Extending the cellulosome paradigm: The modular *Clostridium thermocellum* cellulosomal serpin, PinA, is a broad spectrum inhibitor of subtilisin like proteases

Páraic Ó Cuív\(^1\#\), Rajesh Gupta\(^1\#\), Hareshwar P. Goswami\(^1\) and Mark Morrison\(^{1,2}\#\)

\(^1\)CSIRO Animal, Food and Health Sciences, Queensland Biosciences Precinct, 306 Carmody Road, Queensland 4067, Australia

\(^2\)The Ohio State University, Columbus, Ohio, USA

#Corresponding authors:

Páraic Ó Cuív

CSIRO Animal, Food and Health Sciences, Queensland Biosciences Precinct, 306 Carmody Road, Queensland 4067, Australia

Tel: +61 (0)7 3214 2312/Fax: +61 (0)7 3214 2900

paraic.ocuiv@csiro.au

Mark Morrison

CSIRO Animal, Food and Health Sciences, Queensland Biosciences Precinct, 306 Carmody Road, Queensland 4067, Australia

Tel: +61 (0)7 3214 2216/Fax: +61 (0)7 3214 2900

mark.morrison@csiro.au

Running Title: PinA is an inhibitor of subtilisin like proteases

\(^\d\)RG current address: Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia.
Abstract

*Clostridium thermocellum* encodes a cellulosomal, modular and thermostable serine protease inhibitor (serpin), PinA. PinA stability but not inhibitory activity is affected by the Fn(III) and Doc(I) domains; and PinA is a broad inhibitor of subtilisin like proteases and may play a key role in protecting the cellulosome from protease attack.
Much of the organic carbon and nitrogen in soil is sequestered in cellulosic and protein biomass respectively and microorganisms have consequently evolved specialised strategies to facilitate the release and assimilation of these biomaterials. *Clostridium thermocellum* is an anaerobic, thermophilic soil bacterium that produces a large extracellular multisubunit protein complex termed the cellulosome that functions to deliver organic carbon to the cell. The extracellular location of the cellulosome leaves it vulnerable to protease attack and bacterial and fungal cellulosomes have been shown to carry specific protease inhibitors, including serpins and cyspins, that can confer protection against protease attack (1-4). Serpins function as metastable suicide substrates for their cognate protease(s) with cleavage of the reactive centre loop resulting in the formation of an inactive covalently linked serpin-protease complex. *Clostridium thermocellum* ATCC 27405 produces a cellulosomally targeted modular thermostable serpin termed PinA ((2), GenBank Accession number: YP_001036624) however the contribution of the modular domains to serpin functionality remains unknown, as does the potential broader ecological role of PinA.

To address these questions we expressed a near full length recombinant protein lacking the N-terminal signal sequence (rPinA75) and a truncated protein lacking the N-terminal signal sequence, Fn(III) like and Doc(I) domains (rPinA614) (Figure 1A). The stoichiometry of inhibition (SI) of rPinA75 (1.3±0.1) against subtilisin type VIII from *Bacillus licheniformis* was similar to that previously reported (2) indicating that the N-terminal signal sequence does not affect inhibitory activity. Similarly, the SI for rPinA614 (1.4±0.3; Students independent t-test, *p*>0.05) was not significantly different.
from that of rPinA75 indicating that the Fn(III) like and Doc(I) domains do not contribute to inhibitory activity.

Serpins fold to a metastable state and we next examined the effect of the N-terminal truncation on the conformational stability and functionality of rPinA. Protein samples were harvested in a longitudinal manner and native PAGE analysis revealed that two rPinA614 species, rPinA614 and rPinA614-L, as identified by liquid chromatography electrospray ionization tandem mass spectrometry (LC-EIS MS/MS), were evident after 5 hours with a gradual accumulation up to 24 hours where each species accounted for ~50% of the total recombinant serpin produced (Figure 2A). We observed a single species for rPinA75 over 24 hours as determined by native PAGE and we thus hypothesised that the PinA614-L species represented a cleaved or latent state and we sought to discriminate between these two possibilities as previously outlined by Dafforn et al., (5). Firstly, samples containing both rPinA614 and rPinA614-L were analysed by SDS-PAGE and revealed the presence of a single serpin band at the time points examined, suggesting that both species had similar mobility and that rPinA614-L was therefore uncleaved (Figure 2B). Serpins in the latent conformation are more thermostable than those in the metastable conformation and we next investigated the thermal stability of rPinA614 harvested at 9 hours to allow discrimination between both forms. Native PAGE indicated that rPinA614 underwent thermal denaturation at 70°C over a period of 180 minutes resulting in the accumulation of rPinA614-L (Figure 2C). Subsequent, SDS-PAGE analysis indicated that these differences in mobility were not due to protein cleavage and/or degradation. Finally, the SI of rPinA614 harvested at 24
hours was significantly different (2.5±0.3; \( p<0.01 \)) further suggesting that rPinA614-L was in a non-inhibitory latent conformation. Taken together this suggests that the truncation of the Fn(III) like and Doc(I) domains is not necessary for rPinA614 to fold to the metastable state but it may leave the serpin domain prone to assuming a latent conformation. Thus, the N-terminal region of PinA may function in a manner analogous to that of the N-terminal domain of the thermostable serpin tengpin from *Thermoanaerobacter tengcongensis* where an N-terminal domain that is not required for folding to the metastable state functions to stabilise the serpin and maintain it in that conformation (6).

