Proteases of *Nicotiana benthamiana*: an emerging battle for molecular farming

Philippe V Jutras, Isobel Dodds and Renier AL van der Hoorn

Molecular farming increasingly uses the tobacco relative *Nicotiana benthamiana* for production of recombinant proteins through transient expression. Several proteins are produced efficiently with this expression platform, but yields for other proteins are often very low. These low yields are frequently due to endogenous proteases. The latest genome annotations indicate that *N. benthamiana* encodes for at least 1243 putative proteases that probably act redundantly and consecutively on substrates in different subcellular compartments. Here, we discuss the *N. benthamiana* protease repertoire that may affect recombinant protein production and recent advances in protease depletion strategies to increase recombinant protein production in *N. benthamiana*.

Address
The Plant Chemetics Laboratory, Department of Plant Sciences, University of Oxford, OX1 3RB Oxford, UK

Corresponding author: van der Hoorn, Renier AL (renier. vanderhoorn@plants.ox.ac.uk)

Proteases are ubiquitous in all organisms and fundamental for life. Proteases remove denatured and inactivated proteins and release amino acids for recycling but they also cleave proteins to regulate their activity and subcellular localisation [1,2]. In plants, many cellular functions require proteolytic enzymes, including seed germination, growth, development, and defence [3,4]. Plant genomes encode for hundreds of proteases that are tightly regulated and implicated in different responses to environmental or developmental stimuli, including senescence [5].

Recombinant plant-expressed proteins are frequently targeted by plant proteases, resulting in the partial or complete hydrolysis of proteins. The purified product is, consequently, a mixture of full-length proteins and degradation fragments and these mixtures compromise the commercial value of the product [6•].

Over the past decades, different strategies have been taken to reduce the negative impact of plant proteases on recombinant protein production *in planta* [7]. Here, we review the protease repertoire of *Nicotiana benthamiana* and the most recent strategies used to deplete these protease activities in *N. benthamiana*.
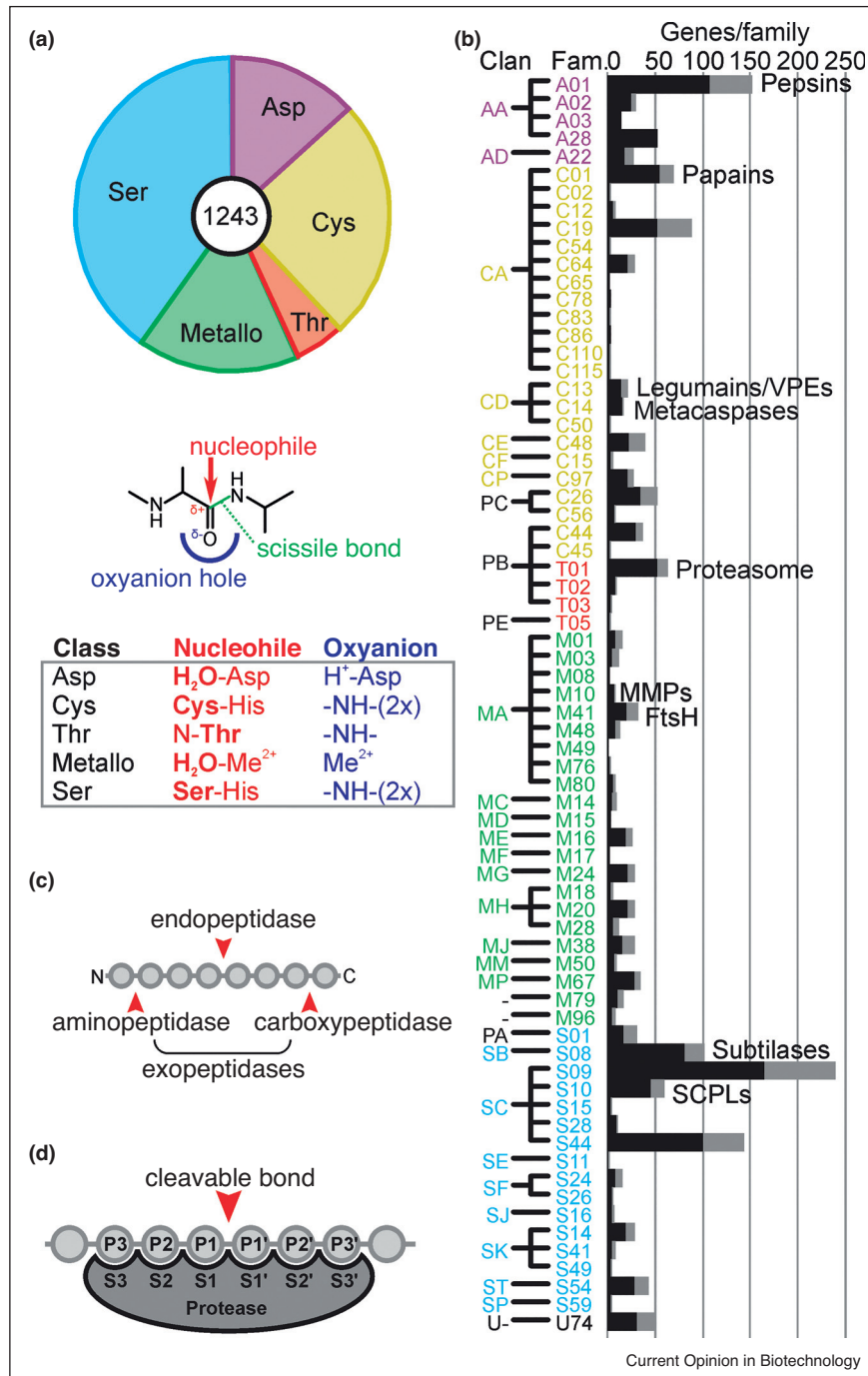
The proteases of *N. benthamiana*

