Supplementary information

An independent evolutionary origin for insect deterrent cucurbitacins in *Iberis amara*

Lemeng Dong[1,6*], Aldo Almeida[1], Jacob Pollier[2,3], Bekzod Khakimov[4], Jean-Etienne Bassard[1], Karel Miettinen[2,3], Dan Stærk[5], Rahimi, Mehran[6], Carl Erik Olsen[1], Mohammed Saddik Motawia[1], Alain Goossens[2,3], and Søren Bak[1*]

[1]Department of Plant and Environmental Science, University of Copenhagen, Thorvaldsensvej 40, DK-1871 Frederiksberg C, Denmark. [2]Ghent University, Department of Plant Biotechnology and Bioinformatics, Technologiepark 71, 9052 Ghent, Belgium. [3]VIB Center for Plant Systems Biology, Technologiepark 71, 9052 Ghent, Belgium. [4]Department of Food Science, University of Copenhagen, Rolighedsvej 26, DK-1958 Frederiksberg C, Denmark. [5]Department of Drug Design and Pharmacology, Faculty of Health and Medical Sciences, University of Copenhagen, Universitetsparken 2, DK-2100 Copenhagen, Denmark. [6]Plant Hormone Biology group, Swammerdam Institute for Life Science, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, the Netherlands

*l.dong2@uva.nl

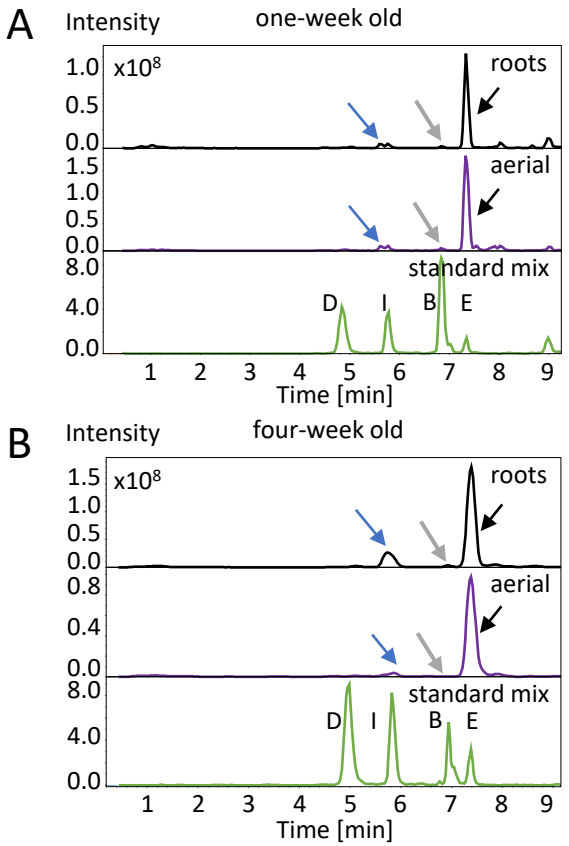
bak@plen.ku.dk

This PDF file includes:

Supplementary Figure 1 to Figure 11,

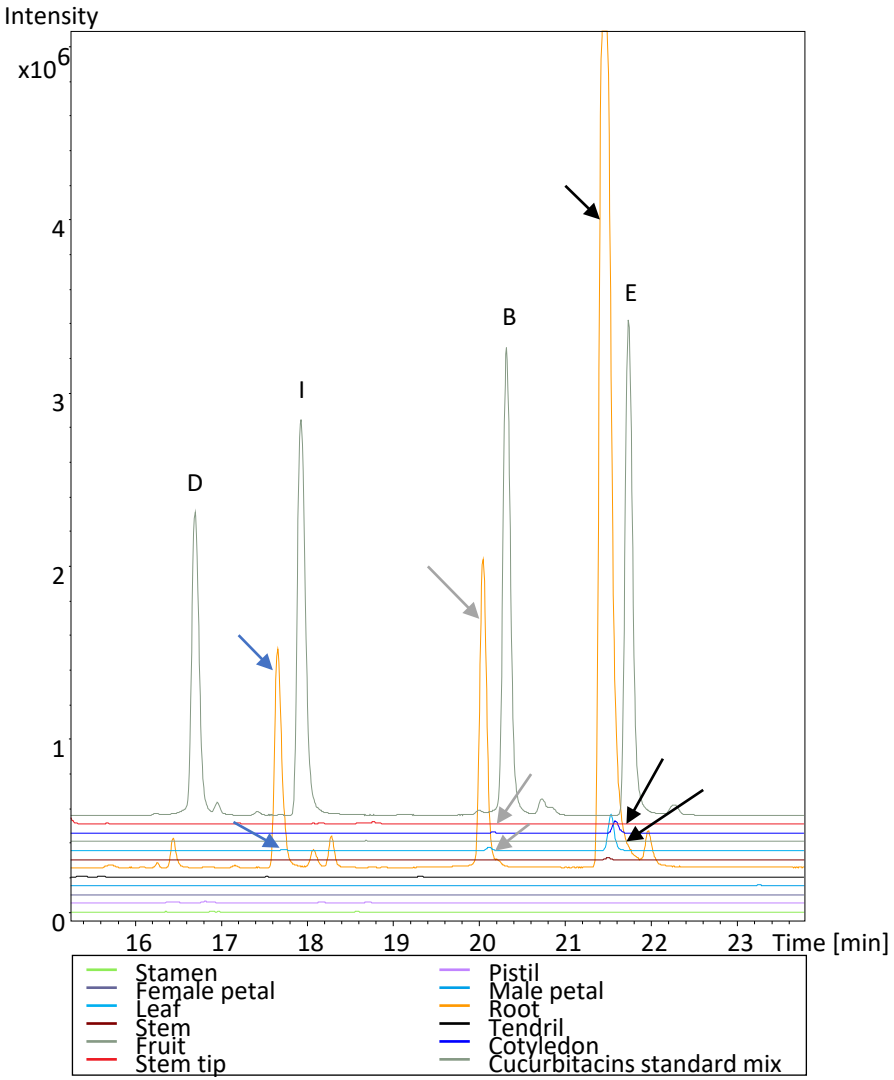
Supplementary Table 1 to Table 6,

Supplementary NMR data (general experimental procedures, data interpretation of compound 1-5, supplementary Figure 12 to Figure 19, Supplementary Table 7 to Table 11), and References

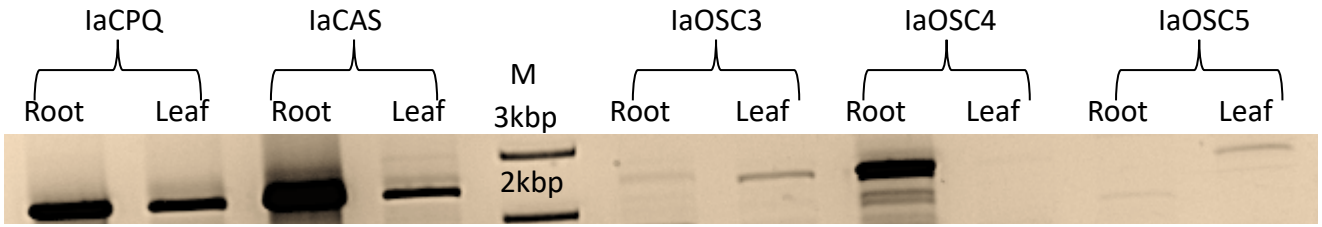


Supplementary Figure 1 Cucurbitacin E, I, and B accumulate in both roots and aerial parts of *I. amara*.

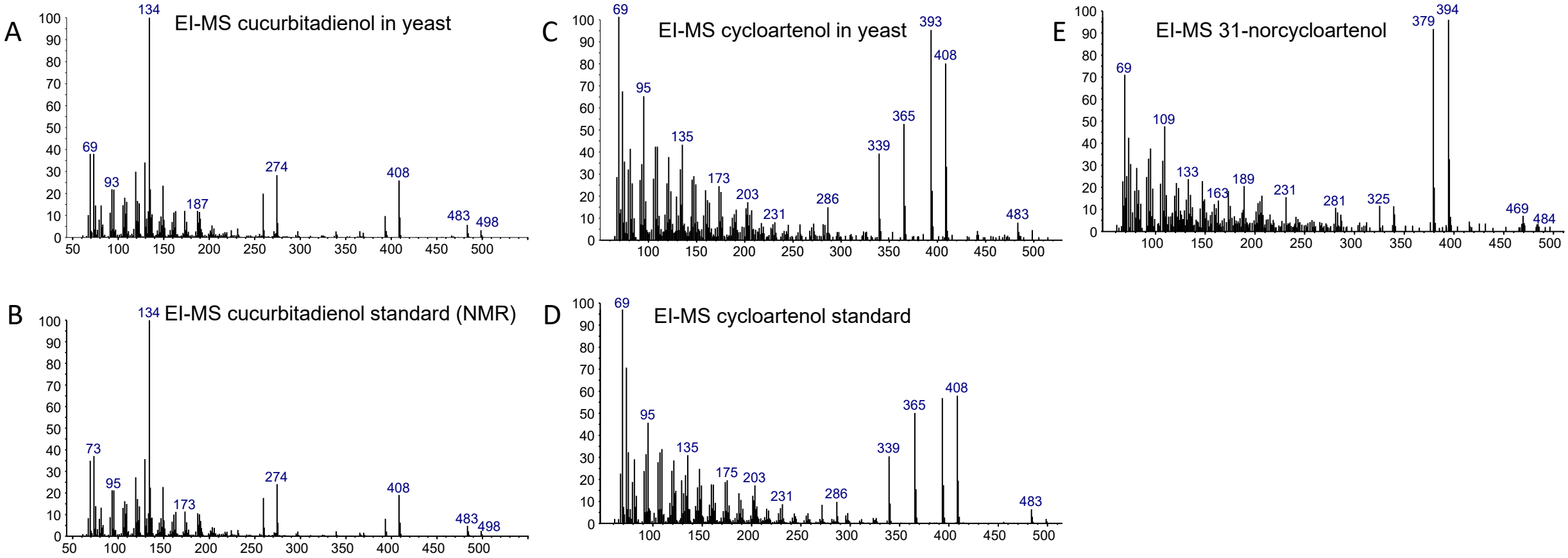
(A) LC-ESI-MS extracted ion chromatograms (EIC) of extracts from roots and aerial parts of one-week-old *I. amara* plants. **(B)** LC-ESI-MS extracted ion chromatograms (EIC) of extracts from roots and aerial parts of four-week-old *I. amara* plants. Extracted ions at *m/z* 539, 537, 581, and 579 are representative for sodium adducts of cucurbitacin D, I, B, E, respectively. Peaks indicated by blue, gray and black arrows verified by the mass spectrum compare to the cucurbitacin standard I, B E, respectively.



Supplementary Figure 2 Cucurbitacins accumulate mainly in roots and leaves in *Citrullus lanatus*. LC-QTOF-MS extracted ion chromatograms (EIC) of extracts from *C. lanatus* stamen, female petals, leaves, stems, fruits, stem tips, pistils, male petals, roots, tendrils, cotyledons. Extracted ions at *m/z* 539, 537, 581, and 579 are representative for sodium adducts of cucurbitacin D, I, B, E, respectively. Peaks indicated by blue, gray and black arrows were verified by comparing their mass spectrum to the corresponding mass spectrum of standards of cucurbitacin I, B E, respectively. Note, the LC-MS analysis program changed for analyzing *C. lanatus* extracts (see materials and methods), giving rise to the observed changes in retention times.



Supplementary Figure 3 Agarose gel image of PCR product of 5 full-length OSCs from *I. amara*. Lane M, 1kb DNA ladder.

