Accepted Manuscript

Two new lignan-iridoid glucoside diesters from the leaves of *Vaccinium bracteatum* and their relative and absolute configuration determination by DFT NMR and TDDFT-ECD calculation

Yong-Mei Ren, Chang-Qiang Ke, Attila Mándi, Tibor Kurtán, Chunping Tang, Sheng Yao, Yang Ye

Graphical Abstract

Two new lignan-iridoid glucoside diesters from the leaves of *Vaccinium bracteatum* and their relative and absolute configuration determination by DFT NMR and TDDFT-ECD calculation

Leave this area blank for abstract info.

Yong-Mei Ren, Chang-Qiang Ke, Attila Mándi, Tibor Kurtán, Chunping Tang, Sheng Yao, Yang Ye∗
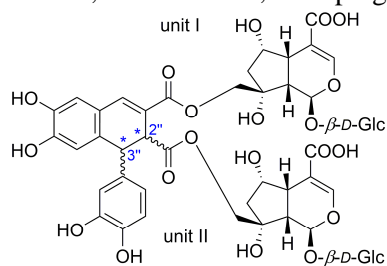
2 2"S, 3"R
3 2"R, 3"S

Two new lignan-iridoid glucoside diesters from the leaves of *Vaccinium bracteatum* and their relative and absolute configuration determination by DFT NMR and TDDFT-ECD calculation

Yong-Mei Ren [a,b,c], Chang-Qiang Ke [a], Attila Mándi [d], Tibor Kurtán [d], Chunping Tang [a], Sheng Yao [a], Yang Ye [a,b,c,∗]

[a]*State Key Laboratory of Drug Research, and Natural Products Chemistry Department, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China*
[b]*School of Life Science and Technology, ShanghaiTech University, Shanghai 201203, China*
[c]*University of Chinese Academy of Sciences, No.19A Yuquan Road, Beijing 100049, China*
[d]*Department of Organic Chemistry, University of Debrecen, P. O. Box 400, H-4002 Debrecen, Hungary*

ARTICLE INFO

ABSTRACT

Two new lignan-iridoid glucoside diesters (**2** and **3**), together with their putative biosynthetic precursor 10-*O-trans*-caffeoyl-6*α*-hydroxyl-dihydromonotropein (**1**), were characterized from the leaves of *Vaccinium bracteatum*. Their planar structures and relative configuration were elucidated by spectroscopic measurements and DFT C-NMR calculations, and their absolute configurations were determined by time-dependent density functional theory (TDDFT) electronic circular dichroism (ECD) calculations. The plausible biosynthetic pathways of new compounds were also proposed.

2009 Elsevier Ltd. All rights reserved.

1. Introduction

Vaccinium bracteatum Thunb. (Ericaceae), known as "Nan Zhu" in Chinese, is an evergreen shrubby tree distributed in mountainous regions of southern China and recognized as an edible and also a medicinal product for use in daily life. Its fruits ("Nan Zhu Zi"), like cranberry and blueberry, can be used as a fruit or beverage material. Its leaves have been used for cooking since the Tang Dynasty, and now it has been a tradition to eat this kind of food during the "Qing Ming Festival" in the south of China. As *V. bracteatum* was reported to possess significant health benefits, such as antifatigue, antianemia, antioxidant and immunomodulate effects,[1] an array of phytochemical investigations have been conducted, revealing the existence of fatty acids, flavonoids and triterpenes as the main chemical components.[2,3] Iridoid glucosides have also been reported as one type of minor constitutes from this plant.[4-6]

Iridoids display an interesting spectrum of biological activity such as cardiovascular,[7,8] antihepatotoxic,[9] antiinflammatory[10,11] and antiviral activities.[12] In order to search for more iridoid compounds, a thorough investigation of the leaves of *Vaccinium*

bracteatum was carried out. Two novel lignan-iridoid glucoside diesters (compounds **2**-**3**, **Fig. 1**), together with their putative biosynthetic precursor (**1**), were isolated and characterized using spectroscopic measurements, DFT C-NMR calculations, and time-dependent density functional theory (TDDFT) electronic circular dichroism (ECD) calculations. In this study, we report the isolation, structural elucidation of these three new compounds, and the proposed biosynthetic pathway as well.

2. Results and discussion

The air-dried leaves of *V. bracteatum* were ground into powder, and then extracted three times with 95% ethanol under ambient temperature to afford a crude extract. The extract was further partitioned in water, and then extracted with petroleum ether (PE) and EtOAc, successively, to give a PE, an EtOAc, and a water soluble fraction. The water soluble fraction was then fractionated by repeated column chromatography (CC) over macroporous resin AB-8 gel, polyamide, octadecyl silane (ODS), Sephadex LH-20, and finally preparative or semipreparative HPLC to yield compounds **1-3**.

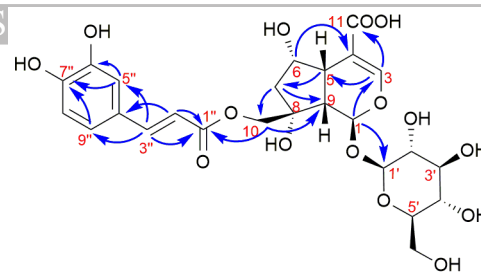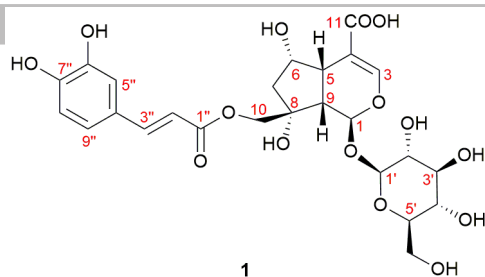∗ Corresponding author. fax: +86 21 50806726; e-mail: yye@mail.shcnc.ac.cn

Fig. 1. Structures of compounds **1–3**.

