Determination of the substrate specificity of turnip mosaic virus NIa protease using a genetic method

Hara Kang, Yong Jae Lee, Jae Hwan Goo and Woo Jin Park

Department of Life Science and National Research Laboratory, Kwangju Institute of Science and Technology, 1 Oryong-dong, Puk-gu, Kwangju 500-712, Korea

The RNA genome of turnip mosaic potyvirus (TuMV) encodes a large polyprotein that is processed to mature proteins by virus-encoded proteases. The TuMV NIa protease is responsible for the cleavage of the polyprotein at seven different locations. These cleavage sites are defined by a conserved sequence motif Val-Xaa-His-Gln, with the scissile bond located after Gln. To determine the substrate specificity of the NIa protease, amino acid sequences cleaved by the NIa protease were obtained from randomized sequence libraries using a screening method referred to as GASP (genetic assay for site-specific proteolysis). Based on statistical analysis of the obtained sequences, a consensus substrate sequence was deduced: Yaa-Val-Arg-His-Gln Ser, with Yaa being an aliphatic amino acid and the scissile bond located between Gln and Ser. This result is consistent with the conserved cleavage sequence motif, and should provide insight into the molecular activity of the NIa protease.

Turnip mosaic virus (TuMV) is a member of the Potyviridae, part of the picornavirus superfamily (Shukla et al., 1994), and is considered to be the most significant and widespread virus infecting cruciferous crops (Shattuck et al., 1989). The genome of TuMV is composed of a positive-sense single-stranded RNA of 9830 nucleotides, and is translated into a polyprotein of 3863 amino acids with a calculated molecular mass of 358 kDa (Nicolas & Laliberté, 1992). This polyprotein is processed into 10 mature viral proteins by virus-encoded proteases, the N-terminal protein (P1), the helper component protease (HC-Pro) and the nuclear inclusion protease (NIa protease) (Riechmann et al., 1992). The NIa protease is a trypsin-like cysteine protease that cleaves the TuMV polyprotein at seven distinct locations. These cleavage sites are marked by the highly conserved sequence motif Val-Xaa-His-Gln, with the scissile bond located after Gln. However, the substrate specificity of the NIa protease has not been investigated intensively.

In this study, we have determined the substrate specificity of the NIa protease by utilizing a recently developed screening method referred to as GASP (genetic assay for site-specific proteolysis) (Hawkins et al., 1999; Steiner et al., 1999; Kim et al., 2000; Kang et al., 2001). The principle of GASP is shown in Fig. 1(a). A fusion protein, SteSubLex, was generated in which a transcription factor, LexA-b42, is linked to the truncated cytoplasmic domain of an integral membrane protein, STE2, by a short linker containing a substrate sequence of the NIa protease. The rationale behind this approach was that: (i) in the absence of NIa protease activity, LexA-b42 remains anchored to the cytoplasmic membrane (left panel); (ii) in the presence of the NIa protease, the substrate sequence is cleaved, resulting in the release of LexA-b42, which can in turn enter the nucleus and activate reporter genes (right panel).

The construction of two backbone vectors, pADH-SteLex and pGAL, has been described previously (Kang et al., 2001). We further constructed a substrate vector, pADH-SteSubLex, in which an oligonucleotide encoding a dodecapeptide spanning the P10–P2′ sites of the 6K1/CI junction of the polyprotein (Nicolas & Laliberté, 1992) was inserted between the STE2 and LexA-b42 coding sequences of pADH-SteLex (Fig. 1b). Constitutive expression of the resultant fusion proteins, SteLex and SteSubLex, was driven by the ADH promoter. A DNA fragment encoding the NIa protease was amplified by PCR and inserted into pGAL, resulting in the protease vector pGAL-NIa. Expression of the NIa protease was placed under the control of the inducible GAL promoter. A Lex2 mutant reporter strain, EGY48 (Estojak et al., 1995), was transformed with both pADH-SteSubLex and pGAL-NIa, and spotted onto either non-selective or selective plates lacking Leu and containing either glucose or galactose (row 4). This indicated that expression of the NIa protease was strictly regulated by the GAL promoter, and that the substrate sequence in SteSubLex was specifically cleaved by the NIa protease. Further, results of the control experiments...
performed with the blank vectors pADH-SteLex and pGAL indicated that the proteolysis was highly specific (rows 1, 2, 3).