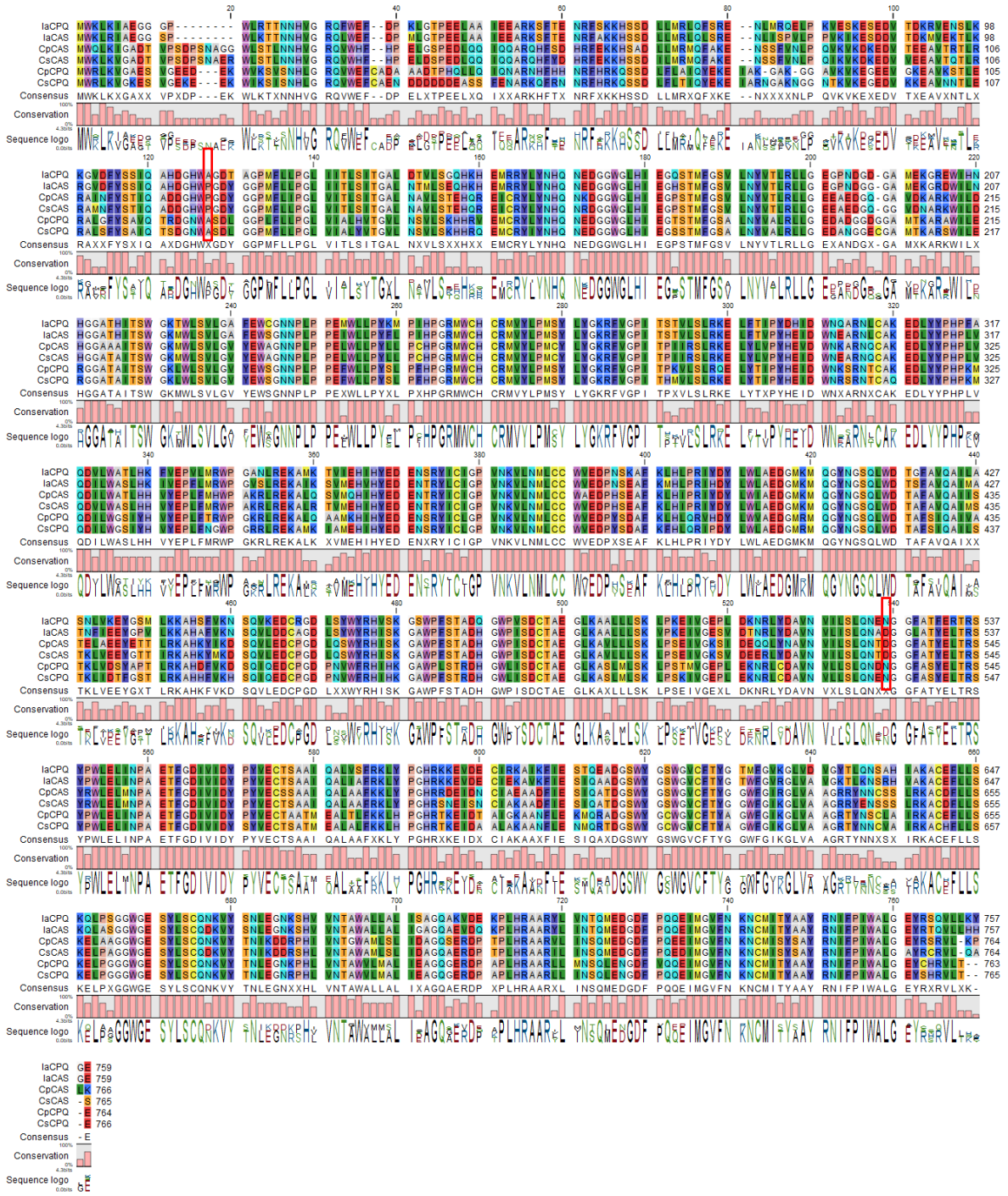


Supplementary Figure 4 EI-MS fragmentation spectrum of trimethylsilylated cycloartenol, 31-norcycloartenol and cucurbitadienol. (A) EI-MS of the cucurbitadienol peak in extracts of spent medium from yeast expressing *IaCPQ*. **(B)** EI-MS of a cucurbitadienol standard confirmed by NMR. **(C)** EI-MS of the cycloartenol peak in extracts of spent medium from yeast expressing with *IaCAS*. **(D)** EI-MS of a cycloartenol standard. **(E)** EI-MS of the 31-norcycloartenol peak in extracts of spent medium from yeast expressing with *IaCAS*.

A

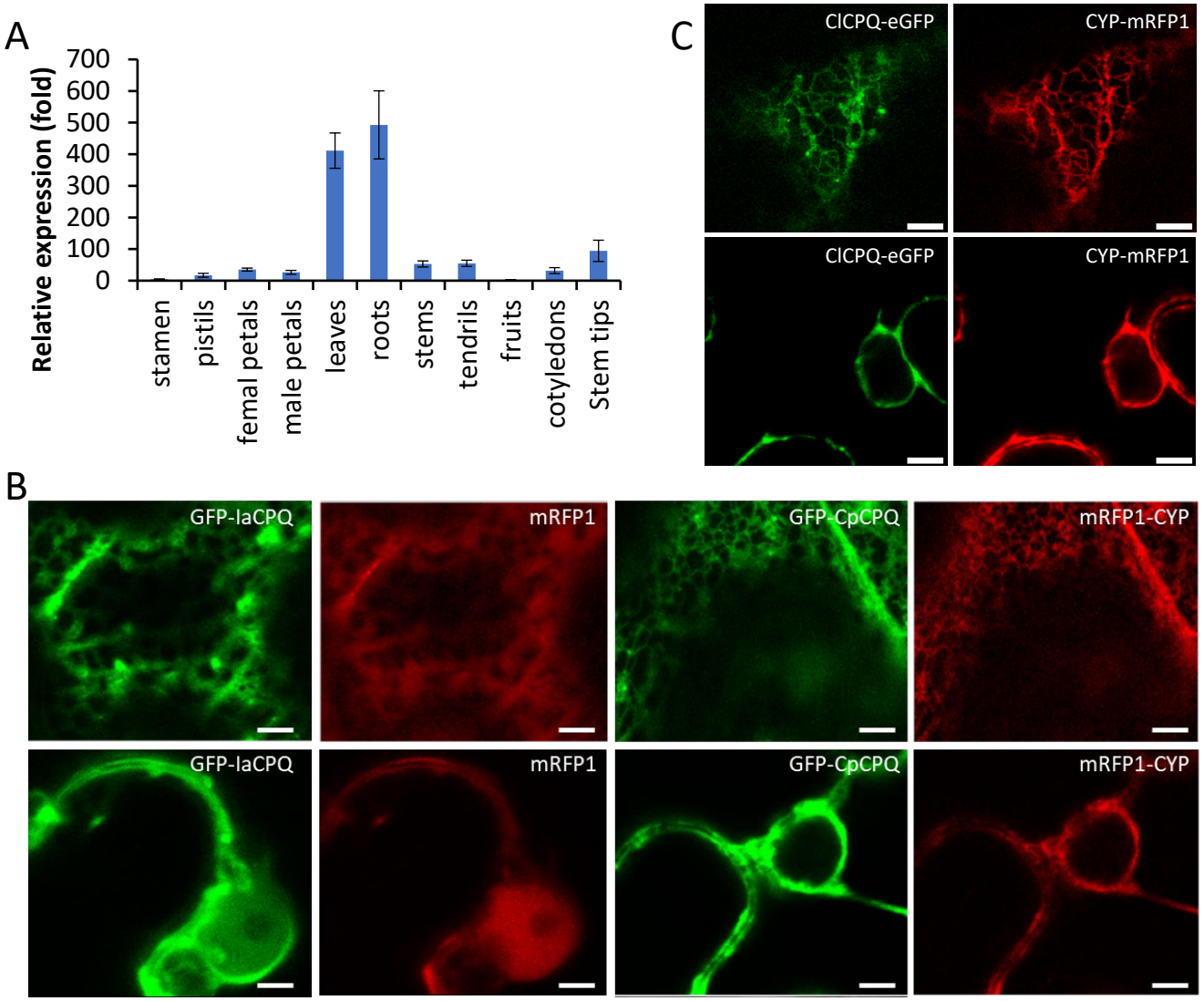
	IaCPQ	IaCAS	CpCPQ	CpCAS	CsCPQ	CsCAS
IaCPQ	100%	86%	68%	70%	68%	72%
IaCAS	86%	100%	68%	73%	68%	74%
CpCPQ	68%	68%	100%	69%	89%	70%
CpCAS	70%	73%	69%	100%	69%	93%
CsCPQ	68%	68%	89%	69%	100%	70%
CsCAS	72%	74%	70%	93%	70%	100%

B

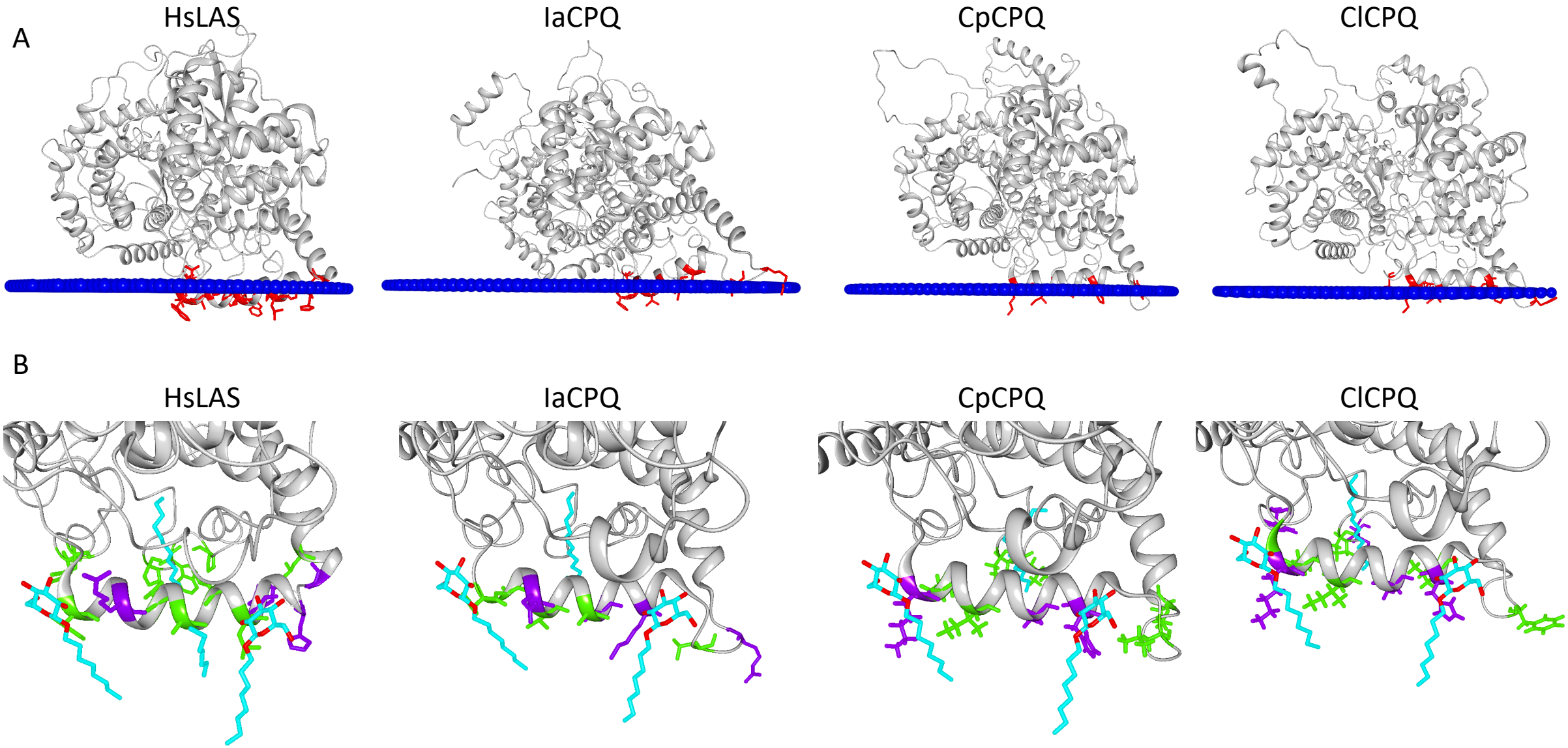


Supplementary Figure 5 Amino acid sequences % identity matrix (A) and alignment (B) between cucurbitadienol synthases (CPQs) and cycloartenol synthases (CASs) from *I. amara*, *C. pepo* and *C. sativus*.

Red boxes indicate the amino acids different between CPQs and CASs.

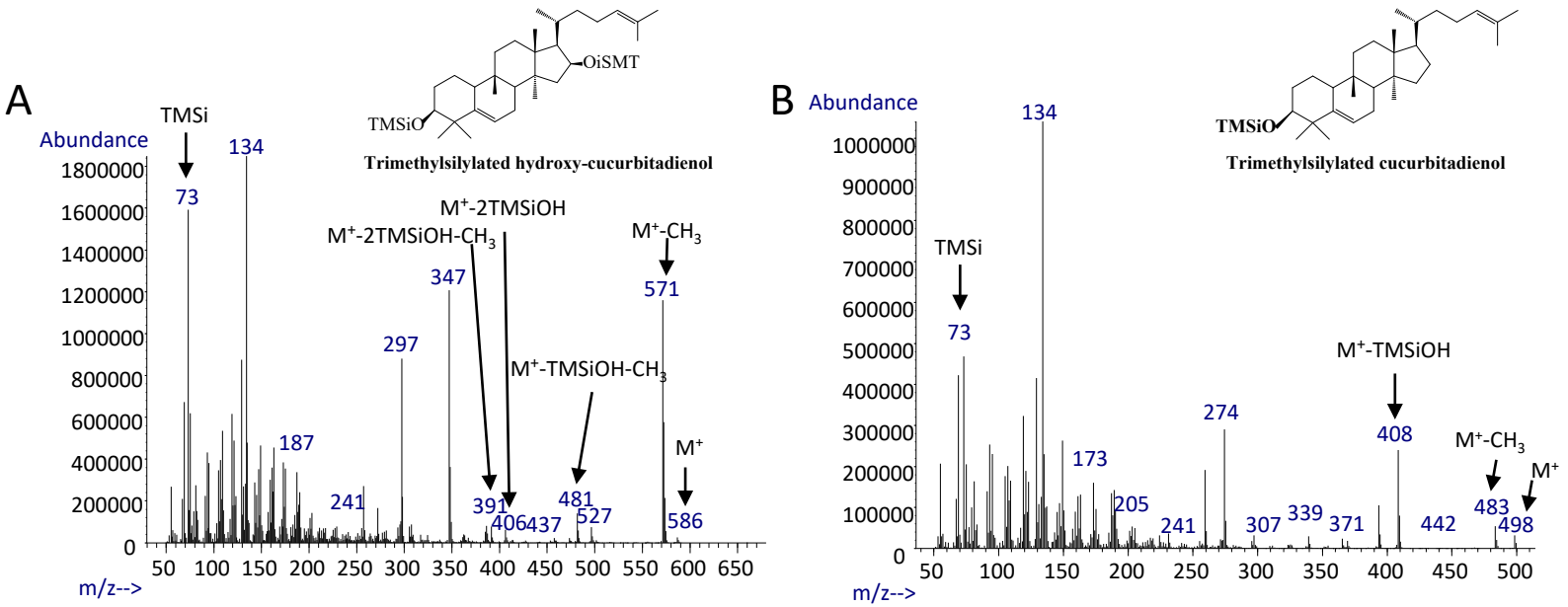


Supplementary Figure 6 Organ-specific expression pattern of *C. lanatus* cucurbitadienol synthase and its subcellular localization to the ER. (A) Relative expression of *ClCPQ* in different *C. lanatus* organs quantified by qPCR. **(B)** Representative confocal images of the epidermal cell space and the nucleus of *N. benthamiana* leaves expressing N-terminally tagged proteins are shown. mRFP1 and CYP98A1 (P450) are used as controls for cytosolic and ER localization, respectively. Soluble mRFP1 is found in the nucleus and fills in gaps between plant cell compartments. ER-localized proteins (CpCPQ and CYP98A1) are restricted to nuclear membranes and a well-defined ER membrane network. Scale bars = 10 mm. **(C)** Representative confocal images of the epidermal cell space and the nucleus of *N. benthamiana* leaves expressing C-terminally tagged proteins are shown. ER-localized CYP98A1 (P450) is used as a control. ClCPQ is restrained to nuclear membranes and the ER membrane network. Scale bars = 10 mm.

