross-linking mass spectrometry provides evidence for coexisting respiratory supercomplexes. *Mol Cell Proteomics* *17*, 216-232.
- Liu, F., Lossl, P., Scheltema, R., Viner, R., and Heck, A.J.R. (2017). Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. *Nat Commun* *8*, 15473.
- Liu, Z., Szarecka, A., Yonkunas, M., Speranskiy, K., Kurnikova, M., and Cascio, M. (2014). Crosslinking constraints and computational models as complementary tools in modeling the extracellular domain of the glycine receptor. *PLoS One* *9*, e102571.
- Lossl, P., van de Waterbeemd, M., and Heck, A.J. (2016). The diverse and expanding role of mass spectrometry in structural and molecular biology. *EMBO J* *35*, 2634-2657.
- Lovelace, L.L., Cooper, C.L., Sodetz, J.M., and Lebioda, L. (2011). Structure of human C8 protein provides mechanistic insight into membrane pore formation by complement. *J Biol Chem* *286*, 17585-17592.
- Mannige, R.V. (2014). Dynamic New World: Refining Our View of Protein Structure, Function and Evolution. *Proteomes* *2*, 128-153.
- Matthew Allen Bullock, J., Schwab, J., Thalassinos, K., and Topf, M. (2016). The Importance of Non-accessible Crosslinks and Solvent Accessible Surface Distance in Modeling Proteins with Restraints From Crosslinking Mass Spectrometry. *Mol Cell Proteomics* *15*, 2491-2500.
- Melo, F., Sanchez, R., and Sali, A. (2002). Statistical potentials for fold assessment. *Protein Sci* *11*, 430-448.
- 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., *et al.* (2019). An integrated workflow for crosslinking mass spectrometry. *Mol Syst Biol* *15*, e8994.
- Menny, A., Serna, M., Boyd, C.M., Gardner, S., Joseph, A.P., Morgan, B.P., Topf, M., Brooks, N.J., and Bubeck, D. (2018). CryoEM reveals how the complement membrane attack complex ruptures lipid bilayers. *Nat Commun* *9*, 5316.
- Muller, M.Q., Dreiocker, F., Ihling, C.H., Schafer, M., and Sinz, A. (2010). Cleavable cross-linker for protein structure analysis: reliable identification of cross-linking products by tandem MS. *Anal Chem* *82*, 6958-6968.
- Poetsch, A., Neff, D., Seelert, H., Schagger, H., and Dencher, N.A. (2000). Dye removal, catalytic activity and 2D crystallization of chloroplast H(+)-ATP synthase purified by blue native electrophoresis. *Biochim Biophys Acta* *1466*, 339-349.
- Quaite-Randall, E., and Joachimiak, A. (2000). Purification of GroEL from an overproducing *E. coli* strain. *Methods Mol Biol* *140*, 29-39.
- Rappsilber, J. (2011). The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J Struct Biol* *173*, 530-540.
- Robinson, C.V. (2019). Mass spectrometry: From plasma proteins to mitochondrial membranes. *Proc Natl Acad Sci U S A* *116*, 2814-2820.
- Ryl, P.S.J., Bohlke-Schneider, M., Lenz, S., Fischer, L., Budzinski, L., Stuiver, M., Mendes, M.M.L., Sinn, L., O'Reilly, F.J., and Rappsilber, J. (2020). In Situ Structural Restraints from Cross-Linking Mass Spectrometry in Human Mitochondria. *J Proteome Res* *19*, 327-336.
- Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* *234*, 779-815.
- Savitski, M.M., Reinhard, F.B., Franken, H., Werner, T., Savitski, M.F., Eberhard, D., Martinez Molina, D., Jafari, R., Dovega, R.B., Klaeger, S., *et al.* (2014). Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* *346*, 1255784.

- Schafer, E., Seelert, H., Reifschneider, N.H., Krause, F., Dencher, N.A., and Vonck, J. (2006). Architecture of active mammalian respiratory chain supercomplexes. *J Biol Chem* *281*, 15370-15375.
- Schagger, H., and Pfeiffer, K. (2000). Supercomplexes in the respiratory chains of yeast and mammalian mitochondria. *EMBO J* *19*, 1777-1783.
- Schatz-Jakobsen, J.A., Pedersen, D.V., and Andersen, G.R. (2016). Structural insight into proteolytic activation and regulation of the complement system. *Immunol Rev* *274*, 59-73.
- Schweppe, D.K., Chavez, J.D., and Bruce, J.E. (2016). XLmap: an R package to visualize and score protein structure models based on sites of protein cross-linking. *Bioinformatics* *32*, 306-308.
- Schweppe, D.K., Chavez, J.D., Lee, C.F., Caudal, A., Kruse, S.E., Stuppard, R., Marcinek, D.J., Shadel, G.S., Tian, R., and Bruce, J.E. (2017). Mitochondrial protein interactome elucidated by chemical cross-linking mass spectrometry. *Proc Natl Acad Sci U S A* *114*, 1732-1737.
- Shen, M.Y., and Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein Sci* *15*, 2507-2524.
- Shevchenko, A., Tomas, H., Havlis, J., Olsen, J.V., and Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* *1*, 2856-2860.
- Sinz, A. (2003). Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. *J Mass Spectrom* *38*, 1225-1237.
- Staros, J.V. (1982). N-hydroxysulfosuccinimide active esters: bis(N-hydroxysulfosuccinimide) esters of two dicarboxylic acids are hydrophilic, membrane-impermeant, protein cross-linkers. *Biochemistry* *21*, 3950-3955.
- Steigenberger, B., Albanese, P., Heck, A.J.R., and Scheltema, R.A. (2020). To Cleave or Not To Cleave in XL-MS? *J Am Soc Mass Spectrom* *31*, 196-206.
- Steigenberger, B., Pieters, R.J., Heck, A.J.R., and Scheltema, R.A. (2019). PhoX: An IMAC-Enrichable Cross-Linking Reagent. *ACS Cent Sci* *5*, 1514-1522.
- Uversky, V.N. (2011). Intrinsically disordered proteins from A to Z. *Int J Biochem Cell Biol* *43*, 1090-1103.
- van Duijn, E., Bakkes, P.J., Heeren, R.M., van den Heuvel, R.H., van Heerikhuizen, H., van der Vies, S.M., and Heck, A.J. (2005). Monitoring macromolecular complexes involved in the chaperonin-assisted protein folding cycle by mass spectrometry. *Nat Methods* *2*, 371-376.
- van Duijn, E., Simmons, D.A., van den Heuvel, R.H., Bakkes, P.J., van Heerikhuizen, H., Heeren, R.M., Robinson, C.V., van der Vies, S.M., and Heck, A.J. (2006). Tandem mass spectrometry of intact GroEL-substrate complexes reveals substrate-specific conformational changes in the trans ring. *J Am Chem Soc* *128*, 4694-4702.
- van Zundert, G.C., and Bonvin, A.M. (2015). DisVis: quantifying and visualizing accessible interaction space of distance-restrained biomolecular complexes. *Bioinformatics* *31*, 3222-3224.
- van Zundert, G.C.P., Rodrigues, J., Trellet, M., Schmitz, C., Kastiris, P.L., Karaca, E., Melquiond, A.S.J., van Dijk, M., de Vries, S.J., and Bonvin, A. (2016). The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol Biol* *428*, 720-725.
- Vizcaino, J.A., Csordas, A., Del-Toro, N., Dianes, J.A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., *et al.* (2016). 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res* *44*, 11033.
- Webb, B., and Sali, A. (2016). Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Protein Sci* *86*, 29.1-29.37.

