BIOENGINEERED
2016, VOL. 0, NO. 0, 1–19
http://dx.doi.org/10.1080/21655979.2016.1208868

Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

Agroinfiltration contributes to VP1 recombinant protein degradation

Priyen Pillay[a], Karl J. Kunert[a], Stefan van Wyk[a], Matome Eugene Makgopa[a], Christopher A. Cullis[b], and Barend J. Vorster[a]

[a]Department of Plant and Soil Sciences, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Hillcrest, Pretoria, South Africa; [b]Department of Biology, Case Western Reserve University, Cleveland, OH, USA

ABSTRACT
There is a growing interest in applying tobacco agroinfiltration for recombinant protein production in a plant based system. However, in such a system, the action of proteases might compromise recombinant protein production. Protease sensitivity of model recombinant foot-and-mouth disease (FMD) virus P1-polyprotein (P1) and VP1 (viral capsid protein 1) as well as *E. coli* glutathione reductase (GOR) were investigated. Recombinant VP1 was more severely degraded when treated with the serine protease trypsin than when treated with the cysteine protease papain. Cathepsin L- and B-like as well as legumain proteolytic activities were elevated in agroinfiltrated tobacco tissues and recombinant VP1 was degraded when incubated with such a protease-containing tobacco extract. *In silico* analysis revealed potential protease cleavage sites within the P1, VP1 and GOR sequences. The interaction modeling of the single VP1 protein with the proteases papain and trypsin showed greater proximity to proteolytic active sites compared to modeling with the entire P1-polyprotein fusion complex. Several plant transcripts with differential expression were detected 24 hr post-agroinfiltration when the RNA-seq technology was applied to identify changed protease transcripts using the recently available tobacco draft genome. Three candidate genes were identified coding for proteases which included the Responsive-to-Desiccation-21 (RD21) gene and genes for coding vacuolar processing enzymes 1a (*Nb*VPE1a) and 1b (*Nb*VPE1b). The data demonstrates that the tested recombinant proteins are sensitive to protease action and agroinfiltration induces the expression of potential proteases that can compromise recombinant protein production.

ARTICLE HISTORY
Received 9 June 2016
Revised 24 June 2016
Accepted 28 June 2016

KEYWORDS
agroinfiltration; cysteine proteases; foot-and-mouth disease; proteases; recombinant protein production; tobacco; VP1 protein

Introduction

In 2014, an experimental drug, called "ZMapp," was used to treat 2 medical workers who had contracted the deadly Ebola virus. The unique characteristic of ZMapp is that its constituents have been produced in *Nicotiana benthamiana* plants.[1] *N. benthamiana* is a model plant species widely used for the transient expression of proteins. Tobacco is sometimes compared to the role that the white mouse has played in mammalian studies.[2-5] The *N. benthamiana* genome sequence has further potential to be useful for gene mining, construct design, and for the assessment of target and non-target gene silencing.[2] A future prospect is also applying RNA-Seq data to fully annotate the tobacco genome and characterize the transcriptome.[2] The large leaves of *N. benthamiana* and its susceptibility to a variety of pathogens have been harnessed as a means to transiently express proteins, using either engineered viruses or *Agrobacterium tumefaciens*.[6] Due to a lower content of secondary compounds interfering in the protein purification process, *N. benthamiana* has been previously applied as a model plant species for heterologous protein expression.[7] It has also been included as a tool in platforms for the production of recombinant proteins for comparative analyses.[8]

Due to proteolysis caused by protease action, plant-expressed recombinant proteins can possibly undergo either complete or partial proteolytic degradation.[9-11] Such degradation can ultimately result in proteins with altered biological activity or no protein production at all.[12,13] The identification of such proteases

CONTACT Barend J. Vorster juan.vorster@up.ac.za Department of Plant and Soil Sciences, University of Pretoria, Private Bag X20, Hillcrest, Pretoria, South Africa.

Color versions of one or more of the figures in this article can be found online at www.tandfonline.com/kbie.
Supplemental data for this article can be accessed on the publisher's website.

involved, particularly in *Nicotiana* species, has therefore been the subject of several recent investigations. The majority of protease families, which might compromise recombinant protein production in *Nicotiana* species, belong to the aspartic and cysteine protease (papain-like) families and, to a lesser extent, the serine and metallo-protease families.[14,15] There is further evidence that such recombinant protein degradation might occur during the extraction process *ex vivo* as a result of proteases being released during the tissue disruption process.[16] However, almost all protease families have also been associated with plant senescence.[17] In *Nicotiana* species, the majority of these proteases are of aspartic or cysteine-type and, to a lesser extent, of serine and metallo-type.[18] However, the *N. benthamiana* leaf contains less protease activity than a *N. tabacum* leaf and is therefore preferred for agroinfiltation.[15] It has been recently reported that agroinfiltration can significantly alter the distribution of cysteine (C1A) and aspartate (A1) protease along the leaf age gradient in *N. benthamiana*. This was further related to the level of proteolysis in whole-cell and apoplast protein extracts.[19] Improvements have been found for various recombinant proteins when protease activity was altered including: bovine serum albumin (BSA), human serum immunoglobulins G (hIgGs), anti-HIV antibodies (2F5) as well as human protease inhibitor, α1-antitrypsin.[13,20,21] However, there is a lack of detailed information on whether induction of plant-derived proteases is among the host responses to agroinfiltration.[22] Therefore, the identification of proteases induced by agroinfiltration might be a key step in the improvement of recombinant protein production when applying the agroinfiltration technique.

The purpose of this study was to investigate protease-sensitivity of various model recombinant proteins with the aim to first establish any protease sensitivity and secondly to identify possible proteases expressed as a consequence of the agroinfiltration process. We specifically hypothesized that cysteine proteases are among these expressed proteases following agroinfiltration based on a previous finding in our group that recombinant *E. coli* glutathione reductase (GOR) was more stable in agroinfiltrated tobacco leaves engineered with a rice cysteine protease inhibitor (OC-I).[23] In our study, we determined the inherent vulnerabilities of recombinant model proteins derived from the foot-and-mouth disease virus (FMDV) which are the VP1 and P1-polyprotein (P1) as well as *Escherichia coli* (*E. coli*)-derived glutathione reductase (GOR) proteins toward proteolysis. We also applied protein modeling to investigate how interacting residues within VP1 would interact with a cysteine and serine protease (papain and trypsin) either individually or as part of a P1-polyprotein to obtain more information on VP1 stability against protease action. Finally by applying transcriptomic profiling using RNA-Seq, *N. benthamiana* leaves were screened for the transcription of proteases due to agroinfiltration. We found that the recombinant model proteins used were sensitive to cysteine and serine protease degradation and that expression of several types of proteases, including cysteine proteases, increased due to the agroinfiltration of tobacco leaves.

Results

Protease sensitivity of model recombinant proteins

Since VP1 was used in the study as one of the model recombinant proteins, the VP1 protein (Fig. 1) was first treated with either a cysteine (papain) or serine (trypsin) protease to determine VP1 sensitivity to protease treatment (Fig. 2). Both proteases degraded VP1 when determined by SDS-PAGE analysis, but with more severe VP1 degradation occurring when treated with trypsin (Fig. 2b). Less degradation occurred when either E64, a cysteine protease inhibitor, or TLCK, a trypsin inhibitor, was added to the reaction mixture (Fig. 2a, b). In order to also investigate the influence of proteases *ex planta*, VP1 was further treated using tobacco extracts with and without the addition of a protease inhibitor (Fig. 3). VP1 band intensity changed indicating possible protease action; also, some smaller sized bands cross-reacted with the His-antiserum possibly indicating some proteolytic degradation products (Fig. 3a). Furthermore, a tobacco extract containing a protease inhibitor resulted in less VP1 degradation (Fig. 3b, c).

Protease cleavage sites

In a second step we investigated *in vitro* if model recombinant proteins VP1, GOR as well as P1-polyprotein (VP2 - 4) have proteolytic cleavage sites (Table 1). Subsite nomenclature was assumed from a model created by Schechter and Berger (1967, 1968) where the amino acid residues in a substrate undergoing proteolytic cleavage are designated P1, P2, P3, P4,

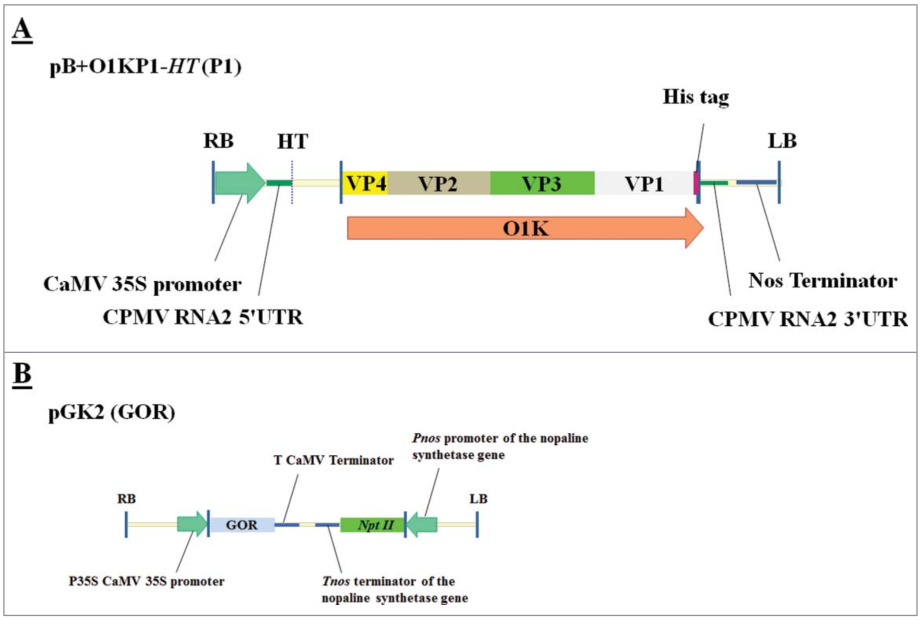


Figure 1. (A) Binary vector pB+O1KP1-*HT* harbouring the coding sequence O1K under control of a duplicated cauliflower mosaic viSSS-rus 35S promoter and a t-nos terminator sequence and O1K consisting of fused VP1 - 4 coding sequences with VP1 fused to a 6xHis coding sequence (P1-polyprotein). (B) Schematic representation of GOR T-DNA used for agroinfiltration. Binary vector pGK2 (GOR) harbouring the GOR gene. RB and LB refers to the right and left border, respectively. *NptII* refers to the neomycin phosphotransferase gene which confers kanamycin resistance.

etc., in the N-terminal direction from the cleaved bond. VP1 was particularly susceptible to papain cleavage with 3 papain amino acid sites (C^{25}, H^{159}, N^{175}) involved in the interaction. When amino acid T (threonine) was used in the P1 substrate position, 62 cleavage sites were found in the polyprotein (Table 1, highlighted in yellow). Within the GOR sequence, 34 cleavage sites were found when amino acid G (glycine) was used in the P1 substrate position (Table 1, highlighted in yellow). Cleavages sites with amino acid W (tryptophan) in the P1 substrate position were, however, not highly abundant in either the P1-polyprotein or GOR sequences with only 3 and 2 sites, respectively, being present (Table 1, highlighted in turquoise). Papain cleaves at TL^{213} at the end of the VP1 sequence as well as AE^{220} at the end of the VP3 sequence while cathepsin L cleaves at KE^{218} at the end of the VP2 sequence (Table 3.1, highlighted in gray). Cathepsin H further cleaves VP1 at L^{213} and at R^{478} of the GOR sequence (Table 1, highlighted in gray). Cathepsin H cleavage sites, with K and L at the P1 substrate position, were further abundant for GOR

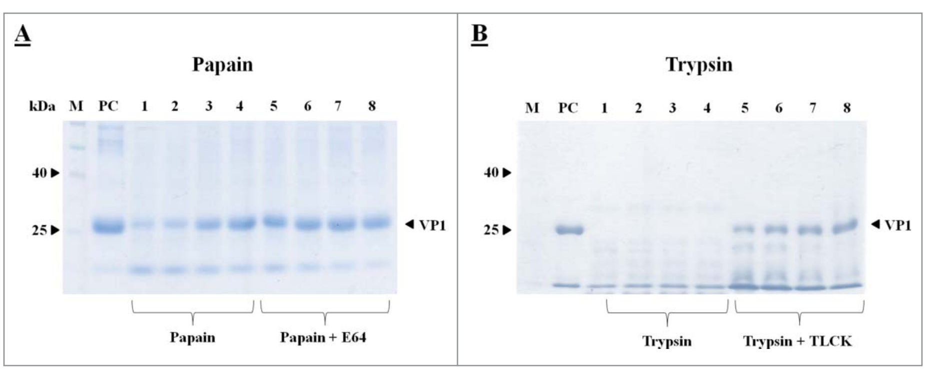


Figure 2. SDS-PAGE analysis of VP1 susceptibility to papain (A) and trypsin (B). Samples were analyzed on a 12% SDS-polyacrylamide gel. All sample lanes contain 44 μg of VP1 treated with either papain or trypsin. Lanes 1 to 4 contain either papain (0.2 μg, 0.3 μg, 0.4 μg and 0.5 μg) or trypsin (0.2 μg, 0.3 μg, 0.4 μg and 0.5 μg). Lanes 5 to 8 contain VP1 treated with papain (0.2 μg, 0.3 μg, 0.4 μg and 0.5 μg) together with the cysteine protease inhibitor E64 (1 μM) or VP1 treated trypsin (0.2 μg, 0.3 μg, 0.4 μg and 0.5 μg respectively) together with the serine protease inhibitor TLCK (40 mM). PC (44 μg purified VP1) represents the positive control and M represents a pre-stained protein ladder.

