Received: 23 April 2019 | Revised: 12 June 2019 | Accepted: 11 July 2019

DOI: 10.1002/pro.3693

ARTICLE

THE PROTEIN SOCIETY WILEY

Streptococcus pneumoniae G5 domains bind different ligands

Natasia Paukovich[1] | Jasmina S. Redzic[1] | Ying-Chih Chi[1] | Jeremy T. Rahkola[2] | Aaron Issaian[1] | Ashley Blue[3] | Kirk C. Hansen[1] | Edward N. Janoff[2] | Elan Z. Eisenmesser[1]

[1]Department of Biochemistry and Molecular Genetics, School of Medicine, University of Colorado Denver, School of Medicine, Aurora, Colorado

[2]Mucosal and Vaccine Research Program Colorado (MAVRC) & Denver Veterans Affairs Medical Center, Aurora, Colorado

[3]National High Magnetic Field Laboratory, Tallahassee, Florida

Correspondence
Elan Z. Eisenmesser, Department of Biochemistry and Molecular Genetics, School of Medicine, University of Colorado Denver, School of Medicine, Aurora, CO.
Email: elan.eisenmesser@ucdenver.edu

Funding information
Division of Materials Research, Grant/Award Number: 1644779; National Science Foundation, Grant/Award Number: 1807326

Abstract

Many bacterial pathogens express small G5 domains that exist in the context of various membrane-anchored proteins and these G5 domains have been associated with colonization, cellular adhesion, and biofilm formation. However, despite over a decade since the computational prediction of these G5 domains, many remain uncharacterized, particularly those from *Streptococcus pneumoniae*. Of five previously predicted G5 domains we found that four of these, all derived from *S. pneumoniae*, are independently folded modules. As one of these exhibits extreme line broadening due to self-association, we were able to use NMR solution studies to probe the potential ligand interactions of the remaining three G5 domains. None of these G5 domains engage N-acetylglucosamine (NAG) as previously predicted but do interact with other small molecules that may modulate adherence to both bacteria and host cells. Specifically, while all G5 domains tested engage Zn, only one of these G5 domains engage heparin. NMR solution structural studies of the IgA1 Protease G5 (IgA1P-G5) and endo-beta-N-acetylglucosaminidase-D G5 (ENDD-G5) also facilitated identification of the ligand binding sites and confirm the typical G5 fold that comprises two connected β-sheets with no canonical core. NMR relaxation experiments indicate flexibility on both ends and within the connecting regions between the β-sheets. Our studies thus establish a basis for future biological experiments to test whether the ligands presented here are involved in bacterial adherence, either to bacteria or to host cells.

KEYWORDS

adherence, colinization, G5 domain, *streptococcal pneumoniae*

1 | INTRODUCTION

Infectious bacterial pathogens rely on mechanisms of adherence for subsequent colonization and for biofilm formation.[1] Beyond the obvious implications for blocking such adherence to block infection, biofilms have become a growing problem in hospitals where they serve as natural reservoirs for many infectious pathogens and in particular have become a systemic problem with the advent of medical devices.[2] For example, infections, and particularly recurring infections, are attributed to biofilm formation in patients with *Streptococcus pneumoniae* (*S. pneumoniae*) that become resistant to antimicrobial agents.[3] *S. pneumoniae* is the primary cause of community-acquired pneumonia in both the United States and Europe[4,5] and remains the leading cause of pneumonia and meningitis in children worldwide.[6] Thus, understanding the early events that lead to colonization and biofilm formation may have wide implications for preventing this infection.

Natasia Paukovich and Jasmina S. Redzic contributed equally to this work.

Many bacterial pathogens, such as *S. pneumoniae*, comprise numerous small modules called G5 domains implicated in cellular adhesion, colonization, and biofilm formation.[7] G5 domains were initially grouped based on the presence of five conserved glycine residues (G5) and were proposed to target N-acetylglucosamine (NAG) on the cell wall of neighboring bacteria and within glycosaminoglycans on host cells.[7] G5 domains generally exist within the context of large membrane-anchored proteins on the bacterial surface and may exist as single domains or highly similar repeats. Several G5 domains from *S. epidermidis* and *S. aureus* have been shown to self-associate in a Zn-dependent manner, thereby expanding the repertoire of G5 interactions.[8,9] However, these G5 domains are found within large tandem arrays of 5–17 repeats and little is known about those G5 domains that exist within just a single copy. For *S. pneumoniae*, there are five different membrane-anchored proteins that contain G5 domains, four of which comprise only one copy and are under investigation here (Figure 1a). While the NMR solution structure of the other *S. pneumoniae* protein comprising two tandem copies was solved several years ago, no specific interaction was tested.[10]

Here, we sought to characterize the single copy, membrane-anchored G5 domains initially identified from the original G5 bioinformatics study that include four potential *S. pneumoniae* G5 domains and one from vancomycin resistant *Enteroccoccus faecium*.[7,11] Our results indicate that the G5 domain from vancomycin resistant *Enteroccoccus faecium* is not competent in folding, however, the remaining four G5 modules derived from *S. pneumoniae* proteins are independently-folded modules. Interestingly, none of the folded *S. pneumoniae* G5 modules specifically engage NAG, as previously predicted, but do exhibit a range of small molecule interactions. Thus, our studies indicate that G5 domains cannot be broadly defined as NAG binding modules and that the specific interactions of each G5 domain are unique.

2 | RESULTS

2.1 | Single copy G5 domains from *S. pneumoniae* are well-folded

Based on the previous bioinformatics identification of G5 domains by Bateman et al.,[7] we expressed a total of five G5 domains that were predicted to be single copy modules within the context of larger membrane-anchored proteins. Four of these G5 domains are derived from *S. pneumoniae* G5 modules (Figure 1a) and include the IgA1 Protease (IgA1P-G5), endo-beta-N-acetylglucosaminidase-D (ENDD-G5), beta-galactosidase (BGAA-G5), and *S. pneumoniae* collagen-like protein (SPR1403-G5). The fifth G5 module expressed in this study is derived from Enteroccoccus faecium,[11] which is a vancomycin-resistant bacteria (VanW-G5). The sequence comparison of all 5 G5 domains are shown here (Figure 1b). All five G5 domains were initially expressed as soluble polypeptides and were ^{15}N-labeled in order to assess their chemical shift dispersion using ^{15}N-heteronuclear single quantum coherence (HSQC) spectra.