Good substrate sequences for the Nla protease were selected basically as described previously (Kim et al., 2000). Briefly, we incorporated randomization into the P6-P5-P4, P3-P2 and P1-P1 positions of the 6K1/CI junction in separate experiments. Each randomized sequence was designed to contain any of the 20 natural amino acids. Yeast cells harbouring pGAL-Nla were transformed with plasmid DNA of these randomized substrate libraries and then plated onto the plates lacking Trp and His (Trp1 and His3 are the marker genes for the substrate and protease vectors, respectively), and then replica-plated to selective plates lacking Trp, His and Leu (Leu2 is the reporter gene that is activated by LexA-b42) and containing either glucose or galactose. They were incubated for 3–4 days at 30 °C. Colonies that grew on selective plates containing galactose but not on plates containing glucose were assumed to harbour substrate sequences that were specifically cleaved by the Nla protease. To select substrate sequences which were cleaved as efficiently as the 6K1/CI junction site, colonies that grew as fast as transformants harbouring pADH-SteSubLex and pGAL-Nla were selected. These colonies were further tested on X-Gal plates and only strong blue colonies were further characterized (lacZ is another reporter gene that is activated by LexA-b42). The substrate regions of the selected positive colonies were amplified by PCR and directly subjected to nucleotide sequencing. The results are shown in Fig. 2, and the deduced consensus sequence is shown in Fig. 3. Val, His and Gln were the most frequently occurring residues at positions P4, P2 and P1, respectively. This result is consistent with the fact that these amino acids are the most highly conserved residues in the cleavage junctions of the TuMV polyprotein (Nicolas & Laliberté, 1992; Yoon et al., 2000). In particular, His appeared to be strictly required at position P2 for efficient cleavage. Previous studies have shown that Gln at P1 is essential for the cleavage by tobacco etch virus Nla proteases (Dougherty et al., 1988). Notably, the type of amino acid at position P1 was less strict according to results from our experiments. Considering that some of the cleavage sites of TuMV do not have Gln at P1 and they are less efficiently cleaved (Kim et al., 1996), it is likely that a small number of suboptimal sequences were selected during screening. Ser or Ile were the two most favoured residues at position P1. Arg was the most frequently occurring residue at P3, although eight other amino acids were also found at this position. Aliphatic amino acids (Gly, Ala, Leu, Ile) were favoured at P5. No obvious preference was observed at position P6.

![Fig. 1. Utilization of GASP to detect the activity of the Nla protease. (a) A fusion protein, SteSubLex, is expressed in a yeast strain that contains LexA-b42-inducible reporter genes. LexA-b42 is anchored in the plasma membrane via an integral membrane protein (STE2) and a substrate sequence for the Nla protease (left panel). When the Nla protease is introduced, the substrate sequence is cleaved, resulting in the release of LexA-b42 and subsequent activation of reporter genes (right panel). (b) The structure of SteSubLex is shown. The dodecapeptide substrate sequence (ATVEPTVYHQTL) is derived from the 6K1/CI cleavage site of the TuMV polyprotein. (c) Yeast cells transformed with substrate and protease vectors as indicated were diluted tenfold each and spotted on the non-selective (THU-) and selective plates (THUL-). Only transformants harbouring pADH-SteSubLex and pGAL-Nla (row 4) were able to grow on selective plates containing galactose but not glucose. This indicates specific cleavage of the substrate sequence by the Nla protease in yeast.](image-url)
Substrate specificity of TuMV NIa protease

Fig. 2. The underlined sequences, PTV, YH, and QT, were randomized in separate experiments. The obtained cleavable sequences are shown in each box.

Fig. 3. (a) The consensus motif for the TuMV NIa protease is shown. (b) The consensus cleavage sequence obtained from this study is shown. The heights of each box and the numbers under each box represent the percentage of the occurrence of the amino acid in specific locations. Yaa indicates aliphatic amino acids (Glu, Ala, Leu, Ile).

employed to detect caspase activity (Hawkins et al., 1999; Steiner et al., 1999) and to determine the substrate specificity of hepatitis C virus NS3 protease (Kim et al., 2000) and hepatitis G virus NS3 protease (Lee et al., 2001). In this study, we have successfully determined the substrate specificity of the TuMV NIa protease by using GASP, and suggest that the technique might be used for the study of other viral proteases. GASP offers a more reliable and simpler method than phage display or standard in vitro biochemical methods in the study of substrate specificity of proteases. The consensus substrate sequence obtained from this study should be useful for substrate-based design of the NIa protease inhibitors.

This work was supported by grants from the National Research Laboratory, Molecular Medicine Research Group Program (98-103-02-03-A-01) from the Ministry of Science and Technology, and BK21 program from the Ministry of Education, Korea. We are grateful to Dr Kwan Yong Choi (POSTEC) for providing us with TuMV NIa cDNA.

References


Received 20 June 2001; Accepted 17 August 2001