Arabinosylation Modulates the Growth-Regulating Activity of the Peptide Hormone CLE40a from Soybean

Leo Corcilius,[1,4] April H. Hastwell,[2,4] Mengbai Zhang,[2] James Williams,[1] Joel P. Mackay,[3] Peter M. Gresshoff,[2] Brett J. Ferguson,[2]* and Richard J. Payne[1,5]*

[1] School of Chemistry, The University of Sydney, NSW 2006, Australia

[2] Centre for Integrative Legume Research, School of Agriculture and Food Sciences, The University of Queensland, Brisbane, QLD 4072, Australia

[3] School of Life and Environmental Sciences, The University of Sydney, NSW 2006, Australia,

[4] These authors contributed equally

[5] Lead contact

*e-mail: richard.payne@sydney.edu.au and b.ferguson1@uq.edu.au.

SUMMARY

Small post-translationally modified peptide hormones mediate crucial developmental and regulatory processes in plants. CLAVATA/ENDOSPERM-SURROUNDING REGION (CLE) genes are found throughout the plant kingdom and encode for 12-13 amino acid peptides that must often undergo post-translational proline hydroxylation and glycosylation with O-β1,2-triarabinose moieties before they become functional. Apart from a few recent examples, a detailed understanding of the structure and function of most CLE hormones is yet to be uncovered. This is mainly owing to difficulties in isolating mature homogeneously modified CLE peptides from natural plant sources. In this study, we describe the efficient synthesis of a synthetic Araf$_3$Hyp glycosylamino acid building block that was used to access a hitherto uninvestigated CLE hormone from soybean called GmCLE40a. Through the development and

implementation of a novel in vivo root growth assay, we show that the synthetic triarabinosylated glycopeptide enhances suppression of primary root growth in this important crop species.

KEYWORDS

Glycopeptide, glycosylation, arabinosylation, CLE, hormone, soybean

INTRODUCTION

CLAVATA/Endosperm-Surrounding Region (CLE) genes were first discovered in *Arabidopsis thaliana* in 1999, and have since been identified throughout the plant kingdom (Fletcher et al., 1999; Hastwell et al., 2015a; Oelkers et al., 2008; Okamoto et al., 2009; Strabala et al., 2014; Zhang et al., 2014). CLE genes encode short (12-13 amino acid) peptide hormones with up to 3 highly conserved proline residues (Hastwell et al., 2015b; Kucukoglu and Nilsson, 2015). To date, only a handful of mature functional CLE peptides have been isolated and structurally characterized. For those which have been isolated, prolines 4 and 7 are almost invariably post-translationally hydroxylated to form *trans*-4-hydroxy-L-hydroxyproline (Hyp) residues (Ito et al., 2006; Kondo et al., 2006; Ohyama et al., 2009; Shinohara et al., 2012). In addition, Hyp-7 can be side chain *O*-glycosylated with the plant specific β1,2-linked tri-L-arabinofuranosyl oligomer to form a central triarabinosylated Hyp (Ara*f*$_3$Hyp) residue (Ohyama et al., 2009). In recent reports, the Ara*f*$_3$Hyp motif has proven to be crucial for the activity of several plant hormones (Okamoto et al., 2013; Shinohara and Matsubayashi, 2013; Xu et al., 2015).

Triarabinosylated CLE glycopeptides mediate diverse developmental processes in plants (Kucukoglu and Nilsson, 2015). For example, the first triarabinosylated CLE glycopeptide to

be identified in Arabidopsis, *At*CLV3, is responsible for the negative regulation of stem cell differentiation in the shoot apical meristem (SAM), which gives rise to the above ground features of the plant (Clark et al., 1993; Clark et al., 1995; Clark et al., 1997; Ogawa et al., 2008; Schoof et al., 2000). The recently discovered orthologues of AtCLV3, *Sl*CLV3 and *Sl*CLE9, fulfill the same function in tomato and have been identified as a key factors which control the size and number of fruiting organs (Xu et al., 2015). These latter two peptides represent the only two mature functional CLE hormones to be structurally characterized in a commercially important crop species, highlighting a hitherto unknown molecular basis for the artificial selection of tomato cultivars with better fruiting yields.

The legume (*Fabaceae*) family contains several important crop and pasture species including soybean, pea, common bean, mung bean, clover, cowpea, alfalfa, chickpea, lentil and peanut. Recently, a systemic signal responsible for the negative regulation of root nodule production in *Lotus japonicus* was identified as a triarabinosylated CLE glycopeptide, with a structure analogous to that of CLV3 (Okamoto et al., 2013). While several other CLE peptide-encoding genes have been identified in legume species, the mature structures and functions of these CLE hormones have yet to be elucidated (Hastwell et al., 2015a; Hastwell et al., 2015b; Oelkers et al., 2008). Hence, probing the structure and function of CLE hormones from the world's most economically significant legume crop species, soybean, is of considerable interest. Recently, the complete family of soybean CLE peptides was identified and genetically characterized (Hastwell et al., 2015a). This includes the orthologue of *AtCLE40*, a CLV3-related peptide that maintains essential stem cell homeostasis in the root apical meristem (RAM) (Greb and Lohmann, 2016; Yamaguchi et al., 2016). Soybean is a paleopolyploid, having undergone genome-wide duplication events roughly 53 and 19 million years ago, followed by the process of diploidisation (Schmutz et al., 2010). Consequently, most genes in the soybean genome are

duplicated; however, in some instances, one of the gene copies has subsequently undergone genetic variation and/or loss. This is the case with the soybean hormone CLE40, with *GmCLE40a* maintaining a typical CLE prepropeptide sequence, whereas its homeologous duplicate, *GmCLE40b*, encodes a nonsense mutation upstream of the CLE peptide domain that likely renders its ligand un-transcribed and functionless (Hastwell et al., 2015a).

Modifying root architecture via key developmental factors is viewed as a pivotal step in enhancing agricultural sustainability and food security (Meister et al., 2014). CLE40 represents a logical molecular component to evaluate for this purpose due to its central role in root organogenesis and in light of the aforementioned findings with the functionally-related CLV3 in enhancing fruit development in tomato. Importantly, CLE40 contains the amino acid motif present in AtCLV3 that is required for arabinosylation (Ohyama et al., 2009), which we proposed (based on the CLE hormones) would be vital for optimum biological activity (Shinohara and Matsubayashi, 2013). However, to date the investigation of the effect of arabinosylation on CLE40 in any species has not been possible due to the difficulty in isolating homogenously modified hormone in sufficient quantity and purity for biological study. This is owing to the enzymatic nature of the post-ribosomal hydroxylation and glycosylation events that means that peptide hormones are produced as complex mixtures containing both the triarabinosylated and unglycosylated variants and varying proline hydroxylation patterns (Matsubayashi, 2014). Both the inherent challenges associated with purifying such complex mixtures and the low concentration of the hormone in plant tissues mean that chemical synthesis is currently the only viable means to obtain useful quantities of homogeneous triarabinosylated plant glycopeptides for biological study. As such, synthetic methods that enable access to peptides and proteins bearing the Ara*f*$_3$Hyp modification are of enormous

interest to advance the field of plant molecular biology and is a key focus of the work reported herein.

The synthesis of Ara*f*$_3$Hyp is complicated by the presence of contiguous β-arabinofuranosidic linkages which, even in a non-contiguous setting, are difficult to construct stereoselectively (Lowary, 2003; Yin and Lowary, 2001). Although several chemical methodologies have been developed for the direct construction of β-arabinofuranosidic linkages (Crich et al., 2007; Gadikota et al., 2003; Ishiwata et al., 2006; Lee et al., 2005; Li and Singh, 2001; Liu et al., 2013; Zhu et al., 2006), such methodologies do not provide complete stereoselectivity when applied to the construction of plant derived glycans containing Ara*f*-β-Hyp and contiguous Ara*f*-β1,2-Ara*f* linkages, such as those present in Ara*f*$_3$Hyp (Kaeothip and Boons, 2013; Kaeothip et al., 2013; Xie and Taylor, 2010). Recently, the Matsubayashi (Shinohara and Matsubayashi, 2013) and Ito (Kaeothip et al., 2013) groups reported highly stereoselective routes to an Fmoc-protected and peracetylated Ara*f*$_3$Hyp glycosylamino acid building block **1** (in box, Figure 1), and subsequent incorporation into the *At*CLV3 peptide through Fmoc-strategy solid phase peptide synthesis (Fmoc-SPPS). Ito and co-workers have also reported the use of the Ara*f*$_3$Hyp building block for the synthesis of a multiply glycosylated cell wall extensin fragment which is also known to natively bear the triarabinose moiety (Ishiwata et al., 2014). Key to both synthetic strategies was the use of an intramolecular aglycone delivery (IAD) strategy, mediated by *p*-methoxybenzyl (PMB) ether (Désiré and Prandi, 1999) and naphthylmethyl (NAP) ether (Ishiwata et al., 2008) auxiliaries respectively, to construct each β-arabinofuranosidic linkage with complete stereoselectivity.

Herein, we report a highly efficient synthesis of Ara*f*$_3$Hyp glycosylamino acid building block **1** using a modified NAP ether-mediated-IAD (NAP-IAD) strategy (Kaeothip et al., 2013) and

its use for the chemical synthesis of the soybean-derived triarabinosylated CLE glycopeptide, *Gm*CLE40a **2a** (Figure 1). We also report the functional characterization of **2a** through biological evaluation and demonstrate the importance of the central Ara*f*$_3$Hyp moiety through comparison with the synthetic Hyp7 unglycosylated variant **2b**.

RESULTS AND DISCUSSION

Synthesis of Triarabinosylated *Gm*CLE40a Glycopeptide. In order to access sufficient quantities of the homogeneous triarabinosylated *Gm*CLE40a glycopeptide **2a**, we sought a highly stereoselective route to the Fmoc-protected and peracetylated Ara*f*$_3$Hyp building block **1** using an IAD-based synthetic strategy. One major drawback of the IAD route reported previously was the low yielding formation of the two sterically crowded Ara*f*-β1,2-Ara*f* glycosidic linkages when a conventional 'donor-mediated' tethering step was applied to form the crucial mixed acetal intermediates (Kaeothip et al., 2013; Shinohara and Matsubayashi, 2013). Ito and co-workers addressed this issue through recourse to an 'acceptor-mediated' tethering approach, requiring installation of the NAP ether on each acceptor prior to tethering (Kaeothip et al., 2013). While this provided the requisite mixed acetal intermediates in high yield, it added extra linear steps to the synthesis for auxiliary installation and protecting group manipulation. To overcome these issues, we proposed a modification of the NAP-IAD strategy involving less sterically demanding *O*-protecting groups, such as acetyl groups, with the view to improving the yields for 'donor-mediated' tethering steps, thereby enabling access to all three β-arabinofuranosidic linkages in Ara*f*$_3$Hyp using a single 2-*O*-NAP-derived arabinofuranosyl donor.

To this end, we first synthesized thioglycoside donor **3** bearing 3,5-di-*O*-acetyl protection and a 2-*O*-NAP ether auxiliary for IAD (Figure 2, see supplemental synthetic details file for donor

preparation). Donor **3** was tethered to acceptor, Fmoc-Hyp-OBn (1 equiv.), under the promotion of DDQ, affording the bench stable mixed acetal **4**, which was taken forward without purification. The crude acetal **4** was unreactive when treated under conventional promotion conditions [MeOTf (4 equiv.)/TTBP (5 equiv.)] but smoothly underwent IAD in the presence of MeOTf/Me_2S_2 (4 equiv.)/TTBP (5 equiv.) to give the di-*O*-acetylated-Ara*f*-β-Hyp monoglycoside **5** in 56% overall yield after acidolytic workup (10% TFA in $CHCl_3$). This overall yield was comparable to that obtained in a control experiment with flash chromatographic purification of the intermediate mixed acetal. Importantly, the Ara*f*-β-Hyp linkage was formed with complete stereoselectivity by virtue of the IAD transformation.

Next, we attempted the tethering of NAP ether **3** to acceptor **5** to provide the requisite mixed acetal for formation of the first challenging Ara*f*-β1,2-Ara*f* linkage. Pleasingly, mixed acetal **6** was afforded in an excellent isolated yield of 85% using only 1.05 equiv. of NAP ether **3** when the reaction was tested on a small scale. As discussed above, this particular tethering step has been reported to be low yielding in both previously reported syntheses of the Ara*f*$_3$Hyp building block **1** (Kaeothip et al., 2013; Shinohara and Matsubayashi, 2013). We attribute the improvement in tethering yield to the sterically unencumbering acetyl-protected donor and acceptor combination. In a scale up of the tethering procedure, the mixed acetal **6** was used directly in the next IAD step without purification to afford diglycoside acceptor **7** in 65% overall yield with exclusively β-stereoselectivity. An excess of donor **3** (2 equiv.) was employed to push the next tethering reaction to completion to afford mixed acetal **8** in 75% yield after purification by flash column chromatography. Subsequent rearrangement gave triglycoside **9**, containing all three β-arabinofuranosidic linkages, in 85% yield. To complete the synthesis of **1**, treatment of **9** with Ac_2O in pyridine provided the peracetylated glycoside **10**, which was subjected to a mild transfer hydrogenation procedure using Et_3SiH and Pd/C

catalyst to remove the benzyl ester without affecting the Fmoc protecting group (Mandal and McMurray, 2007). This provided the desired glycosylamino acid building block **1** bearing the Ara*f*$_3$Hyp moiety in 86% yield. Overall the Ara*f*$_3$Hyp building block **1** was assembled in 18% yield over 11 steps, which significantly improves on the best synthesis previously reported (Kaeothip et al., 2013) using the longer 'acceptor-mediated' NAP-IAD strategy (12% over 15 steps).

