Validation of Structures of Novel Eudesmane Sesquiterpenes Using Scatter Plots

Taye T. Alawode¹* and Kehinde O. Alawode²

¹Department of Chemical Sciences, Federal University Otuoke, Bayelsa State, Nigeria.
²Department of Electrical and Electronic Engineering, Osun State University, Osogbo, Osun State, Nigeria.

Authors' contributions
This work was carried out in collaboration between both authors. Both authors read and approved the final manuscript.

Article Information
DOI: 10.9734/BJAST/2015/15724
(1) Changle Chen, Department of Polymer Science and Engineering, University of Science and Technology of China, China.
(2) Mohammad Jamshed Ahmad Siddiqui, International Islamic University Malaysia, Kuantan, Malaysia.
(2) Ala Uddin, Department of Chemistry, Bacha Khan University Charsadda, Pakistan.
Complete Peer review History: http://www.sciencedomain.org/review-history.php?id=770&id=5&aid=7843

ABSTRACT

Aim: This study explores the potential of scatter plots as a tool in validating proposed structures for novel Eudesmane Sesquiterpenes.

Methodology: Substituents on the skeletons of several Eudesmane compounds were coded and plotted against the ¹³C chemical shift values for each Carbon position on the skeleton (C₁-C₁₅).

Results: The range of chemical shift values (for each Carbon position) over which each substituent type may be obtained was determined from the scatter plots. The results imply that when the carbon atom C₁ on a novel eudesmane compound is assigned any chemical shift value between 26.1 and 54.0, then that position should definitely be without a substituent. Chemical shift values between 68.1 and 91.3 (on C₁) would indicate that β-OH as the most likely substituent (with 23.29% probability) while values within the 121.7 – 160.4 range indicate with 100% certainty that the substituent is Α¹. Similar conclusions can be drawn for all the chemical shift ranges for the different carbon positions.

Conclusion: These chemical shift ranges could be useful in validating proposed structures for novel Eudesmane sesquiterpenes.

*Corresponding author: E-mail: onatop2003@yahoo.com;
1. INTRODUCTION

Sesquiterpenes are formed from countless biogenetic pathways and therefore produce several types of carbon skeletons. This makes the elucidation of their structures very challenging. The biological activities exhibited by sesquiterpenes (including compounds that are insect growth regulators, antifeedant, antifungal, antitumor, antibacterials) makes relating their structures to function even more imperative. The current study focuses on Eudesmane-type compounds which are one of the most representative skeletons of sesquiterpenes. This class of compounds has been the subject of numerous phytochemical, pharmacological and synthetic studies [1-2].

The structure of any natural product is conventionally divisible into three sub-units: (i) the skeletal atoms; (ii) heteroatoms directly bonded to the skeletal atoms or unsaturations between them; and (iii) secondary carbon chains, usually bound to a skeletal atom through an ester or ether linkage [3]. Procedures that could be employed for the identification of the skeleton and substructures present in a compound have been previously described [4-7]. Artificial Neural Networks (ANNs) methods have been reported to give fast and accurate results for identification of skeletons and for assigning unknown compounds among distinct fingerprints (skeletons) of aporphine alkaloids [8]. In a previous work, we have shown that Generalized Regression Neural Networks (GRNN) could predict substituents types and positions on Eudesmane-type sesquiterpenes [9]. When the chemical shift values proposed for each of the fifteen (15) Carbon positions on the Eudesmane skeleton is used as input for the GRNN, this procedure could be used in validating the structures of novel Eudesmanes. In the current work, we use scatter plots to determine the $^{13}$C chemical shift ranges (for the 15 carbon atoms on the Eudesmane sesquiterpene skeleton—shown in Fig. 1) over which different substituent types may exist. We discuss its potential application in validating structures proposed for natural products using Eudesmane sesquiterpenes as reference.

2. METHODOLOGY

The structural (skeletal) $^{13}$C data, substituents and stereochemical information of 325 compounds (out of 350 compounds) reviewed and published by Olievera et al. [1] was used in this study. Twenty-five of these compounds were left out owing to their structural complexity. This information can be extracted from data of Eudesmane sesquiterpenes published in literature by isolating $^{13}$C values of the skeletal (carbons) from those of the substituents.