As G5 modules exhibit a uniquely conserved β-strand structure with a surprising absence of a canonical core,[9,12,13] these domains are predicted to exhibit broad chemical shift dispersion due to the predominance of β-sheets, as shown for the two tandem G5 domains in StrH.[10] Indeed, all four *S. pneumoniae* G5 domains exhibit broad chemical shift dispersion (Figure 1c-f), although SPR1403-G5 exhibits similar dispersion only at elevated temperatures (further described below). However, the VanW-G5 exhibits a very narrow dispersion indicative of an unfolded protein (Figure 1g).

As the *S. pneumoniae* G5 domains give rise to well dispersed ^{15}N-HSQC spectra and several G5 domains from other bacteria exhibit metal-dependent self-association,[8,9] we questioned whether these modules could self-associate even in the absence

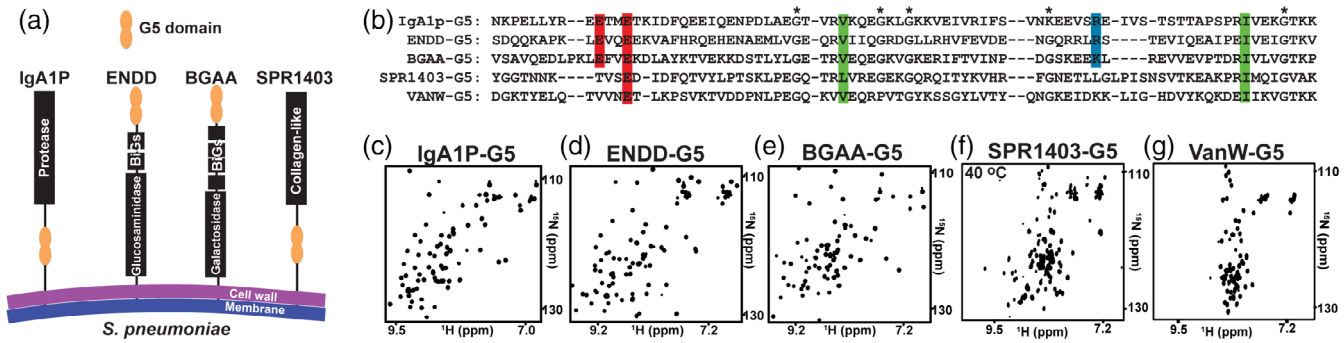

FIGURE 1 Single copy *Streptococcus pneumoniae* G5-containing proteins give rise to well-dispersed spectra. (a) Cartoon representation of the four *S. pneumoniae* proteins for which their G5 modules have been expressed in this study. (b) Sequence alignment of *S. pneumoniae* IgA1P-G5, ENDD-G5, BGAA-G5, SPR1403-G5 and *Enteroccoccus faecium* VanW-G5 with conserved acidic (red), basic (blue), and hydrophobic (green) residues, and the original delineation of 5 glycine residues that are in fact not all conserved (*). ^{15}N-HSQC spectrum of the (c) IgA1p-G5, (d) ENND-G5, (e) BGAA-G5, (f) SPR1403-G5, and (g) VanW-G5. Spectra were recorded at 900 MHz at 25°C, excluding SPR1403-G5 that exhibits extreme line-broadening at such lower temperatures (see Figure 2)

of ligands. SPR1403-G5 is especially prone to self-association and only gives rise to a well dispersed ^{15}N-HSQC at elevated temperatures (Figure S1a), while IgA1P-G5 and BGAA-G5 only moderately self-associate, as indicated by concentration-dependent chemical shifts (Figure S1b,c). In contrast, ENDD-G5 exhibits no such concentration-dependent shifts, suggestive of no self-association (Figure S1d). The resonance assignments and structural elucidation of IgA1P-G5 facilitates identification of those residues involved in its moderate self-association, which are largely confined to one region of this module (Figure S1e,f).

2.2 | *S. pneumoniae* G5 modules exhibit both similarities and differences in their solution structures

Given that all four G5 modules exhibit NMR spectra that are suggestive of folded domains, we further interrogated the high-resolution solution structure of two of the four G5 modules in order to confirm their typical G5 folds. We chose to further investigate IgA1P-G5, considering IgA1 proteases are in general critical virulence factors,[14] and ENDD-G5, which shows no concentration-dependent self-association (Figure S1). Structure calculations included both chemical shifts and long-distance NOE distance restraints that were utilized in RASREC CS-Rosetta, which resulted in two β-sheets and the absence of a conical core (Figure 2a, Tables 1, 2). The overall architecture of these domains is typical of G5 modules with two distinct β-sheets formed by three strands that each comprise an N-terminal and C-terminal β-strand with the third strand woven within the center of the first and second. However, there are differences within the sizes of these secondary structure elements, as the β-sheet near the N-terminus is larger within the IgA1P-G5 (Figure 2b), while the β-sheet near the C-terminus is larger within ENDD-G5 (Figure 2c). Both intervening regions between the two β-sheets are similar in regard to their lengths that each comprise 3–4 residues.