With the suitably protected Ara*f*$_3$Hyp building block **1** in hand, the synthesis of the target triarabinosylated CLE glycopeptide **2a** could now commence (Figure 3). Towards this end, 2-chlorotrityl chloride resin was first loaded with Fmoc-His(Trt)-OH, followed by conventional Fmoc-SPPS conditions to afford resin-bound hexapeptide **11**. The resin bound hexapeptide **11** was subsequently Fmoc-deprotected and treated with a coupling mixture containing 1.2 equiv. of the Ara*f*$_3$Hyp glycosylamino acid building block **1**, 1-[bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU), 1-Hydroxy-7-azabenzotriazole (HOAt) and *i*Pr$_2$NEt. Under these conditions, the glycosylamino acid was incorporated quantitatively (as judged by LC-MS analysis of a test cleavage) to provide resin-bound glycoheptapeptide **12**. Subsequent extension using conventional Fmoc-SPPS gave the resin-bound glycotridecapeptide, which was Fmoc-deprotected and acidolytically released from the resin with concomitant global deprotection of the side chain protecting groups from the amino acids. The carbohydrate moiety was next deacetylated in solution using NaOMe in MeOH before purification by reversed-phase HPLC using 0.1% trifluoroacetic acid in the eluent. After lyophilization, the triarabinosylated *Gm*CLE40a glycopeptide **2a** was isolated in an excellent overall yield of 22% over the 26 linear steps (~94% per step).

Functional characterization of the *GmCLE40a* glycopeptide hormone. Only four putative CLE40 orthologues have been reported to date (in *Arabidopsis*, soybean, common bean and rice) (Hastwell et al., 2015a; Hobe et al., 2003; Kinoshita et al., 2007; Sharma et al., 2003). To better understand the signal provided by the *GmCLE40a* glycopeptide and establish amino acid conservation within the peptide ligand, BLAST searches were conducted across a range of plant species. A total of 26 putative CLE40 orthologues (including *Gm*CLE40b) were identified across 21 different species (Figure 4A). Each contain two introns, consistent with the CLV3 and CLE40 encoding genes of *Arabidopsis*, soybean and common bean. Interestingly, those from Arabidopsis and the Brassicaceae family form a distinct branch with a high bootstrap value (94.2), and those identified in monocot species group within the same clade as CLV3, but on a distinct and well supported (96.9) branch (Figure 4). This includes *Os*FCP1 (also known as *Os*CLE402). The CLE domain of the orthologues is highly conserved, with only four residues showing less than 90% pairwise identity (positions 2,7,9 and 13; Figure 4A and 4B). The amino acid residue at position 7 is a proline in 65% (including *Gm*CLE40a) with a serine in 31% (including *At*CLE40, and orthologues from other species within the Brassicaceae family). Importantly, proline and serine are both residues that can be subjected to *O*-glycosylation (Van den Steen et al., 1998), and all of the CLE40 orthologues identified contain the motif for arabinosylation.

Development of a bioassay to assess *Gm*CLE40a activity in soybean. To evaluate *Gm*CLE40a activity, a novel bioassay was developed to quantify the effect of the peptide on soybean root growth. Initially, *Agrobacterium*-mediated soybean hairy root transformation was carried out to establish where *GmCLE40a* is transcriptionally active, and hence where to apply synthetic peptide hormones in the bioassay. These studies involved driving GUS reporter gene expression with the 2.5 kB promotor region located directly upstream of *GmCLE40a*. GUS

222 expression was observed in the apical region of the root tip (Figure 5), which is in agreement
223 with the peptide's role in regulating the stem cell population of the RAM, and consistent with
224 the expression pattern of *AtCLE40* (Stahl et al., 2009). Based on this, different concentrations
225 of the *Gm*CLE40a glycopeptide **2a** or its unglycosylated variant **2b** were applied every 12 h to
226 the tap root tip of wild-type soybean seedlings, and the length of the root was subsequently
227 recorded. Specialized growth pouches were modified and used to enable precision-feeding of
228 the peptides and to record the development of the tap root in a non-destructive and repetitious
229 manner (Figure S1). It is important to note that whilst most feeding studies broadly apply
230 peptides to the entire root/plant, precision-feeding is a highly localized technique, minimizing
231 unwanted and biologically irrelevant responses.
232
233 **Triarabinosylated *Gm*CLE40a glycopeptide possesses potent root growth inhibition in**
234 **soybean.** To determine the biological activity of *Gm*CLE40a variants, soybean seedlings
235 treated with different concentrations of the triarabinosylated *Gm*CLE40a glycopeptide **2a**, the
236 unglycosylated *Gm*CLE40a variant **2b**, or water (control) were used in the root-growth
237 bioassay. Tap root lengths were measured throughout the experiment to establish the effect of
238 the treatments on growth over time (Figure 6).
239
240 Compared with the control treatment, application of the triarabinosylated *Gm*CLE40a
241 glycopeptide **2a** significantly reduced root growth at concentrations of 10^{-4} ($P<0.0001$) and 10^{-6} M ($P\leq0.05$; Figure 6). In contrast, *Gm*CLE40a peptide **2b**, containing an unmodified Hyp
243 residue, only inhibited root growth when applied at 10^{-4} M ($P\leq0.05$). A significant difference
244 in root length was also observed between the glycosylated **2a** and unglycosylated peptide **2b**
245 treatments, with the former suppressing root growth significantly more than the latter at both
246 10^{-4} and 10^{-6} M ($P\leq0.05$, Figure 6). Moreover, for each concentration tested the growth-rate of

247 the root was significantly reduced by arabinosylated *Gm*CLE40a **2a** compared with **2b** (Figure
248 7). These results demonstrate that arabinosylated *Gm*CLE40a is significantly more potent than
249 the hydroxylated version of the peptide at reducing root growth. This indicates an important
250 role for the carbohydrate moiety in CLE40 activity, possibly acting directly in perception
251 and/or protection of the peptide ligand from peptidase breakdown.

252 Suppression of soybean root growth by *Gm*CLE40a application is consistent with *A. thaliana*
253 studies using semi- or unmodified *At*CLE40 (Fiers et al., 2005). Excess levels of *At*CLE40
254 caused by over-expression can also significantly reduce root growth, and intriguingly so can
255 reduced levels caused by genetic mutation (Hobe et al., 2003). This suggests that homeostasis
256 is required for optimum root growth, where either elevated or reduced levels of the peptide can
257 prevent maximum growth. Interestingly, application of hydroxylated *Gm*CLE40a at 10^{-8} M led
258 to a mild yet significant enhancement of soybean root growth compared with the arabinosylated
259 glycopeptide and water control treatments ($P \leq 0.05$; Fig. 2A). To the best of our knowledge,
260 this is the first report of such an increase and it is tempting to speculate that certain CLE40
261 analogues have the potential to enhance root growth, which would have tremendous
262 commercial and agricultural potential. It should be noted that this reversed effect of promoting
263 or inhibiting plant development when exogenously applying different concentrations of a plant
264 hormone has previously been reported for gibberellin, brassinosteroids, auxin and cytokinin
265 (e.g. Ferguson and Mathesius, 2014; Hayashi et al., 2014; Wei and Li, 2016).
266
267 **NMR conformational analysis of CLE40a (glyco)peptides.** Using NMR and computational
268 techniques, Shinohara and Matsubayashi (Shinohara and Matsubayashi, 2013) have previously
269 shown that the central triarabinose of the CLE40a orthologue, *At*CLV3, causes the C-terminus
270 to bend away from the glycan through a 'kink' at the conserved Gly-6 residue. Based on this
271 data, the authors proposed that the triarabinose moiety on CLV3 may be crucial for maintaining

the correct conformation of the peptide ligand for receptor binding and the downstream biological activity. In order to determine whether the same conformational effect was operative in glycosylated CLE40a, we conducted comparative 2D NOESY analysis of CLE40a (glyco)peptides **2a** and **2b**. Using the homonuclear TOCSY, DQF-COSY and NOESY spectra, we made full ^{1}H resonance assignments for both the peptide and sugar portions of the two molecules at 278 K. A number of NOEs were observed, particularly at the lower temperature, that were consistent with the peptides displaying significant conformational preferences. For example, HN-HN(*i*,*i*+1) NOEs were observed for the residue pairs T5-G6, L10-H11 and H11-H12. As exemplified by the plot of Hα chemical shifts for both **2a** and **2b** (Figure S2), there were no significant chemical shift changes between the glycosylated and non-glycosylated forms of CLE40a, other than changes in the sidechain of HyP7 that are expected from addition of the trisaccharide unit. This, together with the observation that very similar NOE patterns were observed for the two peptides (Figure S3), strongly suggests that glycosylation of CLE40a does not have a significant effect on the conformational preferences of the peptide.

Given that triabrabinosylation of CLE40a does not provide the same conformational changes to the underlying peptide backbone as observed for AtCLV3 it can be deduced that alteration to the conformation or shape of the peptide is not responsible for the increased inhibition of root growth observed for **2a**, compared the unmodified Hyp-containing peptide **2b**. As such, it is possible that the improved activity is owing to improved interaction with the putative receptor through H-bonding interactions with the carbohydrate moiety. Alternatively, the triarabinose unit may provide improved proteolytic stability of the peptide hormone that would enhance the half-life, and therefore activity of the hormone. Studies to address these possibilities will be the subject of future work in our laboratories.

In summary, we have developed an efficient synthetic route to a suitably protected glycosylamino acid building block bearing Ara*f*$_3$Hyp, a post-translational modification that has recently emerged as a common feature of plant peptide hormones and proteins. The building block was used to access a homogeneous arabinosylated *Gm*CLE40a glycopeptide, which functions to control the stem cell population of the root apical meristem of plants. Moreover, a novel and highly effective bioassay was developed to evaluate the peptide's activity in relation to root growth. Findings from this work demonstrate that the Ara*f*$_3$Hyp residue significantly enhances *Gm*CLE40a root growth suppressive activity. This raises a pertinent and fundamental question relating to the use of plant peptide variants in application and binding studies, which are often performed using only semi- or unmodified variants to ascertain peptide perception and function. Whether the Ara*f*$_3$Hyp residue of CLE40 is optimal for receptor binding, enhanced ligand stability, localization, or some other aspect that promotes the peptide's activity is now of great interest to determine. Moreover, the Ara*f*$_3$Hyp residue prepared here can now be used to synthesize additional plant proteins and peptide signals of interest, with a focus on those that could potentially benefit crop development and yields. Studies toward this end will be the subject of future work in our laboratories.

SIGNIFICANCE

The ubiquity and diverse functionality of CLE hormones make them important research targets in the study of plant development. Most CLE hormones isolated to date are triarabinosylated and possess little or no biological activity without this critical post-translational modification. However, the isolation of mature functional CLE hormones is not always possible, meaning that structural and functional characterization is dependent on access to the homogeneous triarabinosylated isoform through chemical synthesis. This paper outlines a more

straightforward and higher yielding route to the Ara*f*3Hyp glycosylamino acid and demonstrates its utility through chemical synthesis and functional characterization of Soybean CLE40a using a novel root growth bioassay. The tools presented herein should assist with the functional characterization of new CLE hormones and other triarabinosylated plant glycopeptides.

AUTHOR CONTRIBUTIONS

L.C. and J.W. performed the chemical synthesis and compound characterization. A.H.H. and M.Z. conducted plant experiments. J.P.M ran structural NMR studies. L.C and R.J.P. conceived and designed the synthesis of the triarabinosylated hydroxyproline building block. L. C. A.H.H, P.M.G, B.J.F and R. J. P. conceived the CLE40 glycopeptide as a synthetic target. A.H.H, P.M.G and B.J.F designed the plant experiments. L.C., A.H.H, B. J. F. and R. J. P. wrote the manuscript with the assistance of all authors. All authors participated in data analysis and discussions.

ACKNOWLEDGEMENTS

This work was supported by an Australian Research Council Future Fellowship (FT130100150) to RJP. The work was also funded by the Hermon Slade Foundation, and Australian Research Council Discovery Project grants (DP130103084 and DP130102266) to BJF and PMG. The Fellowship Fund Inc. is also thanked for provision of a Molly-Budtz Olsen PhD fellowship to AHH. We also gratefully acknowledge the funding provided to LC by the John A. Lamberton research scholarship and the Agnes Campbell postgraduate prize. We would like to thank Huanan Su, Xitong Chu and Dongxue Li for their technical assistance.

REFERENCES

Broughton, W.J., and Dilworth, M.J. (1971). Control of leghaemoglobin synthesis in snake beans. Biochem. J. *125*, 1075-1080.