Each substituent type (on first encounter) was assigned 3 number codes. These codes serve to identify the substituent while also taking into account its possible stereochemistry (α or β) in various positions of the skeletons in other compounds.

Fig. 1. The eudesmane skeleton
Carbon positions without substituents were assigned a code of 0 while α and β positions without substituent(s) were assigned codes of 1 and 2 respectively. For example, OH group was given a code of 3, an α-OH is given a code of 4 while a β-OH was assigned a code of 5. (The different substituent types and the corresponding codes assigned to them are shown in Appendix 1). Thereafter, 30 columns containing, alternately, all the possible $^{13}$C chemical shift data for each of the 15 positions (C_1-C_{15}) on the Eudesmane skeleton for all the 325 compounds and the corresponding codes for the substituents attached to each position in each of the compounds, were prepared on an Excel sheet. A scatter plot of the codes (of the substituents) against their corresponding chemical shift values for each Carbon position on the skeleton (C_1-C_{15}) was plotted. From this, the range of chemical shift values (for each Carbon position) over each substituent type may be obtained was determined. Where there are multiple possible substituent types within a particular carbon range, the probability (in percentages) that a substituent would occupy this position was determined relative to the total number of points within the range.

### 3. RESULTS AND DISCUSSION

Fig. 2 shows the scatter plots of codes of substituents against their corresponding chemical shift values. From this, the chemical shift ranges characteristic of each substituent type on each of the fifteen (15) carbon atoms on the Eudesmane skeleton was obtained.

Scatter plots (in conjunction with other CASE procedures) have previously been used by Elyashberg et al. [10] in the revision of the structure of Asperjinone. The authors performed a search for the (3,3-dimethyloxiran-2-yl) methyl fragment in the ACD/NMR Database containing 425,000 structures with assigned $^{13}$C and $^1$H chemical shifts. The program selected 180 structures of which about 150 structures exhibiting the closest similarity with the environment of the oxirane fragment were chosen. For these structures, a scatter plot was created. Inspection of the scatter plot convincingly confirms the incorrectness of the original structure.
C6

Codes (of substituents)

Chemical Shifts

C7

Codes (of substituents)

Chemical Shifts
Fig. 2. Scatter plots of codes (of substituents) against $^{13}$C chemical shifts
Oliveira et al. [1] described the use of two component programs (TIPCARB and PICKUP) of the system, SISTEMAT, in the search for heuristic rules (practical rules obtained from the experience of specialists, or originated from programs which perform “learning from machine” routines, and are aimed at solving a specific problem). TIPCARB can determine which carbon atom is present in each position on a skeleton whether or not a carbon atom is substituted and the kind of substituent. After the position and types of substituents attached to each carbon atom have been defined, the fragments, denominated substructures, are coded in the PICKUP program that performs the search of the database for the chemical shift range for $^{13}$C data of the carbons in the substructure. The authors then utilized the PICKUP program to determine several chemical shift ranges that characterize several substructures present in eudesmanes. A summary of the substituent types that may be obtained over different $^{13}$C ranges for each of the fifteen (15) positions on the Eudesmane skeleton using scatter plots are presented in Table 1.

