Modular Arrangement and Secretion of a Multidomain Serine Protease: EVIDENCE FOR INVOLVEMENT OF PROLINE-RICH REGION AND N-GLYCANS IN THE SECRETION PATHWAY

Jing Wang¹, Nguan Soon Tan², Bow Ho² and Jeak Ling Ding¹*

¹ Department of Biological Sciences, ² Department of Microbiology, National University of Singapore, 14 Science Drive 4, Singapore 117543 and §Current address: Institut de Biologie Animale, Bâtiment de Biologie, Université de Lausanne, CH-1015 Lausanne, Switzerland

Both authors contributed equally to this work.

Keywords: Proline-rich region; N-glycosylation; Molecular chaperone; Limulus Factor C; Modular arrangement

*Correspondence:
National University of Singapore, Marine Biotechnology Laboratory, Department of Biological Sciences, 14 Science Drive 4, Singapore 117543
Tel: (65)-8742776
Fax: (65)-7792486
E-mail: dbsdj@nus.edu.sg
URL address: http://www.dbs.nus.edu.sg/Staff/ding.html
Summary

The *Limulus* Factor C (FC), a multidomain glycoprotein which binds bacterial endotoxin with high affinity, belongs to the serine protease family of the complement and blood coagulation cascade. Here, we provide compelling evidence for the importance of modular arrangement, and relevance of proline-rich region (PRR) and N-glycosylation to the secretion and function of FC. We propose that PRR could be a universal conformational domain that regulates protein folding and targeting. FCs lacking PRR preceding the serine protease (SP) domain, were localized intracellularly. Misfolded conformers of the intracellular FCs were more susceptible to trypsin digestion. Glycosylation inhibition studies indicate that the presence, but not the exact structure of the N-glycans affects the secretion of FC, although the complexity of glycosylation may influence its endotoxin-induced proteolytic cleavage with resultant enzymatic activity. Disruption of specific N-glycan sites at positions 740, 767 and 912, downstream of the PRR, at/near the SP domain, blocks its secretion. Co-expressed molecular chaperones like canine calnexin associates with glycosylated FCs to increase its solubility and secretion level, but did not alter their expression profiles. Our results clearly demonstrate that the folding and secretion of a multidomain serine protease like FC are determined by its modular domain arrangement and site-specific N-glycans. The secreted FCs containing N-terminal portion of FC are able to detect LPS with high sensitivity. We also identified the lectin-like and sushi 4 domains to contribute to the binding of LPS.
Introduction

The availability of engineered serine proteases allows the study of their activation, substrate specificity and regulation of various physiological processes. As the initiator of the coagulation cascade of LAL (Limulus Amoebocyte Lysate)\(^1\), Factor C (FC) is a multidomain serine protease which is activated by trace amount of its ligand, endotoxin (lipopolysaccharide, LPS) \([1]\). LAL is widely employed to detect the presence of LPS in various industrial and clinical products \([2]\). Analysis of the amino acid sequence indicates that FC is a complex modular protein, comprising 11 domains. \([\text{ref. 3 & Fig. 1A}]\). Many earlier attempts to express FC in heterologous hosts have faced limited success \([4-7]\). The native signal peptide of FC directs this molecule to intracellular large granules of the limulus amoebocytes, and it is unable to secrete the protein out of heterologous expression hosts. Instead, it may serve as a membrane anchor and invariably result in membrane-bound recombinant protein. Thus, an alternative signal peptide is needed to obtain secreted FC for investigation of its domain-function relationship. Using a cross-host secretory signal, ss \([8]\), which leads to the secretion of various recombinant proteins in both prokaryotic and eukaryotic systems, we have recently produced full length and biologically functional domains of FC that binds LPS \([9, 10]\).

The LPS binding and enzymatic activity of Factor C are functionally distinct and are localized in the N-terminal sushi (ES) and C-terminal serine protease domains (SP), respectively \([10, 11]\). However, the importance of interdomain interactions for expression and biological activity of FC is hitherto unknown. Of particular interest is the proline-rich region, PRR, of the FC protein. Proline residues contribute significantly to the conformation and rigidity of the protein structure, and PRR serves as a molecular hinge in a number of proteins \([12, 13, 14]\), affecting protein-protein interaction and protein-ligand binding. In this study, we focus on the potential involvement of PRR in FC secretion and function. By systematic design of FC deletion homologues with and without PRR, we are
able to examine their expression profile, post-translational modifications, localization and LPS-binding activity.

N-linked glycans play pivotal roles in protein folding, oligomerisation, quality control, sorting and transport [15]. These sugar chains are extensively involved in protein conformation and function [16]. There are six potential N-glycosylation sites in FC at positions 523, 534, 624, 740, 767 and 912 (Fig. 1) [17]. However the functions of these glycans has not been examined. We used glycosylation inhibition and site-directed mutagenesis to test the role of N-glycans on expression and secretion of FC, and to map the most critical N-glycans. Further assessment on the involvement of N-glycans in the secretion pathway was performed by co-expression of FC with various molecular chaperones, which are proteins that assist in the correct non-covalent assembly of other polypeptide-containing structures [18].

We report for the first time, that expression and secretion of such a multidomain serine protease requires the proper modular arrangement of the various domains, and PRR is essential for proper folding of the downstream SP domain. In addition, we demonstrated the importance of site-specific N-glycosylation in FC secretion, and provided evidence that N-glycans interact with chaperones and hence manifest their effect on FC secretion. Interestingly, despite having proper protein folding, the secreted FC exhibited only LPS-binding but, lack enzymatic activity. This loss of LPS-induced enzymatic activity can be attributed to different complexity in glycosylation.
Experimental procedures

Unless otherwise stated, all reagents were of molecular biology grade, purchased from Sigma (USA).