Fig. 2. Key HMBC (H→C) correlations of **1**.
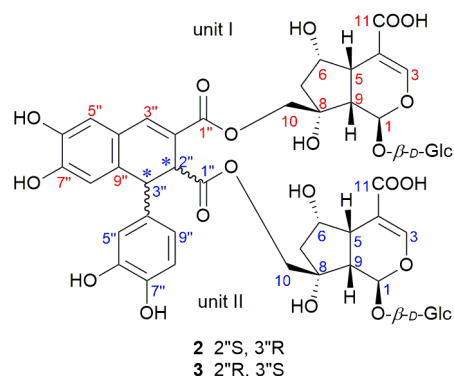
Compound **1** was obtained as yellow amorphous powder and its molecular formula was assigned to be $C_{25}H_{30}O_{15}$ from the pseudo-molecular positive ion at m/z 593.1494 $[M + Na]^+$ (calcd. for $C_{25}H_{30}NaO_{15}$, 593.1477) in the HRESIMS, requiring 11 degrees of unsaturation. The IR spectrum indicated the presence of hydroxyl (3430 cm^{-1}), conjugated carbonyl (1732 cm^{-1}) and aromatic (1636, 1528 cm^{-1}) groups. The ^{1}H and ^{13}C NMR data of **1** (**Tables 1** and **2**) displayed characteristic signals of an iridoid

glucoside.[13-16] The ^{13}C NMR and DEPT spectra revealed 25 carbon resonances including ten ascribed to a 10-carbon iridoid skeleton (δ_C 171.4, 154.4, 110.1, 95.5, 80.1, 76.8, 71.7, 45.8, 45.1, 42.0), six to a glucopyranosyl unit (δ_C 100.4, 78.3, 77.9, 74.6, 71.2, 62.5), and nine to a caffeoyl group[17] (δ_C 169.2, 149.6, 147.4, 146.8, 127.7, 123.1, 116.5, 115.2, 114.9). These signals showed high similarities to those reported for the known compound 10-O-trans-p-coumaroyl-6α-hydroxyl-dihydromonotropein.[5] Their NMR data comparison revealed that singals of a caffeoyl group instead of a p-coumaroyl group were presented in the molecule of **1**. The correlations between the protons resonating at δ_H 4.24, 4.34 (H-10) and the ester carbonyl at δ_C 169.2 (C-1") of the caffeoyl group were observed in the HMBC spectrum (**Fig. 2**), suggesting that the caffeoyl group was attached to C-10 of the iridoid skeleton. The absolute configuration of the glucose was determined as D-glucose according to the method described in the literature[18] (see supporting information). Therefore, the structure of **1** was determined to be 10-O-trans-caffeoyl-6α-hydroxyl-dihydromonotropein.

Compound **2**, obtained as yellow amorphous powder, had a molecular formula of $C_{50}H_{58}O_{30}$ with 22 double-bond equivalents

Table 1
^{1}H NMR Data for Compounds **1–3**.

no.	1[a]	2[b]		3[b]	
		Unit I	Unit II	Unit I	Unit II
1	5.67 (d, 3.3)	5.55 (d, 4.1)	5.46 (d, 4.2)	5.59 (d, 3.7)	5.46 (d, 3.9)
3	7.55 (br s)	7.52 (br s)	7.48 (br s)	7.51 (br s)	7.48 (br s)
5	2.91 (dd, 9.6, 4.0)	2.85 (dd, 9.7, 4.6)	2.67 (dd, 10.0, 5.4)	2.87 (dd, 9.7, 4.5)	2.74 (dd, 10.0, 5.0)
6	4.38 (q, 5.3)	4.29 (m)	4.18 (m)	4.31 (q, 5.6)	4.22 (m)
7	1.95 (dd, 13.8, 5.6) 2.04 (dd, 13.8, 5.6)	1.92 (dd, 13.7, 6.7) 2.02 (dd, 13.7, 5.7)	1.72 (dd, 13.4, 7.3) 1.89 (dd, 13.4, 5.9)	1.91 (dd, 13.6, 6.0) 2.02 (dd, 13.6, 5.7)	1.75 (dd, 13.6, 7.4) 1.91 (dd, 13.6, 6.0)
9	2.68 (dd, 9.6, 3.3)	2.55 (dd, 9.7, 4.1)	2.24 (dd, 10.0, 4.2)	2.60 (dd, 9.7, 3.7)	2.34 (dd, 10.0, 3.9)
10	4.24 (d, 11.2) 4.34 (d, 11.2)	4.18 (d, 11.1) 4.30 (d, 11.1)	4.03 (d, 11.1) 4.13 (d, 11.1)	4.21 (d, 11.2) 4.27 (d, 11.2)	4.04 (d, 11.1) 4.14 (d, 11.1)
1'	4.71 (d, 7.9)	4.69 (d, 7.9)	4.66 (d, 7.9)	4.68 (d, 7.9)	4.64 (d, 7.9)
2'	3.23 (t, 8.3)	3.24 (m)	3.25 (m)	3.26 (m)	3.24 (m)
3'	3.38 (m)	3.35 (m)	3.36 (m)	3.36 (m)	3.34 (m)
4'	3.32 (m)	3.32 (m)	3.35 (m)	3.32 (m)	3.34 (m)
5'	3.31 (m)	3.30 (m)	3.30 (m)	3.32 (m)	3.28 (m)
6'	3.68 (dd, 12.0, 4.7) 3.86 (dd, 12.0, 1.8)	3.65 (m) 3.82 (m)	3.61 (m) 3.79 (m)	3.62 (m) 3.81 (m)	3.64 (m) 3.78 (m)
2"	6.35 (d, 15.9)		3.99 (d, 3.3)		3.98 (d, 3.4)
3"	7.63 (d, 15.9)	7.71 (s)	4.45 (d, 3.3)	7.74 (s)	4.46 (d, 3.4)
5"	7.10 (d, 2.1)	6.88 (s)	6.46 (d, 2.2)	6.88 (s)	6.45 (m)
8"	6.81 (d, 8.2)	6.57 (s)	6.63 (d, 8.2)	6.55 (s)	6.64 (d, 8.7)
9"	7.00 (dd, 8.3, 2.1)		6.42 (dd, 8.2, 2.2)		6.45 (m)

[a] Recorded in methanol-d_4 at 500 MHz, [b] Recorded in methanol-d_4 at 600 MHz, δ in ppm, J in Hz.

