Organic & Biomolecular Chemistry

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/obc

Organic & Biomolecular Chemistry RSCPublishing

ARTICLE

Anti Hepatitis B Virus Activities and Absolute Configurations of Sesquiterpenoid Glycosides from *Phyllanthus emblica*

Cite this: DOI:

Received,
Accepted

DOI:

www.rsc.org/

Jun-Jiang Lv,†[a,b] Ya-Feng Wang,†[c] Jing-Min Zhang,[c,d] Shan Yu,[a,b] Dong Wang,[a] Hong-Tao Zhu,[a] Rong-Rong Cheng,[a] Chong-Ren Yang,[a] Min Xu,[a,*] and Ying-Jun Zhang[a,*]

During the process exploring anti-viral compounds from *Phyllanthus* species, eight new highly oxygenated bisabolane sesquiterpenoid glycosides phyllaemblicins G1-G8 (**1**-**8**) were isolated from *Phyllanthus emblica*, along with three known compounds, phyllaemblicin F (**9**), phyllaemblic acid (**10**) and glochicoccin D (**11**). Phyllaemblicin G2 (**2**), bearing a tricyclo [3.1.1.1] oxygen bridge ring system, is an unusual sesquiterpenoid glycoside, while phyllaemblicins G6-G8 (**6**-**8**) are dimeric sesquiterpenoid glycosides with two norbisabolane units connecting through a disaccharide. All the structures were elucidated by means of extensive analysis of HRMS and NMR data. The relative configuration of phyllaemblicin G2 was constructed based on heteronuclear coupling constants measurement, and the absolute configurations for all new compounds were established by calculated electronic circular dichroism (ECD) using time dependent density functional theory. The sesquiterpenoid glycoside dimers **6 - 9** displayed potential anti-hepatitis B virus (HBV) activities, especially for the new compound **6** with IC_{50} of 8.53 ± 0.97 and 5.68 ± 1.75 μM, respectively, towards the HBV surface antigen (HBsAg) and HBV excreted antigen (HBeAg) secretion.

Introduction

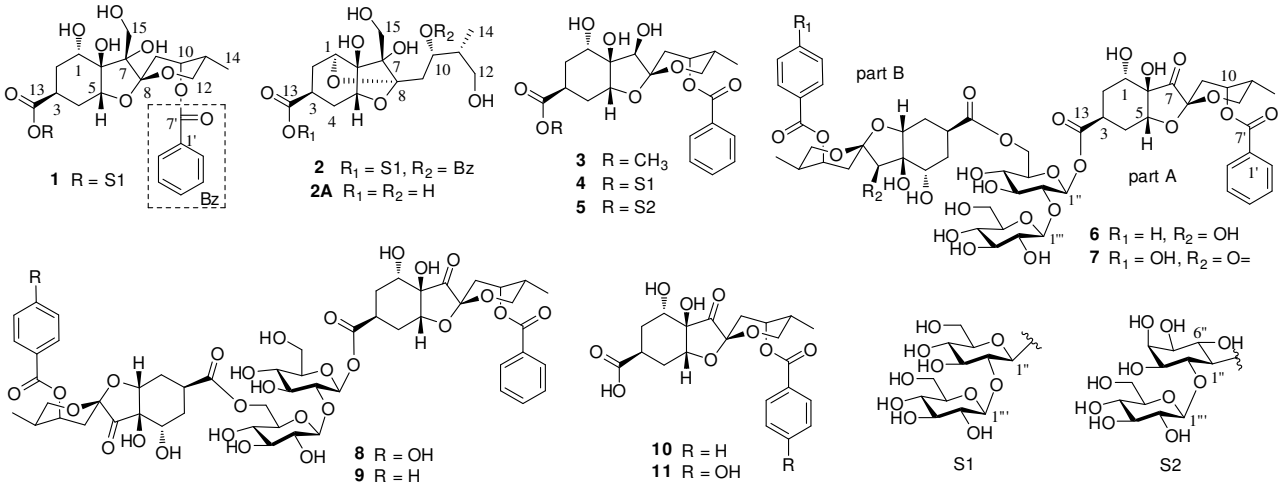
Despite the effective vaccine program's available, hepatitis B virus (HBV) infection is still a major world health problem. About 400 million people in the world suffer from HBV infection.[1] The present approved HBV chemotherapies are all nucleoside or nucleoside based polymerase inhibitors,[2] and the resistance to these drugs stimulated chemists to explore new anti-HBV agents with novel mechanisms. High levels of HBV surface antigen (HBsAg) bearing subviral particles in the serum of chronically infected individuals play a role in suppressing the HBV immune response, and current therapeutics are not directed at reducing this antigenemia.[3-5]

Phyllanthus emblica Linn, an euphorbiaceae plant, is an important traditional medicinal plant in China. In our previous studies, nine bisabolane and norbisabolane sesquiterpenoid glycosides were isolated and reported: phyllaemblic acid,[6] phyllaemblic acids B and C,[7] and phyllaemblicins A-F.[8-9] Among them, phyllaemblicin B exhibited significant anti-coxsackie virus B3 activity.[9] In order to explore novel antiviral lead compounds from *P. emblica*, a further chemical study was carried out, leading to the isolation of eight new highly oxygenated bisabolane sesquiterpenoid glycosides (**1**-**8**) from the roots of the titled plant, along with three known ones (**9**-**11**). Their structures were elucidated by means of HRMS and NMR experiments, and the absolute configurations were established by calculated ECD using time dependent density functional theory (TDDFT). The isolated compounds were evaluated for their anti-HBV activities against HBsAg and HBeAg secretion, and the results obtained are discussed herein.

Results and Discussion

The 70% ethanol extract of the air dried roots of *P. emblica* was subjected to various column chromatography, followed with preparative HPLC (p-HPLC) to afford 11 sesquiterpenoids (**1**-**11**). Compounds **9**-**11** were identified as the known phyllaemblicin F (**9**),[9] phyllaemblic acid (**10**),[6] and glochicoccin D (**11**),[10] respectively, by comparing with authentic samples directly and their spectroscopic and physical data with those previously reported in the literatures.

Phyllaemblicin G1 (**1**) was obtained as a white amorphous powder, with a molecular formula $C_{34}H_{48}O_{20}$, as established from the HRESIMS *m/z* 775.2657 $[M-H]^-$. The ^1H NMR spectrum of **1** (Table 1) displayed the existence of one benzoyl group [δ_H 8.12 and 7.53 (each 2H, J = 8.0 Hz), δ_H 7.68 (1H, t, J = 7.4 Hz)] and two anomeric protons [δ_H 5.55 and 4.26 (each 1H, d, J = 8.0 Hz)]. Besides the signals of one benzoyl and two



hexosyl group, the ^{13}C NMR and DEPT spectra (Table 1) showed 15 carbon signals, arising from one methyl (δ_C 13.0), five methylenes including two oxygen bearing ones (δ_C 63.2 and 63.4), five methines including three oxymethines, one ketal carbon (δ_C 107.7), one carboxyl (δ_C 175.9) group and two oxygen bearing quaternary carbons (δ_C 81.1 and 87.8). These NMR features were comparable to those of phyllaemblicin B, a known norbisabolane glycoside from the titled plant,[8] indicating that **1** was a highly oxygenated bisabolane sesquisterpenoid glycoside. The main difference was the appearance of an oxygenated quaternary carbon (δ_C 87.8) and an additional oxymethylene (δ_C 63.4) in **1**, instead of the C-7 ketone (δ_C 213.7) in phyllaemblicin B. The signals δ_C 63.4 and δ_c 87.8 were assigned as C-15 and C-7, respectively, based on the HMBC correlations from H-15 to C-7 and C-8, and H-1 to C-7 (Figure 1). Other HMBC correlations confirmed the planar structure of **1** as shown in Figure 1.

The relative configuration of **1** was established by proton coupling constants and ROESY experiment (Figure 1). The coupling constants of $J_{3,4a}$ (11.0 Hz), and $J_{4,5}$ (3.8 Hz) indicated the axially α-orientated H-3 and β-orientated H-5. The broad singlet of H-1 suggested its β-orientation. This was further confirmed by the ROESY correlations of H-3 with H-4b (δ_H 2.07), H-1 and H-5 with H-4a. The ROESY correlation of H-15 with H-5 implied that C-15 methylene was β-orientated. The coupling constants of H-12b (δ_H 4.05, 1H, dd, J = 11.2, 11.5 Hz) suggested the axially orientated H-11, and the small coupling constants of H-10 (δ_H 5.28, 1H, brs) implied its equatorial orientation and located on the same face as H-11. The relative configuration of **1** (except C-8) was therefore established as shown in Figure 1.