Construction of expression plasmids - Drosophila expression vector, pAc5.1/V5-HisA (Invitrogen, USA) was used for fusion expression. The native signal peptide sequence of full-length (FL) FC cDNA was replaced by a novel secretion signal, ss, to direct the fusion protein out of the Drosophila cells. Seven expression plasmids were constructed, with various combinations of FC domains inserted downstream of ss and upstream of V5 and 6 × His fusion tag: FL; FL(-PRR); ES-PRR; ES; PRR-SP; SP and SP-PRR. Figure 1 depicts the domain structures of these recombinant constructs. Among the seven constructs, FL, ES-PRR, PRR-SP and SP-PRR contain PRR. The remaining three constructs: FL(-PRR); ES and SP lack PRR. Only ES, which contains the N-terminal LPS-binding region, lacks both PRR and N-glycan sites. Details on cloning strategies and ss sequence are available upon request. Canine calnexin (cCNX) [19] and human calreticulin (hCRT) [20] cDNAs were subcloned into pAc5.1/V5-HisA.

Site-directed mutagenesis of the N-glycan sites of FC was performed using QuikChange™ site-directed mutagenesis kit (Stratagene, USA). Mutations were verified by sequencing.

Transient and stable transfections - Plasmids were isolated using CsCl-ethidium bromide gradient ultracentrifugation. Drosophila S2 cells were maintained at 25 °C in DES™ medium supplemented with 10 % FBS (Clontech, USA). Calcium phosphate-mediated transient transfection was performed according to the manufacturer’s recommendation (Invitrogen, USA). For stable transfection, pCoHYGRO selection plasmid was cotransfected with Factor
C expression construct, selected with 350 μg/ml hygromycin for 4-6 weeks, and adapted to DES™ serum free medium.

For chaperone co-expression studies, Cellfectin-mediated (Gibco-BRL, USA) transfection was performed with DNA mixture containing 1 μg each of FC and chaperone expression construct. Control transfections were performed with FC expression construct and parental pAc5.1/V5-HisA DNA, instead of chaperone expression constructs.

**Partial tryptic digestion of FC** - Two μl of trypsin solution (100 ng/μl) in 50 mM acetic acid was added into 200 μg of FC in 100 μl of 50 mM NH₄HCO₃. The digestion was carried out at 4 °C. At time intervals, 10 μl of the reaction mixture was removed and terminated by 4 μl of the 4 × SDS sample buffer. Western analysis was performed on the digested samples using polyclonal rabbit anti-FC antibody.

**In vivo glycosylation inhibition and ConA-Sepharose binding assay** - Tunicamycin (TM) inhibits N-glycosylation of glycoproteins by blocking the first step in N-glycan biosynthesis, while α-deoxynojirimycin (DNJ) and α-deoxymannojirimycin (DMJ) specifically inhibit α-glucosidase and α-mannosidase activities, respectively, thus blocking the conversion of high mannose oligosaccharides to complex-type oligosaccharides. S2 cells stably expressing FC were exposed to sublethal doses of TM (5 μg/ml), DNJ (5 mM) and DMJ (2 mM). Proteins from treated cells were collected at 24 h and subjected to affinity chromatography on ConA-Sepharose (Pharmacia). Conditioned medium containing 10 μg/ml secreted FC, and cell lysate protein containing 1 μg/μl cellular FC were incubated with 1/5 volume of ConA-Sepharose at 25 °C for 1.5 h. The unbound fraction containing aglycosylated FC was recovered by centrifugation at 2,000 g for 5 min. The resin-bound glycosylated FC was
dissociated via boiling for 5 min in 3 resin volumes of SDS-PAGE loading buffer. Both fractions were subjected to Western analysis.

**Immunoprecipitation and analysis of FC and chaperone association** - S2 cells co-transfected with FC and chaperone expression plasmids were collected on day 5 post-transfection and washed twice with 1.0 ml PBS. Cells were resuspended in 500 μl lysis buffer (0.1 × PBS, 1.0 mM PMSF, 1.0 μM aprotinin), sheared 7 times through 27-gauge needle and centrifuged at 14,000 g for 15 min. Supernatant containing the cell lysate (L) proteins was retrieved. The pelleted membranes (P) were solubilised in 500 μl lysis buffer containing 1 % digitonin, for 30 min at 4 °C with occasional agitation. The samples were clarified at 14,000 g for 15 min to separate aggregates from solubilised proteins, which were used for subsequent immunoprecipitation.

Fifty μl of protein A-Sepharose (Pharmacia) was incubated with 50 μl of PBS containing 2 μl of anti-V5 antibody for 1 h at 4 °C. After incubation, 100 μl of solubilised protein samples was immunoprecipitated at 4 °C overnight on a rotator. The samples were centrifuged and washed 3 times with cold PBS. Forty μl of SDS-PAGE loading buffer was added to each tube, and boiled for 5 min before Western analysis.

**Concentration and partial purification of FC** - Culture medium containing secreted FC was collected after centrifugation of cell culture at 3,000 g for 10 min at 4 °C. Medium was concentrated 10-fold by ultrafiltration using regenerated cellulose membrane cassette (Pellicon XL, 30 kDa molecular weight cut-off, MWCO, Millipore) at 4 °C. Concentrated protein sample was chromatographed on a Sephadex G-100 (1.5 × 90 cm) column. Fractions containing partially purified FC were pooled and concentrated using Centriprep regenerated cellulose membrane (30 kDa MWCO, Amicon YM).
In vitro endoglycosidase digestion and glycan differentiation - 10 μg each of partially purified secreted FC from Sf9-BEVS and S2 was denatured in 5 % SDS, 10 % β-mercaptoethanol for 10 min at 100 °C, and subjected to endo H and PNGase F (New England Biolabs) digestion at 37 °C for 16 h, followed by Western analysis using anti-FC antibody. Control samples were treated in the same way except in the absence of enzyme.

For glycan differentiation assay, FCs from LAL, CAL, Sf9-BEVS and S2 were purified by protein A–FC antibody affinity chromatography. Fifty μl of protein A-Sepharose (Pharmacia) was incubated with 50 μl of PBS containing 10 μl of anti-FC antibody for 1 h at 4 °C. 100 μg of FC samples were then added to the protein A-Sepharose complex and incubated at 4 °C overnight on a rotator. The samples were centrifuged and washed 3 times with PBS and FC was recovered from protein A-Sepharose by boiling for 10 min in PBS. Lectin binding assay of the immunoaffinity-purified FCs was carried out using DIG Glycan Differentiation Kit (Roche), according to the manufacturer’s instruction.

Functional analyses of FCs - endotoxin binding assay - The LPS-binding potential of various FC proteins were analysed using (a) LPS ELISA-based assay and (b) surface plasmon resonance (SPR).