Clark, S.E., Running, M.P., and Meyerowitz, E.M. (1993). CLAVATA1, a regulator of meristem and flower development in Arabidopsis. Development *119*, 397-418.

Clark, S.E., Running, M.P., and Meyerowitz, E.M. (1995). CLAVATA3 is a specific regulator of shoot and floral meristem development affecting the same processes as CLAVATA1. Development *121*, 2057-2067.

Clark, S.E., Williams, R.W., and Meyerowitz, E.M. (1997). The CLAVATA1Gene Encodes a Putative Receptor Kinase That Controls Shoot and Floral Meristem Size in Arabidopsis. Cell *89*, 575-585.

Crich, D., Pedersen, C.M., Bowers, A.A., and Wink, D.J. (2007). On the Use of 3,5-O-Benzylidene and 3,5-O-(Di-tert-butylsilylene)-2-O-benzylarabinothiofuranosides and Their Sulfoxides as Glycosyl Donors for the Synthesis of β-Arabinofuranosides: Importance of the Activation Method. J. Org. Chem. *72*, 1553-1565.

Désiré, J., and Prandi, J. (1999). Synthesis of methyl β-d-arabinofuranoside 5-[1D (and L)-myo-inositol 1-phosphate], the capping motif of the lipoarabinomannan of Mycobacterium smegmatis. Carbohydr. Res. *317*, 110-118.

Ferguson, B.J., Li, D., Hastwell, A.H., Reid, D.E., Li, Y., Jackson, S.A., and Gresshoff, P.M. (2014). The soybean (Glycine max) nodulation-suppressive CLE peptide, GmRIC1, functions interspecifically in common white bean (Phaseolus vulgaris), but not in a supernodulating line mutated in the receptor PvNARK. Plant Biotechnol. J. *12*, 1085-1097.

Ferguson, B.J., and Mathesius, U. (2014). Phytohormone regulation of legume-rhizobia interactions. J. Chem. Ecol. *40*, 770-790.

Fiers, M., Golemiec, E., Xu, J., van der Geest, L., Heidstra, R., Stiekema, W., and Liu, C.-M. (2005). The 14–Amino Acid CLV3, CLE19, and CLE40 Peptides Trigger Consumption of the Root Meristem in Arabidopsis through a CLAVATA2-Dependent Pathway. Plant Cell *17*, 2542-2553.

Fletcher, J.C., Brand, U., Running, M.P., Simon, R., and Meyerowitz, E.M. (1999). Signaling of Cell Fate Decisions by CLAVATA3 in Arabidopsis Shoot Meristems. Science *283*, 1911-1914.

Gadikota, R.R., Callam, C.S., Wagner, T., Del Fraino, B., and Lowary, T.L. (2003). 2,3-Anhydro Sugars in Glycoside Bond Synthesis. Highly Stereoselective Syntheses of Oligosaccharides Containing α- and β-Arabinofuranosyl Linkages. J. Am. Chem. Soc. *125*, 4155-4165.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., and Putnam, N. (2012). Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. *40*, D1178-D1186.

382 Greb, T., and Lohmann, J.U. (2016). Plant Stem Cells. Curr. Biol. *26*, R816-R821.

383 Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O. (2010).
384 New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the
385 performance of PhyML 3.0. Syst. Biol. *59*, 307-321.

386 Hastwell, A.H., Gresshoff, P.M., and Ferguson, B.J. (2015a). Genome-wide annotation and
387 characterization of CLAVATA/ESR (CLE) peptide hormones of soybean (Glycine max) and
388 common bean (Phaseolus vulgaris), and their orthologues of Arabidopsis thaliana. J. Exp. Bot.
389 *66*, 5271-5287.

390 Hastwell, A.H., Gresshoff, P.M., and Ferguson, B.J. (2015b). The structure and activity of
391 nodulation-suppressing CLE peptide hormones of legumes. Funct. Plant Biol. *42*, 229-238.

392 Hayashi, S., Gresshoff, P.M., and Ferguson, B.J. (2014). Mechanistic action of gibberellins in
393 legume nodulation. J. Integr. Plant Biol. *56*, 971-978.

394 Hayashi, S., Reid, D.E., Lorenc, M.T., Stiller, J., Edwards, D., Gresshoff, P.M., and Ferguson,
395 B.J. (2012). Transient Nod factor-dependent gene expression in the nodulation-competent zone
396 of soybean (Glycine max [L.] Merr.) roots. Plant Biotechnol. J. *10*, 995-1010.

397 Hobe, M., Müller, R., Grünewald, M., Brand, U., and Simon, R. (2003). Loss of CLE40, a
398 protein functionally equivalent to the stem cell restricting signal CLV3, enhances root waving
399 in Arabidopsis. Dev. Genes Evol. *213*, 371-381.

400 Ishiwata, A., Akao, H., and Ito, Y. (2006). Stereoselective Synthesis of a Fragment of
401 Mycobacterial Arabinan. Org. Lett. *8*, 5525-5528.

402 Ishiwata, A., Kaeothip, S., Takeda, Y., and Ito, Y. (2014). Synthesis of the Highly Glycosylated
403 Hydrophilic Motif of Extensins. Angew. Chem. Int. Ed. *53*, 9812-9816.

404 Ishiwata, A., Munemura, Y., and Ito, Y. (2008). NAP Ether Mediated Intramolecular Aglycon
405 Delivery: A Unified Strategy for 1,2-cis-Glycosylation. Eur. J. Org. Chem. *2008*, 4250-4263.

406 Ito, Y., Nakanomyo, I., Motose, H., Iwamoto, K., Sawa, S., Dohmae, N., and Fukuda, H.
407 (2006). Dodeca-CLE peptides as suppressors of plant stem cell differentiation. Science *313*,
408 842-845.

409 Kaeothip, S., and Boons, G.-J. (2013). Chemical synthesis of β-arabinofuranosyl containing
410 oligosaccharides derived from plant cell wall extensins. Org. Biomol. Chem. *11*, 5136-5146.

411 Kaeothip, S., Ishiwata, A., and Ito, Y. (2013). Stereoselective synthesis of Arabidopsis
412 CLAVATA3 (CLV3) glycopeptide, unique protein post-translational modifications of secreted
413 peptide hormone in plant. Org. Biomol. Chem. *11*, 5892-5907.

414 Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S.,
415 Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: An integrated and
416 extendable desktop software platform for the organization and analysis of sequence data.
417 Bioinformatics *28*, 1647-1649.

Kereszt, A., Li, D., Indrasumunar, A., Nguyen, C.D.T., Nontachaiyapoom, S., Kinkema, M., and Gresshoff, P.M. (2007). Agrobacterium rhizogenes-mediated transformation of soybean to study root biology. Nat. Protocols *2*, 948-952.

Kinoshita, A., Nakamura, Y., Sasaki, E., Kyozuka, J., Fukuda, H., and Sawa, S. (2007). Gain-of-function phenotypes of chemically synthetic CLAVATA3/ESR-related (CLE) peptides in Arabidopsis thaliana and Oryza sativa. Plant Cell Physiol. *48*, 1821-1825.

Kondo, T., Sawa, S., Kinoshita, A., Mizuno, S., Kakimoto, T., Fukuda, H., and Sakagami, Y. (2006). A Plant Peptide Encoded by CLV3 Identified by in situ MALDI-TOF MS Analysis. Science *313*, 845-848.

Kucukoglu, M., and Nilsson, O. (2015). CLE peptide signaling in plants – the power of moving around. Physiol. Plant. *155*, 74-87.

Larkin, P.J., Gibson, J.M., Mathesius, U., Weinman, J.J., Gartner, E., Hall, E., Tanner, G.J., Rolfe, B.G., and Djordjevic, M.A. (1996). Transgenic white clover. Studies with the auxin-responsive promoter, GH3, in root gravitropism and lateral root development. Transgenic Res. *5*, 325-335.

Lee, Y.J., Lee, K., Jung, E.H., Jeon, H.B., and Kim, K.S. (2005). Acceptor-Dependent Stereoselective Glycosylation: 2‘-CB Glycoside-Mediated Direct β-d-Arabinofuranosylation and Efficient Synthesis of the Octaarabinofuranoside in Mycobacterial Cell Wall. Org. Lett. *7*, 3263-3266.

Li, Y., and Singh, G. (2001). Synthesis of d-arabinofuranosides using propane-1,3-diyl phosphate as the anomeric leaving group. Tetrahedron Lett. *42*, 6615-6618.

Lin, Y.-H., Ferguson, B.J., Kereszt, A., and Gresshoff, P.M. (2010). Suppression of hypernodulation in soybean by a leaf-extracted, NARK- and Nod factor-dependent, low molecular mass fraction. New Phytol. *185*, 1074-1086.

Liu, Q.-W., Bin, H.-C., and Yang, J.-S. (2013). β-Arabinofuranosylation Using 5-O-(2-Quinolinecarbonyl) Substituted Ethyl Thioglycoside Donors. Org. Lett. *15*, 3974-3977.

Lowary, T.L. (2003). Synthesis and conformational analysis of arabinofuranosides, galactofuranosides and fructofuranosides. Curr. Opin. Chem. Biol. *7*, 749-756.

Lux, A., Morita, S., Abe, J.U.N., and Ito, K. (2005). An Improved Method for Clearing and Staining Free-hand Sections and Whole-mount Samples*. Ann. Bot. *96*, 989-996.

Mandal, P.K., and McMurray, J.S. (2007). Pd−C-Induced Catalytic Transfer Hydrogenation with Triethylsilane. J. Org. Chem. *72*, 6599-6601.

Matsubayashi, Y. (2014). Posttranslationally Modified Small-Peptide Signals in Plants. Annu. Rev. Plant Biol. *65*, 385-413.

Meister, R., Rajani, M., Ruzicka, D., and Schachtman, D.P. (2014). Challenges of modifying root traits in crops for agriculture. Trends Plant Sci. *19*, 779-788.

Oelkers, K., Goffard, N., Weiller, G.F., Gresshoff, P.M., Mathesius, U., and Frickey, T. (2008). Bioinformatic analysis of the CLE signaling peptide family. BMC Plant Biol. *8*, 1-15.

Ogawa, M., Shinohara, H., Sakagami, Y., and Matsubayashi, Y. (2008). Arabidopsis CLV3 Peptide Directly Binds CLV1 Ectodomain. Science *319*, 294.

Ohyama, K., Shinohara, H., Ogawa-Ohnishi, M., and Matsubayashi, Y. (2009). A glycopeptide regulating stem cell fate in Arabidopsis thaliana. Nat. Chem. Biol. *5*, 578-580.

Okamoto, S., Ohnishi, E., Sato, S., Takahashi, H., Nakazono, M., Tabata, S., and Kawaguchi, M. (2009). Nod Factor/Nitrate-Induced CLE Genes that Drive HAR1-Mediated Systemic Regulation of Nodulation. Plant Cell Physiol. *50*, 67-77.

Okamoto, S., Shinohara, H., Mori, T., Matsubayashi, Y., and Kawaguchi, M. (2013). Root-derived CLE glycopeptides control nodulation by direct binding to HAR1 receptor kinase. Nat. Commun. *4*.

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., and Cheng, J. (2010). Genome sequence of the palaeopolyploid soybean. Nature *463*, 178-183.

Schoof, H., Lenhard, M., Haecker, A., Mayer, K.F.X., Jürgens, G., and Laux, T. (2000). The Stem Cell Population of Arabidopsis Shoot Meristems Is Maintained by a Regulatory Loop between the CLAVATA and WUSCHEL Genes. Cell *100*, 635-644.

Sharma, V.K., Ramirez, J., and Fletcher, J.C. (2003). The Arabidopsis CLV3-like (CLE) genes are expressed in diverse tissues and encode secreted proteins. Plant Mol. Biol. *51*, 415-425.

Shinohara, H., and Matsubayashi, Y. (2013). Chemical synthesis of Arabidopsis CLV3 glycopeptide reveals the impact of hydroxyproline arabinosylation on peptide conformation and activity. Plant Cell Physiol. *54*, 369-374.

Shinohara, H., Moriyama, Y., Ohyama, K., and Matsubayashi, Y. (2012). Biochemical mapping of a ligand-binding domain within Arabidopsis BAM1 reveals diversified ligand recognition mechanisms of plant LRR-RKs. Plant J. *70*, 845-854.

Sievers, F., Wilm, A. Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539.

Stahl, Y., Wink, R.H., Ingram, G.C., and Simon, R. (2009). A signaling module controlling the stem cell niche in Arabidopsis root meristems. Curr. Biol. *19*, 909-914.

Strabala, T.J., Phillips, L., West, M., and Stanbra, L. (2014). Bioinformatic and phylogenetic analysis of the CLAVATA3/EMBRYO-SURROUNDING REGION (CLE) and the CLE-LIKE signal peptide genes in the Pinophyta. BMC Plant Biol. *14*, 1-16.

Van den Steen, P., Rudd, P.M., Dwek, R.A., and Opdenakker, G. (1998). Concepts and principles of O-linked glycosylation. Crit. Rev. Biochem. Mol. Biol. *33*, 151-208.

Wei, Z., and Li, J. (2016). Brassinosteroids regulate root growth, development, and symbiosis. Mol. Plant *9*, 86-100.

