Journal of Structural Biology xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Contributions of different modules of the plasminogen-binding *Streptococcus pyogenes* M-protein that mediate its functional dimerization

Cunjia Qiu[a,b], Yue Yuan[a], Jaroslav Zajicek[a,b], Zhong Liang[a,b], Rashna D. Balsara[a,b], Teresa Brito-Robinson[a], Shaun W. Lee[a,c], Victoria A. Ploplis[a,b], Francis J. Castellino[a,b,*]

[a] W.M. Keck Center for Transgene Research, University of Notre Dame, Notre Dame, IN 46556, United States
[b] Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN 46556, United States
[c] Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, United States

ARTICLE INFO

Keywords:
Protein domains
Protein chemistry
Protein-protein interaction
Plasminogen
Bacteria
Receptor structure-function
Recombinant protein expression

ABSTRACT

Group A *Streptococcus pyogenes* (GAS) is a causative agent of pharyngeal and dermal infections in humans. A major virulence determinant of GAS is its dimeric signature fibrillar M-protein (M-Prt), which is evolutionarily designed in modules, ranging from a hypervariable extracellular N-terminal region to a progressively more highly conserved C-terminus that is covalently anchored to the cell wall. Of the > 250 GAS isolates classified, only the subset of skin-trophic Pattern D strains expresses a specific serotype of M-Prt, PAM, that directly binds to host human plasminogen (hPg) *via* its extracellular NH_2-terminal variable A-domain region. This interaction allows these GAS strains to accumulate components of the host fibrinolytic system on their surfaces to serve extracellular functions. While structure-function studies have been accomplished on M-Prts from Pattern A-C GAS isolates with different direct ligand binding properties compared to PAM, much less is known regarding the structure-function relationships of PAM-type M-Prts, particularly their dimerization determinants. To examine these questions, PAMs from seven GAS strains with sequence variations in the NH_2-terminal ligand binding domains, as well as truncated versions of PAM, were designed and studied. The results from bioinformatic and biophysical analyses show that the different domains of PAM are disparately engaged in dimerization. From these data, we propose an experimentally-based model for PAM secondary and quaternary structures that is highly dependent on the conserved helical C-terminal C-D-domains. In addition, while the N-terminal regions of PAMs are variable in sequence, the binding properties of hPg and its activated product, plasmin, to the A-domain, remain intact.

1. Introduction

Group A *Streptococcus pyogenes* (GAS) is a spherical Gram-positive β-hemolytic pathogen that selectively infects humans, causing a range of infections, from antibiotic-sensitive pharyngitis and impetigo to serious invasive diseases, such as necrotizing fasciitis and streptococcal toxic shock, with a mortality > 25%. In addition, immune sequela of these infections, such as acute rheumatic fever (ARF), are common causes of heart disease worldwide (Lee et al., 2009a). Thus, the morbid and sometimes mortal GAS infections represent serious burdens to human health.

Numerous virulence factors are present in GAS, ranging from those present on the bacterial chromosome to those expressed by incorporated bacteriophages (Bao et al., 2014). The GAS genome is very plastic with many integration sites, and horizontal transfer with recombination of genetic materials between GAS strains (Bao et al., 2014), and between GAS strains with other β-hemolytic streptococcal strains (McNeilly and McMillan, 2014), lead to new isolates with variable properties that result from selective pressures from the immune system of human hosts.

All known GAS isolates contain a signature high copy number M-protein (M-Prt), encoded by the *emm* gene. This is the most widely recognized virulence determinant of GAS. Approximately 250 different antigenically distinct M-Prt-types of GAS isolates have been identified through sequencing ~90 nucleotides in the hypervariable region (HVR) of the *emm* gene (Beall et al., 1996; Bessen, 2016). M-Prts contain ~350–450 amino acids arranged in a modular fashion (Hollingshead et al., 1986). The N-terminus contains the HVR, which is sequentially followed by variable A- and B- domains and highly conserved C-, D-, and Pro/Gly-modules (Fischetti, 1989, 1991). The C-terminus of mature

* Corresponding author at: W.M. Keck Center of Transgene Research, 230 Raclin-Carmichael Hall, University of Notre Dame, Notre Dame, IN 46556, United States.
E-mail address: fcastell@nd.edu (F.J. Castellino).