N. benthamiana is an Australian relative of tobacco (*Nicotiana tabacum*) that has been embraced by the plant science community as a model plant for over two decades for its ease of manipulation by transient expression and RNA interference. This plant is favourite for agroinfiltration because its large leaves can be infiltrated easily and responses to *Agrobacterium tumefaciens* are relatively weak, whereas its RNAi system is hampered, supporting high transcript levels of transgenes [8•]. *N. benthamiana* is also easy to transform and manipulate by genome editing and virus-induced gene silencing (VIGS) [9,10]. A complication, however, is the complex genome of *N. benthamiana* because it is an ancient alloploid with a double gene set [11]. We have recently improved the annotation of the *N. benthamiana* genome [12••] and here we used this annotation to classify the putative proteases of *N. benthamiana* using PFAM [13] and the MEROPS classification [14].

The core proteome (NbD dataset) of *N. benthamiana* contains 1243 putative proteases, with an additional 512 putative proteases in the supplemental dataset (NbE dataset). These supplemental proteases are >70% identical to proteins of the core proteome and will include homeologs, allelic variants and sequencing errors. The core putative proteases include 165 aspartic (Asp) proteases, 307 cysteine (Cys) proteases, 66 threonine (Thr proteases), 207 metalloproteases and 498 serine (Ser) proteases (Figure 1a). This grouping into different catalytic classes is based on the catalytic mechanism of these enzymes. For instance, Cys, Thr and Ser proteases use a catalytic Cys, Thr or Ser residue to attack the peptide bond, respectively, whereas Asp and metalloproteases use Asp residues or a metal ion to activate a water molecule to perform the nucleophilic attack (Figure 1a).

Following the MEROPS principle [14], the proteases are further subdivided into families that share sufficient sequence homology to the type member of that family. Different families are grouped together in a clan if there is

Figure 1

Protease nomenclature and putative proteases of *Nicotiana benthamiana*.

