



**HAL**  
open science

## **In vivo Cross-Linking MS of the Complement System MAC Assembled on Live Gram-Positive Bacteria**

Hamed Khakzad, Lotta Happonen, Guy Tran van Nhieu, Johan Malmström,  
Lars Malmström

► **To cite this version:**

Hamed Khakzad, Lotta Happonen, Guy Tran van Nhieu, Johan Malmström, Lars Malmström. In vivo Cross-Linking MS of the Complement System MAC Assembled on Live Gram-Positive Bacteria. *Frontiers in Genetics*, 2021, 11, 10.3389/fgene.2020.612475 . hal-04429714

**HAL Id: hal-04429714**

**<https://hal.science/hal-04429714>**

Submitted on 31 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# *In vivo* Cross-Linking MS of the Complement System MAC Assembled on Live Gram-Positive Bacteria

Hamed Khakzad<sup>1,2</sup>, Lotta Happonen<sup>3</sup>, Guy Tran Van Nhieu<sup>1,2</sup>, Johan Malmström<sup>3</sup> and Lars Malmström<sup>3\*</sup>

<sup>1</sup> Equipe Signalisation Calcique et Infections Microbiennes, Ecole Normale Supérieure Paris-Saclay, Gif-sur-Yvette, France,

<sup>2</sup> Institut National de la Santé et de la Recherche Médicale U1282, Gif-sur-Yvette, France, <sup>3</sup> Faculty of Medicine, Department of Clinical Sciences, Division of Infection Medicine, Lund University, Lund, Sweden

## OPEN ACCESS

### Edited by:

Angelo Facchiano,  
Institute of Food Sciences, National  
Research Council (CNR), Italy

### Reviewed by:

Ankush Sharma,  
University of Oslo, Norway  
Vijaykumar Muley,  
Universidad Nacional Autónoma  
de México, Mexico

### \*Correspondence:

Lars Malmström  
lars.malmstrom@med.lu.se

### Specialty section:

This article was submitted to  
Computational Genomics,  
a section of the journal  
Frontiers in Genetics

Received: 30 September 2020

Accepted: 24 November 2020

Published: 08 January 2021

### Citation:

Khakzad H, Happonen L,  
Tran Van Nhieu G, Malmström J and  
Malmström L (2021) *In vivo*  
Cross-Linking MS of the Complement  
System MAC Assembled on Live  
Gram-Positive Bacteria.  
Front. Genet. 11:612475.  
doi: 10.3389/fgene.2020.612475

Protein–protein interactions are central in many biological processes, but they are challenging to characterize, especially in complex samples. Protein cross-linking combined with mass spectrometry (MS) and computational modeling is gaining increased recognition as a viable tool in protein interaction studies. Here, we provide insights into the structure of the multicomponent human complement system membrane attack complex (MAC) using *in vivo* cross-linking MS combined with computational macromolecular modeling. We developed an affinity procedure followed by chemical cross-linking on human blood plasma using live *Streptococcus pyogenes* to enrich for native MAC associated with the bacterial surface. In this highly complex sample, we identified over 100 cross-linked lysine–lysine pairs between different MAC components that enabled us to present a quaternary model of the assembled MAC in its native environment. Demonstrating the validity of our approach, this MAC model is supported by existing X-ray crystallographic and electron cryo-microscopic models. This approach allows the study of protein–protein interactions in native environment mimicking their natural milieu. Its high potential in assisting and refining data interpretation in electron cryo-tomographic experiments will be discussed.

**Keywords:** membrane attack complex, cross-linking mass spectrometry, *Streptococcus pyogenes*, Gram-positive bacteria, *in vivo* cross-linking

## INTRODUCTION

The human immune system plays a crucial role in detecting and clearing bacterial infections. The complement system is an important immunological sensor and effector pathway that in interplay with other components of innate and adaptive immunity defends the host against infection and regulates the inflammatory response. Complement activation can occur through three distinct pathways depending on the molecular trigger, all proceeding through a proteolytic cascade that leads to the formation of the multicomponent membrane attack complex (MAC) capable of disrupting target cell membranes. While Gram-positive bacteria are thought to be able to resist the MAC-induced membrane disruption due to their thick cell wall, several such bacteria have also been shown to secrete small proteins that specifically target this multicomponent

complex (Berends et al., 2013) and to possibly inhibit its formation and function. In *Streptococcus pyogenes*, SIC (Akeson et al., 1996) and the recently characterized ISP (Happonen et al., 2019) perform this function.

The importance of MAC targeting the surface of Gram-positive bacteria is illustrated in various examples. Berends et al. (2013) found that complement activation leads to specific C3-independent deposition of MAC on the Gram-positive bacterial surface. In *S. pyogenes*, this deposition is localized close to the division septum (Berends et al., 2013). Intriguingly, the mammalian peptidoglycan recognition proteins (PGRPs) bind to the bacterial cell wall of Gram-positive *Bacillus subtilis* and kill the bacteria through specific interaction with the CssR–CssS two-component system at the site of daughter cell separation (Kashyap et al., 2011). While we previously reported interactions between MAC components and bacterial CovR, part of CovRS two-component system in *S. pyogenes* (Happonen et al., 2019), the importance of studying the MAC formation on the bacterial surface is evident.

The MAC consists of four single-copy components, C5b, C6, C7, and C8, together with multiple copies of C9, the structural component forming the pore walls (Figure 1A). The MAC has been extensively studied, and the structure has been determined at resolutions ranging from 2.51 to 5.6 Å. Many structures of individual components, partial assemblies, and full assemblies are available (Lovelace et al., 2011; Serna et al., 2016; Menny et al., 2018).