The absolute configuration of the aglycon of **1** was determined by comparing the calculated ECD curve with its experimental ECD spectrum, since there was a benzoate chromaphore in the rigid C ring connecting to the C*-10 hydroxyl group. A negative and a positive Cotton effects appeared at 230 and 250 nm, arising from the π-π* excitation of the benzene ring. The calculated ECD curve agreed well with the experimental result Figure 2). Thus, the absolute configurations of C-10 and C-11 were determined as 10*S*,11*R*. The rigid tetrahydropyran ring C in **1** was of chair conformation with 8*R*,10*S*,11*R* (Figure 1C) and 8*S*,10*S*,11*R* (Figure 1D) configurations, which were confirmed by the Monte Carlo search with MMFF method and DFT/B3LYP 6-311G(d,p) optimization. The 10-H and 11-H in 8*S*,10*S*,11*R* isomer were equatorially and axially orientated, respectively, which were consistant with those of compound **1**. However, they were axially and equatorially orientated in the 8*R*,10*S*,11*R* isomer, respectively, which were opposite to those of **1**. On the basis of the above evidence, the absolute configuration of C-8 was assigned as *S*.

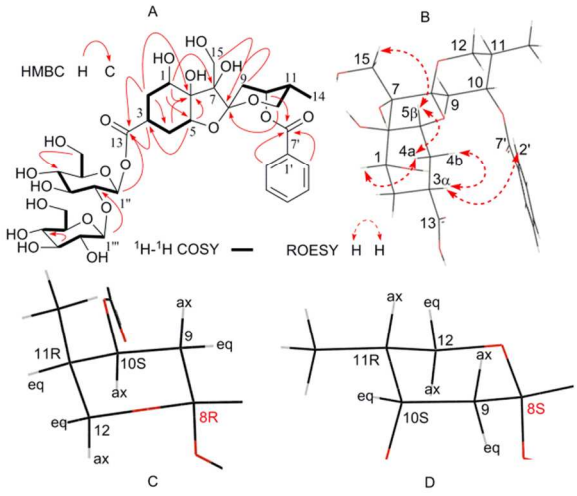


Figure 1. Key ^{1}H-^{1}H COSY, HMBC, ROESY correlations (A-B), and ring C conformations (C-D) of **1**

Phyllaemblicin G2 (**2**) had the same molecular formula $C_{34}H_{48}O_{20}$ as **1**, as deduced from the HRESIMS (*m/z* 775.2653[M-H]$^-$). The ^{1}H and ^{13}C NMR spectroscopic data (Table 1) were similar to those of **1**, suggesting a close relationship between these two compounds. In the ^{13}C NMR spectrum, significant up-field shifts of C-5, C-6 and C-7 ($\Delta\delta_C$ -5.6, -4.0, -5.7) and down-field shifts of C-1 and C-11 ($\Delta\delta_C$ 2.4

Table 1. ^{13}C and ^{1}H NMR Spectroscopic data for compounds **1** – **3** in CD_3OD (δ in ppm)

No.	**1**		**2**		**3**	
	δ_C [a]	δ_H [a]	δ_C [a]	δ_H [c]	δ_C [b]	δ_H [b]
1	73.0, CH	3.90 brs	75.4, CH	4.32 brs	72.1, CH	3.81 dd (4.9, 10.2)
2a	31.7, CH_2	2.02 [d]	27.6, CH_2	1.63 [d]	29.1, CH_2	1.60 ddd (9.2, 10.2, 13.9)
2b				2.24 brd (14.4)		1.96 [d]
3	34.6, CH	2.85 m	33.5, CH	3.03 tt (4.0, 13.0)	34.3, CH	2.58 dddd (5.6, 5.6, 9.2, 11.2)
4a	29.3, CH_2	1.85 ddd (4.0, 11.0, 14.3)	28.4, CH_2	1.49 ddd (1.4, 13.0 14.1)	27.7, CH_2	1.87 ddd (3.3, 11.2, 14.3)
4b		2.07 [d]		1.63 [d]		1.96 [d]
5	80.5, CH	4.09 t (3.8)	74.9, CH	3.84 brs	82.0, CH	4.07 brs
6	81.1, C		77.1, C		76.5, C	
7	87.8, C		82.1, C		76.0, CH	3.82 s
8	107.7, C		107.9, C		102.4, C	
9a	34.2, CH_2	1.95 dd (3.1, 15.1)	32.0, CH_2	2.21 dd (10.0, 15.5)	36.0, CH_2	2.17 [d]
9b		2.28 dd (2.8, 15.1)		2.16 dd (2.3, 15.5)		
10	71.8, CH	5.28 brs	72.7, CH	5.67 ddd (2.3, 5.8, 10.0)	72.0, CH	5.29 brs
11	34.1, CH	2.08 [d]	41.7, CH	2.03 m	34.2, CH	2.15 [d]
12a	63.2, CH_2	3.56 dd (4.8, 11.2)	64.8, CH_2	3.45 [d]	63.1, CH_2	3.64 dd (4.7, 11.4)
12b		4.05 dd (11.2, 11.5)		3.67 dd (5.8, 11.0)		4.07 dd (11.4, 11.4)
13	175.9, C		176.4, C		177.4, C	
14	13.0, CH_3	0.86 d (7.2)	13.6, CH_3	1.03 d (6.9)	13.0, CH_3	0.93 d (6.8)
15a	63.4, CH_2	3.69 [d]	61.0, CH_2	3.81 d (11.8)		
15b		3.92 [d]		3.88 d (11.8)		
1'	132.3, C		132.6, C		132.1, C	
2',6'	130.8, CH	8.12 d (8.0)	130.9, CH	8.06 d (8.0)	130.7, CH	8.12 d (8.0)
3',5'	129.8, CH	7.53 t (8.0)	129.6, CH	7.47 t (8.0)	129.5, CH	7.50 t (8.0)
4'	134.3, CH	7.64 t (7.4)	134.0, CH	7.57 t (7.4)	134.1, C	7.63 t (7.4)
7'	168.0, C		168.4, C		168.1, C	
1''	93.7, CH	5.55 d (8.0)	94.3, CH	5.59 d (7.5)	52.3, CH_3	3.37 s
2''	82.6, CH	3.35 [d]	82.0, CH	3.66 [d]		
3''	77.8, CH	3.61 dd (9.0, 9.0)	77.9, CH	3.64 dd (9.0, 9.0)		
4''	70.7, CH	3.43 dd (9.0, 9.0)	71.1, CH	3.42 [d]		
5''	78.8, CH	3.37 [d]	79.0, CH	3.44 ddd (3.0, 5.0, 9.0)		
6''a	62.3, CH_2	3.72 dd (5.0, 12.1)	62.7, CH_2	3.76 dd (5.0, 12.2)		
6''b		3.87 dd (2.1, 12.1)		3.86 dd (3.0, 12.2)		
1'''	105.8, CH	4.26 d (8.0)	105.4, CH	4.71 d (8.0)		
2'''	75.8, CH	3.13 dd (8.0, 9.2)	76.5, CH	3.21 dd (8.0, 9.5)		
3'''	77.8, CH	3.28 [d]	78.0, CH	3.46 dd (9.5, 9.5)		
4'''	70.8, CH	3.28 [d]	71.6, CH	3.38 dd (9.5, 9.5)		
5'''	77.7, CH	2.88 ddd (2.9, 2.9, 9.2)	77.8, CH	3.46 [d]		
6'''a	61.9, CH_2	3.62 m	62.7, CH_2	3.77 dd (4.8, 12.2)		
6'''b		3.62 m		3.85 dd (3.0, 12.2)		

[a] Data were recorded for ^{13} C NMR at 125 MHz, or ^{1}H NMR at 500 MHz. [b] Data were recorded for ^{13} C NMR at 150 MHz, or ^{1}H NMR at 600 MHz. [c] Data were recorded at 800 MHz. [d] overlapped signals