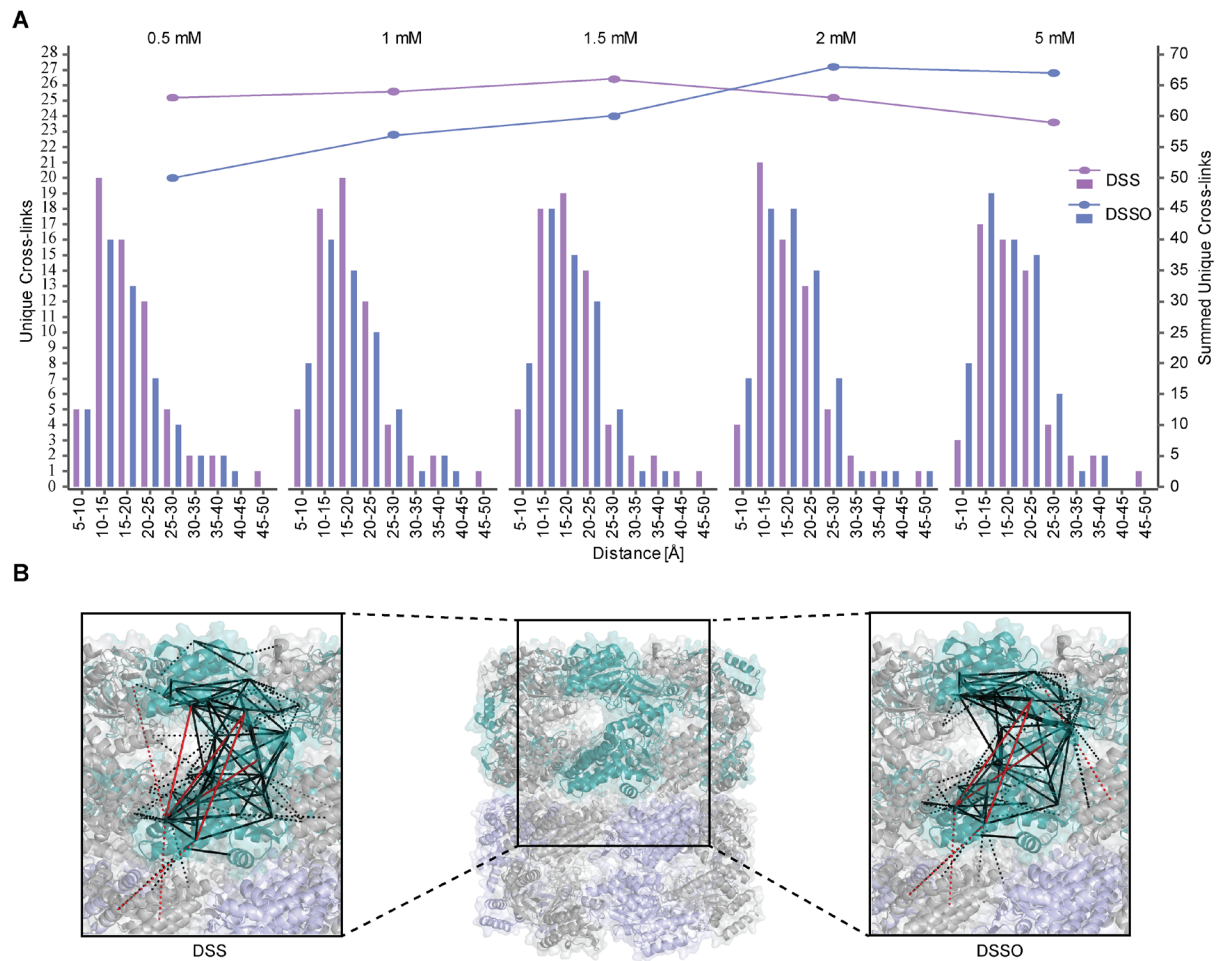
- Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., *et al.* (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* *67*, 235-242.
- Wittig, I., Beckhaus, T., Wumaier, Z., Karas, M., and Schagger, H. (2010). Mass estimation of native proteins by blue native electrophoresis: principles and practical hints. *Mol Cell Proteomics* *9*, 2149-2161.
- Wittig, I., Braun, H.P., and Schagger, H. (2006). Blue native PAGE. *Nat Protoc* *1*, 418-428.
- Wittig, I., Karas, M., and Schagger, H. (2007). High resolution clear native electrophoresis for in-gel functional assays and fluorescence studies of membrane protein complexes. *Mol Cell Proteomics* *6*, 1215-1225.
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nat Methods* *12*, 7-8.
- Zheng, J., Strutzenberg, T., Pascal, B.D., and Griffin, P.R. (2019). Protein dynamics and conformational changes explored by hydrogen/deuterium exchange mass spectrometry. *Curr Opin Struct Biol* *58*, 305-313.

## Main Text Figures



**Figure 1 – Combining BN-PAGE with IGX-MS.**

In the first step, proteins and their assemblies are loaded and separated by BN-PAGE. The bands representing different protein assemblies are visualized after running the gel by Coomassie in the upper running buffer. The band(s) of interest are then excised from the gel and incubated with a cross-linking reagent in the cross-linking buffer. The cross-linking reaction is quenched and subsequently subjected to standard in-gel digestion. The extracted peptides are finally analyzed using cross-linking optimized parameters for the MS analysis.