Table 2
^{13}C NMR Data for Compounds **1**–**3**.

no	1	2 Unit	2 Unit	3 Unit	3 Unit
1	95.5	95.6	95.6	95.4	95.5
3	154.4	154.2	154.0	154.0	153.8
4	110.1	110.5	110.8	110.6	111.3
5	42.0	42.3	42.2	42.2	42.2
6	76.8	77.1	77.3	77.1	77.4
7	45.1	45.1	44.9	45.1	45.1
8	80.1	80.0	79.8	80.1	79.7
9	45.8	45.7	45.0	45.7	45.3
10	71.7	71.6	70.5	71.8	70.5
11	171.4	171.8	172.0	171.8	172.1
1′	100.4	100.6	100.4	100.4	100.1
2′	74.6	74.5	74.5	74.5	74.5
3′	77.9	77.9	77.9	77.9	77.9
4′	71.2	71.3	71.3	71 .3	71.3
5′	78.3	78.3	78.3	78.3	78.3
6′	62.5	62.6	62.6	62.6	62.6
1″	169.2	168.6	174.5	168.6	174.4
2″	114.9	122.4	49.4	122.3	49.2
3″	147.4	140.4	46.9	140.5	46.9
4″	127.7	124.9	135.7	124.8	135.8
5″	115.2	117.3	115.9	117.5	115.9
6″	146.8	145.7	146.1	145.7	146.1
7″	149.6	149.3	145.1	149.4	145.1
8″	116.5	117.5	116.4	117.3	116.4
9″	123.1	131.3	120.1	131.4	120.1

Recorded in methanol-d_4 at 125 MHz, δ in ppm.

(DBEs) as determined by the HRESIMS. The IR spectrum showed absorption bands for hydroxyl (3428 cm^{-1}), conjugated carbonyl (1730 cm^{-1}) and aromatic (1640, 1526 cm^{-1}) functionalities. Analysis of the ^{1}H and ^{13}C NMR data with the aid of DEPT experiment revealed the presence of two sets of iridoid glucoside signals, similar to those of **1** (**Tables 1** and **2**). The remaining signals included those of two doublet methine protons (δ_H 3.99 and δ_H 4.45) and a singlet olefinic proton (δ_H 7.71), corresponding to two methine carbons (δ_C 49.4 and δ_C 46.9) and one olefinic carbon (δ_C 140.4). In addition, signals at δ_H 6.63 (1H, d, J = 8.2), 6.46 (1H, d, J = 2.2) and 6.42 (1H, dd, J = 8.2, 2.2), indicative of an aromatic ring with a typical ABX spin system, and two singlet signals at δ_H 6.88 and 6.57, corresponding to carbons at 117.3 and 117.5 and assigned as two para-positioned aromatic protons of another aromatic ring, were observed. The HMBC correlations (**Fig. 3**) from δ_H 3.99 (H-2″, unit II) to δ_C 168.6 (C-1″, unit I), 140.4 (C-3″, unit I), 135.7 (C-4″, unit II) , and 131.3 (C-9″, unit I), and from δ_H 4.45 (H-3″, unit II) to δ_C 174.5 (C-1″, unit II), 124.9 (C-4″, unit I), 122.4 (C-2″, unit I), 115.9 (C-5″, unit II), and 117.5 (C-8″, unit I), and from δ_H 7.71 (H-3″, unit I) to δ_C 117.3 (C-5″, unit I), along with the assignment information aforementioned, revealed the presence of a lignan group.[19,20] This lignan group and the iridoid glucoside moieties was finally connected by ester groups, which was deduced from the HMBC correlations of the methylene protons (H-10) to the carbonyl carbon (C-1″). Therefore, compound **2** was established as a lignan-iridoid glycoside diester.
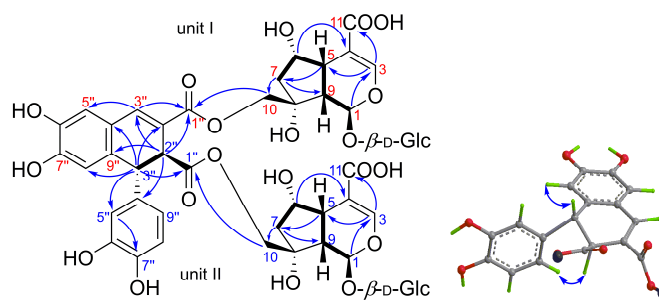

Fig. 3. ^{1}H–^{1}H COSY (▬), key HMBC (H→C) and ROESY (H↔H) correlations of compound **2**.

The moderate $^3J_{2''\text{-}H,3''\text{-}H}$ coupling constant (3.3 Hz) of **2** suggested either cis[21-23] relative configuration of the C-2″ and C-3″ substituents provided that one of them is equatorial, while the other is axial or trans[19,20,24] relative configuration provided that both of them is axial. However, correlations between H-2″ (unit II) and H-9″ (unit II), and between H-3″ and H-8″ in the ROESY spectrum (**Fig. 3**) suggested a trans relative configuration between H-2″ (unit II) and H-3″ (unit II).[25,26] To further confirm both the relative and the absolute configuration of those two chirality centers of the lignan moiety, ECD and C-NMR calculations were performed on a model compound (vide infra).