Biological mass spectrometry (MS) has undergone rapid evolution and has become the method of choice to analyze complex protein samples (Aebersold and Mann, 2003). One of the most powerful features of MS is handling complex samples with ease, permitting the identification and quantification of proteins in biopsies and other clinical samples (Guo et al., 2015; Zhang et al., 2019). Other important aspects are its high capacity, a high degree of automated operation, and reliability. Modern instruments and acquisition strategies can identify and quantify thousands of proteins with high accuracy and high sensitivity (Röst et al., 2014). The data obtained by MS can be categorized depending on acquisition strategies into high-resolution MS1 proteomics (hrMS1), shotgun proteomics, or data-dependent acquisition (DDA) and data-independent acquisition (DIA). Shotgun proteomics measure the *m/z* of all peptide ions introduced into the MS, the produced spectra being called an MS1 or precursor spectra, then select up to tens of prominent ions, which are isolated and fragmented. The *m/z* of all produced fragments is measured (the produced spectra referred to as MS2 or tandem spectra) (Nesvizhskii and Aebersold, 2005). In hrMS1, the MS only measures MS1 spectra at the highest possible resolution. In DIA acquisition, systematic fragment intensities are collected with many peptides being fragmented and measured at once. This results in throughput to levels similar to shotgun MS but increases the quantitative accuracy (Gillet et al., 2012).

Cross-linking MS (XL-MS) has been shown to be a powerful technique to measure protein–protein interactions directly in complex samples (Sinz, 2006; Herzog et al., 2012; Liu et al., 2015; Hauri et al., 2019). Here, a bi-functional cross-linking reagent,

such as disuccinimidyl suberate (DSS), is used to covalently link two lysine residues when the proteins are in their native states. The proteins are subjected to endoproteolytic digestion using, for example, trypsin, and peptides linked by DSS are analyzed. Cross-links between polypeptide chains are called inter-cross-links, as opposed to intra-cross-links. As the proteins are in their native state when cross-linked, the 11.4 Å long DSS arm determines the maximum distance between the cross-linked amino acids. The maximum distance information is then used to identify and characterize protein–protein interactions. A model of the protein–protein interaction can be created using macromolecular docking software, such as RosettaDock (Gray, 2006; Koehler Leman et al., 2020), MEGADOCK (Ohue et al., 2014b), or HADDOCK (Zundert et al., 2016), in cases where enough cross-links are identified and reliable models of the interacting proteins are available.

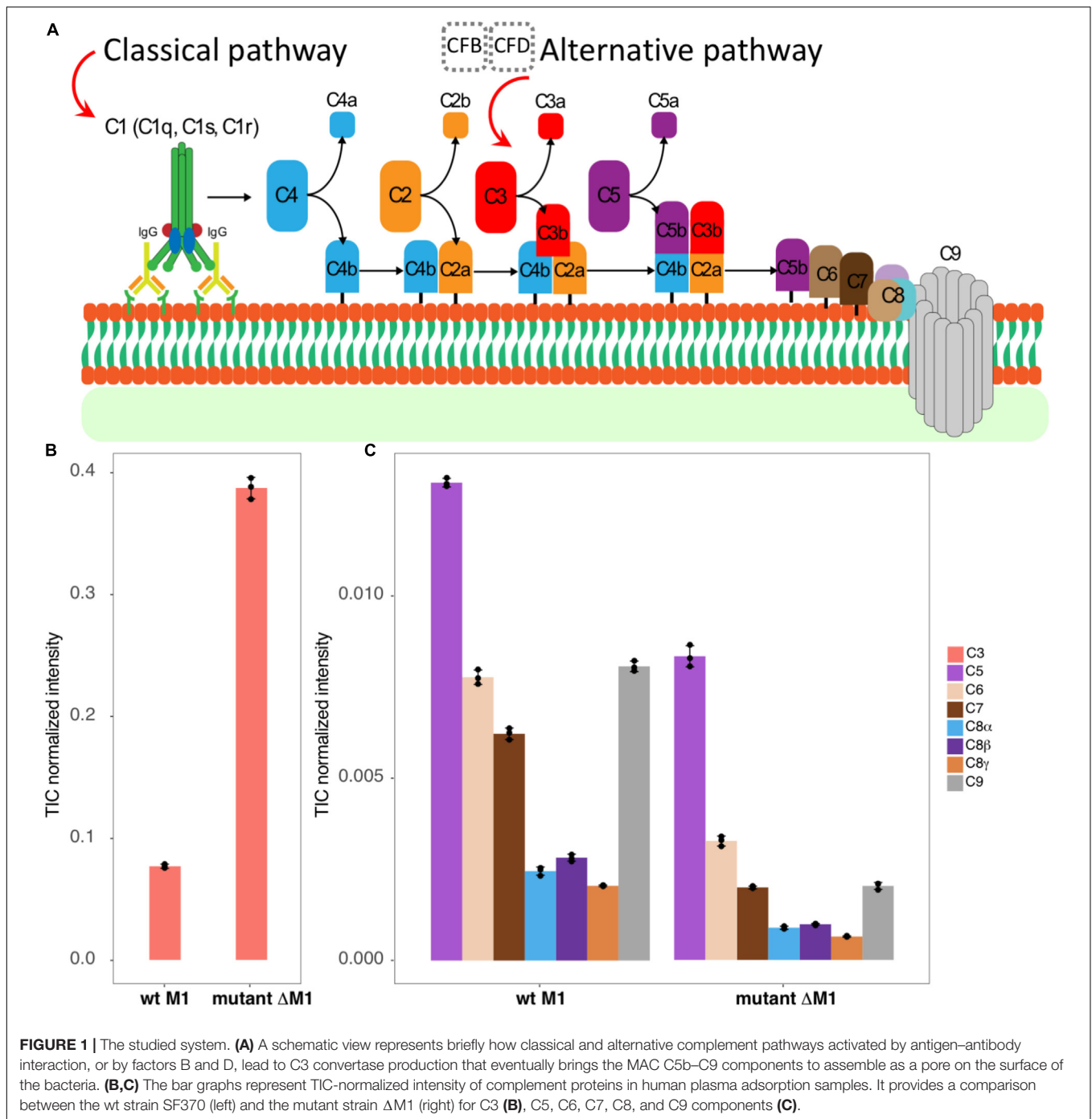
Here, we applied a simplified version (see the Materials and Methods section for details) of the recently developed targeted cross-linking mass spectrometry (TX-MS) approach (Hauri et al., 2019) to study the formation of MAC on the surface of live *S. pyogenes* bacteria in human plasma. We show that our integrative mass spectrometric approach can detect cross-linked peptides in complex unfractionated samples. This integrative approach is based on MS/MS acquisitions filtered by hrMS1 isotopic patterns combined with macromolecular docking encompassing the entire MAC and recapitulating the interaction interfaces previously described by X-ray crystallography (Lovelace et al., 2011) and single-particle electron cryo-microscopy (Serna et al., 2016; Menny et al., 2018). These results further highlight that MAC can assemble in its active, pore-like form on the surface of Gram-positive bacteria with identified cross-linked peptides between adjacent C9 proteins.