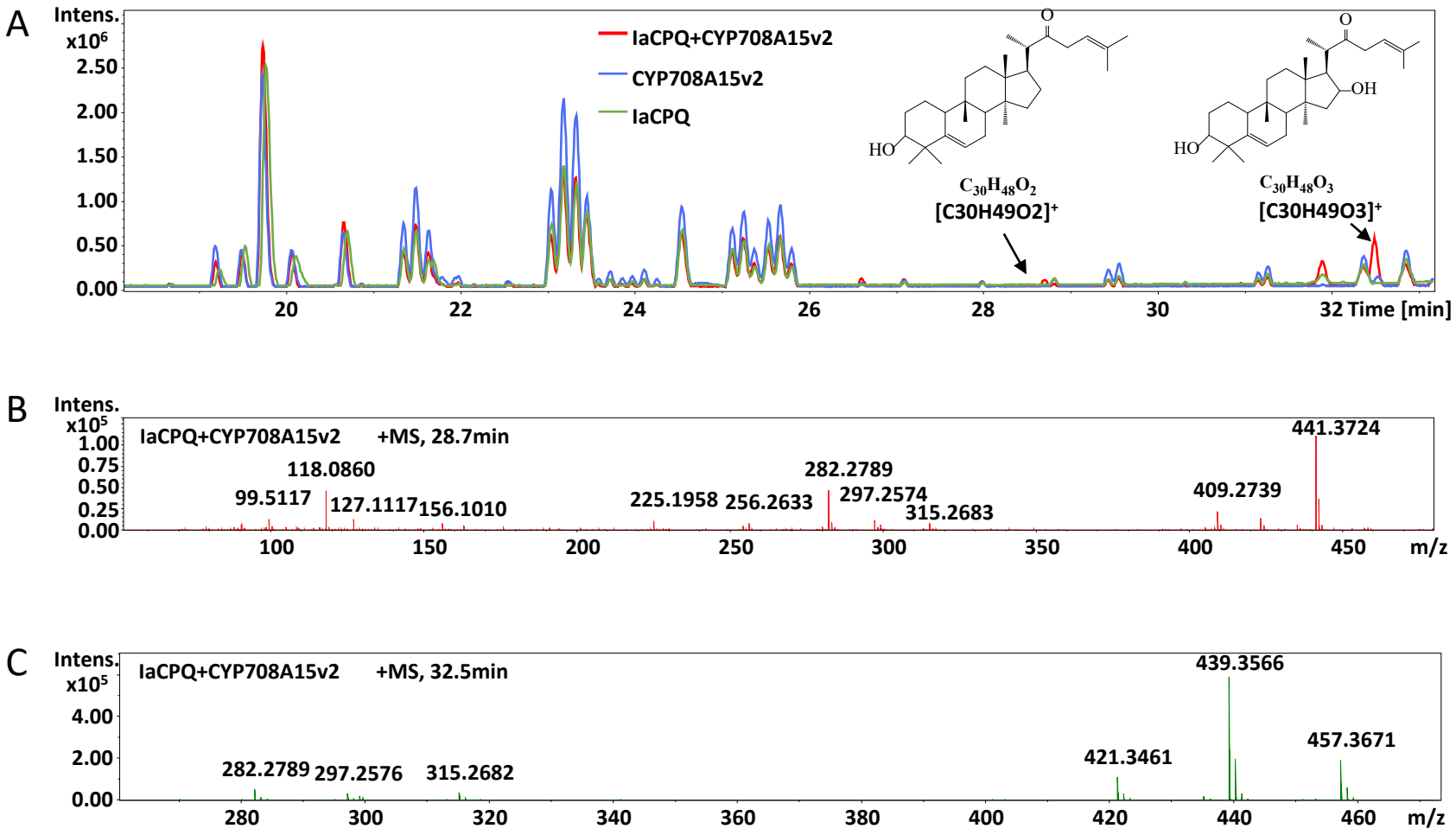


Supplementary Figure 7 Homology modeling of OSCs and membrane interaction analysis using Positioning of Proteins in Membrane server.

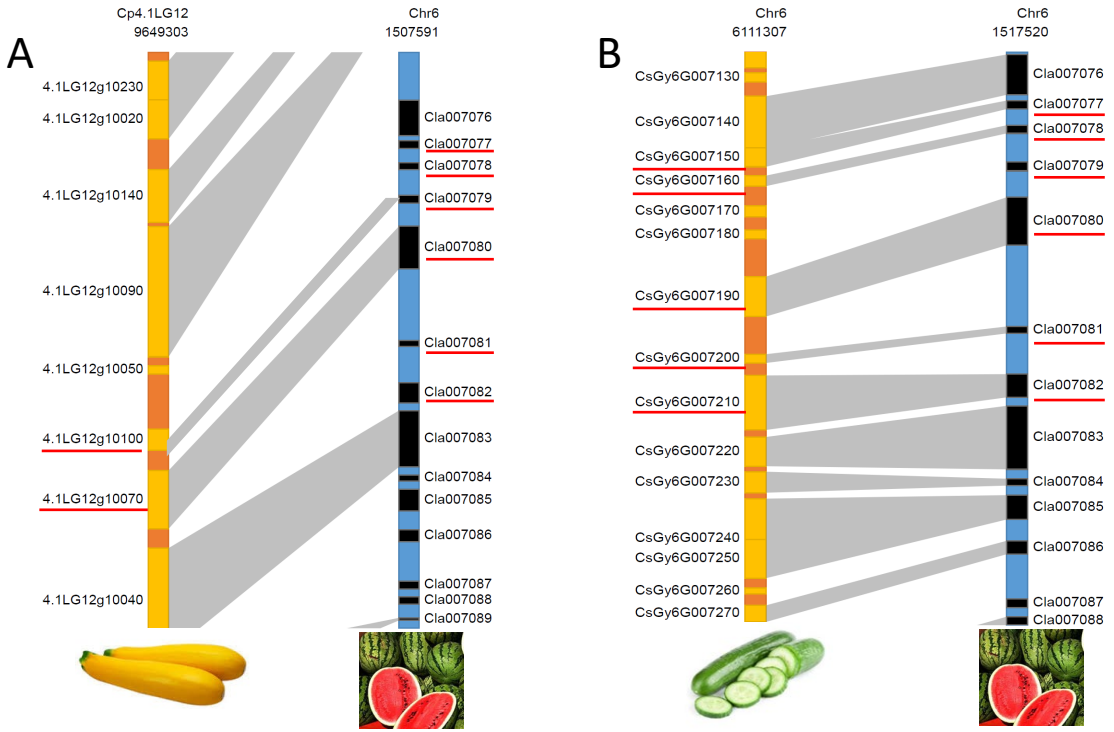
(A) The structure of LAS (PDB:1W6J) and homology models of IaCPQ, CpCPQ, and ClCPQ and the interaction with an empirical membrane (blue). Amino acid resides which interact with the membrane are marked in red. **(B)** The amino acid residues of membrane inserting sites of LAS, IaCPQ, CpCPQ, and ClCPQ are marked in green (hydrophobic) and purple (polar), and interact with fragments of lipids (cyan).



Supplementary Figure 8 EI-MS fragmentation patterns of trimethylsilylated 16β-hydroxy-cucurbitadienol and cucurbitadienol. (A) 16b-hydroxy-cucurbitadienol peak in *N. benthamiana* leaves coexpressing *IaCPQ* and *CYP708A16*. **(B)** Cucurbitadienol peak in *N. benthamiana* leaves expressing *IaCPQ.* M^+, parent ion; TMSi, trimethylsilyl group; TMSiOH, trimethylsilanol. 2TMSiOH, two trimethylsilanol groups.



Supplementary Figure 9 LC-MS chromatograms and spectrums of extracts from *N. benthamiana* leaves agro-infiltrated with vectors expressing *IaCPQ*, *CYP708A15v2*, and *IaCPQ+CYP708A15v2*. (A) LC-ESI-MS total ion chromatograms (TIC) indicating two new peaks in *N. benthamiana* leaves coexpressing *IaCPQ+CYP708A15v2*, and their postulated structures and formulas based on the accurate masses. **(B)** MS spectrum of the first new peak at the retention time 28.7 min. **(C)** MS spectrum of the second new peak at the retention time 32.5 min.

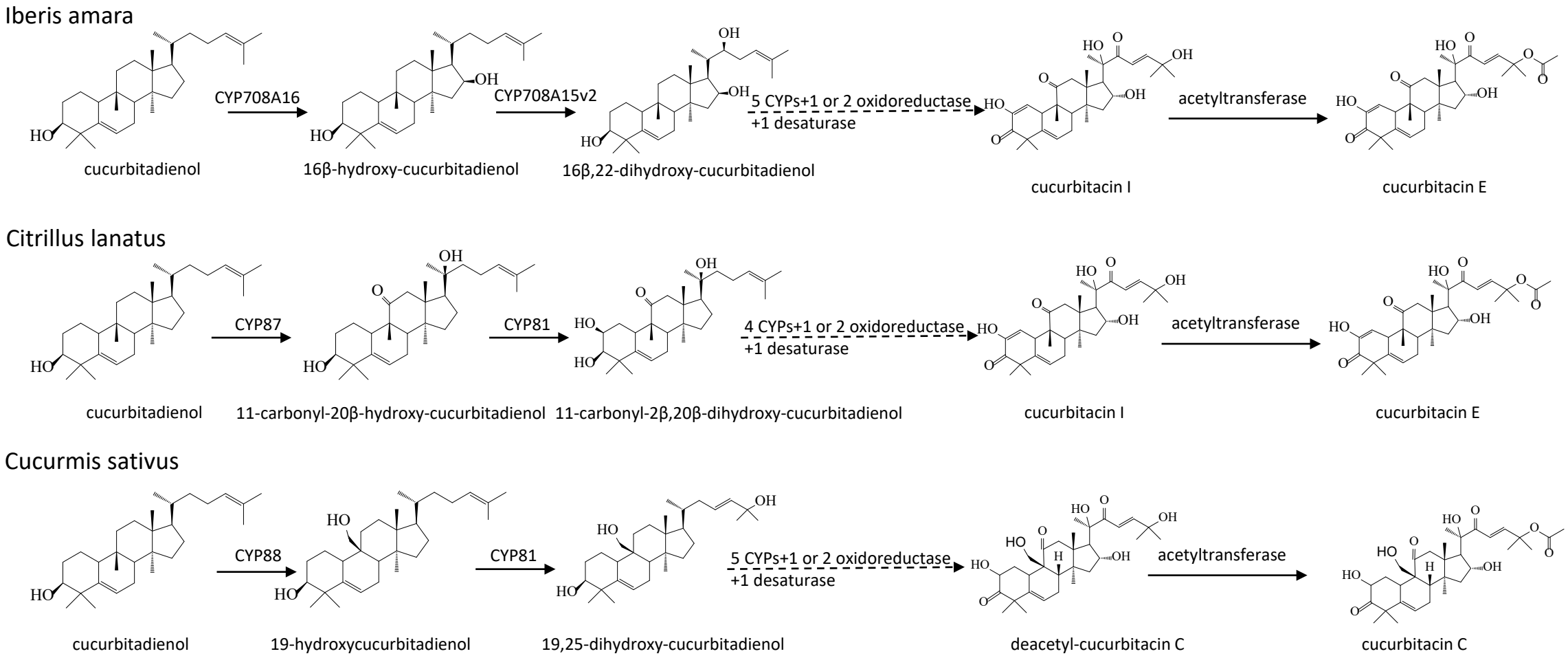


Supplementary Figure 10 Genome localization between cucurbitacin biosynthetic genes in *I. amara* and cucurbitaceous species. (A) Synteny between cucurbitacin biosynthetic gene clusters in *C. pepo* and *C. lanatus.* The proposed cucurbitacin biosynthetic genes are underlined with red lines (according to Zhou *et al.* (2016)).