Compound **3** was obtained as yellow amorphous powder. The HRESIMS indicated a molecular formula of $C_{50}H_{58}O_{30}$ (m/z 1137.2948, calcd for $C_{50}H_{57}O_{30}$, 1137.2940), the same as that of **2**. Its IR spectrum indicated the existence of hydroxyl (3424 cm^{-1}), conjugated carbonyl (1726 cm^{-1}) and aromatic (1638, 1527cm^{-1}) groups. The NMR data of **3** (**Tables 1** and **2**) showed high similarities to those of compound **2**, suggesting that the molecule of **3** also contained two iridoid glucoside moieties and a lignan moiety. A detailed comparison of 1D and 2D NMR data of these two compounds further confirmed that these two compounds shared the same planar structure and relative configuration, and only differed at the stereochemistry of C-2″ and C-3″ of unit II.
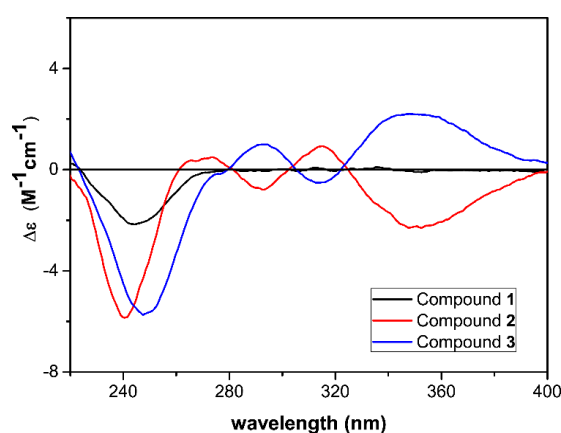

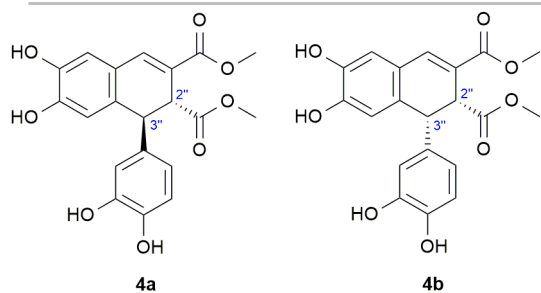**Fig. 4.** Experimental ECD spectra of compounds **1**–**3** recorded in MeOH.

Fig. 5. Structure of truncated model compounds **4a** and **4b**.

In order to determine the relative and the absolute configuration of C-2" and C-3" of the lignan moiety, first the ECD spectra of **2** and **3** in MeOH were recorded, showing opposite Cotton effects (CEs) at 293, 315 and 350 nm and the same negative CEs around 245 nm (**Fig. 4**). Due to the high conformational flexibility and large molecular weight of **2** and **3**, a large number of low-energy conformational isomers are expected in the conformational search, the optimization and ECD calculation of which would impose high uncertainty in the determination of the absolute configuration.[27,28] Since the three high-wavelength transitions (above 280 nm) of **2** and **3** could be attributed to the trisubstituted 1,2-dihydronaphthalene unit and the iridoid moieties contain only α,β-unsaturated carboxylic acid chromophores, the truncated model compounds **4a** and **4b** with *trans* and *cis* relative configurations, respectively, could be used for the ECD calculations, in which the complex ester groups are simplified to methyl esters (**Fig. 5**).[27,29] The experimental ECD spectrum of **1** had a single negative ECD transition at *ca.* 245 nm justifying the truncation (**Fig. 4**).

Thus the conformationally flexible iridoid part was truncated and model compounds **4a** and **4b** with a (2"*R*, 3"*S*) and a (2"*R*, 3"*R*) absolute configurations were selected for conformational analysis and ECD calculations. The Merck Molecular Force Field (MMFF) conformational search of the model compound **4a** resulted in 46 low-energy conformers within a 21 kJ/mol energy window and 73 conformers for **4b**. These conformers were reoptimized at the B3LYP/6-31G(d) level *in vacuo*, the B97D/TZVP[30,31] PCM/MeOH and the CAM-B3LYP/TZVP[32,33] PCM/MeOH levels providing eight low-energy conformers ($\geq$ 1%) at each level for **4a** (**Fig. 6**) and 14, 16 and 16 low-energy conformers for **4b** (**Fig. 7**). It is interesting to note that both substituents of **4a** (C-2" and C-3") adopted axial orientation in all the low-energy conformers, while for **4b** the
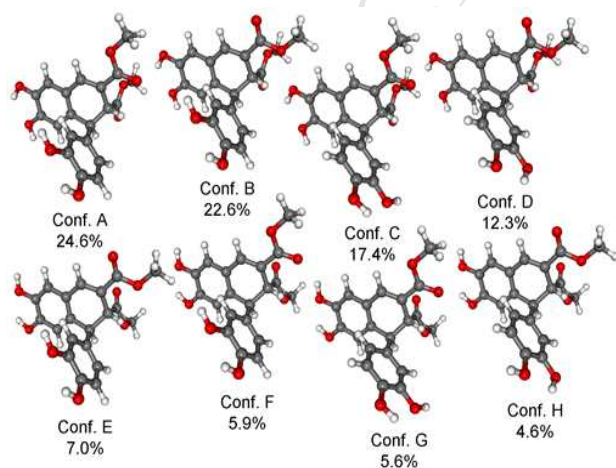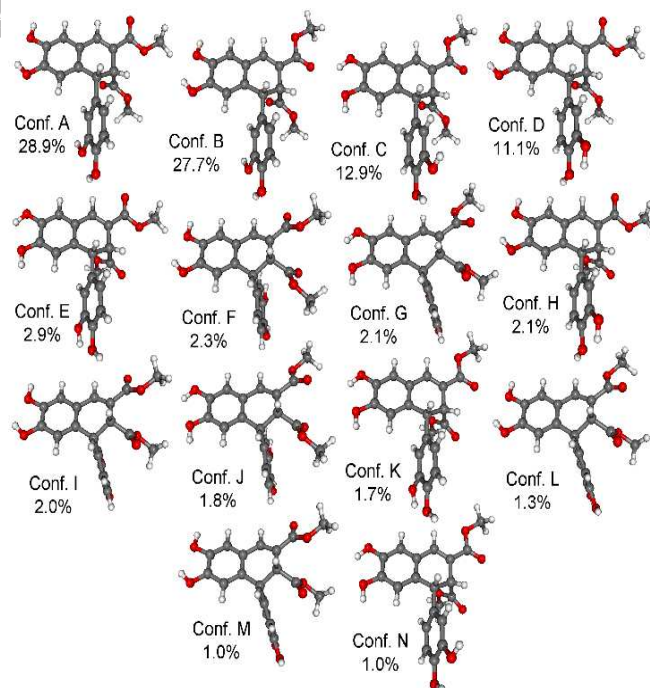

Fig. 7. Structure and population of the low-energy B3LYP/6-31G(d) conformers of **4b**.