2.3 | *S. pneumoniae* G5 modules exhibit localized dynamics

To address the potential similarities and differences between G5 domain dynamics, we collected standard R1 relaxation rates, R2 relaxation rates, and heteronuclear NOEs for both IgA1P-G5 (Figure 3a) and ENDD-G5 (Figure 3b). While all of these rates are sensitive to the overall tumbling of the molecule, their individual rates are sensitive measures of local dynamics as well. For example, R1 relaxation rates and heteronuclear NOEs monitor ps–ns motions while R2 relaxation rates are sensitive to both ps–ns and μs–ms motions. For IgA1P-G5, average R1 relaxation rates are 1.46 ± 0.16 Hz, R2 relaxation rates are 14.64 ± 3.54 Hz, and heteronuclear NOEs are 0.74 ± 0.19. For ENDD-G5, average R1 relaxation rates are 1.63 ± 0.2 Hz,

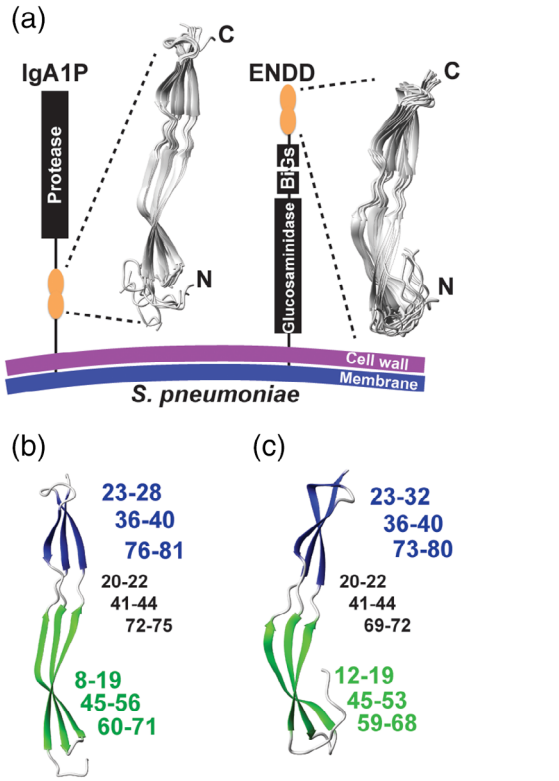

FIGURE 2 *Streptococcus pneumoniae* G5 domain structures from IgA1P-G5 and ENDD-G5. (a) Cartoon representation of the *S. pneumoniae* IgA1P and ENDD with 10 solution structural ensembles calculated from chemical shifts and NOE distance restraints of the IgA1P-G5 and ENDD-G5 (Tables 1 and 2, respectively). (b) Both G5 β-sheets are colored (green and blue) with the corresponding residues for each, numbered in the same color and the intervening coils numbered (in black). (c) Both ENDD-G5 β-sheets are colored (green and blue) with the corresponding residues numbered the same as in b

R2 relaxation rates are 9.32 ± 1.91 Hz, and heteronuclear NOEs are 0.67 ± 0.12. All three data sets were also collected for the IgA1P-G5 at a lower concentration, in order to distinguish potential differences in relaxation due to intermolecular interactions rather than internal dynamics, which suggest that intermolecular interactions do not contribute (Figure S2).

Relaxation data indicate that regions at both ends of the G5 domains are inherently flexible on multiple timescales with at least several residues within the connecting regions between the two β-sheets also exhibiting slower timescale motions. Specifically, for R2 relaxation rates, high rates are indicative of localized μs–ms timescale motions from chemical exchange while, in contrast, lower rates such as those within the N-termini of both G5 domains are due to elevated ps–ns timescale motions. Considering that R1 relaxation rates and heteronuclear NOEs report on ps–ns motions (see below), we focused first on slower motions that contribute to higher R2 relaxation rates. These elevated R2 relaxation rates include IgA1P-G5 Y10, D30, and residues 57–58 and ENDD-G5 Q7, residues 30–34,

TABLE 1 Structural Statistics for the IgA1P-G5 RASREC Rosetta Structures

No. of residues	84
NOE distance restraints	472
Avg violation per structure	37±3
Avg violation distance(Å)	0.59±0.04
Intra-residue	133
Medium-range (li-jl<4)	206
Long-range (li-jl>5)	133
Average rmsd (residues 7-84)	
Backbone (Å)	1.14±0.44
Heavy atom (Å)	1.57±0.49
Ramachandran plot summary (%)	
Most favored regions	87.0
Allowed regions	12.6
Generously allowed regions	0.4
Disallowed regions	0.0

G42, and 61–64 (Figure 3, top panels). Most of these residues reside in loop regions that either include the termini of both G5 domains or are adjacent to these termini. However, ENDD-G5 G42 is located within the connecting region of the two β-sheets and IgA1P-G5 residues 41–42 were the only amides within both G5 domains that we were unable to assign, potentially suggestive of chemical exchange. For ps–ns motions, we utilized R1 relaxation rates and heteronuclear NOEs as a proxy for localized motions. We note further in the Discussion that many of the slow R1 relaxation rates observed are likely due to locally elevated ps–ns motions based on the relatively small size of these G5 domains that places them close to the extreme narrowing limit (as opposed to larger proteins where faster R1 relaxation rates would be synonymous with elevated ps–ns motions). Slow R1 relaxation rates and small heteronuclear NOEs indicate the existence of localized ps–ns dynamic regions at predominantly loop regions that cap the ends of both G5 domains (Figure 3a,b middle panels). As the IgA1P-G5 does weakly self-associate (Figure S1), R2 relaxation rates, R1 relaxation rates, and heteronuclear NOEs were also collected at a lower concentration for IgA1P-G5 (Figure S2), which are similar to those described in the main text.

TABLE 2 Structural Statistics for the ENDD-G5 RASREC Rosetta Structures

No. of residues	81
NOE distance restraints	361
Avg violation per structure	35±3
Avg violation distance(Å)	0.55±0.07
Intra-residue	97
medium-range (li-jl<4)	150
long-range (li-jl>5)	114
average rmsd (residues 7-81)	
backbone (Å)	1.65±0.65
heavy atom (Å)	2.15±0.65
Ramachandran plot summary (%)	
most favored regions	87.5
allowed regions	12.5
generously allowed regions	0.0
disallowed regions	0.0

Additionally, the R2/R1 ratio can be used for proteins within the extreme narrowing limit to estimate local correlation times predictive of internal dynamics,[15] which highlights several regions that comprise exchange contributions (μs–ms motions). Several of these residues are highlighted for both G5 domains and mapped onto the NMR structures, which identifies several additional residues within the connecting region between the two β-sheets.