(B) Synteny between cucurbitacin biosynthetic gene clusters in *C. sativus* and *C. lanatus.* The proposed cucurbitacin biosynthetic genes are underlined with red lines (according to Zhou *et al.* (2016)). **(C)** Genome localization of cucurbitacin biosynthetic genes in *I. amara* and selected the Cucurbitacea species. Data were mined from *I. amara* draft genome and the Cucurbit Genomics Database (CuGenDB).

C

Brassicaceae		*Iberis amara*			Cucurbitaceae		*Cucurbita maxima*		*Cucurbita moschata*		*Cucurbita pepo*		*Lagenaria siceraria*		*Citrullus lanatus*		*Cucurmis melo*		*Cucurmis sativus* (Chinese long)		*Cucurmis sativus* (Gy14)	
Annotation	Function	Gene ID	Position	Scaffold	Annotation	Function	Gene ID	Chr	Gene ID	Chr	Gene ID	Chr	Gene ID	Chr	Gene ID	Chr	Gene ID	Chr	Gene ID	Chr	Gene ID	Chr
CYP708A15	C22-hydroxylation	Ia138495_g9_i2	36455..40316 (-)	scaffold 711	CYP81	C25-hydroxylation	CmaCh08G000360	8	CmoCh08G000370	8	Cp4.1LG17g09460	17	Lsi09G017880	9	Cla007077	6	MELO3C022377	11	Csa6G088160	6	CsGy6G007150	6
CYP708A14	Unknown	Ia570	133487..140620 (-)	scaffold 711	CYP89	Unknown	CmaCh17G013860	17	CmoCh17G013390	17	Unknown gene ID	12	Lsi09G017870	9	Cla007078	6	MELO3C022376	11	Csa6G088170	6	CsGy6G007160	6
Cucurbitadienol synthase	Cucurbitadienol synthase	Ia2462	85361...87211 (+)	scaffold 16365	CYP81	C2-hydroxylation	CmaCh17G013870	17	CmoCh17G013400	17	Cp4.1LG12g10100	12	Lsi09G017850	9	Cla007079	6	MELO3C022375	11	Csa6G088180	6	CsGy6G007170	6
CYP708A16	C16-hydroxylation				Cucurbitadienol synthase	Cucurbitadienol synthase	CmaCh17G013880	17	CmoCh17G013410	17	Cp4.1LG12g10070	12	Lsi09G017840	9	Cla007080	6	MELO3C022374	11	Csa6G088690	6	CsGy6G007190	6
					Acetyltransferase	Acetyltransferase	CmaCh08G000350	8	CmoCh08G000360	8	Cp4.1LG17g09410	17	Lsi09G017810	9	Cla007081	6	MELO3C022373	11	Csa6G088700	6	CsGy6G007200	6
					CYP87	Unknown	CmaCh08G000340	8	CmoCh08G000350	8	Cp4.1LG17g09580	17	Lsi09G017800	9	Cla007082	6	MELO3C022372	11	Csa6G088710	6	CsGy6G007210	6
					CYP87	C11-carbonylation and C20-hydroxylation							Lsi02G002970	2	Cla008354	1						
					CYP87	C11-carbonylation and C20-hydroxylation	CmaCh18G001160	18			Cp4.1LG09g10760	9	Lsi02G002970	2	Cla008355	1	MELO3C002192		Csa1G044890	1	CsGy1G007640	1
					CYP88	C19-hydroxylation										9			Csa3G903540	3	CsGy3G010130	3



Supplementary Figure 11 Proposed biosynthetic routes of cucurbitacin E in *I. amara* and *C. lanatus* and cucurbitacin C in *C. sativus*

Supplementary Table 1 Miseq data assembling from leaves and roots

Contig measurement	Leaf	Root
N75	475	460
N50	850	852
N25	1,490	1,558
Minimum contig length	200	200
Maximum contig length	7,694	10,710
Average contig length	664	654
Count	43,548	60,975
Total number of bases	28,913,534	39,849,046

Supplementary Table 2 Evaluation of one-ratio, free ratio, and two-ratio models using PAML for *OSC* genes

ω, nonsynonymous/synonymous substitution rate ratios (dN/dS).

Model	Log likelihood	Number of parameters	ω value
One-ratio model	-75765.939912	133	0.16574
Free-ratio model	-75449.905430	263	0.05522
Two-ratio model	-75760.995364	134	background (0.16480), IaCPQ branch as foreground (0.28234)

Supplementary Table 3 Hiseq data assembling from roots, petals, and stems

Total trinity contigs	167,939
Percent GC	41.11
N90	271
N80	383
N70	551
N60	776
N50	1,057
N40	1,374
N30	1,743
N20	2,206
N10	3,005
Maximum contig length	14,757
Minimum contig length	201
Average contig length	670
Total assembled bases	152,588,251

Supplementary Table 4 P450 genes showed the same expression pattern as *IaCPQ*, *CYP708A15v2*, and *CYP708A16* in the hierarchical cluster analysis

Name	Annotation	Root_FPKM	Petal_FPKM	Stem_FPKM
c138495_g9_i2	**CYP70815v2**	**289.77**	**1042.14**	**674.93**
c138495_g8_i2	CYP708A14	254.78	742.04	358.15
c138495_g9_i1	CYP70815v1	168.08	564.14	430.54
c143740_g2_i3	**IaCPQ**	**19.34**	**105.59**	**96.64**
c134847_g1_i1	CYP91A2	19.27	57.76	42.36
c141458_g2_i2	CYP72A13	5.39	29.92	18.79
c51282_g1_i1	CYP708A2	0.19	26.67	25.99
c141272_g2_i1	CYP81D8	0.02	25.45	15.53
c144136_g7_i2	CYP90A1	3.4	18.61	9.05
c143985_g4_i2	CYP76C2	7.71	14.66	9.82
c143285_g2_i1	CYP86A8	3.81	12.65	11.85
c138467_g1_i2	CYP71B6	7.32	11.06	9.42
c144174_g1_i3	CYP71B7	7.36	10.07	9.62
c144995_g2_i6	CYP72A11	3.28	10.02	8.92
c96586_g1_i1	CYP96A9	0.01	8.2	3.41
c142167_g2_i5	CYP71B11	0.08	6.78	6.36
c137075_g2_i1	CYP72A14	1.24	6.68	2.61
c110176_g2_i1	CYP76C1	0.26	6.54	0.48
c137636_g2_i7	CYP718	0.28	6.25	5.52
c117846_g1_i1	CYP715A1	0.32	6.07	3.67
c142167_g2_i3	CYP71B11	0.18	5.36	3.85
c142167_g2_i1	CYP71B11	0.09	5.31	4.6
c134323_g1_i6	CYP97B3	1.76	3.74	3.66
c137371_g2_i5	CYP708A3	0.78	1.55	1.39
c137371_g2_i2	CYP708A2	0.68	1.5	1.15
c137371_g2_i1	CYP708A3	0.6	1.34	1.33
c129976_g2_i1	CYP78A5	0.29	1.2	0.33
c88613_g1_i1	CYP710A1	0.03	1.14	0.71
c97057_g1_i1	**CYP708A16**	**0.07**	**0.84**	**0.69**
c139206_g1_i6	CYP81K1	0.48	0.72	0.68
c123163_g1_i2	CYP82G1	0.05	0.51	0.46
c139206_g1_i1	CYP81K1	0.28	0.46	0.33
c140397_g1_i4	CYP75B1	0.08	0.28	0.14

Supplementary Table 5 P450 gene/contig numbers comparison between *Arabidopsis thaliana* and *I. amara.* Subfamilies with expanded P450 members from *I. amara* compare to *A. thaliana* were highlighted with different colours corresponding to Figure 6B. Subfamilies containing P450s catalyzing triterpenoids reported in other literatures were labeled with red colour.

Sub Family	Gene/contig number comparison	
	Arabidopsis thaliana	*Iberis amara*
CYP71B	37	36
CYP72A	9	26
CYP705A	31	25
CYP708A	4	17
CYP81D	8	16
CYP96A	14	11
CYP707A	4	10
CYP76C	8	10
CYP706A	7	9
CYP71A	17	9
CYP78A	6	9
CYP86A	8	8
CYP709B	3	6
CYP712A	2	5
CYP77A	5	5
CYP79B	3	5
CYP710A	4	4
CYP86C	4	4
CYP714A	2	3
CYP715A	1	3
CYP75B	1	3
CYP81F	4	3
CYP82G	1	3
CYP98A	3	3
CYP702A	8	2
CYP703A	1	2
CYP721A	1	2
CYP735A	2	2
CYP81K	2	2
CYP83B	1	2
CYP84A	1	2
CYP86B	1	2
CYP89A	8	2
CYP51G	2	1
CYP701A	1	1
CYP704A	2	1
CYP704B	1	1
CYP711A	1	1
CYP718	1	1
CYP720A	1	1

CYP722A	1	1
CYP734A	1	1
CYP72C	1	1
CYP73A	1	1
CYP74A	1	1
CYP74B	1	1
CYP76G	1	1
CYP77B	1	1
CYP79A	3	1
CYP79F	2	1
CYP81H	1	1
CYP82C	3	1
CYP82F	1	1
CYP83A	1	1
CYP85A	1	1
CYP87A	1	1
CYP88A	2	1
CYP90A	1	1
CYP90B	1	1
CYP90C	1	1
CYP90D	1	1
CYP93D	1	1
CYP94B	3	1
CYP94C	1	1
CYP94D	2	1
CYP97A	1	1
CYP97B	1	1
CYP97C	1	1
CYP716A	2	0
CYP724A	1	0
CYP79C	5	0
CYP81G	1	0

Supplementary Table 6 Primer sequences used in this study

Name	Primer Sequence	Purpose
IaCAS-userF	GGCTTAAXATGTGGAAACTGAGAATCGCC	Cloning, *N.benthamiana* expression, subcellular localization
IaCAS-userR	GGTTTAAXTCATTCCCCATGGTGCAATAA	Cloning, *N.benthamiana* expression, subcellular localization
IaCPQ-userF	GGCTTAAXATGTGGAAACTGAAAATCGC	Cloning, *N.benthamiana* expression, subcellular localization
IaCPQ-userR	GGTTTAAXTTATTCTCCATATTTCAATAAAACCTG	Cloning, *N.benthamiana* expression, subcellular localization
IaOSC3-userF	GGCTTAAXATGTGGAGGCTAAGGATTGGAG	Cloning
IaOSC3-userR	GGTTTAAXTTACATAAGTAAACGCATGGCTTTAG	Cloning
IaOSC4-userF	GGCTTAAXATGTGGAGGCTGAAGATCGGAGC	Cloning
IaOSC4-userR	GGTTTAAXTCAGAGCATGCGCAAAGCCTTGG	Cloning
IaOSC5-userF	GGCTTAAXATGTGGAAACTGAAGATAGGAG	Cloning
IaOSC5-userR	GGTTTAAXCTATAGATGATGTTGATCTGTTG	Cloning
IaOSC1-pdnorF	GGGGACAAGTTTGTACAAAAAAGCAGGCTCCATGTGGAAACTGAGAATCGCC	Yeast expression
IaOSC1-pdnorR	GGGGACCACTTTGTACAAGAAAGCTGGGTTCATTCCCCATGGTGCAATAA	Yeast expression
IaOSC2-pdnorF	GGGGACAAGTTTGTACAAAAAAGCAGGCTCCATGTGGAAACTGAAAATCGC	Yeast expression
IaOSC2-pdnorR	GGGGACCACTTTGTACAAGAAAGCTGGGTTTATTCTCCATATTTCAATAAAACCTG	Yeast expression
ClCPQ-userR-stop	GGTTTAAXCCATCAGTCAAAACCCGATGGA	Subcellular localization
CpCPQ-userR-stop	GGTTTAAXCCTTCAGTAAGAACCCGATG	Subcellular localization
IaOSC2-userR-stop	GGTTTAAXCCTTCTCCATATTTCAATAAAACCTG	Subcellular localization
Iaactin-qpcrF	GGACATCCAACCCCTTGTCT	qPCR
Iaactin-qpcrR	CCATCCCAACCATGACACCA	qPCR
IaCPQ-qpcrF	CAAATTGCACCTACCCCGGA	qPCR
IaCPQ-qpcrR	CCCACGGCAATCCTCTTTCA	qPCR
Clactin-qpcrF	AGCTGCAGGAATCCACGAAA	qPCR
Clactin-qpcrR	GTGGCGCAACGACCTTAATC	qPCR
ClCPQ-qpcrF	TGGAAAACGGCTGAGGGAAA	qPCR
ClCPQ-qpcrR	TAGTCGGGGACACGTTGAAG	qPCR