## MATERIALS AND METHODS

### Computational Workflow

The UniProt accession IDs for human complement proteins considered in this study, including C3, C5, C6, C7, C8 $\alpha$ , C8 $\beta$ , C8 $\gamma$ , and C9, were P01024, P01031, P13671, P10643, P07357, P07358, P07360, and P02748, respectively. The computational modeling method designed for this study was a modified version of the TX-MS workflow (Hauri et al., 2019). Accordingly, we first generated a list of all hypothetical lysine–lysine pairs based on the sequence of partner proteins for each pairwise protein–protein interaction. Considering all studied components and with the minimum peptide length restricted to five residues, this list included a total number of 27,873 hypothetical XLs.

The generated XL list was analyzed for each interface through all MS/MS, DDA samples using the TX-MS workflow. To do that, DDA mzML data were first converted to Mascot Generic Format (MGF) files using *msconvert* (Chambers, 2012). Then, precursor *m/z* values for each XL were calculated in six different charge states ranging from + 3 to + 8. These values were used to filter the MGF files with a delta window of 0.05 *m/z*. After that, a list of product ions ( $\gamma$  and  $b$  ions with charge states + 1 to + 3)



were produced for each selected XL and searched through all filtered spectra to find a similar pattern. Finally, by considering the number of detected fragment ions, those spectra with >30 matched product ions were selected for the second phase and to be validated by hrMS1 data.

To further validate selected XLs from DDA samples, we used Dinosaur software (Teleman et al., 2016) to extract the isotopic pattern for each selected XL and searched them through hrMS1 samples. The final filtered list of XLs (those detected in hrMS1

data) in this stage was used as distance constraints for protein–protein docking. For structural docking, we used the input protein structures deposited in protein data bank with PDB IDs: 6H04, 2WII, 3OJY, and 4A5W.

To generate docking models, we used MEGADOCK 4.0, a fast Fourier transform (FFT)-based rigid-body docking software (Ohue et al., 2014b). Its scoring function is a combinatorial score based on electrostatics and desolvation free energy combined with shape complementarity (Ohue et al., 2014a). We produced

2,000 docking models per interface. So, by considering the six evaluated pairwise interfaces, a total number of 12,000 docking models were generated where for each interface, the top 100 models based on the MEGADOCK score were selected and then filtered out using distance restraints provided by XLs. The threshold cut-off to evaluate models based on XLs was set to 32 Å between C $\beta$ -C $\beta$  as we used DSS (with 11.4 Å arm length) to prepare XL-MS samples. Finally, the top five models based on the number of XLs they fulfill and the average length of all fulfilled XLs in the model were selected as the final models.

## Cross-Linking of Plasma Adsorption Samples

For cross-linking of MAC to the surface of *S. pyogenes*, we used pooled normal human plasma adsorbed onto the surface of *S. pyogenes* bacteria (Happonen et al., 2019; Hauri et al., 2019). Briefly, the *S. pyogenes* M1 serotype strain SF370 from the American Type Culture Collection (ATCC; strain reference 700294) and its derived  $\Delta$ M1 mutant lacking the major surface antigen, the M1 protein, was grown at 37°C 5% CO<sub>2</sub> to mid-exponential phase (OD<sub>620 nm</sub> ~0.4–0.5) in Todd-Hewitt (TH) broth supplemented with 0.3% (w/v) yeast extract. The cells were harvested by centrifugation (1,900 × g, 10 min, 22°C), washed once with phosphate-buffered saline [PBS; 10 mM phosphate buffer, 2.7 mM potassium chloride, 137 mM sodium chloride (Sigma)], recentrifuged (1,900 × g, 5 min, 22°C), and resuspended to an approximate concentration of 1 × 10<sup>9</sup> colony forming units ml<sup>-1</sup>. For plasma adsorption, 400 μl of pooled normal human plasma from healthy donors (Innovative Research) was supplemented with a final concentration of 10 μM argatroban (Sigma) to prevent any plasma clotting and was mixed with 100 μl of bacteria with subsequent incubation at 37°C for 30 min at 500 rpm.