(a) ELISA-based assay for LPS-binding - This assay was performed as described by Tan et al [10] with some modifications. Coating of polysorp 96-well plate (NUNC) was achieved by incubation at 37 °C for 2 h and blocking was performed overnight at 25 °C. 10 μg of protein sample was added into each well and incubated for 16 h at 25 °C. Bound FCs were detected by sequential incubation with primary rabbit anti-FC antibody (1: 500 dilution, 3h at 37 °C) and goat anti-rabbit secondary antibody (1: 2000 dilution) conjugated to horseradish peroxidase (Dako).
(b) **SPR to determine the Kd of FC for lipid A** - Anti-V5 antibody at 20 μg/ml was immobilised to CM5 chip (Pharmacia) according to the manufacturer’s specification. A low-capacity coating for kinetic study was achieved as indicated by an increase of 1418 resonance units. FC samples were injected at 10 μl/min over the anti-V5 antibody surface, resulting in capture of 5.0 pmole FC (determined by densitometric scanning using pure ES as standard and normalised based on their peptide size) from the conditioned medium. Subsequently, lipid A at 5 different concentrations was injected at 20 μl/min to examine interaction between FC and lipid A. The affinity constant was calculated using BIAevaluation program (ver.-3.0.2) and reconfirmed using CLAMP [21]. For regeneration, 100 mM NaOH was used. Values represent the mean of three independent experiments.
Results

**PRR is critical for the correct folding and secretion of downstream SP domain** - Seven expression plasmids were transfected into S2 cells. Immunoblot analysis shows that all the stable S2 transfectants expressed FCs, albeit displaying different amount and expression profiles (Fig. 1A). Interestingly, although subcloned downstream of a potent secretion signal, not all the FC proteins were secreted, suggesting that the localisation of FCs might be signal sequence-independent. We observed that the FL could be detected in the culture medium. However, the deletion of PRR from FL, yielding FL(-PRR), is not secreted (Fig. 1A, constructs 1 and 2). This indicates that PRR is important for the secretion competency of FCs. We next investigated the relative contribution of PRR to the functionally distinct LPS-binding and serine protease domains, using paired truncated FCs (ES·PRR vs ES and PRR·SP vs SP) containing or lacking the PRR (Fig. 1A). Neither the preservation nor removal of PRR affects the secretion competency of the LPS-binding domain, ES (ES·PRR and ES; Fig. 1A, constructs 3 and 4). Similarly, when PRR was localized upstream of SP (PRR·SP), as in the native FC, the PRR·SP protein was easily detectable in the medium. Interestingly, when the PRR was deleted, the serine protease domain alone (SP) is not longer detectable in the culture medium (Fig. 1A, constructs 5 and 6). This clearly indicates that PRR is critical for the secretion of the serine protease domain.

A major difference between secreted and cellular FCs that harbor SP was that the secreted FCs (FL and PRR·SP) have the PRR domain localized upstream of SP, without which SP is not secreted into the medium. To verify the importance of this modular arrangement between PRR and SP, the PRR was subcloned C-terminal to SP (SP·PRR). The PRR domain in this reciprocal construct was not able to rescue the intracellular retention of SP·PRR. Taken together, the PRR region of FC is important for correct folding of the downstream SP domain and that intracellularly localized FCs might be misfolded.
To shed more light on the role of PRR on folding of the downstream SP domain, partial tryptic digestion followed by Western analysis was carried out on two pairs of FCs: [FL and FL(-PRR)] and [PRR·SP and SP]. Over the observed time, distinct protein fragments of FL (132 kDa, 80 kDa, 52 kDa and 38 kDa), and fragments of PRR·SP (56 kDa, 52 kDa and 38 kDa) were clearly detectable by anti-FC antibody (Fig. 1B), suggesting that they were properly folded into a series of individual compact domains, which are relatively trypsin-resistant. It is interesting to note that two of the tryptic fragments, 80 and 52 kDa, corresponded to LPS-activated autocleavage of Factor C. Importantly, both PRR-deficient mutants, FL(-PRR) and SP, were rapidly digested by trypsin and were not detectable by anti-FC antibody after 1 h. Two isoforms of SP, of ~38 kDa were observed, of which the larger isoform was rapidly digested by trypsin. Interestingly, the minor smaller isoform increased slightly over the tested time period. This increase could be generated from a small subpopulation of the larger isoform, and were relatively resistant to trypsin digestion. These results show differences in protein folding between intracellular PRR-deficient FCs and secretion competent FCs. It is envisaged that lack of PRR upstream of the SP domain tend to result in misfolded FCs, which expose more accessible trypsin cleavage sites that render them more susceptible to protease digestion.

Clearly, PRR is important for the folding and hence, secretion of the serine protease. Essentially, PRR can only exert its effect when the serine protease domain is localized downstream of it.

N-glycosylation is required for proper post-translational processing - We also observed a differential dependence of the two functional domains, viz, LPS-binding and serine protease domains, on PRR for folding and secretion (Fig. 1). The secretion competency of LPS-binding ES is unaffected by PRR, which is in contrast to the SP. A distinguishing feature of ES is that it lacks N-glycosylation sites compared to SP or other SP-harboring FCs, which
contain multiple N-glycosylation sites. Thus, we next investigated the significance of N-glycosylation on the expression and secretion of FC.

The secreted FL protein is glycosylated and is able to bind to ConA-Sepharose (data not shown). FL proteins from S2 cells treated with TM show progressively decreasing size based on their mobility (Fig. 2A), suggesting loss of N-linked sugar chains. Aglycosylated FL was detected only in the L (cell lysate) and P (particulate) fractions, which clearly indicates that loss of N-glycosylation inhibited FL secretion. Absence of N-glycosylation of FL after TM treatment is also confirmed by its inability to bind ConA-Sepharose (Fig. 2B). In contrast to the effects of TM on FC secretion, cells treated with DNJ and DMJ, exhibited no difference in expression level or distribution pattern of FCs (data not shown). Thus, the conversion of high mannose oligosaccharide to complex oligosaccharide is dispensable for protein secretion.