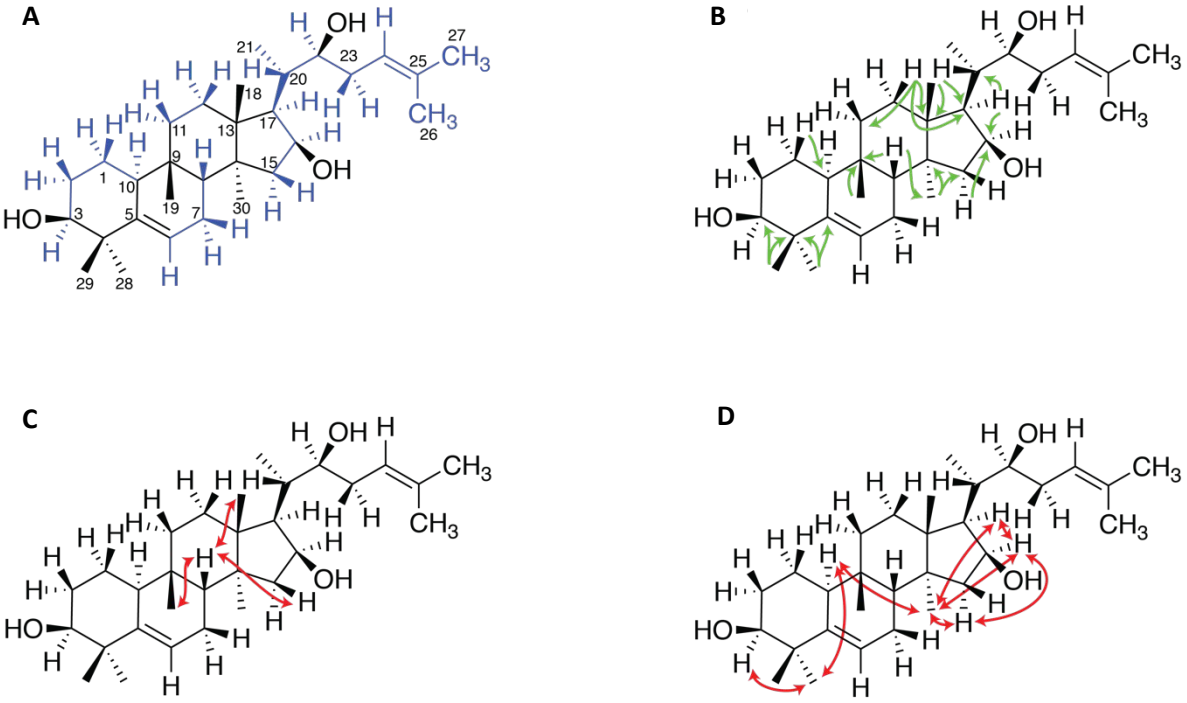
Supplementary NMR data

General experimental procedures

NMR experiments were performed at 300 K using a 600 MHz Bruker Avance III instrument (operating frequency of 600.13 MHz) equipped with a Bruker SampleJet sample changer and a 5-mm room temperature BBI probe head with z-gradient and automated tuning and matching (Bruker Biospin, Rheinstetten, Germany). Approximately 1 mg sample was dissolved in methanol-d_4, and the acquired data were calibrated according to the residual solvent peak at δ 3.31 for ^{1}H and δ 49.01 ppm for ^{13}C. All pulse programs were from Bruker's standard experiment library, and 1D ^{1}H spectra were acquired with 30°-pulses, 1.5 s relaxation delay, 20 ppm spectral width, and 64k data points. 2D homonuclear experiments were acquired with 10 ppm spectral width and 8k (COSY) or 2k (NOESY/ROESY) data points in the direct dimension, collecting 256 FIDs for the indirect dimension. 2D heteronuclear experiments were acquired with 10 ppm spectral width and 8k (HSQC) or 4k (HMBC) data points in the direct dimension, collecting 256 (multiplicity edited HSQC) or 512 (HMBC) FIDs for a spectral width of 240 ppm in the indirect dimension. The HSQC experiments were optimized for $^{1}J_{C,H}$ = 145Hz and HMBC experiments were optimized for optimized for $^{n}JC,H$ = 8Hz. Automated tuning and matching, optimization of lock parameters, gradient shimming, setting of receiver gain, and acquisition of NMR data were performed using topspin 3.5 (Bruker Biospin, Rheinstetten, Germany). Before measurement, each sample was pre-heated at 300 K for 60 s and kept for 5 min inside the NMR probe head to reach temperature equilibrium of 300±1 K.

Compound 1: 16β,22-dihydroxycucurbitadienol

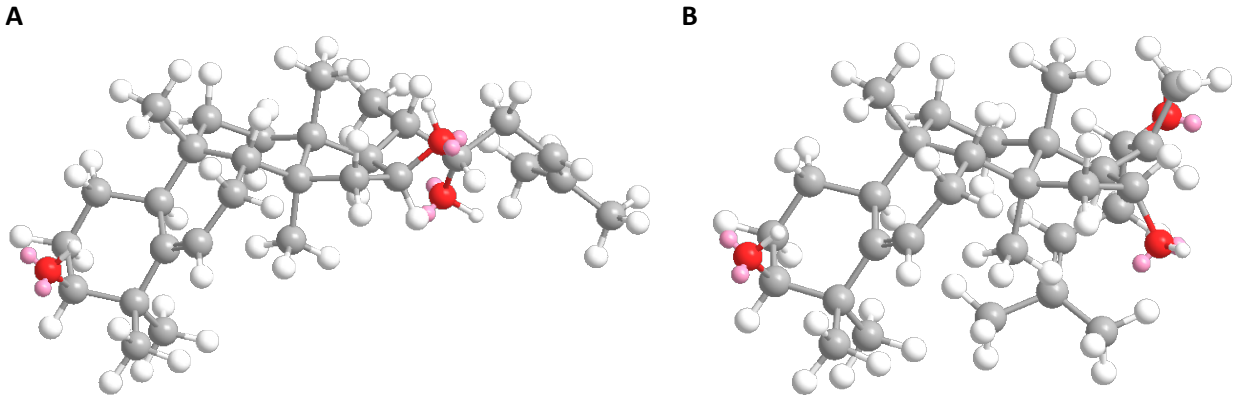
Approximately 1 mg of compound **1** was dissolved in 650 µl methanol-d_4, and 1D ^{1}H and 2D COSY, HSQC, HMBC, and ROESY NMR spectra were acquired as described above. The ^{1}H NMR spectrum showed five methyl singlets at δ 0.89 (H-30), 0.95 (H-19), 1.01 (H-28), 1.09 (H-18), and 1.11 (H-29), a methyl doublet at δ 0.95 (3H, d, 6.6 Hz, H-21), two signals for double bond-bound methyl groups at δ 1.63 (H-27) and 1.70 (H-26), two olefinic signals at δ 5.54 (1H, m, H-6) and 5.25 (1H, br t, 7.1 Hz, H-24), and a downfield shifted methine triplet at δ 3.44 (1H, br t, 2.7 Hz, H-3), which indicated the presence of a cucurbitadienol core skeleton(Shang, et al. 2014). However, with the three oxygen according to the molecular formula and two downfield shifted methine signals at δ 4.42 (1H, td, 7.6, 5.8 Hz, H-16) and 3.76 (1H, ddd, 7.3, 6.3, 1.5 Hz, H-22) indicated two additional hydroxyl groups in compound **1**. Analysis of the COSY spectrum led to identification of the CHO(3)-CH_2(2)-CH_2(1)-CH(10), CH(6)-CH_2(7)-CH(8), CH_2(11)-CH_2(12), and CH_2(15)-CHO(16)-CH(17)-CH(20)(CH_3(21))-CHO(22)-CH_2(23)-CH(24)(CH_3(26))-CH_3(27) spin systems indicated with blue in Supplementary Figure 12A. These spin systems, the remaining methyl groups, and quaternary carbon atoms C-4, C-5, C-9, C-13, C-14 and C-25 were assembled based on HMBC correlations, of which selected are shown in Supplementary Figure 12B. In the ROESY spectrum, a network of correlations between H-19 ↔ H-8 ↔ H-18 ↔ H-15β showed these to be β-positioned (above the plane of the core skeleton) as seen in Supplementary Figure 12C. Similarly, network of correlations between H-3 ↔ H-28 ↔ H-10 ↔ H-30 ↔ H-15α ↔ H-16 ↔ H-17 ↔ H-30 showed these to be α-positioned (below the plane of the core skeleton) as seen in Supplementary Figure 12D. The pseudo-equatorial α-position of H-3 (as seen from the above-mentioned ROEs and the triplet structure arising from pseudo-equatorial to pseudo-equatorial and pseudo-equatorial to pseudo-axial couplings of 2.7 Hz (Supplementary Table 7)) leads to a β-positioned (pseudo-axial) 3-hydroxyl group, as also reported for cucurbitadienol(Shang, et al. 2014) and related analogs(Itkin, et al. 2016).



Supplementary Figure 12 **(A)** Structure and numbering of 16β,22-dihydroxycucurbitadienol with correlations found in the COSY spectrum indicated in blue. **(B)** Selected HMBC correlations used for structure elucidation with $^{n}J_{C,H}$ optimized for 8 Hz (arrows pointing from H to C). **(C)** Correlations from the ROESY spectrum used to identify β-oriented protons, i.e., pointing up, starting with H-19 as the anchoring point. **(D)** Correlations from the ROESY spectrum used to identify α-oriented protons, i.e., pointing down, starting with H-3 as the anchoring point.

Whereas cucurbitacin C and its deacetyl analog has been reported to have an α-positioned hydroxyl group at C-16(Shang, et al. 2014), compound **1** has a β-positioned hydroxyl group at C-16. This is seen from the above-mentioned ROEs from H-30, H-15α and H17 to H-16, where especially the ROE between the 1,3-pseudodiaxial H-30 and H-16 would not have been observed with an α-positioned hydroxyl at C-16. Furthermore, the coupling pattern of H-16 (td, 7.6, 5.8 Hz) supports the β-orientation of the proton and the α-orientation of the hydroxyl group at C-16. Thus, after initial energy minimization of compound **1** using MM2 force field, ten repeated cycles of MM2 molecular dynamics followed by MM2 energy minimization in Chem3D ver. 18.1 (PerkinElmer, Waltham, MA, USA), lead to convergence of minimum energy conformations, of which the last is shown in Supplementary Figure 13A. The following dihedral angles were found for the ten minimization cycles: H16-C16-C17-H17 = 27.6 ± 0.1°, H15α-C15-C16-H16 = 16.7 ± 0.1°, and H15β-C15-C16-H16 = 135.0 ± 0.1°. For its 16-epimer, where the minimum energy conformation is shown in Supplementary Figure 13B, the following dihedral angles were measured: H16-C16-C17-H17 = 147.5 ± 0.7°, H15α-C15-C16-H16 = -100.6 ± 0.7°, and H15β-C15-C16-H16 = 17.3 ± 0.7°. These dihedral angles were used for calculating coupling constants according to the Karplus equation using SweetJ version 4.2 desktop calculator(Balacco 1996) (www.inmr.net/sweetj.html), which resulted in $^{3}J_{H16,H17} = {}^{3}J_{H15\beta,H16}$ = 7.4 Hz and $^{3}J_{H15\alpha,H16}$ = 6.8 Hz for compound **1** (β-positioned hydroxyl group), whereas the calculations resulted in $^{3}J_{H16,H17}$ = 7.2 Hz, $^{3}J_{H15\beta,H16}$ = 7.4 Hz, and $^{3}J_{H15\alpha,H16}$ = 1.9 Hz for its 16-epimer (α-positioned hydroxyl group). Thus, the triplet of doublet pattern with coupling constants of 7.6 and 5.8 Hz observed for H-16 in the ^{1}H NMR spectrum of **1** is in relative good agreement with the theoretical triplet of doublet pattern with 7.4 and 5.8 Hz calculated for a β-positioned

hydroxyl group at C-16 - whereas the theoretical doublet of doublet of doublet pattern with coupling constants of 7.4, 7.2, and 1.9 Hz for its 16-epimer is far from the observed coupling pattern. Thus, the 16-hydroxyl group of **1** is in β-position.



Supplementary Figure 13 Minimum energy conformations after initial energy minimization of compound 1 using MM2 force field, ten repeated cycles of MM2 molecular dynamics followed by MM2 energy minimization. **(A)** Compound **1** (16β,22-dihydroxycucurbitadienol). **(B)** the 16-epimer of compound **1** (16α,22-dihydroxycucurbitadienol)

The last hydroxyl group is positioned at C-22, as seen from COSY cross peaks from the downfield shifted H-22 (δ 3.76) to H-20 and H-23. The free rotation around the C17-C20 bond does not allow assignment of configuration at C-20 and C-22, but **1** is tentatively assigned the *S*-configuration at C-20 in agreement with previous reported cucurbitadienol analogs(Shang, et al. 2014; Itkin, et al. 2016).

Compound **1** is a previously unreported compound. Full assignment of all ^{1}H and ^{13}C resonances as well as correlations observed in COSY, ROESY and HMBC experiments are provided in Supplementary Table 7.

Supplementary Table 7 ^{1}H (600 MHz) and ^{13}C NMR (150 MHz) spectroscopic data and COSY, HSQC, HMBC, and ROESY correlations of compound 1 acquired in methanol-d_4.