(a) Classification of 1243 putative proteases of *N. benthamiana* into the five main catalytic classes, explained mechanistically below the pie-graph. **(b)** Further grouping of *N. benthamiana* into families and clans, following the MEROPS principles. The number of genes per family is shown for the core proteome (NbD, black) and supplemental proteome (NbE, grey). The latest proteome annotation of *N. benthamiana* [12••] was searched for PFAM domains that define the different protease families. Several relevant protease families are highlighted. **(c)** Nomenclature of endo/exo and amino/carboxy peptidases. **(d)** Nomenclature of P-sites and S-sites relative to the cleavable bond.

evidence that they are evolutionary related, for example, because they share the same fold or carry similar sequence motifs. Like most angiosperms, *N. benthamiana* has representatives of proteases of 70 families that group into 29 different clans (Figure 1b). The S8 subtilases, S9 prolyl oligopeptidases and A1 pepsins comprise the largest families of putative proteases of *N. benthamiana* (Figure 1b).

Proteases are also often classified into endopeptidases and exopeptidases (Figure 1c) but both versions can exist within the same protease family and so this annotation requires experimentation. A classification based on cleavage site specificity is not possible because cleavage sites are notoriously difficult to predict. Because proteins are folded, proteases do not act like restriction enzymes cleaving DNA. Proteases rather attack unstructured regions, often loops between structured regions in proteins and select cleavage sites using substrate binding pockets (S-pockets) that recognise residues before and after the cleavage site (residues P and P', respectively, Figure 1d). However, not every substrate residue flanking the cleavage site is recognised by every protease family. C1A papain-like proteases, for instance, select for residues at the P2 position and do not interact much with P1 residues, whereas S8 subtilase-like proteases often select for specific residues at the P1 position. In addition, the substrate binding pockets are often promiscuous binding sites, making substrate prediction by motif searches notoriously challenging.

Not all proteases are thought to affect recombinant protein degradation. Organelle-specific proteases, for instance, are unlikely to affect degradation of secreted recombinant proteins. Also, many proteases are not expressed in leaves, or not active at molecular farming conditions. The proteases that seem to affect the accumulation of recombinant proteins the most are papain-like Cys proteases (PLCPs, family C1A), subtilisins (SBTs, family S8), and pepsin-like Asp proteases (family A1). These proteases are abundant in leaves, can have a broad substrate specificity, and accumulate in subcellular compartments where glycosylated recombinant proteins reside [15•]. Different strategies have been taken to deplete these and other proteases from *N. benthamiana*, with varying success. These strategies are discussed in the following sections.

Protease depletion with protease inhibitors

Several studies have shown that co-expression of protease inhibitors increases the yield of recombinant proteins (Table 1). Protease inhibitors can have a relatively broad activity spectrum and can inhibit populations of functionally related proteases in plant tissues [16,17]. For instance, PLCPs are inhibited by cystatins (I12 family), which are protease inhibitors harbouring a conserved Gln-Xaa-Val-Xaa-Gly (QxVxG) motif [18]. The tomato cystatin *Sl*CYS8 was used to improve the yield of fully assembled and biologically active fragments of IgG antibodies transiently expressed in *N. benthamiana* [19–21]. An inactive version of *Sl*CYS8 showed no protective effect on recombinant proteins, indicating that the stabilising effect is accomplished through protease inhibition [20,22]. A chimeric version of *Sl*CYS8, the 'Cysta-tag', has also been designed to combine its inhibition potential with routine protein purification techniques [23]. The Cysta-tag provides a convenient way to efficiently and cost-effectively purify recombinant proteins from plants.

Other classes of protease inhibitors targeting Ser proteases and metalloproteases also increase the accumulation of recombinant proteins. Recently, three protease inhibitors of these classes significantly increased the accumulation of three unrelated recombinant proteins: α-galactosidase (a glycoenzyme), erythropoietin (a glycohormone) and VRC01 (an IgG antibody) [21]. *N. benthamiana* NbPR4, NbPot1 and human HsTIMP are thought to inhibit Cys, Ser and metalloproteases, respectively [21]. However, in contrast to *Sl*CYS8, NbPR4, NbPot1 and HsTIMP do not affect activity profiles of Ser proteases or PLCPs, indicating that perhaps another, yet uncharacterised mechanism may explain how unrelated protease inhibitors can improve recombinant protein accumulation.