To determine the critical N-glycan sites, a series of glycomutations were introduced (Table I). The position of PRR is such that it partitioned the FL into the N-terminal LPS-binding and the C-terminal serine protease regions. Thus, we decided to mutate the N-glycosylation sites sequentially from each terminal. Single disruption of any one of the six N-glycan sites has no observable impact on the FC secretion (Fig. 2C, Series I). However, elimination of two sites, particularly N523, 534Q and N767, 912Q, impaired the expression of FL (Fig. 2C, lanes 1 and 5 of Series II). Although mutation of the three consecutive N-glycan sites located at the N-terminal caused drastic reduction of the expression of FC, this mutant FL is still secreted into the culture medium (Fig. 2C, lane 1 of Series III). In contrast and importantly, the mutation of the three consecutive C-terminal N-glycan sites located downstream of PRR, at/near the SP domain totally abolished the secretion (Fig. 2C, lane 2 of Series III). The removal of all six glycosylation sites, i.e. aglycosylated FL, is not secretion competent (Fig. 2C, Series IV). Thus, in addition to PRR, the N-glycosylation sites in the SP are also critical for protein expression and secretion. We next sought to assess the
contribution of N-glycans to secretion competency and also to determine a possible functional dependence on PRR.

**N-glycans interact with molecular chaperones to enhance solubility and secretion of FC** - To further examine the involvement of N-glycans in the secretion pathway, co-expression study of FC with two molecular chaperones was carried out. cCNX and hCRT were expressed in S2 cells and their expected protein sizes and cellular localisation were detected with their respective antibodies. The expression levels of the cCNX and hCRT in S2 were also comparable (Fig. 2D).

The overall expression level of FL, as determined by densitometric scanning of the immunoblot, increased over time, and peaked between 5-7 days (data not shown). cCNX caused a 3-fold increase in the level of secreted FL, whereas hCRT minimally enhanced FL expression. Next, the effects of chaperone on the expression level of cellular FCs lacking PRR: FL(-PRR) and SP, were examined. Although overexpression of cCNX and hCRT enhanced the solubility of FL(-PRR) and SP to variable degree (Fig. 2E), none of these chaperones was able to rescue the non-secretion of PRR-deficient FCs: FL(-PRR) and SP. The chaperones did not affect ES expression and secretion. This is not surprising since the LPS-binding ES domain is not glycosylated (Fig. 2E). Co-immunoprecipitation experiments showed that cCNX co-precipitated with FL(-PRR) and SP (Fig. 2F, left panel), but no apparent interaction was observed with hCRT (data not shown). Since hCRT consistently gives a small increase in FC expression, it is conceivable that the association between FC and hCRT is transient. Specific interaction between the N-glycans of FC and cCNX was further confirmed using N-glycosylation inhibitors, TM, DNJ and DMJ. Consistent with the above observation, treatment with TM and DNJ abolished the interaction between FC and cCNX, while DMJ had no effect (Fig. 2F, right panel). This clearly indicates that the N-glycans on the FC molecules are responsible for interaction with cCNX. In addition, cCNX probably
recognises certain structures of FC N-glycans, and binds specifically to de-glucosylated form of FC. Taken together, the results indicate that N-glycans are important for interaction with cCNX, which increases the solubility of the FC protein, and that its observed effect on protein secretion is secondary to the lack of this interaction. Furthermore, the PRR region, although critical for the correct conformation of downstream serine protease domain, is dispensable for the interaction with these chaperones.

**Specific glycosylation modification is required for FC enzymatic activity** - In addition to its contribution to protein solubility, N-glycans have been reported to influence protein functions [16]. As glycosylation is the only post-translational modification known so far in FCs, we attempted to precisely investigate the extent and type of glycosylation in FC that can alter LPS-induced autocleavage and its enzymatic activity. Sf9-BEVS FC was previously demonstrated to exhibit LPS-induced cleavage and enzymatic activity [22].

Endo H and PNGase F digestion analyses show the presence of both high mannose and complex glycans on S2 FC and Sf9-BEVS FC. Differences in their mobility indicate that Sf9-BEVS FC contains 12 kDa of complex N-glycans, and S2 FC harbors 8 kDa of complex N-glycans (Fig. 3A). Thus, it appears that S2 FC is less glycolytically-processed. GNA-treatment reveals the presence of high mannose glycans in all the four FCs. SNA-treatment yielded positive results for FCs from LAL, CAL and Sf9-BEVS, and only weakly reacted with S2 FC. Interestingly, reaction with MAA yielded strongest signal with S2 FC, pointing to a difference in the branching of terminal sialic acids. S2 FC contain greater sialic acid linked $\alpha(2-3)$ to galactose, which is in contrast to Sf9-BEVS FC, LAL and CAL. The latter three FCs contain more sialic acid linked $\alpha(2-6)$ to galactose. In agreement with the Endo H digestion results, DSA yielded weakest signal with S2 FC, confirming that S2 FC is the most poorly processed amongst the four FCs (Fig. 3B). Collectively, we have shown that the
glycan structures of FC expressed in S2 are different from FCs derived from Sf9, CAL and LAL. Such differences in glycomodifications can translate into disparity in protein functions.

LPS-activation of the FC expressed in Sf9-baculovirus expression system was cleaved into two protein fragments of 80 and 52 kDa (Fig. 3C), with resultant enzymatic activity [22]. In contrast, attempts at measuring LPS-inducible enzymatic activity of S2 FC showed that it does not autocleave, and exhibited no enzymatic activity. The abovementioned host cell-specific complexity of glycosylation of the FCs offers plausible explanations for this lack of cleavage and enzymatic activity.