C-2" substituent had axial orientation and the C-3" one equatorial in most of the conformers (*ca.* 88.2% overall population *vs.* 10.6% for equatorial C-2" and axial C-3"). With these orientation of C-2" and C-3" substituents, the small value of $3J_{2"-H,3"-H}$ and NOE correlations did not allow distinguishing the *cis* and *trans* relative onfiguration. ECD spectra computed at various levels for all sets of conformers for both *trans* and *cis* model compounds of **4a** (**Fig. 8**) and **4b** (**Fig. 9**) reproduced the main features of the experimental spectrum of **3** down to *ca.* 235 nm (the highest-energy experimental 247 nm transition is an overlap of the core part and the truncated parts) with the best agreement at the PBE0/TZVP level on the gas-phase conformers. Although relative configuration of C-2" and C-3" could not be determined on the basis of the above ECD results, the consistent results obtained at various levels for both diastereomers allowed the elucidation of the absolute configuration of C-2" of the 1,2-dihidronaphthalene moiety as (*R*) in **3** and (*S*) in **2**.

Fig. 6. Structure and population of the low-energy B3LYP/6-31G(d) conformers of **4a**.

Fig. 8. Experimental ECD spectrum of **3** compared with the Boltzmann-weighted PBE0/TZVP ECD spectrum of model compound **4a** computed for the B3LYP/6-31G(d) conformers. Bars represent the rotational strengths of conformer A.
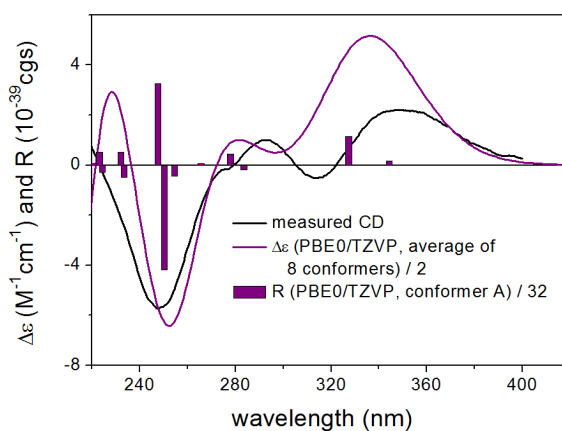
Fig. 9. Experimental ECD spectrum of **3** compared with the Boltzmann-weighted PBE0/TZVP ECD spectrum of model compound **4b** computed for the B3LYP/6-31G(d) conformers. Bars represent the rotational strengths of conformer A.
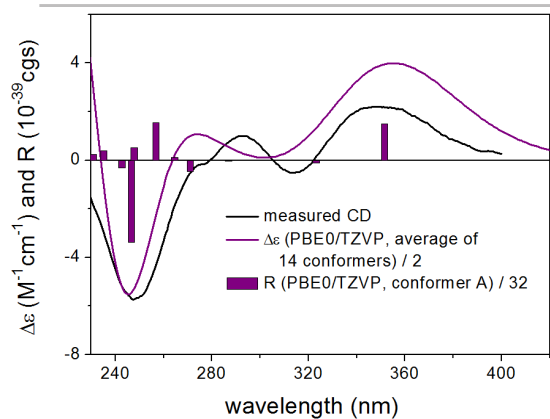
In order to determine the relative configuration of the two chirality centers and hence the absolute configuration, C-NMR calculations were performed on the same model compounds at the mPW1PW91/6-311+G(d,p) level[34] computing for the low-energy B3LYP/6-31+G(d,p) reoptimized MMFF conformers. **Table 3** summarizes the results; from 11 relevant carbons (including ring B and the first connecting carbon atoms) in the vicinity of the two chirality centers, 9 suggested *trans* and only 2 *cis* relative configuration. On the basis of the deviations and average errors, the *trans* relative configuration could be determined and thus (2"*S*,3"*R*)-**2** and (2"*R*,3"*S*)-**3** absolute configurations could be elucidated.[41,43,44] The C-NMR results are in line with the computed *J* values, which also indicated *trans* relative configuration (see SI for details) in line with the ROESY correlations making elucidation of the relative and absolute configuration solid. Based on the above results, it is advisable to similarly prove the *cis* relative configuration of other tetrahydronaphthalene lignans having moderate *J* values.

Table 3
Comparison of the computed C-NMR data of 11 relevant carbons in the vicinity of the two chirality centers of model compounds **4a** and **4b** with the experimental ^{13}C-NMR data. Average $\Delta\delta_{trans}$ = 1.20 while average $\Delta\delta_{cis}$ = 2.24 suggesting *trans* relative configuration.

