Plant Physiology and Biochemistry 53 (2012) 27–32

Contents lists available at SciVerse ScienceDirect

Plant Physiology and Biochemistry

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

Short communication

The PA domain is crucial for determining optimum substrate length for soybean protease C1: Structure and kinetics correlate with molecular function

Anna Tan-Wilson [a], Basel Bandak [b], Moses Prabu-Jeyabalan [b,c,*]

[a] Department of Biological Sciences, State University of New York at Binghamton, 4400 Vestal Parkway East, Binghamton, NY 13902, USA
[b] Department of Chemistry, University of Scranton, 800 Linden Street, Scranton, PA 18510, USA
[c] Division of Basic Sciences, The Commonwealth Medical College, 525 Pine Street, Scranton, PA 18509, USA

ARTICLE INFO

Article history:
Received 8 October 2011
Accepted 3 January 2012
Available online 10 January 2012

Keywords:
PA domain
Proteolytic enzyme
Storage protein mobilization
Substrate binding
Subtilisin
Homology modeling
Active site

ABSTRACT

A subtilisin-like enzyme, soybean protease C1 (EC 3.4.21.25), initiates the degradation of the β-conglycinin storage proteins in early seedling growth. Previous kinetic studies revealed a nine-residue (P5–P4′) length requirement for substrate peptides to attain optimum cleavage rates. This modeling study used the crystal structure of tomato subtilase (SBT3) as a starting model to explain the length requirement. The study also correlates structure to kinetic studies that elucidated the amino acid preferences of soybean protease C1 for P1, P1′ and P4′ locations of the cleavage sequence. The interactions of a number of protease C1 residues with P5, P4 and P4′ residues of its substrate elucidated by this analysis can explain why the enzyme only hydrolyzes peptide bonds outside of soybean storage protein's core double β-barrel cupin domains. The findings further correlate with the literature-reported hypothesis for the subtilisin-specific protease-associated (PA) domain to play a critical role. Residues of the SBT3 PA domain also interact with the P2′ residue on the substrate's carboxyl side of the scissile bond, while those on protease C1 interact with its substrate's P4′ residue. This stands in contrast with the subtilisin BPN′ that has no PA domain, and where the enzyme makes stronger interaction with residues on the amino side of the cleaved bond. The variable patterns of interactions between the substrate models and PA domains of tomato SBT3 and soybean protease C1 illustrate a crucial role for the PA domain in molecular recognition of their substrates.

1. Introduction

In early seedling growth of the soybean (*Glycine max* (L.) Merrill), a subtilisin-like enzyme, protease C1 (EC 3.4.21.25), initiates the proteolytic reactions to mobilize the β-conglycinin storage protein [1]. The β-conglycinins are homo- and hetero-trimers of α, α′ and β subunits (UniProt ID: P11827, P13916 and P25974, respectively), with the N-termini of the α and α′ subunits being longer than the β subunit by 169–183 residues [2]. These subunits have in common core cupin domains, two compact β-barrels connected by a 30-residue linker [3–5]. Five sequenced cleavage sites of protease C1 were all within the N-terminal extensions of the α and α′ subunits [6], regions that extend outside the cupin domains [4]. The β subunit is not attacked, and the core structures of the α′ and α subunits that are homologous to the β subunit are the final products [1]. Protease C1 also targets the C-terminal region of the glycinins, the other major soybean storage protein [7] composed of subunits that also have core cupin domains [8]. The targeted cleavage sites are also in regions that are outside the barrel structures.

Protease C1 mobilizes protein reserves within the first 24 h of seedling growth, its early action made possible by its presence in active form in protein storage vacuoles along with its protein substrates. Proteolysis is initiated after germination by an increase in acidity [9]. Degradation of the cupin domains begins 6 days after seedling growth, initiated by another enzyme that is synthesized only after germination [10]. The inability of protease C1 to cleave peptide bonds within the cupin domains contributes to the preservation of these reserves for a later stage of growth. Based on the sequences of its cleavage sites [6], protease C1 was thought to have strict cleavage specificity for E↓EXXE [6]. However, kinetic studies

Abbreviations: Ac, acetylated; PA, protease-associated (The nomenclature of Schechter and Berger (1968) is used to denote residues of the substrates and subsites on the proteases. Amino acid residues on the proteases are counted beginning from the N-terminal methionine.).