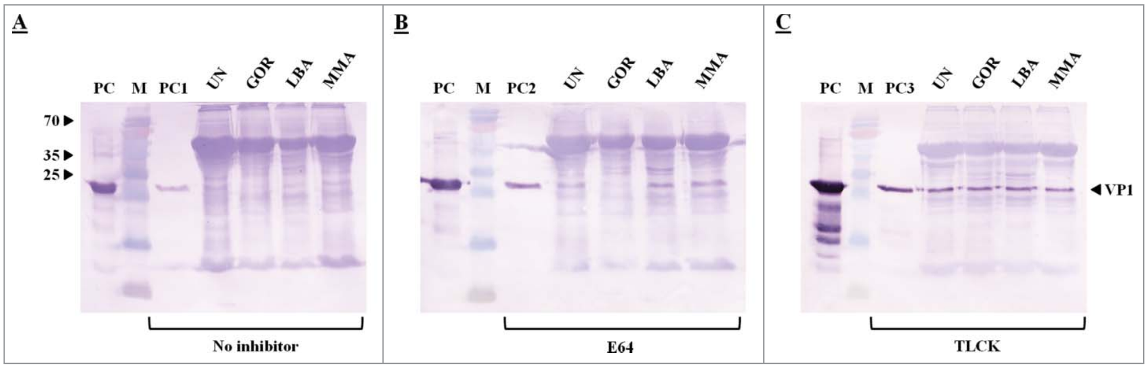


Figure 3. Western blot analysis of VP1 susceptibility with Anti-His antiserum of exogenous VP1 incubated with *N. benthamiana* leaf extracts with or without protease inhibitors. Lanes 1-4 represent 15 μg of VP1 treated with a tobacco extract (230 μg) with either no protease inhibitor (A), a cysteine protease inhibitor E64 (100 μM) (B) or a serine protease inhibitor TLCK (100 μM) (C) added to Arakawa extraction buffer. Lanes UN, GOR, LBA and MMA represent VP1 (15 μg) incubated with tobacco extract only, GOR agroinfiltrated extract, LBA agroinfiltrated extract and MMA agroinfiltrated extract, respectively. Lane M represents a Pageruler pre-stained protein ladder. PC represents VP1 positive control (60 μg). PC1, PC2 and PC3 represent 15 μg of VP1 incubated with Arakawa extraction buffer.

and VP1 sequences with 34 and 33 sites, respectively (Table 1, highlighted in yellow). VP1 seems to be particular susceptible to cathepsin H cleavage with a total of 61 potential cleavage sites (Table 1, highlighted in yellow).

To further determine VP1 susceptibility to papain, or trypsin action, ZDOCK protein modeling was then applied between VP1 and the 2 proteases and the distances (in Ångstroms) between interacting VP1 and protease residues (VP1 cleavage site R^{67} for papain and VP1 cleavage site R^{26} for trypsin) were determined. The distance in Ångstroms (Å) decreased when VP1 was not modeled together with the additional capsid proteins (VP2, VP3 and VP4) indicating higher VP1 sensitivity to protease action (Fig. 4, bottom panels). In contrast, when all other binding sites for VP2, VP3 and VP4 within the P1-polyprotein were permitted in the interaction model with the protease, the distance increased with a weaker VP1-protease interaction and better stability of VP1 (Fig. 4, top panels).

Transcriptome analysis of agroinfiltrated tobacco leaves

Since we found no significant statistical differences in proteolytic activities between experimental (LBA, P1, GOR) and control (UN & MMA) groups when various cysteine protease activities (cathepsin L and H, legumain proteolytic activities) were measured with fluorogenic substrates 24 hr post infection (pi) (Fig. 5), possible expression of proteases due to agroinfiltration was also investigated using RNA-seq analysis. When a limit of at least 2-times higher protease expression in agroinfiltrated tissues based on FPKM (fragments per kilobase of transcript per million mapped reads) values was set, the most expressed proteases were cysteine proteases (7 proteases) belonging to 3 different cysteine protease families (C1, C13 and C14) (Fig. 6, Table 2) with most cysteine proteases belonging to the C1 family of cysteine proteases (4 proteases). Further, other proteases expressed as a consequence of agroinfiltration belonged to metalloproteases (2 proteases) in 2 families (M38 and M67), aspartic proteases (2 proteases) in 2 families (A1 and A22) and threonine proteases (1 protease) in 1 family (T1) (Table 2). In comparison to all other proteases, the cysteine protease RD21 (XM_009614860.1) was the highest expressed cysteine protease. Agroinfiltration increased expression of this cysteine protease about 4-times (Table 2). The protease most induced was *Nb*VPE-1b, a vacuolar processing enzyme, belonging to the C13 cysteine protease family (Table 2). This protease was expressed about 95-times higher (based on FPKM-values) in agroinfiltrated leaves than in non-infiltrated leaves. Such a high expression was also found when tobacco leaves were infiltrated with a construct to produce the P1-polyprotein (75-times) and with a construct for GOR production (92-times). For all other identified proteases belonging to different classes, expressions in non-treated leaves were much lower compared to agroinfiltration-induced expression which was in the range of 2-30-times more. This increase was irrespective of infiltration with *Agrobacterium* alone or with a construct allowing P1-polyprotein or GOR expression.

Table 1. *In silico* proteolytic cleavage site analysis for P1-polyprotein, VP1 and glutathione reductase (GOR) using primarily the P1 and P2 substrate positions within amino acid sequences.

Protease *Cysteine protease*	P1-polyprotein	GOR	Cleavage	Preferential Cleavage
Papain (C^{25}, H^{159}, N^{175})	VP1 chain – AR$^{157, 200}$, LR$^{67, 145}$, YR^{179}, TR$^{114, 172}$, PR^{189}, DR^{38}, SR^{138}, QR^{26}VP2 chain – TR$^{18, 54}$, MR^{102}, PR^{151}, DR^{13}, ER^{60}VP3 chain – R^{56}, R^{120}, R^{212}, R^{218}, PR^{34}, DR^{72}VP4 chain – none	AR$^{37, 45}$, LR$^{224, 478}$, IR$^{127, 218}$, WR^{97}, SR^{109}	AR↓, VR↓, LR↓, IR↓, FR↓, WR↓, YR↓, TR↓, MR↓, PR↓, DR↓, SR↓, ER↓, QR↓	(A, V, L, I, F, W, Y – hydrophobic amino acids in P2 position)(↓ no Val after cleavage sites)
	VP1 chain - VK$^{81, 210}$, IK^{169}, MK^{181}, EK^{96}, QK^{204}VP2 chain – FK^{63}, IK^{198}, DK^{2}, SK^{217}, QK^{134}VP3 chain – AK$^{84, 118}$, TK^{67}, PK$^{20, 134}$, SK^{154}VP4 chain – SK^{70}	VK$^{252, 255, 324, 397, 452}$, LK^{247}, IK^{100}, YK^{357}, TK$^{120, 296, 412, 457}$, MK^{420}, PK$^{66, 137}$, DK$^{221, 310}$, SK$^{212, 361}$, EK^{102}	AK↓, VK↓, LK↓, IK↓, FK↓, WK↓, YK↓, TK↓, MK↓, PK↓, DK↓, SK↓, EK↓, QK↓	
	VP1 chain – IH^{195}, YH^{108}, PH^{123}, SH^{59}, QH^{29}VP2 chain – AH^{157}, VH^{174}, TH^{65}, PH^{145}, DH^{87}VP3 chain – AH^{141}, LH^{108}, IH^{144}, TH^{191} VP4 chain – TH^{58}	AH^{351}, VH^{75}, IH$^{389, 434, 467}$, YH^{408}, MH^{80}, PH$^{151, 164}$, DH^{82}, SH$^{52, 122, 374}$	AH↓, VH↓, LH↓, IH↓, FH↓, WH↓, YH↓, TH↓, MH↓, PH↓, DH↓, SH↓, EH↓, QH↓	
	VP1 chain – AE^{184}, VE^{16}, LE^{77}, TE$^{175, 198}$, PE^{95}VP2 chain – AE$^{40, 59}$, VE^{108}, LE$^{11, 52, 82}$, TE$^{5, 192}$, PE^{128}, PE^{137}, EE^{6}VP3 chain – AE$^{49, 146, 176, 220}$, FE$^{58, 210}$, ME^{131}VP4 chain – none	AE^{41}, VE$^{50, 201}$, LE$^{184, 239}$, IE$^{124, 141, 394}$, FE^{355}, TE$^{236, 384}$, DE$^{317, 386, 442}$, SE$^{77, 472}$, EE$^{185, 237, 428, 473}$, QE^{5}	AE↓, VE↓, LE↓, IE↓, FE↓, WE↓, YE↓, TE↓, ME↓, PE↓, DE↓, SE↓, EE↓, QE↓	
	VP1 chain – IN^{49}, FN^{164}, YN^{131}, TN^{103}, PN$^{91, 143}$, DN^{100}, EN^{17}, QN^{46}VP2 chain – AN^{202}, VN$^{166, 190}$, IN^{149}, FN^{117}, TN$^{153, 207}$, PN^{47}VP3 chain – LN^{152}, IN^{106}, FN^{31}, TN$^{43, 179}$, SN^{88}VP4 chain – IN^{23}, TN^{61}, DN^{41}, SN^{48}, QN$^{13, 32, 64}$	AN^{425}, LN$^{111, 301}$, FN^{95}, WN^{71}, YN^{365}, TN^{233}, PN^{294}, DN^{462}, EN$^{240, 395}$, QN^{116}	AN↓, VN↓, LN↓, IN↓, FN↓, WN↓, YN↓, TN↓, MN↓, PN↓, DN↓, SN↓, EN↓, QN↓	
	VP1 chain – AQ^{153}, IQ^{25}, TQ^{23}, PQ^{45}VP2 chain – VQ^{57}, IQ^{133}, YQ^{139}, TQ^{27}, PQ^{196}, DQ^{170}VP3 chain – AQ$^{76, 96}$, VQ^{181}, FQ^{188}, TQ^{100}VP4 chain – YQ^{31}, TQ^{37}, MQ^{28}, SQ^{12}, QQ^{29}	LQ^{445}, IQ^{306}, FQ$^{182, 319}$, YQ^{115}, MQ^{436}, PQ^{6,9}, SQ^{167}	AQ↓, VQ↓, LQ↓, IQ↓, FQ↓, WQ↓, YQ↓, TQ↓, MQ↓, PQ↓, DQ↓, SQ↓, EQ↓, QQ↓	
	VP1 chain – AS^{69}, FS^{74}, YS^{137}, TS$^{3, 162}$, PS^{58}, ES^{7}VP2 chain – VS^{44}, TS$^{24, 49, 72}$, PS^{216}, DS$^{74, 97}$, QS^{28}VP3 chain – AS$^{172, 204}$, LS^{163}, FS^{158}, YS^{102}, MS$^{80, 87}$, DS^{70}VP4 chain – AS^{73}, IS^{44}, FS$^{69, 77}$, TS^{56}, SS$^{6, 74}$, QS$^{5, 15}$	AS$^{20, 35, 172}$, VS$^{108, 143, 264}$, LS$^{207, 259, 299}$, IS^{231}, FS$^{249, 373}$, YS^{400}, TS$^{134, 177, 214, 402}$, PS$^{89, 161}$, DS$^{228, 360}$, SS^{471}, ES$^{51, 166}$	AS↓, VS↓, LS↓, IS↓, FS↓, WS↓, YS↓, TS↓, MS↓, PS↓, DS↓, SS↓, ES↓, QS↓	
	VP1 chain – AT$^{128, 171}$, VT$^{12, 43, 174}$, LT$^{87, 113}$, YT^{120}, TT$^{2, 13, 14, 102}$, PT$^{105, 161, 197}$, ST^{70}, ET$^{22, 185}$, QT^{212}VP2 chain – AT^{38}, VT$^{33, 71, 110}$, LT$^{16, 95, 141, 188}$, IT^{159}, WT^{178}, TT$^{8, 17, 23, 26}$, MT^{155}, PT$^{85, 206}$, ST^{25}, ET$^{7, 53}$VP3 chain – VT$^{16, 65}$, IT^{190}, FT$^{42, 112, 156}$, YT$^{99, 168, 170}$, TT$^{17, 66, 178}$, PT$^{53, 115}$, DT^{149}, ET^{177}VP4 chain – AT^{9}, TT$^{55, 60}$, DT$^{36, 54}$, ST$^{52, 57}$	AT$^{156, 456}$, VT$^{267, 411}$, LT$^{119, 339, 383}$, IT^{176}, FT$^{133, 404}$, YT^{148}, MT^{277}, PT$^{139, 369, 469}$, ST$^{162, 232, 401}$, QT^{307}	AT↓, VT↓, LT↓, IT↓, FT↓, WT↓, YT↓, TT↓, MT↓, PT↓, DT↓, ST↓, ET↓, QT↓	
	VP1 chain – AG^{5}, VG^{63}, YG$^{19, 166}$, EG^{84}VP2 chain – AG^{212}, VG$^{31, 113, 164}$, FG^{76}, YG$^{35, 92}$, SG$^{45, 50}$, EG^{193}VP3 chain – AG$^{93, 206}$, YG$^{12, 27}$, TG$^{113, 150}$, PG$^{39, 129}$, DG$^{10, 196}$, SG^{103}, EG^{59}, QG^{182}VP4 chain – AG^{3}, LG^{39}, FG^{81}, TG$^{10, 19}$, SG$^{16, 45, 78}$, EG^{50}	AG$^{16, 196, 204, 242, 346}$, VG$^{62, 194, 330, 381, 432}$, LG$^{43, 55, 174, 210, 304, 439}$, IG$^{27, 290, 378}$, YG$^{86, 392}$, TG^{157}, MG^{454}, PG$^{170, 188, 271}$, DG^{179}, SG$^{31, 144, 260}$, EG^{92}, QG$^{10, 437, 446}$	AG↓, VG↓, LG↓, IG↓, FG↓, WG↓, YG↓, TG↓, MG↓, PG↓, DG↓, SG↓, EG↓, QG↓	
	VP1 chain – YF^{73}, SF$^{34, 163}$VP2 chain – LF$^{67, 143}$, FF^{62}, PF^{162}, DF^{42}, SF^{75}, EF^{214}, QF$^{116, 147}$ VP3 chain – VF^{30}, LF^{187}, IF^{3}, TF$^{54, 90, 157}$, MF^{111}, DF^{209}, QF^{77}VP4 chain – AF^{76}, LF^{80}, WF^{68}	AF^{132}, VF^{372}, LF^{354}, FF^{181}, DF^{460}, SF$^{226, 403}$, EF$^{78, 318}$	AF↓, VF↓, LF↓, IF↓, FF↓, WF↓, YF↓, TF↓, MF↓, PF↓, DF↓, SF↓, EF↓, QF↓	
	VP1 chain – noneVP2 chain – PW^{177}VP3 chain – EW^{147}VP4 chain – DW^{67}	LW^{287}, MW^{70}	AW↓, VW↓, LW↓, IW↓, FW↓, WW↓, YW↓, TW↓, MW↓, PW↓, DW↓, SW↓, EW↓, QW↓	
	VP1 chain – AY^{107}, VY^{130}, LY^{178}, YY^{72}, TY$^{71, 186}$, PY^{119}VP2 chain – AY^{100}, VY$^{91, 200}$, LY^{138}, TY^{34}, SY^{98}, QY^{171}VP3 chain – AY^{125}, VY^{26}, YY^{98}, TY^{169}, PY$^{63, 161}$, DY^{167}, QY$^{97, 101}$VP4 chain – YY^{26}, QY^{30}	AY^{106}, IY$^{114, 327}$, TY^{399}, MY^{407}, DY$^{23, 85, 364}$, SY^{21}, EY^{356}	AY↓, VY↓, LY↓, IY↓, FY↓, WY↓, YY↓, TY↓, MY↓, PY↓, DY↓, SY↓, EY↓, QY↓	