Xie, N., and Taylor, C.M. (2010). Synthesis of a Dimer of β-(1,4)-L-Arabinosyl-(2S,4R)-4-hydroxyproline Inspired by Art v 1, the Major Allergen of Mugwort. Org. Lett. *12*, 4968-4971.

Xu, C., Liberatore, K.L., MacAlister, C.A., Huang, Z., Chu, Y.-H., Jiang, K., Brooks, C., Ogawa-Ohnishi, M., Xiong, G., Pauly, M., et al. (2015). A cascade of arabinosyltransferases controls shoot meristem size in tomato. Nat. Genet. *47*, 784-792.

Yamaguchi, Y.L., Ishida, T., and Sawa, S. (2016). CLE peptides and their signaling pathways in plant development. J. Exp. Bot. *67*, 4813-4826.

Yin, H., and Lowary, T.L. (2001). Synthesis of arabinofuranosides via low-temperature activation of thioglycosides. Tetrahedron Lett. *42*, 5829-5832.

Zhang, Y., Yang, S., Song, Y., and Wang, J. (2014). Genome-wide characterization, expression and functional analysis of CLV3/ESR gene family in tomato. BMC Genomics *15*, 827.

Zhu, X., Kawatkar, S., Rao, Y., and Boons, G.-J. (2006). Practical Approach for the Stereoselective Introduction of β-Arabinofuranosides. J. Am. Chem. Soc. *128*, 11948-11957.

MAIN FIGURE TITLES AND LEGENDS

Figure 1. Structure of the suitably protected Araf_3Hyp building block target **1** (in box) and triarabinosylated and unglycosylated *Gm*CLE40a **2a** and **2b**.

Figure 2. Synthesis of glycosylamino acid building block **1**. Reaction conditions: (a) Fmoc-Hyp-OBn, DDQ, 4Å mol. Sieves, CH_2Cl_2, rt, 18 h (b) Me_2S_2, MeOTf, TTBP, 4Å mol. sieves, CH_2Cl_2, rt, 12-18 h (c) TFA:$CHCl_3$ 1:9 v/v, rt, 30 min (d) **3**, DDQ, CH_2Cl_2, rt, 18 h (e) Ac_2O, pyridine, rt, 16 h (f) Et_3SiH, 10% Pd/C, MeOH, 2 h.

Figure 3. Synthesis of triarabinosylated *Gm*CLE40a peptide **2a**.

Figure 4. CLE40 orthologues in various species. **A** Multiple sequence alignment of CLE40 prepropeptides showing highest conservation between orthologues in the CLE domain at positions 136 to 148. Outside of the CLE domain, sequence conservation is typically seen between closely related species. **B** Sequence logo diagram representing amino acid conservation in the CLE domain of the CLE40 orthologues. **C.** Phylogenetic tree of the CLE40 orthologues, along with some *At*CLV3 orthologues which group separately, and *At*RGF1 as an outgroup. CLE40 orthologues from monocots group within the same clade as CLV3 orthologues, but on a distinct branch with OsFCP1, which is known to function in the root apical meristem of rice. GmCLE40b has been excluded as it is truncated before the CLE domain.

Figure 5. Expression pattern of pro*GmCLE40a::GUS* in two-week-old soybean hairy roots. **A** and **B** show activation of the *GmCLE40a* promoter in the apical region of the tap root. **C** exemplifies its activity in lateral roots.

Figure 6. Soybean root growth following treatment with 10^{-8} to 10^{-4} M *Gm*CLE40a glycopeptide **2a** and peptide **2b**. The tip of the tap root was treated directly every 12 h for a total of 228 h. **A** Total tap root length after 228 h of treatment. Different letters above the bars represent significant statistical differences (Student's *t* test, $P \leq 0.05$). **B** Total length of the tap root recorded throughout the experiment. Some error bars are not presented as they appear smaller than icons displayed. **C** Fifteen day-old soybean plants following treatment for 228 h. n=9 to 15 plants per treatment.

Figure 7. Rate of tap root growth of soybean plants treated every 12 hours with GmCLE40a variants, including glycopeptide **2a,** peptide **2b** and water control. **A** 10^{-4} M, **B** 10^{-6} M, and **C** 10^{-8} M. Some error bars are not presented as they appear smaller than icons displayed. n=9 to 15 plants per treatment.

STAR METHODS

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Wild type soybean, *Glycine max* [L.] Merr. cv. Bragg, was used in this study. For experiments using pouches, chlorine gas sterilized seeds were germinated for 2 days in Grade 3 sterilized vermiculite and autoclaved Milli-Q® water. Germinated seedlings having a straight radicle of a similar length (2-3 cm) were transplanted to modified CYG germination pouches (Mega International, Newport, MN, USA) (Hayashi et al., 2012). The pouch length was increased as required to prevent roots from reaching the bottom. Pouches were watered with autoclaved Milli-Q® water, making sure to avoid excess water build-up or drying out of the filter paper.

Seedlings grown for genetic transformation were first ethanol sterilized (Ferguson et al., 2014) and germinated in Grade 2 vermiculite for 4 days prior to *A. rhizogenes* stab-inoculation

(Ferguson et al., 2014; Kereszt et al., 2007; Lin et al., 2010). Three days after inoculation, additional vermiculite was added to cover the wound site and this was covered with cellophane wrap to enhance humidity and promote transgenic hairy-root growth. Plants were watered every three days, alternating between water and B&D nutrient solution containing 1mM KNO_3 (Broughton and Dilworth, 1971). Two weeks after inoculation, hairy roots were harvested for histochemical beta-glucuronidase (GUS) staining.

All plants were grown under 16:8 day:night conditions. For peptide feeding, plants were grown at 28°C:25°C respectively in a E-75L1 growth chamber (Percival Scientific, Perry, IA, USA); and for hairy-root transformation, 25°C:22°C respectively in a TPG-1260-TH growth chamber (Thermoline, Wetherill Park, NSW, Australia).

E. coli XL1-Blue was cultured at 37°C overnight LB with 50 μg/ml kanamycin and *Agrobacterium rhizogenes* K599 cultured for genetic transformation of soybean was grown at 28°C on Solid LB medium with 50 μg/ml rifampicin and 100 μg/ml ampicillin as described in Reid et al. (2011).

METHOD DETAILS

General Synthetic and Analytical Procedures: Commercial materials, including solvents were used as received unless otherwise noted. Anhydrous MeOH, DMF and CH_2Cl_2 were obtained from a PURE SOLV^{TM} solvent dispensing unit. Solution-phase reactions were carried out under an atmosphere of dry nitrogen or argon.

Flash column chromatography was performed using 230–400 mesh Kieselgel 60 silica eluting with gradients as specified. Analytical thin layer chromatography (TLC) was performed on commercially prepared silica plates (Merck Kieselgel 60 0.25 mm F254). Compounds were visualized using UV at 254 nm and 5% H_2SO_4 in ethanol charring solution.

572 ^{1}H NMR, ^{13}C NMR, DEPT-135 and 2D NMR spectra were recorded at 300 K using a Bruker
573 DRX500, DRX400 or AVANCE300 spectrometer. Chemical shifts are reported in parts per
574 million (ppm) and are referenced to solvent residual signals: $CDCl_3$ δ 7.26 [^{1}H], and δ 77.16
575 [^{13}C]; and D_2O δ 4.79 [^{1}H]. ^{1}H NMR data is reported as chemical shift, multiplicity, relative
576 integral, coupling constant, and assignment where possible. Signal assignments and
577 regiochemical information were obtained through standard 2D experiments (HSQC, HMBC
578 and phase-sensitive COSY). Glycosylamino acid ^{1}H NMR signal assignments marked with the
579 superscripts ‘ and “ indicate signals corresponding to the central arabinoside (Ara*f*-β1,2-**Ara*f*-**
580 β1,2-Ara*f*-β-Hyp) and terminal arabinoside (**Ara*f*-**β1,2-Ara*f*-β1,2-Ara*f*-β-Hyp), respectively.
581 Unmarked ^{1}H NMR sugar signal assignments refer to the reducing terminal arabinoside (Ara*f*-
582 β1,2-Ara*f*-β1,2-**Ara*f*-**β-Hyp).
583 High resolution ESI+ mass spectra were measured on a Bruker–Daltonics Apex Ultra 7.0T
584 Fourier transform mass spectrometer (FTICR). Infrared (IR) absorption spectra were recorded
585 on a Bruker ALPHA Spectrometer with Attenuated Total Reflection (ATR) capability.
586 Compounds were deposited as films on the ATR plate *via* a CH_2Cl_2 solution. Optical rotations
587 were recorded at ambient temperature (293K) on a Perkin–Elmer 341 polarimeter at 589 nm
588 (sodium D line) with a cell path length of 1 dm, and the concentrations are reported in g/100
589 mL.
590 UPLC chromatograms and low resolution ESI mass spectra were obtained on a Shimadzu
591 NexeraX2 UPLC equipped with a SPD-M30A diode array detector and a LCMS-2020 ESI
592 mass spectrometer operating in positive ion mode.
593 Preparative reverse-phase HPLC was performed using a Waters 600 Multisolvent Delivery
594 System and pump with Waters 486 Tuneable absorbance detector operating at 214 nm.
595 Analytical reverse-phase HPLC was performed on a Waters 2695 separations module equipped
596 with a 2996 DAD detector operating at 214 nm.

Synthesis of thioglycoside donor (3): Thioglycoside donor **3** was prepared in 8 steps from L-arabinose. Please see supplemental synthetic details file for detailed synthetic methods and characterization data for **3** and synthetic intermediates.

General procedure for β-arabinofuranosylation *via* IAD. (a) Mixed acetal formation: A mixture of acceptor Fmoc-Hyp-OBn, **5** or **7** (1.0 equiv.), donor **3** (1-2 equiv.) and activated powdered 4Å molecular sieves (1 g.mmol^{-1} of acceptor) in anhydrous CH_2Cl_2 (20 mL.mmol^{-1} of acceptor, 50 mM) was stirred for 2 h at rt before addition of DDQ (2.0-2.5 equiv.) in a single portion. The resulting dark green-blue reaction mixture was stirred at rt under argon for 18 h and then cautiously (CO_2 evolution) treated with *ca.* 5 volume equivalents of sat. aq. $NaHCO_3$ solution. The biphasic mixture was vigorously stirred until complete hydrolysis of DDQ, as indicated by almost complete decolourization of the organic layer (*ca.* 20 min). The mixture was filtered through celite, and the celite pad was washed with additional CH_2Cl_2. The organic layer was separated, and the red aqueous layer was extracted with additional equivalents of CH_2Cl_2. The combined organic extracts were dried over $MgSO_4$, filtered and concentrated under reduced pressure, affording crude mixed acetal as a mixture of diastereoisomers (see Figure S4 for exemplar UPLC data), which was either purified by silica gel column chromatography or used directly in the next step without further purification.

(b) IAD: A mixture of mixed acetal **4**, **6** or **8** (1.0 equiv.), TTBP (4.0 equiv), Me_2S_2 (4.0 equiv.) and activated powdered 4Å molecular sieves (2.5 g.mmol^{-1} of acetal) in anhydrous CH_2Cl_2 (100 mL.mmol^{-1} of acetal, 10 mM) was stirred at rt for 2 h before addition of MeOTf (4.0 equiv.). The mixture was stirred at rt for 12-18 h and then treated with sat. aq. $NaHCO_3$ solution (*ca.* 0.5 volume equiv.) and vigorously stirred for an additional 30 min. After filtering the mixture through celite, the organic layer was removed and concentrated. The residue,

containing mostly naphthylidene adducts (see Figure S4-S6 for exemplar UPLC data), was dissolved in TFA:$CHCl_3$ (1:9 v/v), and stirred at rt for 30 min before co-evaporation with toluene. The residue was purified by silica gel chromatography (eluent as specified), affording β-arabinofuranoside as a white foam.

IAD reactions could be monitored by UPLC-MS using the following procedure. Analytical samples were diluted with MeCN, filtered, and eluted with a linear gradient of 50 to 100% MeCN [0.1% formic acid] in H_2O [0.1% formic acid] over 8 min (Acquity UPLC® BEH C18 1.7 µm, 2.1 x 50 mm, 0.6 mL/min). Compounds were visualized with UV absorbance at 265 nm.

Synthesis of Ara*f*$_3$Hyp Building block (1).