Protease depletion by changing pH

The hydrolytic activity of broad-spectrum Cys, Ser and Asp proteases is generally influenced by the pH. For instance, Vacuolar Processing Enzymes (VPEs) have a

---

Table 1

Protease inhibitors used for molecular farming in *N. benthamiana* and other Solanaceae

Family	Name	Origin	Reference
Cys proteases	*Sl*CYS8	*Solanum lycopersicum*	[20]
	Oryzacystatin I	*Glycine max*	[24]
	NbPR4	*Nicotiana benthamiana*	[21]
Ser proteases	Bowman-Birk Serine protease inhibitor	*Glycine max*	[25]
	Protease Inhibitor II	*Nicotiana alata*	[26]
	NbPot1	*Nicotiana benthamiana*	[21]
Ser / Asp proteases	Cathepsin D inhibitor	*Solanum lycopersicum*	[27]
Metalloproteases	HsTIMP	*Homo sapiens*	[21]

unique pH-dependent two-chain state which regulates their protease and ligase activity [28]. Likewise, PLCPs are produced as inactive pro-proteases that often auto-catalytically mature in low-pH environments [29].

Regulating pH in the plant secretory pathway has recently been used to reduce the proteolytic degradation of recombinant proteins in plants. Partial neutralisation of the Golgi lumen pH by ectopic expression of Influenza virus M2 proton channel stabilises acid-labile recombinant proteins and peptides in leaf cells [30]. The M2 protein forms tetrameric pH-activated transmembrane channels and increases pH in the Golgi lumen of infected mammalian and plant cells [30,31]. Transient co-expression assays with fusion protein hybrids showed a significant alteration of host protease activities upon M2 channel expression [32]. However, M2 also alters the steady-state levels of proteins in different cellular compartments and attenuates the plant defence response upon agroinfiltration [33].

Protease depletion by gene knockdown/out
Several proteases have been depleted by RNA interference approaches to improve recombinant protein production in *N. benthamiana*. For instance, Mandal *et al.* depleted the most abundant aspartic, cysteine and metallo proteases simultaneously in tobacco BY-2 cell cultures using multitarget antisense silencing to obtain a cell culture that produces higher levels of antibody 2F5 [34]. Similarly, Duwadi *et al.* generated antisense tobacco plants for ten different Cys proteases and found that silencing of Cys6 could increase levels of interleukin IL-10 expression [35].

Genome editing of higher plants has significantly improved over the past years. Sequence-specific nuclease systems, such as TALEN and CRISPR/Cas, can target multiple genes and precisely modify the plant cell environment [36]. To date, this technology has been applied mainly to suppress the production of *N. benthamiana*-specific glycans, which are often undesirable in molecular farming [37]. For instance, CRISPR/Cas-mediated knockout of six glycosyltransferase genes in *N. benthamiana* recently allowed the production of a glyco-engineered antibody lacking plant *N*-linked glycans [36]. The same strategy could be used to target protease genes and to generate plants with protease depleted environments.

Future perspectives
Protease characterisation remains challenging as their roles in plants, including their subcellular localisation and target proteins, are mostly unknown. A major limiting factor to our understanding of protease roles is the lack of identified relevant substrates [38•]. Combinations of different experimental strategies are necessary to reveal the physiological substrates and hence, the molecular functions of plant proteases. Quantitative mass spectrometry-based proteomics enables large-scale interrogation of plant proteomes and allows the identification of protease cleavage sites and determination of protease sequence specificity [39,40]. Biochemical profiling of active sites using proteome-derived peptide libraries in combination with quantitative proteomics is useful to simultaneously identify N-terminal and C-terminal cleavage motifs [41]. Activity-based protein profiling (ABPP) is also increasingly used to uncover the active proteome using tagged chemical probes that react covalently and irreversibly with the active site of proteins [22,42]. These collective efforts to identify substrates and decipher protease functions will create new opportunities for plant biotechnology applications.