SF370 and  $\Delta$ M1 bacteria with adsorbed plasma proteins—including the MAC—were harvested by centrifugation (1,900 × g, 5 min, 22°C), washed twice with PBS, and finally resuspended in 200 μl of PBS for cross-linking. For cross-linking, heavy/light disuccinimidyl-suberate cross-linker (DSS-H12/D12, Creative Molecules Inc.) resuspended in 100% dimethylformamide (Sigma) was added to final concentrations of 0, 0.25, 0.5, 1.0, and 2.0 mM in duplicates, and the samples were incubated at 37°C for 1 h at 900 rpm. The cross-linking reaction was quenched with a final concentration of 50 mM ammonium bicarbonate (Sigma) at 37°C for 15 min at 900 rpm. The MAC complex adsorbed to the bacterial surface was released by limited proteolysis with 2 μg trypsin (Promega) at 37°C for 1 h at 800 rpm prior to cell debris removal by centrifugation (1,900 × g, 15 min) and subsequent supernatant recovery. Two hundred microliters of the supernatant containing the MAC was recovered, and any remaining bacteria were killed by heat inactivation (85°C, 5 min) prior to sample preparation for MS.

## Sample Preparation

For MS, 100 μl of the heat-inactivated supernatant containing the MAC was denatured in 200 μl 8 M urea–100 mM ammonium bicarbonate (both Sigma), and the cysteine bonds

were reduced with 5 mM *tris*(2-carboxyethyl)phosphine (Sigma) at 37°C for 2 h at 500 rpm. The cysteines were alkylated with 5 mM iodoacetamide (Sigma) at 22°C for 30 min, and the samples were subsequently digested using sequencing-grade lysyl endopeptidase (Wako) at 37°C for 4 h at 500 rpm. The samples were diluted with 100 mM ammonium bicarbonate to a final urea concentration of 1.5 M and subsequent trypsin (Promega) digestion at 37°C for 16 h. Digested samples were acidified with 10% formic acid to a final pH of 3.0, and the peptides were purified with C18 reverse-phase spin columns according to the manufacturer's instructions (Macrospin columns; Harvard Apparatus). Dried peptides were reconstituted in 2% acetonitrile and 0.2% formic acid prior to MS analyses.

## MS for Cross-Linking Analysis

All peptide analyses were performed on a Q Exactive HFX mass spectrometer (Thermo Scientific) connected to an EASY-nLC 1200 ultra-high-performance liquid chromatography system (Thermo Scientific). For analysis of cross-linked samples, we used DDA-MS and hrMS1. For DDA analysis, the peptides were separated on an EASY-Spray column (Thermo Scientific; ID 75 μm × 25 cm, column temperature 45°C) operated at a constant pressure of 800 bar. A linear gradient from 4 to 45% of 0.1% formic acid in 80% acetonitrile was run for 50 min at a flow rate of 300 nl min<sup>-1</sup>. One full MS scan (resolution 60,000 at 200 m/z; mass range 350–1,600 m/z) was followed by MS/MS scans (resolution 15,000 at 200 m/z) of the 15 most abundant ion signals. The precursor ions were isolated with 2 m/z isolation width and fragmented using higher-energy collisional-induced dissociation (CID) at a normalized collision energy (NCE) of 30. Charge state screening was enabled, and precursors with an unknown charge state and singly charged ions were excluded. The dynamic exclusion window was set to 15 s and limited to 300 entries. The automatic gain control was set to 3e6 for MS and 1e5 for MS/MS with ion accumulation times of 110 and 60 ms, respectively. The intensity threshold for precursor ion selection was set to 1.7e4.

In the hrMS1 analysis, peptides were separated using an EASY-Spray column (Thermo Scientific; ID 75 μm × 25 cm, column temperature 45°C) operated at a constant pressure of 800 bar. A linear gradient from 4 to 45% of 0.1% formic acid in 80% acetonitrile was run for 60 min at a flow rate of 300 nl min<sup>-1</sup>. High-resolution MS scans (resolution 240,000 at 200 m/z; mass range from 400 to 2,000 m/z) were acquired using automatic gain control set to 3e6 and a fill time of 500 ms.

## MS for Protein Quantification

The MAC enriched to the streptococcal surface was quantified using DIA-MS. For DIA-MS, the peptides from the non-cross-linked samples were separated using an EASY-Spray column (Thermo Scientific; ID 75 μm × 25 cm, column temperature 45°C) operated at a constant pressure of 800 bar. A linear gradient from 4 to 45% of 0.1% formic acid in 80% acetonitrile was run for 110 min at a flow rate of 300 nl min<sup>-1</sup>. A full MS scan (resolution 60,000 at 200 m/z; mass range from 390 to 1,210 m/z) was followed by 32 MS/MS full fragmentation scans (resolution 35,000 at 200 m/z) using an isolation window

**TABLE 1** | The Identified XLs for individual components of pairwise protein–protein interactions with previously solved structures.