2.4 | *S. pneumoniae* G5 modules selectively bind ligands but not NAG

The recombinant production of *S. pneumoniae* G5 domains facilitates the direct assessment of whether they bind specific ligands previously proposed. We exploited the high sensitivity of NMR to identify the *S. pneumoniae* G5 interactions of the two modules that do not self-associate (ENDD-G5 and BGAA-G5) and the one that self-associates only weakly (IgA1P-G5). Because resonances were assigned for both IgA1P-G5 and ENDD-G5 for a full structure and dynamics characterization above, we were also able to identify the specific sites by chemical shift perturbation (CSP) analysis, but we used the unassigned spectrum of BGAA-G5 to also determine whether ligands specifically engage this G5 domain.

No CSPs were observed within any of the three G5 domains when NAG was titrated at stoichiometric concentrations (Figure 4a), which was surprising considering that NAG was proposed to be the primary target of G5 modules.[7] Additional titrations with the cell wall component choline and purified cell wall polysaccharides also did not induce any CSPs within any of these G5 modules (data not shown). However, as host cell glycosaminaoglycans (GAGs) are potential targets for bacterial cell adhesion and form gradients in the extracellular matrix,[16] we also tested whether these G5 domains bind a highly expressed representative GAG member, heparin. Indeed, heparin does induce CSPs for IgA1P-G5 (Figure 4b, top panel) and binds weakly, with a dissociation constant (K_d) of 0.64 ± 0.043 mM (Figure 4d), although heparin did not induce CSPs for either ENDD-G5 and BGAA [Figure 4b, middle and bottom panels]. In contrast, Zn induces CSPs for all tested G5 domains (Figure 4c, S3b–e), although the relative affinities

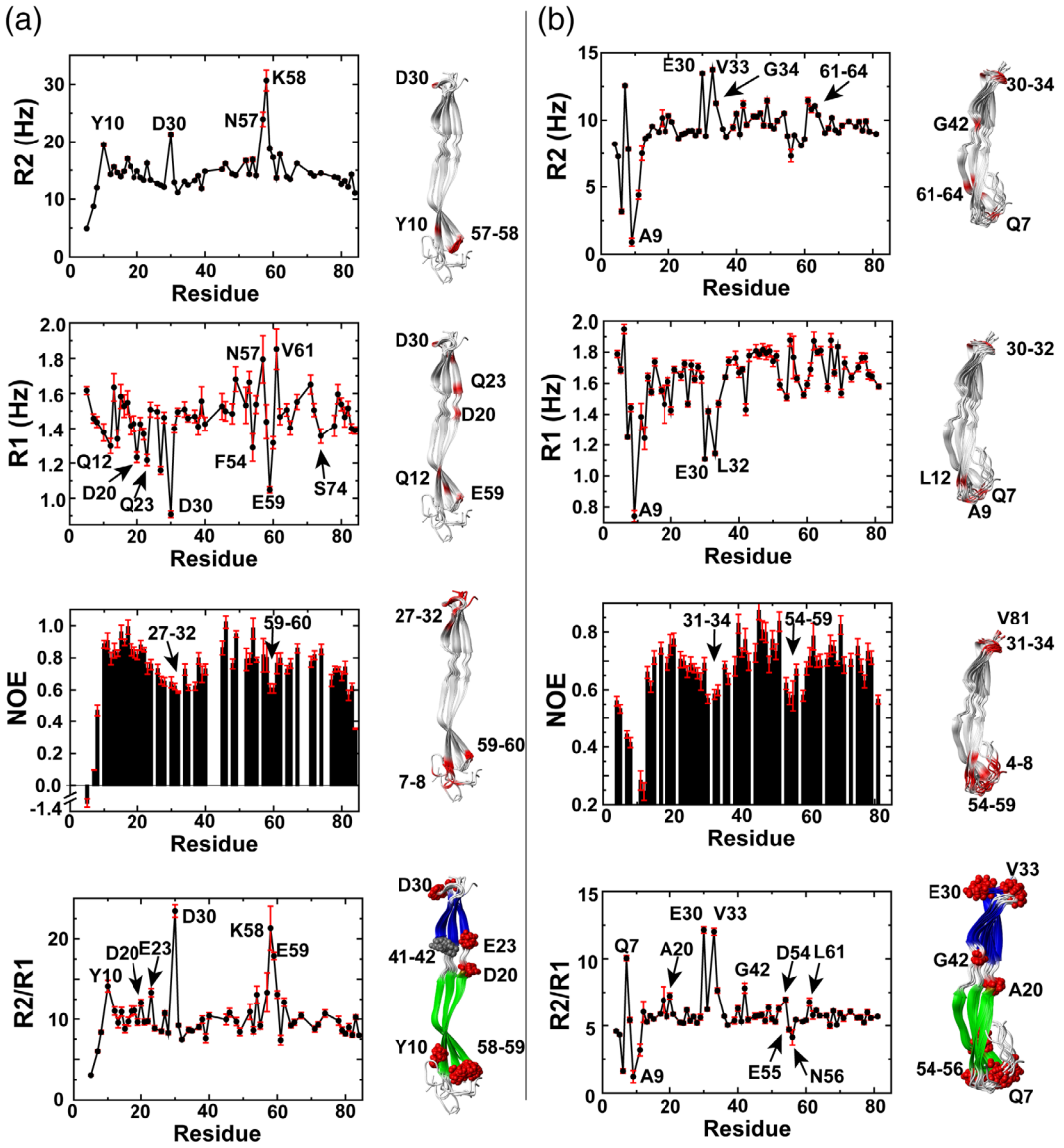

FIGURE 3 *Streptococcus pneumoniae* G5 domains exhibit localized motions. R2 relaxation rates, R1 relaxation rates, heteronuclear NOEs, and R2/R1 ratios are shown for (a) IgA1P-G5 at 1 mM and (b) ENDD-G5 at 0.86 mM. Outliers are mapped onto the NMR solution structures (red). These include regions with elevated R2 relaxation rates above one standard deviation, regions with R1 relaxation rates below one standard deviation, regions with low heteronuclear NOEs that are below ½ standard deviation, and specific R2/R1 ratios