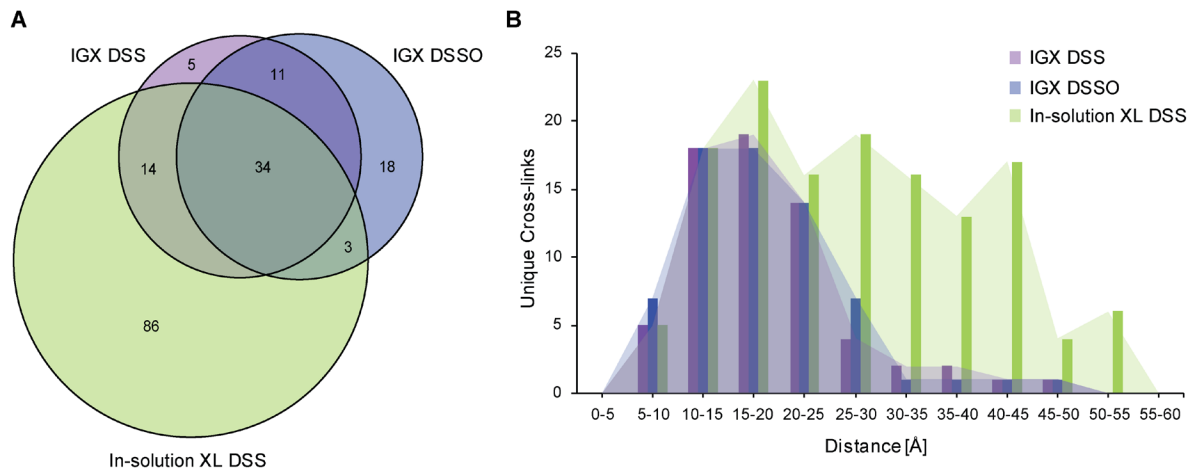


**Figure 2 - IGX-MS of GroEL using either DSS or DSSO at varying concentrations.**

- A. *E. Coli* GroEL was cross-linked by IGX-MS using either DSS or DSSO at concentrations ranging from 0.5 to 5 mM. The identified cross-links were placed on the reported GroEL structure (PDB ID:1KP8), whereby the number of unique cross-linked lysine  $C_{\alpha}$ - $C_{\alpha}$  distances (using the left y-axis) were binned and as shown in bars. The summed number of unique cross-links obtained at each concentration is shown in lines using the right y-axis.
- B. Cross-links obtained by IGX-MS using DSS or DSSO plotted onto one subunit of GroEL (intra-links; solid lines) (PDB ID:1KP8) and neighboring subunits (inter-links; dashed lines). Cross-links agreeing with the set distance restraint of 30 Å are colored black, links exceeding the restraint are colored red.

Data information: The presented data is summed from triplicates.

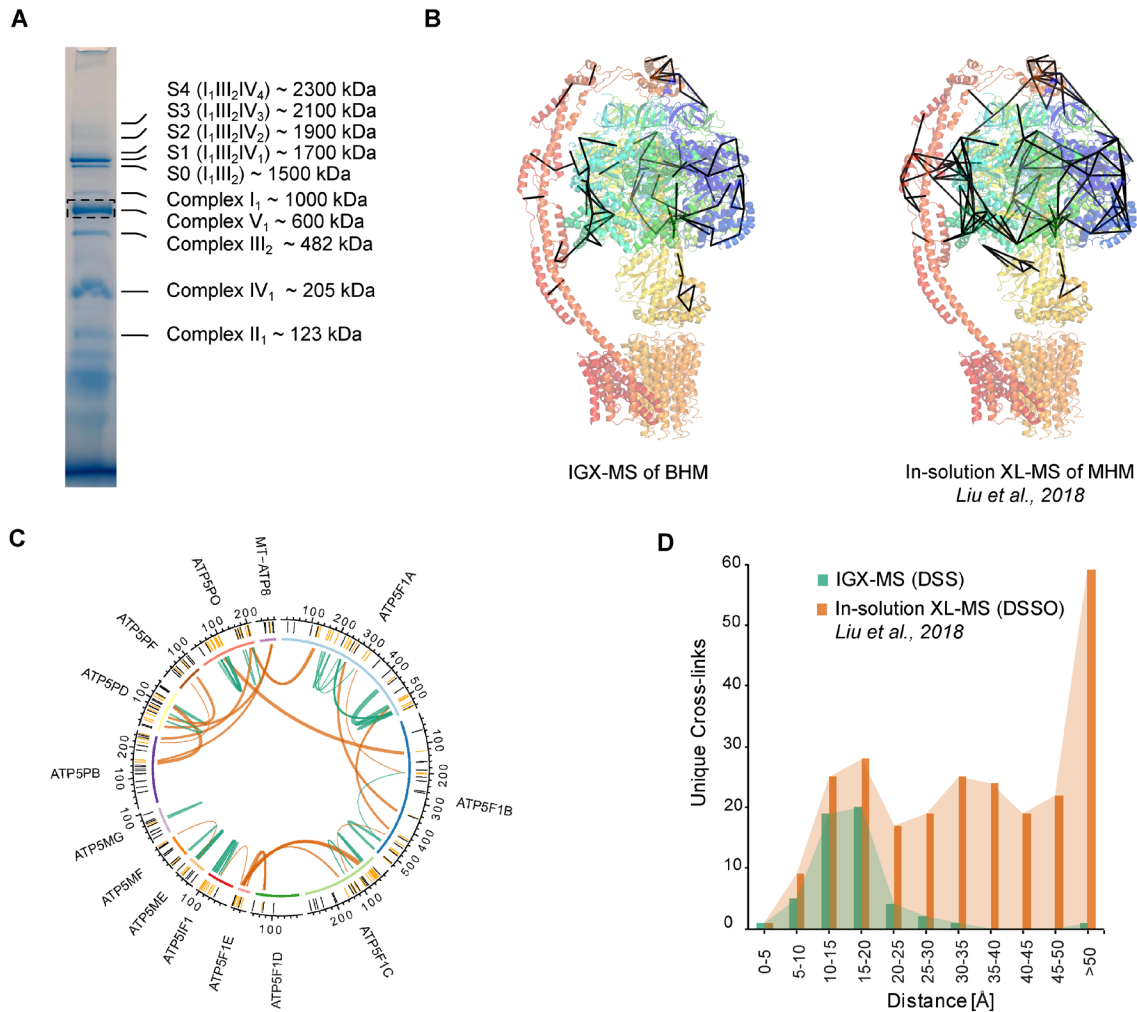




**Figure 3 - Comparison of IGX-MS and in-solution XL-MS of GroEL.**

- A. *E.Coli* GroEL was subjected to in-solution XL-MS using 0.75 mM DSS. The resulting cross-links were compared to cross-links obtained by IGX-MS using DSS (1.5 mM) or DSSO (2 mM). The Venn diagram shows the overlap in the cross-links identified.
- B. Distribution of lysine  $C_{\alpha}$ - $C_{\alpha}$  distances of unique cross-links identified by IGX-MS (DSS or DSSO) or in-solution XL-MS plotted on the GroEL structure (PDB ID:1KP8).