(a) Fmoc-Hyp(Ac$_2$Ara*f*)-OBn (5): Monoarabinoside **5** was synthesized following the general method for β-arabinosylation using 700 mg (1.58 mmol) of acceptor Fmoc-Hyp-OBn, 1.0 equiv. of donor **3**, and 2.0 equiv. of DDQ. The intermediate mixed acetal **4** was used crude for IAD (see Figure S4 for UPLC traces collected during reaction monitoring). Acidolytic workup and purification by silica gel chromatography (eluent: EtOAc:toluene 2:3 →1:1 v/v), afforded **5** as a white foam (588 mg, 56% over 3 steps). **[α]$_D$** +1.4° (c 0.77, $CHCl_3$). **^{1}H NMR** (500 MHz, $CDCl_3$) *ca.* 1:1 mixture of rotational isomers: δ 7.78-7.73 (m, 2H, ArH), 7.59-7.51 (m, 2H, ArH), 7.42-7.36 (m, 2H, ArH), 7.34-7.24 (m, 7H, ArH), 5.22, 5.16, 5.13 (3d, 1.5H, J = 12.3 Hz, 1.5PhCH), 5.08-5.01 (m, 2.5H, H1, H3, 0.5PhCH), 4.55 (t_{apt}, 0.5H, $J_{\alpha,\beta}$ = 7.5 Hz, 0.5Hα), 4.52-4.39 (m, 2.5H, 0.5Hα, Hγ, FmocCH_{2a}), 4.36-4.18 (m, 4.5H, H2, H5$_a$ H5$_b$, FmocCH_{2b}, 0.5FmocCH), 4.11-4.08 (m, 1H, H4), 4.00 (t_{apt}, 0.5H, J = 6.8 Hz, 0.5FmocCH), 3.74-3.69 (m, 1.5H, 0.5Hδ$_a$, Hδ$_b$), 3.57-3.55 (m, 0.5H, 0.5Hδ$_a$), 2.60-2.49 (m, 2H, Hβ$_a$, OH), 2.26-2.17 (m, 1H, Hβ$_b$), 2.12, 2.11 (2s, 3H, CH_3CO), 2.04 (s, 3H, CH_3CO) ppm. **^{13}C NMR** (125 MHz,

$CDCl_3$) *ca.* 1:1 mixture of rotational isomers: δ 172.15, 172.12, 170.9, 170.8, 170.7, 154.8, 154.6, 144.3, 144.2, 144.0, 143.7, 141.51, 141.46, 141.4, 135.6, 135.4, 129.2, 128.7, 128.6, 128.5, 128.4, 128.3, 127.9, 127.8, 127.2, 125.3, 125.2, 125.1, 120.12, 120.08, 100.7, 79.4, 79.3, 78.7, 78.5, 76.3, 76.2, 76.1, 75.4, 67.9, 67.7, 67.3, 67.2, 65.3, 65.2, 58.2, 57.9, 52.0, 51.6, 47.31, 47.27, 37.7, 36.6, 21.0, 20.9 ppm. **FTIR**: 3470, 2951, 1740, 1704, 1451, 1420, 1351, 1230, 1189, 1166, 1126, 1077, 1032, 992, 758, 738, 699 cm^{-1}. **HRMS** (ESI+): calcd. for $C_{36}H_{37}NO_{11}Na$ 682.2259, found 682.2260 (M+Na).

(b) Fmoc-Hyp(Ac$_4$Ara*f*$_2$)-OBn (7): Diarabinoside **7** was synthesized following the general method for β-arabinosylation using 588 mg (0.891 mmol) of acceptor **7**, 1.05 equiv. of donor **3**, and 2.0 equiv. of DDQ. The intermediate mixed acetal **6** was used crude for IAD (see Figure S5 for the UPLC trace collected during reaction monitoring. Acidolytic workup and purification by silica gel chromatography (eluent: EtOAc:toluene 1:1 → 3:2 v/v), afforded **7** as a white foam (507 mg, 65% over 3 steps). **[α]$_D$** +34.6° (c 2.15, $CHCl_3$). **^{1}H NMR** (500 MHz, $CDCl_3$) *ca.* 2:1 mixture of rotational isomers: δ 7.78-7.73 (m, 2H, ArH), 7.59-7.56 (m, 2H, ArH), 7.42-7.21 (m, 9H, ArH), 5.23-5.19 (m, 1.33H, H3, 0.33PhCH), 5.15-5.08 (m, 2.33H, 0.67H1, 1.67PhCH), 5.03 (d, 0.33H, $J_{1,2}$ = 4.3 Hz, 0.33H1), 4.99 (t_{apt}, 0.33H, $J_{2',3'/3',4'}$ = 6.0 Hz, 0.33H3'), 4.93-4.90 (m, 1.33H, 0.67H1', 0.67H3'), 4.85 (d, 0.33H, $J_{1',2'}$ = 4.4 Hz, H1'), 4.61, 4.57 ($2t_{apt}$, 1H, Jα,β = 4.6, Hα), 4.51-4.46 (m, 1.33H, Hγ, 0.33FmocCH_2), 4.45-4.41 (m, 1H, H2), 4.37-4.23 (m, 5H, H5a, H5b, H5a', 1.67 FmocCH_2, 0.33FmocCH), 4.21-4.13 (m, 2H, H2', H5'), 4.11-4.01 (m, 2.67H, H4, H4', 0.67FmocCH), 3.84-3.82 (m, 0.67H, 0.67Hδ_a), 3.67 (dd, 0.33H, $J_{\gamma,\delta}$ = 4.5 Hz, $J_{\delta a,\delta b}$ = 11.5 Hz, 0.33Hδ_b), 3.60-3.56 (m, 1H, 0.33Hδ_a, 0.67Hδ_b), 2.94-2.88 (m, 1H, 2'-OH), 2.55, 2.48 (2m, 1H, Hβ_a), 2.24-2.15 (m, 1H, Hβ_b), 2.12, 2.11, 2.069, 2.065, 2.06, 2.053, 2.049, 1.93 (8s, 12H, 4 x CH_3CO) ppm.

672 **^{13}C NMR** (125 MHz, $CDCl_3$) *ca.* 2:1 mixture of rotational isomers: δ 172.4, 172.2,
673 170.7, 170.64, 170.57, 170.5, 170.4, 155.0, 154.9, 144.2, 144.11, 144.08, 144.0, 141.5,
674 141.4, 141.31, 141.28, 135.6, 153.4, 128.6, 128.5, 128.4, 128.2, 127.8, 127.74, 127.71,
675 127.21, 127.17, 127.15, 125.34, 125.27, 125.2, 125.08, 120.1, 119.9, 101.5, 101.4, 98.7,
676 98.3, 80.1, 80.0, 79.62, 79.57, 79.55, 79.4, 78.9, 78.7, 76.34, 76.25, 75.4, 74.2, 68.1,
677 67.8, 67.2, 67.1, 66.0, 65.5, 65.2, 58.0, 57.9, 51.5, 50.8, 47.2, 37.7, 36.5, 20.91, 20.87,
678 20.8, 20.7 ppm. **FTIR**: 3492, 2956, 2918, 2850, 1742, 1706, 1452, 1422, 1368, 1233,
679 1194, 1164, 1125, 1034, 905, 760, 741, 699 cm^{-1}. **HRMS** (ESI+): calcd. for
680 $C_{45}H_{49}NO_{17}Na$ 898.2893, found 898.2889 (M+Na).
681
682 **(c) Fmoc-Hyp(Ac$_6$Ara*f*$_3$)-OBn (9)**: Triarabinoside **9** was synthesized following the
683 general method for β-arabinosylation using 674 mg (0.770 mmol) of acceptor **7**, 2.0
684 equiv. of donor **3**, and 2.5 equiv. of DDQ. The intermediate mixed acetal **8** (786 mg,
685 75%) was obtained after purification by silica gel chromatography (EtOAc:hexane 1:2
686 →1:1 v/v). Subsequent IAD, acidolytic workup, and purification by silica gel
687 chromatography (eluent: EtOAc:toluene 1:1 → 2:1 v/v), afforded **9** as a white foam
688 (533 mg, 84% [63% over 3 steps]). See Figure S6 for the UPLC trace collected during
689 IAD reaction monitoring. **$[\alpha]_D$** +56.1° (c 1.18, $CHCl_3$). **^{1}H NMR** (500 MHz, $CDCl_3$)
690 *ca.* 1:1 mixture of rotational isomers: δ 7.78-7.73 (m, 2H, ArH), 7.60-7.53 (m, 2H,
691 ArH), 7.42-7.22 (m, 9H, ArH), 5.22-5.06 (m, 5.5H, H1, H3, 0.5H1', H3', $PhCH_2$), 5.02-
692 4.97 (m, 2H, 0.5H1', 0.5H", H3"), 4.94 (d, 0.5H, $J_{1",2"}$ = 4.5 Hz, 0.5H1"), 4.59, 4.55
693 (2dd, 1H, $J_{\alpha,\beta}$ = 6.7, 8.3 Hz, Hα), 4.52 -4.16 (m, 11.5H, Hγ, H2, H2', H2", 2H5, 2H5',
694 $H5_a$", $FmocCH_2$, 0.5FmocCH), 4.12-4.06 (m, 3H, H4, H4', $H5_b$"), 4.03-3.99 (m, 1H,
695 0.5H4", 0.5FmocCH), 3.87 (m, 0.5H, 0.5H4"), 3.73-3.65 (m, 1.5H, 0.5Hδ_a, Hδ_b), 3.54
696 (m, 0.5H, 0.5Hδ), 2.59, 2.51 (2m, 1H, Hβ_a), 2.25, 2.17 (2m, 1H, Hβ_b), 2.12, 2.10, 2.09,

2.08, 2.055, 2.049, 2.01, 1.982, 1.976, 1.94 (10s, 18H, 6 x CH_3CO) ppm. **^{13}C NMR** (125 MHz, $CDCl_3$) *ca.* 1:1 mixture of rotational isomers: δ 172.3, 172.2, 170.81, 170.76, 170.73, 170.65, 170.6, 170.48, 170.46, 170.38, 170.37, 154.8, 154.7, 144.3, 144.2, 144.0, 143.7, 141.5, 141.41, 141.40, 141.3, 135.6, 135.3, 128.7, 128.6, 128.5, 128.4, 128.2, 127.89, 127.87, 127.85, 127.7, 127.24, 127.21, 125.4, 125.18, 125.15, 125.08, 120.1, 120.0, 120.1, 101.0, 100.9, 99.6, 99.1, 98.5, 98.3, 80.1, 79.82, 79.78, 79.5, 79.3, 78.4, 78.0, 76.6, 76.3, 76.2, 75.3, 74.9, 68.1, 67.8, 67.23, 67.21, 66.3, 66.1, 65.8, 65.7, 65.6, 65.2, 58.1, 57.8, 51.9, 51.1, 47.3, 37.5, 36.5, 20.94, 20.91, 20.89, 20.88, 20.86, 20.79, 20.75, 20.74, 20.71 ppm. **FTIR**: 3514, 2955, 2922, 1739, 1707, 1451, 1421, 1367, 1227, 1164, 1124, 1033, 905, 759, 738, 700 cm^{-1}. **HRMS** (ESI+): calcd. for $C_{54}H_{61}NO_{23}Na$ 1114.3527, found 1114.3524 (M+Na).

(d) Fmoc-Hyp(Ac$_7$Ara*f*$_3$)-OBn (10): To a solution of free alcohol **9** (513 mg, 0.470 mmol) in pyridine (4 mL) was added Ac_2O (2 mL). The solution was stirred at rt for 16 h and then co-evaporated with toluene. Purification by silica gel chromatography (eluent: EtOAc:hex 1:1 → 3:2 v/v) afforded peracetylated trisaccharide **10** as a white foam (478 mg, 90%). **$[\alpha]_D$** +63.8° (c 1.00, $CHCl_3$). **^{1}H NMR** (500 MHz, $CDCl_3$) *ca.* 1:1 mixture of rotational isomers: δ 7.77-7.72 (m, 2H, ArH), 7.61-7.58 (m, 1.5H, ArH), 7.53 (m, 0.5H, ArH), 7.41-7.20 (m, 9H, ArH), 5.36, 5.28 (2d, 1H, $J_{1'',2''}$ = 4.3 Hz, H1''), 5.26-5.09 (m, 5H, H1/H1', H3/H3', H3'', $PhCH_2$), 5.05-4.99 (m, 2.5H, H1/H1', H3/H3', 0.5H2''), 4.93 (dd, 0.5H, $J_{2'',3''}$ = 7.3 Hz, 0.5H2''), 4.71 (m, 1H, Hα), 4.53-4.22 (m, 9H, H2, H2', $H5_a$, H5a' H5b/H5b', $H5_a$'', Hγ, $FmocCH_{2a}$, 0.5FmocCH, $0.5FmocCH_{2b}$), 4.16-3.94 (m, 5.5H, H4, H4', 0.5H4'', $H5_b$/$H5_b$', $H5_b$'', $0.5FmocCH_{2b}$, 0.5FmocCH), 3.78-3.69 (m, 2H, 0.5H4'', $H\delta_a$, $0.5H\delta_b$), 3.58 (m, 0.5H, $0.5H\delta_b$), 2.80, 2.63 (2m, 1H, $H\beta_a$), 2.31-2.20 (m, 1H, $H\beta_b$), 2.110, 2.107, 2.09, 2.081, 2.076, 2.063, 2.061, 2.04, 1.95,

1.92, 1.91, 1.82 (12s, 21H, 7 x CH_3CO) ppm. **^{13}C NMR** (125 MHz, $CDCl_3$) *ca.* 1:1 mixture of rotational isomers: δ 172.6, 172.4, 170.81, 170.77, 170.74, 170.65, 170.6, 170.4, 170.3, 169.9, 154.8, 154.6, 144.6, 144.3, 144.0, 143.7, 141.44, 141.39, 141.3, 135.7, 135.4, 128.7, 128.54, 128.50, 128.4, 128.2, 127.9, 127.7, 127.24, 127.17, 125.6, 125.3, 125.20, 125.16, 120.1, 120.04, 120.00, 98.9, 98.6, 97.82, 97.76, 97.7, 97.6, 80.6, 80.5, 79.8, 79.6, 79.2, 79.0, 77.7, 77.30, 77.25, 77.2, 77.1, 76.8, 76.6, 76.5, 76.4, 75.6, 75.4, 75.2, 74.6, 68.0, 67.8, 67.2, 67.1, 66.4, 66.3, 66.2, 65.5, 65.4, 58.2, 57.9, 51.9, 51.4, 47.3, 37.6, 36.4, 20.93, 20.91, 20.78, 20.77, 20.74, 20.70, 20.6, 20.52, 20.48 ppm. **FTIR**: 2955, 2918, 1741, 1708, 1451, 1421, 1369, 1225, 1164, 1120, 1036, 909, 760, 740, 700 cm^{-1}. **HRMS** (ESI+): calcd. for $C_{56}H_{63}NO_{24}Na$ 1156.3632, found 1156.3633 (M+Na).