carbon	exp (**3**)	calcd **4a** (*trans*)	calcd **4b** (*cis*)	$\Delta\delta_{trans}$	$\Delta\delta_{cis}$
UI-1"	168.6	167.51	167.36	1.09	1.24
UI-2"	122.3	123.91	125.16	1.61	2.86
UI-3"	140.5	139.58	141.98	0.92	1.48
UI-4"	124.8	127.45	126.42	2.65	1.62
UI-5"	117.5	114.85	114.29	2.65	3.21
UI-8"	117.3	114.62	115.05	2.68	2.25
UI-9"	131.4	132.15	135.16	0.75	3.76
UII-1"	174.4	174.64	173.79	0.24	0.61
UII-2"	49.2	49.15	48.44	0.05	0.76
UII-3"	46.9	47.26	50.49	0.36	3.59
UII-4"	135.8	135.95	132.49	0.15	3.31
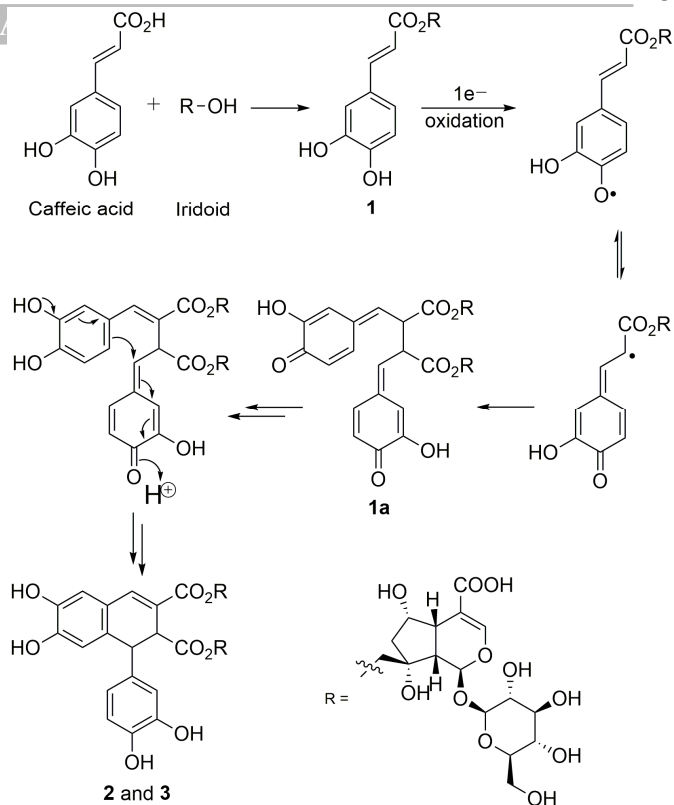
Scheme 1. Proposed biosynthetic pathways for compounds **2** and **3**.

On the basis of literatures and our findings, the biosynthetic pathways of **2** and **3** could be proposed (Scheme 1). In brief, the mechanism might involve an esterification reaction between caffeic acid and iridoid to produce compound **1**. One-electron oxidation, resonance hybridization and dimerization of **1** results in bimolecular radical coupling to yield **1a**.[35,36] Intermediate **1a** underwent a rearrangement reaction, and then further constructed the bicyclic ring system to afford **2** and **3**.

3. Conclusions

In summary, three novel iridoid glucoside derivatives were isolated and characterized from the leaves of *V. bracteatum*. 10-*O*-*trans*-caffeoyl-6α-hydroxyl-dihydromonotropein (**1**) is the first iridoid glucoside from the genus *Vaccinium* that contains a caffeoyl group instead of a *p*-coumaroyl group while compounds **2** and **3** represent a new type of compounds that is lignan-iridoid glucoside diesters. Furthermore, for the first time, DFT NMR and TDDFT ECD calculation was applied to the relative and absolute configuration determination of this kind of lignan derivatives.

4. Experimental Section

4.1. General experimental procedures

TLC was carried out on precoated silica gel 60 F254 25 Aluminium sheets (Merck KGaA, Darmstadt, Germany) and the TLC spots were viewed at 254 nm and visualized with 5% H_2SO_4 in EtOH containing 10 mg/mL vanillin. Macro porous resin AB-8 gel (Shandong Lu Kang Chemical Industrials, Jinan, Shandong, China), ODS (YMC Co., Ltd., Japan), and Sephadex LH-20 (Pharmacia Biotech AB, Uppsala, Sweden) were used for column chromatography (CC). Analytical HPLC was applied on a Waters 2695 instrument (Milford, MD, USA) coupled with a 2998 PDA, a Waters 2424 ELSD, and a Waters 3100 MS detector. Preparative HPLC was performed on a Varian PrepStar

instrument with an Alltech 3300 ELSD detector (Columbia, MD, USA) using a Waters Sunfire RP C18, 5 μm, 30 × 150 mm column. Semipreparative HPLC was performed on a Waters 2690 instrument (Milford, MD, USA) coupled with a 996 photodiode array detector using a YMC-pack RP C18, 5 μm, 10 × 250 mm column. Optical rotations were measured on a Rudolph Autopol VIAutomatic polarimeter (Hackettstown, NJ, USA). IR spectra were recorded on a Nicolet Magna FT-IR 750 spectrophotometer (Waltham, MA, USA) using KBr disks. ECD spectra were recorded on a JASCO J-810 spectrometer. ESIMS and HRESIMS data were recorded on Waters 2695-3100 LC-MS and Agilent G6520 Q-TOF mass spectrometers (Santa Clara, CA, USA), respectively. NMR spectra were recorded on a Bruker Avance III (Bruker, Zurich, Switzerland) for 500 and 600 M NMR spectrometer with TMS as internal standard. The chemical shift (δ) values were given in ppm and coupling constants (J) are in Hz. All solvents used for CC were of at least analytical grade (Shanghai Chemical Reagents Co., Ltd., Shanghai, China), and solvents used for HPLC were of HPLC grade (Merck KGaA, Darmstadt, Germany).

4.2. Plant material

The leaves of *V. bracteatum* were collected in Jiangsu Province, China, in 2015, and identified by Professor Jin-Gui Shen from Shanghai Institute of Materia Medica. A voucher specimen (No. 20150926) was deposited at the Herbarium of the Shanghai Institute of Materia Medica, Chinese Academy of Sciences.