https://doi.org/10.1016/j.jsb.2018.07.017
Received 14 May 2018; Received in revised form 26 July 2018; Accepted 28 July 2018

M-Prts includes a sortase A (LPXTG)-sensitive membrane spanning site, which, when fully processed, is covalently linked to cell wall peptidoglycans. Conformationally, the mature M-Prt is a fibrillar protein consisting primarily of α-helices interrupted by loops, with its N-terminal regions (HVR-A-B-C-D) protruding through the cell wall into extracellular solutions. This provides GAS with a hair-like surface and allows the binding of many different types of host proteins to variably interact with strain-specific M-Prts.

The *emm gene* is located within a virulence-determining region of the bacterial chromosome, viz., the multiple gene activator (Mga) regulon (Hondorp and McIver, 2007), which can contain up to five other sortase A-anchored proteins (Liang et al., 2013). This entire regulon is situated between the *mga* and laminin binding protein (*lmb*) genes in Pattern D strains. Thus, the environmentally-sensitive transcriptional factor, Mga, is an important regulator of M-Prt expression (McIver and Scott, 1997; McIver et al., 1995). The number and arrangement of the genes in the Mga regulon is used to type the pattern of this chromosomal region. Five patterns (A-E) have been identified with different gene arrangements in this regulon (Bessen et al., 1996). The characterization of this region of the chromosome is important for association of the subject GAS strain with the epidemiology of GAS infections, as well as their tissue selectivities and virulences.

In many pattern A-C strains, GAS utilizes M-Prts for adherence and colonization on oral and skin epithelial cells (Berge and Sjobring, 1993; Walker et al., 2014) to provide the inflammatory response accompanying infection. In addition to direct causation of the infection, GAS has evolved survival mechanisms in different host niches to assist the organism in evasion of the host innate immune response. For example, several Pattern A-C M-Prts interact with hemostasis system proteins, *e.g.,* fibrinogen (Ghosh, 2011; Glinton et al., 2017), to provide fibrin that encapsulates the bacteria to allow their colonization. This serves as an initial protective antiphagocytosis mechanism. Some other M-Prts of the Pattern A-C class react with complement binding proteins, e.g., C4BP, and Factor H (FH), through the HVR-region, and inhibit C3b deposition on GAS cells (Berggård et al., 2001; Gustafsson et al., 2013).

Pattern D GAS strains are unique with regard to binding of host proteins and are the only GAS subtype that directly interact with human plasma plasminogen (hPg), a key component of the fibrinolytic system, through their M-Prts (PAM). This specific and very tight interaction occurs *via* the variable A-domains of PAM (Ringdahl et al., 1998; Rios-Steiner et al., 2001; Wang et al., 2010b; Wistedt et al., 1995). The hPg-PAM interaction is central to the pathogenesis of this microorganism (Sanderson-Smith et al., 2006b). PAM-bound hPg is activated by a secreted subtype (2b) of streptokinase (SK2b), produced only by Pattern D strains (Zhang et al., 2012, 2013). This process results in formation of a serine protease plasmin (hPm) (Lahteenmaki et al., 2001), that remains bound to GAS, thereby providing GAS with a surface-bound protease that is employed to solubilize host fibrin that encapsulates the GAS cells to permit ultimate dissemination. In addition, cell-bound hPm serves to degrade extracellular matrix (ECM) components and cellular tight junctions thereby facilitating the invasion of GAS into deep tissues of the host (Sumitomo et al., 2016; Wong et al., 1992). Thus, the exclusive PAM-hPg interaction generally favors virulence of GAS (Sanderson-Smith et al., 2008; Sun et al., 2004).