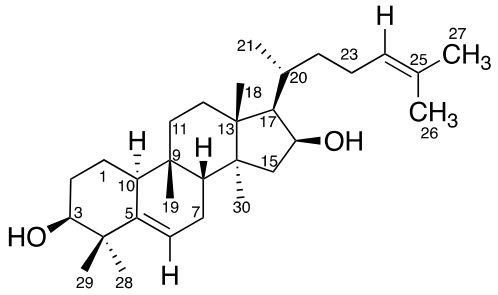
Pos.	δ_C[a]	δ_H (nH, multiplicity, *J* in Hz)[b]	COSY	HMBC[c]	ROESY
1	22.1	1.55 (2H, m)	H-2B, H-10	C-10	
2	30.1	A: 1.67 (1H, m) B: 1.91 (1H, m)	A: H-2B, H-3 B: H-2A, H-3		A: H-3 B: H-3, H-10
3	77.2	3.44 (1H, br t, $J_{H3,H2A}$ = $J_{H3,H2B}$ = 2.7)	H-2A, H-2B		H-2A, H2B, H-28, H-29
4	41.7				
5	143.1				
6	120.5	5.54 (1H, m)	H-7α, H-7β, H-10	C-4, C-7, C-8, C-10	H-7α, H-7β
7	25.1	α: 1.81 (1H, br dd, $J_{H7α,H7β}$ = 19.2, dd, $J_{H7α,H6}$ = 6.3) β: 2.41 (1H, br dd, $J_{H7α,H7β}$ = 19.2, $J_{H7β,H8}$ ≈ 8.1)	α: H-6, H-7β, H-8 β: H-6, H-7α, H-8	α: C-5, C-6, C-9, C-14 β: C-9	α: H-6, H-7β β: H-6, H-7α, H-8, H-19

8	44.7	1.85 (1H, m)	H-7α, H-7β	C-6, C-7, C-10, C-14, C-15, C-19, C-30	H-7β, H-15β, H-18, H-19
9	35.2				
10	39.5	2.30 (1H, br m)	H-1, H-6,		H-2B, H-12B, H-28, H-30
11	32.9	A: 1.50 (1H, m) B: 1.69 (1H, m)	A: H-11B B: H-11A	A: C-10, C-12 B: C-9	A: H-11B, H-18 B: H-11A
12	31.5	A: 1.52 (1H, m) B: 1.73 (1H, m)	A: H-12B B: H-12A	A: C-9, C-10 B: C-15, C-18	A: H-15α, H-12B B: H-10, H-12A
13	48.5				
14	48.5				
15	47.3	α: 1.92 (1H, dd, $J_{H15\alpha,H15\beta}$ = 12.1, $J_{H15\alpha,H16}$ = 7.6) β: 1.23 (1H, dd, $J_{H15\alpha,H15\beta}$ = 12.1, $J_{H15\beta,H16}$ = 5.8)	α: H-16, H-15β β: H-16, H-15α	α: C-14, C-17, C-30 β: C-14, C-16, C-30	α: H-12A, H-16, H-15β, H-30 β: H-7β, H-15α
16	72.6	4.42 (1H, td, $J_{H16,H15\alpha}$ = $J_{H16,H17}$ = 7.6, $J_{H16,H15\beta}$ = 5.8)	H-15α, H-15β, H-17	C-15	H-15α, H-17, H-22, H-30
17	51.3	2.04 (1H, dd, $J_{H17,H20}$ = 11.2, $J_{H17,H16}$ = 7.6)	H-16, H-20	C-15, C-18, C-20	H-16, H-30
18	16.0	1.09 (3H, s)		C-12, C-13, C-17	H-8, H-11A, H-20
19	28.3	0.95 (3H, s)		C-8, C-9, C-10, C-11	H-7β, H-8
20	36.0	2.12 (1H, m)	H-17, H-21		H-18, H-22
21	14.4	0.95 (3H, d, $J_{H21,H20}$ = 6.6 Hz)	H-20	C-20, C-17, C-22	H-22, H-23
22	76.2	3.76 (1H, br ddd, $J_{H22,H23}$ ≈ 7.3, $J_{H22,H23}$ ≈ 6.3, $J_{H22,H20}$ = 1.5)	H-23	C-17, C-21, C-24	H-16, H-20, H-21, H-23, H-24
23	33.3	2.17 (2H, m)	H-22, H-24, H-26, H-27	C-20, C-22, C-24, C-25	H-21, H-22, H-24
24	122.9	5.25 (1H, br t, $J_{H24,H23}$ = 7.2)	H-23, H-26, H-27	C-27	H-22, H-23, H-27
25	133.1				
26	17.8	1.63 (3H, br s)	H-23, H-24	C-24, C-25, C-27	
27	25.7	1.70 (3H, q, $J_{H27,H23}$ ≈ $J_{H27,H24}$ ≈ 1.2)	H-23, H-24	C-24, C-25, C-26	H-24
28	28.0	1.01 (3H, s)		C-3, C-4, C-5, C-29	H-3, H-10
29	25.8	1.11 (3H, s)		C-3, C-4, C-5, C-28	H-3
30	19.5	0.89 (3H, s)		C-8, C-14, C-15	H-10, H-15α, H-16, H-17

[a] ^{13}C chemical shift data obtained from HSQC and HMBC spectra. [b] Multiplicities reported as apparent splittings with s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad with spin-spin coupling constants listed in the order of the reported multiplicity. [c] Correlations from H to indicated C in this column.

Compound 2: 16β-hydroxycucurbitadienol

Approximately 1 mg of **2** was dissolved in 650 µl methanol-d_4, and 1D ^{1}H NMR as well as 2D COSY and HSQC spectra were acquired and compared with the spectra of **1**. The overlayed spectra of **1** and **2** were almost identical, except that **2** lacked the signal for the oxymethine H-22 observed as a br ddd at δ 3.76 in **1** but instead contained an additional pair of non-oxygenated diastereotopic methylene protons at δ 1.64 and δ 1.03 in **2**. The configuration of the hydroxy groups was assigned 3β at C-3 and 16β at C-16, based on identical coupling patterns seen for H-3 and H-16 in **1** and **2**.



Supplementary Figure 14 Structure and numbering of 16β-hydroxycucurbitadienol (2)

Full assignment of all ^{1}H and ^{13}C resonances as well as correlations observed in COSY, ROESY and HMBC experiments are provided in Supplementary Table 8.

Supplementary Table 8 ^{1}H (600 MHz) and ^{13}C NMR (150 MHz) spectroscopic data and COSY correlations of compound 2 acquired in methanol-d_4.

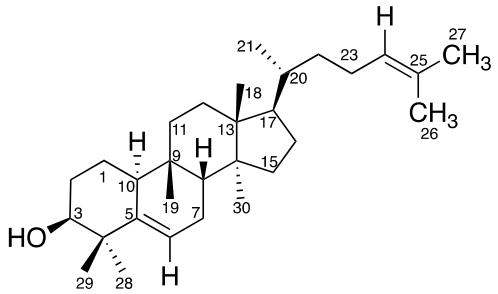
Pos.	δ_C[a]	δ_H (nH, multiplicity, *J* in Hz)[b]	COSY
1	22.1	1.54 (2H, m)	H-2A, H-2B, H-10
2	30.0	A: 1.67 (1H, m) B: 1.91 (1H, m)	A: H-2B, H-3 B: H-2A, H-3
3	77.1	3.44 (1H, t, $J_{H3,H2A} = J_{H3,H2B} = 2.8$)	H-2A, H-2B
4	n.d.		
5	n.d.		
6	120.5	5.53 (1H, dt, $J_{H6,\ H7\alpha} = 6.1$, $J_{H6,\ H7\beta} \approx J_{H6,\ H10} \approx 1.8$)	H-7α, H-7β, H-10
7	25.0	α: 1.81 (1H, br dd, $J_{H7\alpha,H7\beta} = 18.9$, dd, $J_{H7\alpha,H6} = 6.1$) β: 2.41 (1H, dddd, $J_{H7\alpha,H7\beta} = 18.9$, $J_{H7\beta,H8} = 8.0$, $J_{H7\beta,H16} = 1.8$, $J_{H7\beta,H10} = 3.3$)	α: H-6, H-7β β: H-8, H-7α
8	44.5	1.83 (1H, br d, $J_{H8\text{-}H7\beta} = 8.0$)	H-7β
9	n.d.		
10	39.4	2.28 (1H, br m)	H-1, H-6,
11	32.9	A: 1.48 (1H, m) B: 1.68 (1H, m)	A: H-11B B: H-11A
12	31.2	A: 1.53 (1H, m) B: 1.71 (1H, m)	A: H-12B B: H-12A
13	n.d.		
14	n.d.		
15	47.7	α: 1.92 (1H, dd, $J_{H15\alpha,H15\beta} = 12.1$, $J_{H15\alpha,H16} = 7.8$)	α: H-16, H-15β

		β: 1.21 (1H, ddd, $J_{H15\alpha,H15\beta}$ = 12.1, $J_{H15\beta,H16}$ = 5.8, $J_{H15\beta,H30}$ = 1.2)	β: H-16, H-15α
16	72.7	4.38 (1H, td, $J_{H16,H15\alpha}$ = $J_{H16,H17}$ = 7.7, $J_{H16,H15\beta}$ = 5.8)	H-15α, H-15β, H-17
17	55.4	1.65 (1H, m)	H-16, H-20
18	16.0	1.05 (3H, s)	
19	28.3	0.94 (3H, s)	
20	31.1	1.86 (1H, m)	H-17, H-21, H-22
21	18.5	0.96 (3H, d, $J_{H21,H20}$ = 6.6 Hz)	H-20, H-22
22	37.4	A: 1.03 (1H, m) B: 1.65 (1H, m)	H-20, H-21, H-23
23	25.8	A: 1.94 (1H, m) B: 2.08 (1H, m)	H-22, H-24
24	126.2	5.15 (1H, triplet of septets, $J_{H24,H23}$ = 7.2, $J_{H24,H26}$ = $J_{H24,H27}$ = 1.4)	H-23, H-26, H-27
25	n.d.		
26	17.4	1.60 (3H, br s)	H-24
27	25.6	1.66 (3H, q, $J_{H27,H23} \approx J_{H27,H24} \approx$ 1.1)	H-24
28	27.9	1.00 (3H, s)	
29	25.8	1.11 (3H, s)	
30	19.3	0.84 (3H, d, $J_{H30\text{-}H15\beta}$ = 1.2)	

[a] ^{13}C chemical shift data obtained from HSQC and HMBC spectra. [b] Multiplicities reported as apparent splittings with s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad with spin-spin coupling constants listed in the order of the reported multiplicity.

Compound 3: Cucurbitadienol

Approximately 1 mg of compound **3** was dissolved in 650 µl methanol-d_4, and 1D ^{1}H NMR as well as 2D COSY and HSQC spectra were acquired and compared with the spectra of **1** and **2** as well as NMR data of cucurbitadienol reported in the literature(Itkin, et al. 2016). Thus unambiguously established the structure of 3 as cucurbitadienol (Supplementary Figure 15).



Supplementary Figure 15 Structure and numbering of cucurbitadienol (3)

Full assignment of all ^{1}H and ^{13}C resonances as well as correlations observed in COSY, ROESY and HMBC experiments are provided in Supplementary Table 9.

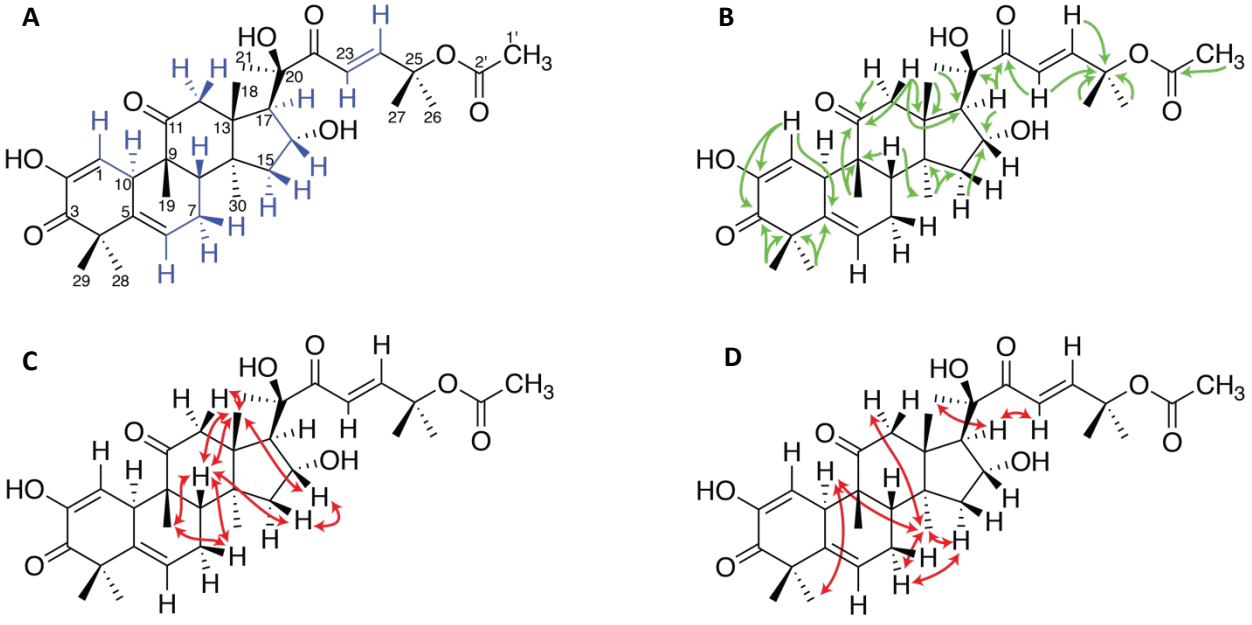
Supplementary Table 9 ^{1}H (600 MHz) and ^{13}C NMR (150 MHz) spectroscopic data of compound 3 acquired in methanol-d_4 and comparison with data from literature(Itkin, et al. 2016).