Components	Reference PDB	# K–K PAIRS	# INTRA-XLS	# INTER-XLS
C3B	2WII (X-ray, 2.7 Å)	5,886	43	–
C5B	4A5W (X-ray, 3.5 Å)	6,105	32	–
C5B–C6	6H04 (EM, 5.6 Å)	6,882	–	5
C6–C7	6H04 (EM, 5.6 Å)	2,728	–	8
C7–C8 ( $\alpha$ , $\beta$ , and $\gamma$ )	6H04 (EM, 5.6 Å)	2,640	–	16
C8 ( $\alpha$ , $\beta$ , and $\gamma$ )	3OJY (X-ray, 2.51 Å), 6H04 (EM, 5.6 Å)	1,031	10	–
C8 ( $\alpha$ , $\beta$ , and $\gamma$ )–C9	3OJY-6DLW (EM, 3.9 Å)	2,040	–	5
C9–C9	6H04 (EM, 5.6 Å)	561	3	4

The individual components analyzed, the reference PDB (method, resolution), number of intra-XLs (threshold < 32 Å), and number of inter-XLs (threshold < 32 Å) are shown in the table columns, respectively. For C9–C9 interaction, detected XLs that could not be explained on a single C9 protein (above threshold), but in C9–C9, were considered as inter-XLs.

of 26 m/z (including 0.5 m/z overlap between the previous and next window). The precursor ions within each isolation window were fragmented using higher-energy CID at a NCE of 30. The automatic gain control was set to 3e6 for MS and 1e6 for MS/MS with ion accumulation times of 100 and 120 ms, respectively. For protein abundance comparison across the different streptococcal strains, the data were normalized based on measured total ion current (TIC).

## RESULTS

### The Human MAC Assembles on the Surface of Live Gram-Positive Bacteria

We studied the MAC assembly on the surface of live bacteria. The MAC has previously been shown to associate at the division septum of *S. pyogenes*, where the cell wall might not yet be fully mature, in a complement system C3-independent manner, possibly mediated through the interaction with ISP and CovR (Happonen et al., 2019). Here, we demonstrated that the MAC is assembled on the surface of *S. pyogenes*. In order to study the MAC assembly on a Gram-positive bacterium, we performed TX-MS using human plasma on the surface of live wt *S. pyogenes* bacteria expressing the main surface antigen, the M1 protein, and an isogenic mutant lacking M1 expression, without any subsequent enrichment of cross-linked peptides.

By analyzing non-cross-linked control samples by DIA-MS, we found the complement protein C3 and all MAC components (C5b, C6, C7, C8, and C9) to be associated with the bacterial surface (Figure 1B). Intriguingly, the mutant lacking M1 expression bound complement C3 to a higher degree than the wt strain (Figure 1B). These results could be explained by the lack of steric hindrance imposed by the extended, coiled-coil M1 protein in the M1-deficient strain (Happonen et al., 2019). Moreover, the MAC components were more enriched in the wt strain than in the M1 mutant (Figure 1C), consistent with MAC binding to *S. pyogenes* being complement C3-independent.

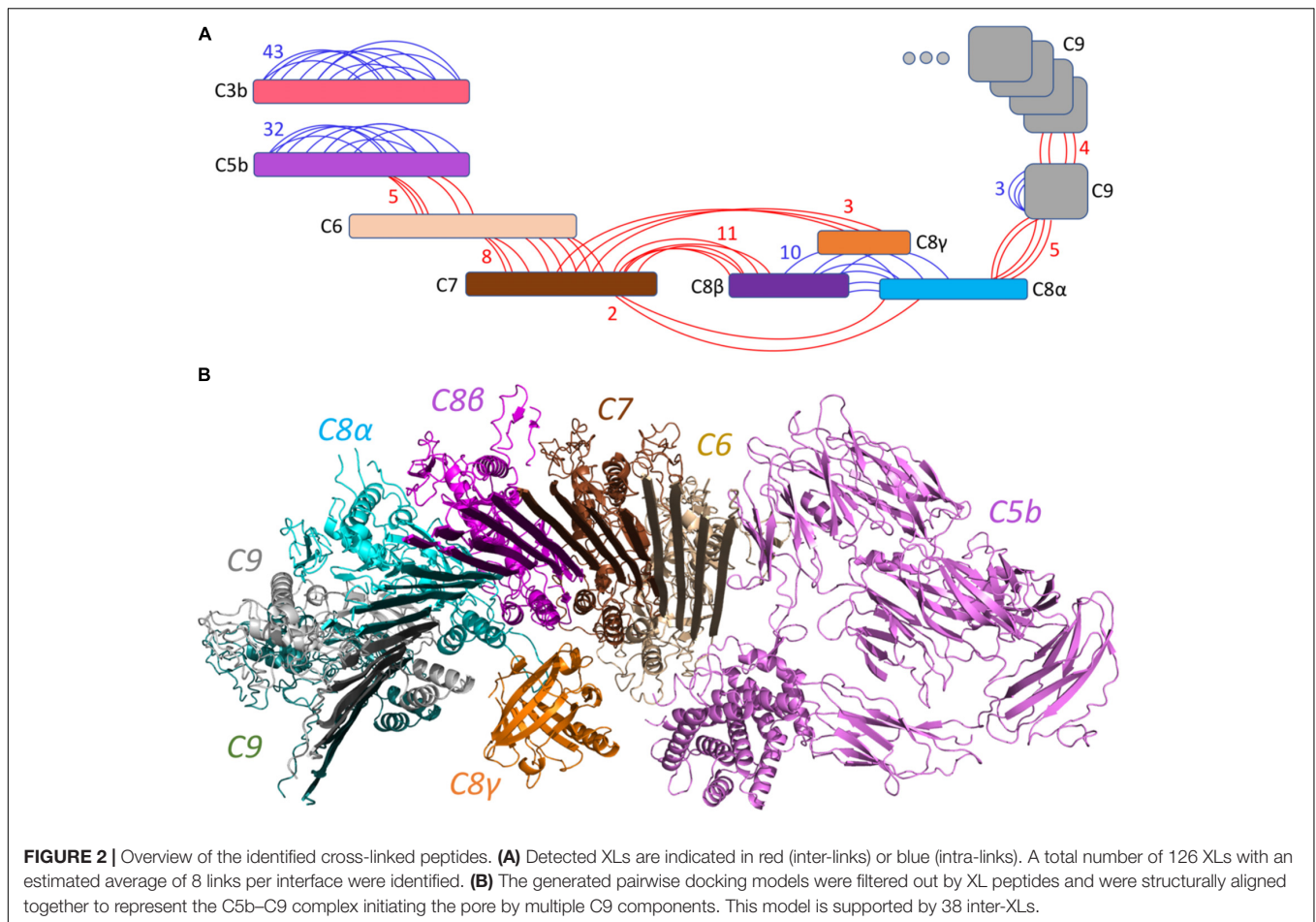
We validated the structure of some of the intact components from human blood plasma enriched at the bacterial surface by analyzing intra-protein cross-links. For this, we analyzed the

