crystals

MDPI

Article

Crystallographic and NMR Study of *Streptococcus pneumonia* LCP Protein Psr_{Sp} Indicate the Importance of Dynamics in Four Long Loops for Ligand Specificity

Tatyana Sandalova [1,2], Benedetta Maria Sala [1,2], Martin Moche [3], Hans-Gustaf Ljunggren [4], Evren Alici [5,6], Birgitta Henriques-Normark [7,8], Tatiana Agback [9], Dmitry Lesovoy [10,11], Peter Agback [9] and Adnane Achour [1,2,*]

1 Science for Life Laboratory, Department of Medicine, Solna, Karolinska Institute, SE-17165 Solna, Sweden; tatyana.sandalova@ki.se (T.S.); sala.benedettamaria@gmail.com (B.M.S.)
2 Division of Infectious Diseases, Karolinska University Hospital, SE-17176 Stockholm, Sweden
3 Department of Medical Biochemistry and Biophysics, Protein Science Facility, Karolinska Institute, SE-17177 Stockholm, Sweden; martin.moche@ki.se
4 Center for Infectious Medicine, Department of Medicine Huddinge, Karolinska Institute, SE-14152 Huddinge, Sweden; hans-gustaf.ljunggren@ki.se
5 Centre for Hematology and Regenerative Medicine, Department of Medicine Huddinge, Karolinska Institute, SE-14152 Huddinge, Sweden; evren.alici@ki.se
6 Hematology Centre, Karolinska University Hospital, SE-14152 Huddinge, Sweden
7 Department of Microbiology, Tumor and Cell Biology, Karolinska Institute, SE-17177 Stockholm, Sweden; birgitta.henriques@ki.se
8 Clinical Microbiology, Karolinska University Hospital Solna, SE-17176 Stockholm, Sweden
9 Department of Molecular Sciences, Swedish University of Agricultural Sciences, SE-75007 Uppsala, Sweden; tatiana.agback@slu.se (T.A.); peter.agback@slu.se (P.A.)
10 Swedish NMR Centre, University of Gothenburg, Box 465, SE-40530 Gothenburg, Sweden; lesovoydm@gmail.com
11 Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry RAS, 117997 Moscow, Russia
* Correspondence: adnane.achour@ki.se

check for updates

Citation: Sandalova, T.; Sala, B.M.; Moche, M.; Ljunggren, H.-G.; Alici, E.; Henriques-Normark, B.; Agback, T.; Lesovoy, D.; Agback, P.; Achour, A. Crystallographic and NMR Study of *Streptococcus pneumonia* LCP Protein Psr_{Sp} Indicate the Importance of Dynamics in Four Long Loops for Ligand Specificity. *Crystals* **2024**, *14*, 1094. https://doi.org/10.3390/cryst14121094

Academic Editor: Emilio Parisini

Received: 14 November 2024
Revised: 13 December 2024
Accepted: 17 December 2024
Published: 19 December 2024

Abstract: The crystal structure of the extracellular region of the second pneumococcal LCP, a polyisoprenyl-teichoic acid-peptidoglycan teichoic acid transferase Psr_{Sp}, was determined and refined to 2.15 Å resolution. Despite the low sequence homology with other LCP proteins, the Psr_{Sp} maintains the fold of the LCP domain, and the positions of the residues suggested to participate in the transferase function are conserved. The tunnel found in the Psr_{Sp} between the central β-sheet and three α-helices is wide enough to accommodate polyisoprenyl-teichoic acid. Comparison of the crystallographic temperature factors of LCP from distinct bacteria demonstrated that the four long loops located close to the teichoic acid and peptidoglycan binding sites have different relative mobilities. To compare the dynamics of the Psr_{Sp} in crystalline state and in solution, NMR spectra were recorded, and 88% of the residues were assigned in the ^{1}H-^{15}N TROSY HSQC spectra. Perfect accordance in the secondary structure of the crystal structure of Psr_{Sp} with NMR data demonstrated correct assignment. Moreover, the relative mobility of the essential loops estimated from the crystallographic B-factor is in good agreement with order parameter S^2, predicted from chemical shift. We hypothesize that the dynamics of these loops are important for the substrate promiscuity of LCP proteins.

Keywords: *S. pneumoniae*; LCP; cell wall teichoic acid; capsular polysaccharide; crystal structure; NMR; protein dynamics

1. Introduction

Glycopolymers in Gram-positive bacteria play a crucial role in the maintenance of the structural integrity and functionality of the cell wall, significantly contributing to the bacterial ability to interact with their environment and to resist external stresses [1–3]. In addition to the defense of the cell against the immune system of the host, bacterial

glycopolymers bind to a large ensemble of other cell surface proteins, regulating their functional activity. They also participate in transport processes, cation homeostasis, resistance to antimicrobial peptides and lysozymes, regulation of cell elongation and cell division, as well as many other additional processes [4]. Despite the varied composition of these glycopolymers across different bacteria, they share several key features in their structures and their biosynthesis [3].

Teichoic acids (TA) are important cell wall polymers present in almost all Gram-positive bacteria [5]. TAs are commonly divided into two classes; (i) lipoteichoic acid (LTA), which are anchored in the cytoplasmic membrane using a lipid anchor, and (ii) wall TA (WTA), which are covalently bound to the peptidoglycan (PG). Importantly, TA in *Streptococcus pneumoniae* is a major cell wall polymer with a unique identity, including several unusual sequence and structural features compared to other Gram-positive bacteria. First, the chemical composition of both LTA and WTA is identical in *S. pneumoniae*. Secondly, the *S. pneumoniae* TA is unusually complex and contains the rare positively charged amino sugar 2-acetamido-4-amino-2,4,6-trideoxy-D–galactose (AATGal) [6]. Third, pneumococcal TA contains phosphorylcholine attached to the N-acetyl-galactose-amine residues, which is a binding site for several choline-binding proteins [7]. Altogether, pneumococcal TA comprise four to eight repeating units (RU), all composed of AATGal, D-glucose, ribitol-phosphate, and two N-acetyl-galactose-amine residues modified by phosphorylcholine [8].

The biosynthesis of streptococcal TA RUs and their polymerization takes place on the cytoplasmic side of the bacterial cell membrane, and involves at least 18 different cytosolic and membrane-associated enzymes [8]. The polymerization of TA-RUs occurs on the membrane-bound C55-undecaprenyl lipid carrier (Figure S1). When the synthesis of these RUs is completed, the transmembrane transporter TacF presumably moves the C55-PP-linked TA chains across the membrane. The last step of the WTA synthesis is implemented by a specific set of transferases from the LytR-CpsA-Psr (LCP) enzyme family, which catalyzes the transfer of TA from the lipid carrier to the PG [9]. Most often, covalent linkage occurs via the formation of a phosphodiester bond to the C6-OH group of the N-acetylmuramic acid of the PG [10]. LCP transferases can catalyze not only TA attachment to the PG in *S. pneumoniae* but are also responsible for the linkage of many capsule polysaccharides to the PG, which plays a crucial role in the bacteria structure and pathogenicity [6,11]. The LCP proteins are transmembrane proteins with a catalytically active extracellular C-terminus and cytoplasmic N-terminus. They have been described as important potential antimicrobial targets for Gram-positive pathogens due to the importance of their multiple functions, as well as their conserved structures [11]. Several independent studies have revealed that a large array of different Gram-positive bacteria commonly have multiple copies of LCP enzymes that are seemingly functionally redundant. Three different genes encoding for LCP proteins have been hitherto identified in *Streptococcus pneumoniae*, *Streptococcus agalactiae*, *Staphylococcus aureus*, and *Bacillus subtilis* [11,12]. It was shown for *B. subtilis* that the deletion of a single LCP member had no or only a minor effect on cell growth under normal conditions. However, the removal of all three LCP proteins is lethal to the bacteria [12]. In contrast, it should be noted that the deletion of even a single LCP gene in *S. pneumoniae* or *S. agalactiae* results in a reduced capsule volume, significantly decreased viability over time, and reduced cell wall integrity [11,13].

The three-dimensional structures of several LCP molecules from a large ensemble of bacterial species have been previously determined, including LcpA from *S. aureus* (6eux.pdb) [14], Cps2A from *S. pneumoniae* (2xxq.pdb, 4de8.pdb, 3tep.pdb) [11,12], *S. agalactiae* (3okz.pdb), TagT (4de9.pdb, 6mps.pdb) [11,15], TagV (6uf3.pdb) and TagU (6uf6.pdb), all from *B. subtilis* [14]. Although the sequence identity among LCPs is low, ranging between 20 and 40% (Figure S2), the overall fold of the determined LCP domains is similar. The core of each determined LCP molecule has an α–β–α fold that contains a central β-sheet surrounded on both sides by several α-helices. Importantly, since the active site of different LCP proteins is surrounded by several flexible loops with different sequences, an enhanced

understanding of the dynamic behavior of the active site is, in our opinion, crucial for adequate drug discovery targeting LCP proteins and/or a specific LCP protein.

The crystal structure of the *S. pneumoniae* LCP protein Cps2A has been previously determined, revealing how it binds to polyisoprenyl–pyrophosphate—a crucial step for attaching capsular polysaccharides to the cell wall PG [12]. However, although the 3D-structure of Cps2A and LCP proteins from other bacteria provided fundamental insights into how this enzyme catalyzes the final steps of the polysaccharide attachment process, the molecular bases underlying the importance of each *S. pneumoniae* LCP for attaching TA or other glycopolysaccharides on PG remain unclear. Here, in an effort to enhance our understanding of these functions, we determined the crystal structure of the extracellular region of the second pneumococcal LCP, the polyisoprenyl-teichoic acid-peptidoglycan teichoic acid transferase Psr_{Sp}, (previously named MsrM), combining static high-resolution X-ray structural information with dynamic NMR analyses. We report the near-complete assignment of the ^{1}H, ^{13}C, and ^{15}N backbone resonances for the Psr_{Sp} domain, as well as the $^{13}C\beta$ side chain assignment of this protein. To resolve ambiguities in the crowded areas of the Psr_{Sp} LCP domain spectra, we compared the assigned resonances with those at different stages of deuterium exchange in the amide groups. Based on these data, the secondary structure of the Psr_{Sp} in solution was determined and compared with the crystal structure of the same protein, allowing us to evaluate the flexibility of this protein. Our results provide a structural basis for understanding the intra- and intermolecular interactions of the LCP domain of Psr_{Sp}.