(*Continued on next page*)

Table 1. (*Continued*)

Protease *Cysteine protease*	P1-polyprotein	GOR	Cleavage	Preferential Cleavage
	VP1 chain – AL$^{98, 117}$, VL$^{126, 151}$, LL$^{66, 177, 192}$, IL^{51}, TL$^{61, 159, 213}$, PL$^{112, 191}$, DL$^{53, 76, 86, 148}$, EL^{176}VP2 chain – LL$^{10, 81, 122}$, IL^{15}, TL$^{9, 142, 179}$, PL^{187}, SL^{94}, EL$^{83, 129, 137}$, QL^{140}VP3 chain – AL^{199}, VL^{74,202}, LL^{45}, FL$^{55, 91}$, YL^{162}, SL^{81}, EL^{211}, QL^{37}VP4 chain – AL^{83}, LL^{84}, QL^{38}	AL$^{209, 337}$, VL$^{223, 246}$, LL$^{286, 338}$, IL^{153}, TL^{258}, ML^{444}, DL^{298}, SL$^{173, 215, 300}$, EL$^{42, 186, 238}$, QL^{183}	AL↓, VL↓, LL↓, IL↓, FL↓, WL↓, YL↓, TL↓, ML↓, PL↓, DL↓, SL↓, EL↓, QL↓	
	VP1 chain – LM^{54}, IM^{36}VP2 chain – AM^{125}, YM^{101}VP3 chain – FM^{110}, YM^{122}, DM^{79}, SM^{34}VP4 chain – YM^{27}	VM$^{69, 276, 419}$, LM^{216}, FM^{79}, TM^{278}, PM^{406}, SM$^{229, 265}$, EM$^{202, 443}$	AM↓, VM↓, LM↓, IM↓, FM↓, WM↓, YM↓, TM↓, MM↓, PM↓, DM↓, SM↓, EM↓, QM↓	
Cathepsin L	VP1 chain – KR^{182}, KH^{82}, KQ^{203}, KA$^{97, 110, 170}$VP2 chain – KK^{3}, KR^{135}, KE^{218}, KT^{4}, KG^{89} VP3 chain – KH^{85}, KT$^{21, 68, 135}$, KF^{155}, KA$^{119, 194}$VP4 chain – KL^{71}	KK$^{67, 146, 256}$, KR$^{103, 413}$, KE$^{101, 253, 358, 427}$, KS^{121}, KT$^{213, 257, 398, 415}$, KG$^{311, 325}$, KF$^{94, 248}$, KY^{147}, KA$^{336, 458}$, KL$^{54, 303, 349, 362}$, KM^{453}	KK↓, KR↓, KH↓, KE↓, KN↓, KQ↓, KS↓, KT↓, KG↓, KF↓, KW↓, KY↓, KA↓, KL↓, KM↓,	• *Generally the same as papain with these additional specificities* • *Exceptions at the P2 position with respect to papain are EX$_{aa}$, QX$_{aa}$, TX$_{aa}$, AX$_{aa}$ where Cathepsin-L does not have specificity for*
	VP1 chain – noneVP2 chain – WT^{178}VP3 chain – noneVP4 chain – WF^{68}	WR^{97}, WN^{71}, WA^{288}	WK↓, WR↓, WH↓, WE↓, WN↓, WQ↓, WS↓, WT↓, WG↓, WF↓, WW↓, WY↓, WA↓, WL↓, WM↓	
	VP1 chain – HK$^{109, 202}$, HE^{83}, HT$^{30, 60}$VP2 chain – HK$^{88, 175}$, HQ^{146}, HT^{22}, HL$^{66, 80}$VP3 chain – HG^{192}, HF^{109}, HA^{145}, HM^{86}VP4 chain – HT^{59}	HK$^{53, 390}$, HR^{352}, HE^{165}, HS^{76}, HA$^{83, 130, 409}$, HM^{435}	HK↓, HR↓, HH↓, HE↓, HN↓, HQ↓, HS↓, HT↓, HG↓, HF↓, HW↓, HY↓, HA↓, HL↓, HM↓	
	VP1 chain – FI^{35}VP2 chain – HI^{158}, SI^{132}, FI^{148}VP3 chain – SI^{159}VP4 chain – SI^{21}, II^{22}	HI$^{123, 152, 313}$, PI^{377}, YI^{198}, VI$^{26, 99}$, II^{126}, LI^{154}, MI$^{217, 230, 279}$	KI↓, HI↓, DI↓, SI↓, PI↓, FI↓, WI↓, YI↓, VI↓, II↓, LI↓, MI↓	
	VP1 chain – KV$^{42, 155}$, DV^{32}, PV$^{11, 209}$, FV^{40}, WV^{89}, IV^{206}, LV^{62}VP2 chain – KV$^{173, 199}$, HV^{210}, DV^{107}, SV^{30}, FV$^{43, 163}$, WV^{70}, VV$^{56, 181}$, LV$^{123, 180}$, MV^{126}VP3 chain – KV^{29}, DV$^{47, 174}$, PV$^{5, 25, 215}$, WV^{184}, YV^{64}, VV^{201}, LV^{15}, MV^{123}VP4 chain – none	KV$^{68, 222}$, DV$^{282, 332}$, SV^{191}, PV$^{275, 341}$, YV^{107}, VV$^{49, 371, 431}$, IV$^{193, 315}$, LV$^{25, 475}$, MV$^{266, 422}$	KV↓, HV↓, DV↓, SV↓, PV↓, FV↓, WV↓, YV↓, VV↓, IV↓, LV↓, MV↓	
Cathepsin H	VP1 chain – R$^{26, 27, 38, 67, 114, 124, 135, 138, 145, 157, 172, 179, 182, 189, 200}$, K$^{41, 81, 96, 109, 154, 169, 181, 202, 204, 210}$, F$^{34, 39, 73, 163}$, W^{88}, L$^{51, 53, 61, 65, 66, 76, 86, 98, 112, 115, 117, 126, 144, 148, 151, 159, 176, 177, 191, 192, 213}$, Y$^{18, 71, 72, 107, 119, 130, 136, 165, 178, 186}$ VP2 chain – R$^{13, 18, 54, 60, 77, 102, 135, 151, 167}$, K$^{2, 3, 63, 88, 134, 172, 175, 198, 217}$, F$^{42, 61, 62, 67, 75, 116, 143, 147, 162, 214}$, W$^{69, 105, 177}$, L$^{9, 10, 15, 51, 66, 80, 81, 83, 94, 121, 122, 129, 137, 140, 142, 179, 187}$, Y$^{34, 36, 91, 98, 100, 138, 168, 171, 200}$ VP3 chain – R$^{34, 40, 56, 72, 120, 212, 218}$, K$^{20, 28, 67, 84, 118, 134, 154, 193, 207}$, F$^{3, 30, 41, 54, 57, 77, 90, 109, 111, 155, 157, 187, 209}$, W$^{147, 183}$, L$^{14, 37, 44, 45, 55, 74, 81, 91, 94, 107, 151, 162, 186, 199, 202, 211, 213}$, Y$^{11, 26, 63, 97, 98, 101, 121, 125, 161, 167, 169}$ VP4 chain – K^{70}, F$^{68, 76, 80}$, W^{67}, L$^{38, 71, 79, 83, 84}$, Y$^{25, 26, 30}$	R$^{3, 37, 38, 45, 97, 103, 109, 127, 189, 218, 224, 272, 291, 347, 352, 413, 478}K^{53, 66, 67, 93, 100, 102, 120, 137, 145, 146, 212, 221, 247, 252, 255, 256, 296, 302, 310, 324, 335, 348, 357, 361, 390, 397, 412, 414, 416, 420, 426, 429, 452, 457}F^{78, 87, 94, 132, 180, 181, 226, 248, 318, 354, 372, 403, 447, 460,}W^{70, 96, 287}L^{24, 33, 42, 54, 110, 118, 153, 173, 183, 186, 206, 209, 215, 223, 238, 246, 258, 261, 273, 285, 286, 298, 300, 303, 337, 338, 349, 353, 362, 382, 438, 444, 474}Y^{21, 23, 85, 106, 114, 147, 197, 327, 356, 364, 391, 399, 407}$	R↓, K↓, F↓, W↓, L↓, Y↓	
Cathepsin B	VP1 chain – RR^{27}VP2 chain – noneVP3 chain – noneVP4 chain – none	RR^{38}	RR↓	• *Generally the same as papain and Cathepsin-L with these additional specificities* • *Exceptions at the P2 position with respect to papain and Cathepsin-L are PX$_{aa}$- where Cathepsin B does not have specificity for*
	VP1 chain – RP^{190}, HP^{196}, NP^{104}, TP^{44}, AP^{94}, VP$^{57, 90, 142}$, LP$^{118, 160}$VP2 chain – KP^{176}, NP^{150}, GP^{46}, FP$^{144, 215}$, AP$^{186, 195, 205}$, VP$^{127, 161}$, LP^{84}, VP3 chain – DP$^{19, 24}$, EP^{132}, NP^{32}, TP^{136}, GP^{114}, AP^{127}, VP^{62}, IP^{160}, LP^{38,}VP4 chain – SP^{7}	KP^{138}, HP$^{375, 468}$, DP^{136}, EP^{6}, QP^{8}, TP$^{163, 340, 405}$, GP^{11}, FP^{88}, AP^{150}, VP^{65}, IP$^{169, 280, 368}$, LP$^{187, 274}$, MP^{160}	KP↓, RP↓, HP↓, DP↓, EP↓, NP↓, QP↓, SP↓, TP↓, GP↓, FP↓, WP↓, YP↓, AP↓, VP↓, IP↓, LP↓, MP↓	
	VP1 chain – GE^{6}, GG$^{20, 64}$, GA$^{93, 167}$, GL^{65}VP2 chain – GR^{77}, GH^{21}, GE^{213}, GN^{114}, GS^{93}, GG^{119}, GP^{46}, GW^{105}, GY^{36}, GA^{194}, GL^{51}VP3 chain – GK$^{193, 207}$, GR^{40},	GK$^{93, 145, 335}$, GR$^{189, 272, 347}$, GH$^{129, 312}$, GS$^{30, 211}$, GT^{57}, GG$^{29, 32, 56, 158}$, GP^{11}, GF$^{87, 180, 447}$,	GK↓, GR↓, GH↓, GE↓, GN↓, GQ↓, GS↓, GT↓, GG↓, GP↓ GF↓,	