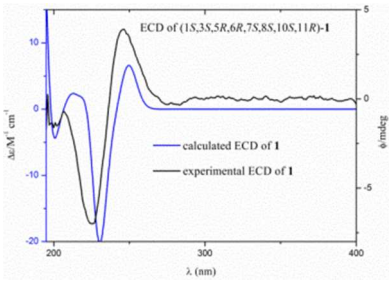


Figure 2. Experimental (**1**) and calculated (the aglycon of **1**) ECD curves

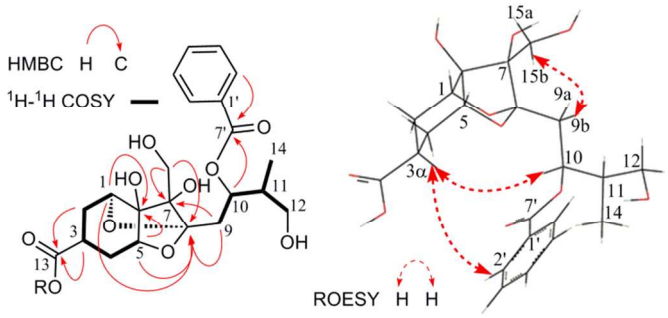


Figure 3. Key ^{1}H-^{1}H COSY, HMBC and ROESY correlations of **2**

and 7.6) in **2**, related to **1**, were observed. In the HMBC experiment, the strong correlation from H-12 to C-8 in **1** disappeared in **2**, instead of weak correlations from H-1 and H-5 to C-8 (δ_C 107.9) (Figure 3). This suggested that two oxygen bridges were formed between C-1 and C-5 with C-8 in **2**, and ring C was opened to form an aliphatic chain. In order to confirm the rearranged ring system in **2**, the aglycon (**2A**) was obtained after hydrolysis using 0.3 M NaOH aqueous solution. The ^{1}H-^{1}H COSY spectrum of **2A** in DMSO-d_6 (see ESI Figure S23-27) indicated that both H-12 and H-15 were mutually coupled with their corresponding hydroxyl protons at δ_H 4.29 (1H, t, J = 6.4 Hz, 12-OH) and δ_H 4.82 (1H, brs, 15-OH), respectively, indicating that both C-12 and C-15 connected with a free hydroxy group. Together with other ^{1}H-^{1}H COSY and HMBC correlations (Figure 3) of **2**, the planar structure of **2** was constructed as shown in Figure 3.

The large coupling constant 13.0 Hz of H-3 suggested its axial orientation, while the small coupling constants of H-1 and H-5 revealed their equatorial orientations. H-3 in A ring with chair conformation was on the opposite face to H-1 and H-5. The stereochemistry of C-7 could be deduced as the same as compound **1**, considering the biosynthetic reason. This was

confirmed by the ROESY correlations of H-15b (δ_H 3.88) and H-9b. Together with the ROESY correlations of H-3 with H-10 and H-2',6' of the benzoyl group, it was allowed the construction of the relative configuration of **2** (Figure 3). The value of $J_{9a,10}$ = 10.0 Hz indicated an anti orientation of H-10 and H-9a (δ_H 2.21). The relative configurations of C-10 and C-11 in the aliphatic chain were determined using Murata's method,[11] since this is a methyl and oxygen substituted 1,2-methine system in aliphatic chain. All $^{2,3}J_{C,H}$ values were accurately measured by 2D NMR of HETLOC (see ESI Figure S15-18). As shown in Figure 4, the medium $^3J_{H\text{-}11,H\text{-}10}$ (5.8 Hz), medium $^3J_{H\text{-}10,\ C\text{-}12}$ (4.2 Hz), and small $^3J_{H\text{-}11,C\text{-}9}$ (3.0 Hz) indicated that rotamers B2/B3 were the correct rotamer pair based on Murata's rules. This was confirmed by the ROESY correlations of H-10 with H-11 and H-14 in B2, and H-9 with H-14 in B3. Thus, the relative configurations of C-10 and C-11 in compound **2** were established as 10*S** and 11*R**, respectively. In this case, the third rotamer B1 existed in comparable population, which made the $^3J_{C\text{-}14,H\text{-}10}$ and $^2J_{C\text{-}10,\ H\text{-}11}$ do not accurately correspond to the rotamers B2/B3. This situation was mentioned by Murata etal.,[11] and the predominant B2/B3 rotamers deduced from $^3J_{H\text{-}11,C\text{-}9}$ and $^3J_{H\text{-}10,\ C\text{-}12}$ could led to correct answer. Using Monte Carlo search with MMFF method and DFT/B3LYP 6-311G(d,p) optimization, the populations of the three rotamers B1, B2 and B3 could be estimated as 44.1%, 34.2%, and 21.5%, respectively (see ESI Figure S74).

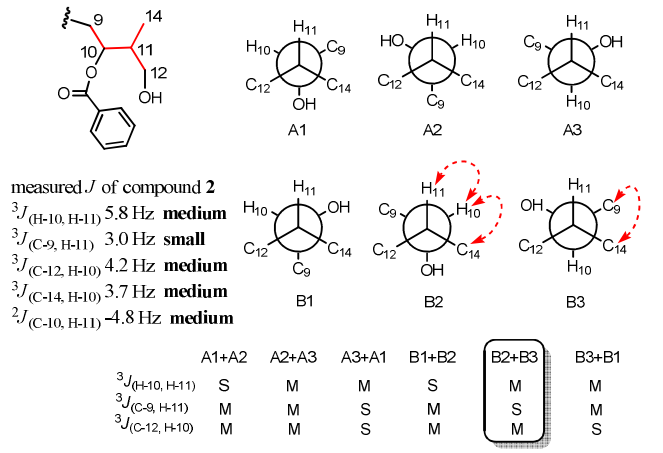


Figure 4. The Newman projection of all possible staggered rotamers of threo and erythro configurations viewed down bonds C11-C10. The $^3J_{H,H}$ and $^3J_{C,H}$ values allowed the assignment of rotamers of B2/B3, and the ROESY correlations were drawn in double-sided arrows

Although the aliphatic chain, on which the benzoate chromophore was located, was not a rigid system, all lower energy conformers displayed similar Cotton effects as that of the Boltzmann averaged curve (Figure 5 and ESI Figure S74). The configurations of the aglycon part of phyllaemblicin G2 (**2**) was determined on the basis of ECD calculation. The absolute configurations of C-10 and C-11 were consistent with those of **1** and the other bisabolane sesquiterpenoid glycosides.[6,8]

Phyllaemblicin G3 (**3**) had a molecular formula $C_{22}H_{28}O_9$, as determined from the HRESIMS. Its NMR data were similar to

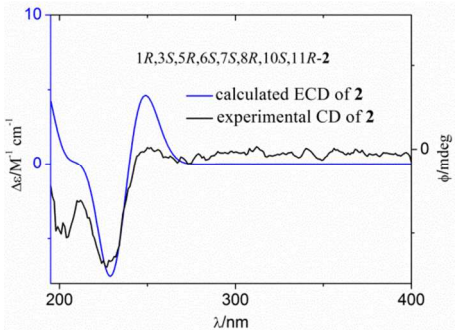


Figure 5. Experimental and calculated ECD curves of **2**

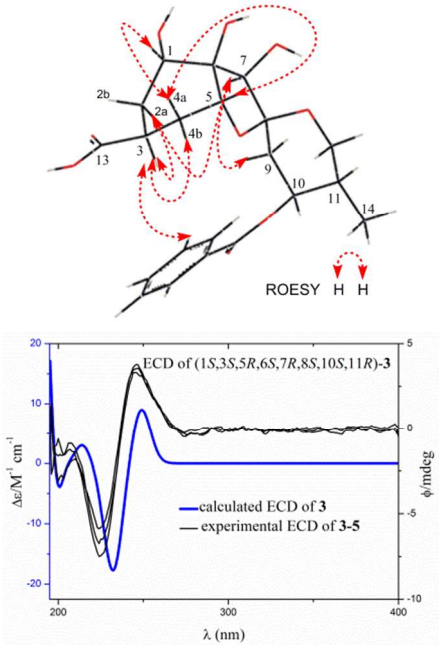


Figure 6. Key ROESY correlations of **3**, and calculated (**3**) and experimental ECD spectra of **3**-**5**

those of phyllanemblic acid (**10**). However, instead of a ketone signal (δ_C 213.9, C-7) in **10**, the ^{13}C NMR spectrum displayed an oxymethine signal (δ_C 76.0) in **3**. In the HMBC spectrum, its corresponding proton at δ_H 3.82 (1H, s) showed correlations with C-6, C-8 and C-9, and the signal δ_C 76.0 was assigned to be C-7. Moreover, the methoxy group (δ_C 52.3, δ_H 3.37) was assigned to connect to C-13 as a methyl ester, by the HMBC correlation from δ_H 3.37 to the carboxyl carbon C-13. Ring C in compound **3** had the same relative configurations as **1**. The large coupling constants of $J_{H\text{-}1,H\text{-}2a}$ (10.2 Hz), $J_{H\text{-}2a,H\text{-}3}$ (9.2 Hz), and $J_{H\text{-}4a,H\text{-}3}$ (11.2 Hz) indicated the axial orientations of H-1 and H-3, while H-5 with a small coupling constant was assigned to be equatorial orientation. The NOE effects of H-1 and H-5 with H-4a (δ_H 1.87), H-3 with H-2a (δ_H 1.60) and H-4b allowed the assignment of H-1 and H-5 on the opposite face to H-3, indicating that ring A was of boat conformation (Figure 6). Taking into account of the ROESY correlations of H-7 with H-9 and H-2a, the configuration of **3** was determined. The calculated ECD curve agreed well with the experimental ECD spectrum. Thus, the configuration of **3** was determined as 1*S*,3*S*,

5*R*,6*S*,7*R*,8*S*,10*S*,11*R*.