| Skeletal carbon | Chemical shift range | Codes of substituents (%) |
|-----------------|----------------------|---------------------------|
| C₁              | 26.1 - 54.0          | Nil(100)                  |
|                 | 68.1 - 91.3          | OH(2.74), α-OH(4.11), β-OH(23.29), α-Ocin(2.05), β-Ocin(2.05), OAc(13.70), α-OAc(3.42), β-OAc(13.01), α-Oxy(1.37), β-Oxy(0.68), α-Ogly(0.68), α-Ogly (OAc)$_4$(0.68), OBzt(3.42), α-OBzt(8.22), β-O Bzt (5.48), α-ONic(2.74), β-ONic (2.05), α-OEpcin(0.68), β-OFur(1.37), Oib(0.68), β-Oib(0.68), β-OBut(2.05), β-O2MeBu(0.68), β-O(O-H--Dihydrocou)-2.05, β-O(α-OH--Iva)-0.68, OPro(0.68), β-O(Val-2‘OH)-0.68 |
|                 | 121.7 - 160.4        | Δ '(100)                  |
|                 | 206.3 - 216          | Oxo(100)                  |
| C₂              | 17.0 - 46.4          | Nil(100)                  |
|                 | 67.4 - 78.4          | OH(3.45), α-OH(6.90), β-OH(10.34), OAc(5.17), α-OAc(8.62), β-OAc(15.52), β-Ogly(oac)$_x$(1.72), OBzt(1.72), α-OBzt(1.72), β-OBzt (10.34), OFur(1.72), α-OFur(1.72) β-OFur(3.45), Oib(3.45), β-Oib(3.45), OBzt(1.72), β-OBzt (8.62), β-OBzt --(2’-Me)-1.72, β-OMeBu(3.45), OHex(1.72), β-OGly--(2’-OAc)-1.72, α-OMe(1.72), α-Peroxy(1.72) |
|                 | 120.5-128.5          | Nil(minor)                |
|                 | 126.2-133            | Δ '(minor)                |
|                 | 192.1 - 210          | Oxo                       |
| C₃              | 21.6 - 54.7          | Nil(100)                  |
|                 | 69.9 -85.1           | OH(3.45), α-OH(17.24), β-OH(10.34), α-OAc(3.45), β-OAc(6.90), Ogly(3.45), β-Ogly(3.45), α-EPang(10.34), β-OEpa(3.45), α-OOH(3.45), α-OAng(6.90), β-OAng(20.69), α-O2MeBu-(2’OAc,3’OH)(6.90) |
|                 | 127.0-134.0          | Nil(12.5)                 |
|                 | 116.8-152.7          | Δ '(87.5)                 |
|                 | 180.8 – 213.1        | Oxo(Nil)                  |
| C₄              | 31.5 – 48.6          | Nil(100)                  |
|                 | 65.9- 87.7           | OH(4.17), α-OH(29.17), β-OH(37.5), β-Ocin(4.17), α-OAc(8.33), β-OAc(4.17), α-Oxy(3.13), β-Oxy(1.04), α-Eopoxy(2.08), β-OFuc(1.04), β-OFuc(2’OmeBu)-1.04, β-OFuc(OAc)$_3$-1.04, β-O Fuc(2’OmeBu,3’OAc)(1.04), β-OFuc(3’4’-o-isopropylidene)-1.04, β-OFuc(2’OmeBu,3’4’-o-isopropylidene)-1.04 |
|                 | 123.8 – 167.0        | Nil(27.59), Δ ' (40.23), Δ '' (32.18) |

Table 1. Chemical shift ranges for substituents on the eudesmane skeleton
| Skeletal carbon | Chemical shift range | Codes of substituents (%) |
|-----------------|----------------------|---------------------------|
| C₁₁             | 23.1 – 42.6          | Nil(20), α(54.5), β(58.18), Δ₁₁(3.64), OH, β(54.5), Δ₁₁, β-(7.27) |
| C₁₂             | 14.4 – 31.9          | Nil(100)                  |
| C₁₃             | 8.2 – 32.4           | Nil(94.85), α(5.15)       |
| C₁₄             | 9.9 – 29.9           | Nil(24.1), α(31.93), β(43.98) |
|                 | 51.4 – 76.5          | OH(9.09), OAc(9.09), Oxy(9.09), OGly(9.09), Peroxy(9.09), Oxy, α(9.09), OGly, α(2.73), OH, α(9.09), Epoxy, β(9.09) |
|                 | 105.3 – 118.5        | Nil(100)                  |
|                 | 168.7 – 170.9        | Oxo, OMe(minor)           |
|                 | 190.7-194.8          | Oxo(minor)                |
From the Table, it can be inferred, for example that when the carbon atom on position 1 (C₁) on a novel eudesmane compound is assigned any chemical shift value between 26.1 and 54.0, then that position should definitely be without a substituent. Chemical shift values between 68.1 and 91.3 would indicate that β-OH as the most likely substituent (with 23.29% probability) while other substituents shown on the Table would have lesser probabilities of occurrence. Chemical shift values within the 121.7 – 160.4 range indicate with 100% certainty that the substituent is Δ¹. When the carbon on position 2 (C₂) on a novel eudesmane compound is assigned a chemical shift value between 17.0 and 46.4, the position is definitely without a substituent. It could be observed that when a chemical shift value between 67.4 and 78.4 is assigned to this position, all the possible substituents (listed against this position on the Table) have very similar probabilities of occupying this position. It should be noted, however, that this procedure has successfully reduced the number of likely substituents for this position from about 215 to the 22 reflecting on the Table. Again, the stereochemistry of each substituent type has been taken into consideration in assigning codes to the substituents (reflected as α or β on the Table). The user would be able to reach a conclusion (without regard to stereochemistry) that the substituents OH and OAc have 20.69% and 29.31% probabilities of occurring in this position.