2. Materials and Methods

2.1. Cloning, Expression and Purification of $Psr_{sp130-424}$

The sequence of the extracellular LCP domain of Psr_{Sp} (residues 130-424; UNIPROT Q8DPD6_STRR6) was cloned into the vector pET28 in frame with a N-terminal His-tag for NMR studies and in frame with a C-terminal His-tag for crystallization assays. Plasmids were transformed into the BL21 (DE3) pLys *E. Coli* strain, and cells were cultured at 37 °C in LB broth supplemented with 50 µg/mL kanamycin until an OD_{600} of 0.8 was reached. Protein transcription was then initiated by reaching a final concentration of 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG), which was added to the culture after reducing the culture temperature to 20 °C before overnight incubation. Cells were harvested by centrifugation and thereafter resuspended in a lysis buffer (20 mM Tris pH 7.5, 250 mM NaCl, 20 mM Imidazole) supplemented with a Complete protease inhibitor (Roche). Cells were lysed by sonication during a 10 min period with 30 s ON and 30 s OFF, with 30% of the maximal amplitude, and the cell debris were pelleted by centrifugation. The supernatant was loaded on a 5 mL HisTrap FF column (Cytiva) and after column wash, the protein construct $Psr_{Sp130-424}$ was eluted with 500 mM Imidazole and concentrated down to 5 mL with an Amicon Ultra centrifugal filter. The protein was further purified via size exclusion chromatography with a HiLoad 16/60 Superdex, 75 pg and eluted in 20 mM Tris pH 7.5, 150 mM NaCl as a monomeric peak. Peak fractions were pooled and concentrated to 25 mg/mL for crystallization assays. Protein purity was assessed using SDS-PAGE. For NMR studies, the buffer of $Psr_{Sp130-424}$ was exchanged into 25 mM NaPhosphate, 100 mM NaCl, and concentrated to 200 µM.

2.2. Crystallization and Structure Determination

Psr_{Sp} crystals were obtained at 20 °C using the sitting drop technique, mixing equal amounts of the protein Psr_{Sp} and the reservoir solution (25 mM NaPhosphate pH 6.8, 100 mM NaCl), in a 400 nL drop. Plate-like crystals of about 150 µm × 60 µm were cryo-protected by adding 20% ethylene glycol to the crystallization drop and thereafter flash-frozen in liquid nitrogen. Diffraction data were collected at the BioMax Beamline at the Max IV Laboratory (Lund, Sweden). The crystal diffraction was anisotropic and the integrated data were shared with the STARANISO web server (http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi, accessed in 20 October 2020) for elliptical truncation after data

processing in XDSAPP3 to 2.15 Å. We were able to determine the crystal structure of $Psr_{Sp130-424}$ by molecular replacement only by using a three-dimensional molecular model generated with AlphaFold [16] as a search template. Three molecules were present in the asymmetric unit of the crystal according to the calculated Matthews coefficient. The first two monomers were found with PHASER [17], while the third monomer was found by MOLREP [18], based on the PHASER output. Model building and refinement were performed with Phenix (Version 1.19.1-4122) [19], CCP4 [20] and Coot (Version 0.9) [21]. All structural figures as well as electrostatic potential calculations were performed using PyMOL (The PyMOL. Molecular Graphics System. Version 1.8 Schrödinger, LLC).

2.3. Expression of Isotope-Labeled Psr_{Sp} and Preparation of NMR Samples

The DNA sequence encoding for $Psr_{Sp130–420}$ and an N-terminal $MHis_6$ENLYFQ-tag was cloned into the pET28a expression vector (Novagen). The tag-Psr_{Sp} construct was transformed into the BL21 (DE3) pLys *E. coli* express competent cells, and the protein was thereafter expressed in different isotopic labeling combinations in $^{1/2}$H, ^{15}N, $^{12/13}$C-labeled M9 medium. Chemicals for isotope labeling including ammonium chloride, ^{15}N (99%), D-glucose, ^{13}C (99%) and deuterium oxide were purchased from Cambridge Isotope Laboratories. Psr_{Sp} for NMR experiments was expressed in 1 L D_2O M9 medium using 3 g/L of U-[^{13}C,^{2}H]-glucose (CIL, Andover, MA, USA) as the main carbon source, and 1 g/L of $^{15}NH_4Cl$ (CIL, Andover, MA, USA) as the nitrogen source. Bacterial growth was continued for 16 h at 16 °C, and cells were harvested by centrifugation. Cells were resuspended in a lysis buffer (20 mM Tris pH 7.5, 250 mM NaCl, 20 mM Imidazole) supplemented with a complete protease inhibitor (Roche) and lysed using an ultrasonicator, followed by centrifugation at 40,000× g for 30 min to remove cell debris. The purification steps were the same as for the construct used for the crystallographic study. The final $Psr_{sp130-424}$ protein sample was subsequently exchanged to a buffer composed of 25 mM sodium phosphate ($Na_2HPO_4{}^+NaH_2PO_4$) $PO_4{}^{3-}$, 100 mM NaCl, pH 6.8, 1 mM NaN_3, 10 (*v*/*v*) % D_2O, suitable for NMR experiments using gravity flow PD10 desalting columns (GE Healthcare). All NMR experiments were performed by adding 0.1 mM DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) as an internal ^{1}H chemical shift standard. Protein concentration was about 0.7 mM, and spectra were acquired in a 3 mm tube. ^{13}C and ^{15}N chemical shifts were referenced indirectly to the ^{1}H standard using a conversion factor derived from the ratio of NMR frequencies [22].

2.4. NMR Experiments

NMR experiments on U-[^{15}N,^{13}C,^{2}H], 75%U-[^{15}N,^{13}C,^{2}H] and U-[^{15}N,^{13}C] labeled samples of $Psr_{sp130-424}$ were acquired either on a 700 MHz Bruker Avance III spectrometer equipped with a 3 mm cryo-enhanced QCI-F probe, or on a 600 MHz Bruker Avance III spectrometer equipped with a 5 mm cryo-enhanced QCI-P probe. The experiments were performed at 308 K to improve relaxation parameters. Backbone resonance assignments for $Psr_{sp130-424}$ were obtained as previously described using the conventional strategy [23–25] and were based on a set of 3D TROSY or HSQC triple resonance experiments from the Bruker library. In summary, to increase resolution in the indirect dimensions, and to reduce acquisition time for the 3D experiments, the iterative non-uniform sampling protocol (NUS) [26] was used in the experiments comprising TROSY-HNCO, TROSY-HNCA and TROSY-HN(CO)CA, TROSY-HN(CA)CO, TROSY-HN(CO)CACB, TROSY HNCACB, H(CC)(CO)NH, TOCSY-^{15}N HSQC, HCACO and TOCSY-^{13}C HSQC experiments. Additionally, NOESY ^{15}N-HSQC, NOESY-^{13}C- HSQC spectra were collected [27–29]. The combined 3D NUS NMR target acquisition (TA) data were processed using the IST algorithm in the NUS module in TopSpin4.0.6 (Bruker, Billerica, MA, USA), and analyzed using CcpNmr2.4.2. [30] and Dynamics Center2.8 (Bruker, Billerica, MA, USA). Data submitted to the BioMagResBank with accession code BMRB ID is 52556. The chemical shifts (CS) of ^{1}H(N), ^{15}N, ^{13}Cα, ^{13}Cβ and ^{13}C′ nuclei for every amino acid of $Psr_{sp130-424}$ were analyzed with the TALOS-N 4.12 software [31] to extract the secondary structure prediction (S.S.

Prediction) index, random coil index (RCI) [32] order parameter S2, and confidence index. In case of no chemical shift, TALOS-N uses a database of sequences to predict the secondary structure. Standard nomenclature for amino acids of the carbon atoms was used in both the main text and the figures, where $^{13}C_{\alpha}$ is the carbon next to the carbonyl group $^{13}C'$ and $^{13}C_{\beta}$ is the carbon next to $^{13}C_{\alpha}$ [33]. Bruker standard pseudo 2D pulse sequence tractf3gpphwg was used to obtain a quick estimate of the rotational correlation time of the Psr_{sp} [34].

3. Results

3.1. The Extracellular Domain of Psr_{Sp} Is a Monomer with a Classical LCP Fold

The *S. pneumoniae* Psr_{Sp} has a short N-terminal cytoplasmic region (residues 1–90), a single transmembrane helix (comprising residues 91–114) and an extracellular catalytic LCP domain encompassing residues 115–428. The construct used within the present X-ray crystallography study of Psr_{Sp}, comprises almost the entire extracellular region of this protein (residues 130–424) and a C-terminal His-tag. Size exclusion and circular dichroism analyses demonstrated that Psr_{Sp} is a monomer with a molecular weight of 33 kDa and significant overall stability, with a melting temperature T_m of about 63 °C. The crystal of Psr_{Sp} belongs to the space group P321 (Table S1), and it contains, according to solvent content analysis [20], three Psr_{Sp} molecules in the asymmetric unit. The crystal structure of Psr_{Sp} was determined by molecular replacement. Interestingly, the program Phaser [17] was unable to find any solution with initial models from any of the previously determined crystal structures of LCP domains, including Cps2A, the first structurally determined member of the pneumococcal LCP (2xxq, [12] nor LcpA from *S. aureus* (6eux, [14]). However, two molecules of Psr_{Sp} were unambiguously identified by Phaser using a molecular model obtained from the AlphaFold server [16]. The third molecule in the asymmetric unit was found only upon using MolRep [18], after establishing the first two solutions with Phaser. The final crystal structure of Psr_{Sp} was refined to 2.15 Å resolution.