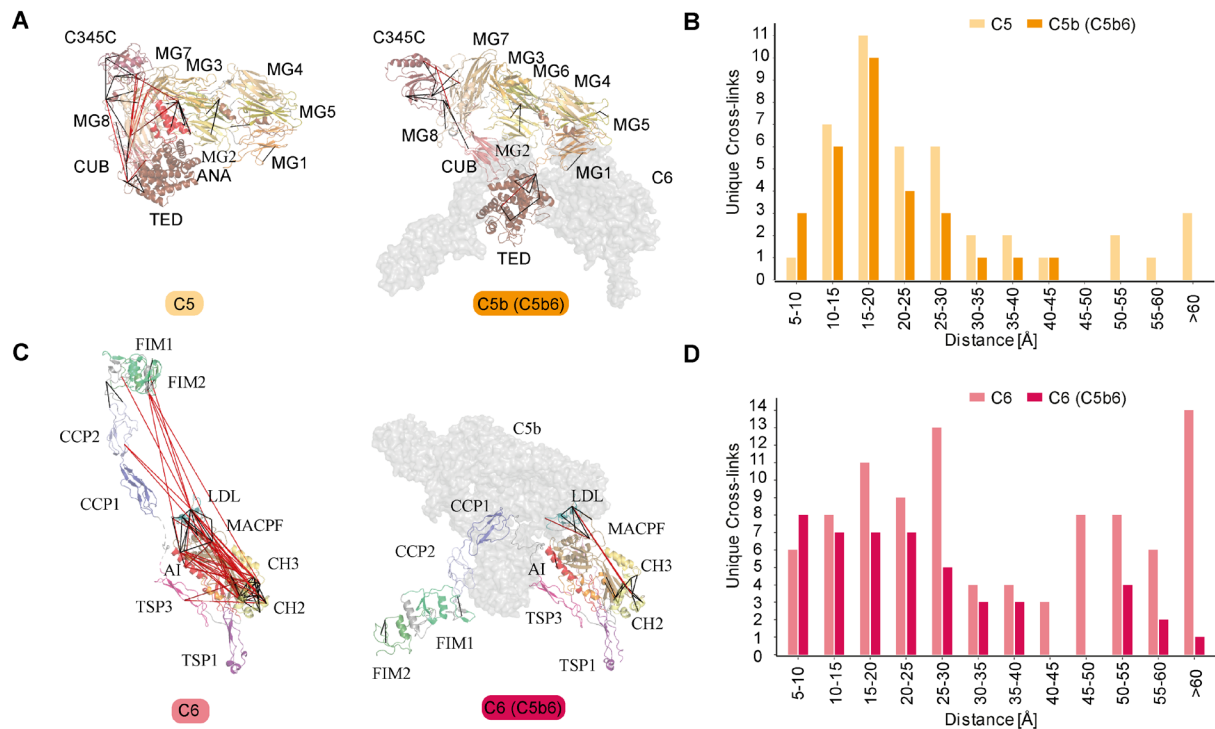
Data information: The presented data is summed from triplicates.



**Figure 4 - Selective cross-linking of the ATP synthase monomer from bovine heart mitochondria (BHM) by IGX-MS.**

- BN-PAGE of 20  $\mu$ g solubilized BHM with several of the distinct complexes annotated. The dashed box indicates the monomeric ATP synthase (complex V) band subjected to IGX-MS. Annotation of protein bands was done in accordance to a previously published study (Wittig et al., 2010).
- Unique cross-links < 30 Å identified by IGX-MS in this study (left) or by Liu et al., 2018 (Liu et al., 2018) using in-solution XL-MS plotted on the ATP synthase structure (PDB ID:5ARA).
- Circos plot of cross-links identified by IGX-MS of ATP synthase. The position of the lysine residues is shown in the outer-ring, and cross-linked residues are colored dark orange. Green lines represent intra-link while inter-links are colored orange. Thickness of the cross-link lines correlates to the number of detected cross-linked spectra matches (CSMs).
- Distribution of lysine C $\alpha$ -C $\alpha$  distances of unique cross-links identified by IGX-MS or in-solution XL-MS (Liu et al., 2018) using DSS (1.5 mM) or DSSO (0.5 mM), respectively.

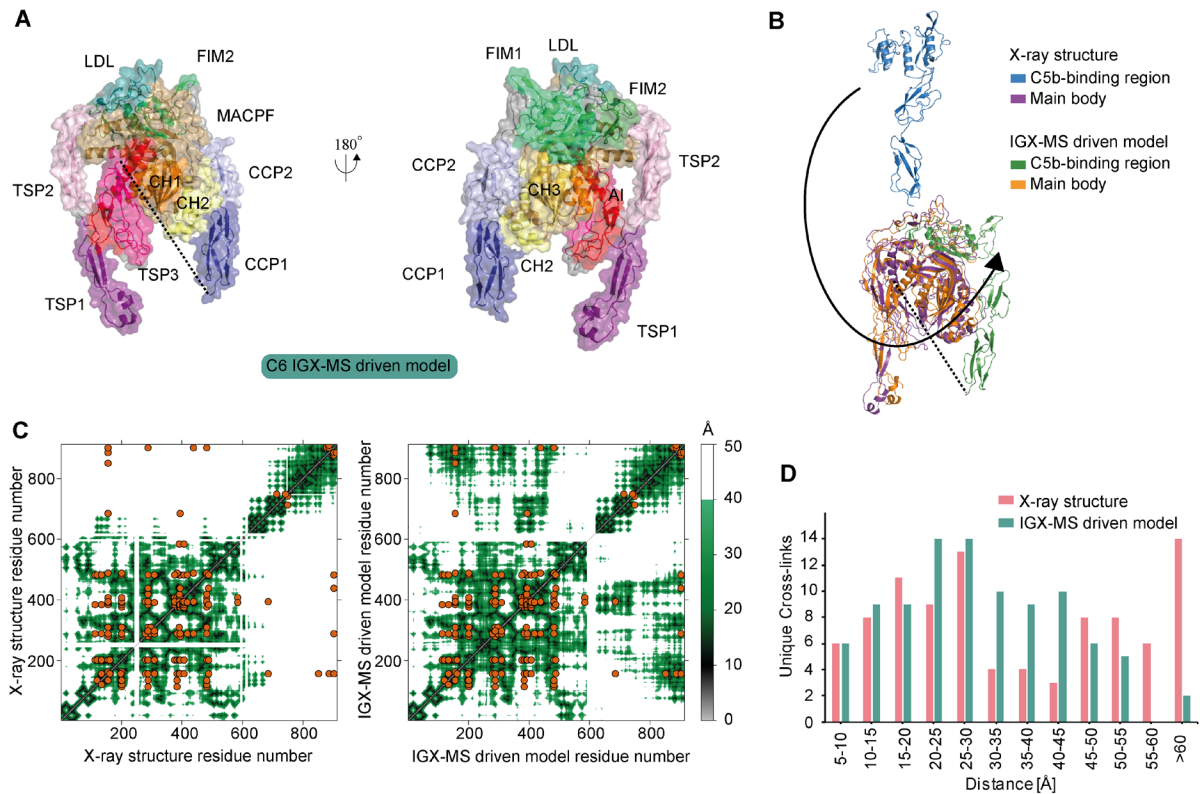
Data information: The presented IGX-MS data is summed from triplicates.