4. CONCLUSION

With the availability of sufficiently broad database on the ¹³C chemical shift values of the carbon atoms on the Eudesmane skeleton, scatter plots may be a useful complementary tool in the elucidation of structure of this class of compounds.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Oliveira FC, Ferreira MJP, Nuñez CV, Rodriguez GV, Emerenciano VP. ¹³C NMR spectroscopy of Eudesmane Sesquiterpenes. Progress in Nuclear Magnetic Resonance Spectroscopy. 2000; 37: 1-45.
2. Ferreira MJP, Oliveira FC, Rodrigues GV, Emerenciano VP. ¹³C NMR pattern recognition of Guaiane Sesquiterpenes. Internet Electr. J. Mol. Des. 2004;3(11):737-749.
3. Rodrigues GV, Campos IPA, Emerenciano VP. Applications of artificial intelligence to structure determination of organic compounds**. Determination of groups attached to skeleton of natural products using ¹³C nuclear magnetic resonance spectroscopy. Spectroscopy. 1997;191-200.
4. Ferreira MJP, Rodrigues GV, Brant AJC, Emerenciano VP. REGRAS: An auxiliary program for pattern recognition and substructure elucidation of monoterpenes. Spectroscopy. 2000;15:65–98.
5. Ferreira MJP, Brant AJC, Rodrigues GV, Emerenciano VP. Automatic identification of terpenoid skeletons through 13C nuclear magnetic resonance data disfunctionalization. Analytica Chimica Acta. 2001;429:151–170.

6. Ferreira MJP, Oliveira FC, Alvarenga SAV, Macari PAT, Rodrigues GV, Emerenciano VP. Automatic identification by 13C NMR of substituent groups bonded in natural product skeletons. Computers & Chemistry. 2002;26:601–632.

7. Ferreiraa MJP, Emerenciano VP, Linia GAR, Romoff P, Macarib PAT, Rodrigues GV. 13C NMR spectroscopy of monoterpenoids. Progress in Nuclear Magnetic Resonance Spectroscopy. 1998; 33:153–206.

8. Rufino AA, Brant AJC, Santos JBO, Ferreira MJP, Emerenciano VP. Simple method for identification of aporphine alkaloids from 13C NMR data using artificial neural networks. J. Chem. Inf. Model. 2005;45:645-651.

9. Alawode TT, Alawode KO. Prediction of substituent types and positions on skeleton of Eudesmane-type sesquiterpenes using generalized regression neural networks. African Journal of Pure and Applied Chemistry. 2014;8(7):102-109.

10. Elyashberg M, Blinov K, Molodstov S, Williams AJ. Structure Revision of Asperjinone Using Computer-Assisted Structure Elucidation Methods. Journal of Natural Products. 2013;76(1):113-116.
### APPENDIX