Secreted FL and ES-PRR retained their LPS-binding activities – Secreted S2 FCs were tested for their LPS-binding efficacy using both ELISA and SPR. Using gel-filtration purified FCs, significant lipid A binding activity of FL and ES-PRR was observed, as compared to the partially purified conditioned serum-free medium from wild-type S2 cells. Both secreted FCs were capable of detecting as low as 0.05 μg/ml LPS (Fig. 4A). The secreted PRR·SP was unable to detect LPS, indicating that there is no LPS-binding domain in this molecule. The Kd values of FL and ES-PRR for binding to *E. coli* lipid A was determined using SPR. Figure 4B shows a sensorgram of lipid A binding to FL. The Kd values were $2.23 \times 10^{-7}$ M and $9.58 \times 10^{-7}$ M for FL and ES-PRR, respectively. It is interesting to observe that the absence of the lectin-like and sushi 4 domains leads to 4-fold decrease in affinity for LPS.
Discussion

In this study, we have identified two important determinants of protein secretion of a mosaic serine protease – the PRR and certain strategically located N-glycans. PRR is necessary for proper folding of the downstream SP domain. The N-glycans, particularly those residing downstream of PRR, at/near the SP domain, contribute to stable interaction with specific chaperone(s) and consequently increase the solubility of FCs. Although both factors are involved in different aspects of protein expression, they are equally critical for the maturation and secretion of the protein (Fig. 5). We also identified two potential domains that contribute to FC’s LPS binding activity. Absence of the lectin-like and sushi 4 domain in SP·PRR leads to an apparent decrease in affinity for LPS by 4-fold compared to FL, indicating that these two domains could either affect the conformation of the upstream LPS binding domains, or they also directly contribute to the binding of LPS. Finally, we provide evidence that the N-glycans are dispensable for LPS-binding, but that the differential complexity in glycosylation is critical for LPS-induced cleavage and enzymatic activity.

PRR governs the folding of downstream SP domain - The significance of modular arrangement to the structural integrity and functionality of multidomain proteins such as Factor XII and FC may provide us with important insights into their regulation, activation and substrate specificity. The availability of secreted recombinant proteases in a heterologous system is usually preferable for ease of subsequent purification and functional studies. As the secretion of a protein not only relies on its signal peptide, but also the intrinsic amino acid sequence which determines the final tertiary structure of the mature protein, we examined the secretion of FC constructs containing different modular domains in their native and reciprocal positions.

In the middle of the FC molecule is the basic PRR (amino acid 641 - QNPPVPSYGSEIKPPSRTNSISRVGSPFLRLPRLPLPLARAARKPPKPR -690), which...
contains three Src homology 3 (SH3) binding core motifs [23], PXXP. SH3 domains affect protein targeting, enzymatic activity and play a role in signal transduction [24]. It is possible that the PRR of Factor C is involved in transducing LPS-induced activation of the serine protease enzymatic activity as well as contributing to the LPS-dependent multi-activation of Factor C via protein-protein interaction [25]. In addition, the PRR of Factor C shares homology with the hinge region of human coagulation Factor XII [26]. Hence, it is possible that the PRR acts as a crucial conformational motif necessary for the secretion and function of such serine proteases.

In this study, we constructed and expressed a pair of FCs: [FL and FL(-PRR)] with a view to examine the functional significance of PRR. The FL was efficiently secreted and fully capable of binding LPS. In contrast, the FL(-PRR) which lack the PRR were expressed intracellularly and non-functional. The expression profile of three additional FCs: SP, PRR·SP and SP·PRR confirmed that PRR is essential for folding and secretion of the downstream SP domain. The introduction of PRR C-terminal to ES has no observable effect on ES expression profile, suggesting that PRR only exerts its influence on the downstream domains. Thus, PRR itself may act as an intra-molecular chaperone for correct folding of the downstream SP region for its subsequent transportation through the organelle. In support of our data, it was reported that the minimal portion of secreted Factor XII, which is also a multidomain serine protease similar to FC, also contains PRR [27].

The differential susceptibility to trypsin clearly demonstrates the importance of PRR in the correct folding of the downstream SP region and in directing the secretion of glycosylated FC (Fig. 1B). Trypsin acts preferentially at the unfolded regions in proteins, such as the domain boundaries of a mosaic protein. It is worth noting that on the basis of sequence specificity alone, there are close to 100 and 50 potential trypsin cleavage sites for FL and PRR·SP, respectively. However, only 3-4 characteristic bands, representing two sets of relatively protease-resistant fragments were detected after partial tryptic digestion. Therefore,
the stable fragments are attributable to the compactly folded individual domains. The 38 kDa stable fragment observed after digestion of both FL and PRR·SP could be the properly folded SP domain. On the contrary, PRR-deficient mutants, FL(-PRR) and SP failed to give rise to characteristic patterns of stable products, suggesting that they are incorrectly folded and are more susceptible to proteolytic digestion.

The crystal structure of a catalytic fragment from human C1s, which shares high similarity to Factor C modular architecture, has been determined [28]. The CCP2 module (sushi domain) of C1s is oriented perpendicularly to the surface of the SP domain, and the connection is maintained through a rigid module-domain interface involving intertwined proline- and tyrosine-rich polypeptide segments [28]. Since the residues maintaining the interface framework are highly conserved in the sushi-SP family, it is believed that the rigid sushi-SP assembly is also present in Factor C [29]. Consistent with the role of a module-domain interface, our data show that the absence of PRR changes the sushi-SP assembly, resulting in misfolding of the deletion homologues, rendering them non-secretory.

Many other studies have documented the importance of PRR to the biological functions of different proteins, and that its removal adversely affects the localisation and activity of such proteins [30, 31, 32]. Thus, rather than restricting to a small subset of serine proteases, the PRR could be a universal conformational motif that regulates protein targeting.

**N-glycans contribute to the interaction with chaperones and solubility of the recombinant proteins** - While the data above clearly demonstrate that PRR is an important secretory determinant of FC, PRR also exerts a preferential effect on the serine protease domain, SP, and not on the LPS-binding ES domain. A distinctive characteristic is that ES does not harbor any N-glycosylation sites. Using specific glycosylation inhibitors and site-directed mutagenesis, we demonstrate that inhibition of glycosylation blocked the secretion of FL. The aglycosylated FL became intracellularly localized even though their corresponding PRR
was present, thus indicating that N-glycosylation is also critical for PRR-dependent secretion (Fig. 5). The importance of N-glycans for the secretion of FC was recapitulated using site-directed mutagenesis of the N-glycosylation sites. Interestingly, the ablation of the three N-glycan sites at the C-terminus at positions N740, N767, and N912 completely abolished secretion. Coincidently, these three sites, reside downstream of PRR, at/near the SP domain. The partial dependence of FC secretion on N-linked oligosaccharide appendices suggests that the individual N-glycans does not have local effects on protein folding. However, when the N-glycosylation sites at positions 740, 767 and 912 are modified at the same time, folding and secretion of FC are compromised even though none of the glycans are needed individually. In view of this observation, the oligosaccharides of FC seem to have global effect on the folding/secretion process.