Although the three Psr_{Sp} molecules found in the asymmetric unit display highly similar three-dimensional structures, specific differences were identified in their backbones, testifying to the flexibility of this protein (Figure 1a). The superposition of molecules A and B on C resulted in root mean square deviation (rmsd) values of 0.5 and 0.49 Å, respectively. The main difference between the three molecules relates to the conformation of the two loops, A (residues 153–162) and B (residues 285–292), which are similar in molecules A and B but very different in molecule C (Figure 1a). In addition, the backbone conformation of residues 342–348 is different in molecule B compared to molecules A and C, due to close contacts with symmetry-related molecules in the crystal asymmetric unit.

All three subunits form crystallographic trimers around the three-fold axis of P_{321}. The 700 $Å^2$ interface between neighboring subunits is formed by the N-terminal stretches of residues 130–139 and 368–390. However, we believe that this hypothetical trimer formation mainly reflects a crystallographic artifact since Psr_{Sp} behaves as a monomer during purification. NMR provides an alternative way, compared to gel filtration analyses, to discriminate monomeric and multimeric protein forms in a solution, through the direct estimation of the rotational correlation time τ_C. The experimental τ_C can be compared with the values predicted using an empiric equation that takes into account the temperature and molecular weight [35]. Such estimates at 35 °C gave τ_C values of 15.6, 31.1, and 46.6 ns for the monomeric, dimeric, and trimeric forms of Psr_{Sp}, respectively. The former value agrees well with 14.7 ns obtained at 35 °C by using [^{1}H-^{15}N] TRACT type NMR experiment [34], thus confirming the predominantly monomeric form of Psr_{Sp} in solution.

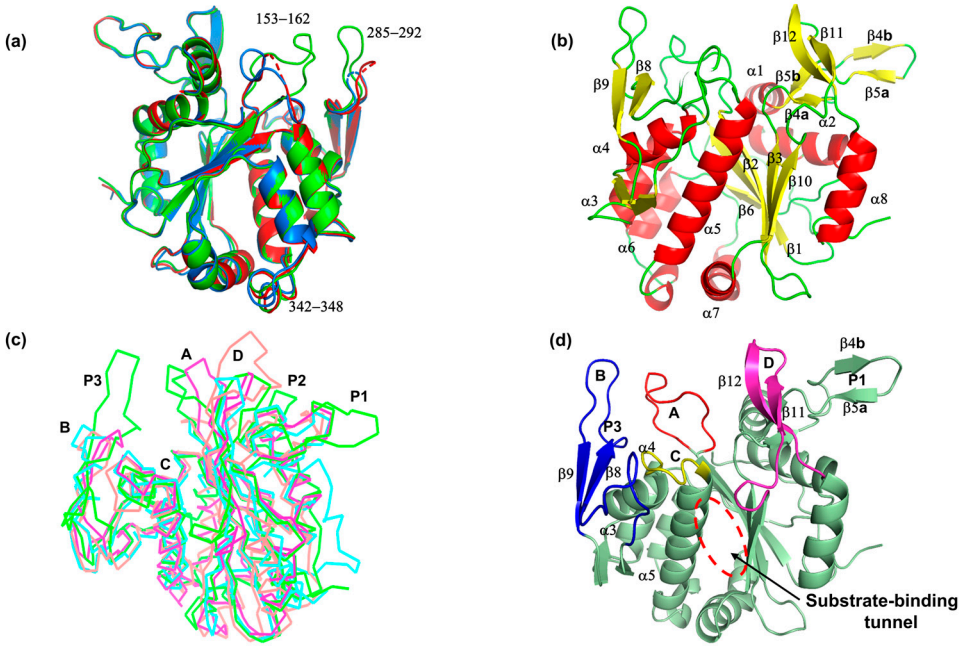

Figure 1. Crystal structure of the *S. pneumoniae*-associated Psr_{Sp}. (**a**). Superposition of the three Psr_{Sp} molecules found in the asymmetric unit of the crystal structure indicates flexibility in three different regions annotated by the corresponding residues. The three Psr_{Sp} molecules A, B and C found in the asymmetric unit are in red, blue, and green, respectively. (**b**). The crystal structure of Psr_{Sp} reveals a typical overall LCP fold despite low sequence homology to other members of the LCP protein family. (**c**). Superposition of Psr_{Sp} (in green) on the crystal structures of the three most similar LCP, $LcpA_{Sa}$ from *Staphylococcus aureus* (in blue, 6uex.pdb), TagT from *Bacillus subtilis* (in red, 6uf5.pdb) and Cps2A from *S. pneumoniae* (in orange, 4de8.pdb) demonstrates that the main difference between these LCP molecules resides in the length of the P1–P3 protrusions and in the conformations of four loops, that we named A, B, C, and D. (**d**). The A, B, C, and D regions of Psr_{Sp}, suggested to be important for substrate binding, are displayed in red, blue, yellow and magenta, respectively. The localization of the substrate binding tunnel is indicated.

3.2. Psr_{Sp} Takes a Classical LCP Fold with Some Unique Features

The central core of the extracellular LCP domain of Psr_{Sp} is composed of a five-stranded β-sheet (β6–β–β–β–β10) with the middle β2-strand running antiparallel to all other β-strands (Figure 1b). The central β-sheet is surrounded by eight α-helices, three of which, α1, α2, and α8, are located on one side of the β-sheet, and five additional α-helices on the other side. Analysis of the crystal structure of Psr_{Sp} using the Dali server [36] demonstrated the similarity of its fold to other previously determined LCP domain crystal structures. The structures that are most similar to Psr_{Sp} are the *S. aureus*-associated $LcpA_{Sa}$ (6uex.pdb) [14], TagT from *B. subtilis* (6mps.pdb) [15] and Cps2A from *S. pneumoniae* (4de8.pdb) [11]. The cores of these three proteins, consisting of 230 residues, align well to Psr_{Sp}, with rmsd values of 2.0 Å for $LcpA_{Sa}$, 2.7 Å for $TagT_{Bs}$, and 2.3 Å for Cps2A (Figure 1c). Superposition of these four proteins clearly demonstrates that despite their similarity in the overall fold, the sequence identity among LCP is not high, and pairwise alignment shows only 37% of identical residues between Psr_{Sp} and $LcpA_{Sa}$, and 21–22% with TagT and Cps2A. Multiple sequence alignment of the six proteins with the most similar three-dimensional structures to Psr_{Sp} confirms the overall low sequence homology among these members of the LCP family, with only 15 residues that are absolutely conserved (Figure S2). Superposition of the crystal structure of Psr_{Sp} with three other LCP proteins illustrates that there are three β-hairpin-like protrusions in Psr_{Sp}, which are not present in other LCP proteins (Figure S1 and Figure 1d). Protrusion P1 composed of the stretch of residues 188–206, with the two antiparallel β-strands β4a-β5b and β4b-β5a, separates the

β3-strand from the helix α1. Despite the insertion of nine residues, the conserved essential lysine residue K208 superposed very well with lysine residues on other LCP proteins (Figure S2). The relatively short protrusion P2 (residues 217–222) is located between helices α1 and α2. Finally, the longest protrusion P3 comprises residues 271–288 (Figure 1c). It should be noted that the length of these protruding chains is also significantly different among these LCP proteins.

3.3. The Substrate-Binding Site of Psr_{Sp} Comprises Common LCP Features

A wide and long empty tunnel is present in Psr_{Sp} between the central β-sheet and helices α3, α4, and α5 (Figure 1d). It should be noted that unlike $LcpA_{Sa}$ and TagT, Psr_{Sp} was expressed and crystallized without any endogenous hydrophobic undecaprenyl molecule. However, the compound C_{40}-PP-GlcNAc found in $LcpA_{Sa}$ can be easily modeled inside the tunnel of Psr_{Sp} by simple superposition of Psr_{Sp} onto $LcpA_{Sa}$. Small changes in the conformation of only three residues, D165, R314 and L337, are enough to remove all steric clashes for the appropriate binding of the ligand C_{40}-PP-GlcNAc docked to substrate-free Psr_{Sp} (Figure 2). While the side chains of residues D165 and R314 localized in the binding site have to move to allow adequate binding to the polar moiety of the substrate binding, the conformation of the side chain of residue L337 positioned on helix α5 has to be changed to allow binding to the lipid part of C_{40}-PP-GlcNAc. A detailed description of the substrate-binding site of $LcpA_{Sa}$ and probable mechanisms underlying the function of LCP transferases has been previously provided [12,14]. The present crystal structure demonstrates that the catalytic mechanism of Psr_{Sp} is most probably highly similar to other LCP proteins. Indeed, despite the low sequence identity within LCP proteins, all twelve amino acids that have been suggested as important for the teichoic acid transferase activity are conserved in Psr_{Sp} (Figures S1 and 2b,c). Though the residues that bind the pyrophosphate moiety of the ligand are highly conserved in LCP, the four loop-like regions that surround the active site are very different. Following previous descriptions [14], we will call these regions A (residues 152–164), B (267–298), C (315–322), and D (390–410) (Figures S1 and 1d). The B region of Psr_{Sp} includes the protrusion P3, which makes region B the longest among LPC proteins (Figure 1c,d). Regions A and C form loops that connect strands β1 and β2, and helices α4 and α5, respectively. Multiple sequence alignment (Figure S2) and the superposition of 3D structures (Figure 1c) demonstrate that the length and conformation of all these four regions are different among LCP proteins. Several residues suggested to be essential for the catalysis activity, such as D152, R163, D165 and S166, are located on the N- and C-termini of region A. The shorter region C is also located between the four essential residues R314 and R316 from helix α4, and D322 and R325 from helix α5. Regions B and D are the most divergent among LCPs, since they do not comprise any conserved residue, which can reflect the different chemical structure of TA from different bacterial species.

