The impact of macrocycle conformation on the taxadiene-forming carbocation cascade: Insight gained from sobralene, a recently discovered verticillene isomer.

Christopher J. Hayes*, Matthew J. Palframan and Gerald Pattenden.

School of Chemistry, University of Nottingham, University Park, Nottingham, NG7 2RD U.K.

ABSTRACT: DFT calculations on the carbocation intermediates that connect the biosynthetic pathways leading to the sand fly pheromone sobralene and taxadiene have been made. Establishment of the conformation of the macrocyclic carbocation intermediate required to produce the cis-C8,C9 alkene bond in sobralene, has identified new conformations of the verticillyl carbocation intermediates on the taxadiene biosynthetic pathway. These "sobralene-like" carbocation conformations provide an exothermic pathway to taxadiene, and are validated by comparison to closely-related structures (x-ray, NMR).

INTRODUCTION

Sobralene **1** is a recently discovered sex-aggregation pheromone produced by populations of the sand fly *Lutzomyia longipalpis* from Sobral, Brazil,[1] which is the main carrier of the Protist parasite *Leishmania infantum*, the causative agent of visceral leishmaniasis (VL). Visceral leishmaniasis is a tropical disease transmitted to humans through the bite of the female sand fly (*Lutzomyia longipalpis*) and it is fatal if untreated. Ninety five percent of new VL cases (estimated at 200-400 thousand per year) occur in just ten countries (Bangladesh, Brazil, China, Ethiopia, India, Kenya, Nepal, Somalia, South Sudan and Sudan), and it is the second largest parasite killer in the world after malaria, resulting in 20,000-40,000 fatalities per year.[2,3] Sobralene **1** has the potential to be used in pheromone attractant traps as a vector control measure. However, it has only been isolated in very small quantities from the sand flies, and methods need to be developed for its large scale production. Inspired by this challenge, we initiated studies exploring sobralene's biosynthetic origin as a first step to producing **1** using well known enzyme over-expression techniques. This approach has been successfully employed for the production of closely related diterpenes such as taxadiene **3**[4-10] and a collection of verticillenes[11,12] (e.g. verticillene **2**), but the production of sobralene **1** has yet to be explored.

The synthase responsible for transforming geranyl geranyl pyrophosphate (GGPP) (**5**) into sobralene **1** is currently not known. However, the fact that taxadiene **3** (a biosynthetic precursor to the anti-cancer drug Taxol **4**[13]) was co-isolated with **1** from sand flies[1] provides strong evidence for a close biosynthetic relationship between **1** and **3** and suggests that sobralene synthase is similar to taxadiene synthase (TXS)[14] (Figure 1). Taxadiene synthase mediates a pathway from GGPP **5** to taxadiene **3** *via* a series of carbocation-mediated cyclisations, *i.e.* macrocyclisation (**A→B**), bridged bicycle formation (**B→C**), transannulation (**D→E**)

and termination by loss of a proton from carbocation **E** (Scheme 1). It is likely that sobralene **1** has its origins in the same carbocation **D** that sits on the taxadiene biosynthetic pathway, with stereoselective loss of a proton from C9 in **D** producing the required *cis*-C8,C9 alkene bond in **1**.
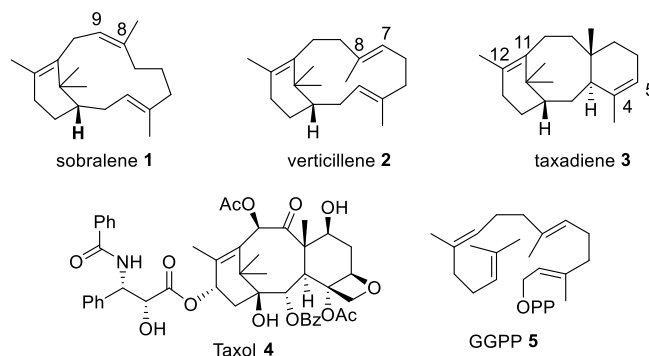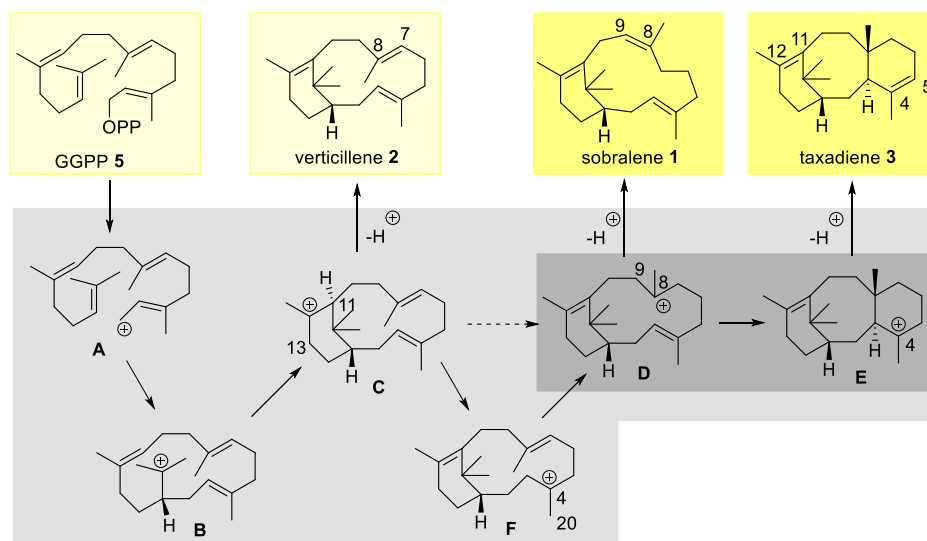

Figure 1: Structures of sobralene **1**, verticillene **2**, taxadiene **3**, Taxol **4** and GGPP **5** (n.b. Taxane atom numbering used throughout).

Motivated by the recent discovery of sobralene **1** and its biosynthetic connection to taxadiene **3**, we decided to investigate the cyclisation cascade shown in Scheme 1 in more detail using computational methods, paying particular attention to the role of carbocation **D**.

Although the carbocation cascade leading to sobralene **1** has not been studied, the pathway to taxadiene **3** from GGPP, outlined in Scheme 1, has been examined computationally, and a variety of methodologies have been explored. Oikawa *et al.*[15] first used an *ab initio* (HF/6-31G(d)) approach to study carbocations **C**, **D** and **F**, but the complete pathway was not studied and transition state structures were not determined. A more comprehensive QM study (MPW1PW91/6-31+G(d,p)//B3LYP/6-31+G(d,p)) of the taxadiene pathway was reported by Tantillo *et al.*[16]

Scheme 1: Terpene synthase-mediated cascades from GGPP 5 leading to sobralene 1, verticillene 2, and taxadiene 3. Proposed structures for key carbocation intermediates are highlighted in grey and isolated products are highlighted in yellow.

where all the carbocations **A** to **E** were examined, along with the transition states connecting them. A key result of this study was that the direct conversion of **C** into **D** was unfavourable in comparison to the two step intramolecular proton transfer sequence **C→F→D**, thereby supporting the involvement of carbocation **F** in the taxadiene synthesis pathway. However, their most intriguing finding was that the lowest energy point on the calculated pathway was the carbocation intermediate **C**, and that the tricyclic carbocation intermediate **E** (the precursor to taxadiene **3**) was significantly higher in energy (+3.8 kcal/mol). A later QM/MM study by Major *et al.*[17] reexamined the **C→F→D** proton transfer process within a model of TXS and these authors concluded that the one step (**C→D**) and the two step (**C→F→D**) processes were energetically similar. In contemporaneous studies, Thiel *et al.*[18] and Major *et al.*[19] used molecular dynamics simulations to study the interactions of the intermediate carbocations with TXS, which showed that pyrophosphate (PPi) plays a crucial role in favouring the formation of taxadiene, and subsequent QM/MM studies provided energy profiles for the whole process.[19,20]

The absence of a protein structure for sobralene synthase precluded QM/MM calculations and MD simulations for our studies, so we selected a QM (DFT) approach. By identifying a conformation of the carbocation intermediate **D** required to install the *cis*-C8,C9 alkene in sobralene **1** we have discovered a significantly more energetically favourable cylisation cascade (**C→D→E**) on the taxadiene pathway than that previously disclosed.[16] In this paper we report the details of our study.

