Published in final edited form as:

Magn Reson Chem. 2017 March ; 55(3): 210–213. doi:10.1002/mrc.4466.

Application of a Computer-Assisted Structure Elucidation Program for the Structural Determination of a New Terpenoid Aldehyde with an Unusual Skeleton

Xing-Nuo Li^{1,2}, Clark D. Ridge^{3,*}, Eugene P. Mazzola⁴, Jianghao Sun², Pei Chen^{2,*}, Osvaldo Gutierrez⁴, Arvin Moser⁵, Joseph C. DiMartino⁵, and Scott A. MacDonald⁵

¹College of Pharmaceutical Science, Zhejiang University of Technology, Hangzhou 310014, P. R. China

²Food Composition and Methods Development Laboratory, Beltsville Human Nutrition Research Center, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, MD 20705, USA

³Office of Regulatory Science, Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, College Park, MD 20740, USA

⁴Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742

⁵Advanced Chemistry Development, Inc. (ACD/Labs) Toronto Department, 8 King Street East, Suite 107, Toronto, Ontario M5C 1B5, Canada

Abstract

The structure of a novel compound from *Adansonia digitata* has been elucidated and its ¹H and ¹³C NMR spectra has been assigned employing a variety of 1- and 2-dimensional NMR techniques without degradative chemistry. The Advanced Chemistry Development ACD/Structure Elucidator software was important for determining part of this structure that contained a fused bicyclic system with very few hydrogen atoms, which in turn, exhibited essentially no discriminating HMBC connectivities throughout that portion of the molecule.

Graphical Abstract

*Correspondence to: Pei Chen (compound isolation, natural products), U.S. Department of Agriculture, Beltsville, MD 20705, pei.chen@ars.usda.gov or Clark Ridge (NMR, CASE), U.S. Food and Drug Administration, College Park, MD 20740, Clark.Ridge@fda.hhs.gov.

$d_N(^{13}\text{C})$: 2.044

Keywords

NMR; ^1H NMR; ^{13}C NMR; 1D NMR; 2D NMR; chemical shift assignments; computer-assisted structure elucidation (CASE); natural products; *Adansonia digitata*; baobab

Introduction

Adansonia digitata commonly known as “Baobab” is found mainly in Africa.¹ The tree occurs naturally in dry areas, mainly in the Sahelian, Soudano-Sahelian, and Soudanian zones. The baobab is a multi-purpose tree with products having numerous food uses and medical properties. The fruits, seeds and leaves are all utilized and consumed daily by rural populations in Africa.^{2,3} Baobab fruit pulp is approved by statutory bodies for use in certain nutritional products. Since 2008, the European Commission recognizes the dried fruit pulp of baobab as a novel food.⁴

In our continuing search for pharmacological and structurally interesting substances from the fruits of the baobab, a new terpenoid aldehyde, AD-10 (**1**), with an unusual fused 5/6-ring skeleton has been isolated. The complete proton and carbon NMR assignments of this compound were determined by a series of 1D and 2D NMR experiments including ^1H - ^1H COSY, ^1H - ^{13}C HSQC, ^1H - ^{13}C HMBC and ^1H - ^1H ROESY experiments.

Results and Discussion

AD-10 was isolated as an orange gum and was optically active $\{[\alpha]_D^{20} - 15 (c 0.11, \text{CH}_3\text{OH})\}$. Its UV and IR data were identified as follows: UV, (MeOH) λ_{max} : 226, 309, 444 nm and IR, (KBr) ν_{max} : 3334, 3207, 1732, 1692, 1610, 1519, 1448 cm^{-1} . Its HR-ESI-MS exhibited a quasi-molecular ion peak at m/z 281.0812 $[\text{M}-\text{H}]^-$, (calculated for $\text{C}_{17}\text{H}_{13}\text{O}_4$, 281.0814) and corresponding to a molecular formula of $\text{C}_{17}\text{H}_{14}\text{O}_4$, which requires 11 units of unsaturation. The ^1H , ^{13}C , and HSQC NMR spectra of AD-10 suggested that it is predominantly an aromatic compound and contains both a $\text{CH}-\text{CH}_3$ moiety and an aldehyde group (Table 1). The COSY NMR spectra demonstrated that the following three protons:

9.13 brs, 7.87 d ($J = 5.3$ Hz), and 7.17 dd ($J = 5.3, 0.7$ Hz) ppm constitute a 3-spin system. In addition, protons at 9.73 and 7.94 ppm appeared as singlets and displayed no COSY connectivities (Table 2). COSY and HMBC NMR spectra quickly demonstrated that a catechol group (here a 3,4-dihydroxyphenyl moiety), which is common in natural products,⁵ is attached to the methine carbon of the above CH-CH₃ group to give Fragment 1.