#### Appendix 1. Substituents and their corresponding codes

| Substituent | Code | Substituent | Code | Substituent | Code | Substituent | Code |
|-------------|------|-------------|------|-------------|------|-------------|------|
| Nil         | 0    | α-O-cis-(3’-OAc-2-butenoate) | 26   | β-OBut-(2’-Me) | 51   | 4-O(CH$_2$)O-4 | 76   |
| α           | 1    | β-O-cis-(3’-OAc-2-butenoate) | 27   | OMeBu       | 52   | α-4-O(CH$_2$)O-4 | 77   |
| β           | 2    | OBzt        | 28   | α-OMeBu     | 53   | β-4-O(CH$_2$)O-4 | 78   |
| OH          | 3    | α-OBzt      | 29   | β-OMeBu     | 54   | Nor         | 79   |
| α-OH        | 4    | β-OBzt      | 30   | O2MeBu      | 55   | α-Nor       | 80   |
| β-OH        | 5    | ONic        | 31   | α-O2MeBu    | 56   | β-Nor       | 81   |
| OCin        | 6    | α-ONic      | 32   | β-O2MeBu    | 57   | OMe         | 82   |
| α-OCin      | 7    | β-ONic      | 33   | OHex        | 58   | α-OMe       | 83   |
| β-OCin      | 8    | OEpacin     | 34   | α-OHex      | 59   | β-OMe       | 84   |
| OAc         | 9    | α-OEpcin    | 35   | β-OHex      | 60   | Oxo, OMe    | 85   |
| α-OAc       | 10   | β-OEpcin    | 36   | Oxo         | 61   | O-trans-Cou | 86   |
| β-OAc       | 11   | OFur        | 37   | α-Oxo       | 62   | α-O-trans-Cou | 87   |
| Oxy         | 13   | α-OFur      | 38   | β-Oxo       | 63   | β-O-trans-Cou | 88   |
| α-Oxy       | 14   | β-OFur      | 39   | OGly-(2’,6’-OAc) | 64  | $\Delta^3$ | 90   |
| β-Oxy       | 15   | OPic        | 40   | α-OGly-(2’,6’-OAc) | 65  | $\Delta^3$ | 90   |
| OGly        | 16   | α-OPic      | 41   | β-OGly-(2’,6’-OAc) | 66  | GOgly[(OAc)$_3$]6’-OTig | 91 |
| α-OGly      | 17   | β-OPic      | 42   | Trinor      | 67   | α-OGly[(OAc)$_3$]6’-OTig | 92 |
| β-OGly      | 18   | OiBu        | 43   | α-Trinor    | 68   | β-OGly[(OAc)$_3$]6’-OTig | 93 |
| OGly-(OAc)$_4$ | 19   | α-OiBu      | 44   | β-Trinor    | 69   | OXyl-(OAc)$_3$ | 94 |
| α-OGly-(OAc)$_4$ | 20   | β-OiBu      | 45   | OEpang      | 70   | OXyl-(OAc)$_3$ | 95 |
| β-OGly-(OAc)$_4$ | 21   | OBut        | 46   | α-OEpang    | 71   | α-OXyl-(OAc)$_3$ | 96 |
| O-trans-(3’-OAc-2-butenoate) | 22   | α-OBut      | 47   | β-OEpang    | 72   | β-OXyl-(OAc)$_3$ | 97 |
| α-O-trans-(3’-OAc-2-butenoate) | 23   | β-OBut      | 48   | Epoxy       | 73   | OOH         | 98   |
| β-O-trans-(3’-OAc-2-butenoate) | 24   | OBut-(2’-Me) | 49   | α-Epoxy     | 74   | α-OOH       | 99   |
| O-cis-(3’-OAc-2-butenoate) | 25   | α-OBut-(2’-Me) | 50   | β-Epoxy     | 75   | β-OOH       | 100  |
### Appendix 1. (Continues): Substituents and their corresponding codes