COMPUTATIONAL METHODS

Geometry optimisations were performed using either B3LYP/6-31+G(d,p) or B3LYP/6-31G(d)[25-28] level of theory using Q-Chem as deployed in Spartan 10 (Windows), and single point MPWB1K/6-31+G(d,p)[29] energies were also calculated for the B3LYP/6-31+G(d,p) and B3LYP/6-31G(d) optimised structures. MPWB1K/6-31+G(d,p) was selected as this hybrid functional has been shown to perform well in energy calculations for carbocations, and it also allows for a direct comparison to the energies reported by Tantillo and Hong in their previous study of the taxadiene carbocation cyclisation cascade.[16a] Zero point vibrational energies were calculated for all structures and the absence of imaginary frequencies was used to characterise the structures as minima on their potential energy surfaces, and transition states were confirmed by the presence of a single imaginary frequency consistent with motion along the reaction coordinate. IRC-like calculations, as described by Thiel *et al.*,[20] were performed to confirm that starting materials and products were connected *via* the transition states identified. The Freezing String Method (FSM)[30,31] was used to locate transition states in cases where the standard approach failed. All representations of calculated structures were generated using CYLview.[32]

RESULTS AND DISCUSSION

We first examined the carbocation **D** because this intermediate directly links sobralene **1** and taxadiene **3** (Scheme 1). Since the formation of sobralene **1** has not been examined before, we investigated the conformation of **D** required to accommodate the *cis*-C8,C9 alkene in the macrocyclic ring of **1**. Molecular mechanics on sobralene **1** provided a good starting conformation for the corresponding carbocation ^{sob}**D** from which to conduct QM calculations (the superscript 'sob' is used here to identify the sobralene-derived conformer of **D**). Subsequent geometry optimization using B3LYP/6-31+G(d,p) identified ^{sob}**D** as a local minimum, and the MPWB1K/6-31+G(d,p) single point energy was calculated for this structure (Figure 2). Examination of the ^{sob}**D** structure shows that elimination of the proton Ha on C9 is stereoelectronically favourable, and that this elimination would naturally lead to the *cis*-C8,C9

alkene found in sobralene **1**. As mentioned earlier, the structure of carbocation **D** has been calculated previously by Tantillo and Hong[16] during their investigation of the taxadiene biosynthetic cascade. In contrast to our work, they identified an alternative conformation for **D** that can be described as adopting a verticillene-like conformation **vertD** (the superscript 'vert' is used here to distinguish this conformation from the sobralene-like **sobD** conformer). Crucially, the conformation of **vertD** precludes the direct formation of sobralene **1** because elimination of a proton from C9 in **vertD** would give the incorrect *trans*-C8,C9 stereoisomer of sobralene, and the alternative elimination of Hb from C7 in **vertD** would give the *trans* C7,8 alkene found in verticillene **2** (Figure 2). We therefore propose that the synthase responsible for forming sobralene **1** must carefully control the conformation of carbocation **D**, favouring the **sobD** conformer prior to elimination of the C9 Ha proton.
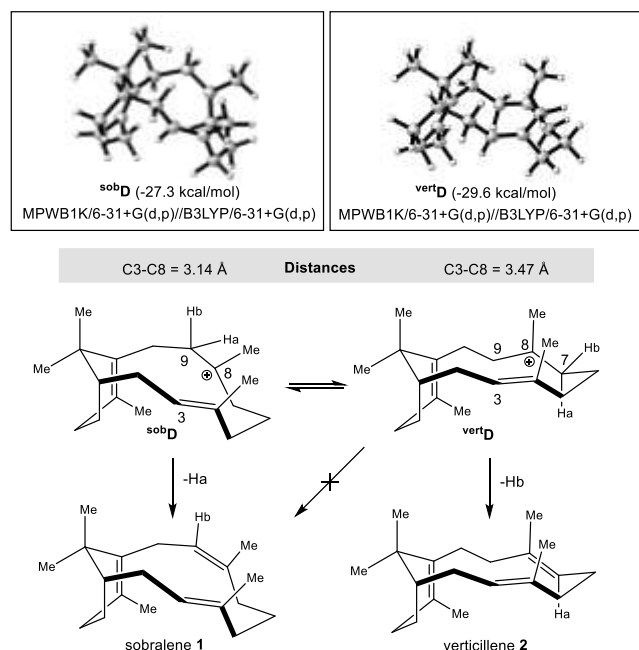

Figure 2: Comparison of the sobralene-like (**sobD**) and verticillene-like (**vertD**)(16) conformations of carbocation **D**. MPWB1K/6-31+G(d,p)//B3LYP/6-31+G(d,p) was used for the calculations and energies are quoted relative to GG+ (cation **A**) at 0.00 kcal/mol.

Recent experiments show that sobralene **1** isomerises to verticillene **2** upon treatment with mild acid *via* cation **D** (i.e. **1**→**sobD**→**2**),[21] thus demonstrating that **2** is more thermodynamically stable than **1**. This finding is consistent with our calculations that show **sobD** is slightly higher (2.3 kcal/mol) in energy than **vertD**. Without the structure of the sobralene synthase, we cannot assess if specific active site residues facilitatate the desired deprotonation at C9 in the carbocation **D** to give sobralene **1**, but it seems likely that the synthase is responsible for guiding carbocation **D** to adopt the **sobD** conformation during the deprotonation to form sobralene **1**. Further inspection of the calculated **sobD** structure showed that the C3-C8 distance was much shorter (3.14 Å) than that previously reported for the **vertD**

conformer (3.47 Å)[16a] (Figure 2), and whilst not relevant to the formation of sobralene **1**, this shortened C3-C8 distance could indicate that transannulation on the taxadiene pathway (*i.e.* **D**→**E**, Figure 3) might be more facile in the sobralene-like versus the verticillene-like conformation (*vide infra*).

Having found the **sobD** conformation, we next explored whether any other key carbocation intermediates on the taxadiene cascade (Scheme 1) could similarly adopt sobralene-like conformations. Following the same methodology, we examined the carbocations **E** (Figure 3) and **C** (Figure 5), and we were able to locate sobralene-like conformations for both. For clarity, the new sobralene-like conformers were named **sobE** and **sobC**, whilst the previously calculated verticillene-like conformers were named **vertE** and **vertC** respectively.

Comparison of our newly calculated sobralene-like conformer **sobE** with the previously reported verticillene-like conformer **vertE**[16] revealed a very significant (and somewhat unexpected) energy difference, with the sobralene-like conformer (**sobE**) being 9.2 kcal/mol *lower* in energy than the verticillene-like equivalent (**vertE**) (Figure 3).

Figure 3: Comparison of the sobralene-like (**sobE**) and verticillene-like (**vertE**)[16] conformations of carbocation **E**. MPWB1K/6-31+G(d,p)//B3LYP/6-31+G(d,p) was used for the calculations and energies are quoted relative to GG+ (cation **A**) at 0.00 kcal/mol.

This important result means that the sobralene-like conformer **sobE** is the lowest energy point on the taxadiene carbocation cascade, which is in contrast to the verticillene-like conformers where **vertC** is the lowest energy point. This means that by utilizing the **sobE** conformation, the carbocation cascade to taxadiene is calculated to be energetically downhill overall. Both **sobE** and **vertE** adopt similar conformations in their A- and C-rings, with the major difference

being between their 8-membered B-rings. As expected for the lowest energy conformer, [sob]E adopts a boat-chair (BC) conformation, and [vert]E adopts a higher energy chair-chair (CC) conformation.[22] The boat-chair conformation of [sob]E allows the C17 and C19 methyl groups to be much further apart (4.80 Å) than in [vert]E (3.98 Å), and the C3-C8 bond length is shorter in [sob]E (1.60 Å) than in [vert]E (1.73 Å) (Figure 3).

Supporting evidence of the relevance and importance of the [sob]E conformation on the taxadiene biosynthetic pathway comes from examining experimental data measured on closely related neutral structures. Firstly, in their early pioneering work, Coates *et al.* measured n.O.e. data on taxadiene **3** itself, and they showed that irradiation of the C19 methyl group (which is attached to C8) gave enhancements to both methylene protons on C9, and also to the β-proton on C2 (Figure 4A).[23] On the basis of these data, Coates *et al.* thereby concluded that taxadiene **3** adopts a sobralene-like, in preference to the alternative verticillene-like, conformation in solution (Figure 4A). In addition, during the first total synthesis of taxadiene **3**, Williams *et al.*[24] measured an X-ray crystal structure of their key synthetic intermediate **9** that contains a ketone at C4 (Figure 4B). The tricyclic intermediate **9** clearly adopts a sobralene-like conformation in the crystal structure, which is strikingly similar to that of [sob]E calculated in this work, with the ketone at C4 in **9** acting as an excellent structural mimic of the carbocation at C4 in [sob]E (Figure 4B).
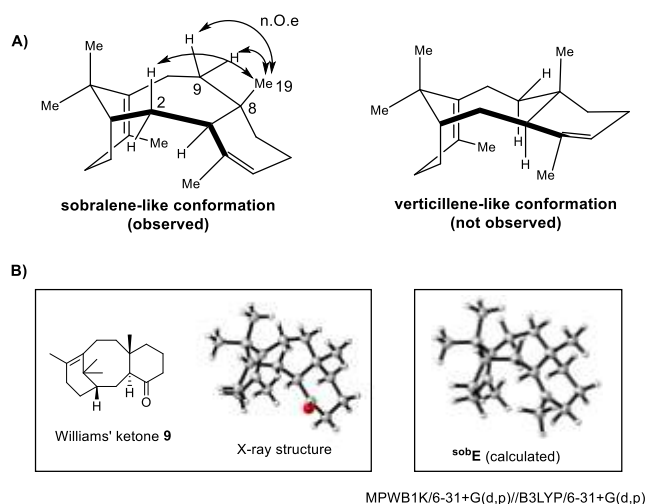
Figure 4: A) n.O.e data showing a sobralene-like conformation in solution for taxadiene **3**;[23] **B)** Comparison of Williams' X-ray crystal structure of ketone **9** (ccdc NAGCON)[24] with the calculated structure of [sob]E.