extracted from binding isotherms differ. For example, Zn binds the IgA1P-G5 with a K_d of 9.5 ± 2.2 mM (Figure 4e), which is an order of magnitude weaker than it binds to BGAA-G5 with a K_d of 1.2 ± 0.38 mM (Figure 4f). Zn binding to ENDD-G5 is on the slow timescale whereby peaks disappear and reappear elsewhere (Figure 4c middle, S3). This indicates that Zn binding is at least somewhat tighter for ENDD-G5. In fact, considering that Zn induces largescale changes to the chemical environment of its entire spectrum (Figure S3), these findings may reflect multiple Zn binding sites or could be due to induction of oligomeric species. However, ENDD-G5 binding to Zn is concentration-dependent, as the migratory behavior of ENDD-G5 does not significantly change in the presence of Zn (Figure S4), suggesting that its tight interactions only occur at the higher concentrations used for NMR studies. Similar results have been found for G5 repeats derived from *S. epidermidis*. For example, although the specific affinity to Zn was not quantified, millimolar concentrations of Zn were necessary to induce dimer formation for these former G5 repeats (3.7–5.4 mM), which depended on the exact repeat.[8] Despite this concentration dependence, Zn did specifically bind these G5 repeats and was observed within the X-ray crystal structure to induce anti-parallel G5 structures.

As both heparin and Zn induce, specific CSPs to IgA1P-G5 that could be followed with our resonance assignments, these site-specific changes were mapped onto our NMR solution structures (Figure 4d,e). Both heparin and Zn binding induce similar CSPs to IgA1P-G5, which suggests that both ligands

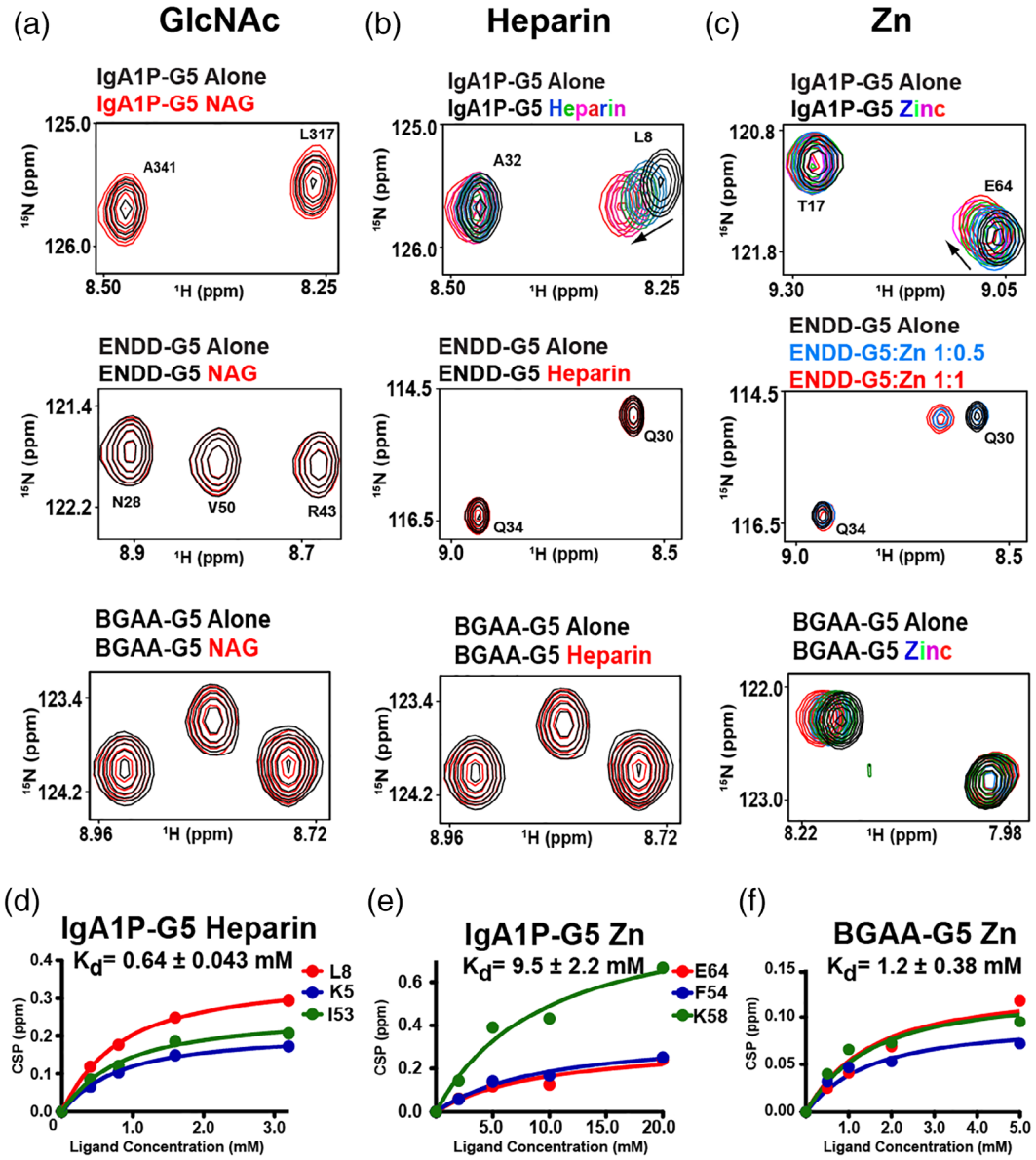

FIGURE 4 G5 domains selectively bind ligands. ^{15}N-labeled G5 domains at 300 μM were titrated with (a) NAG, (b) heparin, and (c) Zn. Unless additional spectra are shown, spectra correspond to 300 μM of the G5 domains alone (black) with 600 μM of indicated ligands (red). Full titrations of IgA1P-G5 with heparin include 0.4, 0.8, 1.6, 3.2 mM and Zn at 2, 5, 10, and 20 mM. Titration of ENDD-G5 with three ratios to Zn are shown along with a full titration of BGAA-G5 with Zn at 0.5, 1, 2, and 5 mM. (d) Heparin binding isotherm to IgA1P-G5 with the extracted dissociation constant, K_d, and specific amides used that include K5, L8, and I63. (e) Zinc binding isotherm to IgA1P-G5 with the extracted dissociation constant, K_d and specific amides used that include E64, F54, and K58. (f) Zinc binding isotherm to BGAA-G5 (unassigned) with extracted dissociation constant, K_d. All data were collected at 900 MHz at 25°C and all binding isotherms were simultaneously fit to a single K_d using GraphPad software to provide the associated uncertainty

engage the same β-sheet and may even have overlapping binding sites.