samples cross-linked with DSS. Taken together, we identified 43, 32, and 10 intra-protein cross-links for C3, C5, and all three chains ( $\alpha$ ,  $\beta$ , and  $\gamma$ ) of C8. **Supplementary Figure 1** represents the detected XLs mapped on the experimentally solved structures (PDB IDs: 6H04, 3OJY, 4A5W, and 2WII).

### Structural Modeling

The structure of all MAC components has been determined *in vitro* at different resolutions (Serna et al., 2016; Menny et al., 2018) ranging from 2.51 to 5.6 Å, as indicated in **Table 1**. To determine the structural model of the complement system MAC in its native environment enriched from human plasma onto live bacteria, we first analyzed the pairwise interactions between the subunits by considering six interfaces including C5b–C6, C6–C7, C7–C8, C8 $\alpha$ –C8 $\beta$ , C8 ( $\alpha$ ,  $\beta$ , and  $\gamma$ )–C9, and C9–C9. For each interface, we generated a list of all hypothetical lysine–lysine pairs (**Table 1**), which were further investigated using MS data generated by two different mass spectrometric acquisition methods, DDA-MS and hrMS1. Accordingly, by retaining the spectrums with high fragment ion coverages from DDA data, we further evaluated the MS/MS output XL list by hrMS1 data to filter out XLs based on their isotopic pattern (see the Materials and Methods section). Applying this double filtration step, the produced list was used to evaluate molecular docking models of each pairwise interface. As a result, we generated docking models for each interface supported with an estimated average of eight inter-XLs per interface (**Figure 2A**).

This structural modeling step was done without considering any of the native structures to guide the modeling (blind modeling). Hence, we were able to calculate the Root Mean Square Deviation (RMSD) of the produced model as an average of all backbone atoms (CA, C, N, O) of all pairwise interfaces to the native conformation (reference structures PDB: 6H04 (Menny et al., 2018) at 5.6 Å resolution and PDB: 3OJY (Lovelace et al., 2011) at 2.51 Å resolution). **Supplementary Figure 2** shows the detailed RMSD calculations on the pairwise interfaces where the average RMSD over the whole C5b–C9 complex was calculated to be 11.6 Å away from the native conformation. While most of the interfaces were predicted with only a minor deviation from the native conformation, two interfaces (C5b–C6 and C9–C9) increased the average RMSD value. Considering the fact



that the MAC is a flexible immune pore complex, this deviation might be either due to conformational differences in the cryoEM maps obtained from MAC complexes assembled from purified components *in vitro*, and the MAC enriched from pooled human plasma, or due to the lower number of detected XLs for these two specific interactions (Table 1).

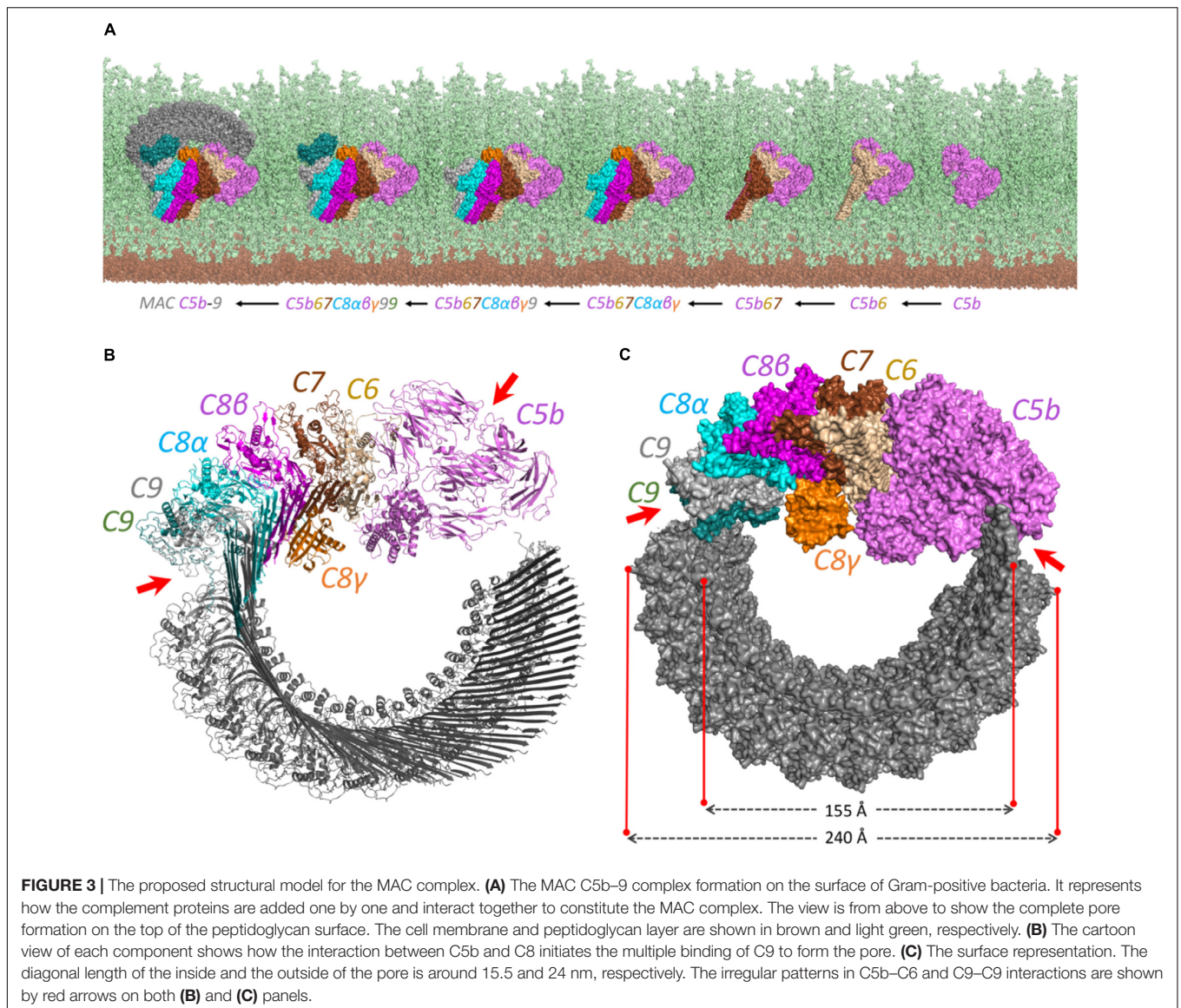
The produced pairwise models were further aligned and docked together to constitute the C5b–C9 complex, as indicated in Figure 2B. The predicted complex is supported by a total number of 126 intra- and inter-XLs, as listed in detail in Table 1. While the initial position of C6–C9 represents the formation of C5b–C9 pore, we further employed multiple copies of C9 protein to complete the predicted model of the MAC complex, as represented in Figure 3.