At this point, the structural elucidation process became especially difficult for the following reasons. First, while Fragment 1 did contain nine of the 17 hydrogens of the above-mentioned 3,4-dihydroxyphenyl, CH-CH₃, and aldehyde groups, the remainder of its structure was relatively deficient of hydrogens. Second, the singlet proton at 7.94 ppm displayed an unusually large number of HMBC connectivities (total of 8) that included correlations to all but one of the six potentially unassigned carbons. These two conditions significantly hindered manual structural determination efforts. In addition, what appeared to be an α -furan or α -pyran proton was exceptionally deshielded (9.13 ppm), having a resonance frequency more than 1–2 ppm greater than would be expected for α -furan or α -pyran protons. In such a small compound, it was difficult to envision a structure that would position an oxygen atom in a proximate and largely fixed orientation to the putative furan or pyran hydrogen.

After arriving at various combinations of fused 5/6-membered ring systems, all of which gave structures that were more or less consistent with the HMBC data and required a great deal of analysis time, it was decided to enter a set of experimental data (molecular formula, ¹H, ¹³C, HSQC, and HMBC) into the ACD/Structure Elucidator program (v.14.04 Jul 30, 2015 running on a Windows 7 64 bit, RAM 8 GB, Dual Core 2.67 GHz computer platform) from Advanced Chemistry Development Inc.^{6–9} The program generated 152,400 candidate structures in just 7 minutes and 32 seconds. Of these preliminary structures, the list of candidates was filtered down to 61 structures using the “fast increment-deviation” statistics, $d_{\text{FI}}(^{13}\text{C})$ 4 ppm/carbon, and these were selected for further evaluation. The remaining set of 61 structures was ranked using *neural net* statistics [$d_{\text{N}}(^{13}\text{C})$] by again comparing predicted ¹³C NMR shifts to experimental NMR data. After this refinement, the eight best structures, *viz.* lowest scores based on the best match of the second iteration of predicted and actual ¹³C chemical shift data, are shown in composite Fig. 1 and in greater detail in Fig. S13 in the Supplementary Material. It was readily seen that the top candidate (Structure #1) of the above eight structures is the only one that is consistent with all of the observed NMR data. The COSY data were submitted after the structure was solved and thus used to help eliminate candidates and verify the top candidate structure.

A closer look at Fig. 1 reveals that Structure #1 is potentially conformationally mobile. Moreover, two pieces of experimental data further supported conformer **1A** over **1B**. First, H-1 (9.13 ppm) was part of a 3-spin system, *vide supra*, and considerably deshielded by the neighboring aldehydic oxygen, relative to H-3 (7.87 ppm) and H-6 (7.94 ppm) in conformer **1A**, in which the aldehyde group was positioned so that H-1 and the aldehydic oxygen were spatially proximate. Second, ROESY data (Table 3) revealed a strong spatial correlation between H-8 and H-6 and a much smaller correlation between H-8 and H-1. The opposite situation with respect to protons 1 and 6 would be expected if the aldehyde group were

rotated by 180 degrees to give conformer **1B**. The ratio of the cross peaks can be seen in the trace of the ROESY contained in Fig. 2.

Molecular models suggested that the steric requirements of conformers **1A** and **1B** should be similar so density functional theory (DFT) calculations were employed to gain insight as to why conformer **1A** is preferred as indicated by the NMR. The calculations were carried out on a model system with the bicyclic pyran scaffold bearing the formyl group (see Supplementary Material). In agreement with experiments, DFT calculations indicated a 1.9–3.4 kcal/mol energetic preference for the conformer with the carbonyl moiety involved in a C-H...O interaction with the H-1 of the pyran ring (**1A**). Mulliken charges indicated that H-1 bears a slightly positive charge that prefers to interact with the electronegative formyl oxygen (distance is 2.64 Å), which effectively locks this conformation.

Also in support of Structure #1 was the fact that the chemical shift of C-7 was observed as a *ca.* 23-Hz doublet in the HMBC spectrum. This required an aldehyde group to be placed at C-7 so that the coupling between the aldehydic proton (H-8) and C-7 covers two bonds since such geminal couplings are known to be extremely large.¹⁰ In addition, H-6 (7.94 ppm) was remarkable in that it exhibited HMBC connectivities to the following eight carbons: the linking C-7' of the two ring systems and all of the possible carbons of the fused-5/6-ring system except C-3.