Calculation of electrostatic surface potentials demonstrates that the entrance to the tunnel contains positively charged residues in all LCP molecules from *S. pneumonia* as well as in $LcpA_{Sa}$ (Figure 3). A molecular model of the Psr_{Sp}/C_{40}-PP-GlcNac complex demonstrates that the five positively charged residues, R154, R186, R314, R316, and R325 are located at the entrance of the active site, ready to interact with the PP moiety of the TA-PP-polyisoprenyl molecule (Figure 2b). A comparison of the electrostatic potential on the surface of other LCP proteins emphasizes also the low conservation of surface residues outside the entrance (Figure 3). Psr_{Sp} is the most acidic within the LCP region compared to $LcpA_{Sa}$, TagT, or Cps2A. This is not surprising since the overall number of acidic residues in the LCP domain of Psr_{Sp} amounts to 41, with only 30 basic residues, which stands in sharp contrast to $LcpA_{Sa}$, which has 30 acidic and 37 basic residues. The main contribution to the overall acidity of the Psr_{Sp} surface is provided by regions B, C, and D, which contain five, five, and four negatively charged non-conserved residues, respectively. Since the WTA of *S. pneumoniae* contains the positively charged sugar AATGal, which must be connected

to the N-acetyl muramic acid (NAM) of PG, the presence of these acidic residues close to the active site is favorable for binding of WTA to Psr_{Sp}.

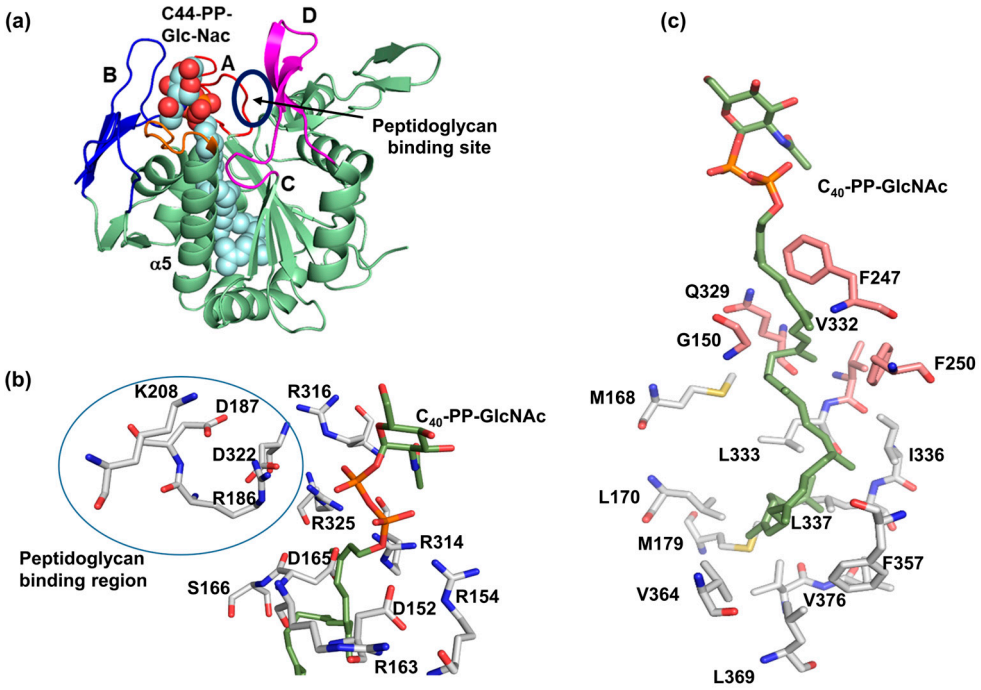

Figure 2. Superposition of $LcpA_{Sa}$/C_{40}-PP-GlcNAc onto ligand-free Psr_{Sp} reveals that the ligand fits well to the substrate-binding site of Psr_{Sp} and that all catalytic residues are located appropriately. (**a**). The C_{40}-PP-GlcNAc ligand found in the active site of $LcpA_{Sa}$, is displayed as light blue spheres inside the tunnel within Psr_{Sp}. The four regions A–D that are suggested to be important for substrate binding are colored in red, blue, orange and magenta, respectively. (**b**). LCP-conserved residues are located in Psr_{Sp} close to the pyrophosphate moiety of the ligand. (**c**). The hydrophobic part of the ligand is surrounded either by conserved residues (pink) or non-conserved but still hydrophobic residues (gray) of Psr_{Sp}.

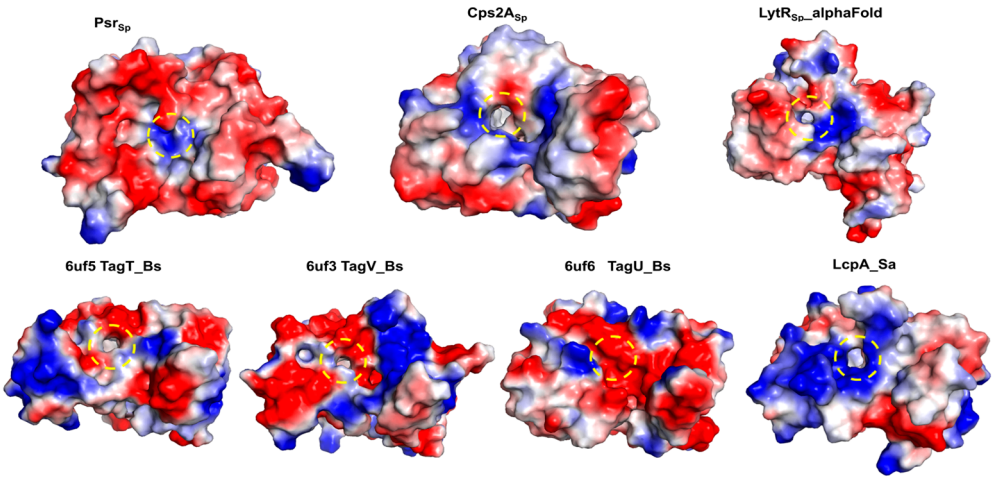

Figure 3. Comparison of electrostatic surface potentials in different LCP indicates different substrate preferences. The surfaces of LCP from different bacterial species, including Psr_{Sp}, $LcpA_{Sa}$, Cps2A, and $TagT_{Bs}$ are colored according to their electrostatic potential. Positively and negatively charged regions

are in blue and red, respectively. The entry of the tunnel for the ligand is indicated by a yellow arrowed circle. This structural comparison indicates differences in sizes at the entry of the tunnel. It also reveals that the entry of the tunnel is surrounded mainly by positively charged residues. Differences in size and in residue charge distribution can be related to different substrate preferences.

It is known that the function of LCP transferases is Mg-dependent [12]. The Mg^{2+} ion is not essential for substrate binding, since no ion has been found in some crystal structures of LCP proteins, even if these bound efficiently to the octaprenyl-pyrophosphate moiety of the ligand. However, the presence of Mg^{2+} significantly increases phosphotransferase activity [12]. The crystal structures of $TagT_{Bs}$ [15] and Cps2A [12] demonstrate that the Mg^{2+} ion binds to six oxygen atoms in perfect tetrahedral symmetry, with two oxygen atoms from the pyrophosphate moiety, two from water molecules, and two from the conserved aspartate residues. Comparison of the structures of the same protein with and without Mg^{2+} ion revealed that ion binding modifies the conformation not only of the two aspartates directly involved in the Mg^{2+} binding but also of several arginine residues that bind the pyrophosphate moiety. Thus, Mg^{2+} stabilizes the active conformation of the active site facilitating catalysis. Multiple sequence alignment as well as comparison of the crystal structures of Psr_{Sp} with Csp2A and TagT clearly illustrate that residues D152 and D165, located at the C-terminus of the β1-strand and at the N-terminus of the β2-strand, respectively, are Mg^{2+} binders (Figure S2).

The poly-isoprenyl-binding tunnel contains several LCP-conserved residues, including G150, F247, F250, Q329, and V332. Although other residues lining the tunnel, such as M168, L170, M179, L333, I336, L337, F357, V364, L369, V376, and L377 are not conserved among LCP proteins, they keep a hydrophobic nature within the tunnel that is favorable for binding to the undecaprenyl moiety (Figure 2c). In addition to the pocket aimed for the hydrophobic carrier bearing the TA, LCPs have a binding site for the second substrate, the PG. Based on comparative analyses of crystal structures of different LCPs, the PG-binding site was proposed to be localized between regions A, C, and D (Figure 2a,b), with the four conserved residues R163, D187, K208, and D322 that could interact with the PG moiety in Psr_{Sp}.

3.4. Differential Flexibility of the Four ABCD Regions Explains the Ability of LCPs to Adapt to Different TA and PG Substrates

The actual substrates of LCP enzymes are much longer than the compounds found in the LCP crystal structures that have been determined until now. To illustrate the relative size of the teichoic acid (TA), the peptidoglycan moiety (PG), and the LCP molecule, a hypothetical molecular model of the Psr_{Sp} complex with these two substrates was created (Figure S3). Our model clearly indicates that regions A and B could be important for TA binding and that region D is positioned close to the PG. Comparison of substrate-free LCPs with different LCP/substrate complexes revealed that the binding of the undecaprenyl-PP compound to LCPs does not require large conformational changes within the protein [12,14,15]. However, the importance of the mobility of regions A, B, C, and D for binding to the full-length LCP substrates remains unclear. Region A is very flexible in the substrate-free Psr_{Sp}, and the conformation of the stretch of residues 153–162 depends on the environment, since it is different in the three molecules found in the asymmetric unit of our Psr_{Sp} crystal (Figure 1a). In other ligand-free LCPs such as TagT (6uf5.pdb), region A is not visible in the electron density. High flexibility of the B-region also follows from the fact that residues 288–291 are not visible in the electron density in two out of the three chains of Psr_{Sp} in the asymmetric unit. Poor electron density represents an extreme case of flexibility, and more limited differences in the mobility of the protein within the crystal can be estimated by differences in temperature B-factors [37]. The analysis of the B-factor values along the protein chain can identify flexible regions in the protein. Figure 4 presents the crystal structures of several LCP proteins colored according to their B factors with rigid residues/sections (low B-factor) displayed in blue and flexible regions (higher B-factor) in red.