The molecular formulas of phyllaemblicins G4-G5 (**4**-**5**) were determined as $C_{33}H_{46}O_{19}$ and $C_{33}H_{46}O_{19}$, respectively, by the HRESIMS. Detailed analysis of NMR data (Table 2) revealed that both compounds **4** and **5** had the same aglycon as compound **3**. The only difference was the different saccharide moieties of **4** and **5**. The saccharide moiety of **4** was determined to be S1, which is the same as that of **1**. In the case of compound **5**, the ^{1}H and ^{13}C NMR spectra displayed the presence of a glucosyl [anomeric proton at δ_H 4.10 (1H, d, J = 8.0Hz)] unit and a set of signals arising from a hexaoxycyclohexane residue, which was linked to C-13, based on the HMBC correlation from H-1" (δ_H 4.80) to C-13 (δ_C 177.1). The hexaoxycyclohexane moiety was determined to be *myo*-inositol on the basis of ^{1}H-^{1}H COSY correlations of H-1"/H-2"/H-3"/H-4"/H-5"/H-6" and their corresponding coupling constants (Table 2). Moreover, the HMBC correlation from δ_H 4.10 (H-1"') to C-2" (δ_C 83.2) confirmed the saccharide moiety in **5** as *myo*-inositol-2-*O*-*β*-glucopyranosyl group. The ECD spectra of **4** and **5** displayed the same Cotton effects as that of **3**, indicating the same absolute configuration as that of **3** (Figure 6).

Table 2. ^{13}C and ^{1}H NMR Spectroscopic data for compounds **4** – **5** in CD_3OD (δ in ppm)

No.	**4**[a]		**5**[b]	
	δ_C	δ_H	δ_C	δ_H
1	72.1, CH	3.85 dd (4.6, 10.2)	72.3, CH	3.84 dd (3.7, 9.4)
2a	28.6, CH_2	1.60 ddd (9.4, 10.2, 14.0)	28.4, CH_2	1.60 ddd (9.7, 9.7, 14.1)
2b		1.91 m		2.20 m
3	34.7, CH	2.67 dddd (5.6, 5.6, 9.4, 11.2)	34.5, CH	2.71 m
4a	28.1, CH_2	1.96 [c]	28.6, CH_2	1.93 m
4b		2.16 [c]		2.20 m
5	82.3, CH	4.11 brs	82.3, CH	4.13 dd (2.5, 2.5)
6	76.7, C		76.6, C	
7	72.0, CH	3.83 s	75.8, CH	3.85 s
8	102.5, C		102.8, C	
9	36.1, CH_2	2.12 [c]	36.1, CH_2	2.15 [c]
10	72.0, CH	5.30 brs	71.9, CH	5.33 m
11	34.1, CH	2.11 m	34.2, CH	2.15 [c]
12a	63.0, CH_2	3.60 [c]	63.1, CH_2	3.64 [c]
12b		4.04 dd (11.6, 11.6)		4.09 dd (12.0, 12.0)
13	175.9, C		177.1, C	
14	13.1, CH_3	0.88 d (6.8)	13.1, CH_3	0.90 d (6.9)
1'	132.1, C		132.2, C	
2',6'	131.0, CH	8.13 d (7.8)	130.9, CH	8.17 d (8.0)
3',5'	129.7, CH	7.52 t (7.8)	129.8, CH	7.56 t (8.0)
4'	134.4, CH	7.64 t (7.4)	134.6, CH	7.76 t (7.4)
7'	168.1, C		168.0, C	
1"	94.0, CH	5.55 d (7.9)	75.4, CH	4.80 t (9.6)
2"	82.9, CH	3.41 dd (7.9, 9.5)	83.2, CH	3.78 dd (9.6, 9.6)
3"	77.6, CH	3.61 [c]	73.6, CH	3.60 dd (2.8, 9.6)
4"	70.7, CH	3.38 dd (9.5, 9.5)	73.5, CH	4.02 dd (2.8, 2.8)
5"	78.8, CH	3.35 m	73.3, CH_2	3.43 dd (2.8, 9.6)
6"a	62.2, CH_2	3.78 [c]	72.7, CH	3.79 dd (9.6, 9.6)
6"b		3.86 [c]		
1"'	105.8, CH	4.24 d (7.7)	106.1, CH	4.10 d (8.0)
2"'	76.0, CH	3.14 dd (7.7, 8.8)	76.0, CH	3.11 dd (8.0, 9.0)
3"'	77.8, CH	3.28 [c]	78.1, CH	3.24 dd (9.0, 9.0)
4"'	70.7, CH	3.30 [c]	70.7, CH	3.31 dd (9.0, 9.0)
5"'	77.8, CH	2.86 ddd (3.2, 6.3, 9.1)	77.6, CH	2.66 dt (3.0, 9.5)
6"'a	62.2, CH_2	3.62 [c]	62.1, CH_2	3.44 [c]
6"'b		3.70 dd (5.2, 12.8)		3.87 [c]

[a] Data were recorded at 400 MHz for ^{1}H NMR and 100 MHz for ^{13}C NMR. [b] Data were recorded at 500 MHz for ^{1}H NMR and 125 MHz for ^{13}C NMR. [c] Signals were overlapped.

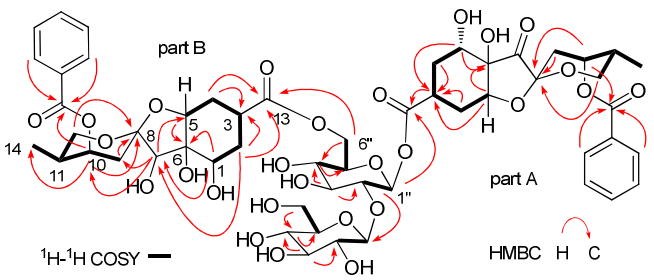


Figure 7. Key ^{1}H-^{1}H COSY and HMBC correlations of **6**

Phyllaemblicins G6-G8 (**6**-**8**) were dimeric sesquiterpenoid glycosides with two norbisabolane units connecting through a disaccharide moiey. Their molecular formulas were assigned as $C_{54}H_{68}O_{27}$ (**6**), $C_{54}H_{66}O_{28}$ (**7**), and $C_{54}H_{66}O_{28}$ (**8**), respectively, based on HRESIMS. The 1D NMR spectra (Tables 3 and 4) of **6** displayed the presence of two hexosyl units [anomeric centers at δ_H 5.64 and 4.14 (each 1H, d, J = 8.0 Hz, H-1",1'"), δ_C 93.8 (C-1") and 106.2 (C-1'")]. The ^{13}C NMR signals arising from the aglycon part were similar to those of phyllaemblic acid (**10**). However, they appeared in pairs or overlapped, suggesting that there were two norbisabolane sesquiterpenoid units in **6**, taking account of the molecular formula. In ^{1}H-^{1}H COSY experiment, two proton spin systems, H-1"/H-2" and H-4"/H-5"/H-6" arising from a hexosyl unit, could be constructed. Together with the HMBC correlations from H-3" to C-2" and C-5", this hexosyl moiety was determined to be a *β*-glucopyranosyl group, based on the coupling constants (Table 3). In HMBC spectrum, H-1" and H-6" were correlated with two aglycon carboxyl carbons at δ_C 176.0 (C-13, part A) and δ_C 177.4 (C-13, part B), respectively, implying that the sesquiterpenoid parts A and B were connected respectively to this glucosyl C-1" and C-6" through ester linkage. Another hexosyl group in **6** was also assigned as *β*-glucopyranosyl unit, and the connectivity of the saccharide was confirmed by the HMBC correlation from H-1"' to C-2". The extensive analysis of the 2D NMR data (Figure 7) allowed the assignment of the signals arising from aglycon moieties (parts A and B), which were determined to be phyllaemblic acid (**10**, part A) and 7-hydroxyphyllaemblic acid (part B), respectively. The structure of **6** was therefore determined as shown.

Phyllaemblicin G7 (**7**) had similar NMR spectra to **6**. The obvious difference was the appearance of a *p*-hydroxybenzoyl [δ_H 7.90 and 7.07 (each 2H, d, J = 8.5 Hz, H-2',6', H-3',5')] and a ketone [δ_C 214.0] group in **7**, instead of the benzoyl and C-7 oxymethine in **6**. The ketone group at δ_C 214.0 was assigned to be C-7 of part B based on its HMBC correlations with H_2-9 (δ_H 1.91 and 2.10). The *p*-hydroxybenzoyl group was assigned to connect to the C-10 hydroxyl group of part B, based on the HMBC correlations from H-10 and H-2',6' to C-7' of part B. Detailed analysis of the 2D NMR data of **7** allowed the assignments of parts A and B as phyllaemblic acid (**10**) and glochicoccin D (**11**), respectively.