An examination of the other seven structures showed that these compounds have many common deficiencies: (i) the unusually deshielded proton (9.13 ppm) is not part of an observed 3-spin system [Structures #2, 3, 6–8], (ii) the centrally located proton (7.94 ppm) does not exhibit nearly enough HMBC connectivities [Structures #2, 4, 5, 7, 8], (iii) certain HMBC connectivities would be expected to be seen between an alkenic proton and an alkoxy-bearing carbon [Structures #2–7], where none are actually observed, and (iv) there is no 3-spin proton system [Structures #4 and 5, the latter of which has two 2-spin systems], where one such system is found. In addition, Structure #7 contains a bis-exocyclic-butene system, which is an unusual ring system.

Therefore, the highest-ranked (*i.e.* the candidate with the lowest deviation score) Structure #1 is the only ACD/Labs-generated structure that is consistent with all of the NMR spectral features observed for **1**. Additionally, the intensities of ROESY correlations (Table 3) were indicative not only of the spatially proximate orientation of the aldehydic proton (8) and H-6, *vide supra*, but also suggested that methyl-8' is situated closer to H-6 and H-2' while H-7' was located closer to H-4 and H-6'.

Thus, AD-10 (**1**) is 5-[1-(3,4-dihydroxyphenyl)ethyl]cyclopenta[*c*]pyran-7-carbaldehyde. A subsequent literature search revealed three manuscripts that described compounds with the same fused-5/6-ring systems, which contained aldehyde groups adjacent to the pyranyl-oxygens.^{11–13}

Experimental

General Experimental Procedures

Medium-pressure liquid chromatography was performed on a Buchi Sepacore chromatography system (Flawil, Switzerland). Semi-preparative HPLC was carried out on an Agilent 1100 series with an Agilent DAD spectrophotometer and Agilent XDB-C₁₈ reversed-phase column (5 μ m, 250 $\times$ 10 mm) with an Eclipse XDB-C₁₈ guard column. Sephadex LH-20 (GE Healthcare Bio-Sciences, Uppsala, Sweden) and ODS (S-50 and 75 μ m, YMC Co., Ltd., Kyoto, Japan) were used for regular and medium-pressure column chromatography fractionations. Optical rotations: PE-341 polarimeter. IR spectra: Nicolet-Magna-FT-IR 750 spectrometer. UV spectra: Shimadzu UV-2450 spectrophotometer. HR-ESI-MS: Thermo Fisher Scientific LTQ Orbitrap XL mass spectrometer. ¹H and ¹³C NMR were recorded on a Varian VNMRs-600MHz spectrometer (Agilent Technologies, Santa Clara, CA) operating at 600 MHz for ¹H NMR and 150 MHz for ¹³C NMR at room temperature. The chemical shifts (δ) were reported in ppm and were referenced to the residual solvent peak. The coupling constants (*J*) were quoted in hertz. NMR data processing was performed using standard Agilent, and ACD/Labs software.

Plant material

The fruit of *A. digitata* were collected from Abagana, Anambra State, Nigeria in July 2013, and identified by Mr. A. Ozioko. A voucher specimen (No. INTERCEDD0613) has been deposited in Food Composition and Methods Development Laboratory, Beltsville Human Nutrition Research Center, Agricultural Research Service.

Extraction and isolation

The air-dried fruits of *A. digitata* (0.8 kg) were powdered and extracted three times with 70% (v/v) EtOH-H₂O (for one hour each time) to give 75 g of crude extract. The extract was dissolved in 750 mL of H₂O to form a suspension and successively partitioned with ethyl acetate (750 mL $\times$ 3) and n-butanol (750 mL $\times$ 3). The ethyl acetate fraction (17.2 g) was subjected to Buchi Sepacore Easy Purification Systems over ODS using a gradient of methanol/H₂O (30% $\rightarrow$ 60% $\rightarrow$ 100%, v/v) to give three fractions 1–3. Fraction 3 (100% methanol; 1.6 g) was purified by semi-preparative-HPLC using acetonitrile/H₂O (30:70, v/v, 1.5 ml/min) to afford this compound (11.2 mg).