* Corresponding author. Division of Basic Sciences, The Commonwealth Medical College, 525 Pine Street Room 2032, Scranton, PA 18509, USA. Tel.: +1 570 504 9669 (office); fax: +1 570 504 9636.
E-mail addresses: mprabu@tcmedc.org, moses_prabu@hotmail.com (M. Prabu-Jeyabalan).

0981-9428/$ – see front matter © 2012 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.plaphy.2012.01.005

of protease C1 with a model peptide substrate Ac-KVEKE↓ESEE-NH_2 (cleavage site denoted by "↓") and its variants showed that other residues besides Glu were accepted to different degrees at the P1 and P1′ sites. Also, many residues were as effective as Glu at the P4′ site [11]. In situations like this, where there is no clear consensus, intermolecular interactions and substrate shape may play a role in molecular recognition. This was found previously for human immunodeficiency virus type 1 (HIV-1) protease that cleaves diverse multiple sequences during viral maturation. Structural studies showed that this aspartyl protease recognizes a consensus "substrate-envelope" motif [12]. The substrate-envelope also explained the molecular consequences of drug resistance and co-evolution of HIV protease substrates [12–16] making it a chief player to confer specificity. If this also holds for protease C1, we infer from kinetic studies [11] that the number of amino acid residues flanking the peptide bond to be cleaved is an important feature of the substrate motif. Acting on Ac-KVEKE-↓ESEE-NH_2 that spans from P5 to P4′, the k_{cat}/K_M value of protease C1 decreases 40-fold when the P5 and P4 residues were eliminated, and an 80-fold decrease when the P4′ and P3′ residues were removed. We hypothesize that a requirement for such extensive interaction between enzyme and substrate residues from P5 to P4′ would be difficult to achieve when the cleavage site is within cupin domains. The compact double-barreled structures are difficult for many proteases to access [17]. The short linker region is where later-acting enzymes initiate the final degradation of the storage proteins [10,18]. Interaction with an extensive portion of the linker anchored to two barrel structures is also likely to be difficult for a protease that requires extensive interaction with the substrate. A structural examination of protease C1 may explain why the enzyme requires a span of as many as 9 residues to achieve optimum cleavage rates. Such explanation would help to correlate the structure of the enzyme with its role of providing some amino acids to the seedling early in growth while preserving the rest of the stored proteins for later use. Such a study is now possible with the report of the first crystal structure of a plant subtilase, tomato (*Solanum lycopersicum*) subtilase 3 (SBT3) [19]. The active site architectures of SBT3 and bacterial (*Bacillus amyloliquefaciens*) subtilisin (BPN′) are well conserved in the S8 catalytic domains, suggesting that homology modeling could elucidate the interaction of protease C1 with its substrates.

A feature that both protease C1 and SBT3 have in common, that subtilisin BPN′ does not have, is a protease-associated domain (PA domain) within the catalytic domain [20]. This insert is also present in 53 of the 56 subtilase genes of *Arabidopsis thaliana* [21] and in 58 of the 63 subtilases of rice [22]. Besides the PA domain of SBT3, there is structural information for the PA domain of another subtilisin enzyme, *Streptococcus agalactiae* C5a peptidase [23]. Both reports describe a 7-stranded beta barrel with three peripheral α-helices for the PA domain. A regulatory function has been ascribed to the PA domain of tomato SBT3. Homo-dimerization brings the PA domain of one protomer in contact with a β-hairpin in the other protomer, lifting presumed autoinhibition [19]. Homology modeling of representative Arabidopsis subtilases from different subfamilies showed that the PA domain may play a role in activation of some but not all of the subtilases analyzed [24]. The crystal structure of *Streptococcus pyogenes* C5a peptidase suggests another role for the PA domain. When this enzyme was docked with the C5a peptide substrate, the PA domain was found to interact with the P4′ and P5′ residues of C5a [25]. The importance of these interactions was reflected in a 200-fold decrease in the value of k_{cat}/K_M for a p-nitroanilide substrate (P9–P1–pNA) compared to the value when the same 9 residues were extended for 7 more residues (P9–P7′) [26]. This analysis will exploit the examples in homologous enzymes to elucidate a potential role for the PA domain of protease C1 either in autoregulation as shown for tomato SBT3, or in molecular recognition of the substrate as in the *S. pyogenes* C5a peptidase.