(*Continued on next page*)

Table 1. (*Continued*)

Protease *Cysteine protease*	P1-polyprotein	GOR	Cleavage	Preferential Cleavage
	GT^{104}, GG^{13}, GP^{114}, GT^{104}, GG^{13}, GP^{114}, GY^{11}, GL$^{94, 151}$, GM^{130}VP4 chain – GN^{17}, GQ^{4}, GS$^{11, 20, 47, 51}$, GG^{46}, GL^{79}	GY^{197}, GA$^{17, 44, 171, 195. 455}$, GL$^{33, 261, 382, 438}$, GM^{159}	GW↓, GY↓, GA↓, GL↓, GM↓	
	VP1 chain – NQ^{47}, NT^{101}, NG$^{92, 132}$, NP^{104}, NY$^{18, 165}$, NL^{144}VP2 chain – NR^{167}, NQ^{115}, NT$^{48, 191}$, NG$^{20, 104, 118}$, NP^{150}, NM^{154} VP3 chain – NQ^{36}, NS^{153}, NT^{89}, NP^{32}, NL$^{44, 107}$VP4 chain – NE^{49}, NN$^{24, 65}$, NQ^{14}, NS^{33}, NT$^{18, 62}$, NY^{25}, NA^{42}	NK$^{302, 426}$, NN$^{117, 366}$, NT$^{72, 295, 321, 463}$, NW^{96}, NA$^{112, 241}$, NL^{118}	NK↓, NR↓, NH↓, NE↓, NN↓, NQ↓, NS↓, NT↓, NG↓, NP↓ NF↓, NW↓, NY↓, NA↓, NL↓, NM↓	
	VP1 chain – RH^{201}, RN^{139}, RQ^{28}, RT^{158}, RG^{146}, RP^{190}, RY^{136}, RA^{183}, RL^{115}, RM^{180}VP2 chain – RE^{136}, RN^{19}, RT^{152}, RF^{61}, RY^{168}VP3 chain – RN^{35}, RF^{41}, RY^{121}, RA^{219}, RL^{213}VP4 chain – none	RK$^{348, 414}$, RH^{219}, RQ^{4}, RS^{225}, RG^{128}, RA^{39}, R$^{110, 273, 353}$	RK↓, RH↓, RE↓, RN↓, RQ↓, RS↓, RT↓, RG↓, RP↓ RF↓, RW↓, RY↓, RA↓, RL↓, RM↓	
Legumain	VP1 chain – TN^{103}VP2 chain – AN^{202}, TN$^{153, 207}$VP3 chain – TN$^{43, 179}$, LN^{152}VP4 chain – TN^{61}	AN^{425}, TN$^{233, 322}$, LN$^{111, 301}$	AAN↓, AN↓, TN↓, LN↓	Based on optimal sequences for schistosome (C197, N197C) and human legumains, and cruzain [49]
Serine protease				
Trypsin (H^{57}, D^{102}, G^{193}, S^{195})	VP1 chain – K$^{41, 81, 96, 109, 154, 169, 181, 202, 204, 210}$, R$^{26, 27, 38, 67, 114, 124, 135, 138, 145, 157, 172, 179, 182, 189, 200}$ VP2 chain – K$^{2, 3, 63, 88, 134, 172, 175, 198, 217}$, R$^{13, 18, 54, 60, 77, 102, 135, 151, 167}$ VP3 chain – K$^{20, 28, 67, 84, 118, 134, 154, 193, 207}$, R$^{34, 40, 56, 72, 120, 212, 218}$ VP4 chain – K^{70}	K$^{53, 66, 67, 93, 100, 102, 120, 137, 145, 146, 212, 221, 247, 252, 255, 256, 296, 302, 310, 324, 335, 348., 357, 361, 390, 397, 412, 414, 416, 420, 426, 429, 452, 457}$, R$^{3, 37, 38, 45, 97, 103, 109, 127, 189, 218, 224, 272, 291, 347, 352, 413, 478}$	K↓, R↓	

Amino acids are designated by the single-letter code
Cleavage sites are designated by arrows (↓)
Numbers appearing after amino acids refer to the position of potential cleavage
Highly abundant cleavage sites are highlighted in yellow
Highly scarce cleavage sites are highlighted in turquoise
Cleavage sites occurring at the end of a protein sequence are highlighted in gray

Protease reverse transcription quantitative real-time PCR (RT-qPCR)

RT-qPCR was carried out to confirm the RNA-Seq data. Three sequences coding for cysteine protease-like proteins (RD21, *Nb*VPE-1a and *Nb*VPE-1b) were selected for confirmation. Significantly elevated gene expression was found in all experimental groups (LBA, P1 and GOR) relative to the control (UN), which was set at 1 (Fig. 7), in general accordance with the RNA-Seq data as shown by the line graphs of the 3 respective genes. There were statistically significant differences ($p < 0.05$) between the expression levels of the different genes assayed for (RD21, *Nb*VPE-1a, *Nb*VPE-1a). The highest fold increase due to agroinfiltration was also found for, *Nb*VPE-1b, comparable to the RNA-Seq result. The discrepancy between the RNAseq and RT-qPCR data for *Nb*VPE-1b can be accounted for by the reduced coverage within the RNA-seq datasets compared to the RT-qPCR data.

Discussion

The effect of proteases on recombinant protein stability in plant-based expression systems is a field of growing interest. Results in this study have clearly demonstrated that 3 selected recombinant proteins (P1-polyprotein, VP1 and GOR) are sensitive to protease action. Such action has also been recently associated with lower antigenicity when heterologous proteins were expressed in a plant host.[13,21] Sensitivity of VP1 against proteases was further confirmed when treated with a protease-containing tobacco extract and when *in silico* proteolytic site analysis was carried out. In particular, cathepsin H-like cysteine proteases might play a major role in VP1 degradation. Our study also provided evidence that proteins, like VP1, are less susceptible to protease degradation when part of a polyprotein.[24] Such greater VP1 stability and better processing when expressed in tobacco as a P1-polyprotein, has previously been found.[25] Less protease sensitivity was also identified in our protein

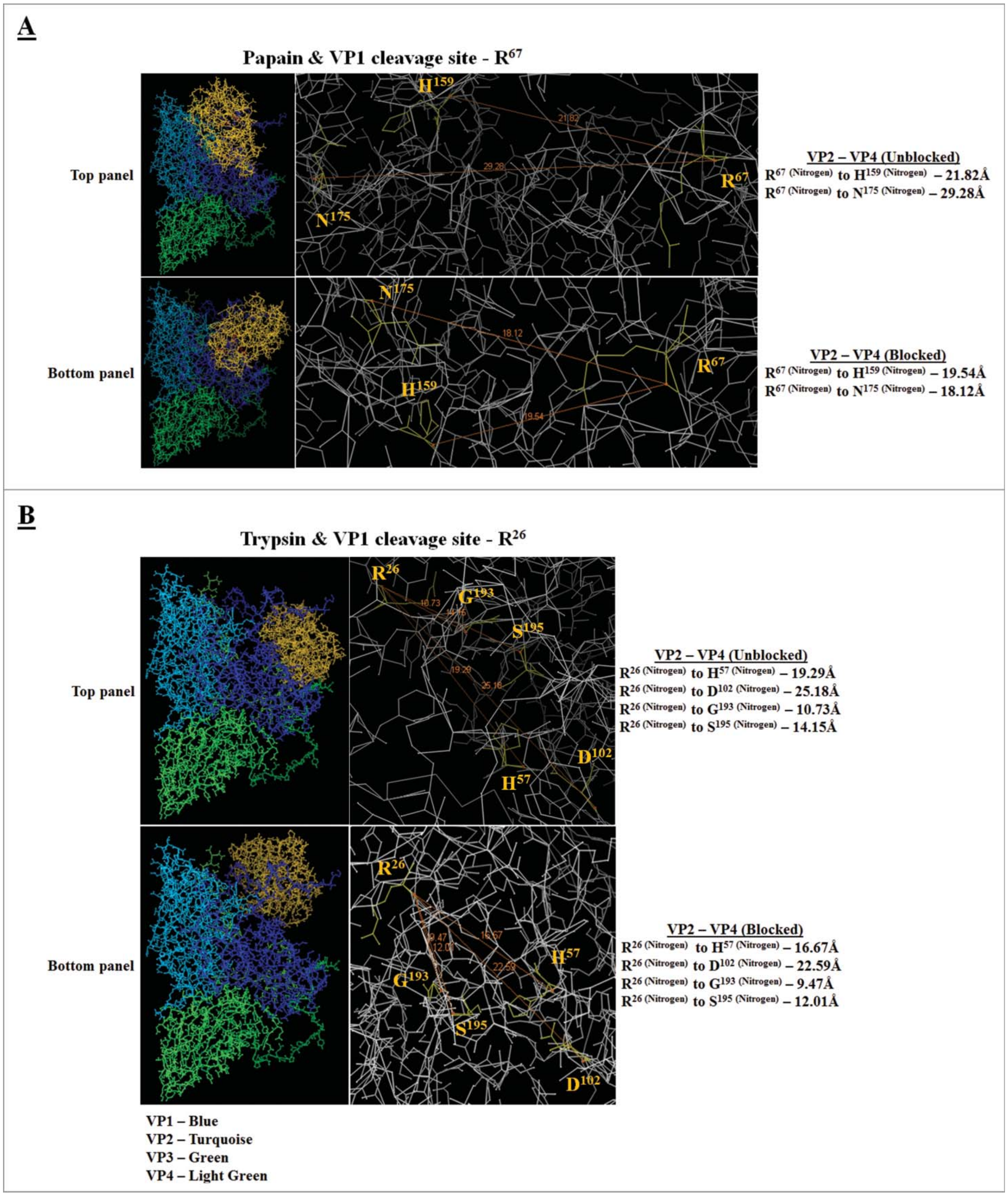


Figure 4. Protein docking model of single VP1 with papain and trypsin. (A) Docking model of VP1 (cleavage site - R^{67}) and papain. (B) Docking model of VP1 (cleavage site - R^{26}) and trypsin. Unblocked means all other binding sites (capsid proteins VP2, VP3 and VP4) within the P1-polyprotein were permitted in the interaction model. Blocked means all other binding sites (capsid proteins VP2, VP3 and VP4) within the P1-polyprotein were blocked in the interaction model. Distances between interacting residues are given in Ångstroms (Å).

docking experiments when the distance between interacting residues of VP1 and proteases increased as a result of being part of a polyprotein.