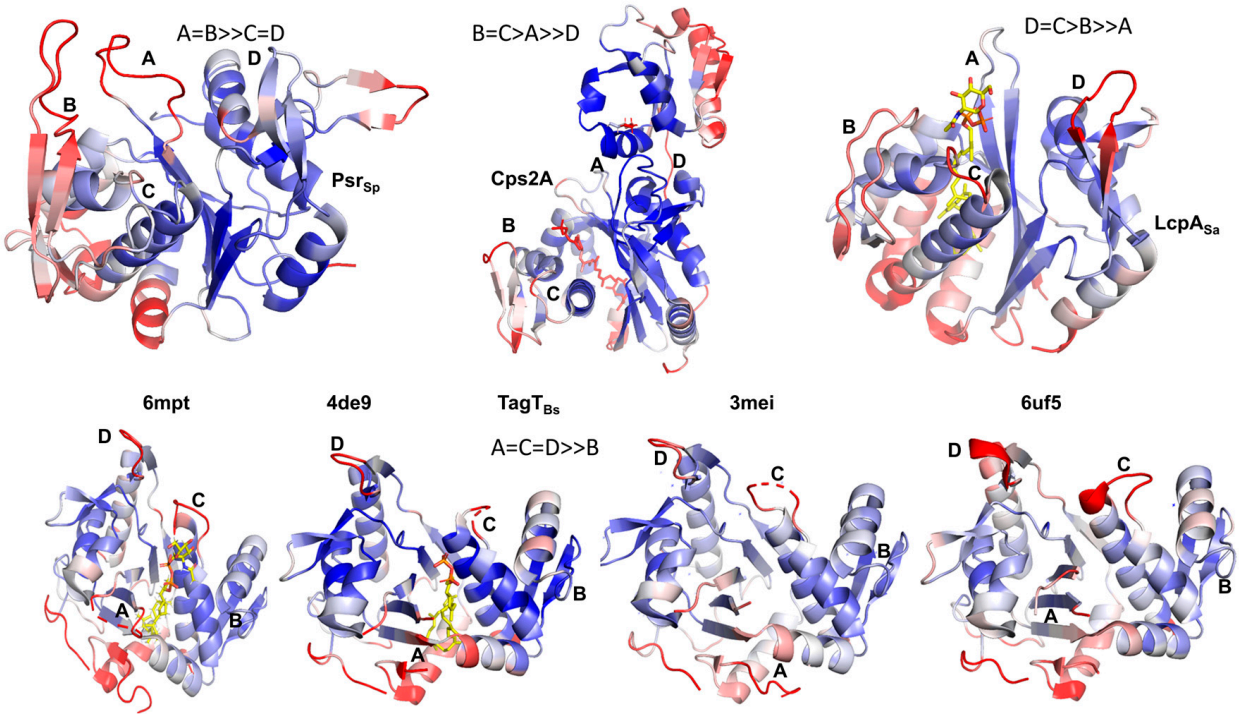

Figure 4. Comparison of crystallographic temperature B-factors in LCP crystal structures from different bacteria demonstrates differences in the mobility of specific loops. The rigid parts of each LCP protein are colored in blue, while the flexible parts are in red. This comparison allowed us to differentiate between the relative flexibility of the loops in the LCP proteins, allowing us to establish patterns. While loops A and B are significantly more flexible than C and D in Psr$_{Sp}$ (A = B >> C = D), loops B and C are more flexible than loop A, which in turn is much more flexible than loop D in Csp2A from *S. pneumoniae* (B = C > A >> D). In contrast, the LCP protein LcpA from *S. aureus* has a pattern in which D = C > B >> A».

This comparative analysis demonstrates the rigidity of the central core of the Psr$_{Sp}$ protein (comprising the central β-sheet and the helices α1, α2, α8. Regions C and D are rigid in Psr$_{Sp}$, while regions A and B are significantly more flexible. Interestingly, all three molecules within the asymmetric unit of the Psr$_{Sp}$ crystal have a similar distribution of temperature B-factors over their structures. Remarkably, similar analyses of other LCP molecules revealed different rankings of mobile regions (Figure 4). Region A is not very flexible compared to region D in Csp2A, whereas region A is the most rigid and region D is the most flexible in LcpA$_{Sa}$. A comparison of four different crystal structures of TagT$_{Bs}$, crystallized without ligands (PDB ID 3mei and 6uf5) and with undecaprenyl-PP (4de9) or with undecaprenyl-PP-sugar (6mpt), demonstrated that the presence of these ligands does not significantly alter the range of flexibility of the ABCD regions (Figure 4, lower panel), possibly due to a short sugar chain.

B-factor values characterize the flexibility of a protein in the crystalline state. Nuclear magnetic resonance (NMR) experiments can allow us to assess the mobility of protein domains in a solution. Therefore, in an effort to evaluate the dynamics of Psr$_{Sp}$ in a solution, we performed NMR experiments to assign the backbone nuclei of this LCP.

3.5. Successful Assignment of Backbone Psr$_{Sp}$ Resonances Allows Dynamic Studies

The validity of the secondary structure analysis based on chemical shifts (CS) depends on the knowledge of the chemical shifts (CS) of the ^{1}H(N), ^{15}N, ^{13}Cα, ^{13}Cβ, ^{13}C´ nuclei. Our crystal structure revealed that the secondary structure of Psr$_{Sp}$ is composed of a five-stranded β-sheet (β6-β1-β2-β3-β10) (Figure 1b), surrounded by eight α-helices and several potentially disordered regions. Folded and intrinsically disordered regions of proteins

exhibit different relaxation properties, requiring optimal NMR conditions tailored to each region. This necessitates finding the best sample preparation conditions to achieve the highest quality NMR experiments. Additionally, it is well established that the amide ^{1}H and ^{15}N chemical shifts in disordered regions and loops are highly sensitive to buffer conditions, including pH and temperature. To optimize data collection, we first performed NMR experiments at two different temperatures, 298 K and 308 K. We also conducted experiments in a physiological buffer at pH 6.8, with a low salt concentration. These conditions are optimal for future analysis of Psr_{Sp} protein interactions with ligands.

NMR data collection showed that the Psr_{Sp} sample remained stable at temperatures up to 313 K. No loss of signal intensity in the NMR spectra was detected over a period of two years. A notable issue was the recovery of amide protons in the U-[^{15}N,^{13}C,^{2}H]-labeled samples of Psr_{Sp}, which took us nearly a year to achieve full proton recovery in some cases. These amide protons belong mainly to amino acids involved in hydrogen bonds within the β-strand core of the protein, shielding them from water exposure. Several refolding processes were used to potentially overcome this problem, but this approach was ultimately not applicable since Psr_{Sp} did not fully fold into its starting conformation. Nevertheless, we used these U-[^{15}N,^{13}C,^{2}H]-labeled samples of Psr_{Sp} to assign backbone resonances in less crowded, simplified NMR spectra due to the absence of ^{2}H-N amide resonances. As a next step, we made use of 75% U-[^{15}N,^{13}C,^{2}H]-labeled samples of Psr_{Sp} to detect and assign the missing ^{1}H-^{15}N backbone resonances. As a control for the recovery of all ^{1}H-^{15}N amide cross peaks in the U-[^{15}N,^{13}C,^{2}H]-labeled samples of Psr_{Sp}, we additionally analyzed the 2D TROSY spectrum of these U-[^{15}N,^{13}C]-labeled Psr_{Sp} samples.

Since no program is available for the automatic assignment of large proteins with both folded and disordered regions, we used a conventional manual assignment strategy based on an approach described in the Section 2. To achieve the best resolution in the 3D experiments, especially for disordered regions, NMR experiments were mostly performed with the NUS option [26]. The ^{1}H-^{15}N TROSY HSQC spectrum at 308 K shows well-dispersed and narrow line widths for the amide signals for the 100% U-[^{15}N,^{13}C,^{2}H]-labeled samples of Psr_{Sp} (Figure 5). Under these conditions, with the back-exchange of ^{2}H to ^{1}H, we observed and assigned 259 out of 295 amino acids, including prolines (Figure 5d). In total, 79% ^{1}H(N) and ^{15}N excluding proline, 87% of ^{13}Cα, 93% of ^{13}Cβ and 85% of ^{13}C′ were assigned successfully. All ^{1}H, ^{15}N and ^{13}C chemical shifts in the LCP domain of Psr_{Sp} at pH 6.8 and at 308 K have been deposited in BioMagResBank (http://www.bmrb.wisc.edu) accessed on 13 August 2024 under the accession code BMRB ID 52556. Although the assignment of 36 Psr_{Sp} amino acid residues is still missing and distributed throughout the structure (marked in red in Figure 5d), we successfully assigned signals for the residues in the structural regions of interest. Backbone resonances were fully assigned in the B and D regions (Figure 5d), while all backbone resonances were also assigned for region A, except for the NH groups of residues D152 and S159. Finally, most of the backbone resonances were also assigned for the C region, Additionally, all LCP-conserved residues, which are located close to the polar moiety of the ligand (including D152, R154, R163, D165, S166, R186, D187, K208, R314, R316, and R325), were successfully assigned, except for D322. Similarly, most of the Psr_{Sp} hydrophobic residues that surround the hydrophobic part of the ligand (including M168, L170, M179, F247, F250, Q329, L333, I336, L337, and V364) were fully assigned. In contrast, the hydrophobic residues L369, V376, G150, V332, and F357 were not found in the spectra, likely due to incomplete recovery of amide protons.