(e) Fmoc-Hyp(Ac$_7$Ara*f*$_3$)-OH (1): Benzyl ester deprotection was accomplished using a chemoselective transfer hydrogenation method reported by Mandal and McMurray (2007). Specifically, Et_3SiH (508 µL, 3.18 mmol) was added dropwise to a stirred suspension of benzyl ester **10** (361 mg, 318 mmol) and 10% Pd/C (36 mg) in MeOH (5 mL) in a vessel equipped with an argon balloon. After effervescence had ceased, the reaction was allowed to stir for an additional 1 h at rt, before filtering through celite. The filtrate was evaporated to dryness and purified by silica gel chromatography (eluent: MeOH:AcOH:CH_2Cl_2 0:1:99 → 9:1:90 v/v/v, slow gradient) affording glycosylamino acid building block **1** as a white foam (285 mg, 86%). $[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\boldsymbol{20}}$ +65.0° (c 1.00, $CHCl_3$), lit:[2] $[\boldsymbol{\alpha}]_{\boldsymbol{D}}^{\boldsymbol{26}}$ +75.2°. **^{1}H NMR** (500 MHz, $CDCl_3$) *ca.* 1:1 mixture of rotational isomers: δ 7.75 (d, 1H, J = 7.7 Hz, ArH), 7.64 (t_{apt}, 1H, J = 8.0 Hz, ArH), 7.60-7.57 (m, 1.5H, ArH), 7.52 (d, 0.5H, J = 7.5 Hz, ArH), 7.39 (t_{apt}, 1H, J = 7.4 Hz, ArH), 7.34-7.23 (m, 3H, ArH), 5.33 (d, 0.5H, $J_{1",2"}$ = 4.3 Hz, 0.5H1"), 5.26-5.24 (m,

1H, 0.5H1", 0.5H3"), 5.19-5.16 (m, 1.5H, 0.5H1/H1', 0.5H3/H3', 0.5H3"), 5.14-5.11 (m, 1H, 0.5H1/H1', 0.5H3/H3'), 5.04 (dd, 0.5H, J = 4.5, 5.9 Hz, 0.5H3/H3'), 5.02-4.99 (m, 2H, H1/H1', 0.5H3/H3' 0.5H2"), 4.93 (dd, 0.5H, $J_{2",3"}$ = 7.1 Hz, 0.5H2"), 4.69, 4.64 ($2t_{apt}$, 1H, $J_{\alpha,\beta}$ = 7.5 Hz, Hα), 4.53-4.05 (m, 14H, H2, H2', H4, H4', 0.5H4", 2H5, 2H5', $H5_a$", 0.5$H5_b$", Hγ, FmocCH, FmocCH2), 3.97 (dd, 0.5H, $J_{4",5b"}$ = 8.9 Hz, $J_{5a",5b"}$ = 11.4 Hz, 0.5$H5_b$"), 3.75 (m, 0.5H, 0.5Hδa), 3.71 (m, 0.5H, 0.5H4"), 3.67-3.62 (m, 1H, Hδb), 3.57 (m, 0.5H, 0.5Hδa), 2.80, 2.61 (2m, 1H, Hβa), 2.37-2.27 (m, 1H, Hβb), 2.102, 2.096, 2.09, 2.07, 2.06, 2.04, 1.96, 1.94, 1.90, 1.80 (10s, 21H, 7 x CH_3CO) ppm. **^{13}C NMR** (125 MHz, $CDCl_3$) *ca.* 1:1 mixture of rotational isomers: δ 177.0, 175.4, 171.0, 170.9, 170.82, 170.76, 170.7, 170.6, 170.5, 170.4, 170.3, 170.0, 169.9, 155.6, 154.7, 144.4, 144.0, 143.9, 143.7, 141.42, 141.37, 141.32, 141.26, 127.9, 127.8, 127.6, 127.3, 127.21, 127.15, 125.4, 125.2, 125.10, 125.06, 120.1, 120.02, 119.98, 98.7, 98.6, 97.81, 97.75, 97.7, 97.5, 80.5, 80.4, 79.7, 79.6, 79.2, 79.0, 77.7, 77.29, 77.27, 77.2, 77.1, 76.8, 76.6, 76.4, 75.6, 75.4, 75.0, 74.5, 68.1, 68.0, 66.34, 66.30, 66.2, 65.6, 65.4, 58.0, 57.5, 51.8, 51.4, 47.3, 47.2, 37.6, 36.1, 20.93, 20.91, 20.75, 20.65, 20.51, 20.48 ppm. **FTIR**: 2955, 2928, 1423, 1368, 1220, 1164, 1120, 1032, 994, 908, 761, 736, 702, 603, 545, 427 cm^{-1}. **HRMS** (ESI+): calcd. for $C_{49}H_{57}NO_{24}Na$ 1066.3163, found 1066.3164 (M+Na). The data is in agreement with that reported by Kaeothip et. al (2013).

Synthesis of CLE40a Glycopeptide (2a)

Preloading of 2-chlorotrityl chloride resin: 2-Chlorotrityl chloride resin (1.6 mmol/g resin substitution, 2 equiv.) was swollen in dry CH_2Cl_2 for 30 min then washed with CH_2Cl_2 (5 × 2 mL) and DMF (5 × 2 mL). A solution of Fmoc-Asn(Trt)-OH (1 equiv.) and *i*Pr_2NEt (2 equiv.) in 1:1 v/v DMF:CH_2Cl_2 (10 $mL.mmol^{-1}$ of amino acid) was added and the resin and shaken at rt for 16 h. After filtering, the resin was washed with DMF (5 × 2 mL) and CH_2Cl_2

(5 × 2 mL), and then treated with a capping solution of CH_2Cl_2/MeOH/*i*Pr_2NEt (17:2:1 v/v/v, 10 mL.mmol^{-1} of amino acid) for 3 h. The resin was again washed with DMF (5 × 2 mL), CH_2Cl_2 (5 × 2 mL), and DMF (5 × 2 mL) before submitting to iterative peptide assembly (Fmoc-SPPS).

General Fmoc deprotection: The resin was shaken with piperidine:DMF (1:9 v/v, 2 mL, 2 × 3 min) then filtered off and washed with DMF (5 × 2 mL), CH_2Cl_2 (5 × 2 mL) and DMF (5 × 2 mL).**General amino acid coupling:** A solution of Fmoc-AA(PG)-OH (4 equiv.), PyBOP (4 equiv.) and *N*-methylmorpholine (8 equiv.) in DMF (10 mL.mmol of peptide) was added to the resin-bound peptide (1 equiv). The resin was shaken for 45 min, then filtered off and washed with DMF (5 × 3 mL), CH_2Cl_2 (5 × 3 mL) and DMF (5 × 3 mL). Any amino acid directly following a hydroxyproline or glycosylhydroxyproline was double-coupled using 10 equiv. of Fmoc-AA(PG)-OH, 10 equiv PyBOP and 20 equiv. *N*-methylmorpholine for each coupling. **Glycosylamino acid coupling:** A solution of glycosylamino acid **1** (1.2 equiv.), HATU (1.2 equiv.), HOAt (1.5 equiv.) and *i*Pr_2NEt (2.4 equiv.) in DMF (10 mL.mmol^{-1} of peptide) was added to the resin-bound peptide (1.0 equiv.) and shaken for 12 h. The resin was filtered off and washed with DMF (5 × 2 mL), CH_2Cl_2 (5 × 3 mL), and DMF (5 × 2 mL).**Capping:** Following coupling of an Fmoc-AA(PG)-OH or glycosylamino acid **1**, the resin was treated with acetic anhydride/pyridine (1:9 v/v, 2 mL) and shaken for 3 min. The resin was filtered off and washed with DMF (5 × 2 mL), CH_2Cl_2 (5 × 2 mL) and DMF (5 × 2 mL).**Cleavage:** After washing thoroughly with CH_2Cl_2 (7 × 2 mL), the resin was suspended in a mixture of TFA, triisopropylsilane and water (90:5:5 v/v/v, 40 mL.mmol^{-1} of peptide). The suspension was shaken for 2.5 h and then filtered. The resin was washed with additional TFA, and the combined filtrates were concentrated under a stream of nitrogen gas. The residue was dispersed in toluene with the aid of sonication and then evaporated to dryness on a rotary evaporator. The residue was dried for 16 h under high vacuum before deacetylation.

Deacetylation and purification: The cleaved glycopeptide residue was dissolved in anhydrous MeOH (160 mL.mmol of peptide) under argon, and adjusted to ~pH 10 (wet universal indicator paper) with 0.5 M NaOMe in MeOH (~4-8 mL.mmol^{-1} of peptide). The solution was stirred until deacetylation was complete (~30 min as judged by LC-MS) and then neutralized with a drop of formic acid. The reaction mixture was concentrated under a stream of nitrogen, and the resulting solid residue re-suspended in H_2O, filtered and purified by preparative reverse phase HPLC (Waters X-bridge BEH300 C18 5 μm, 19 × 150 mm, 7 mL min^{-1}, 0 to 25% MeCN [0.1%TFA] in H_2O [0.1% TFA] over 45 min, rt ~ 33 min).

Glycopeptide **2a** was prepared on a 12.5 μmol scale according to the general procedures outlined above. After preparative HPLC and lyophilization, glycopeptide **2a** was obtained as a fluffy white solid as the *tetrakis*(trifluoroacetate) salt (6.39 mg, 22%). **Analytical HPLC**: R_t 20.2 min (0 to 30% MeCN [0.1% TFA] in H_2O [0.1% TFA] over 30 min, sample dissolved in H_2O, λ = 214 nm). **LRMS**: m/z 1898 $[M+H]^+$, 949 $[M+2H]^{2+}$, 633 $[M+3H]^{3+}$, 589 [M+3H-Ara*f*]$^{3+}$, 545 [m+3H-2Ara*f*]$^{3+}$, 501 [M+3H-3Ara*f*]$^{3+}$, 475 $[M+4H]^{4+}$, 442 [m+4H-Ara*f*]$^{4+}$, 409 [M+4H-2Ara*f*]$^{4+}$, 376 [M+4H-3Ara*f*]$^{4+}$. **HRMS**: calcd. for $C_{78}H_{121}N_{21}O_{34}$ 1896.8458 $[M+H]^+$, found 1896.8453. See supplemental data file for analytical HPLC trace and low resolution ESI-MS spectrum.

Synthesis of unglycosylated CLE40a peptide (2b)

SPPS: Automated Fmoc-SPPS was carried out on a Biotage Initiator$^+$ Alstra microwave peptide synthesizer equipped with an inert gas manifold. General synthetic protocols for Fmoc-deprotection and capping were carried out in accordance with the manufacturer's specifications. Standardized amino acid couplings were performed for 20 min at 50 °C under

microwave irradiation in the presence of amino acid (0.3 M in DMF), Oxyma (0.5 M in DMF) and DIC (0.5 M in DMF).

Cleavage and purification: After washing thoroughly with CH_2Cl_2 (7 × 2 mL), the resin was suspended in a mixture of TFA, triisopropylsilane and water (90:5:5 v/v/v, 40 mL.mmol^{-1} of peptide). The suspension was shaken for 2.5 h and then filtered. The resin was washed with additional TFA, and the combined filtrates were concentrated under a stream of nitrogen gas. The residue was suspended in Et_2O and centrifuged. After pouring off the supernatant, the pellet was dissolved in H_2O, filtered and purified by preparative reverse phase HPLC (Waters Sunfire C18 5 μm, 30 × 150 mm, 40 mL.min^{-1}, 0 to 25% MeCN [0.1%TFA] in H_2O [0.1% TFA] over 30 min).