Extensive analysis of 2D NMR spectra of phyllaemblicin G8 (**8**) revealed that **8** had S1 as saccharide moiety, and

Table 3. ^{1}H NMR Spectroscopic data for compounds **6** – **8** in CD_3OD (δ in ppm)

No.	**6**[a]		**7**[a]		**8**[a]	
	Part A	Part B	Part A	Part B	Part A	Part B
1	3.91 brs	3.83 m	3.90 brs	3.92 brs	3.90 brs	3.90 brs
2a	1.76 dt (1.2, 13.1)	1.60 ddd (9.1, 10.1, 14.0)	1.82 [b]	1.78 [b]	1.77 [b]	1.77 [b]
2b	2.04 [b]	2.03 [b]	1.91 dd (3.3, 12.5)	2.02 [b]	1.94 [b]	1.94 [b]
3	2.90 tt (3.0, 12.5)	2.65 m	2.89 tt (3.5, 13.5)	2.98 tt (3.5, 13.0)	2.85 tt (3.0, 13.0)	2.99 tt (3.0, 13.0)
4a	1.86 ddd (3.9, 12.4, 14.3)	1.98 [c]	1.72 ddd (4.0, 13.1, 14.6)	2.00 [b]	1.84 ddd (3.6, 13.0, 15.5)	2.01 m
4b	2.29 brd (14.3)	2.23 dd (3.2, 14.5)	2.26 [b]	2.91 m	2.29 brd (15.5)	2.59 brd (15.5)
5	4.22 brs	4.04 brs	4.07 brs	4.29 brs	4.17 brs	4.25 brs
7		3.80 s				
9a	1.98 [b]	2.01 [b]	1.92 [b]	1.91 [b]	1.94 [b]	1.90 dd (3.3, 14.8)
9b	2.23 [b]	2.12 [b]	2.00 [b]	2.10 [b]	2.15 dd (3.3, 15.2)	2.11 dd (3.3, 14.8)
10	5.33 brs	5.27 brs	5.28 brs	5.29 brs	5.31 brs	5.24 brs
11	2.13 [b]	2.04 [b]	2.06 [b]	2.00 [b]	2.11 m	2.02 m
12a	3.51 [b]	3.46 [b]	3.40 dd (4.5, 11.5)	3.15 [b]	3.45 [b]	3.26 [b]
12b	3.86 [b]	3.77 [b]	3.60 t (11.5)	3.38 [b]	3.78 t (11.5)	3.45 [b]
14	0.84 d (7.0)	0.78 d (7.0)	0.73 d (7.0)	0.60 d (7.0)	0.79 d (7.0)	0.68 d (7.0)
2',6'	8.09 d (8.0)	8.12 d (8.0)	7.90 d (8.5)	8.25 d (8.0)	8.09 d (8.0)	8.05 d (8.5)
3',5'	7.52 t (8.0)	7.56 t (8.0)	7.07 d (8.5)	7.70 t (8.0)	7.63 t (8.0)	6.97 d (8.5)
4'	7.56 [b]	7.65 t (7.1)		7.65 t (7.5)	7.70 t (7.5)	
1"	5.64 d (8.0)		5.80 d (8.1)		5.72 d (8.1)	
2"	3.44 dd (8.0, 9.0)		3.78 dd (8.0, 9.2)		3.43 dd (8.1, 9.3)	
3"	3.64 dd (9.0, 9.0)		3.70 dd (9.2, 9.2)		3.67 dd (9.3, 9.3)	
4"	3.50 [b]		3.83 dd (9.2, 9.2)		3.21 dd (9.3, 9.3)	
5"	3.62 dt (2.4, 9.5)		3.70 m		2.49 dt (2.5, 9.3)	
6"a	4.28 brd (11.7)		4.23 dd (1.5, 12.5)		3.39 dd (2.3, 12.1)	
6"b	4.50 dd (5.1,11.7)		4.76 dd (3.0,12.5)		3.46 m	
1"'	4.14 d (8.0)		4.05 d (7.7)		4.02 d (8.0)	
2"'	3.11 dd (8.0, 8.0)		3.11 dd (7.7, 9.2)		3.09 dd (8.0, 9.0)	
3"'	3.22 dd (9.0, 9.0)		3.16 dd (9.2, 9.2)		3.18 dd (9.0, 9.0)	
4"'	3.23 dd (9.0, 90)		3.24 dd (9.2, 9.2)		3.59 dd (9.0, 9.0)	
5"'	2.65 ddd (3.2, 5.5, 9.0)		2.27 m		3.70 m	
6"'a	3.49 m		3.21 m		4.36 dd (1.5, 12.0)	
6"'b	3.49 m		3.36 m		4.59 dd (4.5, 12.0)	

[a] Data were recorded at 500 MHz. [b] Signals were overlapped.

Table 4. ^{13}C NMR Spectroscopic data for compounds **6** – **8** in CD_3OD (δ in ppm)

No.	**6**[a]		**7**[b]		**8**[b]	
	Part A	Part B	Part A	Part B	Part A	Part B
1	71.7, CH	72.3, CH	71.5, CH	71.8, CH	71.4, CH	71.4, CH
2	32.9, CH_2	29.2, CH_2	31.3, CH_2	32.6, CH_2	32.6, CH_2	31.7, CH_2
3	32.4, CH	34.6, CH	32.2, CH	32.3, CH	32.1, CH	32.2, CH
4	29.6, CH_2	28.0, CH_2	30.3, CH_2	29.6, CH_2	29.5, CH_2	29.2, CH_2
5	76.3, CH	81.8, CH	75.9, CH	76.5, CH	76.1, CH	76.2, CH
6	75.6, C	76.5, C	75.4, C	75.9, C	75.4, C	75.6, C
7	213.6, C	76.3, CH	213.6, C	214.0, CH	213.6, C	213.9, CH
8	100.7, C	102.5, C	100.5, C	100.6, C	100.4, C	100.5, C
9	32.9, CH_2	36.3, CH_2	32.9, CH_2	33.1, CH_2	32.7, CH_2	32.7, CH_2
10	71.1, CH	72.0, CH	70.1, CH	71.0, CH	70.9, CH	70.3, CH
11	34.4, CH	34.2, CH	34.3, CH	34.0, CH	34.2, CH	34.0, CH
12	63.5, CH_2	63.2, CH_2	63.3, CH_2	63.1, CH_2	63.3, CH_2	63.2, CH_2
13	176.0, C	177.4, C	175.9, C	176.9, C	175.7, C	176.6, C
14	13.8, CH_3	13.8, CH_3	13.3, CH_3	13.6, CH_3	13.2, CH_3	13.3, CH_3
1'	132.4, C	132.3, C	123.2, C	132.2, C	132.5, C	123.1, C
2',6'	131.1, CH	131.0, CH	133.6, CH	131.6, CH	131.1, CH	133.4, CH
3',5'	129.7, CH	130.0, CH	117.1, CH	129.9, CH	130.1, CH	116.2, CH
4'	134.4, CH	134.6, CH	163.5, C	134.8, CH	134.2, CH	163.4, C
7'	168.0, C	168.2, C	168.0, C	167.2, C	167.8, C	168.1, C
1"	93.8, CH		94.0, CH		93.5, CH	
2"	83.3, CH		84.6, CH		83.8, CH	
3"	77.9, CH		77.4, CH		77.4, CH	
4"	71.0, CH		70.0, CH		70.3, CH	
5"	76.6, CH		76.9, CH		77.4, CH	
6"	64.3, CH_2		63.4, CH_2		61.4, CH_2	
1"'	106.2, CH		107.2, CH		106.4, CH	
2"'	76.1, CH		75.9, CH		75.8, CH	
3"'	78.0, CH		78.1, CH		77.8, CH	
4"'	70.9, CH		70.0, CH		70.6, CH	
5"'	77.9, CH		77.5, CH		76.3, CH	
6"'	62.1, CH_2		61.1, CH_2		63.7, CH_2	

[a] Data were recorded at 125 MHz. [b]Data were recorded at 100 MHz.

phyllaemblic acid (**10**, part A) and glochicoccin D (**11**, part B) as aglycon parts, which were the same as those of **7**. In the HMBC spectrum of **8**, correlations from H-1" to C-13 of part A (δ_C 175.7), and

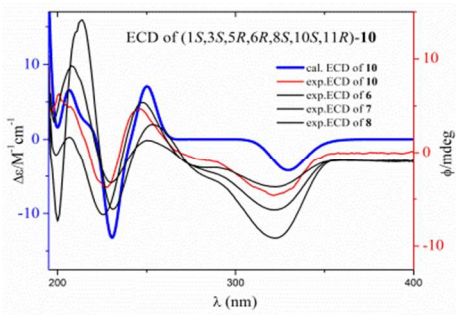


Figure 8. Experimental ECD curves of **6**-**8** and **10**

the H-6"' to C-13 of part B (δ_C 176.6) indicated that part B of **8** located on the second glucosyl C-6"', which is different from that of **7**.