We next sought to determine whether the modular nature of PinA contributed to its thermostability. As expected, the thermostability profile of rPinA75 at 60°C was similar to that previously reported by Kang *et al.*, (2) with a calculated half-life for the serpin of approximately 180 min (Figure 2D). Surprisingly, deletion of the Fn(III) like and Doc(I) domains enhanced the thermostability profile of rPinA at 60°C with no significant loss in activity observed over 180 minutes. These results are consistent with previous reports that revealed that protein thermostability can be influenced by inter domain interactions and that truncated protein domains can exhibit enhanced thermostability (7, 8). Native PAGE did not reveal any change to the migration profile of either rPinA75 or rPinA614 over the 180 minutes (data not shown) suggesting that the overall protein conformations were unaffected following incubation at 60°C. Thermostable multi-domain proteins can undergo cooperative unfolding during thermal denaturation (9-11) and the Fn(III) like and Doc(I) domains may affect the thermostability of the reactive centre loop in rPinA75.
although the overall protein structure remains unaffected. Thus, while the N-terminal
domains of PinA may stabilise the protein in the metastable state they may also increase
its thermolability.

Subtilisin like proteases are amongst the most diverse and prevalent proteases found in
soil environments where they have been proposed to play a critical role in the cycling of
organic nitrogen (12-14), and we finally investigated the ability of rPinA to function as a
general inhibitor of MEROPS family S8, subfamily A proteases. rPinA75 inhibited
Savinase® (1.2±0.2) and Esperase® (1.1±0.1) produced by Bacillus clausii and Bacillus
halodurans respectively ((15); liquid enzyme formulations available from Sigma-Aldrich
Corp, St. Louis, USA), and subtilisin type XXIV (1.3±0.3) produced by B. licheniformis
(Sigma-Aldrich Corp, St. Louis, USA) effectively, and the SI was not significantly
different to that for subtilisin type VIII (p>0.05). In addition, rPinA75 was capable of
inhibiting Proteinase K from Tritirachium album Limber (1.3±0.2; p>0.05) although it is
more distantly related to subtilisin (16, 17). The ability of PinA to function as a broad
inhibitor of MEROPS family S8, subfamily A proteases is of particular interest as these
proteases show significant sequence variability (76%, 70% and 40% similarity between
Savinase®, Esperase® and Proteinase K respectively with subtilisin type VIII; no
sequence available for subtilisin type XXIV) however, the protease active site and
substrate recognition site are highly conserved which likely explains the ability of rPinA
to effectively inhibit these proteases. Furthermore, specific serpin domains termed
exosites also contribute to serpin specificity (reviewed by (18)) and as subtilisin like
proteases are characterised by a common tertiary structure (e.g. (19, 20)) it is possible that specific PinA exosites also contribute to the inhibition process.

Our results suggest that PinA functions to protect the cellulosome from attack by MEROPS family S8, subfamily A proteases. *C. thermocellum* ATCC 27405 encodes a second putative cellulosomal serpin, PinB ((2), GenBank Accession number: YP_001036625), whose cognate protease has not yet been identified but that may also function to protect the cellulosome from protease attack, and this remains to be examined. Separately, while the role of the cellulosome vis-à-vis carbon has been well described little has been postulated regarding the potential role of the cellulosome in nitrogen acquisition. The genome sequences of several *C. thermocellum* strains have revealed the presence of putative cellulosomally targeted subtilisin like proteases indicating that the cellulosome may also function to provide organic nitrogen to the cell. *C. thermocellum* is capable of synthesising all 20 amino acids (21, 22) however it also encodes numerous putative peptidases and an oligopeptide transport system suggesting that it can acquire exogenous amino acids and/or peptides. In this context, the tethering of a protease to the cellulosome may serve to prevent diffusion away from the cell and ensure that any organic nitrogen produced will be located proximal to the host cell and not readily available to competing microorganisms. As production of the cellulosome is metabolically demanding future studies should examine the role of cellulosomal proteases and protease inhibitors in relation to nitrogen bioavailability and their contribution to the ecological fitness of the host.
Acknowledgments

We gratefully acknowledge the support provided by a CSIRO Science Leader award to Mark Morrison.

We thank colleagues for providing critical reading and suggestions to improve the manuscript.
References


Figure 1. (A) Schematic representation of the modular structure of PinA with the rPinA75 and the rPinA614 constructs as indicated. Briefly, pinA was PCR amplified from pET21-SK2 (2) and cloned into pET-28c(+) (Novagen) placing the gene under the control of the vector borne T7 promoter and facilitating the fusion of a C-terminal 6xHis-tag. The final rPinA75 and rPinA614 expressing plasmid constructs were confirmed by Sanger sequencing. The PCR primers used facilitated the addition of an N-terminally located Strep II tag and a Factor Xa cleavage site to the serpin and a C-terminally located Factor Xa cleavage site. The signal sequence (grey), Strep II tag/Factor Xa site (Strep II/Xa), Factor Xa site/6x His tag (Xa/6XHis), fibronectin III (Fn III), dockerin (Doc I) and serpin region are indicated. (B) SDS-PAGE analysis of rPinA75 (1) and rPinA614 (2).

Figure 2. (A) Native PAGE analysis of recombinant rPinA614 harvested at 4, 5, 6, 7, 8, 9 and 24 hours. The metastable (rPinA614) and latent (rPinA614-L) forms are indicated. The identity of rPinA614 and rPinA614-L from 5 to 24 hours was confirmed by LC-EIS MS/MS. (B) SDS-PAGE analysis of recombinant rPinA614 harvested at 5, 6, 7, 8, 9 and 24 hours. (C) Native PAGE analysis of the thermostability profile of rPinA614 heated at 70°C and harvested as indicated. The identity of rPinA614 and rPinA614-L at all time points was confirmed by LC-EIS MS/MS with the exception of rPinA614-L at 60 minutes. (D) Thermostability analysis of rPinA75 and rPinA614. The thermal stability of rPinA was determined essentially as described by Kang et al., (2) except that the reactions were performed at 60°C. The thermostability analyses were performed at least twice independently per sample.