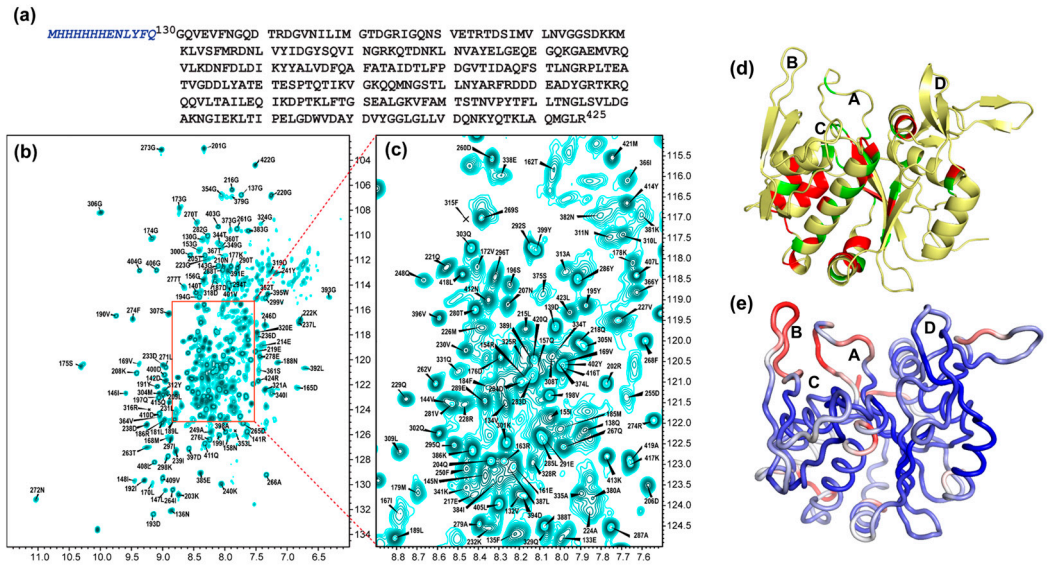

Figure 5. ^{1}H-^{15}N TROSY HSQC spectra of Psr$_{Sp}$ at T = 308 K with backbone amide assignments. (**a**). The amino acid sequence of the Psr$_{Sp}$ construct used in our NMR analyses is presented. Residues from the His-tag are indicated in italic blue. (**b**). ^{1}H-^{15}N TROSY HSQC spectrum at T = 308 K with its extended crowded part (**c**). from the ^{15}N, ^{13}C, ^{2}H labeled Psr$_{Sp}$ at 0.7 mM concentration in a 25 mM sodium phosphate buffer ($Na^2HPO_4{}^+NaH_2PO_4$) $PO_4{}^{3-}$, pH 6.8 containing 100 mM NaCl, 1 mM NaN_3, 10 (v/v) % D_2O and 0.1 mM DSS. The chemical shift assignment of NH backbone is shown by the number and amino acid symbols corresponding to the sequence. The assignment is presented only for residues which have cross-peaks observed at T = 308 K. (**d**). Residues of Psr$_{Sp}$ with fully assigned peaks in the NMR spectrum are displayed in yellow. Carbon-assigned residues are shown in green, and non-assigned residues in red. (**e**). The overall structure of Psr$_{Sp}$ colored in red–white–blue according to the predicted order parameter S^2 with flexible regions in red, and rigid regions in blue.

3.6. The Flexibility of the Psr$_{Sp}$ Regions in Solution Is in Agreement with Crystallographic B-Factors

The chemical shifts in the backbone of the full-length Psr$_{Sp}$ protein were analyzed using the TALOS N program [31] (Figure S4). Our results demonstrate that the secondary structure analysis of the folded domain of Psr$_{Sp}$ from our NMR data closely matches the crystal structure of Psr$_{Sp}$, validating our resonance assignments. Twelve β-strands and seven α-helices were identified in the NMR-derived secondary structure. Small variations at the beginning and end of the α and β structural elements of Psr$_{Sp}$ can be attributed to environmental differences, including a significant temperature difference of approximately 100 K between our NMR measurements and the crystallographic conditions. Notably, the fully unassigned region between residues 242–258, corresponding to the β6 strand and α3 helix in the crystal structure, showed no amide signals in the NMR spectra due to incomplete deuterium exchange. However, TALOS N predicted the secondary structure of this region based only on its sequence, and it matched the expected structure (Figure S4). Next, we focused on the regions A–D, where amino acid assignments were either fully complete (regions B and D) or nearly complete (regions A and C) (Figure S4). Notably, some residues in region B (285–292) could not be observed in the crystal structure of Psr$_{Sp}$, rendering our NMR data the primary source of structural information for these segments. In regions B and D, the short β-strands β8–β9 and β11–β12, were identified in both the crystal structure and the NMR secondary structure (Figure S4). Region A is predicted by TALOS-N as a random coil element, which is also consistent with the results from the crystal structure.

To further investigate the conformational flexibility in regions A–D of Psr$_{Sp}$, NMR spectroscopy offers detailed, site-specific insights into protein motions across various time

scales. Traditionally, the measure of protein dynamics using NMR involves experimental data like NOEs, T1, and T2, which are interpreted using model-free parameters [38,39]. These parameters describe the spatial and temporal motion of N-H(N) amide bonds in the protein backbone, with order parameters (S^2) ranging from 0 to 1. An S^2 value between 0.8 and 1.0 indicates limited motion, while correlation times describe motions on the ps-ns timescale. A simplified alternative method based on the backbone of ^{15}N and ^{13}C nuclei chemical shifts has been previously proposed [32] and thereafter implemented in TALOS [31]. This enables the accurate, site-specific mapping of protein backbone mobility without the need for extensive NMR relaxation experiments beyond standard backbone assignments. This approach, known as the Random Coil Index (RCI) [32], predicts the model-free backbone order parameters S^2. In the present study, these predicted order parameters are presented in Figure S4 and mapped on the three-dimensional structure of Psr_{Sp} (Figure 5e).

For secondary structural elements like β-strands and α-helices, the predicted order parameters S^2 are typically high, around 0.8, indicating their rigidity. However, in regions A–D of Psr_{Sp}, differences in flexibility are evident. Region D consists of a twisted antiparallel beta sheet formed between the β11–β12 strands, connected by a loop between residues 400–405. The predicted S^2 values for the amide bonds in both the β sheet and the loop are close to 0.8, indicating low mobility on the ps-ns timescale. This suggests that region D has a stable conformation, which may play a role in substrate recognition and selectivity in the Psr_{Sp} protein. Region C, characterized by a sharp turn between helices α4 and α5, has predicted S^2 values ranging from 0.7 to 0.8. Although these values are still relatively high, the prediction confidence from TALOS analysis is slightly lower due to the incomplete assignment in this region compared to the fully assigned region D. Nonetheless, the rigidity of the C-turn might be crucial for maintaining the correct positioning of α4 and α5 in the active site.

A more complex conformational behavior is observed in region B, which spans the longest chain of residues (262–304) and has the most complete NMR assignment. Consistent with both X-ray and NMR data, the secondary structure of this region includes two antiparallel β-sheets, formed by the β7–β9a and β8–β9 strands, with higher predicted S^2 values, indicating rigidity. Region B also contains two coil elements. The first is a loop between residues 286–294 connecting the β8–β9 strands, with low S^2 values, which suggests high flexibility. This loop is part of the substrate recognition site in the protein, and its conformational plasticity allows it to accommodate different substrates through a conformational selection mechanism. It is however possible that the entire element, including the β8–β9 strands and the loop, experiences slow dynamics, which were not fully explored in this study.

An interesting discrepancy between the X-ray and NMR data is observed in the region comprising residues 267–275. This stretch of residues connecting the β8–β9a strands is disordered in the crystal structure. However, the β7 strand is extended by five amino acids in the NMR structure, with higher S^2 values, followed by a short turn between residues 271–275, where the S^2 value decreases to 0.6. These differences could be due to crystal contacts since residues 267–275 are close to the neighboring molecule in the crystal. Region A, which corresponds to a loop that connects the β1 and β2 strands, does not contain any secondary structure elements, as confirmed by both X-ray and NMR data. The predicted S^2 values drop to as low as 0.4, indicating its conformational heterogeneity and plasticity. We suggest that the close spatial proximity of the flexible loop in region A to the similarly flexible loop in region B is key to the capacity of Psr_{Sp} to accommodate various substrate shapes within the active site.

4. Discussion

LCP proteins are attractive therapeutic targets since these enzymes catalyze a reaction essential for Gram-positive bacteria, resulting in the attachment of a secondary cell wall glycopolymer to the peptidoglycan. Inhibition of this process disrupts cell wall assembly,

cell division, and pathogenicity. Importantly, since LCP enzymes are located on the extracellular part of the bacterial cell, inhibitors of LCP enzymes are not required to cross the membrane which renders the creation of antimicrobial compounds significantly easier. Here, we describe the crystal structure and NMR dynamic analyses of Psr_{Sp}, the second LCP enzyme from *S. pneumoniae* that catalyzes attachment of wall teichoic acid as well as capsular polysaccharides to the PG. The crystal structure of Psr_{Sp} reveals a typical LCP fold, despite low sequence homology. A comparison with several other LCP structures confirmed that substrate binding and functional mechanisms are the same as in all other LCP proteins. The conservation of the residues participating in the catalysis suggests a common mechanism for this reaction. Thus, all the hitherto gathered structural and mutational data from different LCP reveal a common binding mode of the pyrophosphate-lipid carrier moiety by LCP transferases. However, it should be noted that the real substrates of LCP enzymes are much larger than the ligands found in crystal structures. Moreover, their chemical composition is different not only for different bacteria but even within the same organism. For example, there are almost 100 serotypes described for *S. pneumoniae*, with different capsular polysaccharides on their surfaces [40]. Many of these are covalently linked to the PG in the same way as for WTA. Though it has still not been demonstrated experimentally that LCP transferases catalyze the reaction for WTA and all capsular polysaccharides, LCP genetic and bioinformatics suggest that it is the case.