Having found sobralene-like conformations of carbocations **D** and **E**, we next calculated the corresponding sobralene-like conformer of carbocation **C** (Figure 5). Examination of the structure of [sob]C, shows that the proton on C11 (*i.e.* the one transferred intramolecularly to either the C3,C4 alkene or the C7,C8 alkene) is closer to both C3 (2.26 vs 2.35 Å) and C7 (2.55 vs 2.66 Å) in [sob]C than in [vert]C, thus suggesting that the intramolecular proton transfer to form either of the carbocations **D** or **F** (see Scheme 1) might be

more facile in the sobralene-like conformation than in the verticillene-like conformation. A similar conformation for carbocation **C** has also been identified by Tantillo and Gutta[16b] during their preliminary study of intramolecular proton transfer during taxadiene biosynthesis, and our data is in close agreement with theirs. In order to establish the barrier heights for the proton transfers, and to fully explore the [sob]D to [sob]E cyclisation, we next calculated the pathway from carbocation [sob]C to [sob]E including transition state structures.
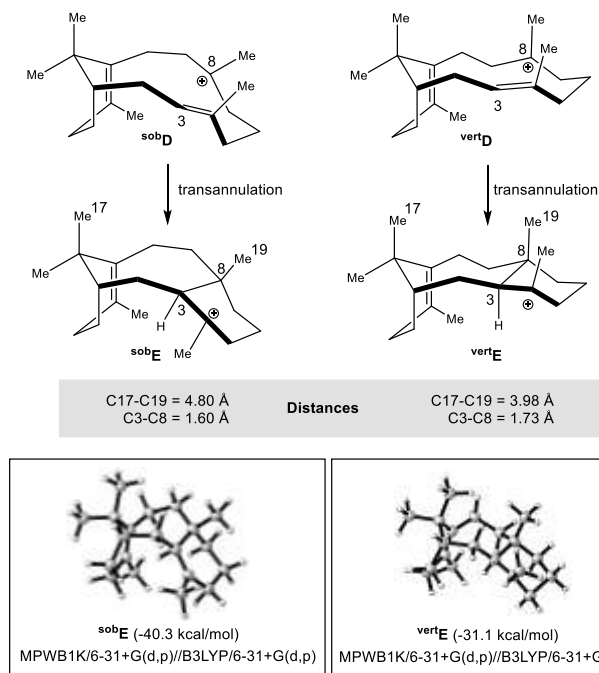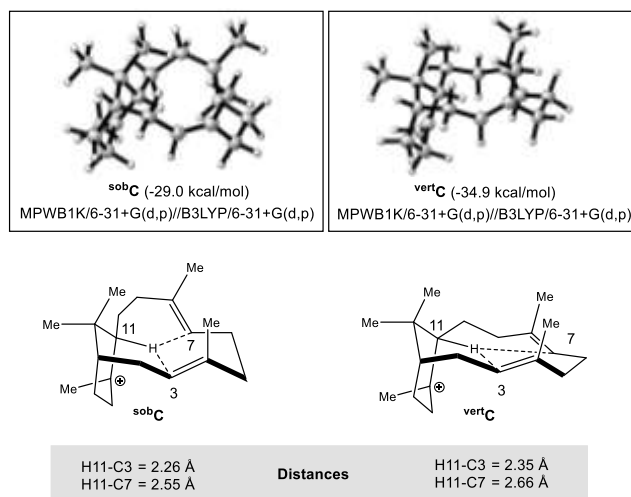

Figure 5: Comparison of the sobralene-like ([sob]C) and verticillene-like ([vert]C)[16] conformations of carbocation **C**. MPWB1K/6-31+G(d,p)//B3LYP/6-31+G(d,p) was used for the calculations and energies are quoted relative to GG+ (cation **A**) at 0.00 kcal/mol.

To expedite these calculations for the nine species (*i.e.* five minima and four transition states) along the carbocation cascade we explored the use of B3LYP/6-31G(d) for geometry optimisations and frequency claculations, followed by MPWB1K/6-31+G(d,p) for single point energies. Using carbocations **C**, **D** and **E**, for which we already had B3LYP/6-31+G(d,p) geometries for comparison, we quickly found that the MPWB1K/6-31+G(d,p)//B3LYP/6-31G(d) energies gave good agreement with the more costly MPWB1K/6-31+G(d,p)//B3LYP/6-31+G(d,p) alternatives (Table S1). The electronic energies were within 1.0 kcal/mol, giving confidence in the B3LYP/6-31G(d) geometries, and scaling of the zero-point vibration energies by a factor of 0.993 allowed direct comparison to the values obtained using the larger 6-31+G(d,p) basis set. This approach resulted in good agreement, so MPWB1K/6-31+G(d,p)//B3LYP/6-31G(d) was used to calculate the sobralene-like carbocation cascade (Figure 6). Using this method we were able to find the sobralene-like conformers of all nine species in the cascade, and for comparison we have plotted their MPWB1K/6-31+G(d,p) energies (in green) alongside the previously-reported verticillene-like equivalents (in blue)[16a] (Figure 6).

Our calculations show that formation of [sob]D from [sob]C directly, *via* proton transfer from C11 to C7, is unfavourable. Instead, intramolecular proton transfer from C11 to C3

within ^{sob}C, results in the formation of ^{sob}F. A second intramolecular proton transfer from C3 to C7 within ^{sob}F next
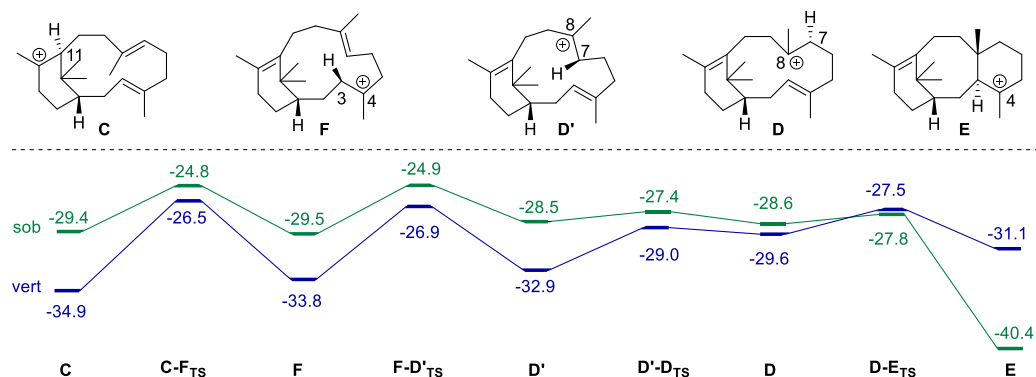

Figure 6: Relative energies, in kcal/mol, of the sobralene-like (green) [MPWB1K/6-31+G(d,p)//B3LYP/6-31G(d)] and verticillene-like (blue) [MPWB1K/6-31+G(d,p)//B3LYP/6-31+G(d,p)][16a] taxadiene carbocation cascades (using GG+, cation **A** at 0.00 kcal/mol as reference).

produces carbocation **D'**, which despite containing the desired carbocation at C8, cannot take part in the transannulation step as it sits in a non-productive conformation.

We could locate a transition state for the direct ^{sob}C→^{sob}D' interconversion (-18.8 kcal/mol, not plotted in Figure 6), but this was considerably higher in energy than the two step pathway involving carbocation ^{sob}F (Figure 6). These findings are consistent with those of Tantillo and Gutta who also reported a similar conformations for carbocations **F** and **D**.[16b] As expected for a simple conformational change, the barrier to formation of the productive conformer ^{sob}D from ^{sob}D' is low (1.1 kcal/mol), and the subsequent transannulation step (^{sob}D→^{sob}E) proceeds with a very low activation barrier (0.9 kcal/mol) to give the carbocation ^{sob}E that contains the fully formed taxane ring system. In general, and with the notable exception of ^{sob}E, the sobralene-like conformers are slightly higher in energy than their verticillene-like equivalents, but it is also apparent that the transition-state barriers are consistently lower than on the verticillene-like pathway.

As discussed above, it is significant that the sobralene-like pathway is energetically favourable going from ^{sob}C to ^{sob}E (-11.0 kcal/mol) whereas the verticillene-like pathway is uphill going from ^{vert}C to ^{vert}E (+3.8 kcal/mol). Since the cascade from ^{sob}C to ^{sob}E would be spontaneous, it is possible that taxadiene synthase (TXS) carefully controls the conformations of the carbocations involved to ensure production of the desired product (taxadiene **3**) resulting from ^{sob}E. It is interesting to speculate that recent site directed mutagenesis studies on taxadiene synthase by Brück *et al.* might provide evidence of this conformational control, as the V584M and V584L TXS mutants both favour formation of verticillene-like structures (*viz.* **2**) instead of taxadiene **3**.[11] Thus, it is possible that the longer amino acid side chains (Met and Leu) interact with the carbocation **C** (or one of the other downstream carbocations) to make the active site unable to accommodate the sobralene-like conformation that is favourable for taxadiene **3** production. In this situation, the next lowest point of the cascade ^{vert}C

would then be favoured leading to verticillene **2** production.