Since the sesquiterpenoid parts in compounds **6**-**8** were also obtained in the study, their relative and absolute configurations can be proposed to be the same as those of their corresponding monomers. This was supported by the similar ECD spectra of **6**-**8** to those of phyllaemblic acid (**10**) (Figure 8).

Antiviral activities of sesquiterpenoid glycosides: The isolated compounds **1**, **2**, **6** - **10** were evaluated for their antiviral activities against four kinds of virus, e.g. herpes simplex virus type 1 (HSV-1), enterovirus 71 (EV71), hepatitis B (HBV) and influenza A virus (H3N2). The sesquiterpenoid dimers **6** - **9** can selectively affect HBsAg and HBeAg secretions, and phyllaemblicin G6 (**6**) in particular showed strong inhibition towards HBsAg and HBeAg secretion with

IC_{50} values of 8.53 ± 0.97 and 5.68 ± 1.75 μM, compared with the positive control, lamivudine with inhibition 44.66 ± 2.09% and 25.96 ± 9.59% towards HBsAg and HBeAg, respectively, even at a concentration of 100 μM, and the inhibitory ratio of lamivudine can't reach 50% below 100 μM. However, the aglycon **10** only displayed weak inhibition effect. The sesquiterpenoid **1** and **2** were both ineffective to HBsAg and HBeAg secretion above the cytotoxic concentrations, implying that the structure charateristics of the sesquiterpenoid glycoside dimers (**6** - **9**) were related to the HBsAg and HBeAg secretion activity. It is believed that high levels of HBsAg bearing particles in the serum of chronically infected individuals play a role in suppressing the HBV immune system.[3] Discovery anti-HBV compounds aimed at potentiating the immune response by suppressing antigenemia became an alternation of anti-HBV therapeutic research. However, there are rare examples except that anti-HBV lead compound PBHBV-2-15 was discovered based on the target of HBsAg secretion.[4] The sesquiterpenoid glycosides dimers **6** - **9**, with unusual structures, displayed significant HBsAg and HBeAg reduction activity, which were promising to development as antigenemia agents for anti-HBV therapeutics.

Table 5. Anti-HBV activities of compounds **6** - **10** (μM)

No.	HBV[a]		CC_{50}[b]
	HBsAg IC_{50}/SI[c]	HBeAg IC_{50}/SI	
6	8.53 ± 0.97 / 4.46	5.68 ± 1.75 / 6.70	38.04 ± 2.38
7	26.30 ± 3.72 / 2.97	21.38 ± 2.56 / 3.65	78.15 ± 7.03
8	25.80 ± 3.18 / 4.65	20.98 ± 3.23 / 5.72	119.97 ± 5.93
9	11.54 ± 1.35 / 3.73	14.08 ± 2.02 / 3.06	43.03 ± 3.32
10	149.87 ± 8.28 / 6.18	568.80 ± 15.91 / 1.63	926.24 ± 17.58

[a]Lamivudine was tested as the positive control, the inhibition ratio against HBsAg and HBeAg were 44.66 ± 2.09 and 25.96 ± 9.59 (100 μM); 31.56 ± 9.21 and 18.67 ± 3.16 (10 μM). [b]Compound concentration reducing the viability of HepG2.2.2.15 cells culture by 50%. [c]SI = CC_{50}/IC_{50}

Conclusions

During the process of exploring novel antiviral compounds from *Phyllanthus* spp., eight new (**1**-**8**) and three known (**9**-**11**) highly oxygenated bisabolane sesquiterpenoids were isolated from *P. emblica*. Phyllaemblicin G2 (**2**), bearing a tricyclo [3.1.1.1] oxygen bridge ring system, is an unusual sesquiterpenoid glycoside, while phyllaemblicins G6-G8 (**6**-**8**) are dimeric sesquiterpenoid glycosides with two norbisabolane units connecting through a disaccharide. All structures were elucidated by detailed analysis of HRESIMS and NMR data. The absolute configurations were determined by means of TDDFT calculated ECD. The relative configuration of compound **2** was resolved by *J*-based analysis and quantum chemical calculations. The sesquiterpenoid dimers **6** - **9** displayed potential inhibitory activity towards HBsAg and HBeAg secretion.

Experimental section

General procedures: Optical rotations were performed on a P-1020 polarimeter (JASCO, Tokyo, Japan). IR and UV spectra were measured on a Bruker Tensor 27 spectrometer with KBr pellets and shimadzu UV 2401PC, respectively. 1D and 2D NMR spectra were run on Bruker DRX-400, 500, AVANCE III-600 and AVANCE-800 NMR spectrometers operating at 400, 500 and 600 and 800 MHz for ^{1}H, and 100, 125 and 150 and 200 MHz for ^{13}C, respectively. Coupling constants are expressed in Hertz and chemical shifts are given on ppm scale with solvents as internal standard. ESI-MS and HRESIMS were measured at Bruker HCT/ Esquire and Agilent G6230. ECD was detected at Applied Photophysics. The apparatus of p-HPLC was an Agilent 1260 with DAD detector. Column chromatography (CC) was performed with Sephadex LH-20 (Pharmacia Fine Chemical Co., Ltd. Uppsala, Sweden), Diaion HP20SS (Mitsubishi Chemical Co., Tokyo, Japan), Rp-8 gel (40-60 μm, Merck, Darmstadt, Germany), Toyopearl HW-40C (TOSOH, Japan), Silica gel (200-300 mesh, Qingdao Hailang Group Co., Ltd. Qingdao, People's Republic of China) and a 250 × 9.4 mm, i.d., 5 μm Sunfire C_{18} column (Waters). TLC was carried out on precoated silica gel GF254 plates, which were visualized by ultraviolet and spraying with 10% sulphuric ethanol solution. The quantum chemical calculations were carried out at HPC Center, Kunming Institute of Botany (KIB), Chinese Academy of Sciences (CAS), China.

Plant materials: The roots of *P. emblica* were collected in Baoshan City, Yunnan Province, People's Republic of China, in 2010, and identified by Prof. C. R. Yang (KIB, CAS). A voucher specimen (KIB-ZL-0100020) has been deposited in State key laboratory of phytochemistry and plant resource in west China of Kunming Institute of Botany, Chinese Academy of Sciences.

Extraction and Isolation. The air dried roots of *P. emblica* (109 kg) were extracted with 70% reflux ethanol solution for three times to give the extract (7.8 Kg), which was suspended into H_2O (22.5 L) and participated with *n*-BuOH. The organic layer was concentrated in vacuum and subjected to a Diaion HP20SS column chromatography (CC), eluting with CH_3OH/H_2O (0-100%), to afford four major fractions (Fr.1-4). Fr.3 (400 g) was applied to a silica gel column, eluting with $CHCl_3/CH_3OH/H_2O$ (50:1:0-7:3:0.5) to give seven sub-fractions (Fr.A–Fr.G). Fr.A was subjected to Sephadex LH-20 (CH_3OH/H_2O, 0-70%) to afford two fractions, Fr.A1 and Fr.A2. Fr.A1 (3.2 g) was chromatographed over MCI CHP-20P (CH_3OH/H_2O, 30%-90%) to give four fractions (Fr.A1.1-Fr.A1.4). Fr.A1.2 was separated by Toyopearl HW-40C (CH_3OH/H_2O, 0-30%), followed by p-HPLC (CH_3CN/H_2O, 15-30%) to yield **2** (32 mg). Fr.A1.4 was purified by Toyopearl HW-40C (CH_3OH/H_2O, 0-30%) to give **5** (5 mg). Fr.B passaged over Sephadex LH-20 (CH_3OH/H_2O, 30-70%) to afford Fr.B1-Fr.B2. Fr.B1 was fractioned on MCI CHP-20P column (CH_3OH/H_2O 30%-80%) to afford Fr.B1.1-Fr.B1.6. Fr.B1.5 was purified by Toyopearl HW-40C (CH_3OH/H_2O 0-30%) and p-HPLC (CH_3CN/H_2O, 15-30%) to afford **3** (1 mg) and **4** (115 mg). Fr.C was fractioned on Sephadex LH-20 (CH_3OH/H_2O, 30-70%) to afford Fr.C1-Fr.C4. Fr.C1 was separated over MCI CHP-20P (CH_3OH/H_2O, 30%-70%) to give Fr.C1.1-1.6, and Fr.C1.5 was chromatographed on Rp-8

(CH_3OH/H_2O, 30%-80%) to afford Fr.C1.5.1-1.5.4. Fr.C1.5.4 was purified by Toyopearl HW-40C (CH_3OH/H_2O, 0-30%) and p-HPLC (CH_3CN/H_2O, 15-30%) to afford **1** (8 mg). Fr.E (24.4 g) passaged over Sephadex LH-20 (CH_3OH/H_2O, 30-100%) to afford Fr.E1-Fr.E3. Fr.E1 was fractioned on Rp-8 (CH_3OH/H_2O, 30%-80%) to afford Fr.E1.1-Fr.E1.8. Fr.E1.2 and Fr.E1.3 were separately chromatographed on Toyopearl HW-40C (CH_3OH/H_2O, 0-30%), followed by p-HPLC (CH_3CN/H_2O, 15-30%) to afford **7** (30 mg), **8** (24 mg), **6** (8 mg) and **9** (500 mg). Fr.G was chromatographed on Toyopearl HW-40C (CH_3OH/H_2O, 0-30%) and p-HPLC (CH_3CN/H_2O 15-30%) to afford **10** (51 mg) and **11** (15 mg).