Most of the bacteria contain three genes codings for three different LCP enzymes. Differences in the electrostatic potential on the surface of the three LCP from *S. pneumoniae* (Figure 3) could reflect the different specificity of each LCP. It has been shown that all three LCP from *S. aureus* can catalyze the attachment of the staphylococcal WTA to the PG [10]. However, LcpA is the most important enzyme among these three LCPs for the attachment of WTA [15]. Interestingly, another LCP enzyme, LcpC, is the most important for the PG attachment of the type 5 capsular polysaccharide in this bacterium [41]. Similarly, all three LCP proteins contribute to the catalysis of the reaction between PG and secondary glycopolymers in *S. pneumoniae* [11]. However, we still do not know the variation nor the basis in the preference of the catalysis performed by the different LCP for different polysaccharides.

We hypothesize here that the mobility of the loops interacting with the polysaccharide could be of the greatest importance to create a promiscuous binding. The connection between flexibility of proteins and promiscuity of their interactions is well-established for intrinsically disordered proteins [42,43]. Interactions via flexible regions of separate molecules, which become ordered during assembly, allows the same protein to interact with multiple partners. A similar relation was found for several enzymes when increased flexibility of their active sites resulted in increased numbers of possible substrates [44,45]. Two methods were used in the present study to estimate the mobility of the essential regions of Psr_{Sp}. First, flexibility was estimated using B-factor values obtained for the crystal structure of Psr_{Sp}. Second, flexibility was also estimated for Psr_{Sp} in a solution from predicted order parameter S^2 obtained from chemical shifts in the NMR spectra.

There was consistency in the secondary structure elements obtained from the crystal structure and NMR experiments, despite a temperature difference of 200 °C between the conditions used in the different experimental procedures, which is evidence for the stability of the Psr_{Sp} molecule and for the correct assignment of the peaks in our ^{1}H-^{15}N TROSY HSQC spectrum. In this study, we found that the chemical shifts predicted by ShiftX2 for rigid secondary structures in proteins closely match those obtained through NMR, particularly for the $^{13}C_{\alpha}$, $^{13}C_{\beta}$, and $^{13}C'$ nuclei. This finding suggests that it may not always be necessary to aim for a full assignment of all parts of very large proteins. In large protein production, deuteration is essential but often problematic because fully recovering the amide protons is difficult, expensive, and can lead to protein loss. Since amide recovery mainly affects the most stable parts of the protein, the results from the present study indicate that it might be more efficient for large proteins to focus on recovering amide

protons only from flexible regions. This approach could simplify the analysis by reducing spectral crowding.

Our results reveal that loops A and B in Psr_{Sp}, responsible for interactions with WTA or capsular polysaccharides, are significantly more mobile compared to loop D, which is suggested to be involved in the interaction with the PG. This is not true for the other LCP protein Cps2A from *S. pneumoniae*. Analysis of the B-factors in the crystal structure of Cps2A shows that this LCP has a very rigid region A (Figure 5) and a short flexible region B, which contrasts significantly with Psr_{Sp}. Furthermore, the staphylococcal protein $LcpA_{Sa}$ has a rigid region A, a B region with some flexibility, and highly mobile C and D regions (Figure 4). We find it interesting that the number of serotypes and the chemical variety of capsular polysaccharides is much smaller in *S. aureus* compared to *S. pneumonia* [46,47]. One of the many remaining questions is thus, how can this flexibility be converted to substrate specificity?

Unfortunately, we do not have enough biological data on the substrate specificity nor on the promiscuity of the three different LCP phosphotransferases. It has been previously shown that all three LCP molecules in *S. pneumoniae* contribute to the maintenance of the normal volume of polysaccharides on the cell surface [11]. Deletion of Cps2A is not defective in capsule attachment [48]. While a Cps2A mutant of the strain D39 of *S. pneumoniae* produces only 30–40% of the capsule, a Psr_{Sp} mutant produces about 70% of the capsule [11]. D39 has a serotype 2 capsule, with its repeating units formed by glucose, glucuronic acid, and rhamnose. Although the previous results of Eberhardt et al. demonstrate that Cps2A is more active with a serotype 2 capsule compared to Psr_{Sp}, it still does not provide a definitive answer in regard to which protein is active with more different serotypes.

5. Conclusions

LCP proteins are an important antibiotic target. The crystal structure of the extracellular region of the second pneumococcal LCP, the polyisoprenyl-teichoic acid-peptidoglycan teichoic acid transferase Psr_{Sp}, can help to understand the substrate specificity of pneumococcal LCP. The present study, based on a combination of X-ray crystallography and NMR analyses, revealed that regions A and B in Psr_{Sp} are significantly more flexible compared to other parts of this key LCP protein. We speculate that the flexibility of the A and B loops could be important, providing the LCP Psr_{Sp} with the possibility to catalyze PG attachment for a large array of different glycopolymers. Identification of the most promiscuous LCP enzyme could be of great importance for the development of future antibiotics.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/cryst14121094/s1, Figure S1. Teichoic acid biosynthesis pathway suggested for *S. pneumonia* [8]. Figure S2: Multiple sequence alignment of the LCP domain of Psr_{Sp} with the five best structural hits found by DALI demonstrates low sequence conservation; Figure S3: Hypothetical molecular model of the undecaprenol-WTA/PG/Psr_{Sp} complex; Figure S4: The secondary structures of Psr_{Sp} and predicted order parameter obtained from NMR are in agreement with the crystal structure of Psr_{Sp}; Table S1: Data collection and refinement statistics for Psr_{Sp}.

Author Contributions: Conceptualization, A.A. and T.S.; methodology, B.M.S., T.S. and T.A.; software, M.M., P.A. and D.L.; validation, T.S., M.M. and T.A.; formal analysis, P.A., D.L. and B.M.S.; investigation, B.M.S., T.S., D.L., T.A. and P.A.; resources, B.M.S.; data curation, T.A., T.S. and M.M.; writing—original draft preparation, T.S., T.A. and A.A.; writing—review and editing, P.A., T.A., H.-G.L., E.A., B.H.-N. and A.A.; visualization, T.S. and T.A.; supervision, A.A.; project administration, A.A.; funding acquisition, A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Swedish Foundation for Strategic Research grant ITM17-0218 to TA and PA, grant RSF 23-44-10021 to D.M.L, Swedish Cancer Society (21 1605 Pj 01 H), Cancer och Allergi Fonden (10399), and Swedish Research Council (2021-05061 and 2018-02874) to AA.

Data Availability Statement: The crystal structure coordinates and structural factors for Psr_{Sp} have been deposited to the Protein Data Bank under accession code 9GN3. Assignment data were submitted to the BioMagResBank with accession code BMRB ID is 52556.