## DISCUSSION

Protein–protein interactions play crucial roles in biological systems yet remained challenging to study. XL-MS, coupled to macromolecular modeling, is a method capable of elucidating high-accuracy models of protein–protein interactions. We previously showed the power of TX-MS method, which is a combinatorial approach based on different MS acquisition

techniques and macromolecular modeling to study host–pathogen interactions (Hauri et al., 2019). Here, by representing a simplified and faster version of TX-MS, we applied this method to the many interfaces of the MAC complex while all the complement proteins detected to be highly abundant and associated to the bacterial surface based on DIA-MS analysis. We obtained an estimated average of eight cross-links per interface for the MAC complex, which allowed us to assemble a full model of MAC and compare it with available X-ray crystallography and electron cryo-microscopy models (Lovelace et al., 2011; Serna et al., 2016; Menny et al., 2018).

Here, we elucidate how *in vivo* cross-linking of human plasma to the surface of live bacteria can provide detailed information of protein–protein complexes in their native environment, with possible implications for the studies of conformational differences between *in vitro* crystallized and electron cryo-microscopically imaged and reconstructed macromolecular complexes and *in vivo* cross-linked ones. The approach detailed here should also open up new avenues for the future integration of XL-MS data with *in vivo* prokaryotic electron cryo-tomographic imaging and model reconstruction of host–pathogen complexes (Nans et al., 2014, 2015; Chowdhury et al., 2020).



## DATA AVAILABILITY STATEMENT

The authors confirm that all mass spectrometry mzML files underlying the findings in this study are available from the Zenodo.org database (accession doi: 10.5281/zenodo.4289947).

## AUTHOR CONTRIBUTIONS

HK, LH, and LM conceived the study. HK and LH wrote the manuscript. HK generated the computational models, analyzed and interpreted the MS data, and wrote the first draft. LH designed and performed the XL-MS sample preparation and analysis. LM, JM, and GT interpreted the data, contributed to the manuscript preparation, and approved the final draft. All authors have read and approved the final manuscript.

## FUNDING

This study was supported by Swiss National Science Foundation (grant no. P2ZHP3\_191289) to HK and GT and the Foundation of Knut and Alice Wallenberg (2016.0023 and 2019.0353) to JM and LM.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.612475/full#supplementary-material>

**Supplementary Figure 1 |** Detected intra- and inter-XLs were mapped on the experimentally solved structures (PDB ids: 6H04, 3OJY, 4A5W, and 2WII). The two structures above (C3b and C5b) shows how the intra XLs are mapped while structures below mainly represent inter-XLs mapping.



**Supplementary Figure 2 |** RMSD calculation between pairwise docking models and the native structures derived from protein data bank. The average RMSD for each pairwise interaction is calculated between all backbone atoms (CA, C, N, O) and is an average of five PyMOL alignment cycles. The average RMSD for

C5b-C6, C6-C7, C7-C8, C8 $\alpha$ -C8 $\beta$ , C8( $\alpha$ ,  $\beta$ )-C9, and C9-C9 interfaces were 16.06, 6.63, 2.16, 0.97, 15.19, and 28.76 Å, respectively. The outliers here are C5b-C6, and C9-C9 interfaces where we see a rotation on C5b and a shift in C9 resulting in high RMSD values (indicated with red arrows).