Phyllaemblicin G1 (1): white amorphous powder; $[\alpha]_D^{25}$ = +29.8 (c = 1.2 in methanol); UV (MeOH) λ_{max} (log ε) 199.8 (1.06), 228.6 (1.08), 272.6 (0.02) nm; ECD (in MeOH, λ_{max} [nm], Φ[mdeg]) 224 (-7.3), 245 (3.6), 280 (-1.1); IR (KBr) ν_{max} 3426, 2931, 1715, 1630, 1283, 1074 cm^{-1}; ^1H NMR (CD_3OD, 500 MHz) and ^{13}C NMR (CD_3OD, 125 MHz) data, see Table 1; MS (ESI): *m/z*: 775 $[M-H]^-$, HRMS (ESI): *m/z* 775.2657 $[M-H]^-$ (calcd for $C_{34}H_{47}O_{20}$, 775.2661).

Phyllaemblicin G2 (2): white amorphous powder; $[\alpha]_D^{25}$ = -24.9 (c = 1.1 in methanol); UV (MeOH) λ_{max} (log ε) 199.8 (1.07), 228.4 (1.08), 271.8 (0.08) nm; ECD (in MeOH, λ_{max} [nm], Φ [mdeg]) 227 (-3.7), 249 (-0.2), 274 (-1.0); IR (KBr) ν_{max} 3427, 2923, 1710, 1630, 1285, 1077 cm^{-1}; ^1H NMR (CD_3OD, 800 MHz) and ^{13}C NMR (CD_3OD, 125 MHz) data, see Table 1; MS (ESI): *m/z*: 799 $[M+Na]^+$; HRMS (ESI): *m/z* 775.2653$[M-H]^-$ (calcd for $C_{34}H_{47}O_{20}$, 775.2661).

Phyllaemblicin G3 (3): white amorphous powder; $[\alpha]_D^{25}$ = +31.3 (c = 0.8 in methanol); UV (MeOH) λ_{max} (log ε) 201.0 (1.07), 227.4 (1.09), 270.8 (0.02) nm; ECD (in MeOH, λ_{max} [nm], Φ[mdeg]) 225 (-5.6), 245 (3.4), 280 (-0.8); ^1H NMR (CD_3OD, 600 MHz) and ^{13}C NMR (CD_3OD, 150 MHz) data, see Table 1; MS (ESI): *m/z*: 459 $[M+Na]^+$; HRMS (ESI): *m/z* 459.1637 $[M+Na]^+$ (calcd for $C_{22}H_{28}O_9Na$, 459.1631).

Phyllaemblicin G4 (4): white amorphous powder; $[\alpha]_D^{25}$ = +21.1 (c = 0.8 in methanol); UV (MeOH) λ_{max} (log ε) 199.8 (1.07), 228.6 (1.09), 272.4 (0.02) nm; ECD (in MeOH, λ_{max} [nm], Φ[mdeg]) 225 (-7.7), 246 (3.3), 281 (-1.1); IR (KBr) ν_{max} 3429, 2930, 1712, 1632, 1281, 1077 cm^{-1}; ^1H NMR (CD_3OD, 400 MHz) and ^{13}C NMR (CD_3OD, 100 MHz) data, see Tables 2; MS (ESI): *m/z*: 781 $[M+Cl]^-$; HRMS (ESI): *m/z* 791.2602 $[M+HCOO]^-$ (calcd for $C_{34}H_{47}O_{21}$, 789.2610).

Phyllaemblicin G5 (5): white amorphous powder; $[\alpha]_D^{25}$ = +5.3 (c = 1.3 in methanol); UV (MeOH) λ_{max} (log ε) 200.4 (1.03), 228.0 (0.92) nm; ECD (in MeOH, λ_{max} [nm], Φ[mdeg]) 225 (-6.7), 247 (3.0), 281 (-1.1); IR (KBr) ν_{max} 3424, 2921, 1716, 1628, 1281, 1116, 1075 cm^{-1}; ^1H NMR (CD_3OD, 500 MHz) and ^{13}C NMR (CD_3OD, 125 MHz) data, see Tables 2; MS (ESI): *m/z*: 769 $[M+Na]^+$; HRMS (ESI): *m/z* 745.2547 $[M-H]^-$ (calcd for $C_{33}H_{45}O_{19}$, 745.2633).

Phyllaemblicin G6 (6): white amorphous powder; $[\alpha]_D^{25}$ = +19.1 (c = 1.2 in methanol); UV (MeOH) λ_{max} (log ε) 200.4 (1.4), 227.8 (1.5), 269.6 (0.6) nm; ECD (in MeOH, λ_{max} [nm], Φ[mdeg]) 226 (-7.4), 247 (5.7), 322 (-3.6); IR (KBr) ν_{max} 3433, 2931, 1718, 1278, 1078 cm^{-1}; ^1H NMR (CD_3OD, 500 MHz) and ^{13}C NMR (CD_3OD, 125 MHz) data, see Tables 4 and 3; MS (ESI): *m/z*: 1183 $[M+Cl]^-$; HRMS (ESI): *m/z* 1193.3915 $[M+HCOO]^-$ (calcd for $C_{55}H_{69}O_{29}$, 1193.3925).

Phyllaemblicin G7 (7): white amorphous powder; $[\alpha]_D^{25}$ = +5.7 (c = 1.0 in methanol); UV (MeOH) λ_{max} (log ε) 201.0 (1.5), 231.2 (1.2), 258.4 (1.2) nm; ECD (in MeOH, λ_{max} [nm], Φ [mdeg]) 212 (19.1), 231 (-6.2), 253 (3.0), 322 (-9.2); IR (KBr) ν_{max} 3435, 2929, 1714, 1610, 1280, 1079 cm^{-1}; ^1H NMR (CD_3OD, 500 MHz) and ^{13}C NMR (CD_3OD, 100 MHz) data, see Tables 4 and 3; MS (ESI): *m/z*: 1161 $[M-H]^-$; HRMS (ESI): *m/z* 1161.3659 $[M-H]^-$ (calcd for $C_{54}H_{65}O_{28}$, 1161.3662).

Phyllaemblicin G8 (8): white amorphous powder; $[\alpha]_D^{25}$ = +6.0 (c = 1.5 in methanol); UV (MeOH) λ_{max} (log ε) 201.0 (1.5), 230.2 (1.2), 258.0 (1.2) nm; ECD (in MeOH, λ_{max} [nm], Φ [mdeg]) 209 (9.8), 228 (-3.5), 253 (1.2), 321 (-6.2); IR (KBr) ν_{max} 3433, 2931, 1713, 1609, 1281, 1079 cm^{-1}; ^1H NMR (CD_3OD, 500 MHz) and ^{13}C NMR (CD_3OD, 100 MHz) data, see Tables 4 and 3; MS (ESI): *m/z*: 1161 $[M-H]^-$; HRMS (ESI): *m/z* 1161.3660 $[M-H]^-$ (calcd for $C_{54}H_{65}O_{28}$, 1161.3662).