The isolation of taxadiene **3**, in addition to sobralene **1**, from sand flies[1] provides a further piece of tentative evidence for conformational control by the sobralene-forming synthase, as our study shows that both sobralene **1** and taxadiene **3** could be produced from the same conformation of carbocation **D** (i.e. ^{sob}D, Figure 2). As mentioned briefly above, in addition to imparting conformational control, it is likely that the yet-to-be identified sobralene synthase is capable of facilitating the deprotonation reaction at C9 of carbocation ^{sob}D. The presence of a suitably positioned basic residue in the active site would direct the flux of carbocation **D** towards sobralene **1**, rather than taxadiene **3** production. Due to the urgent need to treat and/or prevent visceral leishmaniasis (VL), genome sequencing is being conducted on the sand fly carrier (*Lutzomyia longipalpis*), and the identity of the sobralene synthase is likely to be revealed by these studies in due course. However, in the absence of genome sequence data, we feel that based upon our study and the clear biosynthetic link between sobralene **1** and taxadiene **3**, it should be possible to produce new mutants of taxadiene synthase that are capable of over-producing sobralene **1** in a host organism. This metabolic engineering approach could provide improved yields of the sand fly pheromone (sobralene **1**) for use in attractant traps or other applications, and we have initiated our own studies in this area based upon our successful over-production of taxadiene **3** in tomato plants.[7]

CONCLUSIONS

Sobralene **1** is a newly discovered isomer of verticillene **2** and we propose that it has a biosynthetic connection to taxadiene **3** *via* the verticillyl carbocation intermediate **D**. Quantum chemical calculations have shown that **D** can adopt a conformation, ^{sob}D, that allows a stereoelectronically favoured deprotonation leading to the *cis*-C8,C9 alkene bond in sobralene, and it is likely that the same ^{sob}D

is adopted in the active site of sobralene synthase. Furthermore, the conformation ^{sob}D permits an energetically favourable transannulation reaction which leads to the sobralene-like conformer (^{sob}E) of the penultimate carbonium ion precursor E to taxadiene 3. Importantly, the calculated conformations of ^{sob}D and ^{sob}E are in very good agreement with both NMR data collected on taxadiene 3, and with X-ray data measured on an advanced synthetic precursor to taxadiene 9, thus confirming their significance on the taxadiene cascade. Significantly, our combined results provide an overall exothermic pathway to taxadiene 3 that proceeds *via* lower activation barriers than those previously reported. It is reasonable to conclude that taxadiene synthase favours taxadiene production by guiding carbocation D to adopt a sobralene-like conformation thereby facilitating the transannulation (^{sob}D→^{sob}E) reaction; subsequent deprotonation from C5 in ^{sob}E then provides taxadiene 3. The previously determined gas-phase verticillene-like conformations of the carbocations have been used in QM/MM studies of the taxadiene cascade that include taxadiene synthase, and we feel that the sobralene-like conformations identified in this study should also be considered during any future computational studies of the pathway. Although the synthase responsible for the formation of sobralene in the sand fly *Lutzomyia longipalpis* has not yet been identified, the isolation of taxadiene 3 alongside sobralene 1 from *L. longipalpis* indicates a close relationship between the biosynthetic pathways to sobralene and taxadiene, and suggests that mutants of taxadiene synthase could be engineered[4-11] to allow for the overproduction of sobralene 1 in a suitable host organism.

ASSOCIATED CONTENT

Supporting Information. Details of all calculated structures (x,y,z coordinates, energies, zero point vibrational energies and imaginary frequencies for transition states). This information is available free of charge *via* the internet at http://pubs.acs.org/.

AUTHOR INFORMATION

Corresponding Author

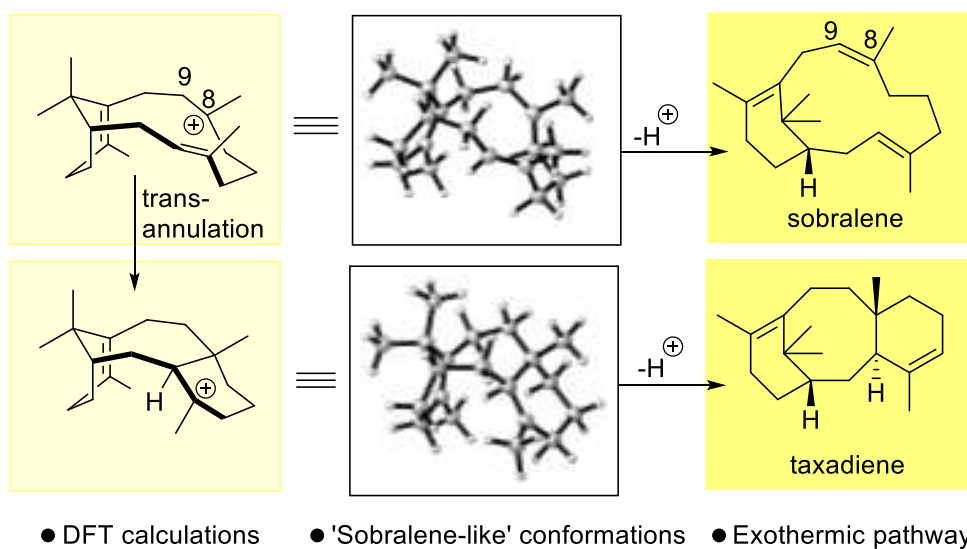*chris.hayes@nottingham.ac.uk

ACKNOWLEDGMENT

REFERENCES

1. Palframan, M. J.; Bandi, K. K.; Hamilton, J. G. C.; Pattenden, G. Sobralene, a new sex-aggregation pheromone and likely shunt metabolite of the taxadiene synthase cascade, produced by a member of the sand fly *Lutzomyia longipalpis* species complex. *Tetrahedron Lett.* **2018**, *59*, 1921-1923.

2. Bezerra, J. M. T.; de Araujo, V. E. M.; Barbosa, D. S.; Martins-Melo, F. R.; Werneck, G. L.; Carneiro, M. Burden of leishmaniasis in Brazil and federated units, 1990-2016: Findings from Global Burden of Disease Study 2016. *PLoS Neglected Trop. Dis.* **2018**, *12*, e0006697.

3. Alvar, J.; Velez, I. D.; Bern, C.; Herrero, M.; Desjeux, P.; Cano, J.; Jannin, J.; den Boer, M.; Team, W. L. C. Leishmaniasis worldwide and global estimates of its incidence. *PLoS One* **2012**, *7*, e35671.

4. Huang, K. X.; Huang, Q. L.; Wildung, M. R.; Croteau, R.; Scott, A. I. Overproduction, in *Escherichia coli*, of soluble taxadiene synthase, a key enzyme in the taxol biosynthetic pathway. *Protein Expression Purif.* **1998**, *13*, 90-96.

5. Huang, Q. L.; Roessner, C. A.; Croteau, R.; Scott, A. I. Engineering *Escherichia coli* for the synthesis of taxadiene, a key intermediate in the biosynthesis of taxol. *Biorg. Med. Chem.* **2001**, *9*, 2237-2242.

6. DeJong, J. M.; Liu, Y. L.; Bollon, A. P.; Long, R. M.; Jennewein, S.; Williams, D.; Croteau, R. B. Genetic engineering of Taxol biosynthetic genes in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* **2006**, *93*, 212-224.

7. Kovacs, K.; Zhang, L.; Linforth, R. S. T.; Whittaker, B.; Hayes, C. J.; Fray, R. G. Redirection of carotenoid metabolism for the efficient production of taxadiene [taxa-4(5),11(12)-diene] in transgenic tomato fruit. *Transgenic Res.* **2007**, *16*, 121-126.

8. Engels, B.; Dahm, P.; Jennewein, S. Metabolic engineering of taxadiene biosynthesis in yeast as a first step towards Taxol (Paclitaxel) production. *Metab. Eng.* **2008**, *10*, 201-206.

9. Rontein, D.; Onillon, S.; Herbette, G.; Lesot, A.; Werck-Reichhart, D.; Sallaud, C.; Tissier, A. CYP725A4 from yew catalyzes complex structural rearrangement of taxa-4(5),11(12)-diene into the cyclic ether 5(12)-oxa-3(11)-cyclotaxane. *J. Biol. Chem.* **2008**, *283*, 6067-6075.

10. Ajikumar, P. K.; Xiao, W. H.; Tyo, K. E. J.; Wang, Y.; Simeon, F.; Leonard, E.; Mucha, O.; Phon, T. H.; Pfeifer, B.; Stephanopoulos, G. Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* **2010**, *330*, 70-74.

11. Schrepfer, P.; Buettner, A.; Goerner, C.; Hertel, M.; van Rijn, J.; Wallrapp, F.; Eisenreich, W.; Sieber, V.; Kourist, R.; Brück, T. Identification of amino acid networks governing catalysis in the closed complex of class I terpene synthases. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, E958-E967.

12. Palframan, M. J.; Pattenden, G. The verticillenes. Pivotal intermediates in the biosynthesis of the taxanes and the phomactins. *Nat. Prod. Rep.* **2019**, *36*, 108-121.

13. Walker, K.; Croteau, R. Taxol biosynthetic genes. *Phytochemistry* **2001**, *58*, 1-7.

14. Koksal, M.; Jin, Y. H.; Coates, R. M.; Croteau, R.; Christianson, D. W. Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. *Nature* **2011**, *469*, 116-U138.

15. Tokiwano, T.; Endo, T.; Tsukagoshi, T.; Goto, H.; Fukushi, E.; Oikawa, H. Proposed mechanism for diterpene synthases in the formation of phomactatriene and taxadiene. *Org. Biomol. Chem.* **2005**, *3*, 2713-2722.

16. (a) Hong, Y. J.; Tantillo, D. J. The taxadiene-forming carbocation cascade. *J. Am. Chem. Soc.* **2011**, *133*, 18249-18256; (b) Gutta, P.; Tantillo, D. J. A promiscuous proton in taxadiene biosynthesis?. *Org. Lett.* **2007**, *9*, 1069-1071.