Plant molecular farming has proven its potential to express recombinant proteins at a high level, and many complex proteins are now produced in plants. Plant proteases are key players in recombinant protein degradation, limiting the development of plant-based expression systems. Future research addressing the problem of unwanted proteolysis will undoubtedly make a high impact on the commercialisation of relevant pharmaceutical and non-pharmaceutical products expressed in plants.

Conflict of interest statement
Nothing declared.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- •• of outstanding interest

1. Pesquet E: **Plant proteases - from detection to function**. *Physiol Plant* 2012, **145**:1-4.

2. van der Hoorn RAL: **Plant proteases: from phenotypes to molecular mechanisms**. *Annu Rev Plant Biol* 2008, **59**:191-223.

3. Shen W, Yao X, Ye T, Ma S, Liu X, Yin X, Wu Y: **Arabidopsis aspartic protease ASPG1 affects seed dormancy, seed longevity and seed germination**. *Plant Cell Physiol* 2018, **59**:1415-1431.

4. Thomas EL, van der Hoorn RAL: **Ten prominent host proteases in plant-pathogen interactions**. *Int J Mol Sci* 2018, **19**:1-12.

5. Roberts IN, Caputo C, Criado MV, Funk C: **Senescence-associated proteases in plants**. *Physiol Plant* 2012, **145**:130-139.

6. Schillberg S, Raven N, Spiegel H, Rasche S, Buntru M: **Critical**
• **analysis of the commercial potential of plants for the production of recombinant proteins**. *Front Plant Sci* 2019, **10**:1-10
An excellent review addressing the challenges facing recombinant protein production in plants. It highlights the bottlenecks that must be overcome before plants can compete with conventional systems.

7. Mandal MK, Ahvari H, Schillberg S, Schiermeyer A: **Tackling unwanted proteolysis in plant production hosts used for molecular farming**. *Front Plant Sci* 2016, **7**:1-6.

8. Bally J, Jung H, Mortimer C, Naim F, Philips JG, Hellens R,
• Bombarely A, Goodin MM, Waterhouse PM: **The rise and rise of *Nicotiana benthamiana*: a plant for all reasons**. *Annu Rev Phytopathol* 2018, **56**:405-426
This review gives a detailed overview of *Nicotiana benthamiana* as an outstanding model plant in molecular research and biotechnology. It describes the origin of the plant and highlights features that have strongly contributed to the rise in popularity of *N. benthamiana*.

9. Nekrasov V, Staskawicz B, Weigel D, Jones JDG, Kamoun S: **Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease**. *Nat Biotechnol* 2013, **31**:691-693.

10. Hayward A, Padmanabhan M, Dinesh-Kumar SP: *Virus-Induced Gene Silencing in* **Nicotiana benthamiana** *and Other Plant Species*. Totowa, NJ: Springer. Humana Press; 2011, 55-63.

11. 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 Interact* 2012, **25**:1523-1530.

12. Kourelis J, Kaschani F, Grosse-Holz FM, Homma F, Kaiser M, van
•• der Hoorn RAL: **Homology-guided re-annotation improves the gene models of the alloploid *Nicotiana benthamiana***. *BMC Genomics* 2019, **20**:722
This paper describes the re-annotation of the gene models of the alloploid *Nicotiana benthamiana*. This annotation greatly facilitates phylogenetic analysis of gene families and the datasets provide the research community with improved capacity to annotate spectra during proteomics experiments.

13. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A *et al.*: **The Pfam protein families database in 2019**. *Nucleic Acids Res* 2019, **47**: D427-D432.

14. Rawlings ND, Barrett AJ, Finn R: **Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors**. *Nucleic Acids Res* 2016, **44**:D343-D350.