**Figure 5 - IGX-MS of the monomeric complement proteins C5 and C6, and the heterodimeric C5b6 complex.**

- Cross-links of monomeric C5 (left) and from C5b incorporated in the C5b6 complex (right) plotted on the respective available structural models (PDB ID: 3CU7 and 4A5W, respectively). The red lines indicate distances  $>30$  Å. The different domains of C5 are indicated, and C6 within the C5b6 complex is shown in grey.
- Distribution of lysine  $C_{\alpha}$ - $C_{\alpha}$  distances of unique cross-links identified by IGX-MS in monomeric C5 (light orange) and C5b when part of the C5b6 complex (dark orange).
- Cross-links of monomeric C6 (left) and C6 incorporated in the C5b6 complex (right) plotted on the respective structural models (PDB ID: 3T50 and 4A5W, respectively). The red lines indicate distances  $>30$  Å. The different domains of C6 are indicated, and C5b within the C5b6 complex is shown in grey.
- Distribution of lysine  $C_{\alpha}$ - $C_{\alpha}$  distances of unique cross-links identified by IGX-MS in monomeric C6 (light red) or C6 when incorporated within the C5b6 complex.

Data information: Only cross-links identified in at least two out of three replicates were included in the analysis.

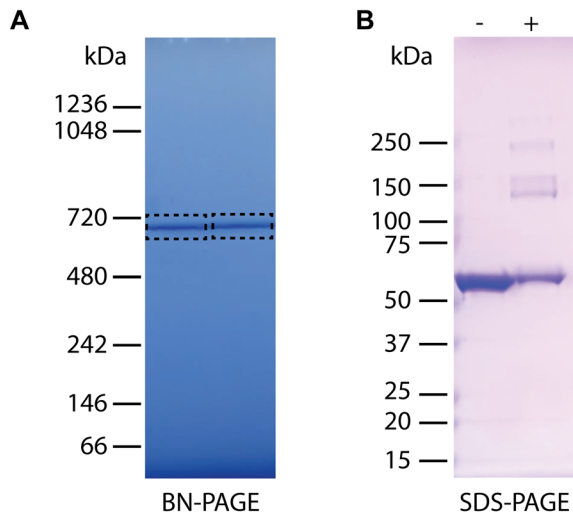


**Figure 6 - IGX-MS driven refined structural model of monomeric C6**

- IGX-MS driven homology model of monomeric free C6 depicted in three different orientations. Black dots indicate the few missing amino acids (residue 591-619) covering the linker region between the main body and C5b-binding region.
- Super-positioning of C6 IGX-MS driven model (purple and green surface) and C6 X-ray structure (orange and blue surface; PDB ID: 3T5O). Black dots indicate the few missing amino acids (residue 591-619) covering the linker region between the main body and C5b-binding region.
- Contact maps with cross-linked residues (orange dots) of C6 X-ray structure (PDB ID: 3T5O, left panel) and IGX-MS driven C6 model (right panel). The colored density represents a contact relationship smaller than 40 Å of individual residues. White density represents a contact relationship bigger than 40 Å of individual residues.
- Distribution of lysine C $\alpha$ -C $\alpha$  distances of unique cross-links identified by IGX-MS for monomeric C6 when plotted on the reported X-ray structure (pink bars, PDB ID: 3T5O) and the IGX-MS driven refined structural model (purple bars).

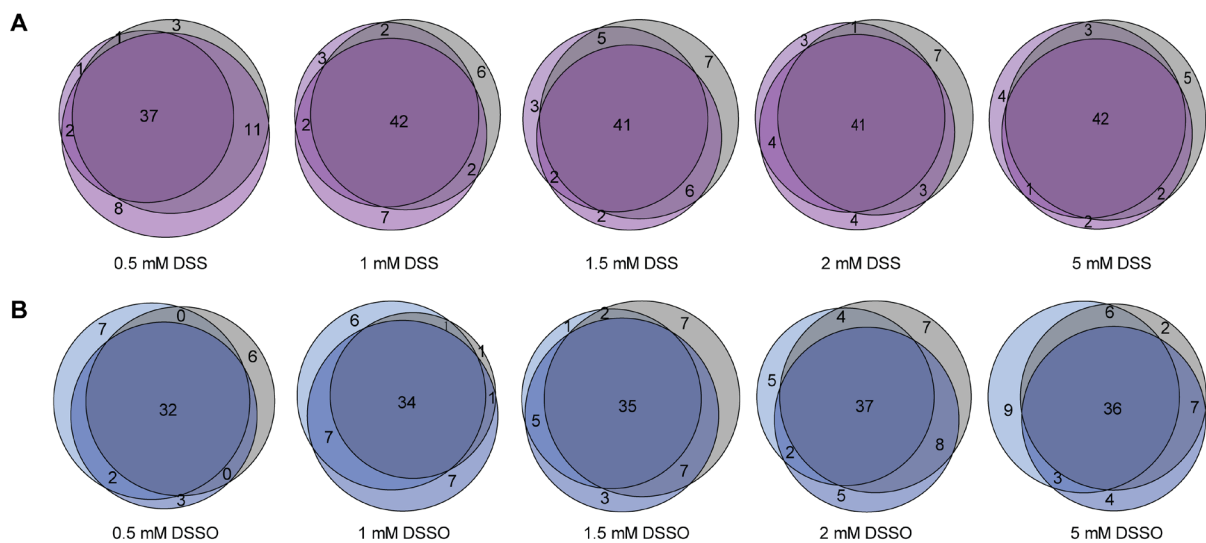
Data information: Only cross-links identified in at least two out of three replicates were included in the analysis.

## Supplementary Figures



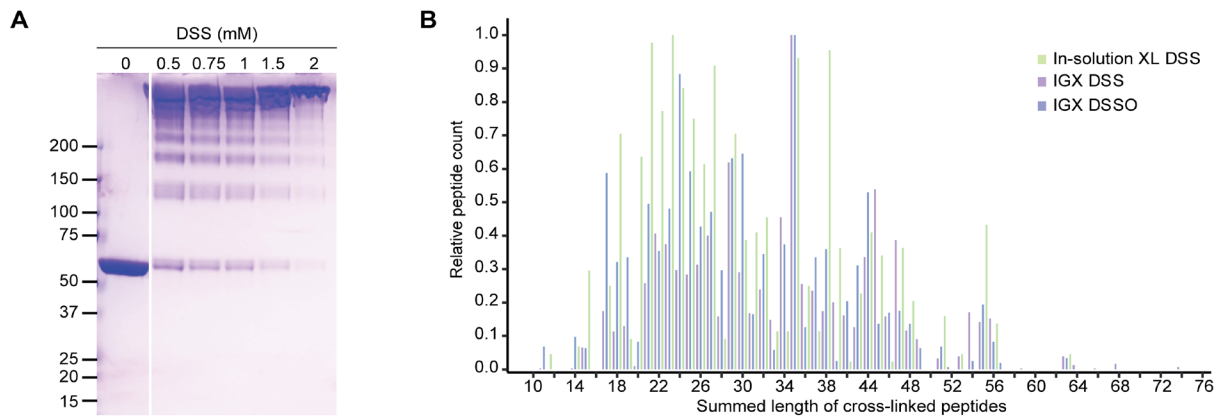
### SI Figure 1 – Verification of in-gel cross-linking (IGX) by SDS-PAGE

- BN-PAGE of *E. Coli* (10  $\mu$ g). Respective bands (dashed boxes) were excised and incubated with or without the cross-linker DSS (1.5 mM).
- SDS-PAGE of GroEL, extracted from respective gel band (see A). The non-cross-linked control showed only a band at 57 kDa of the GroEL subunit, whereas the DSS-cross-linked sample showed several bands at higher Mw.