3 | DISCUSSION

We have experimentally confirmed that four previously proposed *S. pneumoniae* G5 modules are folded,[7] however, the previously proposed G5 module from a vancomycin-resistant bacteria is not competent in folding. From the four folded *S. pneumoniae* G5 modules, SPR1403-G5 was the only module that exhibited marked self-association based on the observation of extreme line-broadening. The remaining three *S. pneumoniae* G5 domains exhibit little to no self-association, which include IgA1P-G5, BGAA-G5, and ENDD-G5. Considering that these three *S. pneumoniae* G5 modules give rise to high-quality NMR spectra, we utilized NMR to probe their ligand interactions previously proposed.[7]

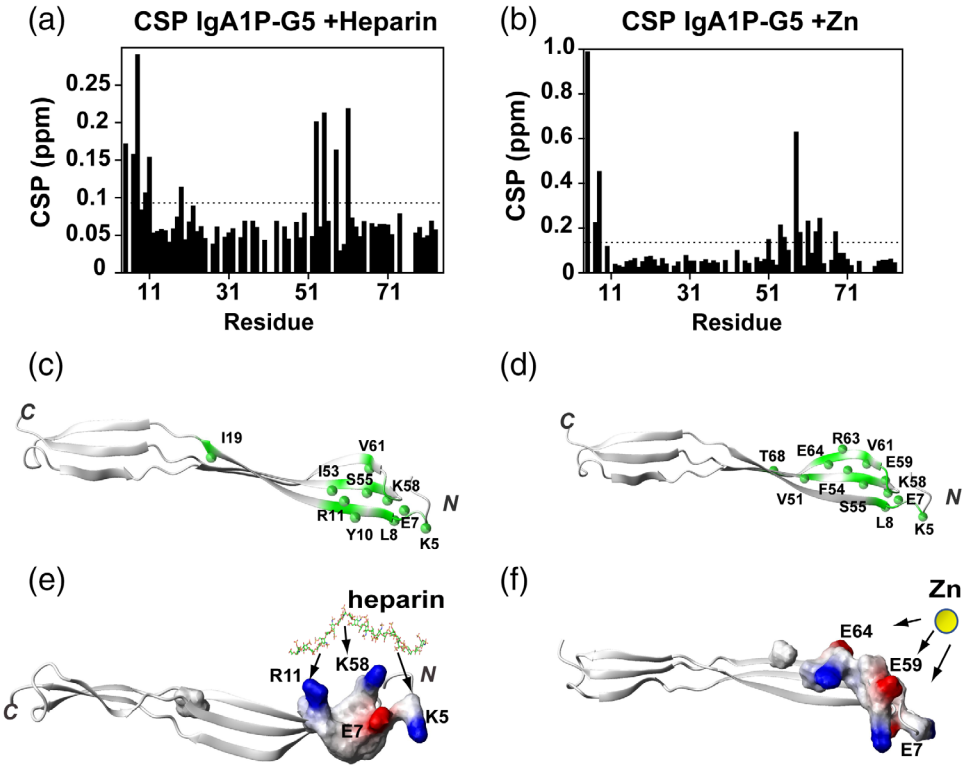

FIGURE 5 IgA1P-G5 ligand interactions localize to the N-terminal β-sheet. (a) CSPs of IgA1P-G5 induced by heparin. The average CSP is 0.077 ppm and standard deviation of 0.050 with the dashed line indicating the average plus ¼ *SD*. (b) CSPs of IgA1P-G5 induced by Zn and binding isotherm with extracted K_d. The average CSP is 0.104 ppm and standard deviation of 0.150 with the dashed line indicating the average plus ¼ *SD*. (c) Heparin-induced resonances above the average plus ¼ *SD* is mapped onto the NMR structure of IgA1P-G5 (green with amides as spheres). (d) Zinc-induced resonances above the average plus ¼ *SD* are mapped onto the NMR structure of IgA1P-G5 (green with amides as spheres). (e) Electrostatic surface representation of heparin-induced CSPs with potential interactions to IgA1P-G5 basic residues annotated. Heparin shown as a stick model simply to illustrate potential interactions. (f) Electrostatic surface representation of zinc-induced CSPs with potential interactions to IgA1P-G5 acidic residues annotated. Zinc is shown as a yellow sphere simply to illustrate potential interactions

The most surprising result from this study is that none of the three *S. pneumoniae* G5 domains tested engage NAG, which has been proposed to be a general G5 ligand.[7] In contrast, all three G5 domains do engage Zn, which is consistent with *Staphylococcus* G5 containing proteins, Aap and SasG that comprise a dozen G5 repeats and oligomerize through specific Zn interactions.[8,9] While these *Staphylococcus* G5 repeats engage Zn with higher affinities than the *S. pneumoniae* G5 affinities, their multimeric structure may enhance their affinities. Zn-mediated adherence is a viable means of modulating bacterial self-association, as host infection increases Zn concentrations in order to stimulate the host cytokine immune response and thus, bacteria may have evolved to exploit such a response for adhesion and/or biofilm formation.[17] Finally, the IgA1P-G5 was the only G5 module to be identified as a heparin interacting module. Although the IgA1P-G5 binding site for both heparin and Zn are located within the same region, as identified by CSPs, the coordination of these very different ligands is likely driven by different interactions (Figure 5). For example, IgA1P-G5 does not actually comprise five glycine residues for which "G5 domain" was initially ascribed. Instead, IgA1P-G5 comprises K58 in place of one of the glycine residues (Figure 1b), which defines the heparin binding site. The term, extracellular polymeric substance (EPS), has been used to broadly encompass substances utilized by bacteria for biofilm formation and heparin has been proposed to mimic extracellular DNA that is a component of the EPS.[18] However, heparin itself may have a function in biofilm formation that is likely complicated and context dependent. For example, heparin induces cell–cell interactions in *Staphylococcus aureus* through the induction of an unknown signaling cascade and not necessarily direct involvement.[19] Thus, it remains a possibility that in some cases G5 modules may comprise receptor complexes that mediate these signaling events to in turn modulate cell–cell interactions or they may play a direct role in other cases. Finally, it is important to address the biological importance of the relatively weak affinities for both Zn and heparin interactions quantified here for these G5 domains. Specifically, the initial events of *Streptococcus* adherence to both surfaces and host cells are known to be relatively weak and facilitate a highly dynamic attachment and dissociation. For example, the initiating weak events of attachment have been proposed to be a