17. Freud, Y.; Ansbacher, T.; Major, D. T. Catalytic control in the facile proton transfer in taxadiene synthase. *ACS Catal.* **2017**, *7*, 7653-7657.

18. Escorcia, A. M.; van Rijn, J. P. M.; Cheng, G. J.; Schrepfer, P.; Bruck, T. B.; Thiel, W. Molecular dynamics study of taxadiene synthase catalysis. *J. Comput. Chem.* **2018**, *39*, 1215-1225.

19. Ansbacher, T.; Freud, Y.; Major, D. T. Slow-starter enzymes: role of active-site architecture in the catalytic control of the biosynthesis of taxadiene by taxadiene synthase. *Biochemistry* **2018**, *57*, 3773-3779.

20. van Rijn, J. P. M.; Escorcia, A. M.; Thiel, W. QM/MM study of the taxadiene synthase mechanism. *J. Comput. Chem.* **2019,** *40*, 1902-1910.

21. Palframan, M. J.; Bandi, K. K.; Hamilton, J. G. C.; Pattenden, G. Acid-catalysed rearrangement of the sandfly pheromone sobralene to verticillenes, consolidating its relationship inter alia to the taxanes and phomactins. *Synlett* **2019,** *30*, 1899-1903.

22. Hendrickson, J. B. Molecular Geometry VII. Modes of interconversion in medium rings. *J. Am. Chem. Soc.* **1967,** *89*, 7047-7061.

23. Jin, Q. W.; Williams, D. C.; Hezari, M.; Croteau, R.; Coates, R. M. Stereochemistry of the macrocyclization and elimination steps in taxadiene biosynthesis through deuterium labeling. *J. Org. Chem.* **2005,** *70*, 4667-4675.

24. Rubenstein, S. M.; Williams, R. M. Studies on the biosynthesis of taxol - Total Synthesis of taxa-4(20),11(12)-diene and taxa-4(5),11(12)-diene - the first committed biosynthetic intermediate. *J. Org. Chem.* **1995,** *60*, 7215-7223.

25. Becke, A. D. Density-functional thermochemistry .3. The role of exact exchange. *J. Chem. Phys.* **1993,** *98*, 5648-5652.

26. Lee, C. T.; Yang, W. T.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron-density. *Phys. Rev. B* **1988,** *37*, 785-789.

27. Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin-density calculations - A critical analysis. *Can. J. Phys.* **1980,** *58*, 1200-1211.

28. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab-Initio calculation of vibrational absorption and circular-dichroism spectra using density-functional force-fields. *J. Phys. Chem.* **1994,** *98*, 11623-11627.

29. Zhao, Y.; Truhlar, D. G. Hybrid meta density functional theory methods for thermochemistry, thermochemical kinetics, and noncovalent interactions: The MPW1B95 and MPWB1K models and comparative assessments for hydrogen bonding and van der Waals interactions. *J. Phys. Chem. A* **2004,** *108*, 6908-6918.

30. Behn, A.; Zimmerman, P. M.; Bell, A. T.; Head-Gordon, M. Efficient exploration of reaction paths via a freezing string method. *J. Chem. Phys.* **2011,** *135*, 224108.

31. Sharada, S. M.; Zimmerman, P. M.; Bell, A. T.; Head-Gordon, M. Automated transition state searches without evaluating the hessian. *J. Chem. Theory Comput.* **2012,** *8*, 5166-5174.

32. CYLview, 1.0b; Legault, C. Y., Université de Sherbrooke, 2009 (http://www.cylview.org)

Insert Table of Contents artwork here

trans-
annulation

-H$^{\oplus}$

9 8

H
sobralene

-H$^{\oplus}$

H

H
H
taxadiene

● DFT calculations   ● 'Sobralene-like' conformations   ● Exothermic pathway

The impact of macrocycle conformation on the taxadiene-forming carbocation cascade: Insight gained from sobralene, a recently discovered verticillene isomer.

Christopher J. Hayes*, Matthew J. Palframan and Gerald Pattenden

School of Chemistry, University of Nottingham, University Park, Nottingham, NG7 2RD U.K.

Supporting Information

The data for structures ^{vert}C, ^{vert}F, $^{vert}D'$, ^{vert}D, ^{vert}E, and the transition states connecting them, are reported in: Hong, Y. J.; Tantillo, D. J., The taxadiene-forming carbocation cascade. *J. Am. Chem. Soc.* **2011,** *133*, 18249-18256. This is reference 16a in the main text.

SUPPORTING INFORMATION

Geometry Optimisations using B3LYP/6-31+G(d,p)

^{sob}C is rendered as: **sobC**
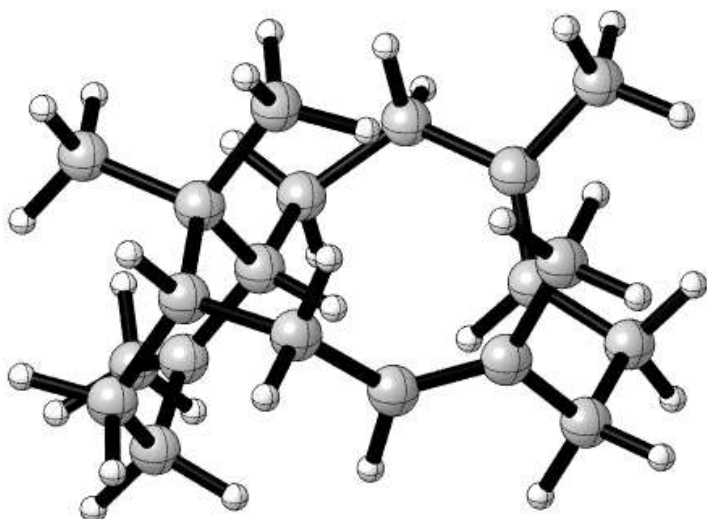
C	-1.548138	0.064562	1.160420
C	-2.080437	0.564631	-1.215198
C	-2.800519	-1.661277	-0.266025
C	-2.481104	-0.817861	-1.541866
C	-1.791451	-1.451100	0.873456
C	-1.024132	0.783214	-0.237497
H	-3.808395	-1.390745	0.060297
H	-1.616416	-1.280533	-2.039957
H	-2.840702	-2.714214	-0.559529
H	-3.317974	-0.841018	-2.243087
H	-2.293543	-1.819630	1.778904
C	-2.847714	0.754460	1.628225
H	-2.690450	1.824899	1.791035
H	-3.172314	0.330487	2.584263
H	-3.676986	0.645987	0.921618
C	-0.488359	0.237815	2.259301
H	-0.488477	1.254642	2.655969
H	0.518761	-0.004092	1.917188
H	-0.736122	-0.428213	3.093450
C	-0.525679	-2.368807	0.750700
H	-0.917383	-3.387708	0.619820
H	-0.016360	-2.371084	1.715948
C	-0.490323	2.240215	-0.207231
H	-1.232807	2.900688	0.255815
C	0.458893	-2.072835	-0.353072
C	1.792496	-1.872764	-0.260376
C	3.187179	-0.056670	-1.422931
H	3.600931	0.179506	-2.411475
C	3.219186	1.861568	1.064270
H	4.134780	1.356823	0.754137
H	3.461278	2.923900	1.195270
H	0.056613	-2.082160	-1.366393
C	2.585572	-1.497069	-1.496466

S2

H	3.410729	-2.204838	-1.648909
H	1.939961	-1.568922	-2.380126
C	2.589436	-1.978847	1.013952
H	3.060063	-1.022884	1.269303
H	3.405515	-2.699234	0.877982
H	2.001610	-2.307823	1.872293
H	2.938290	1.485808	2.056751
H	-0.417131	2.566583	-1.251191
C	-2.770888	1.671505	-1.910677
H	-2.803482	2.596713	-1.332897
H	-3.765337	1.384145	-2.258827
H	-2.176239	1.898996	-2.814750
H	4.027956	-0.045610	-0.725515
H	-0.198453	0.085511	-0.483559
C	0.870503	2.545103	0.454867
H	0.772347	2.500947	1.543505
H	1.085882	3.602900	0.244547
C	2.086011	1.713531	0.074439
C	2.137099	0.948715	-1.031198
H	1.288073	0.981652	-1.714435

E = -781.761688 au (B3LYP/6-31+G(d,p)//B3LYP/6-31+G(d,p))
This Molecule has 0 Imaginary Frequencies
Zero point vibrational energy: 305.301 kcal/mol

Single point E = -781.3119844 au (mPWB1K 6-31+G(d,p))

sobD

C	-2.098331	0.143488	0.937516
C	-1.575357	0.503458	-1.481456
C	-2.547514	-1.787924	-0.655612
C	-2.289836	-0.794141	-1.816495
C	-1.906653	-1.387023	0.688526
C	-1.405971	0.927774	-0.205713
H	-3.625861	-1.883428	-0.501789