15. Grosse-Holz F, Kelly S, Blaskowski S, Kaschani F, Kaiser M, van
• der Hoorn RAL: **The transcriptome, extracellular proteome and active secretome of agroinfiltrated *Nicotiana benthamiana* uncover a large, diverse protease repertoire**. *Plant Biotechnol J* 2018, **16**:1068-1084
The authors characterise the protease repertoire of agroinfiltrated *Nicotiana benthamiana* leaves. To identify the proteases, the authors study the transcriptome, extracellular proteome and active secretome of *N. benthamiana* and highlight abundant extracellular proteases in leaves upon agroinfiltration.

16. Grosse-Holz FM, van der Hoorn RAL: **Juggling jobs: roles and mechanisms of multifunctional protease inhibitors in plants**. *New Phytol* 2016, **210**:794-807.

17. Clemente M, Corigliano MG, Pariani SA, Sánchez-López EF, Sander VA, Ramos-Duarte VA: **Plant serine protease inhibitors: biotechnology application in agriculture and molecular farming**. *Int J Mol Sci* 2019, **20**:1-21.

18. Tremblay J, Goulet MC, Michaud D: **Recombinant cystatins in plants**. *Biochimie* 2019, **166**:184-193 http://dx.doi.org/10.1016/j.biochi.2019.06.006.

19. Robert S, Jutras PV, Khalf M, D'Aoust M-A, Goulet M-C, Sainsbury F, Michaud D: **Companion protease inhibitors for the *in situ* protection of recombinant proteins in plants**. *Methods Mol Biol* 2016, **1385**:115-126.

20. Jutras PV, Marusic C, Lonoce C, Deflers C, Goulet MC, Benvenuto E, Michaud D, Donini M: **An accessory protease inhibitor to increase the yield and quality of a tumour-targeting mAb in *Nicotiana benthamiana* leaves**. *PLoS One* 2016, **11**: e0167086.

21. Grosse-Holz F, Madeira L, Zahid MA, Songer M, Kourelis J, Fesenko M, Ninck S, Kaschani F, Kaiser M, van der Hoorn RAL: **Three unrelated protease inhibitors enhance accumulation of pharmaceutical recombinant proteins in *Nicotiana benthamiana***. *Plant Biotechnol J* 2018, **16**:1797-1810.

22. Jutras PV, Grosse-Holz F, Kaschani F, Kaiser M, Michaud D, Hoorn RAL: **Activity-based proteomics reveals nine target proteases for the recombinant protein-stabilizing inhibitor SlCYS8 in *Nicotiana benthamiana***. *Plant Biotechnol J* 2019, **17**:1670-1678.

23. Sainsbury F, Jutras PV, Vorster J, Goulet M-C, Michaud D: **A chimeric affinity tag for efficient expression and chromatographic purification of heterologous proteins from plants**. *Front Plant Sci* 2016, **7**:1-11.

24. Pillay P, Kibido T, Du Plessis M, Van Der Vyver C, Beyene G, Vorster BJ, Kunert KJ, Schluter U: **Use of transgenic oryzacystatin-I-expressing plants enhances recombinant protein production**. *Appl Biochem Biotechnol* 2012, **168**:1608-1620.

25. Komarnytsky S, Borisjuk N, Yakoby N, Garvey A, Raskin I: **Cosecretion of protease inhibitor stabilizes antibodies produced by plant roots**. *Plant Physiol* 2006, **141**:1185-1193.

26. Kim TG, Kim HM, Lee HJ, Shin YJ, Kwon TH, Lee NJ, Jang YS, Yang MS: **Reduced protease activity in transformed rice cell suspension cultures expressing a proteinase inhibitor**. *Protein Expr Purif* 2007, **53**:270-274.

27. Goulet C, Benchabane M, Anguenot R, Brunelle F, Khalf M, Michaud D: **A companion protease inhibitor for the protection of cytosol-targeted recombinant proteins in plants**. *Plant Biotechnol J* 2010, **8**:142-154.