4.3. Extraction and isolation

The air-dried leaves of *V. bracteatum* (20 kg) were ground into powder, and then extracted three times with 95% ethanol at room temperature to afford a crude extract (1.5 kg). The extract was further partitioned in water, and then extracted with petroleum ether (PE) and EtOAc, successively, to give a PE, an EtOAc, and a water soluble fraction. The water soluble fraction (Fr. A) was then fractionated by a column chromatography (CC) over macroporous resin AB-8 gel (EtOH/H$_2$O, from 30 to 95%), yieling fractions A1–A3. Fraction A1 (200 g) was then separated on polyamide (MeOH/H$_2$O, from 20 to 95%) to give four subfractions (A1A-A1D). Then subfraction A1A was subjected to CC over octadecyl silane (ODS) (MeOH/H$_2$O, from 10 to 45%) to afford fractions A1A1-A1A10. Fraction A1A1 (180.0 mg) was applied to preparative HPLC (MeOH/H$_2$O, from 5 to 20%, containing 0.2% formic acid; 0-120 min, 25 mL/min) and then semipreparative HPLC (MeCN/H$_2$O, from 9 to 14%, containing 0.2% formic acid; 0-60 min, 3 mL/min) to afford compounds **2** (3.2 mg) and **3** (2.4 mg) as trace components. Fraction A1A4 (2.0 g) was chromatographed on Sephadex LH-20 (MeOH) to yield subfractions A1A4C (800 mg). Finally, compound **1** (115 mg) was obtained from fraction A1A4C by preparative HPLC (MeCN/H$_2$O, from 10 to 28%, containing 0.2% formic acid; 0-120 min, 25 mL/min).

4.3.1. Determination of sugar configuration

Compound **1** (11.4 mg) and cellulase (11.4 mg) were dissolved in HOAc – NaOAc buffered solution (PH = 4.5, 2 mL) and stirred at room temperature for a week. The reaction mixture was evaporated by rotary evaporator, and then dissolved in pyridine (2 mL) containing L-cysteine methyl ester hydrochloride (2 mg) and heated at 60 °C for 60 min. Then *o*-tolyl isothiocyanate (5 μL) was added to the mixture and heated at 60 °C for 60 min. After evaporation of the solvent, the residue was dissolved in methanol, and then analyzed by LCMS. The authentic D/L-glucose samples were treated with the same method

aforementioned, and analyzed by LCMS, too. The retention time of the derivatives of compound **1**, D-glucose and L-glucose were 13.13, 13.17 and 12.42 min, respectively. Therefore, the sugar moiety of compound **1** was determined as D-glucose (details see Supporting Information).

4.3.2. Compound characteristics

Compound 1

yellow amorphous powder; [α]$_D^{20}$ –69.7 (c = 1.5, MeOH); UV (MeOH) λ_{max} (log ε) 324 (4.06), 219 (4.15) nm; ECD (MeOH) ($\Delta\varepsilon$) 244 (–2.16) nm; IR (KBr, cm^{-1}): ν_{max} 3430, 1732, 1636, 1528 cm^{-1}; ^{1}H and ^{13}C NMR data, see **Tables 1** and **2**; ESIMS m/z 569 [M – H]$^-$; HRESIMS m/z 593.1494 [M + Na]$^+$ (calcd. For C$_{25}$H$_{30}$NaO$_{15}$, 593.1477).

Compound 2

yellow amorphous powder; [α]$_D^{20}$ –72.5 (c = 0.13, MeOH); UV (MeOH) λ_{max} (log ε) 342 (2.80), 315 (2.73), 290 (2.66), 230 (3.25) nm; ECD (MeOH) ($\Delta\varepsilon$) 348 (–2.30), 315 (0.93), 293 (–0.79), 240 (–5.85), nm; IR (KBr, cm^{-1}): ν_{max} 3428, 1730, 1640, 1526 cm^{-1}; ^{1}H and ^{13}C NMR data, see **Tables 1** and **2**; ESIMS m/z 1137 [M – H]$^-$; HRESIMS m/z 1137.2955 [M – H]$^-$ (calcd. for C$_{50}$H$_{57}$O$_{30}$, 1137.2940).

Compound 3

yellow amorphous powder; [α]$_D^{20}$ +11.4 (c = 0.098, MeOH); UV (MeOH) λ_{max} (log ε) 344 (2.64), 312 (2.62), 292 (2.57), 230 (3.07) nm; ECD (MeOH) ($\Delta\varepsilon$) 348 (2.21), 314 (–0.52), 293 (1.00), 247 (–5.74) nm; IR (KBr, cm^{-1}): ν_{max} 3424, 1726, 1638, 1527 cm^{-1}; ^{1}H and ^{13}C NMR data, see **Tables 1** and **2**; ESIMS m/z 1137.6 [M – H]$^-$; HRESIMS m/z 1137.2948 [M – H]$^-$ (calcd for C$_{50}$H$_{57}$O$_{30}$, 1137.2940).

4.4. Computational section

Mixed torsional/low-frequencymode conformational searches were carried out by means of the Macromodel 10.8.011 software using the Merck Molecular Force Field (MMFF) with an implicit solvent model for CHCl$_3$.[37] Geometry reoptimizations were carried out at the B3LYP/6-31G(d) level in vacuo, the B3LYP/6-31+G(d,p) level in vacuo, the B97D/TZVP[30,31] and the CAMB3LYP/TZVP[32,33] levels with the PCM solvent model for MeOH. DFT optimized geometries were clustered for all non-hydrogen atoms. TDDFT ECD calculations were run with various functionals (B3LYP, BH&HLYP, CAM-B3LYP, PBE0) and the TZVP basis set as implemented in the Gaussian 09 package with the same or no solvent model as in the preceding DFT optimization step.[38] NMR calculations were performed at the mPW1PW91/6-311+G(2d,p) level.[34] ECD spectra were generated as sums of Gaussians with 2400 and 3000 cm^{-1} widths at half-height (corresponding to ca. 15 and 19 nm at 250 nm), using dipole-velocity-computed rotational strength values.[39] Computed NMR data were corrected with I = 185.4855 and S = -1.0306.[40,41] Boltzmann distributions were estimated from the ZPVE-corrected B3LYP/6-31G(d) energies in the B3LYP/6-31G(d) gas-phase calculations, and from the uncorrected B3LYP/6-31+G(d,p), B97D/TZVP and CAM-B3LYP/TZVP energies in the other cases. The MOLEKEL software package was used for visualization of the results.[42]