These data suggest that correct folding of the SP domain is the most critical for secretion of FC -- absence of PRR preceding this domain, or absence of N-glycans within this domain leads to misfolding, which impairs FC secretion.

The mechanism by which N-glycans aid in FC secretion is attributable to its ability to interact with chaperones. Over-expression of chaperones has been proven to improve protein folding and stimulate secretion of certain target proteins [34-37]. In this study, the secretion of FL and solubility of the cellular glycosylated FCs [FL(-PRR) and SP] were greatly enhanced by the presence of cCNX. Specific interaction between certain structures of N-glycans and calnexin was further confirmed by glycosylation inhibition study. So, it is evident that the N-glycans of FC functions through interaction with molecular chaperones to aid in protein folding, and subsequent secretion. On the other hand, co-expressed chaperones did not overcome the non-secretory problem of FL(-PRR) and SP, suggesting that their intracellular localization was determined by their conformation, and not the lack of certain molecular chaperones in the host cell secretion pathway.
Functional FC ought to display both LPS-binding and LPS-activated enzymatic activity. However, these two properties were uncoupled in S2 produced FC. S2 FC binds LPS but it is not cleaved and not enzymatically-activated. This is in contrast to Sf9-BEVS FC, which we demonstrated previously to harbor both functions [22]. Further analyses reveal that the loss/gain of enzymatic activity stems from subtle differences in the complexity of glycosylation of FC. In agreement with our results, over the last decade, various reports [38-41] have described that viral infection of Sf9 cells temporally effects protein glycosylation ie: the activation of glycosyltransferase genes takes place under proper conditions, depending on the virus type, promoter and medium. Normal glycosylation characteristics of insect cells were thus altered to allow complex type oligosaccharide processing to occur. Thus, during baculoviral infection of Sf9 cells for expression of FC, the viral-infection process affects the glycosylation of FC, consequently allowing its conformational change upon LPS binding, leading to limited proteolysis and activation of the serine protease enzyme. In contrast, being non-virally infected, the S2 FC bears differently processed glycans, especially those at/near the catalytic domain. Thus, due to the non-native branching of the adjacent N-glycans, the S2 FC is not proteolytically cleaved even though it possesses LPS binding properties. This is despite the correct overall protein folding and successful passage of the S2 FC through the quality control of the ER and Golgi, which allows its successful secretion. However, as FC is such a demanding glycoprotein, retaining the LPS binding activity in these secreted FCs does not satisfy the requirement for their conformational change induced by LPS binding, to yield enzymatic activity. This reveals another surprising impact of N-glycans to FC functions. Although differences in recombinant protein localisation [42] and N-glycosylation heterogeneity [43] have been reported in these two insect cell expression systems, FC from these two hosts is so far the first recombinant protein to exhibit such marked functional difference. Future understanding and manipulations in post-translational processing pathway
with exoglycosidase / glycotransferases could possibly resolve this problem faced by FC as well as other serine proteases.

In summary, we confirm that PRR of a serine protease is essential for the folding and secretion of the downstream catalytic domain. These results suggest that the PRR serves a simple but essential role in tethering the structurally-distinct but interacting domains of the protein. In addition, the secretion ability of FCs was impaired in the absence of N-glycans, especially those residing at/near the catalytic domain. We also report for the first time that co-expression of chaperones significantly increases soluble and secreted recombinant protein level in S2 cells, but has no appreciable effect on the localization of the protein. These results help to elucidate the function of different structural domains of Factor C and how the intrinsic properties of various domains of such a mosaic protein influence its destination (illustrated in Fig. 5). This general concept of the modular arrangement may be applicable to other members of multi-domain protein family, such as blood coagulation Factor XII, in particular, the influence of domain interactions on their cell sorting, secretion and function. Finally, PRR is therefore a potential universal conformational domain that regulates protein folding and secretion, and it is probably applicable as domain fusion junctions for engineered chimaeric multidomain proteins expressed in eukaryotic expression systems.
Acknowledgements

We thank Prof. D. Williams (University of Toronto, Canada) and Prof. D. Llewellyn (University of Wales College of Medicine, UK) for kindly providing the cDNAs of canine calnexin and human calreticulin. This work was supported by the National Science and Technology Board Grant (NSTB/LS/99/004). Wang J is a postgraduate research scholar of National University of Singapore.
References


Footnotes

1The abbreviations used are: FC, Factor C; cCNX, canine calnexin; hCRT, human calreticulin; FL, full length Factor C; ES, Eco47III-SalI fragment of Factor C cDNA; SP, serine protease; PRR, proline-rich region; TM, tunicamycin; DMJ, α-deoxymannojirimycin; DNJ, α-deoxynojirimycin; LPS, lipopolysaccharide; LAL, *Limulus polyphemus* amoebocyte lysate; SPR, surface plasmon resonance; CAL, *Carcinoscorpius rotundicauda* amoebocyte lysate; DSA, *Datura stramonium* agglutinin; MAA, *Maackia amurensis* agglutinin; PNA, Peanut agglutinin; Endo H, endoglycosidase H; PNGase F, Peptide : N-glycosidase F; S2 FC, full length Factor C derived from *Drosophila melanogaster* S2 expression system; Sf9-BEVs FC, full length Factor C derived from *Spodoptera frugiperda*, Sf9-baculovirus expression vector system; SNA, *Sambucus nigra* agglutinin; TtFC, *Tachyleus tridentatus* Factor C.
**Figure Legends**