### SI Figure 2 – In-gel cross-linking is highly reproducible and not much dependent on the concentration of the cross-linking reagent.

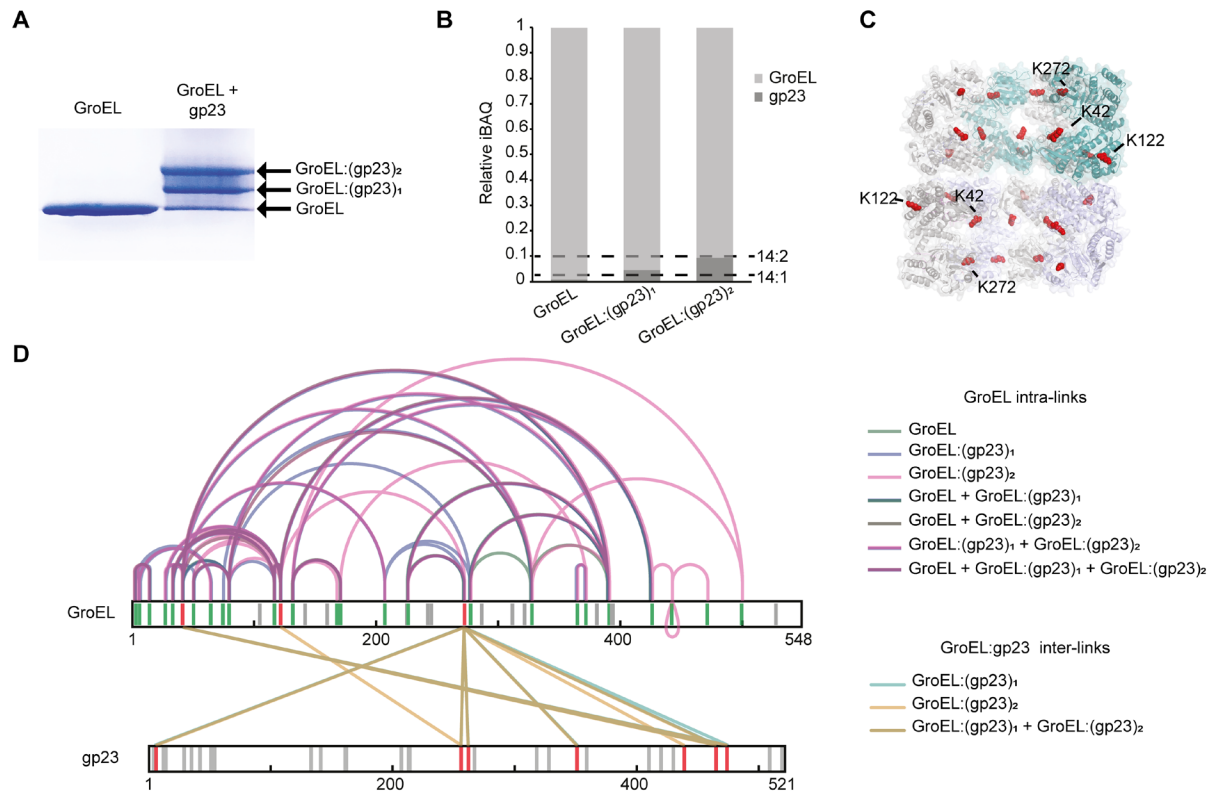
- Venn diagrams displaying the overlap of detected unique cross-links in triplicate measurements of GroEL cross-linked in gel with different DSS concentrations
- Venn diagrams displaying the overlap of detected unique cross-links in triplicate measurements of GroEL cross-linked in gel with different DSSO concentrations.



**SI Figure 3 - Optimization of cross-linker concentration for in-solution XL-MS and observed cross-linked peptide length distributions for IGX-MS and in-solution XL-MS.**

- A. GroEL was cross-linked in-solution using different DSS concentrations for the optimization of the cross-linker concentration. The samples were then loaded onto a SDS-PAGE for visualization. A concentration of 0.75 mM DSS was selected for further experiments.
- B. Comparison of summed length of cross-linked peptide-pairs from IGX-MS (1.5 mM DSS or 2 mM DSSO) or in-solution XL-MS (0.75 mM DSS) experiments.