Besides establishing protease sensitivity of model recombinant proteins, sequencing of RNA (RNA-Seq) was applied in our study as a powerful technique to identify any possible proteases expressed during the agroinfiltration process which might compromise recombinant protein production. RNA-Seq was conceived about 10 years ago and has become a preferred technology for transcriptome and gene analysis with a number of advantages.[26,27] RNA-Seq is not dependent on sequence knowledge being available *a priori* and provides a direct measure of RNA abundance. The

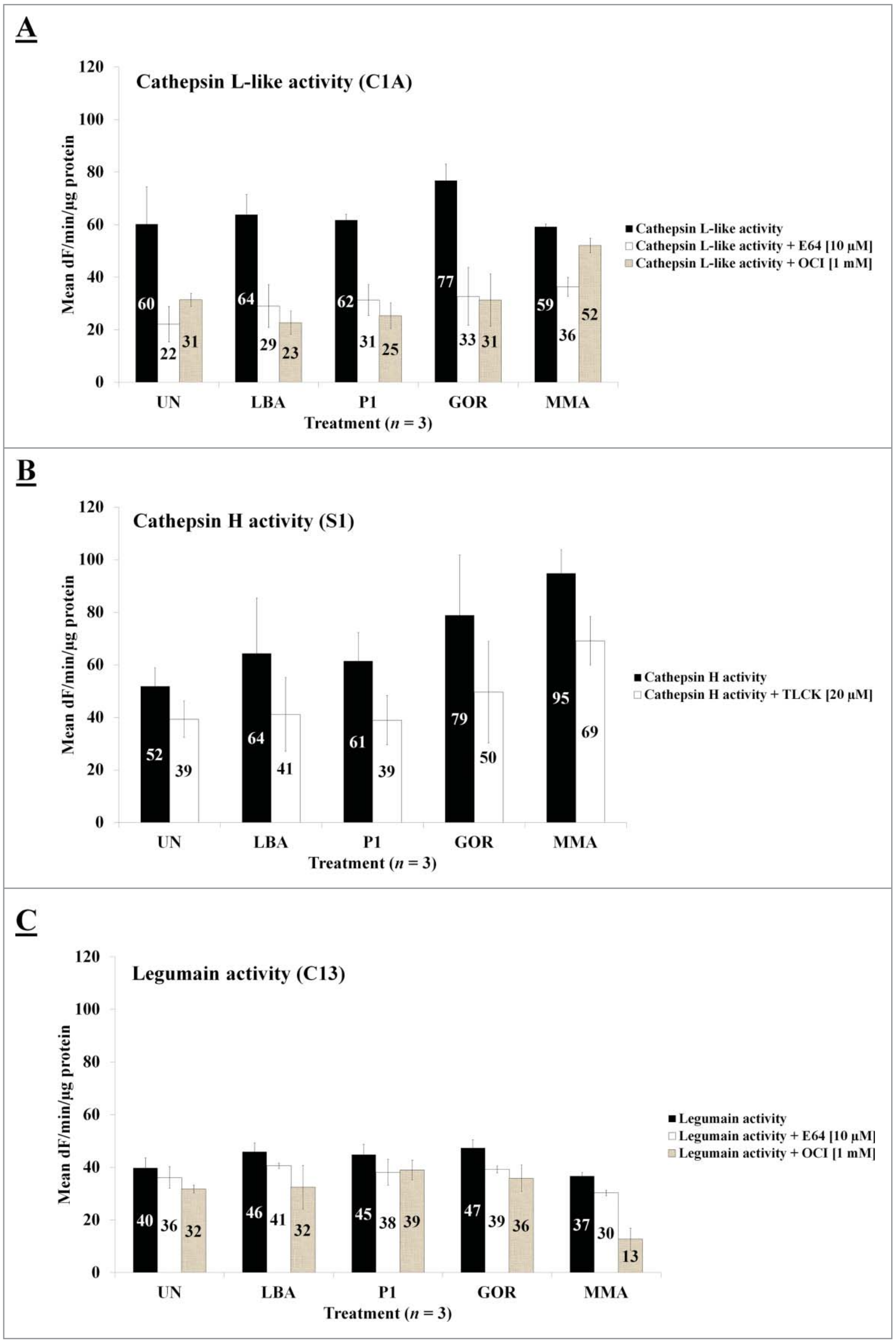


Figure 5. Protease activities of cathepsin L-like, cathepsin H-like and legumain-like in control (UN & MMA) and experimental groups (LBA, P1 and GOR). (A) Cathepsin L-like activity, (B) cathepsin-H like activity, (C) legumain-like activity. The y-axis represents the mean activities expressed as fluorescence units (dF) per min per μg protein. Mean activities of 3 biological replicates are shown within bars. Error bars indicate standard error of the mean (SEM).

technology, although powerful, still has the problem of introducing a certain degree of bias due to the sequencing of pooled samples. Rigorous post-sequencing bioinformatics data analysis, as done in our study, is therefore required as well as a final validation of determined gene expression via alternative methods such as RT-qPCR (reverse transcription quantitative real-time PCR). However, a problem faced in our study, when working with tobacco, was mismatches between gene identifiers and existing gene annotations rendering the annotation process an arduous and challenging task.

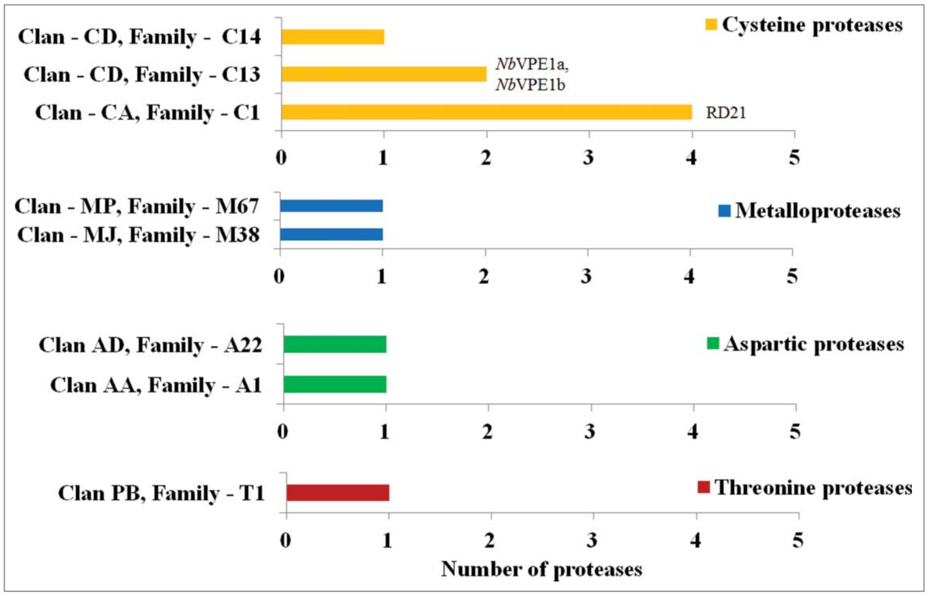


Figure 6. Proteolytic class distribution bearing the clans and families of the 5 major classes of proteases in *N. benthamiana* from the MEROPS database found in RNA-Seq datasets and number of protease(s) identified in various classes to be expressed at 2-times higher based on FPKM values after agroinfiltration. Genes further validated by RT-qPCR (RD21 and *Nb*VPE). Yellow – Cysteine proteases. Blue – Metalloproteases. Green – Aspartic proteases. Maroon – Threonine proteases.

By performing RNA-seq analyses, we were able to obtain a transcriptomic profile for proteases in operation during the early stages of agroinfiltration. Members of the C1 and C13 cysteine protease families were more expressed in agroinfiltrated leaves. This result supports our original hypothesis that proteases and, in particular, cysteine proteases are expressed during agroinfiltration. This is consistent with previous findings that C1 proteases are specifically up-regulated in response to pathogenic microbes.[19,28] Studies have also shown that C1 family RD21-like cysteine proteases, XM_009614860.1 in our study, are important components of plant immunity.[29,30] These proteases are also induced in senescing leaves in conjunction with the induction of vacuolar processing enzymes (VPEs).[31] Endogenous cysteine protease inhibitors might further be possible interaction partners of RD21-like proteases to prevent RD21 activity. Our previous finding of better stability of a recombinant protein (GOR) produced via agroinfiltration in tobacco leaves engineered with the rice cysteine proteases inhibitor OC-I supports the idea that RD21-like proteases might have been inhibited by OC-I expression in a transgenic tobacco leaf resulting in better GOR production.[23]

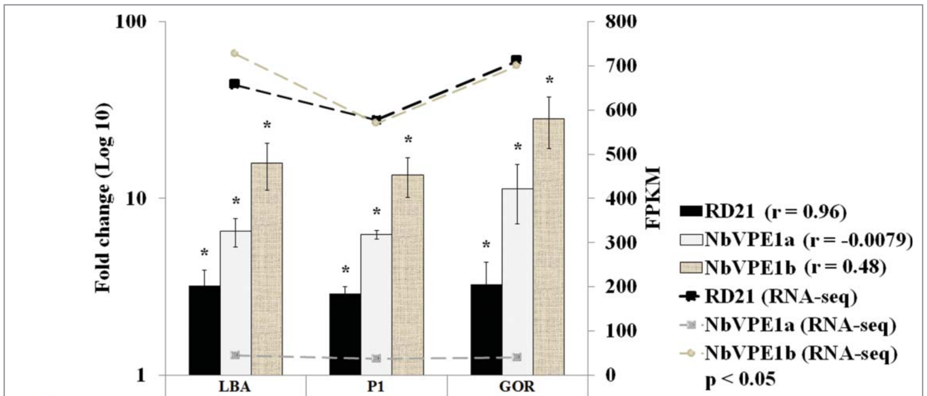
A further C1 cysteine protease (ALP; XM_009594290.1) with aleurain-like activity was found to be expressed 4-times more. Plant aleurain, first isolated from barley and localized in the vacuole, is an aminopeptidase with a number of similarities to animal cathepsin H. These similarities include heterogeneity of charge forms, position of the NH_2-terminus of the mature protein, and a similar pH-activity profile.[32] Several cleavage sites were found in this study for cathepsin H-like proteases in our model recombinant proteins possibly compromising their stability. *N. tabacum* contains cathepsin H-like proteases, such as *Nt*CP-23 and *Nt*CP1, with higher expression during natural senescence.[23,33,34]

The most prominent cysteine proteases induced in our study by agroinfiltration were 2 VPEs (*Nb*VPE-1a & *Nb*VPE-1b; AB181187.1 & AB181188.1). VPEs, so far, have not been extensively investigated in the context of degrading recombinant proteins. However, it is already known that prolonged incubation of anti-HIV antibodies 2F5 with high amounts of a VPE results in the formation of a 30-kDa degradation product implicating the involvement of VPEs in the degradation of this antibody.[21] Such degradation might also be relevant for VP1 degradation in our study. VPE cleavage sites were identified in *in silico* cleavage analyses of VP1 and P1-polyprotein as well as that of GOR. Vegetative-type VPEs are generally expressed during senescence or the pathogen-induced hypersensitive response (HR).[35] VPEs belong to the CD clan of cysteine proteases and within the clade they form part of the C13 family.[36] VPEs contribute to the senescence

Table 2. Identification, expression and function of genes due to agroinfiltration.

Clan†	Family‡	tracking_id	UN FPKM§	LBA	P1	GOR	Genbank Accession number	GenBank Description	TAIR locus name	Tair description
CA	C1	Niben101Scf01369Ctg011:1023-3226	2.2	31.6 (14.3)	17.5 (7.9)	18.1 (8.2)	XM_009798142.1	xylem cysteine proteinase 1-like	AT1G20850	XCP2, XYLEM CYSTEINE PEPTIDASE 2
		Niben101Scf04007Ctg010:1046-4801	140.4	655.6 (4.6)	577.1 (4.1)	710.5 (5.06)	XM_009614860.1	low-temperature-induced cysteine proteinase-like	AT1G47128	RD21, RD21A, RESPONSIVE TO DEHYDRATION 21, RESPONSIVE TO DEHYDRATION 21A
		Niben101Scf01701Ctg053:1304-2925	11.6	30.8 (2.6)	36.9 (3.1)	25.3 (2.2)	XM_009772991.1	zingipain-2	AT1G09850	XBCP3, XYLEM BARK CYSTEINE PEPTIDASE 3
		Niben101Scf03514Ctg004:2555-6476	11.7	45.8 (3.9)	45.8 (3.9)	49.6 (4.2)	XM_009594290.1	cysteine proteinase 3-like	AT5G60360	AALP, ALEURAIN-LIKE PROTEASE, ALP, SAG2, SENESCENCE ASSOCIATED GENE2
CD	C13	Niben101Scf04539Ctg027:143-4747	7.1	44.4 (6.3)	38.0 (5.3)	40.9 (5.7)	AB181187.1	NbVPE-1a mRNA for vacuolar processing enzyme 1a	AT4G32940	GAMMA VACUOLAR PROCESSING ENZYME, GAMMA-VPE, GAMMAVPE
		Niben101Scf04675Ctg067:1577-4581	7.6	726.7 (95.6)	568.9 (74.9)	699.1 (91.9)	AB181188.1	NbVPE-1b mRNA for vacuolar processing enzyme 1b	AT4G32940	GAMMA VACUOLAR PROCESSING ENZYME, GAMMA-VPE, GAMMAVPE
CD	C14	Niben101Scf06902Ctg013:5476-8453	0.6	10.3 (17.1)	8.1 (13.5)	17.8 (29.6)	XM_009795875.1	metacaspase-1	AT1G02170	ARABIDOPSIS THALIANA METACASPASE 1
MJ	M38	Niben101Scf09860Ctg01:1005-5383	1.9	18.5 (9.7)	21.3 (11.2)	24.0 (12.6)	XM_009796689.1	Dihydro-pyrimidinase	AT5G12200	PYD2, PYRIMIDINE 2
MP	M67	Niben101Scf07364Ctg013:20621-21147	20.9	110.7 (5.3)	126.6 (6.0)	116.0 (5.6)	XM_009779021.1	26S proteasome non-ATPase regulatory subunit 14 homolog	AT5G23540	MOV34/MPN/PAD-1 FAMILY PROTEIN
AA	A1	Niben101Scf08709Ctg013:6889-8887	1.7	4.8 (2.8)	2.9 (1.7)	2.6 (1.5)	XM_009764962.1	aspartic proteinase PCS1-like	AT5G02190	ASPARTIC PROTEASE 38, ATASP38
AD	A22	Niben101Scf08817Ctg011:344-1385	1.7	7.2 (4.2)	7.9 (4.6)	6.5 (3.8)	XM_009593783.1	signal peptide peptidase-like 4	AT1G01650	SIGNAL PEPTIDE PEPTIDASE-LIKE 4, ATSPPL4
PB	T1	Niben101Scf07066Ctg019:2266-2528	14.6	59.2 (4.0)	45.1 (3.1)	63.7 (4.4)	XM_009797260.1	proteasome subunit α type-7	AT5G66140	PAD2, PROTEASOME ALPHA SUBUNIT D2