Alkaline hydrolysis of 2: Compound **2** (5 mg) was added to 2 mL NaOH solution (0.2 M), and incubated at 60 °C water bath for 30 min. Then the solution was cooled to room temperature, and neutralized by Amberlite IR 120. The elution was evaporated to dryness and subjected to chromatography over silica gel ($CHCl_3$/MeOH/H_2O 8:2:0.2) to afford **2A** (1.5 mg): white amorphous powder; $[\alpha]_D^{25}$ = -4.9 (c = 1.0 in methanol); IR (KBr) ν_{max} 3431, 2926, 1620, 1423, 1384, 1077, 1027 cm^{-1}; ^1H NMR (DMSO-d_6, 800 MHz): δ_H 3.92 (1H, brs, H-1), 1.48 (1H, m, H-2a), 1.87 (1H, brd, H-2b), 2.00 (1H, m, H-3), 1.52 (1H, m, H-4a), 1.96 (1H, m, H-4b), 4.22 (1H, brs, H-5), 1.68 (1H, dd, J = 9.6, 15.2 Hz, H-9a), 1.83 (1H, brd, J = 15.2 Hz, H-9b), 3.88 (1H, m, H-10), 1.56 (1H, m, H-11), 3.24 (1H, dt, J = 10.4, 6.4 Hz, H-12a), 3.45 (1H, dt, J = 10.4, 4.8Hz, H-12b), 0.80 (3H, d, J = 7.2 Hz, H-14), 3.64 (1H, d, J = 12 Hz, H-15a), 3.75 (1H, d, J = 12.0 Hz, H-15b), 4.82 (1H, brs, 15-OH), 4.29 (1H, t, J = 5.1 Hz, 12-OH); ^{13}C NMR (CD_3OD, 200 MHz): 75.9 (C-1), 29.4 (C-2), 35.4 (C-3), 28.9 (C-4), 76.0 (C-5), 77.2 (C-6), 81.9 (C-7), 108.8 (C-8), 34.9 (C-9), 71.0 (C-10), 42.5 (C-11), 65.8 (C-12), 175.1 (C-13), 13.6 (C-14), 61.0 (C-15); MS(ESI) *m/z* 347 $[M-H]^-$; HRMS (ESI): *m/z* 347.1347 $[M-H]^-$ (calcd for $C_{15}H_{23}O_9$, 347.1348).

Acid hydrolysis of compound 7: Hydrochloride solution (2 M, 400 μL) was added to compound **7** (4 mg) in a vial and the vial was sealed. After incubated at 80 °C for 6 h, the reaction mixture was cooled to room temperature, and extracted with $CHCl_3$. The aqueous layer was neutralized by passage a small column with Amberlite IRA-401. TLC was used to identify the type of monosaccharide by comparing with the authentic sugars, with elution system chloroform/*n*-butanol/methanol/acetic acid/water 17:10:6:2:3, and Rfs of 0.35 for glucose, $[\alpha]_D^{25}$ = +94 (c = 1.5 in H_2O); ^1H NMR (pyridine-d_5, 800 MHz) spectrum see ESI Figure S72.

Quantum chemical calculations: The structure models of the compounds were constructed based on NOE analysis. The conformation analysis was carried out using Monte Carlo searching with molecular mechanics MMFF[12] in Sparton'06

(Wavefunction Inc. Irvine, CA). The resulted conformers were reoptimized using DFT at B3LYP-SCRF/6-311G (d,p) level using the integral equation formalism variant of the polarizable continuum model (IEF-PCM). The free energies and vibrational frequencies were calculated at the same level to confirm their stability, and no imaginary frequencies were found. The optimized low energy conformers with energy < 2 kcal/mol were considered for ECD calculation. The TD-DFT/B3LYP SCRF/6-311G (d,p) method was applied to calculate the excited energies, oscillator strength and rotational strength, with 50 states. All the calculations were run with Gaussian '09.[13] The excited energies and rotational strength were used to simulate ECD spectrum of each conformer by introducing the Gaussian Function. The final ECD spectrum of each compound was obtained by averaging all simulated ECD spectra of all conformers according to their excited energies and Boltzmann distribution.[14] The band shape of the calculated ECD curves were all 0.3 eV.

Anti-virus assay as previously reported.[15]

Acknowledgements

The authors are grateful to the members of the analytical group at State Key Laboratory of Phytochemistry and Plant Resources in West China, Kunming Institute of Botany, Chinese Academy of Sciences, for measuring the spectroscopic data. We also thank Fudan University for part of the antiviral assay. This work was supported by the NSFC 21002105, the 973 Program of Ministry of Science and Technology of P. R. China (2011CB915503), the National Science and Technology Support Program of China (2013BAI11B02), the Fourteenth Candidates of the Young Academic Leaders of Yunnan Province (Min Xu, 2011CI044), and the West Light Foundation of the Chinese Academy of Sciences. The calculation sections were supported by the High Performance Computing Center of Kunming Institute of Botany, Chinese Academy of Sciences.

Notes and references

[a] *State Key Laboratory of Phytochemistry and Plant Resources in West China, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, People's Republic of China.*

[b] *University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China.*

[c] *School of Pharmaceutical Sciences, Zhengzhou University, Zhengzhou 450001, People's Republic of China.*

[d] *Henan Key Laboratory for Pharmacology of Liver Diseases, Academy of Medical and Pharmaceutical Sciences, Zhengzhou University, 40 Daxue Road, Zhengzhou 450052, People's Republic of China.*

* *Tel: +86-871-6522-3235. Fax: +86-871-6522-3235. E-mail: zhangyj@mail.kib.ac.cn (Ying-Jun Zhang) and minxu@mail.kib.ac.cn (Min Xu).*

† These authors contributed equally to this work.

‡ Electronic Supplementary Information (ESI) available: MS, 1D and 2D NMR spectra of compounds **1–8** and **2A**, together with the ECD calculation data are provided.

1 Ferir, G.; Kaptein, S.; Neyts, J. and De Clercq, E. *Rev. Med. Virol.,* 2008, **18**, 19.
2 Mailliard, M. E.; Gollan, J. L. *Annu. Rev. Med.,* 2006, **57**, 155.
3 Dougherty, A. M.; Guo, H.; Westby, G.; Liu, Y.; Simsek, E.; Guo, J. T.; Mehta, A.; Norton, P.; Gu, B.; Block, T.; Cuconati, A. *Antimicrob. Agents Ch.,* 2007, **51**, 4427.
4 Yu, W.; Goddard, C.; Clearfield, E.; Mills, C.; Xiao, T.; Guo, H.; Morrey, J. D.; Motter, N. E.; Zhao, K.; Block, T. M.; Cuconati, A.; Xu, X. *J. Med. Chem.,* 2011, **54**, 5660.
5 Shin, M. S.; Kang, E. H.; Lee, Y. I. *Antivir. Res.,* 2005, **67**, 163.
6 Zhang, Y. J.; Tanaka, T.; Iwamoto, Y.; Yang, C. R.; Kouno, I. *Tetrahedron Lett.,* 2000, **41**, 1781.
7 Zhang, Y. J.; Tanaka, T.; Iwamoto, Y.; Yang, C. R.; Kouno, I. *J. Nat. Prod.,* 2001, **64**, 870.
8 Zhang, Y. J.; Tanaka, T.; Iwamoto, Y.; Yang, C. R.; Kouno, I. *J. Nat. Prod.,* 2000, **63**, 1507.
9 Liu, Q.; Wang, Y.F.; Chen, R.-J.; Zhang, M.Y.; Wang, Y.F.; Yang, C.-R.; Zhang, Y.-J. *J. Nat. Prod.,* 2009, **72**, 969.
10 Xiao, H. T.; Hao, X. Y.; Yang, X. W.; Wang, Y. H.; Lu, Y.; Zhang, Y.; Gao, S.; He, H.-P.; Hao, X.J. *Helv. Chim. Acta,* 2007, **90**, 164.
11 Matsumori, N.; Kaneno, D.; Murata, M.; Nakamura, H.; Tachibana, K. *J. Org. Chem.,* 1999, **64**, 866.
12 Stephens, P. J.; Pan, J. J.; Devlin, F. J.; Krohn, K.; Kurtán, T. *J. Org. Chem.,* 2007, **72**, 3521.
13 Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewsk, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09,* Revision A.02; Gaussian, Inc., Wallingford CT, 2009.
14 Zhao, J. Q.; Lv, J. J.; Wang, Y. M.; Xu, M.; Zhu, H. T.; Wang, D.; Yang, C. R.; Wang, Y. F.; Zhang, Y. J. *Tetrahedron Lett.,* 2013, **54**, 4670.
15 Lv, J. J.; Yu, S.; Wang, Y. F.; Wang, D.; Zhu, H. T.; Cheng, R. R.; Yang, C. R.; Xu, M.; Zhang, Y. J. *J. Org. Chem.,* 2014, **79**, 5432.

Graphical and textual abstract for

Anti Hepatitis B Virus Activities and Absolute Configurations of Sesquiterpenoid Glycosides from *Phyllanthus emblica*

Jun-Jiang Lv, Ya-Feng Wang, Jing-Min Zhang, Shan Yu, Dong Wang, Hong-Tao Zhu, Rong-Rong, Cheng, Chong-Ren Yang, Min Xu,* and Ying-Jun Zhang*

The sesquiterpenoid glycoside dimers isolated from *Phyllanthus emblica* have anti-HBV activities towards HBsAg and HBeAg secretions inhibition.