Acknowledgments: We gratefully acknowledge access to the synchrotron beam line BioMax at MAXIV (Lund, Sweden) and to the crystallization facility within the Protein Science Facility, Karolinska Institutet, (http://ki.se.proxy.kib.ki.se/en/mbb/protein-science-facility, accessed on 1 September 2020).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Rajagopal, M.; Walker, S. Envelope Structures of Gram-Positive Bacteria. *Curr. Top. Microbiol. Immunol.* **2017**, *404*, 1–44. [CrossRef] [PubMed]
2. Schade, J.; Weidenmaier, C. Cell wall glycopolymers of Firmicutes and their role as nonprotein adhesins. *FEBS Lett.* **2016**, *590*, 3758–3771. [CrossRef] [PubMed]
3. Stefanović, C.; Hager, F.F.; Schäffer, C. LytR-CpsA-Psr Glycopolymer Transferases: Essential Bricks in Gram-Positive Bacterial Cell Wall Assembly. *Int. J. Mol. Sci.* **2021**, *22*, 908. [CrossRef]
4. Weidenmaier, C.; Peschel, A. Teichoic acids and related cell-wall glycopolymers in Gram-positive physiology and host interactions. *Nat. Rev. Microbiol.* **2008**, *6*, 276–287. [CrossRef] [PubMed]
5. Sewell, E.W.; Brown, E.D. Taking aim at wall teichoic acid synthesis: New biology and new leads for antibiotics. *J. Antibiot.* **2014**, *67*, 43–51. [CrossRef]
6. Vollmer, W.; Massidda, O.; Tomasz, A. The Cell Wall of *Streptococcus pneumoniae*. *Microbiol. Spectr.* **2019**, *7*, 1–25. [CrossRef] [PubMed]
7. Aceil, J.; Avci, F.Y. Pneumococcal Surface Proteins as Virulence Factors, Immunogens, and Conserved Vaccine Targets. *Front. Cell Infect. Microbiol.* **2022**, *12*, 832254. [CrossRef]
8. Denapaite, D.; Brückner, R.; Hakenbeck, R.; Vollmer, W. Biosynthesis of teichoic acids in *Streptococcus pneumoniae* and closely related species: Lessons from genomes. *Microb. Drug Resist.* **2012**, *18*, 344–358. [CrossRef]
9. Hübscher, J.; Lüthy, L.; Berger-Bächi, B.; Stutzmann Meier, P. Phylogenetic distribution and membrane topology of the LytR-CpsA-Psr protein family. *BMC Genom.* **2008**, *9*, 617. [CrossRef]
10. Schaefer, K.; Matano, L.M.; Qiao, Y.; Kahne, D.; Walker, S. In vitro reconstitution demonstrates the cell wall ligase activity of LCP proteins. *Nat. Chem. Biol.* **2017**, *13*, 396–401. [CrossRef]
11. Eberhardt, A.; Hoyland, C.N.; Vollmer, D.; Bisle, S.; Cleverley, R.M.; Johnsborg, O.; Håvarstein, L.S.; Lewis, R.J.; Vollmer, W. Attachment of capsular polysaccharide to the cell wall in Streptococcus pneumoniae. *Microb. Drug Resist.* **2012**, *18*, 240–255. [CrossRef] [PubMed]
12. Kawai, Y.; Marles-Wright, J.; Cleverley, R.M.; Emmins, R.; Ishikawa, S.; Kuwano, M.; Heinz, N.; Bui, N.K.; Hoyland, C.N.; Ogasawara, N.; et al. A widespread family of bacterial cell wall assembly proteins. *EMBO J.* **2011**, *30*, 4931–4941. [CrossRef] [PubMed]
13. Rajaei, A.; Rowe, H.M.; Neely, M.N. The LCP Family Protein, Psr, Is Required for Cell Wall Integrity and Virulence in. *Microorganisms* **2022**, *10*, 217. [CrossRef] [PubMed]
14. Li, F.K.K.; Rosell, F.I.; Gale, R.T.; Simorre, J.P.; Brown, E.D.; Strynadka, N.C.J. Crystallographic analysis of *Staphylococcus aureus* LcpA, the primary wall teichoic acid ligase. *J. Biol. Chem.* **2020**, *295*, 2629–2639. [CrossRef] [PubMed]
15. Schaefer, K.; Owens, T.W.; Kahne, D.; Walker, S. Substrate Preferences Establish the Order of Cell Wall Assembly in *Staphylococcus aureus*. *J. Am. Chem. Soc.* **2018**, *140*, 2442–2445. [CrossRef] [PubMed]
16. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef] [PubMed]
17. McCoy, A.J.; Grosse-Kunstleve, R.W.; Adams, P.D.; Winn, M.D.; Storoni, L.C.; Read, R.J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, *40*, 658–674. [CrossRef] [PubMed]
18. Vagin, A.; Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 22–25. [CrossRef]
19. Adams, P.D.; Afonine, P.V.; Bunkóczi, G.; Chen, V.B.; Davis, I.W.; Echols, N.; Headd, J.J.; Hung, L.W.; Kapral, G.J.; Grosse-Kunstleve, R.W.; et al. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 213–221. [CrossRef]
20. Agirre, J.; Atanasova, M.; Bagdonas, H.; Ballard, C.B.; Baslé, A.; Beilsten-Edmands, J.; Borges, R.J.; Brown, D.G.; Burgos-Mármol, J.J.; Berrisford, J.M.; et al. The CCP4 suite: Integrative software for macromolecular crystallography. *Acta Crystallogr. D Struct. Biol.* **2023**, *79*, 449–461. [CrossRef]
21. Emsley, P.; Lohkamp, B.; Scott, W.G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 486–501. [CrossRef]
22. Wishart, D.S.; Bigam, C.G.; Yao, J.; Abildgaard, F.; Dyson, H.J.; Oldfield, E.; Markley, J.L.; Sykes, B.D. 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* **1995**, *6*, 135–140. [CrossRef] [PubMed]

23. Agback, P.; Lesovoy, D.M.; Han, X.; Sun, R.H.; Sandalova, T.; Agback, T.; Achour, A.; Orekhov, V.Y. H-1, C-13 and N-15 resonance assignment of backbone and IVL-methyl side chain of the S135A mutant NS3pro/NS2B protein of Dengue II virus reveals unique secondary structure features in solution. *Biomol. Nmr Assign.* **2022**, *16*, 135–145. [CrossRef]
24. Unnerstale, S.; Nowakowski, M.; Baraznenok, V.; Stenberg, G.; Lindberg, J.; Mayzel, M.; Orekhov, V.; Agback, T. Backbone Assignment of the MALT1 Paracaspase by Solution NMR. *PLoS ONE* **2016**, *11*, e0146496. [CrossRef]
25. Agback, T.; Dominguez, F.; Frolov, I.; Frolova, E.I.; Agback, P. 1H, 13C and 15N resonance assignment of the SARS-CoV-2 full-length nsp1 protein and its mutants reveals its unique secondary structure features in solution. *PLoS ONE* **2021**, *16*, e0251834. [CrossRef] [PubMed]
26. Orekhov, V.; Jaravine, V.A. Analysis of non-uniformly sampled spectra with multi-dimensional decomposition. *Progress. Nucl. Magn. Reson. Spectrosc.* **2011**, *59*, 271–292. [CrossRef]
27. Kay, L.E.; Ikura, M.; Tschudin, R.; Bax, A. 3-Dimensional Triple-Resonance Nmr-Spectroscopy of Isotopically Enriched Proteins. *J. Magn. Reson.* **1990**, *89*, 496–514. [CrossRef]
28. Kay, L.E.; Keifer, P.; Saarinen, T. Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy with Improved Sensitivity. *J. Am. Chem. Soc.* **1992**, *114*, 10663–10665. [CrossRef]
29. Schleucher, J.; Schwendinger, M.; Sattler, M.; Schmidt, P.; Schedletzky, O.; Glaser, S.J.; Sorensen, O.W.; Griesinger, C. A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *J. Biomol. NMR* **1994**, *4*, 301–306. [CrossRef] [PubMed]
30. Vranken, W.F.; Boucher, W.; Stevens, T.J.; Fogh, R.H.; Pajon, A.; Llinas, M.; Ulrich, E.L.; Markley, J.L.; Ionides, J.; Laue, E.D. The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins* **2005**, *59*, 687–696. [CrossRef]
31. Shen, Y.; Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. Nmr* **2013**, *56*, 227–241. [CrossRef]
32. Berjanskii, M.V.; Wishart, D.S. A simple method to predict protein flexibility using secondary chemical shifts. *J. Am. Chem. Soc.* **2005**, *127*, 14970–14971. [CrossRef]
33. Markley, J.L.; Bax, A.; Arata, Y.; Hilbers, C.W.; Kaptein, R.; Sykes, B.D.; Wright, P.E.; Wuthrich, K. Recommendations for the presentation of NMR structures of proteins and nucleic acids--IUPAC-IUBMB-IUPAB Inter-Union Task Group on the standardization of data bases of protein and nucleic acid structures determined by NMR spectroscopy. *Eur. J. Biochem.* **1998**, *256*, 1–15. [CrossRef]
34. Lee, D.; Hilty, C.; Wider, G.; Wüthrich, K. Effective rotational correlation times of proteins from NMR relaxation interference. *J. Magn. Reson.* **2006**, *178*, 72–76. [CrossRef]
35. Cavanagh, J.; Fairbrother, W.; Palmer III, A.; Rance, M.; Skelton, N. *Principles and Practice: Protein NMR Spectroscopy*; Elsevier Academic Press: Amsterdam, The Netherlands, 2007; ISBN 9780121644918.
36. Holm, L.; Laiho, A.; Törönen, P.; Salgado, M. DALI shines a light on remote homologs: One hundred discoveries. *Protein Sci.* **2023**, *32*, e4519. [CrossRef]
37. Sun, Z.; Liu, Q.; Qu, G.; Feng, Y.; Reetz, M.T. Utility of B-Factors in Protein Science: Interpreting Rigidity, Flexibility, and Internal Motion and Engineering Thermostability. *Chem. Rev.* **2019**, *119*, 1626–1665. [CrossRef] [PubMed]
38. Lipari, G.; Szabo, A. Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules. 2. Analysis of Experimental Results. *J. Am. Chem. Soc.* **1982**, *104*, 4559–4570. [CrossRef]
39. Lipari, G.; Szabo, A. Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *J. Am. Chem. Soc.* **1982**, *104*, 4546–4559. [CrossRef]
40. Paton, J.C.; Trappetti, C. *Streptococcus pneumoniae* capsular polysaccharide. *Microbiol. Spectr.* **2019**, *7*, 1–15. [CrossRef] [PubMed]
41. Chan, Y.G.; Kim, H.K.; Schneewind, O.; Missiakas, D. The capsular polysaccharide of *Staphylococcus aureus* is attached to peptidoglycan by the LytR-CpsA-Psr (LCP) family of enzymes. *J. Biol. Chem.* **2014**, *289*, 15680–15690. [CrossRef] [PubMed]
42. Tompa, P.; Schad, E.; Tantos, A.; Kalmar, L. Intrinsically disordered proteins: Emerging interaction specialists. *Curr. Opin. Struct. Biol.* **2015**, *35*, 49–59. [CrossRef] [PubMed]
43. Piovesan, D.; Monzon, A.M.; Quaglia, F.; Tosatto, S.C.E. Databases for intrinsically disordered proteins. *Acta Crystallogr. D Struct. Biol.* **2022**, *78*, 144–151. [CrossRef] [PubMed]
44. McAuley, M.; Kristiansson, H.; Huang, M.; Pey, A.L.; Timson, D.J. Galactokinase promiscuity: A question of flexibility? *Biochem. Soc. Trans.* **2016**, *44*, 116–122. [CrossRef]
45. Gatti-Lafranconi, P.; Hollfelder, F. Flexibility and reactivity in promiscuous enzymes. *Chembiochem* **2013**, *14*, 285–292. [CrossRef]
46. O'Riordan, K.; Lee, J.C. Staphylococcus aureus capsular polysaccharides. *Clin. Microbiol. Rev.* **2004**, *17*, 218–234. [CrossRef] [PubMed]
47. Visansirikul, S.; Kolodziej, S.A.; Demchenko, A.V. Staphylococcus aureus capsular polysaccharides: A structural and synthetic perspective. *Org. Biomol. Chem.* **2020**, *18*, 783–798. [CrossRef]
48. Morona, J.K.; Morona, R.; Paton, J.C. Attachment of capsular polysaccharide to the cell wall of Streptococcus pneumoniae type 2 is required for invasive disease. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8505–8510. [CrossRef]