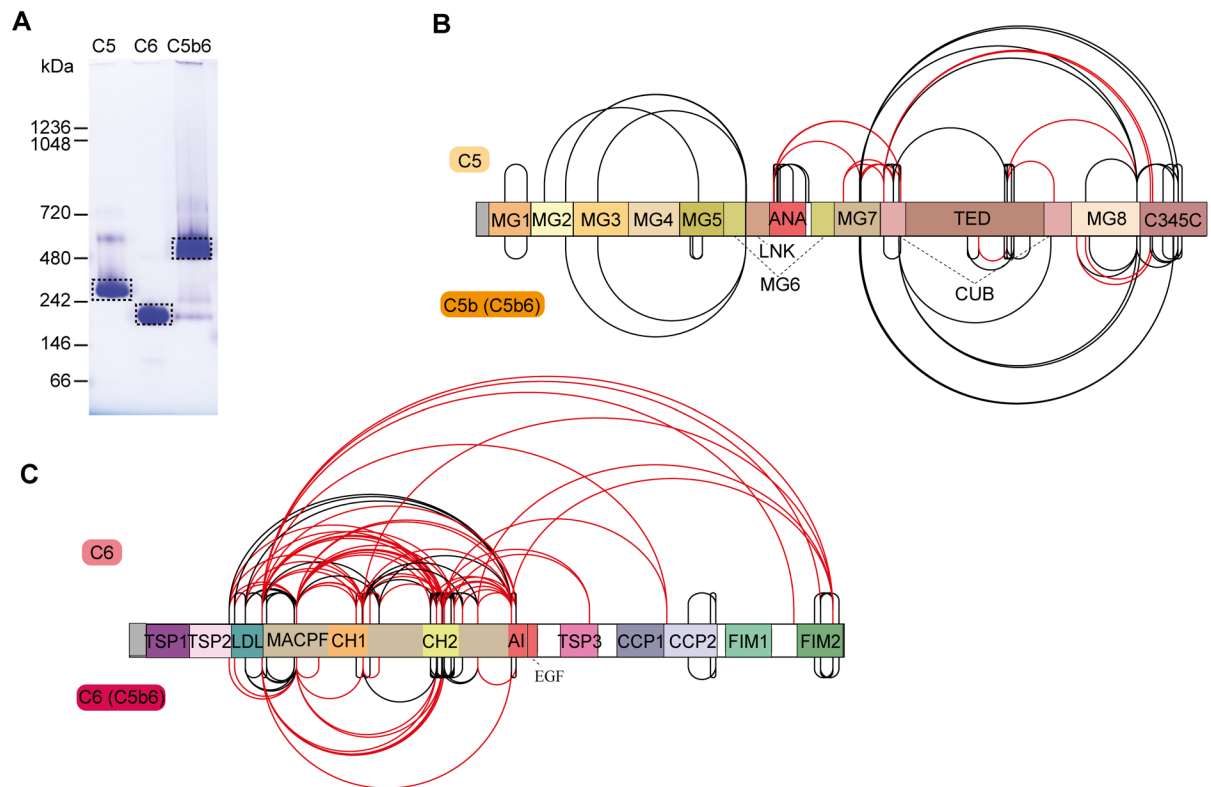




**SI Figure 4 - IGX-MS of GroEL bound to the gp23 substrate.**

- BN-PAGE of GroEL incubated with or without unfolded gp23. The arrows indicate free GroEL or GroEL bound to one or two molecules of gp23.
- Label-free quantification for the estimation of the stoichiometry of the formed complexes. Relative iBAQ values of GroEL and gp23 in the three bands. Dashed lines indicate the GroEL:gp23 ratios for the theoretically expected 14:1 and 14:2 ratios.
- Cross-section of the structural model of GroEL (PDB ID: 1KP8) with lysine residues that were found to be cross-linked to gp23 shown in red spheres.
- Overlay of cross-links identified in the three bands, representing the GroEL, the GroEL:gp23, and GroEL:(gp23)<sub>2</sub> complex. GroEL intra-links are colored green, purple, and pink. Inter-links are colored turquoise and sand. For clarity, intra-links in gp23 are not depicted. Grey lines in the sequence indicate lysine residues not cross-linked, red lines indicate inter-linked lysine residues, and green lines indicate intra-linked residues.

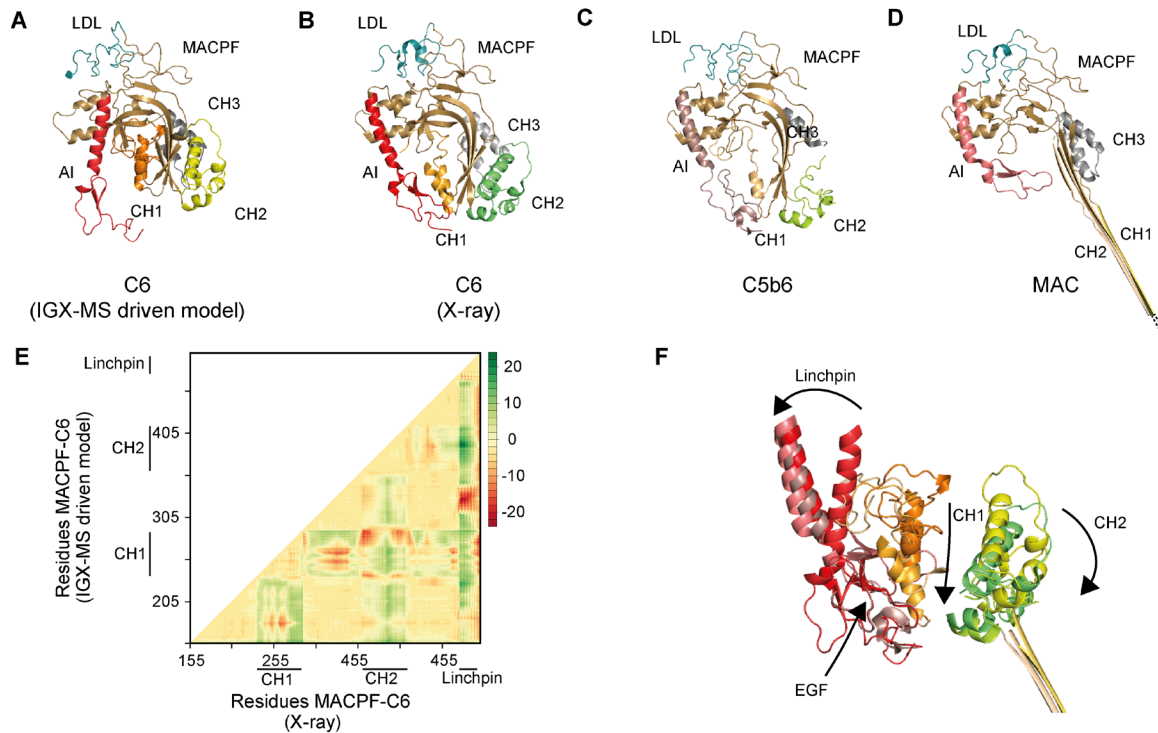
Data information: Only cross-links identified in at least two of three replicates were included in the analysis.