## REFERENCES

- Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. *Nature* 422, 198–207.
- Akesson, P., Sjöholm, A. G., and Björck, L. (1996). Protein SIC, a novel extracellular protein of streptococcus pyogenes interfering with complement function. *J. Biol. Chem.* 271, 1081–1088. doi: 10.1074/jbc.271.2.1081
- Berends, E. T. M., Dekkers, J. F., Nijland, R., Kuipers, A., Soppe, J. A., Strijp, J. A. G. v., et al. (2013). Distinct localization of the complement C5b-9 complex on gram-positive bacteria. *Cell. Microbiol.* 15, 1955–1968. doi: 10.1111/cmi.12170
- Chambers, M. C. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* 30, 918–920.
- Chowdhury, S., Happonen, L., Khakzad, H., Malmström, L., and Malmström, J. (2020). Structural proteomics, electron cryo-microscopy and structural modeling approaches in bacteria–human protein interactions. *Med. Microbiol. Immunol.* 209, 265–275. doi: 10.1007/s00430-020-00663-5
- Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., et al. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* 11:O111.016717.
- Gray, J. J. (2006). High-resolution protein–protein docking. *Curr. Opin. Struct. Biol.* 16, 183–193. doi: 10.1016/j.sbi.2006.03.003
- Guo, T., Kouvonen, P., Koh, C. C., Gillet, L. C., Wolski, W. E., Röst, H. L., et al. (2015). Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps. *Nat. Med.* 21, 407–413. doi: 10.1038/nm.3807
- Happonen, L., Hauri, S., Birkedal, G. S., Karlsson, C., Neergaard, T. d., Khakzad, H., et al. (2019). A quantitative streptococcus pyogenes–human protein–protein interaction map reveals localization of opsonizing antibodies. *Nat. Commun.* 10:2727.
- Hauri, S., Khakzad, H., Happonen, L., Telemán, J., Malmström, J., and Malmström, L. (2019). Rapid determination of quaternary protein structures in complex biological samples. *Nat. Commun.* 10:192.
- Herzog, F., Kahraman, A., Boehringer, D., Mak, R., Bracher, A., Walzthoeni, T., et al. (2012). Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry. *Science* 337, 1348–1352.
- Kashyap, D. R., Wang, M., Liu, L.-H., Boons, G.-J., Gupta, D., and Dziarski, R. (2011). Peptidoglycan recognition proteins kill bacteria by activating protein-sensing two-component systems. *Nat. Med.* 17, 676–683. doi: 10.1038/nm.2357
- Koehler Leman, J., Weitzner, B., Lewis, S., Adolf-Bryfogle, J., Alam, N., Alford, R. F., et al. (2020). Macromolecular modeling and design in rosetta: new methods and frameworks. *Nat. Methods* 17, 665–680.
- Liu, F., Rijkers, D. T. S., Post, H., and Heck, A. J. R. (2015). Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat. Methods* 12, 1179–1184. doi: 10.1038/nmeth.3603
- 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. doi: 10.1074/jbc.m111.219766
- Menny, A., Serna, M., Boyd, C. M., Gardner, S., Joseph, A. P., Morgan, B. P., et al. (2018). CryoEM reveals how the complement membrane attack complex ruptures lipid bilayers. *Nat. Commun.* 9:5316.
- Nans, A., Kudryashev, M., Saibil, H. R., and Hayward, R. D. (2015). Structure of a bacterial type III secretion system in contact with a host membrane in situ. *Nat. Commun.* 6:10114.
- Nans, A., Saibil, H. R., and Hayward, R. D. (2014). Pathogen–host reorganization during Chlamydia invasion revealed by cryo-electron tomography. *Cell. Microbiol.* 16, 1457–1472. doi: 10.1111/cmi.12310
- Nesvizhskii, A. I., and Aebersold, R. (2005). Interpretation of shotgun proteomic data. *Mol. Cell. Proteomics* 4, 1419–1440. doi: 10.1074/mcp.r500012-mcp200
- Ohue, M., Matsuzaki, Y., Uchikoga, N., Ishida, T., and Akiyama, Y. (2014a). MEGADOCK: an all-to-all protein–protein interaction prediction system using tertiary structure data. *Protein Pept. Lett.* 21, 766–778. doi: 10.2174/092986665113209990050
- Ohue, M., Shimoda, T., Suzuki, S., Matsuzaki, Y., Ishida, T., and Akiyama, Y. (2014b). MEGADOCK 4.0: an ultra-high-performance protein–protein docking software for heterogeneous supercomputers. *Bioinformatics* 30, 3281–3283. doi: 10.1093/bioinformatics/btu532
- Röst, H. L., Rosenberger, G., Navarro, P., Gillet, L., Miladinović, S. M., Schubert, O. T., et al. (2014). OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat. Biotechnol.* 32, 219–223. doi: 10.1038/nbt.2841
- Serna, M., Giles, J. L., Morgan, B. P., and Bubeck, D. (2016). Structural basis of complement membrane attack complex formation. *Nat. Commun.* 7:10587.
- Sinz, A. (2006). Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein–protein interactions. *Mass Spectrom. Rev.* 25, 663–682. doi: 10.1002/mas.20082
- Telemán, J., Chawade, A., Sandin, M., Levander, F., and Malmström, J. (2016). Dinosaur: a Refined open-source peptide MS feature detector. *J. Proteome Res.* 15, 2143–2151. doi: 10.1021/acs.jproteome.6b00016
- Zhang, B., Whiteaker, J. R., Hoofnagle, A. N., Baird, G. S., Rodland, K. D., and Paulovich, A. G. (2019). Clinical potential of mass spectrometry-based proteogenomics. *Nat. Rev. Clin. Oncol.* 16, 256–268. doi: 10.1038/s41571-018-0135-7
- Zundert, G. C. P. v., Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastrius, P. L., Karaca, E., et al. (2016). The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* 428, 720–725. doi: 10.1016/j.jmb.2015.09.014

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Khakzad, Happonen, Tran Van Nhieu, Malmström and Malmström. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.