	NMR data acquired of **3** in this study		NMR data of **3** found in the literature(Itkin, et al. 2016)	
Pos.	δ_H (nH, multiplicity, *J* in Hz)[b]	HSQC	1D ^{1}H	^{13}C
1	1.56 (2H, m)	22.0	1.46, 1.57	21.14
2	A: 1.68 (1H, m) B: 1.92 (1H, m)	29.9	1.71, 1.88	28.90
3	3.45 (1H, t, $J_{H3,H2A} = J_{H3,H2B} = 2.8$)	77.1	3.47	76.66
4		n.d.		41.45
5		n.d.		141.23
6	5.53 (1H, br dt, $J_{H6,H7} = 5.9$, $J_{H6,H7} \approx J_{H6,H10} \approx 1.8$)	120.6	5.59	121.51
7	α: 1.83 (1H, br dd, $J_{H7\alpha,H7\beta} = 18.8$, dd, $J_{H7\alpha,H6} = 5.9$) β: 2.39 (1H, dddd, $J_{H7\alpha,H7\beta} = 18.8$, $J_{H7\beta,H8} = 7.9$, $J_{H7\beta,H16} = 2.2$, $J_{H7\beta,H10} = 3.3$)	25.0	1.80, 2.38	24.39
8	1.77 (1H, br d, $J_{H8\text{-}H7\beta} = 8.1$)	44.9	1.76	43.67
9				34.49
10	2.32 (1H, m)	39.2	2.27	37.85
11	A: 1.41 (1H, m) B: 1.70 (1H, m)	33.2	1.44, 1.66	32.34
12	A: 1.54 (1H, m) B: 1.72 (1H, m)	31.4	1.51, 1.68	30.46
13		n.d.		46.27
14		n.d.		49.17
15	A: 1.17 (1H, m) B: 1.24 (1H, m)	35.5	1.12, 1.20	34.77
16	A: 1.29 (1H, m) B: 1.92 (1H, m)	28.8	1.27, 1.88	27.94
17	1.53 (m)	51.5	1.50	50.47
18	0.89 (3H, s)	18.1	0.85	15.37
19	0.94 (3H, s)	28.2	0.92	28.06
20	n.d.	n.d.	1.43	35.82
21	0.93 (3H, d, $J_{H21,H20} = 6.5$ Hz)	18.9	0.90	18.67
22	A: 1.06 (1H, m) B: 1.46 (1H, m)	37.1	1.03, 1.42	36.46
23	A: 1.90 (1H, m) B: 2.04 (1H, m)	25.5	1.85, 2.02	24.87
24	5.09 (1H, broad triplet of septets, $J_{H24,H23} = 7.2$, $J_{H24,H26} = J_{H24,H27} = 1.4$)	125.9	5.09	125.25
25		n.d.		130.94
26	1.59 (3H, s)	17.3	1.68	25.73
27	1.67 (3H, s)	25.6	1.60	17.63
28	1.02 (3H, s)	28.1	1.02	27.26
29	1.12 (3H, s)	25.7	1.14	25.44
30	0.86 (3H, br s)	18.1	0.80	17.83

[a] Multiplicities reported as apparent splittings with s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad with spin-spin coupling constants listed in the order of the reported multiplicity. [b] ^{13}C chemical shift data obtained from HSQC spectrum. n.d. = not detected.

Compound 4: Cucurbitacin E

1 mg of a reference sample of cucurbitacin E bought from Sigma-Aldrich was dissolved in 650 µl methanol-d_4, and 1D ^{1}H NMR as swell as 2D COSY, NOESY, ROESY, HSQC, and HMBC spectra were acquired. Analysis of the COSY spectrum led to identification of the CH(1)-CH(10), CH(6)-CH_2(7)-CH(8), CH_2(12), CH_2(15)-CHO(16)-CH(17), and CH(23)-CH(24) spin systems indicated with blue in Supplementary Figure 16A, and the HMBC spectrum allowed assemblance of the molecule (selected HMBC correlations shown in Supplementary Figure 16B). The orientation of all hydrogens and groups were identified based on analysis of coupling patterns and correlations found in the ROESY spectrum. Thus, in the ROESY spectrum, a network of correlations between H-19 ↔ H-8 ↔ H-15β ↔ H-16 ↔ H-18 (↔ H-8) ↔ H-12β ↔ H-8 showed these to be β-positioned (above the plane of the core skeleton) as seen in Supplementary Figure 16C. Similarly, a network of correlations between H-28 ↔ H-10 ↔ H-30 (↔ H-15α ↔ H-7α ↔ H-30) ↔ H-12α showed these to be α-positioned (below the plane of the core skeleton) as seen in Supplementary Figure 16D. The pseudo-equatorial β-position of H-16 (as seen from the above-mentioned ROEs) leads to an α-positioned 16-hydroxyl group, as previously reported for cucurbitacin E(Ayyad, et al. 2012; Chawech, et al. 2015). The doublet of doublet of doublet coupling pattern ($J_{H16,H15\alpha}$ = 8.9, $J_{H16,H17}$ = 7.3, $J_{H16,H15\beta}$ = 1.1) observed for H-16 in **4** (α-oriented 16-hydroxyl) is in close accordance with the coupling constants calculated for an α-oriented 16-hydroxyl in **1**, *vide supra*, and differs from the coupling pattern observed for H-16 in **1** (β-oriented 16-hydroxyl). This further confirms the 16β-hydroxyl in **1** and the 16α-hydroxyl in **4**.



Supplementary Figure 16 (A) Structure and numbering of cucurbitacin E with correlations found in the COSY spectrum indicated in blue. **(B)** Selected HMBC correlations used for structure elucidation with $^nJ_{C,H}$ optimized for 8 Hz (arrows pointing from H to C). **(C)** Correlations from the ROESY spectrum used to identify β-oriented protons, i.e., pointing up, starting with H-8 as the anchoring point. **(D)** Correlations from the ROESY spectrum used to identify α-oriented protons, i.e., pointing down, starting with H-10 as the anchoring point.

Full assignment of all ^{1}H and ^{13}C resonances as well as correlations observed in COSY, HSQC, ROESY and HMBC experiments are provided in Supplementary Table 10, and they are in agreement with ^{1}H and ^{13}C NMR data of cucurbitacins and their glycosides as reported in literature(Ayyad, et al. 2012; Chawech, et al. 2015; Abdelkhalek, et al. 2017).

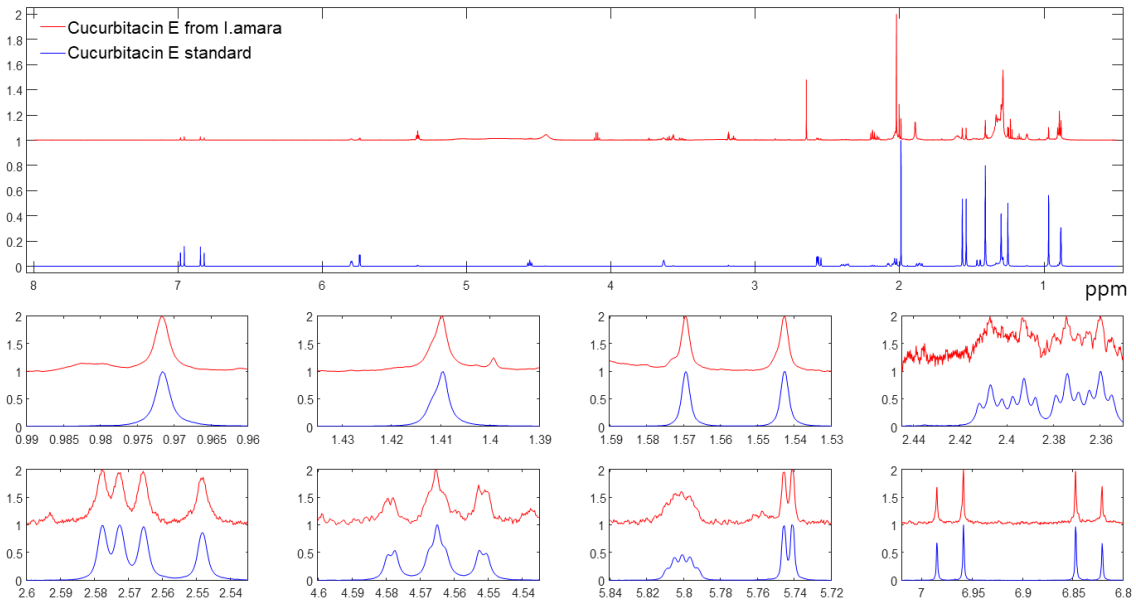
Supplementary Table 10 ^{1}H (600 MHz) and ^{13}C NMR (150 MHz) spectroscopic data and COSY, HSQC, HMBC and ROESY correlations of compound 4 in methanol-d_4.

Pos.	δ_C[a]	δ_H (nH, multiplicity, *J* in Hz)[b]	COSY	HMBC[c]	NOESY
1	116.4	5.75 (1H, d, $J_{H1,H10}$ = 2.7)	H-10	C-2, C-3, C-5, C-9	H-10, H-19
2	146.6	-			
3	199.8	-			
4	49.0	-			
5	138.3	-			
6	121.5	5.80 (1H, br dt, $J_{H6,H7\beta}$ = 4.9, $J_{H6,H7B} \approx$ 2.7, $J_{H6,H10} \approx$ 2.5)	H-7α, H-7β, H-8, H-10		H-7α, H-7β, H-29
7	24.4	α: 2.07 (1H, ddd, $J_{H7\alpha,H7\beta}$ = 19.7, $J_{H7\alpha,H6}$ = 4.9, $J_{H7\alpha,H10}$ = 2.1) β: 2.39 (1H, ddt, $J_{H7\beta,H7\alpha}$ = 19.7, $J_{H7\beta,H8}$ = 8.7, $J_{H7\beta,H6} \approx J_{H7\beta,H10} \approx$ 2.7)	α: H-6, H-7β, H-10 β: H-6, H-7α, H-8, H-10	α: C-5, C-6, C-8, C-9, C-13	α: H-6, H-7β, H-15α, H-30 β: H-6, H-7α, H-8, H-19
8	43.1	2.03 (1H, d, $J_{H8,H7\beta}$ = 8.7)	H-7β	C-6, C-7, C-9, C-10, C-13, C-15, C-19, C-30	H-7β, H-15β, H-18, H-19
9	50.0	-			
10	35.8	3.64 (1H, br p, $J_{H10,H1} \approx J_{H10,H6} \approx J_{H10,H7\alpha} \approx J_{H10,H7\beta} \approx$ 2.5)	H-1, H-6, H-7α, H-7β		H-1, H-12α, H-28, H-30
11	216.0	-			
12	49.5	α: 3.34 (1H, d, $J_{H12\alpha,H12\beta}$ = 14.7) β: 2.56 (1H, d, $J_{H12\beta,H12\alpha}$ = 14.7)	α: H-12β β: H-12α	α: C-11, C-13, C-17, C-18 β: C-11, C-13, C-18	α: H-10, H-12β, H-30 β: H-12α, H-18, H-21
13	51.5	-			
14	48.9	-			
15	46.5	α: 1.46 (1H, br d, $J_{H15\alpha,H15\beta}$ = 13.2) β: 1.87 (1H, dd, $J_{H15\beta,H15\alpha}$ = 13.2, $J_{H15\beta,H16}$ = 8.9)	α: H-15β, H-16 β: H-15α, H-16	α: C-13, C-14, C-16, C-17, C-30 β: C-8, C-30	α: H-7α, H-15β β: H-8, H-15α, H-18
16	71.7	4.57 (1H, ddd, $J_{H16,H15\beta}$ = 8.9, $J_{H16,H17}$ = 7.3, $J_{H16,H15\alpha}$ = 1.1)	H-15α, H-15β, H-17	C-14, C-20	H-15β, H-18, H-21(weak), H-23
17	59.8	2.58 (1H, d, $J_{H17,H16}$ = 7.3)	H-16	C-13, C-16, C-18, C-20, C-21, C-22	H-21, H-23
18	20.4	0.89 (3H, s)		C-12, C-13, C-14, C-17	H-8, H-12β, H-15β, H-16
19	20.0	0.97 (3H, s)		C-8, C-9, C-10, C-11	H-7β, H-8
20	80.1	-			
21	25.0	1.41 (3H, s)		C-17, C-20, C-22	H-16(weak), H-17, H-23, H-24

22	205.0	-			
23	122.5	6.84 (d, $J_{H23,H24}$ = 15.8)	H-24	C-22, C-24, C-25, C-27	H-16, H-17, H-21, H-26, H-27
24	151.3	6.97 (d, $J_{H24,H23}$ = 15.8)	H-23	C-22, C-23, C-25	H-21, H-26, H-27
25	80.8	-			
26	26.6	1.55 (3H, s)		C-24, C-25, C-27	H-23, H-24
27	26.2	1.57 (3H, s)		C-24, C-25, C-26	H-23, H-24
28	28.1	1.26 (3H, s)		C-3, C-4, C-5, C-29	H-10
29	20.5	1.30 (3H, s)		C-3, C-4, C-5, C-6, C-28	H-6
30	18.7	1.41 (3H, s)		C-8, C-13, C-14, C-15	H-7α, H-10, H-12α
1'	21.6	2.00 (3H, s)		C-2'	
2'	171.6	-			

[a] ^{13}C chemical shift data obtained from HSQC and HMBC spectra. [b] Multiplicities reported as apparent splittings with s = singlet, d = doublet, t = triplet, p = pentet, br = broad with spin-spin coupling constants listed in the order of the reported multiplicity. [c] Correlations from H to indicated C in this column.