SUPPORTING INFORMATION

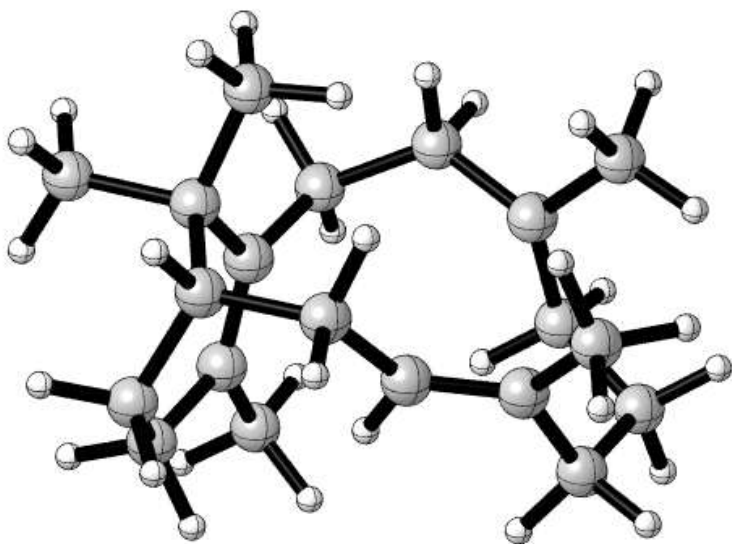|			
H	-1.705405	-1.287206	-2.607185
H	-2.205795	-2.788937	-0.943415
H	-3.241271	-0.544258	-2.306773
H	-2.470861	-1.904120	1.475989
C	-3.612780	0.496412	0.906529
H	-3.761659	1.555632	1.143258
H	-4.162543	-0.090533	1.651296
H	-4.059955	0.314077	-0.071527
C	-1.624888	0.486442	2.367519
H	-1.767731	1.546221	2.601422
H	-0.582247	0.227158	2.559409
H	-2.227160	-0.077994	3.086653
C	-0.449891	-1.905007	0.851665
H	-0.500558	-3.002841	0.763844
H	-0.098081	-1.728230	1.874059
C	-0.657857	2.209302	0.107446
H	-1.298410	2.877515	0.696912
C	0.570808	-1.440758	-0.151760
C	1.863715	1.338582	0.546552
C	1.928830	-1.540384	-0.077891
C	2.353335	1.319761	-0.851527
H	2.855158	2.305595	-0.944043
C	3.375023	0.241703	-1.244330
H	3.775851	0.489484	-2.232059
H	0.179369	-1.116223	-1.108326
C	2.739367	-1.171228	-1.300786
H	3.545965	-1.898581	-1.450403
H	2.099802	-1.222189	-2.187539
C	2.653013	-2.165748	1.086622
H	2.121990	-2.049663	2.035467
H	3.669204	-1.776468	1.202830
H	2.754358	-3.245233	0.910420
H	1.509336	1.355364	-1.545875
H	-0.439097	2.757508	-0.808741
C	-1.097217	1.204968	-2.733750
H	-1.934941	1.326142	-3.431531
H	-0.354612	0.594363	-3.266627
H	-0.668960	2.195912	-2.578359
H	4.230394	0.257173	-0.562590
C	0.652351	2.106561	0.943942
H	0.437942	1.896453	1.991189
H	1.068282	3.135286	0.958193
C	2.775401	1.022507	1.679698
H	3.026619	1.972000	2.178075
H	3.699938	0.528302	1.395883
H	2.257658	0.429245	2.441707

E = -781.759465 au (B3LYP/6-31+G(d,p)//B3LYP/6-31+G(d,p))
This Molecule has 0 Imaginary Frequencies
Zero point vibrational energy: 305.160 kcal/mol

Single point E = -781.3091126 au (mPWB1K 6-31+G(d,p))

S4

sobE
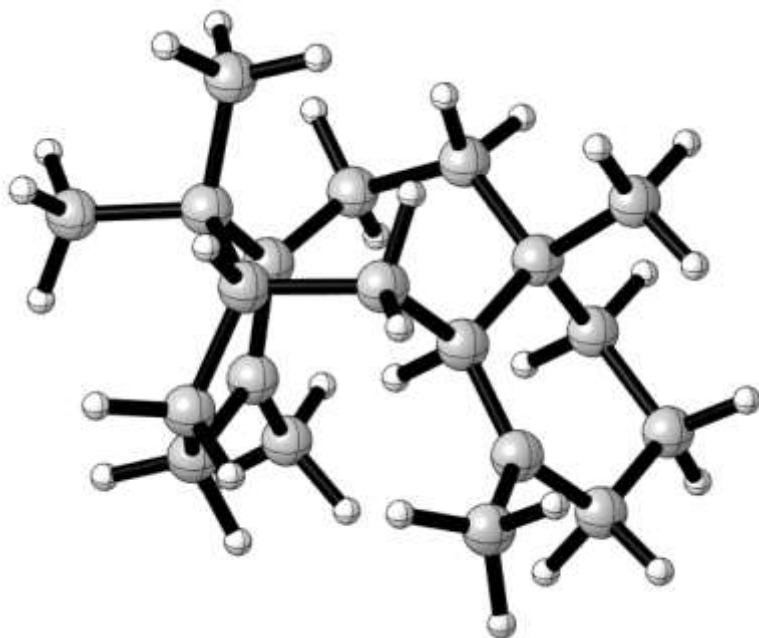
C	-2.330313	0.442150	0.418242
C	-1.172210	-0.597224	-1.526269
C	-1.776951	-2.013220	0.576095
C	-1.544053	-1.956978	-0.961863
C	-1.642611	-0.651506	1.294199
C	-1.442953	0.558183	-0.837238
H	-2.784112	-2.390592	0.767307
H	-0.739714	-2.650711	-1.245681
H	-1.108406	-2.752583	1.027051
H	-2.424609	-2.345078	-1.490584
H	-2.203479	-0.720910	2.235328
C	-3.760275	0.009649	-0.020374
H	-4.256180	0.844225	-0.525849
H	-4.361522	-0.247864	0.858317
H	-3.777087	-0.837926	-0.704712
C	-2.563854	1.732375	1.239784
H	-2.890333	2.565097	0.610639
H	-1.707122	2.066303	1.822898
H	-3.370880	1.541889	1.953631
C	-0.187483	-0.272262	1.731985
H	0.116044	-0.980999	2.506357
H	-0.222835	0.692965	2.241890
C	-0.774452	1.871449	-1.157145
H	-1.504905	2.687856	-1.177366
C	1.452896	1.256978	0.226310
C	1.906016	-1.239360	0.586443
C	2.377186	1.138764	-1.007401
H	2.757019	2.137182	-1.250351
C	3.550898	0.181771	-0.809490
H	4.119613	0.076659	-1.738321
C	2.234409	1.836980	1.428215
H	3.090870	1.214696	1.706493

H	2.614355	2.834439	1.183727
C	3.067706	-1.199598	-0.355945
H	3.881688	-1.803280	0.064168
H	2.727177	-1.789365	-1.226783
C	1.815126	-2.490822	1.386054
H	2.381407	-2.308745	2.316464
H	2.307753	-3.327703	0.883863
H	0.806580	-2.770742	1.682635
H	1.789821	0.814917	-1.875176
H	1.601267	1.937666	2.314130
H	-0.321374	1.846481	-2.149376
C	-0.537468	-0.671994	-2.897172
H	-1.222096	-1.202935	-3.570893
H	0.389589	-1.258991	-2.890097
H	-0.330617	0.298243	-3.348082
H	4.251219	0.579671	-0.068181
C	0.922267	-0.191472	0.632345
H	0.352055	-0.467080	-0.337270
C	0.309348	2.255720	-0.102235
H	-0.175596	2.552484	0.827177
H	0.795036	3.167236	-0.470601

E = -781.771176 au (B3LYP/6-31+G(d,p)//B3LYP/6-31+G(d,p))
This Molecule has 0 Imaginary Frequencies
Zero point vibrational energy: 306.85 kcal/mol

Single point E = -781.3326207au (mPWB1K 6-31+G(d,p))

Geometry Optimisations using B3LYP/6-31G(d)

sobC
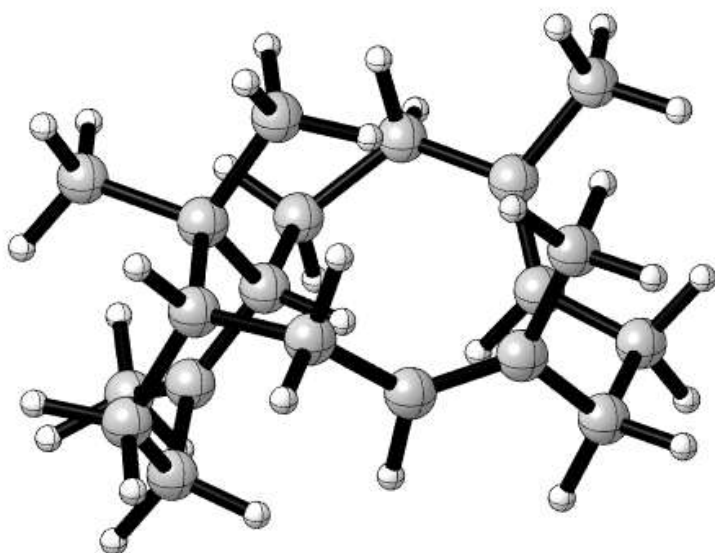
C	-1.537918	0.056814	1.153325
C	-2.092410	0.544448	-1.205949
C	-2.771364	-1.683847	-0.279004