References

1. Chen SK, Chen BY, Li H, eds. *Flora of China* (*Zhongguo Zhiwu Zhi*). Vol. 57. Beijing: Science Press; 1999; 107.
2. Xu Y, Wang L, Chen Z, Chen Y. *Science and Technology of Food Industry (Shipin Gongye Keji).* 2013; 34: 372-376.
3. Yao HY, Wang L, Chen ZX. *Chemistry and Industry of Forest Product (Linchan Huaxue Yu Gongye).* 2007; 27: 121-123.
4. Sakakibara J, Kaiya T, Yasue M. *Chem Pharm Bull.* 1971; 19: 1979.
5. Qu J, Chen X, Niu CS, Yu SS. *China J Chin Mater Med.* 2014; 39: 684-688.
6. Li ZL, Zhou WM. *China J Chin Mater Med.* 2008; 33: 2087-2089.
7. Hansen K, Adsersen A, Christensen SB, Jensen SR, Nyman U, Smitt UW. *Phytomedicine.* 1996; 2: 319-325.
8. Petkov V, Manolov P. *Am J Chin Med.* 1978; 6: 123-130.
9. Dhawan, BN. *Med Chem Res.* 1995; 5: 595-605.
10. María del Carmen Recio, Giner RM, Manez S, Ríos JL. *Planta Med.* 1994; 60; 232-234.
11. Choi J, Lee KT, Choi MY, Nam JH, Jung HJ, Park SK, Park HJ. *Biol Pharm Bull.* 2005; 28: 1915-1918.
12. Rathore A, Rivastava V, Srivastava K. *Phytochemistry.* 1990; 29: 1917-1920.
13. Huang XZ, Wang YH, Yu SS, Fu GM, Hu YC, Liu Y, Fan LH. *J Nat Prod.* 2005; 68: 1646-1650.
14. Turner A, Chen SN, Nikolic D, Farnsworth NR, Pauli GF. *J Nat Prod.* 2007; 70: 253-258.
15. Jensen HD, Krogfelt KA, Cornett C, Hansen SH, Christensen SB. *J Agric Food Chem.* 2002; 50: 6871-6874.
16. Taskova RM, Kokubun T, Ryan KG, Jensen SR. *J Nat Prod.* 2011; 74: 1477-1483.
17. Taskova RM, Kokubun T, Ryan KG, Garnock-Jones PJ, Jensen SR. *J Nat Prod.* 2011; 74:1477-83.
18. Tanaka T, Nakashima T, Ueda T, Tomii K, Kouno I. *Chem Pharm Bull.* 2007; 55: 899-901.
19. Martini U, Zapp J, Becker H. *Phytochemistry.* 1998; 49: 1139-1146.
20. Cullmann F, Becker H. *Phytochemistry.* 1999; 52: 1651-1656.
21. Provan GJ, Waterman PG. *Planta Med.* 1985; 51: 271-272.
22. Fonseca SF, Rúveda EA, McChesney JD. *Phytochemistry.* 1980; 19: 1527-1530.
23. Takei Y, Mori K, Matsui M. *Agric Biol Chem.* 1973; 37: 637-641.
24. Scher JM, Zapp J, Becker H. *Phytochemistry.* 2003; 62: 769-777.
25. Agata I, Hatano T, Nishibe S, Okuda T. *Phytochemistry.* 1989; 28: 2447-2450.
26. Bai H, Li W, Koike K, Dou DQ, Pei YP, Chen YJ, Nikaido T. *Chem Pharm Bull.* 2003; 51: 1095-1097.
27. Mándi A, Mudianta IW, Kurtán T, Garson MJ. *J Nat Prod.* 2015; 78: 2051-2056.
28. Mándi A, Swamy MM, Taniguchi T, Anetai M, Monde K. *Chirality.* 2016; 28: 453-459.
29. Hou XF, Yao S, Mándi A, Kurtán T, Tang CP, Ke CQ, Li XQ, Ye Y. *Org Lett.* 2012; 14: 460-463.
30. Grimme S. *J Comput Chem.* 2006; 27: 1787–1799.
31. Sun P, Xu DX, Mándi A, Kurtán T, Li TJ, Schulz B, ZhangW. *J Org Chem.* 2013; 78: 7030-7047.
32. Yanai T, Tew DP, Handy NC. *Chem Phys Lett.* 2004; 393: 51-57.
33. Pescitelli G, Di Bari L, Berova N. *Chem Soc Rev.* 2011; 40: 4603-4625.
34. Adamo C, Barone V. *J Chem Phys.* 1998; 108: 664-675.
35. Davin LB, Wang HB, Crowell AL, Bedgar DL, Martin DM, Sarkanen S, Lewis NG. *Science.* 1997; 275: 362-367.
36. Katayama T, Masaoka T, Yamada H. *Mokuzai Gakkaishi.* 1997;43: 580-588.
37. MacroModel, Schrödinger, LLC, 2015, http://www.schrodinger.com/MacroModel.
38. Frisch MJ, Trucks GW, Schlegel HB, et al. Gaussian 09, revision B.01; Gaussian, Inc.: Wallingford, CT, 2010.
39. Stephens PJ, Harada N. *Chirality.* 2010; 22: 229-233.
40. CHESHIRE CCAT, the Chemical Shift Repository for computed NMR scaling factors, http://cheshirenmr.info/index.htm.
41. Lodewyk MW, Siebert MR, Tantillo DJ. *Chem Rev.* 2012; 112: 1839-1862.
42. Varetto U. MOLEKEL, v. 5.4; Swiss National Supercomputing Centre: Manno, Switzerland, 2009.
43. Sun P, Yu Q, Li J, Riccio R, Lauro G, Bifulco G, Kurtán T, Mándi A, Tang H, Li TJ, Zhuang CL, Gerwick WH, Zhang W. *J. Nat. Prod.* 2016; 79: 2552-2558.
44. Qiu S, De Gussem E, Tehrani KA, Sergeyev S, Bultinck P, Herrebout W. *J. Med. Chem.* 2013; 56: 8903-8914.