collection of multiple weak interactions that may then be followed by more specific interactions that include lectins.[1] Our direct assessment of *S. pneumoniae* G5 interactions here sets a foundation for further biological experiments to test the role of these ligands in cellular interactions and biofilm formation.

Considering the discrete formation of two β-sheets within G5 domains, we sought to understand whether there was any evidence that could be extracted from relaxation data to indicate flexibility between the two sheets. For example, while R1 relaxation and heteronuclear NOEs are sensitive to ps–ns motions; slower conformational dynamics, that lead to chemical exchange, contribute to an increase in R2 relaxation. Indeed, several residues within the linker regions exhibit locally elevated R2 relaxation rates, which suggests that there are conformational dynamics within this region. Due to the small size of these G5 modules, locally dynamic regions within the faster ps–ns timescale are likely within the "extreme narrowing limit" where $\omega_o\tau_c < 1$, τ_c is the correlation time, and ω_o the spectrometer frequency. Specifically, we can determine this product by estimating τ_c from R2/R1 ratios using a simple approximation.[15] For IgA1P-G5 we obtain 9.8 and 8.6 ns at 1 and 0.5 mM, respectively, and for ENDD-G5 we obtain 7.1 ns, indicating that IgA1P-G5 self-association also slows tumbling relative to ENDD-G5 that does not self-associate (Figure S1). Nonetheless, as the product of the spectrometer frequency and correlation time ($\omega_o\tau_c$ = 2pi*600 MHz*7–10 ns) is only slightly above the extreme narrowing limit for both G5 domains, some of these more flexible regions likely fall below due to the localized nature of dynamics. Thus, slower R1 relaxation rates are expected for such locally dynamic regions.

Cellular adherence is likely complicated and these G5 domains may be directly involved in biofilm formation and/or cellular adherence to epithelial cells. Our studies here provide a foundation for now testing whether these identified ligands are involved at multiple stages of bacterial adherence.

4 | MATERIALS AND METHODS

4.1 | Protein expression and purification

Codon optimized pJ401K (T5 regulation) protein expression plasmids were purchased from DNA2.0 (www.ATUM.bio). Each G5 domain comprised an N-terminal 6xHis-tag, a thrombin cleavage site, and the ORF (Figure 1b), as previously described.[7] A total of five ORFs were included: the IgA1 Protease (IgA1P-G5), endo-beta-N-acetylglucosaminidase-D (ENDD-G5), beta-galactosidase (BGAA-G5), *S. pneumoniae* collagen-like protein (SPR1403-G5) and an Enteroccoccus faecium vancomycin-resistant gene (VanW-G5). Labeled proteins were grown in M9 minimal media (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L NH_4Cl, 2 g/L glucose, 2 mL of 1 M $MgSO_4$ 100 mL of 1 M NaCl $CaCl_2$, 10 mg/L thiamine) in E. Coli BL21/DE3 cells in the presence of 50 μg/mL kanamycin supplemented with either ^{15}N ammonium chloride or both ^{15}N ammonium chloride and ^{13}C glucose. All growths were conducted at 37°C. For a typical growth of 2 L, cells were harvested by sonication and soluble protein applied to Ni-affinity resin (Sigma) using Ni buffer (50 mM Na_2HPO_4, pH 7.5, 500 mM NaCl, 10 mM imidazole). Fractions eluted with Ni buffer B (Ni buffer A supplemented with 400 mM imidazole) were dialyzed against MES buffer (50 mM MES, 150 mM NaCl). Dialysis buffer was changed two times before concentrating, the 6xHis tag removed at room temperature overnight, and then applied to a Superdex-75 (GE Healthcare Life Sciences) equilibrated in the same MES buffer. The G5 domain fractions were further concentrated using Amicon centrifuge filters (3 kDa MWCO) and stored at −80°C until needed. For analytical studies of ENDD-G5, 10 μL of a 1.7 μM stock of ENDD-G5 was applied to a 3 mL Zenix SEC-100 colunn (4.6 × 300 mM) was used with 20 mM Tris, pH 7.5, 150 mM NaCl in the absence and presence of 5 mM Zn.

4.2 | NMR chemical shift assignments, structure determination, and NMR titrations

For structure determination, standard assignment methods were employed on ^{13}C,^{15}N-labeled samples in order to determine the resonance assignments. These include HNCACB, CBCA(co)NH, C(co)NH for backbone assignments and HCCH-TOCSY experiments for side chains. All assignment data for IgA1P-G5 were collected on a Varian 600 MHz spectrometer equipped with a cryo-probe and assignment data for ENDD-G5 were collected on a Bruker 800 at the High Magnetic Field Laboratory (NHMFL) also equipped with a cryo-probe. Chemical shifts are deposited in the BMRB for both IgA1P-G5 (accession number 27839) and ENDD-G5 (accession number 27850). ^{13}C- and ^{15}N-NOESY experiments were all collected on a Varian 900 MHz spectrometer equipped with a cryo-probe. The completeness of assigned resonances for IgA1P-G5 include the following: 98% of CA, 97%, CB, 95% amide nitrogen, 95% amide protons, and 68% of all protons. The completeness of assigned resonances for ENDD-G5 include the following: 100% of CA, 100%, CB, 99% amide nitrogen, 99% amide protons, and 67% of all protons.