NMR experiments

All NMR experiments were performed on 11.2 mg of the compound dissolved in 600 μ L methanol-*d*₄ 99.8% (Cambridge Isotope Laboratories, Inc., Andover, MA.) No further preparation was performed on the sample prior to NMR analysis. The experimental and processing parameters for each experiment are as follows. ¹H 1D: 32 scans (NS), 1.70 s acquisition time (AT) with 2.0 second relaxation delay (RD), 4.0 μ s pulse width (PW) for a 45° excitation, 9615.4 Hz spectral width (SW) with 16 K points (NP) and a 32 K FT size (zero filling). ¹³C 1D: NS 8000, AT 0.865 s, RD 1.0 s, PW 4.0 μ s (45° excitation), SW 37878.8 Hz, NP 32 K, FT size 64 K, with a 0.5 Hz line broadening. HSQC: Agilent pulse sequence gHSQCAD, NS 32, AQ 0.150 s, RD 1.0 s, ¹H PW 8.0 μ s, ¹³C PW 8.0 μ s, ¹H SW

5387.9 Hz, ^{13}C SW 27146.2 Hz, NP 808, number of increments (NI) 258, FT size 1024×1024 , apodization: 90° shifted sine bell in ^1H and 90° shifted sine square in ^{13}C . HMBC: Agilent pulse sequence gHMBCAD, NS 52, AQ 0.150 s, RD 1.0 s, ^1H PW 8.0 μs , ^{13}C PW 8.0 μs , ^1H SW 5387.9 Hz, ^{13}C SW 27146.2 Hz, NP 808, NI 200, FT size 1024×1024 , apodization: 90° shifted sine bell in ^1H and 90° sine square in ^{13}C . HSQC and HMBC experiments were run in echo-antiecho mode. COSY: Agilent pulse sequence gCOSY, NS 4, AQ 0.150 s, RD 1.0 s, PW 8.0 μs , SW 6410.3 by 6410.3 Hz, NP 962, NI 256, FT size 1024×1024 , apodization: sine square in both dimensions. ROESY: Agilent pulse sequence ROESYAD, NS 8, AQ 0.150 s, RD 1.0 s, PW 8.0 μs , SW 6410.3 by 6410.3 Hz, NP 962, NI 128, FT size 1024×1024 , apodization: sine square in both dimensions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research is supported by the Agricultural Research Service of the U.S. Department of Agriculture and an Interagency Agreement with the Office of Dietary Supplements of the National Institutes of Health. It was also financially supported by the National Natural Science Foundation of China (NO. 81502941) and Zhejiang Provincial Natural Science Foundation of China (NO. LY16H300006).

References

1. Kamatou GPP, Vermaak I, Viljoen AM. S. Afr. J. Bot. 2011; 77:908–919.
2. Chadare FJ, Linnemann AR, Hounhouigan JD, Nout MJR, Van Boekel MAJS. Crit. Rev. Food Sci. Nutr. 2009; 49:254–274. [PubMed: 19093269]
3. De Caluwe E, Halamova K, Van Damme P. Afr. Focus. 2010; 23:11–51.
4. Buchmann C, Prehlsler S, Hartl A, Vogl CR. Ecol. Food Nutr. 2010; 49:145–172. [PubMed: 21883078]
5. Ohmori K. The Chemical Record. 2011; 11:252–259. [PubMed: 21898781]
6. ACD/Structure Elucidator, version 14.04. Toronto, Ont., Canada: Advanced Chemistry Development, Inc.; 2015. www.acdlabs.com
7. Naman CB, Li J, Moser A, Hendrycks JM, Benetrehina PA, Chai H, Yuan C, Keller WJ, Kinghorn AD. Org. Lett. 2015; 17:2988–2991. [PubMed: 26030740]
8. Lorenc C, Sauri J, Moser A, Buevich AV, Williams AJ, Williamson RT, Martin GE, Peczu WM. ChemistryOpen. 2015; 4:577–580. [PubMed: 26491634]
9. Williams RB, O'Neill-Johnson M, Williams AJ, Wheeler P, Pol R, Moser A. Org. Biomol. Chem. 2015; 13:9957–9962. [PubMed: 26381222]
10. Shapiro BL, Kopchik RM, Ebersole SJ. J. Chem. Phys. 1963; 39:3154–3155.
11. Joshi KC, Singh P, Taneja S, Cox PJ, Howie RA, Thompson RH. Tetrahedron. 1982; 38:2703–2708.
12. Kumar US, Tiwari AK, Reddy SV, Aparna P, Rao RJ, Ali AZ, Rao JM. J. Nat. Prod. 2005; 68:1615–1621. [PubMed: 16309309]
13. Liu C-L, Cheng L, Kwok H-F, Ko C-H, Lau T-W, Koon C-M, Zhao M, Lau C-P, Lau K-M, Wong C-W, Leung P-C, Fung K-P, Lau CB-S. J. Ethnopharmacology. 2011; 137:1323–1327. [PubMed: 21843616]