†Clan PB contains endopeptidases and self-processing proteins
‡For families C represents cysteine, M metallo, A aspartic, and T threonine
§Increase in expression measured as fold increase of FPKM in brackets



	LBA	P1	GOR
RD21	3.21	2.89	3.27
NbVPE1a	6.51	6.23	11.37
NbVPE1b	15.78	13.54	28.23
RD21 (RNA-seq)	655.69	577.12	710.56
NbVPE1a (RNA-seq)	44.46	38.05	40.94
NbVPE1b (RNA-seq)	726.76	568.96	699.12

Treatment (n = 3)

Figure 7. Fold change and FPKM of gene expression of RD21 (black bars), *Nb*VPE1a (gray bars) and *Nb*VPE1a (canvas bars) in LBA, P1 and GOR experimental groups normalized to reference genes and relative to untreated control (UN) set at 1. Fold changes are represented on left y-axis as log 10 and samples are represented on x-axis. Error bars indicate standard error of mean (SEM) across 3 biological replicates. FPKM values of RD21 (black line), *Nb*VPE1a (gray line) and *Nb*VPE1b (canvas line) are represented on the right y-axis. 'r' refers to the Pearson correlation coefficient between RNA-seq and RT-qPCR gene expression measurements. Statistically significant differences between gene expression levels (RD21, *Nb*VPE1a and *Nb*VPE1a) in treatments (LBA, P1 and GOR) compared to control (UN) were determined by 2-factor ANOVA with replication (p-value$<$ 0.001) and Bonferroni corrected post-hoc t-tests (p-value $<$ 0.05) are represented as asterisks above graphs (*).

process and PCD (programmed cell death) by participating in the collapse of the vacuole membrane with the release of proteases into the cell.[37] VPEs are asparaginyl endopeptidases cleaving peptide bonds on the C-terminal side of asparagine (Asn) and aspartic (Asp) residues of pro-protein precursors to generate mature proteins.[38] They occur along the secretory pathway and are located in the vacuole, except for a single cell wall specific VPE.[15,39] Although VPEs have caspase-1 activity, they are structurally unrelated to caspases. Plants have evolved their own unique alternative regulated cellular suicide strategy that differs from animals with VPEs located in the vacuole as opposed to the cytosol.[40]

In our study, other possible protease candidates were also identified which might compromise recombinant protein production. This included the C1 cysteine proteases XM_009798142.1, a xylem cysteine peptidase 2 (XCP2) as well XM_009772991.1, a xylem bark cysteine peptidase 3 (XBCP3). Both proteases are involved in various cellular proteolytic processes. Their expression and induction was, however, not comparable to XM_009614860.1 (RD21) and any importance of these 2 proteases in recombinant protein degradation has yet to be determined. Also the expression of a metacaspase-1 (XM_009795875.1), involved in apoptosis, was identified in our study, as well as various aspartic, threonine and metallo-proteases. Serine and metallo-proteases have been already found in extracellular root and plant cell culture media. Lallemand et al. (2015) recently proposed that they are prime candidates in compromising protein stability. Since serine- or metallo-proteases were not strongly expressed in our study following agroinfiltration, their importance in compromising recombinant protein stability requires more detailed investigation.

Overall, our study has provided a more detailed insight of protease sensitivity of recombinant proteins, in particular, against cysteine proteases. Through, transcriptomic profiling and gene expression analysis, further evidence was provided that several cysteine proteases including proteases with RD21-like and aleurain-like activity as well as VPEs are upregulated during the early stages of agroinfiltration and might compromise recombinant protein production. However, the identified proteases still require more detailed analyses to confirm their direct involvement in recombinant protein degradation. Ultimately, characterization of identified proteases might also contribute toward establishing a protease library to be screened before any recombinant protein production is envisaged via agroinfiltration.

Methods

Plant material and growth conditions

N. benthamiana seeds were obtained from Dr. Ereck Chakauya (CSIR, Pretoria, South Africa) and were germinated in plastic trays in germination mix. Seedlings were grown at a 12/12 hours light/dark cycle with a day/night temperature of 26°C/20°C and 80% (v/v) relative humidity in a growth chamber (Sanyo, Bensenville, USA). Plants were grown for 12 weeks in order to obtain fully expanded leaves suitable for agroinfiltration.

VP1 expression and purification

Recombinant N-terminal His-tagged VP1 was expressed in *E. coli* M15 cells with bacteria grown in LB medium containing 100 μg mL^{-1} ampicillin, at 37°C up to a cell mass of 0.5 (OD_{600}).[41] Expression was induced with 1 mM IPTG for 5 h at 37°C and cells were harvested by centrifugation and stored at −20°C until use. All purification procedures were carried out as previously described with a purification column (Bio-Rad, CA, USA) and elution with imidazole.[42]

Protein extraction

Whole leaf proteins were extracted in ice-cold Arakawa buffer containing 10 mM L-cysteine, 200 mM Tris-HCl, pH 8.0, 100 mM NaCl, 400 mM Sucrose, 10 mM EDTA, 14 mM 2-Mercaptoethanol, 0.05% Tween-20.[43] Total soluble protein (TSP) amount was determined with a commercial protein determination kit and standardized across samples within experiments (Bio-Rad, Hercules, CA).

VP1 treatment

Varying amounts (0.5, 0.4, 0.3, and 0.2 μg) of either the cysteine protease, papain (Sigma-Aldrich, Germany) or the serine protease, trypsin (Sigma-Aldrich, Germany) were added to 44 μg of purified VP1 protein in 20 μL sodium phosphate buffer (pH 6.0) and incubated for 5 min at 37°C. Tobacco extracts (230 μg total protein) were added to 15 μg of purified VP1 protein and incubated first for 2 hr at 25°C and then for 2 hr at 37°C.

SDS-PAGE and immunoblotting

Protein samples were boiled at 95°C for 5 min in a 4x SDS-containing reducing sample buffer. VP1 stability against proteases was fractionated on a 15% SDS-PAGE gel under reducing conditions with a Mini-PROTEAN® Electrophoresis System (Bio-Rad, Hercules, CA).[44] Western-blotting was carried out with a 0.45 μm Nitrocellulose membrane (Bio-Rad, Hercules, CA) and a Mini-Trans-Blot electrophoretic transfer cell for protein transfer (Bio-Rad, Hercules, CA) at a constant current of 300 mA for 3 hr. For blotting, membranes were blocked overnight in 5% (w/v) skim milk solution in TBST buffer containing 100 mM Tris, 154 mM NaCl, 0.1% Tween-20, pH 7.5. Incubation with the primary antibody (1:7000 dilution of the His-antiserum raised in rabbit) was done overnight in a 5% (w/v) skim milk-TBST solution. Incubation with the secondary antibody (1:10000 dilution of goat anti-rabbit IgG conjugated with alkaline phosphatase; AbD Serotec, UK) was then done for 1 hr. Membranes were finally developed with the AP (Alkaline Phosphatase) Conjugate Substrate Kit as described in the manufacturers protocol (Bio-Rad, Hercules, CA).

Protease activity measurement

Total soluble protein (TSP) from foliar extracts (36 μg) were used for measuring cathepsin L-like protease activity in a 50 mM sodium phosphate buffer, pH 6.0 (Sigma-Aldrich, Germany) containing 10 mM L-cysteine (Sigma-Aldrich, Germany) and cathepsin H-like protease activity in a 50 mM Tris buffer (pH 6.0). Extracts were mixed and transferred into black, flat-bottom polysorp 96 well plates (Nunc, AEC Amersham) for measuring fluorescence. Before measuring fluorescence, 8 μM of the substrate Z-Phe-Arg-7-amido-4-methylcoumarin hydrochloride (Z-Phe-Arg-MCA, Sigma-Aldrich, Germany) for cathepsin L-like activity, or Arginine-7-amido-4-methylcoumarin hydrochloride (Arg-NMec HCl, Sigma-Aldrich, Germany) for cathepsin H-like activity, was added in a final volume of 100 μL. Activity was measured kinetically over a 10 min time period with 20 seconds (sec) of shaking before the first cycle. Fluorescence development was measured with a fluorometer (BMG FluoStar Galaxy, Germany) at 25°C with excitation and emission wavelengths of 360 nm and 450 nm, respectively. Legumain activity in TSP (36 μg) was measured in a legumain assay buffer, pH 5.8, containing 1 mM DTT, 39.5 mM citric acid, 121 mM Na_2HPO_4, 1 mM Na_2EDTA, 0.01% CHAPS (Sigma-Aldrich, Germany). Extracts were mixed and prepared as described above and protease activity was measured fluorometrically after addition of 1 mM of substrate Z-Ala-Ala-Asn-AMC (Z-AAN-AMC, Bachem, Germany) as described above for cathepsin L and H activity. A broad spectrum commercial inhibitor N-[N-(L-3-transcarboxyirane-2-carbonyl)-L-Leucyl]-agmatine (E-64, Sigma-Aldrich, Germany) or 10 μM. purified oryzacystatin-I (OCI) (Department of Plant Sciences, University of Pretoria, South Africa) was used to inhibit cathepsin L-like and legumain protease activity. A commercial inhibitor N-a-tosyl-l-lysine chloromethyl ketone hydrochloride (TLCK, Sigma-Aldrich, Germany) was used to inhibit cathepsin H-like protease activity at 20 μM.

In silico proteolytic cleavage analyses

Proteolytic cleavage assays were conducted *in silico* with CLC Main Workbench 6.6.1 (http://www.clcbio.com) based on various proteases substrate specificities and the Schechter and Berger (1967, 1968) subsite nomenclature.[45-49] Protein structures and amino acid sequences were obtained from the RCSB Protein Data Bank (PDB) (www.rcsb.org).[50] Structures obtained were: P1 – PDB ID: 1FOD,[51] GOR – PDB ID: 3GRS,[52] Papain – PDB ID: 9PAP[53] and Trypsin – PDB ID: 1UTN.[54] Substrate specificity profiling was determined using guidelines as previously described.[47] Protein modeling was conducted between VP1 and papain as well as trypsin, by applying ZDOCK (http://zdock.umassmed.edu/).[55,56] Both the unblocked and blocked settings were used on binding sites within the other P1-polyprotein chains (VP2, VP3 and VP4) when conducting the modeling. Blocking binding sites in the other chains were conducted to simulate an interaction with only VP1 and the respective proteases. Models were visualized in 3D-Mol Viewer (a component of Vector NTI 9.1.0, Invitrogen) and distances between interacting residues were measured using the measure distance tool. The setting structure was used as the color theme. VP2 – 4 were hidden from selections to highlight the interaction between VP1 and the protease. Distances were measured in the measure distance mode in Ångstroms (Å) between defined molecules in amino acids.