Chemical shifts together with NOEs were used for structure determination using Resolution Adapted Structural RECombination (RASREC) in CS-Rosetta software,[20] as we have previously described.[21] Structural coordinates are deposited in the RCSB for both IgA1P-G5 (PDB accession number 6OH1) and ENDD-G5 (PDB accession number 60HF). For titrations, pure choline was purchased from Sigma-Aldrich (catalogue number C7017), NAG purchased from Sigma-Aldrich (catalogue number A8625), Pneumococcal Cell Wall Polysacharides from

Statens Serum Institut Diganostica (catalogue number 3459), and heparin from Sigma-Aldrich (catalogue number 9041-08-1). G5 domains were collected at 300 μM with indicated concentrations of each ligand. For heparin titrated into IgA1P-G5, 0.4, 0.8, 1.6, and 3.2 mM was used. For Zn titrated into IgA1P-G5, 2, 5, 10, and 20 mM was used. For BGAA-G5, 0.5, 1, 2, and 5 mM was used. For ENDD-G5 that exhibited slow exchange, 150 and 300 μM was used. All reported CSPs were calculated as the square root of the sum of squares for both proton and nitrogen shifts with the former normalized by a factor of five ($CSP = \sqrt{(5\Delta\nu_{1H})^2 + (\Delta\nu_{15N})^2}$). GraphPad software was used to simultaneously fit all CSP data in order to provide a single K_d and its associated uncertainty.

ACKNOWLEDGMENTS

EZE was supported by NSF application number 1807326. The National High Magnetic Field Laboratory is supported by National Science Foundation through NSF/DMR-1644779 and the State of Florida.

ORCID

Elan Z. Eisenmesser https://orcid.org/0000-0002-6967-0892

REFERENCES

1. Nobbs AH, Vickerman MM, Jenkinson HF. Heterologous expression of Candida albicans cell wall-associated adhesins in Saccharomyces cerevisiae reveals differential specificities in adherence and biofilm formation and in binding oral Streptococcus gordonii. Eukaryot Cell. 2010;9:1622–1634.
2. Gupta P, Sarkar S, Das B, Bhattacharjee S, Tribedi P. Biofilm, pathogenesis and prevention-a journey to break the wall: A review. Arch Microbiol. 2016;198:1–15.
3. Domenech M, Garcia E, Moscoso M. Biofilm formation in Streptococcus pneumoniae. Microbial Biotechnol. 2012;5:455–465.
4. File TM, Marrie TJ. Burden of community-acquired pneumonia in north American adults. Postgrad Med. 2010;122:130–141.
5. Welte T, Torres A, Nathwani D. Clinical and economic burden of community-acquired pneumonia among adults in Europe. Thorax. 2012;67:71–79.
6. O'Brien KL, Wolfson LJ, Watt JP, et al. Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: Global estimates. Lancet. 2009;374:893–902.
7. Bateman A, Holden MTG, Yeats C. The G5 domain: A potential N-acetylglucosamine recognition domain involved in biofilm formation. Bioinformatics. 2005;21:1301–1303.
8. Conrady DG, Brescia CC, Horii K, Weiss AA, Hassett DJ, Herr AB. A zinc-dependent adhesion module is responsible for intercellular adhesion in staphylococcal biofilms. Proc Natl Acad Sci U S A. 2008;105:19456–19461.
9. Conrady DG, Wilson JJ, Herr AB. Structural basis for Zn2+−dependent intercellular adhesion in staphylococcal biofilms. Proc Natl Acad Sci U S A. 2013;110:E202–E211.
10. Pluvinage B, Chitayat S, Ficko-Blean E, et al. Conformational analysis of StrH, the surface-attached exo-beta-D-N-acetylglucosaminidase from Streptococcus pneumoniae. J Mol Biol. 2013;425:334–349.
11. Lu JJ, Perng CL, Ho MF, Chiueh TS, Lee WH. High prevalence of VanB2 vancomycin-resistant enterococcus faecium in Taiwan. J Clin Microbiol. 2001;39:2140–2145.
12. Yuan ZZ, Yan XJ, Zhang AD, Chen B, Shen YQ, Jin ML. Molecular mechanism by which surface antigen HP0197 mediates host cell attachment in the pathogenic bacteria Streptococcus suis. J Biol Chem. 2013;288:956–963.
13. Gruszka DT, Wojdyla JA, Bingham RJ, et al. Staphylococcal biofilm-forming protein has a contiguous rod-like structure. Proc Natl Acad Sci U S A. 2012;109:E1011–E1018.
14. Mistry D, Stockley RA. IgA1 protease. Intl J Biochem Cell Biol. 2006;38:1244–1248.
15. Larsson G, Martinez G, Schleucher J, Wijmenga SS. Detection of nan-second internal motion and determination of the overall tumbling times independent of the time scale of internal motion in proteins from NMR relaxation data. J Biomol NMR. 2003;27:291–312.
16. Kjellen L, Lindahl U. Specificity of glycosaminoglycan-protein interactions. Curr Opin Struct Biol. 2018;50:101–108.
17. Driessen C, Hirv K, Kirchner H, Rink L. Zinc regulates cytokine induction by superantigens and lipopolysaccharide. Immunology. 1995;84:272–277.
18. Mishra S, Horswill AR. Heparin mimics extracellular DNA in binding to cell surface-localized proteins and promoting Staphylococcus aureus biofilm formation. Msphere. 2017;2:e00135–e00117.
19. Shanks RMQ, Donegan NP, Graber ML, et al. Heparin stimulates Staphylococcus aureus biofilm formation. Infect Immun. 2005;73:4596–4606.
20. Lange OF, Baker D. Resolution-adapted recombination of structural features significantly improves sampling in restraint-guided structure calculation. Proteins Struct Funct Bioinform. 2012;80:884–895.
21. Holliday M, Camilloni C, Armstrong GS, et al. Structure and dynamics of GeoCyp: A thermophilic cyclophilin with a novel substrate binding mechanism that functions efficiently at low temperatures. Biochemistry. 2015;54:3207–3217.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Paukovich N, Redzic JS, Chi Y-C, et al. *Streptococcus pneumoniae* G5 domains bind different ligands. *Protein Science*. 2019;28:1797–1805. https://doi.org/10.1002/pro.3693