C	-2.436582	-0.846106	-1.559939
C	-1.773398	-1.461034	0.868096
C	-1.019573	0.780773	-0.254365
H	-3.785042	-1.416857	0.033827
H	-1.549765	-1.294159	-2.027370
H	-2.802712	-2.738305	-0.570119
H	-3.260759	-0.892562	-2.276307
H	-2.283447	-1.825991	1.771326
C	-2.834577	0.737437	1.636689
H	-2.689671	1.814067	1.772005
H	-3.125808	0.328956	2.610630
H	-3.682236	0.598961	0.956520
C	-0.467230	0.241354	2.238492
H	-0.453936	1.266185	2.615273
H	0.535164	-0.021033	1.895540
H	-0.717360	-0.407331	3.086851
C	-0.504196	-2.374562	0.761023
H	-0.892533	-3.395845	0.632696
H	-0.001893	-2.372977	1.730883
C	-0.518703	2.248348	-0.224123
H	-1.261452	2.891615	0.263196
C	0.485885	-2.082891	-0.339155
C	1.814302	-1.862218	-0.242729
C	3.189558	-0.040476	-1.424396
H	3.604698	0.185414	-2.415426
C	3.189175	1.890047	1.048956
H	4.109854	1.384840	0.752385
H	3.428789	2.954154	1.177059
H	0.086666	-2.111438	-1.353791
C	2.606041	-1.488900	-1.480483
H	3.443319	-2.185594	-1.621419
H	1.966132	-1.581204	-2.366693
C	2.609177	-1.948403	1.034832
H	3.072262	-0.986695	1.284369
H	3.431528	-2.664425	0.907889
H	2.021681	-2.274476	1.895539
H	2.898329	1.519953	2.041513
H	-0.480196	2.583497	-1.268153
C	-2.882218	1.630444	-1.825561
H	-2.863471	2.570638	-1.273376
H	-3.909193	1.312819	-2.032250
H	-2.426917	1.830340	-2.813879
H	4.028717	-0.012073	-0.725041
H	-0.183982	0.098210	-0.496261
C	0.850994	2.572650	0.411666
H	0.759854	2.560296	1.502559
H	1.059817	3.626020	0.171642
C	2.066637	1.733663	0.048361
C	2.131469	0.962898	-1.051141
H	1.293154	0.992372	-1.749183

E = -781.703339 au (B3LYP/6-31G(d)//B3LYP/6-31G(d))

SUPPORTING INFORMATION

BIGGRID keyword used for geom opt and freq calculation.
This Molecule has 0 Imaginary Frequencies
Zero point vibrational energy: 306.850 kcal/mol

Single point E = -781.3117256 au (mPWB1K 6-31+G(d,p))

sobC → sobF TS
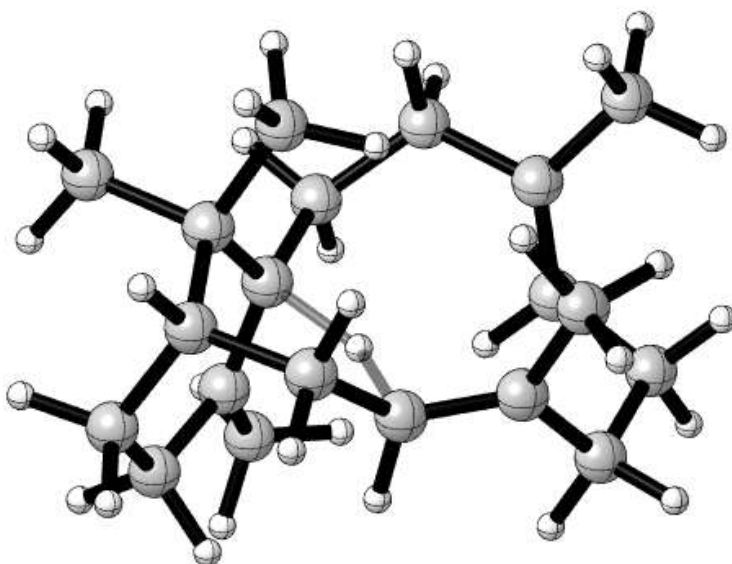
C	-1.689094	0.524927	1.042429
C	-1.630838	0.321728	-1.484451
C	-3.127798	-1.265331	-0.070460
C	-2.627724	-0.798694	-1.445675
C	-2.138105	-0.969382	1.068066
C	-1.004238	0.827741	-0.335923
H	-4.080780	-0.776971	0.149404
H	-2.173992	-1.632359	-2.008709
H	-3.345835	-2.338303	-0.107948
H	-3.469263	-0.496611	-2.086944
H	-2.678961	-1.121109	2.010577
C	-2.918526	1.456444	1.199080
H	-2.615304	2.507651	1.228018
H	-3.429066	1.239438	2.144180
H	-3.643314	1.343538	0.388379
C	-0.762547	0.795373	2.248278
H	-0.659692	1.865704	2.443081
H	0.241351	0.382282	2.121398
H	-1.200109	0.351618	3.149696
C	-0.942935	-1.961362	1.107000
H	-1.345412	-2.981641	1.050261
H	-0.454361	-1.886348	2.081012
C	-0.205845	2.143999	-0.507280
H	-0.892501	2.947570	-0.207802
C	0.083262	-1.791270	-0.007703
C	1.459125	-1.873627	0.140295
C	3.194958	-0.602494	-1.223782

S8

SUPPORTING INFORMATION

|			
H	3.639903	-0.613693	-2.226160
C	3.347394	1.545345	1.068855
H	4.217385	0.915225	0.871561
H	3.705296	2.581232	1.133917
H	-0.263257	-2.094309	-1.000052
C	2.344299	-1.910425	-1.077328
H	3.024994	-2.770559	-1.010465
H	1.733980	-2.047986	-1.976562
C	2.153922	-1.904994	1.465736
H	3.120788	-1.397428	1.427687
H	2.357390	-2.958002	1.716572
H	1.565270	-1.487201	2.284544
H	2.955681	1.293007	2.063506
H	-0.032230	2.310134	-1.570997
C	-1.303215	0.778068	-2.881125
H	-1.683018	1.795502	-3.047288
H	-1.762947	0.125836	-3.627835
H	-0.226152	0.812585	-3.075833
H	4.023701	-0.614297	-0.511984
H	-0.152210	-0.377339	-0.343508
C	1.129726	2.414518	0.220193
H	0.968306	2.511590	1.297431
H	1.451113	3.418269	-0.097286
C	2.273538	1.440247	0.011922
C	2.298983	0.585861	-1.022280
H	1.497199	0.652948	-1.756941

E = -781.690592 au (B3LYP/6-31G(d)//B3LYP/6-31G(d))
This Molecule has 1 Imaginary Frequency (i1120 cm^{-1})
Zero point vibrational energy: 304.205 kcal/mol

Single point E = -781.3002755 au (mPWB1K 6-31+G(d,p))

sobF

S9
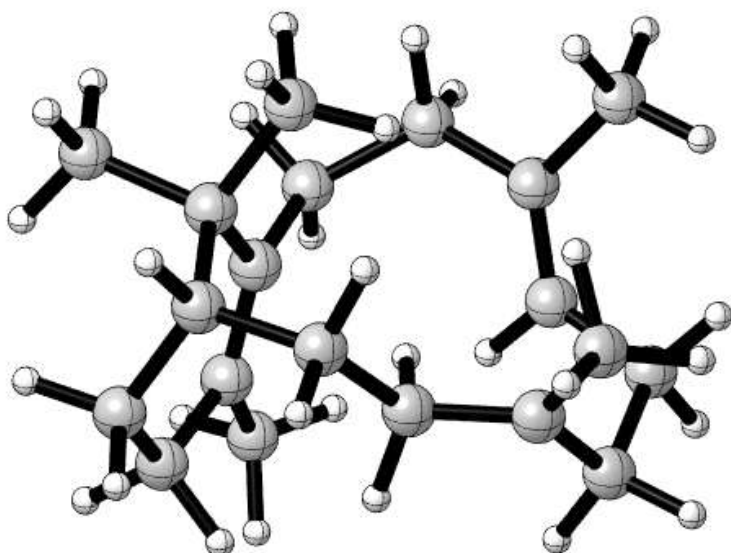
C	-1.756167	0.311237	1.038780
C	-1.635883	0.192190	-1.479043
C	-2.717991	-1.745723	-0.130099
C	-2.314709	-1.160127	-1.491887
C	-1.833359	-1.254322	1.027434
C	-1.312305	0.865210	-0.337934
H	-3.754838	-1.474447	0.087230
H	-1.650730	-1.856817	-2.032121
H	-2.705340	-2.841430	-0.177051
H	-3.197296	-1.093895	-2.144294
H	-2.339471	-1.536527	1.960220
C	-3.170430	0.884793	1.339932
H	-3.136171	1.976931	1.411670
H	-3.543756	0.505410	2.298921
H	-3.897054	0.632330	0.564011
C	-0.855772	0.757643	2.212371
H	-0.961187	1.826693	2.414128
H	0.205901	0.552196	2.045740
H	-1.158600	0.232573	3.125805
C	-0.464014	-1.990719	1.101899
H	-0.666296	-3.068316	1.157507
H	0.027736	-1.725804	2.042920
C	-0.762091	2.287813	-0.435897
H	-1.498784	2.960472	0.025817
C	1.914576	-1.893789	0.009215
C	3.158234	-0.014263	-1.114066
H	3.669590	0.133489	-2.072602
C	2.779936	1.964380	1.242835
H	2.299992	1.719998	2.199202
H	3.699594	1.382160	1.160250
C	2.701092	-1.559017	-1.186778
H	3.620092	-2.150287	-1.243219
H	2.118787	-1.692540	-2.102612
C	2.602348	-2.176691	1.287738
H	3.688921	-2.215401	1.195195
H	2.232835	-3.128788	1.696935
H	2.325574	-1.421097	2.038400
H	3.070445	3.020777	1.313243
H	-0.745923	2.581274	-1.485670
C	-1.393493	0.708096	-2.885256
H	-2.229765	1.339399	-3.215851
H	-1.335103	-0.125312	-3.593854
H	-0.482349	1.300644	-3.000557
H	3.909387	0.079937	-0.328426
C	0.482110	-1.724891	-0.073155
H	0.424126	-0.616705	-0.336434
H	0.122211	-2.150346	-1.020076
C	0.605462	2.694720	0.197015
H	0.464897	2.916391	1.258391
H	0.872723	3.663984	-0.249569
C	2.028465	0.953773	-0.951651