2. Results

2.1. Overall comparison

Soybean protease C1 and tomato SBT3 share a 38% sequence identity (52% homology) throughout the 628 and 649 residues, respectively, of their active forms (Alignment of the sequences of soybean protease C1, tomato SBT3 and the bacterial subtilisin BPN′ are presented, in Fig. S1 of Supplementary data. The residues are numbered starting from the presumed N-terminal methionine residues.). A three-dimensional model for protease C1 was generated by homology modeling using the crystal structure of SBT3 complexed with Ac-Phe-Glu-Lys-Ala-chloromethyl ketone (cmk) [19] (Fig. S1b). The three disulfide bonds of SBT3 (C382–C401, C170–C181, C624–C645) are also seen in the protease C1 model (C383–C400, C165–C173, C607–C626). The sequence corresponding to the autoinhibitory β-hairpin found in SBT3 (F517–L527) [19,24] is different from protease C1 (P510–R519). Since the oligomeric state of protease C1 is unknown, the significance of this sequence difference with respect to autoinhibition cannot be predicted. All except one of the insertion–deletion sites are remote from the active site and less likely to impact the observations of this analysis. The same can be said for the predicted N-glycosylation sites – N170 and N370 on protease C1 corresponding to N177 and N376 on SBT3. All of the N-glycosolation sites are remote from the active site and the changes in locations are less likely to make significant impacts on substrate binding. The modeled catalytic residues – D139, H203 and S530 – are situated in orientations that are consistent with the disposition observed in other serine proteases. With only a few differences in the loop regions, the overall active site geometry appears to be retained between protease C1 and SBT3. Therefore, this model of protease C1 will be suitable for a qualitative assessment of its substrate recognition.

2.2. Active-site loops

The primary (Fig. 1a) and tertiary (Fig. 1b and c) structures of active site Loops 1–5 for SBT3, BPN′ and protease C1 are conserved. Loop 6 of protease C1 follows that of SBT3, except that protease C1 may have an additional disulfide (C244–C249) which is likely to provide structural rigidity to the active site. Large back-bone differences are observed for Loop 6 of both plant subtilases when compared to BPN′; F255–T261 in SBT3 undergoes an outward flip in comparison to L203–S208 in BPN′ by an average linear distance of 4.5 Å. This leads to a conspicuous absence of a β-strand which in the subtilisin BPN′ active site (starting from G207) forms an antiparallel β-sheet arrangement with the P5–P1 region of the chymotrypsin inhibitor 2 (CI2) bound to its active site [27]. The antiparallel β-sheet pattern has also been described between subtilisin BPN′ and the *Streptomyces* subtilisin inhibitor [28]. The large structural difference in Loop 6 of the two plant subtilases in comparison to subtilisin BPN′ suggests that the plant subtilases differ from the bacterial enzyme in their manner of interaction with the P5–P1 region of their substrates. Another difference between the bacterial and the plant enzymes is the insertion of Loop 7 on the carboxyl side of the substrate which is present on both plant subtilases but not on subtilisin BPN′ (Fig. 1c). The residues of the modeled peptide substrates carboxyl to the cleaved bond interact with the PA domains of SBT3 (T343–Y472) and protease C1