Peptide **2b** was prepared on a 16 μmol scale according to the general procedures outlined above. After preparative HPLC (rt ~ 16 min) and lyophilization, peptide **2b** was obtained as a fluffy white solid as the *tetrakis*(trifluoroacetate) salt (9.7 mg, 31%). **Analytical HPLC**: R_t 20.5 min (0 to 30% MeCN [0.1% TFA] in H_2O [0.1% TFA] over 30 min, sample dissolved in H_2O, λ = 214 nm). **LRMS**: m/z 1502 $[M+H]^+$, 751 $[M+2H]^{2+}$, 501 $[M+3H]^{3+}$, 376 $[M+4H]^{4+}$. **HRMS**: calcd. for $C_{63}H_{97}N_{21}O_{22}$ 1500.7200 $[M+H]^+$, found 1500.7153. See supplemental data file for analytical HPLC trace and low resolution ESI-MS spectrum.

Peptide feeding and root length analysis

Five day-old seedlings (three days after being transplanted to growth pouches) were treated with either hydroxylated peptide **2b**, arabinosylated peptide **2a**, or autoclaved MilliQ® water. For each treatment, 10 μL was applied every 12 hours directly to the tip of the tap root via small incisions made in the growth pouch. Incisions were subsequently sealed with tape and the tap root length measured. In all treatments n = 9 to 15 plants.

Cloning the *GmCLE40a* promoter region

A 2.5 kb promoter region located directly upstream of *GmCLE40a* (Glyma.12G054900) was cloned into pGEM®-T easy and subsequently ligated immediately adjacent to the *GUS* coding sequence of modified pCAMBIA1305.1 (pCAMBIA1305.1-Δ35s×2; lacking the duplicated CaMV35s promoter sequence) using T4 DNA ligase (Promega). Positive pro*GmCLE40a*::*GUS* constructs were confirmed by colony PCR and sequencing (Australian Genome Research Facility, Brisbane, Qld, Australia) and transformed from their XL1-Blue *E. coli* strains into electrocompetent *Agrobacterium rhizogenes* K599. Successful transformation was confirmed by PCR prior to use. All primers used in this study are provided in the key resources table.

GUS histochemical assay

GUS activity of transgenic hairy roots was assessed using methods modified from Larkin et al. (Larkin et al., 1996) with X-Gluc staining buffer made using 0.3% (v/v) DMSO in place of 0.1% (v/v) Triton X-100. Harvested hairy roots were treated with fixation buffer (0.5% w/v paraformaldehyde in 100 mM sodium phosphate buffer, pH = 7.2) on ice and under vacuum, rinsed five times with 100 mM sodium phosphate buffer (pH = 7), vacuum infiltrated with X-Gluc staining buffer three times and incubated overnight at 37°C. Stained hairy roots were examined using a clearing solution (Lux et al., 2005) under light microscopes (Nikon models: C-PS/Eclipse E600W).

Bioinformatic analysis

The amino acid sequences of AtCLE40(AT2G27250), and GmCLE40a were used to BLAST for potential orthologues across available genome sequences in Phytozome (https://phytozome.jgi.doe.gov/ (Goodstein et al., 2012). The Phylogenetic tree was created using methods described in Hastwell et al. (Hastwell et al., 2015b) with 1000 bootstrap

replications. Geneious Pro v10.0.2 (Kearse et al., 2012) was used to generate the sequence logo of the CLE domain.

NMR Conformational Analysis

Peptide **2b** was dissolved in 300 µL of 20 mM sodium phosphate (pH 6.5) to a concentration of 3.1 mM. D_2O (15 µL) and 4,4-dimethyl-4-silapentane-1-sulfonic acid (to a final concentration of 10 µM) were added. Peptide **2a** was prepared in the same manner to a final concentration of 2.8 mM.

Spectra were recorded on a Bruker Avance III 600-MHz spectrometer equipped with a TCI cryoprobe. DQF-COSY, 2D TOCSY (mixing time = 70 ms) and NOESY (mixing time = 300 ms) spectra were recorded at both 278 and 298 K. Spectra were analyzed using SPARKY 3.11 (UCSF).

QUANTIFICATION AND STATISTICAL ANALYSIS

Data is expressed as mean ± SEM. Statistical differences between treatments were determined using Student's t-test as described by Ferguson et al. (Ferguson et al., 2014), with the exception of the growth rate analyses, which were done using the Repeated Measures ANOVA. The statistical details of experiments can be found in the figure legends and Results. The n = 9 to 15 plants refers to biological replicates in each treatment group, where each biological replicate is an individual plant. All statistical differences were calculated in GraphPad Prism 7.01 (La Jolla California, USA).

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Prof. Richard J. Payne (richard.payne@sydney.edu.au).

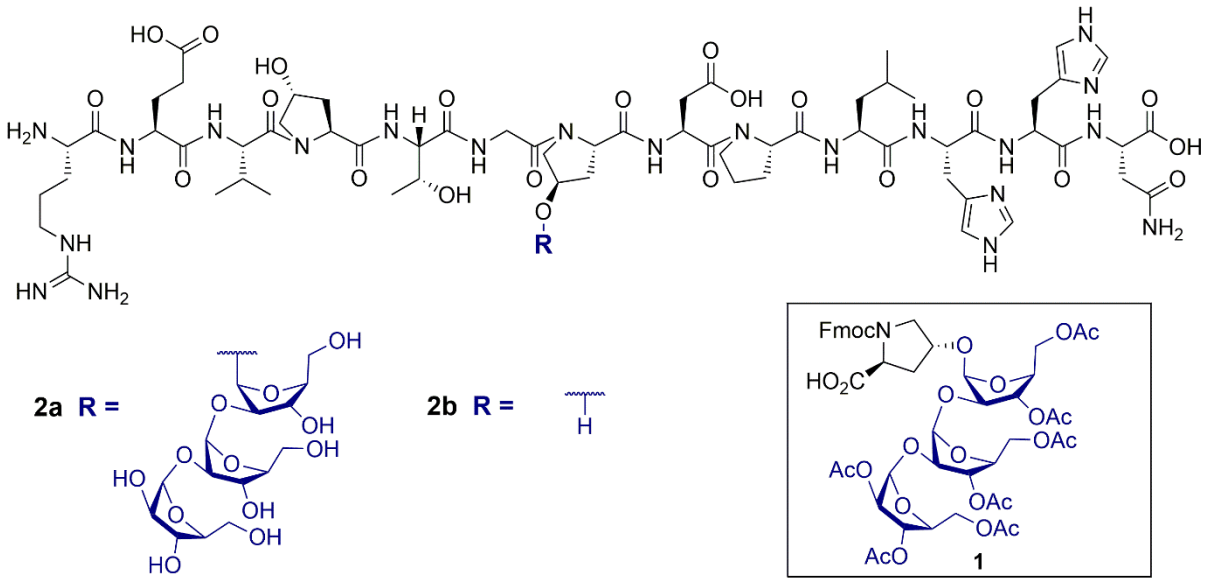


Figure 1

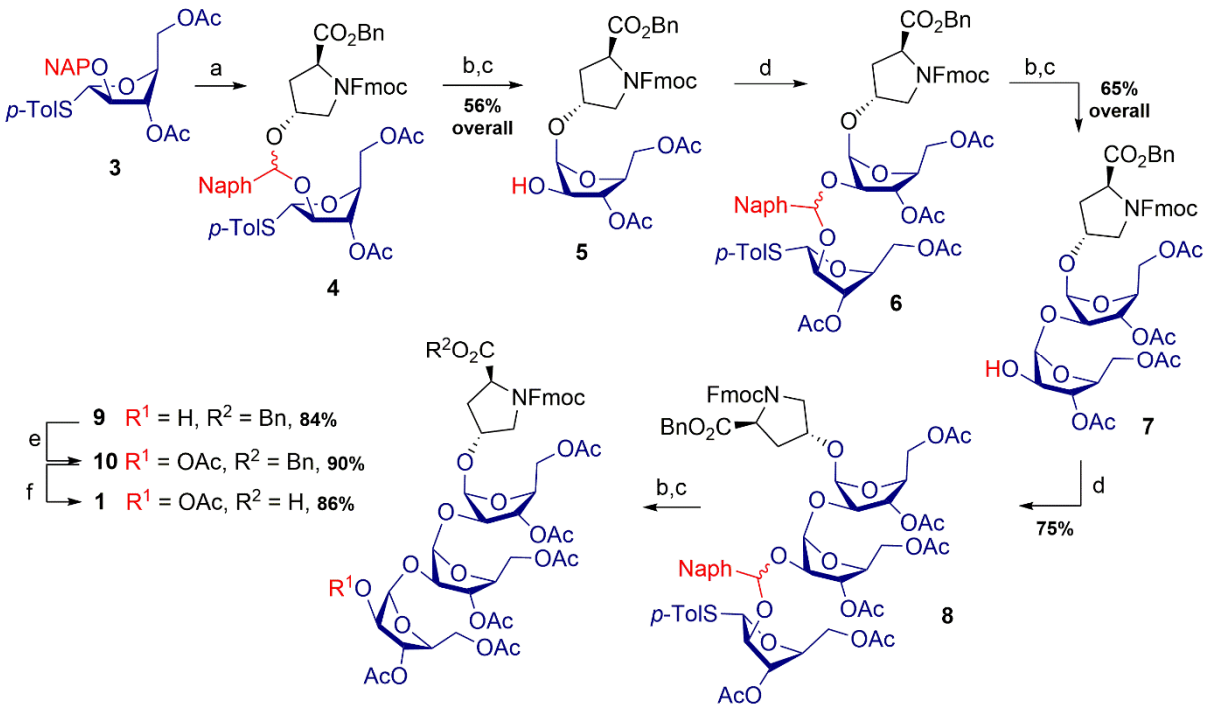
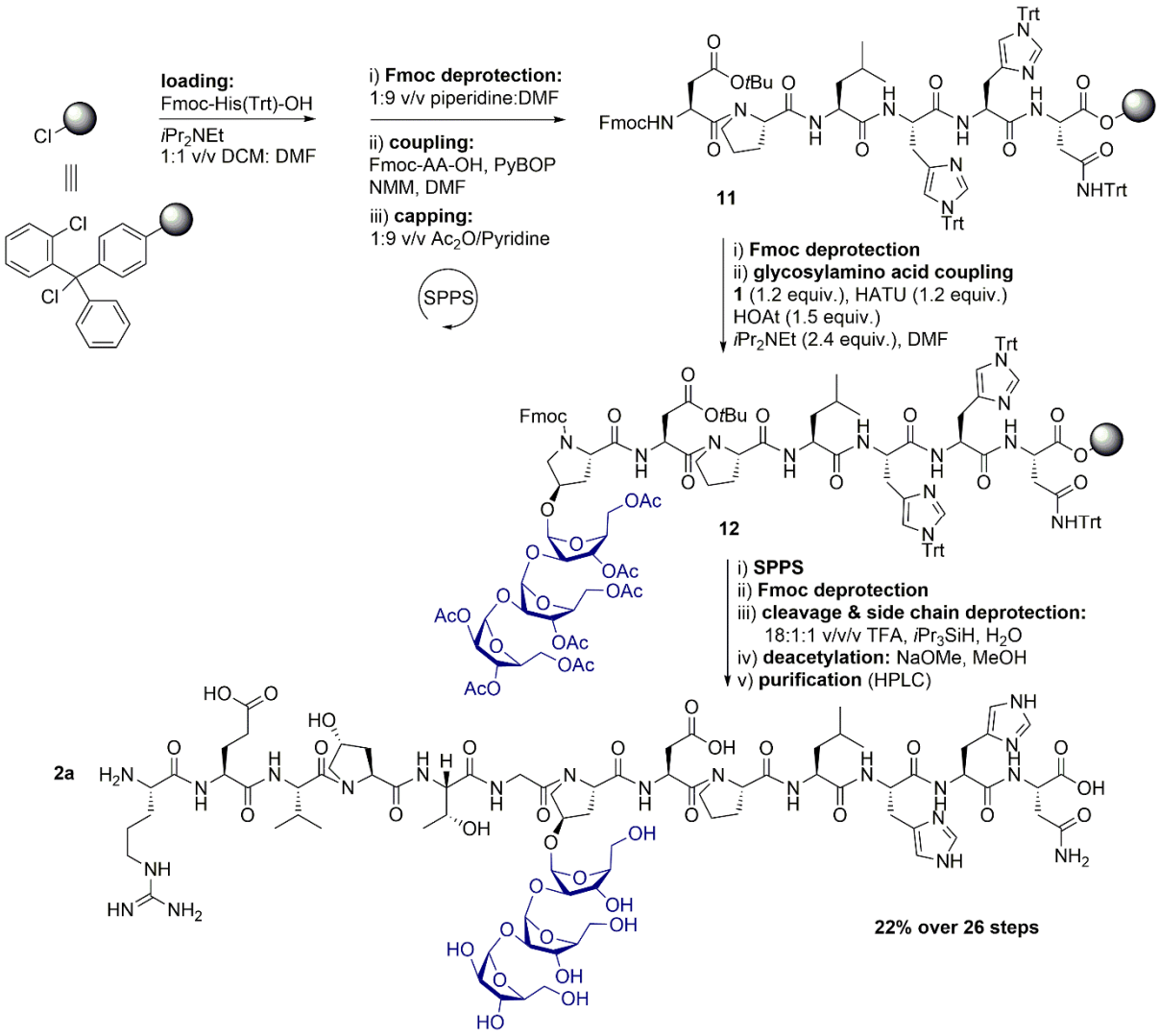
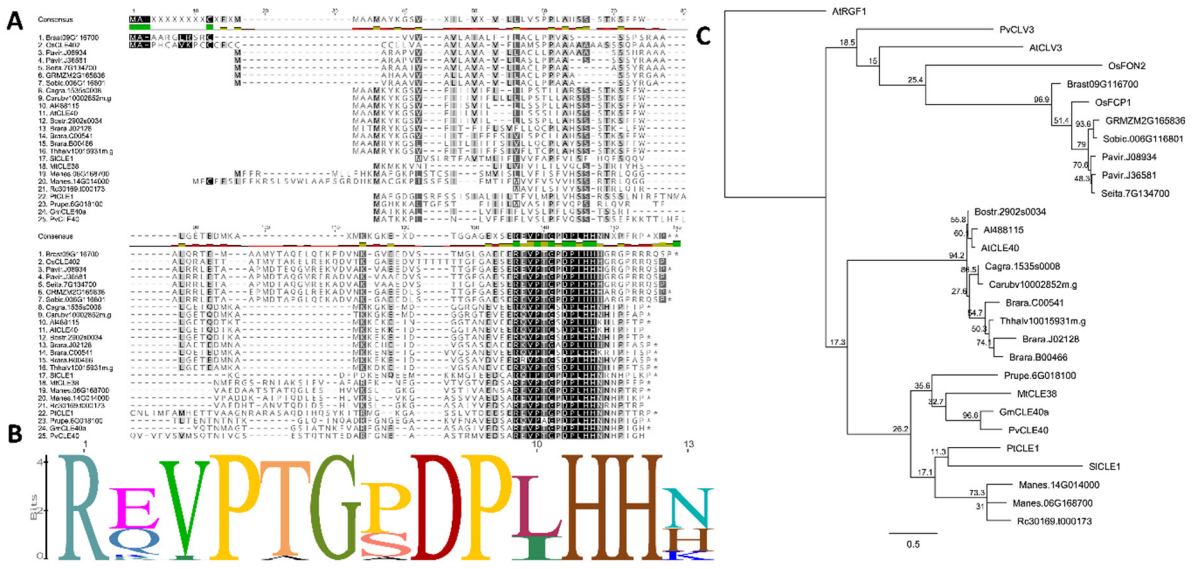