Isolated M-Prts are generally believed to exist as non-ideal coiled-coil dimers in solution (Ghosh, 2011; Phillips et al., 1981), but limited experimental biophysical evidence lies at the basis of this conclusion. Nonetheless, the relative population of dimeric M-Prts on the cell surface is not known but would be expected to be limited by spacing constraints of the covalently bound monomers. While the functions of M-Prts bound to GAS cells have received much attention, M-Prts are also released from the cell surface during invasive infections, principally by a bacterial secreted cysteine protease, SpeB (Berge and Björck, 1995), and by neutrophil proteases (Herwald et al., 2004). The soluble forms of these M-Prts retain much their functional N-termini and thus possess important functions as soluble proteins. Thus, structure-function relationships of M-Prts in solution are equally important to a fuller understanding of these virulence factors. In this regard, Pattern D M-Prts are not nearly as well studied as Pattern A-C M-Prts (Cedervall et al., 1997; Ghosh, 2018), but we know that PAM dimers are present in solution and possess functional properties (Agrahari et al., 2013, 2016; Bhattacharya et al., 2014). However, some Pattern D GAS strains possess PAMs with significant sequence differences in their hPg-binding sites of the A-domain (Sanderson-Smith et al., 2007). It is of central importance to understand whether natural variations in these PAMs result in conformational differences that could affect their structure-function properties, since it is also not expected that the highly variable number of amino acids in the N-terminal domains of M-Prts (Cedervall et al., 1997; Fischetti, 1989; Fischetti et al., 1988; Ghosh, 2011; McNamara et al., 2008; Nilson et al., 1995; Stewart et al., 2016) are compatible with a generalization that these regions all exist as α-helical parallel coiled-coils of similar stability, even in solution. Thus, the determinants for secondary and quaternary structures of PAM, as well as functional interactions of PAMs, were investigated herein using PAMs from different GAS isolates with variable N-terminal sequences.

2. Materials and methods

2.1. Bacterial strains and cultures

The GAS strains investigated are listed by their isolate names (and *emm* serotypes): AP53 (*emm53*), NS88.2 (*emm98.1*), NS223 (*emm91*), NS455 (*emm52*), SS1448 (*emm86.2*), SS1572 (*emm223*), SS1574 (*emm224*). The AP53 strain was provided by G. Lindahl (Lund, Sweden), strains NS88.2, NS223, and NS455 were gifts from M. J. Walker (Queensland, Australia) (Sanderson-Smith et al., 2006a), and strains SS1448, SS1572, and SS1574 were obtained from the Centers for Disease Control and Prevention. These GAS isolates were cultured at 37 °C, 5% CO_2 on sheep blood agar plates (Teknova, Hollister, CA) or in Todd-Hewitt broth (BD Bacto, Franklin Lakes, NJ) supplemented with 1% (w/v) yeast extract (BD Bacto) (THY).

2.2. DNA manipulations, cloning constructs, and protein purification

Genomic DNAs of the GAS isolates described above were extracted as previously reported (Ward and Leigh, 2002). To construct *pam* expression plasmids, the coding sequence of each PAM, excluding the N-terminal signal peptide and C-terminal LPXTG membrane insertion region, was amplified *via* the polymerase chain reaction (PCR). PCR was conducted using Phusion high-fidelity DNA polymerase (New England Biolabs, Ipswich, MA). The forward primer was designed specifically for each PAM, which was necessitated by their 5′-variabilities, while the reverse primer was the same for each *pam* and contained a sequence that encoded a His_6-tag for subsequent protein purification (Table S1). The PCR products as well as the plasmid, pET-28a (EMD4Biosciences, Darmstadt, Germany), were digested with Nco I and EcoR I (New England BioLabs) at 37 °C. The digested PAM inserts and the vector were purified using the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI). The digested PCR products and plasmid pET-28a were next ligated at 14°C overnight using T4 DNA ligase (New England BioLabs). *Escherichia coli* DH5α (Invitrogen, Carlsbad, CA) cells were transformed with ligated mixtures employing the MicroPulser Electroporation System (Bio-Rad, Hercules, CA). The transformed cells were then cultured on Luria Bertani (LB) agar plates supplemented with 40 μg/mL kanamycin. The plasmid DNAs of the surviving colonies were extracted for PCR screening, and sequence analysis using the EZNA plasmid DNA Midi Kit (Omega, Norcross, GA). The same cloning strategy was applied to truncated PAM_{AP53}, viz., PAM_{AP53_short} (residues 42-175), PAM_{AP53_medium} (residues 42-207), and PAM_{AP53_long} (residues 42-338) (Zhang et al., 2012).