<p>1</p> <p>$d_N(^{13}\text{C}): 2.044$</p>	<p>2</p> <p>$d_N(^{13}\text{C}): 3.002$</p>	<p>3</p> <p>$d_N(^{13}\text{C}): 3.367$</p>	<p>4</p> <p>$d_N(^{13}\text{C}): 3.777$</p>
<p>5</p> <p>$d_N(^{13}\text{C}): 3.798$</p>	<p>6</p> <p>$d_N(^{13}\text{C}): 3.984$</p>	<p>7</p> <p>$d_N(^{13}\text{C}): 4.792$</p>	<p>8</p> <p>$d_N(^{13}\text{C}): 5.014$</p>

Fig. 1.

The top eight candidates generated by ACD/Structure Elucidator for attempt #1. Structure ranking is based on overall differences between experimental and predicted ^{13}C chemical shifts. The lower the $d_N(^{13}\text{C})$ value, the better the ranking.

Fig. 2.

A section of the ROESY spectrum showing the correlations for protons H-1, H-6 and H-8 as indicated. The red peaks are negative and the blue are positive. A horizontal trace of H-8 is overlaid in black showing the much stronger correlation between H-6 and H-8 than for H-1 and H-8 indicating conformation 1A over 1B.

Fragment 1.

Table 1Chemical shifts of **1** at 600 MHz (^1H) and 150 MHz (^{13}C)

Position	δ_{C}^a (multiplicity)	δ_{H}^b		HMBC
1	152.09 (CH)	9.13 brs		7.87, 7.94
3	145.15 (CH)	7.87 d (5.3)		7.17, 9.13
4	111.09 (CH)	7.17 dd (5.3, 0.7)		7.87, 7.94
4a	132.83 (C)			7.17, 7.87, 7.94, 9.13
5	132.56 (C)			1.63, 4.22, 7.17, 7.94
6	147.7 br. (CH)	7.94 s		4.22
7	124.25 (C)			7.94, 9.13, 9.73 ^c
7a	125.09 (C)			7.17, 7.94, 9.13, 9.73
8	186.14 (CH)	9.73 s		7.94
1'	139.40 (C)			1.63, 4.22, 6.69
2'	115.61 (CH)	6.66 d (2.1)		4.22, 6.60
3'	146.40 (C)			6.66, 6.69
4'	144.69 (C)			6.60, 6.66
5'	116.43 (CH)	6.69 d (8.2)		None
6'	119.62 (CH)	6.60 dd (8.2, 2.1)		4.22, 6.66
7'	38.12 (CH)	4.22 q (7.2)		1.63, 6.60, 6.66, 7.94
8'	22.85 (CH ₃)	1.63 d (7.2)		4.22

^aReferenced to tetramethylsilane *via* CD₃OD (δ 49.15).^bReferenced to tetramethylsilane *via* CD₃OD (δ 3.31).^cHMBC correlation appears a *ca.* 26-Hz doublet.

Table 2**COSY Correlations of Compound 1**

Position	δ_{H}	COSY Correlations
1	9.13 brs	7.87 (weak), 7.17 (weak)
3	7.87 d (5.3)	9.13 (weak), 7.17
4	7.17 dd (5.3, 0.7)	9.13 (weak), 7.87
6	7.94 s	(none)
8	9.73 s	(none)
2'	6.66 d (2.1)	6.60
5'	6.69 d (8.2)	6.60
6'	6.60 dd (8.2, 2.1)	6.66, 6.69
7'	4.22 q (7.2)	1.63
8'	1.63 d (7.2)	4.22

Table 3**ROESY Correlations of Compound 1**

Position	δ_{H}	ROESY Correlations ^a
1	9.13 brs	9.73 vw
3	7.87 d (5.3)	7.17
4	7.17 dd (5.3, 0.7)	7.87, 4.22 ms, 1.63 w
6	7.94 s	9.73 m, 6.66 w, 1.63 s
8	9.73 s	7.94 s, 9.13 vw
2'	6.66 d (2.1)	7.94 w, 4.22 mw, 1.63 s
5'	6.69 d (8.2)	6.60 s
6'	6.60 dd (8.2, 2.1)	6.69 s, 4.22 m, 1.63 w
7'	4.22 q (7.2)	7.94 w, 7.17 ms, 6.66 mw, 6.60 s, 1.63 s
8'	1.63 d (7.2)	7.94 s, 7.17 w, 6.66 s, 6.60 m, 4.22 s

^a s = strong, m = medium, w = weak, ms = medium strong, mw = medium weak, vw = very weak