Figure 2



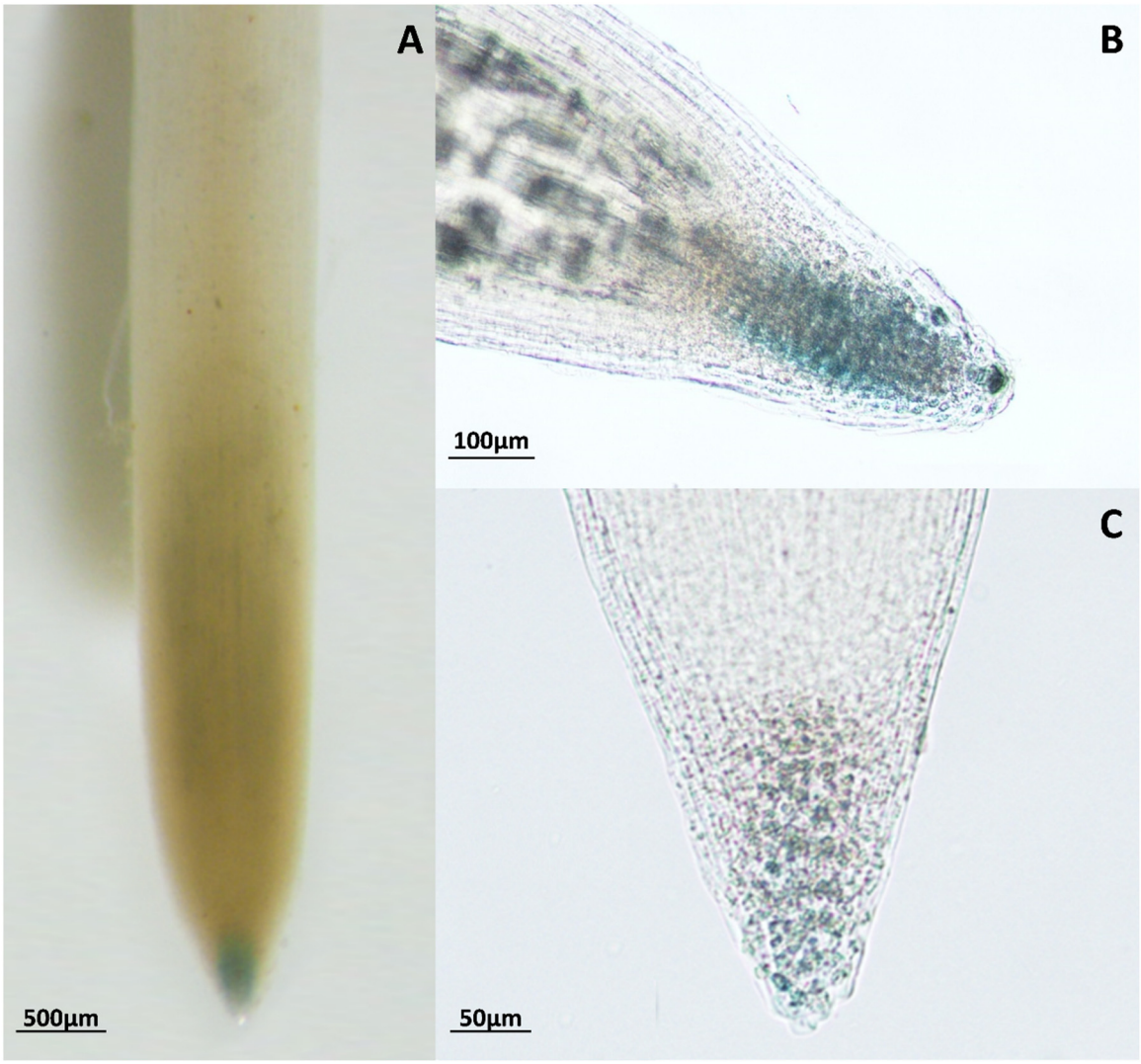
904

905 Figure 3.



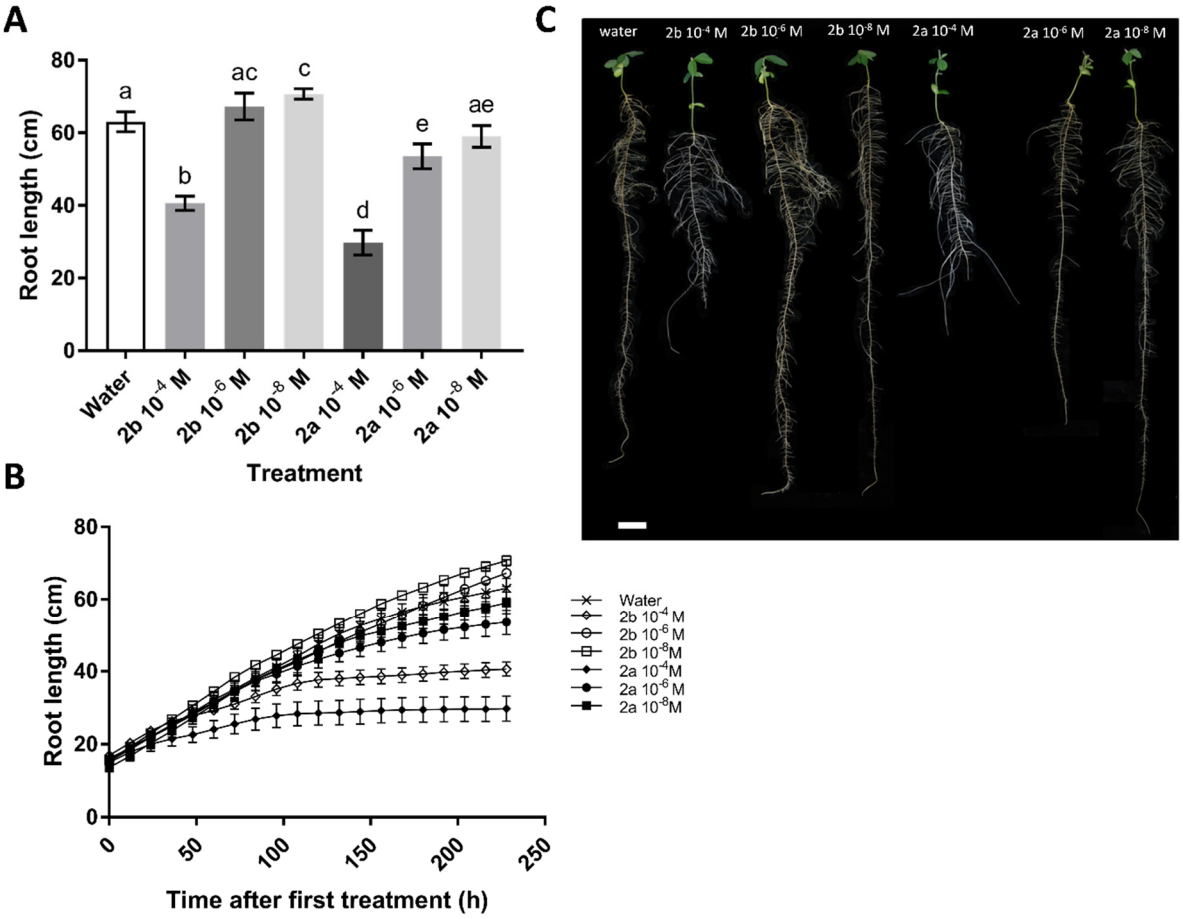
906

907 Figure 4



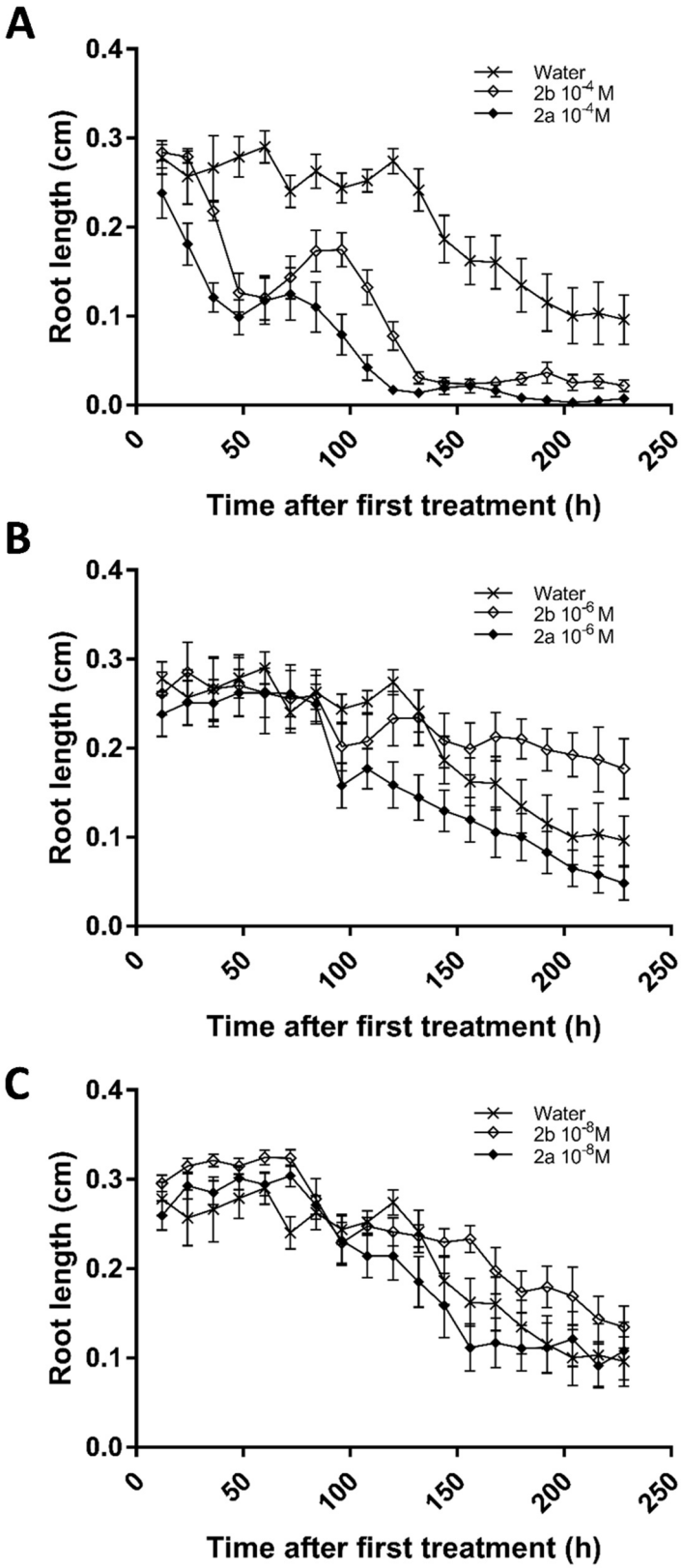
908

909 Figure 5



910

911 Figure 6



912

913 Figure 7.