Agroinfiltration

Syringe agroinfiltration was used to infiltrate the tobacco leaf surface.[7] Cultures were maintained in lysogeny broth (LB) medium supplemented with 50 μg mL^{-1} kanamycin and 50 μg mL^{-1} rifampicin. For agroinfiltration, bacteria were grown to the stable phase at 28°C to an OD_{600} of 1 and collected by centrifugation at 4000*g*. Bacterial pellets were re-suspended in MMA medium (10 mM 2-[N-morpholino] ethanesulfonic acid] (MES) buffer, pH 5.6, containing 100 μM acetosyringone and 10 mM $MgCl_2$. For RNA-seq (RNA sequencing) analyses, the first fully-expanded leaves of the upper 4 individual leaves were infiltrated as previously described using a needle-less syringe with the vector pB+O1KP1-*HT* (gift from Prof George Lommonossoff; John Innes Center, Norwich, UK) containing the P1-polyprotein (P1) coding sequence (Fig. 1a) or with the pGK2 construct (Fig. 1b) containing the glutathione reductase (GOR) coding sequence.[7,19] Leaves were also agroinfiltrated with *A. tumefaciens* strain LBA4404 (LBA) alone. Uninfiltrated leaf material (UN) and infiltration with MMA medium were applied as negative controls to avoid confounding effects due to experimental conditions. Infiltrated plants were kept in a growth cabinet (Sanyo, Bensenville, USA) for 24 h. Three biological (plant) replicates were used for each treatment allowing for statistical treatment of data. Plants were kept in an environmentally controlled growth room and watered daily. Four leaves per plant were harvested after 24 h, as source material for subsequent protein and RNA extraction. Leaf samples were frozen immediately in liquid nitrogen and stored at −80°C until protein or RNA extraction was carried out.

RNA extraction

RNA extraction on leaf tissue samples was carried out with the Trizol method by employing macro-dissection.[57] Ribolock (Thermo Fisher Scientific, Waltham, MA USA) was added to the final volume of RNA in a ratio of 1:10 (v/v). On-column DNase (Thermo Fisher Scientific, Waltham, MA USA) digests were performed with the Qiagen RNeasy Mini Kit (Qiagen, Valencia, CA, USA). RNA samples were initially analyzed on a full-spectrum spectrophotometer Nanodrop® ND-1000 (Thermo Fisher Scientific, Waltham, MA USA) and subsequently sent for RNA quality analyses on the Experion™ Automated Electrophoresis System (Bio-Rad, CA, USA). After quantification, RNA samples from each sample were equivalently pooled for RNA-seq analyses. Biological replicates for each group were kept separate for RT-qPCR (reverse transcription quantitative real-time PCR) analyses.

RNA-seq and mapping

A transcriptomic library was constructed from paired end reads of ~90 bp in size which were generated from the HiSeq 2000 sequencing system (Illumina® sequencing, San Diego, USA) at the Beijing Genomics Institute (BGI Tech Solutions Co., Ltd, Hong Kong, China). Galaxy (Department of Bioinformatics, University of Pretoria), a platform for working with sequencing data, was applied to visualize, interpret, and conduct further analyses on the data generated.[58-60] As part of the filtering process, reads with adaptors, unknown nucleotides larger than 5% and with low quality (more than 20% of the bases' qualities are less than 10 in a read) were removed (BGI Tech Solutions Co., Ltd, Hong Kong, China). The FastQC tool was applied to perform QC checks on the data (Supplementary Fig. 1). The FastQ groomer was then used to convert the data into a format amenable for subsequent interpretation.[61] The draft assembly of the *N. benthamiana* genome was applied as a reference genome.[2,3,62] With TopHat2 RNA-Seq reads were aligned to the tobacco genome available from the Solanaceae Genomics Network (SGN) at url (ftp://ftp.solgenomics.net/genomes/Nicotiana_benthamiana/).[63,64] Tophat2 was carried out with a mean inner distance between mate pairs of 120 and a standard deviation of 30.

Transcript abundances and gene annotation

Cufflinks was applied performing bias correction and in default mode for all other parameters to assemble transcripts and estimate abundances.[65] High-confidence transcripts were obtained from identified transcripts (i.e., transcripts with FPKM value in the case of cufflinks > 0) by filtering for a FPKM 95 % confidence interval lower boundary greater than zero and FPKM value $\geq$ 0.001. With MEROPS, Sol Genomics Network (http://solgenomics.net/organism/Nicotiana_benthamiana/genome), and TAIR databases, proteases were mined from data sets.[2,3,62,66,67] Blast analyses against the draft genome of *N. benthamiana* were conducted with the BLASTN tool.[2,3,62] Tracking IDs were obtained for transcripts of interest and applying the Log10 of FPKM expression data values and transcript abundances were established.[68] Gene annotation was conducted using Blast2GO, TAIR as well as NCBI databases.[67,69,70]

RT-qPCR analyses

mRNA transcripts for 3 proteolytic candidate proteases (RD21, *Nb*VPE1a and *Nb*VPE1b) were assayed by RT-qPCR with a Bio-Rad CFX C1000™ Real-Time PCR Detection System (Bio-Rad, CA, USA). cDNA synthesis was done with the Promega GoScript™ Reverse Transcription System (Madison, Wisconsin, USA). RT-qPCR assays were carried out in accordance with the MIQE guidelines and optimized for annealing temperature and primer efficiency.[71] Reactions contained 10 μM forward and reverse primer and 2.5 μL of cDNA template. Supplementary Table 1 provides information of primer sets used. No-template mixture controls were included in each 96-well plate. Thermocycling parameters included initial denaturation at 95°C for 3 min; 39 cycles of denaturation at 95°C for 10 sec, annealing and extension at the abovementioned ranges for 30 sec, denaturation at 95°C for 10 sec. Melt curve analyses were performed thereafter from 65°C to 95°C with 0.5 increments. Fold changes in gene expression were determined with the Livak method.[72] For comparative purposes, relative gene expression was defined with the value of 1 in control plants.

Statistical analysis

Statistically significant changes in gene expression between control and experimental groups were determined using ANOVA: single/2-factor with replication (p-value < 0.05) applying Microsoft Excel software 2010 version 14 (Microsoft Corporation). Bonferroni corrected post-hoc t-tests (p-value < 0.05) were subsequently performed.

Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Professor Yong Suk Jang, Professor Moon Sik Yang and Dr. Tae Geum Kim for providing us with the VP1 gene and Dr. Huy for assisting with the VP1 purification, Professor George Lomonossoff for providing the P1 construct, and Professor Christine Foyer for providing the GOR construct. We also thank Dr. Francois Maree (Ondersterpoort Veterinary campus, University of Pretoria, South Africa) who kindly provided the Anti-FMDV polyclonal antiserum and Professor Dominique Michaud for assisting us to establish the agroinfiltration technique.

Funding

Our research was supported by the National Research Foundation (NRF) and the Genomics Research Institute (GRI), South Africa as well as NRF incentive funding to Karl Kunert and a NRF bursary to Priyen Pillay.

Notes on contributors

PP and KJK conceptualised and designed the experiment. PP conducted the experiments. BJV financially supported the project and provided analytical tools and scientific intellectual input in data interpretation. MEM helped in running enzymatic assays. CAC and SGVW helped in analyzing the RNA-seq data and manuscript reading. All authors read and approved the manuscript.

References

[1] Zhang Y, Li D, Jin X, Huang Z. Fighting Ebola with ZMapp: spotlight on plant-made antibody. Sci China Life Sci 2014; 57:987-8; PMID:25218825; http://dx.doi.org/10.1007/s11427-014-4746-7

[2] Bombarely A, Rosli HG, Vrebalov J, Moffett P, Mueller LA, Martin GB. A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research. Mol Plant Microbe In 2012; 25:1523-30; http://dx.doi.org/10.1094/MPMI-06-12-0148-TA

[3] Goodin MM, Zaitlin D, Naidu RA, Lommel SA. *Nicotiana benthamiana*: Its history and future as a model for plant-pathogen interactions. Mol Plant Microbe In 2008; 21:1015-26; http://dx.doi.org/10.1094/MPMI-21-8-1015

[4] Faino L, de Jonge R, Thomma BP. The transcriptome of Verticillium dahliae-infected *Nicotiana benthamiana* determined by deep RNA sequencing. Plant Signal Behav 2012; 7:1065; PMID:22899084; http://dx.doi.org/10.4161/psb.21014

[5] Abiri R, Valdiani A, Maziah M, Shaharuddin NA, Sahebi M, Yusof Zy, Atabaki N, Talei D. A critical review of the concept of transgenic plants: insights into pharmaceutical biotechnology and molecular farming. Curr Issues Mol Biol 2015; 18:21-42; PMID:25944541

[6] Wagner B, Fuchs H, Adhami F, Ma Y, Scheiner O, Breiteneder H. Plant virus expression systems for transient production of recombinant allergens in *Nicotiana benthamiana*. Methods 2004; 32:227-34; PMID:14962756; http://dx.doi.org/10.1016/j.ymeth.2003.08.005

[7] D'Aoust MA, Lavoie PO, Belles-Isles J, Bechtold N, Martel M, Vézina LP. Transient expression of antibodies in plants using syringe agroinfiltration. Recombinant Proteins From Plants 2009; 483:41-50; http://dx.doi.org/10.1007/978-1-59745-407-0_3

[8] Van der Hoorn RA, Laurent F, Roth R, De Wit PJ. Agroinfiltration is a versatile tool that facilitates comparative analyses of Avr9/Cf-9-Induced and Avr4/Cf-4-induced necrosis. Phytopathol 2000; 13:439-46

[9] Outchkourov NS, Rogelj B, Strukelj B, Jongsma MA. Expression of sea anemone equistatin in potato. Effects of plant proteases on heterologous protein production. Plant Physiol 2003; 133:379-90; PMID:12970503; http://dx.doi.org/10.1104/pp.102.017293

[10] Donini M, Lombardi R, Lonoce C, Di Carli M, Marusic C, Morea V, Di Micco P. Antibody proteolysis: a common picture emerging from plants. Bioengineered 2015; 6(5):299-302; PMID:26186119; http://dx.doi.org/10.1080/21655979.2015.1067740

[11] Miletic S, Simpson DJ, Szymanski CM, Deyholos MK, Menassa R. A plant-produced bacteriophage tailspike protein for the control of *Salmonella*. Front Plant Sci 2015; 6:1-9; PMID:25653664

[12] Faye L, Boulaflous A, Benchabane M, Gomord V, Michaud D. Protein modifications in the plant secretory pathway: Current status and practical implications in molecular pharming. Vaccine 2005; 23:1770-8; PMID:15734039; http://dx.doi.org/10.1016/j.vaccine.2004.11.003

[13] Castilho A, Windwarder M, Gattinger P, Mach L, Strasser R, Altmann F, Steinkellner H. Proteolytic and N-Glycan processing of Human α1-Antitrypsin expressed in *Nicotiana benthamiana*. Plant Physiol 2014; 166:1839-51; PMID:25355867; http://dx.doi.org/10.1104/pp.114.250720

[14] Delannoy M, Alves G, Vertommen D, Ma J, Boutry M, Navarre C. Identification of peptidases in *Nicotiana tabacum* leaf intercellular fluid. Proteomics 2008; 8:2285-98; PMID:18446799; http://dx.doi.org/10.1002/pmic.200700507

[15] Goulet C, Khalf M, Sainsbury F, D'Aoust MA, Michaud D. A protease activity-depleted environment for heterologous proteins migrating towards the leaf cell apoplast. Plant Biotechnol J 2012; 10:83-94; PMID:21895943; http://dx.doi.org/10.1111/j.1467-7652.2011.00643.x

[16] Benchabane M, Goulet C, Rivard D, Faye L, Gomord V, Michaud D. Preventing unintended proteolysis in plant protein biofactories. Plant Biotechnol J 2008; 6:633-48; PMID:18452504; http://dx.doi.org/10.1111/j.1467-7652.2008.00344.x

[17] Roberts IN, Caputo C, Criado MV, Funk C. Senescence-associated proteases in plants. Physiol Plant 2012; 145:130-39; PMID:22242903; http://dx.doi.org/10.1111/j.1399-3054.2012.01574.x

[18] Goulet C, Goulet C, Goulet MC, Michaud D. 2-DE proteome maps for the leaf apoplast of *Nicotiana benthamiana*. Proteomics 2010; 10:2536-44; PMID:20422621; http://dx.doi.org/10.1002/pmic.200900382

[19] Robert S, Khalf M, Goulet MC, D'Aoust MA, Sainsbury F, Michaud D. Protection of recombinant mammalian antibodies from development-dependent proteolysis in leaves of *Nicotiana benthamiana*. PloS One 2013; 8:1-9

[20] Lallemand J, Bouché F, Desiron C, Stautemas J, De Lemos Esteves F, Périlleux C, Tocquin P. Extracellular peptidase hunting for improvement of protein production in plant cells and roots. Front Plant Sci 2015; 6:1-10; PMID:25653664; http://dx.doi.org/10.3389/fpls.2015.00037