**Fig. 1. Domain structure, targeting, localization and folding of FCs.** (A) Domain structure of the full length Factor C and the deletion homologues. The recombinant fusion sites of different domains are indicated by arrows (↑) with the amino acid numbering in the original *Carcinoscorpius rotundicauda* FC21 (GenBank Accession No: S77063). The glycosylation sites are marked with closed diamonds (♦). ▲: secretory signal (ss); CR: cysteine-rich region; ▼: epidermal growth factor-like domain; S1-5: sushi-like domains 1-5; LECTIN: lectin-like domain; PRR: proline-rich region; SP: serine-protease coding region; ▲: V5 epitope; ▼: 6×His tag. FL annotates full length FC. ES is the N-terminal LPS-binding region encompassing the sushi 1, 2, & 3 domains. Targeting of the seven recombinant FCs was detected by Western blot. M: 1.0 µg culture medium containing secreted proteins; L: 20 µg cell lysates in PBS containing cellular proteins were obtained by 5 cycles of freeze-thawing of cells, followed by centrifugation at 12,000 g for 15 min at 4 °C; P: 20 µg insoluble pellet obtained at this step was also collected, weighed, and solubilised in 1 × SDS-PAGE sample buffer. Protein concentration was measured by Bradford assay. The relative distribution of FC proteins in M, L and P fractions was revealed using anti-V5-HRP antibody (Invitrogen, USA). Membranes were visualised using the Supersignal® West Pico Chemiluminescent Substrate (Pierce). The amount of FC was quantitated densitometrically by using ImageMaster (Pharmacia) software. Pure ES protein [10] was used as standard. Band intensities within the linear range were measured. (B) **Partial proteolysis of FCs.** FCs (FL, FL(-PRR), PRR·SP and SP) were partially digested with trypsin and collected at time points as indicated. Digested samples were analysed by Western analysis using polyclonal rabbit anti-FC antibody. Distinct stable protein fragments of FL (132 kDa, 80 kDa, 52 kDa and 38 kDa), and PRR·SP (56 kDa, 52 kDa and 38 kDa) were detectable.
Fig. 2. N-glycans are required for proper FC synthesis and targeting. (A) Immunoblot of FL from the M, L, and P after treatment of S2 cells with TM. Western analyses used 10 µg medium (M), 20 µg lysate (L) and 20 µg solubilised pellet (P). (B) Immunoblot after ConA-Sepharose binding assay of FL from M and L. Protein samples were collected from cells treated or not with TM for 48 h. UB: unbound proteins; B: proteins bound to ConA-Sepharose. (C) Immunoblots of FC samples from S2 cells transfected with FC glycomutants. Cell culture supernatant was prepared from S2 expressing wild type FL and FL glycomutants. Protein samples were collected as described in legend of Fig. 1. 1.0 µg of medium was resolved in a 8 % gel by SDS-PAGE. All transfections were done in parallel. The blot- shown is representative of three independent experiments. Immunoblotting was performed with anti-V5-HRP antibody. (D) Expression level of FCs co-expressed with molecular chaperones. Expression of: (1) canine calnexin (cCNX) and (2) human calreticulin (hCRT) in S2 cells. S2 cells transfected with pAc5.1/V5-HisA vector or chaperone expression constructs were collected on day 5. Soluble lysate (L) and particulate (P) fractions were resolved by SDS-PAGE and analysed using respective chaperone antibodies. Antibodies for cCNX is from Transduction Laboratories (USA) and hCRT is from Stressgen (Canada). cCNX is detected in the P fraction and hCRT are detected in both the L and P fraction. These expression profiles conform to their original nature, in that cCNX is integrated into the ER membrane while hCRT are ER luminal proteins. (E) Comparison of secreted FL, ES, and cellular FL(-PRR), SP co-expressed with chaperones on day 5 post-transfection. Relative values calculated from the densitometric data are shown, with the value for the S2 cells transfected with FC construct alone, arbitrarily set as one. Error bars represent the S.E. (n=3). (F) left panel : Direct interaction of glycosylated FC with chaperones. Digitonin-solubilised cell lysates (L) from S2 cells transfected with vector-pAc5.1/V5-HisA, cCNX alone, FL(-PRR)+cCNX and SP+cCNX were precipitated overnight with anti-V5 antibody. The immunoprecipitates were subjected to SDS-PAGE and probed
with anti-cCNX antibody. **right panel:** Association of chaperone and FC under glycosylation-inhibited conditions. S2 cells co-transfected with cCNX and SP (lanes 2, 3, 4, 5) were pre-treated overnight with TM (5 µg/ml, lane 3), DNJ (5 mM, lane 4) and DMJ (2 mM, lane 5) before transfection, and incubated with these inhibitors after transfection. FCs expressed were then precipitated with anti-V5 antibody, and the chaperones that associated with FC were analysed by SDS-PAGE and immunoblotting.

**Fig. 3. Comparison of the extent and complexity of glycosylation in S2 FC and Sf9-BEVS FC.** (A) Endoglycosidase treatment revealed Sf9-BEVS FC has more complex N-glycans. The molecular mass shift after Endo H treatment indicates extend of high-mannose N-glycan chains, while the shift after PNGase F treatment indicates total N-glycosylation (including both high-mannose and complex sugar chains). The size difference between PNGase F treated FC sample and FC TnT (transcription coupled translation) product is contributable to O-glycosylation. (B) Glycan differential assay of S2 FC and Sf9-BEVS FC. Lane 1: FC; Lane 2: FC without N-glycans. The N-glycans were detached from FC peptide chain by treatment with PNGase F, and the free glycan chains are removed by centrifugation through 30 kDa MWCO-microcon-filter (Amicon). GNA, SNA, MAA, PNA and DSA has specific affinity for terminal mannose, sialic acid linked α(2-6) to galactose, sialic acid linked α(2-3) to galactose, O-glycosidically linked sugar chain and Galβ-(1-4)GlcNAc in complex and hybrid N-glycans, respectively. GNA-treatment yielded positive signals for all the four FCs, SNA-treatment yielded positive results for FCs from LAL, CAL and Sf9-BEVS, and only weakly reacted with S2 FC. Reaction with MAA yielded strongest signal with S2 FC. SNA did not react with PNGase F treated FC samples. DSA yielded weakest signal with S2 FC. (C) Comparison of LPS activation of Sf9-BEVS FC and S2 FC. Incubation of 10 µg of Sf9-BEVS FC and S2 FC with 50 EU of LPS was carried out at
37°C for 1 h. The activation of FC from Sf9-BEVs was readily observable (autocleavage into 80 and 52 kDa bands). S2 FC remained intact after the LPS treatment.