### SI Figure 5 - IGX of C5, C6, and C5b6.

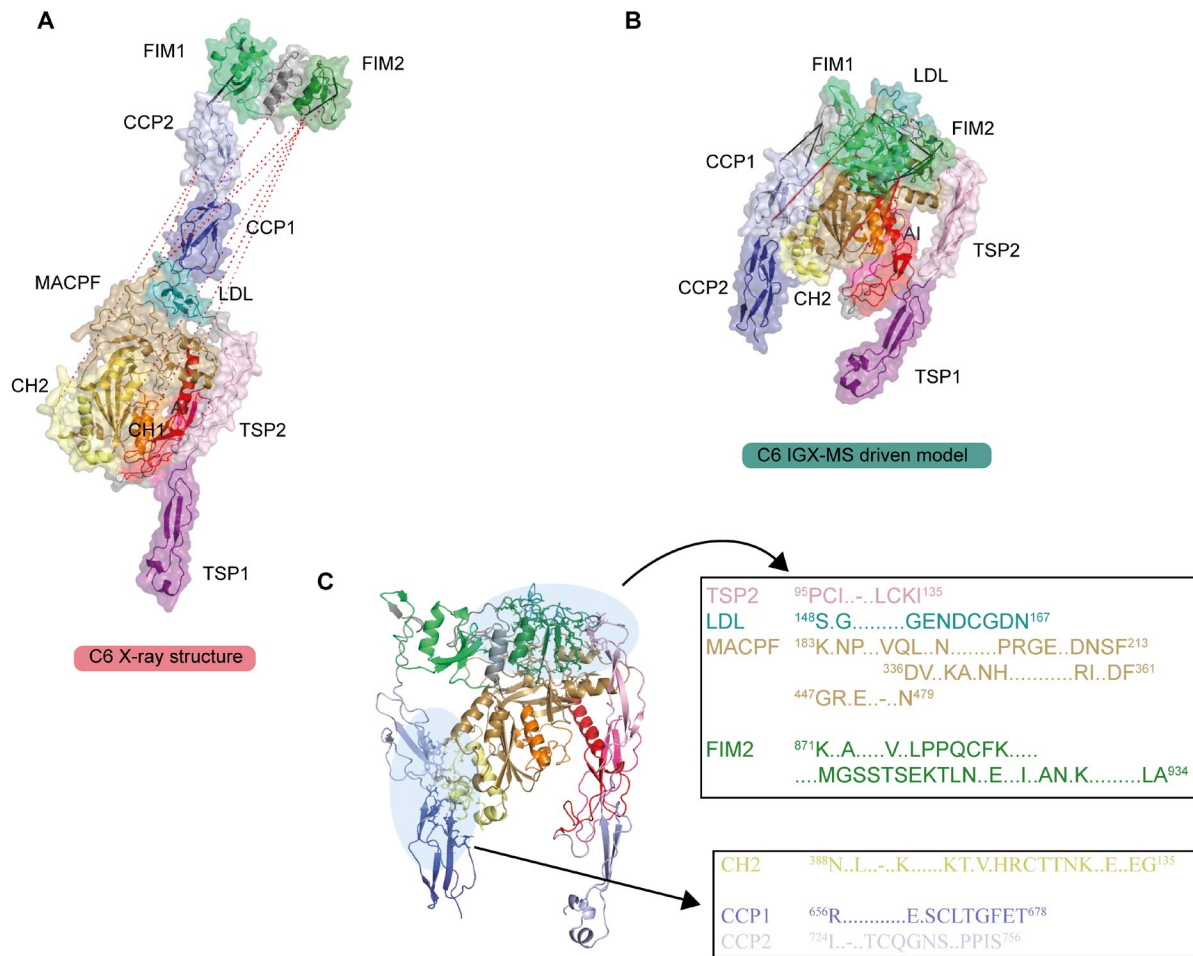
- A. BN-PAGE of complement components C5, C6, and the C5b6 complex. For each lane 5  $\mu\text{g}$  of protein (C5, C6), respectively 10  $\mu\text{g}$  (C5b6) of protein, were applied onto the gel. Dashed boxes indicate the bands cut out for subsequent IGX-MS analysis.
- B, C. Schematic overview for domain-centered cross-link results for C5 and Cb5b in the C5b6 complex (B) or C6 and C6 in the C5b6 complex (C). Black lines indicate cross-links within the distance restraints ( $\leq 30 \text{ \AA}$ ). Red lines indicate cross-links exceeding the distance restraints ( $\geq 30 \text{ \AA}$ ). Notably, for C6, many cross-links to the N-terminal domains are not present when C6 is complexed to C5b.

Data information: Only cross-links identified in at least two of three replicates were included in the analysis.



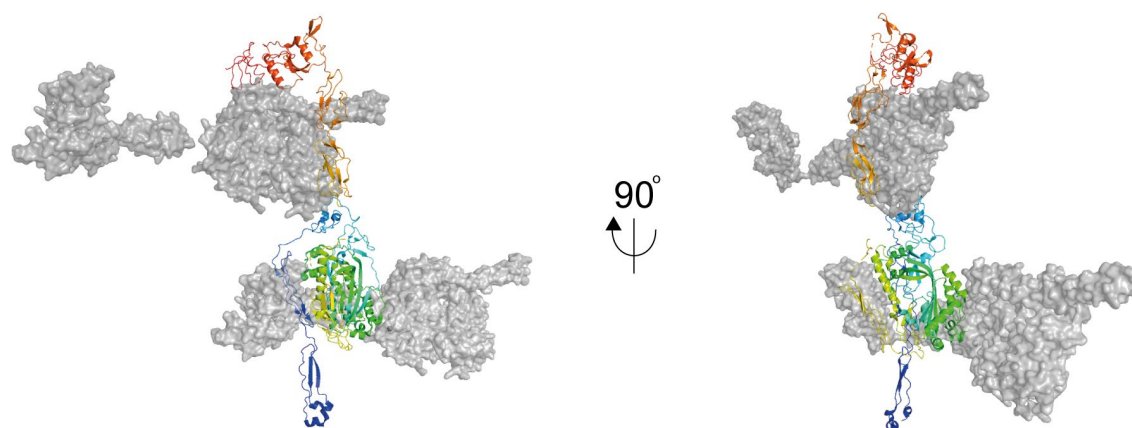
**SI Figure 6 – Structural rearrangements of the AI and CH regions of C6 when going from free C6, the C5b6 intermediate structure to C6 in the fully assembled MAC.**

- A-D.** Cartoon representation of the LDL (blue) and MACPF (light pink) domains within the IGX-MS driven C6 model (A), C6 from the X-ray structure (PDB ID: 3T5O) (B), C6 complexed with C5b (PDB ID: 4A5W) (C), and C6 as part of the MAC (PDB ID: 6H03) (D). The regions CH 1-3 (orange, dark pink, and green) and auto-inhibitory (AI, red) within MACPF are shown.
- E.** The difference distance matrix of superposed C6-MACPF domains. The difference distance was calculated by subtracting the coordinates of aligned MACPF backbones of the IGX-MS driven model and the I-tasser model (for complete sequence coverage) of C6 -X-ray structure (PDB ID: 3T5O). The distance values are plotted with colors representing  $C\alpha$  differences of -20 to 20 Å according to the right-sided scale.
- F.** Overlay of the AI (composed of linchpin helix and EGF domain), CH1, and CH2 regions of MACPF in four distinct conformations of C6, namely the IGX-MS driven model, the deposited C6 X-ray structure, when incorporated in the C5b6 complex, and finally when incorporated in the fully assembled MAC. Arrows indicate the conformational changes of the different regions from free C6 (IGX-MS driven model) to partial activation (PDB ID: 3T5O) and C5b binding (PDB ID: 4A5W) and final assembly into the MAC (PDB ID: 6H03).



**SI Figure 7 – Comparison of the monomeric C6 X-ray structure and IGX-MS driven structural model, wherein the C5b-binding region becomes docked to the main body.**

- Cross-links identified for residues in the C5b-binding region plotted onto the C6 X-ray structure (PDB ID: 3T50). Cross-links that were used to dock the C5b-binding domain (CCP1-2, FIM1-2) to the LDL- and MACPF domain are shown as dotted lines. Black lines indicate cross-links within the distance restraints ( $\leq 30$  Å). Red lines indicate cross-links exceeding the distance restraints ( $\geq 30$  Å).
- IGX-MS driven structural model of free C6. Cross-links obtained for the C5b-binding region are plotted onto the final model of C6. Solid black lines indicate distances below 30 Å, and red dotted lines indicate links with a distance larger than 30 Å.
- Interaction interface analysis for C5b-binding and the LDL-, MACPF domains. Blue circles indicate binding interfaces, for which interacting residues were identified (black boxes).



**SI Figure 8 – Possible interactions between adjacent C6 molecules in the deposited X-ray crystal structure.**

The total electron density acquired for the X-ray structure of C6 (PDB ID: 3T5O) reveals closely packed neighboring C6 molecules. A single C6 molecule is shown here in a rainbow-colored cartoon in two orientations. The surface (grey) of two adjacent C6 molecules in the electron density maps shows the interaction of the C-terminal CCP and FIM doma