[21] Niemer M, Mehofer U, Torres Acosta JA, Verdianz M, Henkel T, Loos A, Strasser R, Maresch D, Rademacher T, Steinkellner H, et al. The human anti-HIV antibodies 2F5, 2G12, and PG9 differ in their susceptibility to proteolytic degradation: Down-regulation of endogenous serine and cysteine proteinase activities could improve antibody production in plant-based expression platforms. Biotechnol J 2014; 9:493-500; PMID:24478053; http://dx.doi.org/10.1002/biot.201300207

[22] Veena Jiang H, Doerge R, Gelvin SB. Transfer of T-DNA and Vir proteins to plant cells by *Agrobacterium tumefaciens* induces expression of host genes involved in mediating transformation and suppresses host defense gene expression. Plant J 2003; 35:219-36; PMID:12848827; http://dx.doi.org/10.1046/j.1365-313X.2003.01796.x

[23] Pillay P, Kibido T, Plessis M, Vyver C, Beyene G, Vorster BJ, Kunert KJ, Schlüter U. Use of transgenic Oryzacystatin-I-expressing plants enhances recombinant protein production. Appl Biochem Biotechnol 2012; 168:1608-20; PMID:22965305; http://dx.doi.org/10.1007/s12010-012-9882-6

[24] Butt TR, Edavettal SC, Hall JP, Mattern MR. SUMO fusion technology for difficult-to-express proteins. Protein Express Purif 2005; 43:1-9; http://dx.doi.org/10.1016/j.pep.2005.03.016

[25] Pan L, Zhang Y, Wang Y, Wang B, Wang W, Fang Y, Jiang S, Lv J, Wang W, Sun Y, et al. Foliar extracts from transgenic tomato plants expressing the structural polyprotein, P1-2A, and protease, 3C, from foot-and-mouth disease virus elicit a protective response in guinea pigs. Vet Immunol Immunop 2008; 121:83-90; http://dx.doi.org/10.1016/j.vetimm.2007.08.010

[26] Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat methods 2008; 5:621-8; PMID:18516045; http://dx.doi.org/10.1038/nmeth.1226

[27] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009; 10:57-63; PMID:19015660; http://dx.doi.org/10.1038/nrg2484

[28] Gilroy EM, Hein I, Van Der Hoorn R, Boevink PC, Venter E, McLellan H, Kaffarnik F, Hrubikova K, Shaw J, Holeva M, et al. Involvement of cathepsin B in the plant disease resistance hypersensitive response. Plant J 2007; 52:1-13; PMID:17697096; http://dx.doi.org/10.1111/j.1365-313X.2007.03226.x

[29] Gu C, Shabab M, Strasser R, Wolters PJ, Shindo T, Niemer M, Kaschani F, Mach L, van der Hoorn RA. Post-translational regulation and trafficking of the granulin-containing protease RD21 of *Arabidopsis thaliana*. PloS One 2012; 7:1-11

[30] Shindo T, Misas-Villamil JC, Hörger AC, Song J, van der Hoorn RA. A role in immunity for Arabidopsis cysteine protease RD21, the ortholog of the tomato immune protease C14. PloS One 2012; 7:1-9; http://dx.doi.org/10.1371/journal.pone.0029317

[31] Kinoshita T, Yamada K, Hiraiwa N, Kondo M, Nishimura M, Hara-Nishimura I. Vacuolar processing enzyme is up-regulated in the lytic vacuoles of vegetative tissues during senescence and under various stressed conditions. Plant J 1999; 19:43-53; PMID:10417725; http://dx.doi.org/10.1046/j.1365-313X.1999.00497.x

[32] Holwerda BC, Rogers JC. Purification and characterization of aleurain A plant thiol protease functionally homologous to mammalian cathepsin H. Plant Physiol 1992; 99:848-55; PMID:16669011; http://dx.doi.org/10.1104/pp.99.3.848

[33] Ueda T, Seo S, Ohashi Y, Hashimoto J. Circadian and senescence-enhanced expression of a tobacco cysteine protease gene. Plant Mol Biol 2000; 44:649-57; PMID:11198425; http://dx.doi.org/10.1023/A:1026546004942

[34] Beyene G, Foyer CH, Kunert KJ. Two new cysteine proteinases with specific expression patterns in mature and senescent tobacco (*Nicotiana tabacum* L.) leaves. J Exp Bot 2006; 57:1431-43; PMID:16551685; http://dx.doi.org/10.1093/jxb/erj123

[35] Yamada K, Shimada T, Nishimura M, Hara-Nishimura I. A VPE family supporting various vacuolar functions in plants. Physiol Plant 2005; 123:369-75; http://dx.doi.org/10.1111/j.1399-3054.2005.00464.x

[36] Mosolov V, Valueva T. Participation of proteolytic enzymes in the interaction of plants with phytopathogenic microorganisms. Biochemistry (Moscow) 2006; 71:838-45; http://dx.doi.org/10.1134/S0006297906080037

[37] Hara-Nishimura I, Hatsugai N, Nakaune S, Kuroyanagi M, Nishimura M. Vacuolar processing enzyme: an executor of plant cell death. Curr Opin Plant Biol 2005; 8:404-8; PMID:15939660; http://dx.doi.org/10.1016/j.pbi.2005.05.016

[38] Hara-Nishimura I, Hatsugai N, Nakaune S, Kuroyanagi M, Nishimura M. Vacuolar processing enzyme: an executor of plant cell death. Curr Opin Plant Biol 2005; 8:404-8; PMID:15939660; http://dx.doi.org/10.1016/j.pbi.2005.05.016

[39] Müntz K, Blattner FR, Shutov AD. Legumains-a family of asparagine-specific cysteine endopeptidases involved in propolypeptide processing and protein breakdown in plants. J Plant Physiol 2002; 159:1281-93; http://dx.doi.org/10.1078/0176-1617-00853

[40] Hatsugai N, Kuroyanagi M, Nishimura M, Hara-Nishimura I. A cellular suicide strategy of plants: vacuole-mediated cell death. Apoptosis 2006; 11:905-11; PMID:16547592; http://dx.doi.org/10.1007/s10495-006-6601-1

[41] Pillay P. Expression of the VP1 antigen from foot-and-mouth disease virus in a bacterial and plant-based expression system. Plant Sciences. South Africa: University of Pretoria, 2012:149

[42] Crowe JH. The QIA expressionist. Chatsworth, CA 1992

[43] Arakawa T, Chong DK, Merritt JL, Langridge WH. Expression of cholera toxin B subunit oligomers in transgenic potato plants. Transgenic Res 1997; 6:403-13; PMID: 9423288; http://dx.doi.org/10.1023/A:1018487401810

[44] Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 1970; 227:680-5; PMID:5432063; http://dx.doi.org/10.1038/227680a0

[45] Schechter I, Berger A. On the size of the active site in proteases. I. Papain. Biochem Bioph Res Co 1967; 27:157-62; http://dx.doi.org/10.1016/S0006-291X(67)80055-X

[46] Schechter I, Berger A. On the active site of proteases. III. Mapping the active site of papain; specific peptide inhibitors of papain. Biochem Bioph Res Co 1968; 32:898-902; http://dx.doi.org/10.1016/0006-291X(68)90326-4

[47] Choe Y, Leonetti F, Greenbaum DC, Lecaille F, Bogyo M, Brömme D, Ellman JA, Craik CS. Substrate profiling of cysteine proteases using a combinatorial peptide library identifies functionally unique specificities. J Biol Chem 2006; 281:12824-32; PMID:16520377; http://dx.doi.org/10.1074/jbc.M513331200

[48] Guncar G, Podobnik M, Pungercar J, Strukelj B, Turk V, Turk D. Crystal structure of porcine cathepsin H determined at 2.1 Å resolution: location of the mini-chain C-terminal carboxyl group defines cathepsin H aminopeptidase function. Structure 1998; 6:51-61; PMID:9493267; http://dx.doi.org/10.1016/S0969-2126(98)00007-0

[49] Mathieu MA, Bogyo M, Caffrey CR, Choe Y, Lee J, Chapman H, Sajid M, Craik CS, McKerrow JH. Substrate specificity of schistosome versus human legumain determined by P1-P3 peptide libraries. Mol Biochem Parasitol 2002; 121:99-105; PMID:11985866; http://dx.doi.org/10.1016/S0166-6851(02)00026-9

[50] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat T, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res 2000; 28:235-42; PMID:10592235; http://dx.doi.org/10.1093/nar/28.1.235

[51] Logan D, Abu-Ghazaleh R, Blakemore W, Curry S, Jackson T, King A, Lea S, Lewis R, Newman J, Parry N, et al. Structure of a major immunogenic site on foot-and-mouth disease virus. Nature 1993; 362:566-8; PMID:8385272; http://dx.doi.org/10.1038/362566a0

[52] Karplus PA, Schulz GE. Refined structure of glutathione reductase at 1.54 Å resolution. J Mol Biol 1987; 195:701-29; PMID:3656429; http://dx.doi.org/10.1016/0022-2836(87)90191-4

[53] Kamphuis IG, Kalk K, Swarte M, Drenth J. Structure of papain refined at 1.65 Å resolution. J Mol Biol 1984; 179:233-56; PMID:6502713; http://dx.doi.org/10.1016/0022-2836(84)90467-4

[54] Leiros HKS, Brandsdal BO, Andersen OA, Os V, Leiros I, Helland R, Otlewski J, Willassen NP, Smalås AO. Trypsin specificity as elucidated by LIE calculations, X-ray structures, and association constant measurements. Protein Sci 2004; 13:1056-70; PMID:15044735; http://dx.doi.org/10.1110/ps.03498604

[55] Pierce BG, Wiehe K, Hwang H, Kim B-H, Vreven T, Weng Z. ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. Bioinformatics 2014; 30:1771-3; PMID:24532726; http://dx.doi.org/10.1093/bioinformatics/btu097

[56] Chen R, Li L, Weng Z. ZDOCK: An initial-stage protein-docking algorithm. Proteins: Struct Funct Bioinf 2003; 52:80-7; http://dx.doi.org/10.1002/prot.10389

[57] MacRae E. Extraction of plant RNA. Methods Mol Biol. Protocols for Nucleic Acid Analysis by Nonradioactive Probes: Springer 2007; 353:15-24

[58] Goecks J, Nekrutenko A, Taylor J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol 2010; 11:1-13; http://dx.doi.org/10.1186/gb-2010-11-8-r86

[59] Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, et al. Galaxy: a platform for interactive large-scale genome analysis. Genome Res 2005; 15:1451-5; PMID:16169926; http://dx.doi.org/10.1101/gr.4086505

[60] Blankenberg D, Kuster GV, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J. Galaxy: a web-based genome analysis tool for experimentalists. Curr Protoc Mol Biol. 2010 Jan; Chapter 19:Unit 19.10.1-21. http://dx.doi.org/10.1002/0471142727.mb1910s89

[61] Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, Nekrutenko A. Manipulation of FASTQ data with Galaxy. Bioinformatics 2010; 26:1783-5; PMID:20562416; http://dx.doi.org/10.1093/bioinformatics/btq281

[62] Knapp S, Chase MW, Clarkson JJ. Nomenclatural changes and a new sectional classification in Nicotiana (Solanaceae). Taxon 2004; 53:73-82; http://dx.doi.org/10.2307/4135490

[63] Bombarely A, Menda N, Tecle IY, Buels RM, Strickler S, Fischer-York T, Pujar A, Leto J, Gosselin J, Mueller LA. The Sol Genomics Network (solgenomics. net): growing tomatoes using Perl. Nucleic Acids Res 2011; 39:D1149-D55; PMID:20935049; http://dx.doi.org/10.1093/nar/gkq866

[64] Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 2013; 14:1-13; http://dx.doi.org/10.1186/gb-2013-14-1-r1

[65] Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnol 2010; 28:511-5; http://dx.doi.org/10.1038/nbt.1621

[66] Rawlings ND, Waller M, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res 2013; 42:D503-D9; PMID:24157837; http://dx.doi.org/10.1093/nar/gkt953

[67] Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, et al. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res 2012; 40:D1202-D10; PMID:22140109; http://dx.doi.org/10.1093/nar/gkr1090

[68] Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 1998; 95:14863-8; PMID:9843981; http://dx.doi.org/10.1073/pnas.95.25.14863

[69] Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 2005; 21:3674-6; PMID:16081474; http://dx.doi.org/10.1093/bioinformatics/bti610

[70] Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997; 25:3389-402; PMID:9254694; http://dx.doi.org/10.1093/nar/25.17.3389

[71] Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem 2009; 55:611-22; PMID:19246619; http://dx.doi.org/10.1373/clinchem.2008.112797

[72] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. Methods 2001; 25:402-8; PMID: 11846609; http://dx.doi.org/10.1006/meth.2001.1262