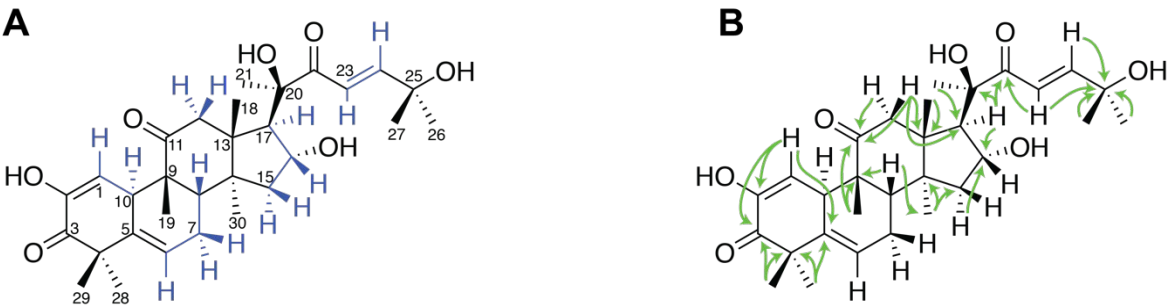
A sample of approximately 0.6 mg cucurbitacin E isolated from *Iberis amara* by using HPLC-UV/Vis based fraction collection was dissolved in 650 µl methanol-d_4, and a 1D ^{1}H NMR spectrum was acquired. Supplementary Figure 17 show superimposed 1D ^{1}H NMR spectra of the authentic cucurbitacin E standard (bottom, blue) and the sample isolated from *I. amara*. (top, red). This clearly demonstrates the two samples to be identical and thereby that the sample isolated from *I. amara* is cucurbitacin E.



Supplementary Figure 17 Bottom in blue: 1D ^{1}H NMR spectra of the authentic cucurbitacin E standard bought from Sigma-Aldrich. **Top in red**: 1D ^{1}H NMR spectra of the sample isolated from *I. amara*.

Compound 5: Cucurbitacin I

1 mg of a reference sample of cucurbitacin I bought from Sigma-Aldrich was dissolved in 650 µl methanol-d_4, and 1D ^{1}H NMR as swell as 2D COSY, HSQC, and HMBC spectra were acquired. All spectra were identical with the spectra of compound **4**, except for the lack of signals for the acetyl group at C-25 and a downfield shift of C-25 from δ 80.8 ppm in **4** to δ 71.2 in **1** due to the lack of acetyl group. Spin systems found in the COSY spectrum are shown in blue in Supplementary Figure 18A and selected HMBC correlations used to assemble the structure are shown in Supplementary Figure 18B.



Supplementary Figure 18 (A) Structure and numbering of cucurbitacin I with correlations found in the COSY spectrum indicated in blue. **(B)** Selected HMBC correlations used for structure elucidation with $^nJ_{C,H}$ optimized for 8 Hz (arrows pointing from H to C).

Full assignment of all ^{1}H and ^{13}C resonances as well as correlations observed in COSY, HSQC, and HMBC experiments are provided in Supplementary Table 11, and they are in agreement with ^{1}H and ^{13}C NMR data of cucurbitacins and their glycosides as reported in literature(Seger, et al. 2005; Ayyad, et al. 2012; Chawech, et al. 2015).

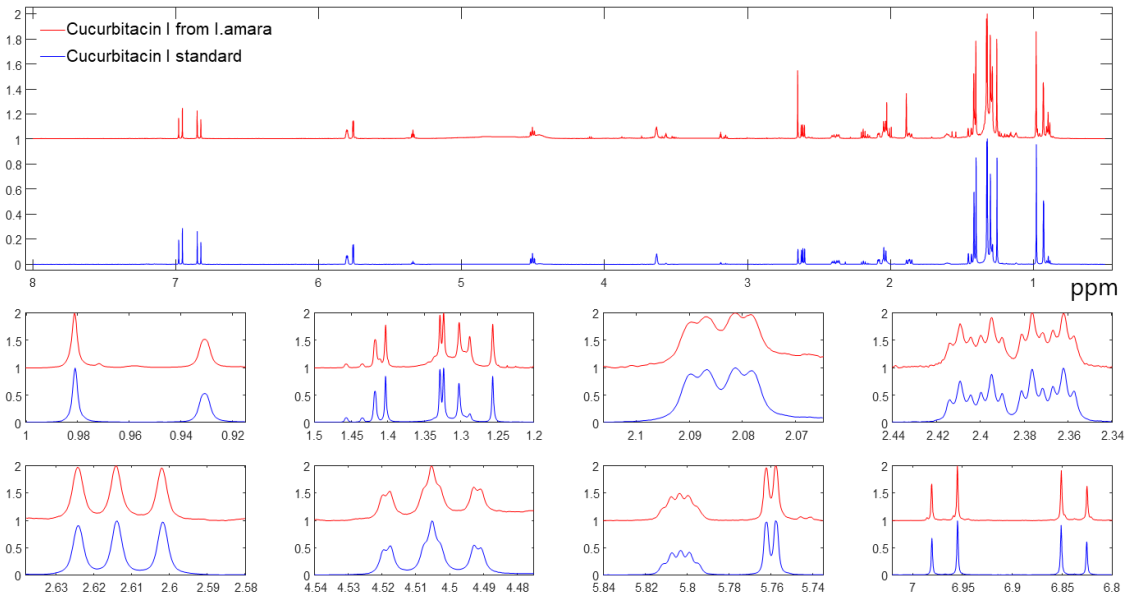
Supplementary Table 11 ^{1}H (600 MHz) and ^{13}C NMR (150 MHz) spectroscopic data and COSY, HSQC, and HMBC correlations of compound **5** in methanol-d_4.

Pos.	δ_C[a]	δ_H (nH, multiplicity, J in Hz)[b]	COSY	HMBC[c]
1	116.5	5.76 (1H, d, $J_{H1,H10}$ = 2.7)	H-10	C-2, C-3, C-5, C-9
2	146.5	-		
3	199.7	-		
4	49.0	-		
5	138.2	-		
6	121.5	5.81 (1H, br dt, $J_{H6,H7\alpha}$ = 4.9, $J_{H6,H7\beta} \approx J_{H6,H10} \approx$ 2.7)	H-7α, H-7β, H-10	C-4, C-8, C-10
7	24.3	α: 2.07 (1H, ddd, $J_{H7\alpha,H7\beta}$ = 19.7, $J_{H7\alpha,H6}$ = 4.9, $J_{H7\alpha,H10}$ = 2.0) β: 2.39 (1H, ddt, $J_{H7\beta,H7\alpha}$ = 19.7, $J_{H7\beta,H8}$ = 8.6, $J_{H7\beta,H6} \approx J_{H7\beta,H10} \approx$ 2.7)	α: H-6, H-7β, H-10 β: H-6, H-7α, H-8, H-10	α: C-5, C-6, C-8, C-9, C-14
8	43.0	2.04 (1H, d, $J_{H8,H7\beta}$ = 8.6)	H-7β	C-6, C-7, C-9, C-10, C-13, C-14, C-15, C-19, C-30
9	50.0	-		

10	35.8	3.64 (1H, br p, $J_{H10,H1} \approx J_{H10,H6} \approx J_{H10,H7\alpha} \approx J_{H10,H7\beta} \approx 2.7$)	H-1, H-6, H-7α, H-7β	
11	215.9	-		
12	49.7	α: 3.37 (1H, dq, $J_{H12\alpha,H12\beta} = 14.7$, $J_{H12\alpha,H18} = 1.0$) β: 2.64 (1H, d, $J_{H12\beta,H12\alpha} = 14.7$)	α: H-12β, H-18 β: H-12α	α: C-11, C-13, C-17, C-18 β: C-11, C-13, C-14, C-18,
13	51.5	-		
14	48.9	-		
15	46.6	α: 1.45 (1H, br d, $J_{H15\alpha,H15\beta} = 13.2$) β: 1.87 (1H, dd, $J_{H15\beta,H15\alpha} = 13.2$, $J_{H15\beta,H16} = 8.9$)	α: H-15β, H-16 β: H-15α, H-16	α: C-13, C-14, C-16, C-17, C-30 β: C-8, C-14, C-16, C-30
16	71.5	4.51 (1H, ddd, $J_{H16,H15\beta} = 8.9$, $J_{H16,H17} = 7.3$, $J_{H16,H15\alpha} = 1.1$)	H-15α, H-15β, H-17	C-14, C-17, C-20
17	59.4	2.61 (1H, d, $J_{H17,H16} = 7.3$)	H-16	C-12, C-13, C-16, C-18, C-20, C-21, C-22
18	20.4	0.93 (3H, br s)	H-12α	C-12, C-13, C-14, C-17
19	20.0	0.98 (3H, s)		C-8, C-9, C-10, C-11
20	79.7	-		
21	25.1	1.41 (3H, s)		C-17, C-20, C-22
22	204.8	-		
23	121.1	6.84 (d, $J_{H23,H24} = 15.4$)	H-24	C-22, C-24, C-25, C-26, C-27
24	155.1	6.97 (d, $J_{H24,H23} = 15.4$)	H-23	C-22, C-23, C-25, C-26, C27
25	71.2	-		
26	29.0	1.33 (3H, s)		C-23, C-24, C-25, C-27
27	29.0	1.33 (3H, s)		C-23, C-24, C-25, C-26
28	28.1	1.26 (3H, s)		C-3, C-4, C-5, C-29
29	20.5	1.31 (3H, s)		C-3, C-4, C-5, C-6, C-28
30	18.6	1.42 (3H, br s)		C-8, C-13, C-14, C-15

[a] ^{13}C chemical shift data obtained from HSQC and HMBC spectra. [b] Multiplicities reported as apparent splittings with s = singlet, d = doublet, t = triplet, q = quartet, p = pentet, br = broad with spin-spin coupling constants listed in the order of the reported multiplicity. [c] Correlations from H to indicated C in this column.

Approximately 0.6 mg cucurbitacin I isolated from *Iberis amara* by using HPLC-UV/Vis based fraction collection was dissolved in 650 µl methanol-d_4, and a 1D ^{1}H NMR spectrum was acquired. Supplementary Figure 19 show superimposed 1D ^{1}H NMR spectra of the authentic cucurbitacin I standard (bottom, blue) and the sample isolated from *I. amara* (top, red). This clearly demonstrates the two samples to be identical and thereby that the sample isolated from *I. amara* is cucurbitacin I.



Supplementary Figure 19 Bottom in blue: 1D ^{1}H NMR spectra of the authentic cucurbitacin I standard bought from Sigma-Aldrich. **Top in red**: 1D ^{1}H NMR spectra of the sample isolated from *I. amara*.

References

Abdelkhalek AA, Sharaf A-MM, Rabie M, El-Subbagh HI. 2017. Derivatives of Cucurbitacin-E-glucoside produced by Curvularia lunata NRRL 2178: anti-inflammatory, antipyretic, antitumor activities, and effect on biochemical parameters. Future J Pharm Sci 3:124-130.

Ayyad S-EN, Abdel-Lateff A, Alarif WM, Patacchioli FR, Badria FA, Ezmirly ST. 2012. In vitro and in vivo study of cucurbitacins-type triterpene glucoside from *Citrullus colocynthis* growing in Saudi Arabia against hepatocellular carcinoma. Environ Toxicol Pharmacol 33:245-251.

Balacco G. 1996. A desktop calculator for the Karplus equation. J Chem Inf Comput Sci 36:885-887. Chawech R, Jarraya R, Girardi C, Vansteelandt M, Marti G, Nasri I, Racaud-Sultan C, Fabre N. 2015. Cucurbitacins from the leaves of *Citrullus colocynthis* (L.) Schrad. Molecules 20:18001-18015.

Itkin M, Davidovich-Rikanati R, Cohen S, Portnoy V, Doron-Faigenboim A, Oren E, Freilich S, Tzuri G, Baranes N, Shen S. 2016. The biosynthetic pathway of the nonsugar, high-intensity sweetener mogroside V from *Siraitia grosvenorii*. Proc Natl Acad Sci 113:E7619-E7628.

Seger C, Sturm S, Mair ME, Ellmerer EP, Stuppner H. 2005. 1H and 13C NMR signal assignment of cucurbitacin derivatives from *Citrullus colocynthis* (L.) Schrader and Ecballium elaterium L.(Cucurbitaceae). Magn Reson Chem 43:489-491.

Shang Y, Ma Y, Zhou Y, Zhang H, Duan L, Chen H, Zeng J, Zhou Q, Wang S, Gu W. 2014. Biosynthesis, regulation, and domestication of bitterness in cucumber. Science 346:1084-1088.