**Fig. 4. Functional analysis of FC.** (A) Ligand binding assays of FC samples. ELISA of 10 μg each of protein samples: control serum-free medium; FL; ES-PRR and PRR-SP partially purified through Sephadex G-100 (void fraction). Three concentrations of LPS: 0, 0.05, 5 μg/ml were used. Error bars represent the S.E. (n=3). (B) A sensorgram depicting the interaction (in response units, RU) of captured FL with various concentrations of lipid A. ES-PRR gave similar response profiles. 100μl of 5 concentrations of (5.00, 2.50, 1.25, 0.62, 0.31 μM) of lipid A were injected. The affinity constants were calculated using BIAevaluation program version 3.0.2 and reconfirmed using CLAMP [21].

**Fig. 5. An illustration of PRR and N-glycans involved in secretion of FC.**  ![Cysteine-rich region](image1.png); ![EGF-like domain](image2.png); ![Sushi-like domains](image3.png); ![Lectin-like domain](image4.png); ![Proline-rich region](image5.png); ![Serine protease native conformer](image6.png); ![Misfolded serine protease](image7.png); ![N-glycan](image8.png); ![Proteins involved in folding, assembly and secretion, e.g. chaperones, foldases, signal peptidase, and glycosidase](image9.png). The hypothesis was confirmed by: (a) removal of PRR from full length Factor C (FL), which completely switches its expression profile and protein conformation, and reconfirms PRR to be an important secretory determinant for glycosylated FC; (b) mutation of N-glycan sites which impaired the secretion competency of FC; and (c) co-expressed chaperone (e.g: calnexin) which associates with glycosylated FL protein and assist the reversal of this impairment to a certain extent, to enhance its solubility and secretion. Thus, whilst PRR is crucial for correctly directing folding and secretion of downstream glycosylated SP domain, chaperones and foldases also play a role in enhancing protein folding, to ensure that the modular arrangement of the glycosylated protein is properly folded to remain soluble, or targeted for secretion and exhibit full functionality.
<table>
<thead>
<tr>
<th>Series I FC mutant</th>
<th>Consensus sequence(s) disrupted</th>
</tr>
</thead>
<tbody>
<tr>
<td>N&lt;sup&gt;523&lt;/sup&gt;Q</td>
<td>Asn-523</td>
</tr>
<tr>
<td>N&lt;sup&gt;534&lt;/sup&gt;Q</td>
<td>Asn-534</td>
</tr>
<tr>
<td>N&lt;sup&gt;624&lt;/sup&gt;Q</td>
<td>Asn-624</td>
</tr>
<tr>
<td>N&lt;sup&gt;740&lt;/sup&gt;Q</td>
<td>Asn-740</td>
</tr>
<tr>
<td>N&lt;sup&gt;767&lt;/sup&gt;Q</td>
<td>Asn-767</td>
</tr>
<tr>
<td>N&lt;sup&gt;912&lt;/sup&gt;Q</td>
<td>Asn-912</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Series II FC mutant</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N&lt;sup&gt;523, 534&lt;/sup&gt;Q</td>
<td>Asn-523 and Asn-534</td>
</tr>
<tr>
<td>N&lt;sup&gt;534, 624&lt;/sup&gt;Q</td>
<td>Asn-534 and Asn-624</td>
</tr>
<tr>
<td>N&lt;sup&gt;624, 740&lt;/sup&gt;Q</td>
<td>Asn-624 and Asn-740</td>
</tr>
<tr>
<td>N&lt;sup&gt;740, 767&lt;/sup&gt;Q</td>
<td>Asn-740 and Asn-767</td>
</tr>
<tr>
<td>N&lt;sup&gt;767, 912&lt;/sup&gt;Q</td>
<td>Asn-767 and Asn-912</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Series III FC mutant</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N&lt;sup&gt;523, 534, 624&lt;/sup&gt;Q</td>
<td>Asn-523, Asn-534 and Asn-624</td>
</tr>
<tr>
<td>N&lt;sup&gt;740, 767, 912&lt;/sup&gt;Q</td>
<td>Asn-740, Asn-767 and Asn-912</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Series IV FC mutant</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N&lt;sup&gt;1-6&lt;/sup&gt;Q</td>
<td>All six sites</td>
</tr>
</tbody>
</table>
(A)

**FC constructs containing PRR**

1. FL: Full Length Factor C
   - FL(-PRR)

2. M L P
   - 132 kDa

3. ES-PRR
   - 62 kDa

4. ES
   - 38 kDa

5. PRR-SP
   - 56 kDa

6. SP
   - 38 kDa

7. SP-PRR
   - 62 kDa

(B)

![Figure 1](http://www.jbc.org/)

- **FL kDa**
  - 132 → 80 → 52 → 38

- **FL(-PRR) kDa**
  - 120 → 80 → 52 → 38

- **PRR-SP**
  - 56 → 52 → 38

- **SP**
  - 38
Figure 3

(A) S2 FC

<table>
<thead>
<tr>
<th>MW (kDa)</th>
<th>MW (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>132</td>
<td>132</td>
</tr>
<tr>
<td>126</td>
<td>128</td>
</tr>
<tr>
<td>118</td>
<td>116</td>
</tr>
<tr>
<td>112</td>
<td>112</td>
</tr>
</tbody>
</table>

N-glycosylation=14 kDa
High mannose N-glycan=6 kDa
Complex-glycan=8 kDa
O-glycosylation=6 kDa

N-glycosylation=16 kDa
High mannose N-glycan=4 kDa
Complex-glycan=12 kDa
O-glycosylation=4 kDa

(B) GNA SNA MAA PNA DSA

1 2 12 12 12 12

LAL CAL Sf9 S2 +ve/-ve

GNA SNA MAA PNA DSA

(C) Sf9-BEVS FC S2 FC

LPS - + - +
Figure 4

(A) LPS-ELISA

(B) Association Dissociation

[Lipid A]
Figure 5

Secretion/Functional

FL

(a) – PRR

(b) – N-glycans

(c) Association of chaperones

FL(-PRR)

Misfolded
Unable to secrete

Glycomutants

Misfolded
Unable to secrete
Modular arrangement and secretion of a multidomain serine protease: Evidence for involvement of proline-rich region and N-glycans in the secretion pathway

Jing Wang, Nguan Soon Tan, Bow Ho and Jeak Ling Ding

J. Biol. Chem. published online June 27, 2002

Access the most updated version of this article at doi: 10.1074/jbc.M202837200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/early/2002/06/27/jbc.M202837200.citation.full.html#ref-list-1