28. Zauner FB, Dall E, Regl C, Grassi L, Huber CG, Cabrele C, Brandstetter H: **Crystal structure of plant legumain reveals a unique two-chain state with pH-dependent activity regulation**. *Plant Cell* 2018, **30**:686-699.

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

30. Jutras PV, D'Aoust M-A, Couture MJ, Vézina L-P, Goulet M-C, Michaud D, Sainsbury F: **Modulating secretory pathway pH by proton channel co-expression can increase recombinant protein stability in plants**. *Biotechnol J* 2015, **10**:1478-1486.

31. Cady SD, Luo W, Hu F, Hong M: **Structure and function of the influenza A M2 proton channel**. *Biochemistry* 2009, **48**:7356-7364.

32. Jutras PV, Goulet MC, Lavoie PO, D'Aoust MA, Sainsbury F, Michaud D: **Recombinant protein susceptibility to proteolysis in the plant cell secretory pathway is pH-dependent**. *Plant Biotechnol J* 2018, **16**:1928-1938.

33. Jutras PV, Sainsbury F, Goulet M-C, Lavoie P-O, Tardif R, Hamel L-P, D'Aoust M-A, Michaud D: **pH gradient mitigation in the leaf cell secretory pathway alters the defense response of *Nicotiana benthamiana* to agroinfiltration**. *bioRxiv* 2018 http://dx.doi.org/10.1101/431767.

34. Mandal MK, Fischer R, Schillberg S, Schiermeyer A: **Inhibition of protease activity by antisense RNA improves recombinant protein production in *Nicotiana tabacum* cv. Bright Yellow 2 (BY-2) suspension cells**. *Biotechnol J* 2014, **9**:1065-1073.

35. Duwadi K, Chen L, Menassa R, Dhaubhadel S: **Identification, characterization and down-regulation of cysteine protease genes in tobacco for use in recombinant protein production**. *PLoS One* 2015, **10**:e0130556.

36. Armario Najera V, Twyman RM, Christou P, Zhu C: **Applications of multiplex genome editing in higher plants**. *Curr Opin Biotechnol* 2019, **59**:93-102.

37. Li J, Stoddard TJ, Demorest ZL, Lavoie P-O, Luo S, Clasen BM, Cedrone F, Ray EE, Coffman AP, Daulhac A *et al.*: **Multiplexed, targeted gene editing in *Nicotiana benthamiana* for glyco-engineering and monoclonal antibody production**. *Plant Biotechnol J* 2016, **14**:533-542.

38. Jansing J, Sack M, Augustine SM, Fischer R, Bortesi L: **CRISPR/**
• **Cas9-mediated knockout of six glycosyltransferase genes in *Nicotiana benthamiana* for the production of recombinant proteins lacking β-1,2-xylose and core α-1,3-fucose**. *Plant Biotechnol J* 2019, **17**:350-361
This article shows the capacity of genetically manipulate *Nicotiana benthamiana* to improve recombinant protein production. The authors

use CRISPR/Cas9 genome editing to generate *N. benthamiana* production lines deficient in plant-specific α-1,3-fucosyltransferase and β-1,2-xylosyltransferase activity.

39. Demir F, Niedermaier S, Villamor JG, Huesgen PF: **Quantitative proteomics in plant protease substrate identification**. *New Phytol* 2018, **218**:936-943.

40. Bhagwat SR, Hajela K, Kumar A: **Proteolysis to identify protease substrates: cleave to decipher**. *Proteomics* 2018, **18**:1-17.

41. Chen C, Mayer B, Schilling O: **Profiling of protease cleavage sites by proteome-derived peptide libraries and quantitative proteomics**. In   *Protein Terminal Profiling: Methods and Protocols*. Edited by Schilling O. New York: Springer; 2017: 197-204.

42. Morimoto K, van der Hoorn RAL: **The increasing impact of activity-based protein profiling in plant science**. *Plant Cell Physiol* 2016, **57**:446-461.