| Substituent | Code | Substituent | Code | Substituent | Code |
|-------------|------|-------------|------|-------------|------|
| Δ¹(14) β-α-OMeAcr- (4’OH) | 101 | Δ²(10) O(α-2’OH) | 126 | Δ³(10) [O Gly(OAc)₃⁻ (2’O Gly(OAc)₄)] | 152 |
| O Ara | 102 | O(α-2’OH-iVa) | 127 | Cin | 153 |
| α-O Ara | 103 | α-α(O-α-2’OH-iVa) | 128 | α-Cin | 154 |
| β-O Ara | 104 | β-α(O-α-2’OH-iVa) | 129 | β-Cin | 155 |
| Δ⁴ | 105 | O Ac | 130 | Br | 156 |
| NC | 106 | α-OFuc | 131 | α-Br | 157 |
| α-NC | 107 | β-OFuc | 132 | β-Br | 158 |
| β-NC | 108 | OFuc-(2’OMeBu) | 133 | OFuc | 159 |
| Δ⁵ | 109 | OFuc-(2’OMeBu) | 134 | α-OPro | 160 |
| O Ang | 110 | β-OFuc-(2’OMeBu) | 135 | β-OPro | 161 |
| α-O Ang | 111 | OFuc-(2’OMeBu3’4’ OAc) | 136 | H | 162 |
| β-O Ang | 112 | α-OFuc-(2’OMeBu3’4’ OAc) | 137 | α-H | 163 |
| O2 MeBu-(2’OMeBu3’4’ OAc) | 113 | β-OFuc-(2’OMeBu3’4’ OAc) | 138 | β-H | 164 |
| α-O2 MeBu-(2’OMeBu3’4’ OAc) | 114 | OFuc-(2’OMeBu3’4’ OAc) | 139 | OFuc-(2’OMeBu3’4’ OAc) | 165 |
| β-O2 MeBu-(2’OMeBu3’4’ OAc) | 115 | α-OFuc-(2’OMeBu3’4’ OAc) | 140 | α-OFuc-(2’OMeBu3’4’ OAc) | 166 |
| Δ⁶ | 116 | β-OFuc-(2’OMeBu3’4’ OAc) | 141 | β-OFuc-(2’OMeBu3’4’ OAc) | 167 |
| Δ⁷ | 117 | OFuc(3’4’ Oisopropylidene) | 142 | OFuc(3’4’ Oisopropylidene) | 168 |
| Δ⁷(11) | 118 | β-OH, α-Oxy | 143 | β-OFuc(3’4’ Oisopropylidene) | 169 |
| β-OH, α-Oxy | 119 | β-OFuc(3’4’ Oisopropylidene) | 144 | X: Oxo, OH | 170 |
| Δ⁸ | 120 | Peroxy | 145 | O (Val-2’OH) | 171 |
| O(α-2’OH-di hydrocou) | 121 | α-Peroxy | 146 | α-O(Val-2’OH) | 172 |
| α-O(α-2’OH-dihydrocou) | 122 | β-Peroxy | 147 | β-O(Val-2’OH) | 173 |
| β-O(α-2’OH-dihydrocou) | 123 | OTig | 148 | O Gly(2’-O Gly) | 174 |
| OMeAcr-(4’OH) | 124 | α-OTig | 149 | α-O Gly(2’-O Gly) | 175 |
| α-OMeAcr-(4’OH) | 125 | β-OTig | 150 | β-O Gly(2’-O Gly) | 176 |
### Appendix 1. (Continues): Substituents and their corresponding codes

| Substituent | Code  | Substituent | Code  | Substituent | Code  | Substituent | Code  |
|-------------|-------|-------------|-------|-------------|-------|-------------|-------|
| \(\Delta^{11}, 11\alpha\) | 202   | 14OH, 14\(\alpha\) | 206   | 11OCin, 11\(\alpha\) | 210   | 15-\(\alpha\)OAc, 15\(\alpha\) | 214   |
| \(\Delta^{11}, 11\beta\)  | 203   | 14OH, 14\(\beta\) | 207   | 11OCin, 11\(\beta\) | 211   | 15-\(\alpha\)OAc, 15\(\beta\) | 215   |
| 14-OGly, 14\(\alpha\)  | 204   | 14Epoxy, 14\(\alpha\) | 208   | 15-\(\alpha\)OH, 15\(\alpha\) | 212   |
| 14-OGly, 14\(\beta\)  | 205   | 14Epoxy, 14\(\beta\) | 209   | 15-\(\alpha\)OH, 15\(\beta\) | 213   |

© 2015 Alawode and Alawode; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

---

**Peer-review history:**

The peer review history for this paper can be accessed here:

http://www.sciencedomain.org/review-history.php?id=770&id=5&aid=7843