H	1.309507	0.963401	-1.769072
C	1.817381	1.784630	0.091185

E = -781.70368 au (B3LYP/6-31G(d)//B3LYP/6-31G(d))
BIGGRID keyword used for geom opt and freq calculation.
This Molecule has 0 Imaginary Frequencies
Zero point vibrational energy: 306.005 kcal/mol

Single point E = -781.3105568 au (mPWB1K 6-31+G(d,p))

sobF → sobD' TS
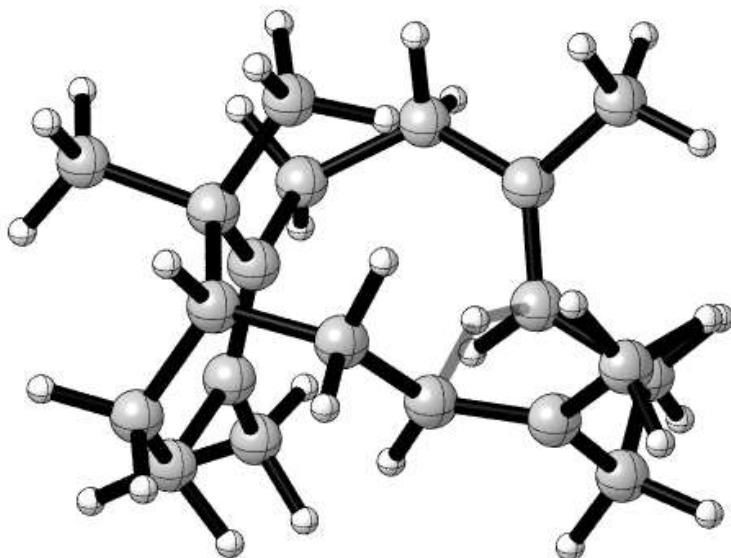
C	-1.830948	0.197855	1.016428
C	-1.649110	0.038741	-1.491247
C	-2.448726	-2.008449	-0.097552
C	-2.156093	-1.393792	-1.476964
C	-1.656739	-1.354860	1.047862
C	-1.446493	0.770124	-0.368180
H	-3.515126	-1.908872	0.123651
H	-1.428130	-2.013443	-2.025100
H	-2.252480	-3.086999	-0.121293
H	-3.060275	-1.442759	-2.099602
H	-2.123747	-1.690748	1.983638
C	-3.322826	0.550684	1.271473
H	-3.460531	1.636583	1.314028
H	-3.662101	0.138967	2.229794
H	-3.975113	0.171958	0.481788
C	-1.042803	0.807514	2.197602
H	-1.264045	1.870893	2.325592
H	0.042783	0.693105	2.106535
H	-1.341391	0.310546	3.127753
C	-0.196443	-1.879509	1.163594
H	-0.268031	-2.974544	1.241589
H	0.231055	-1.550197	2.116150
C	-1.050673	2.241006	-0.453730

H	-1.821057	2.835647	0.055257
C	2.146460	-1.757681	0.059323
C	3.063831	0.224754	-1.146481
H	3.436115	0.535274	-2.127991
C	2.542070	2.244186	1.171043
H	2.108118	2.024437	2.155601
H	3.499466	1.731083	1.085975
C	2.899493	-1.345572	-1.168194
H	3.893994	-1.798001	-1.211837
H	2.354884	-1.631907	-2.073682
C	2.934762	-2.073833	1.289586
H	3.774921	-1.380391	1.419198
H	3.380730	-3.071644	1.166140
H	2.334360	-2.091495	2.200519
H	2.732131	3.325755	1.165525
H	-1.097231	2.566194	-1.492483
C	-1.406289	0.502096	-2.913284
H	-2.352972	0.530785	-3.468661
H	-0.770049	-0.219542	-3.445358
H	-0.942841	1.484166	-3.016091
H	3.832097	0.488255	-0.416738
C	0.772149	-1.585986	0.031306
H	1.132681	-0.129367	-0.107297
H	0.334703	-1.585392	-0.964679
C	0.297685	2.744935	0.150986
H	0.157496	3.048297	1.192353
H	0.551276	3.683466	-0.370542
C	1.749801	0.924822	-0.855934
H	1.029934	0.912227	-1.673662
C	1.552288	1.917574	0.093017

E = -781.691212 au (B3LYP/6-31G(d)//B3LYP/6-31G(d))
This Molecule has 1 Imaginary Frequency (i1011 cm^{-1})
Zero point vibrational energy: 304.180 kcal/mol

Single point E = -781.3003682 au (mPWB1K 6-31+G(d,p))

sobD'
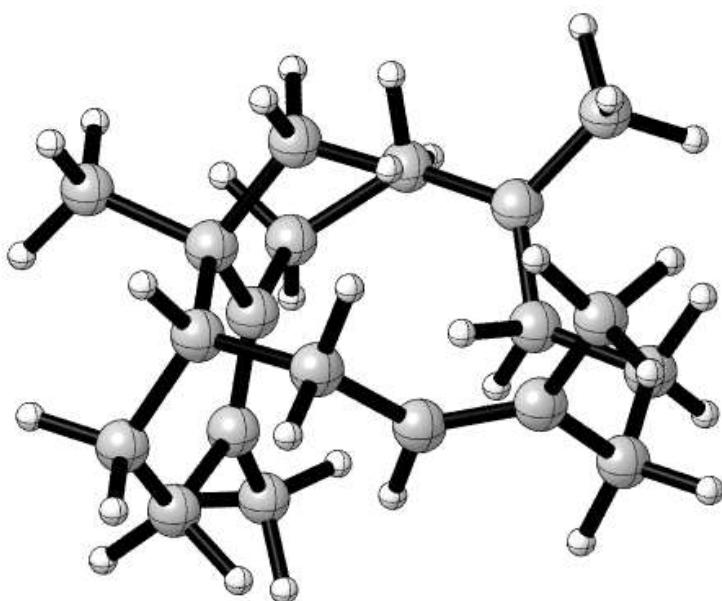
C	-1.620493	0.227845	1.112750
C	-1.678398	0.411559	-1.417675
C	-2.878114	-1.534852	-0.230750
C	-2.435982	-0.891871	-1.549006
C	-1.880968	-1.307087	0.917981
C	-1.280851	0.929688	-0.226290
H	-3.849441	-1.123348	0.061115
H	-1.822670	-1.589588	-2.139424
H	-3.042527	-2.608620	-0.377526
H	-3.314007	-0.708718	-2.185559
H	-2.384106	-1.635538	1.837519
C	-2.912480	0.885682	1.678967
H	-2.745945	1.941821	1.918577
H	-3.218711	0.388517	2.606467
H	-3.745661	0.839390	0.973272
C	-0.535894	0.418659	2.197933
H	-0.455830	1.463378	2.515684
H	0.453383	0.067583	1.885452
H	-0.810091	-0.153886	3.090899
C	-0.628548	-2.242584	0.837954
H	-1.028022	-3.255384	0.674843
H	-0.153255	-2.275233	1.822740
C	-0.748786	2.365955	-0.196810
H	-1.425487	2.968238	0.422428
C	1.752069	-2.020959	-0.123603
C	3.064418	-0.088382	-1.082427
H	3.634295	0.256547	-1.951824
C	2.977852	2.029221	1.106492
H	3.014284	1.147973	1.767950
H	3.940446	2.049655	0.583863
C	2.596631	-1.557739	-1.295990
H	3.484797	-2.187270	-1.424126

|			
H	2.021214	-1.621318	-2.226919
C	2.509128	-2.445434	1.109386
H	3.357793	-1.782284	1.322998
H	2.938634	-3.444456	0.952516
H	1.884706	-2.499244	2.003953
H	2.867162	2.915614	1.734141
H	-0.836331	2.795755	-1.194058
C	-1.456992	1.063333	-2.769349
H	-2.317061	1.686372	-3.050869
H	-1.365983	0.295404	-3.545552
H	-0.566414	1.694895	-2.835329
H	3.745901	-0.059868	-0.227127
C	0.403975	-1.965478	-0.223157
H	1.024374	0.170784	-0.404782
H	0.008756	-1.737812	-1.209665
C	0.671114	2.731140	0.352115
H	0.639214	2.988887	1.417575
H	0.965937	3.687549	-0.124652
C	1.855920	0.825013	-0.842243
H	1.348047	1.150753	-1.760187
C	1.851834	1.861868	0.154656

E = -781.699147 au (B3LYP/6-31G(d)//B3LYP/6-31G(d))
This Molecule has 0 Imaginary Frequencies
Zero point vibrational energy: 305.561 kcal/mol

Single point E = -781.3082248 au (mPWB1K 6-31+G(d,p))

sobC → sobD'TS
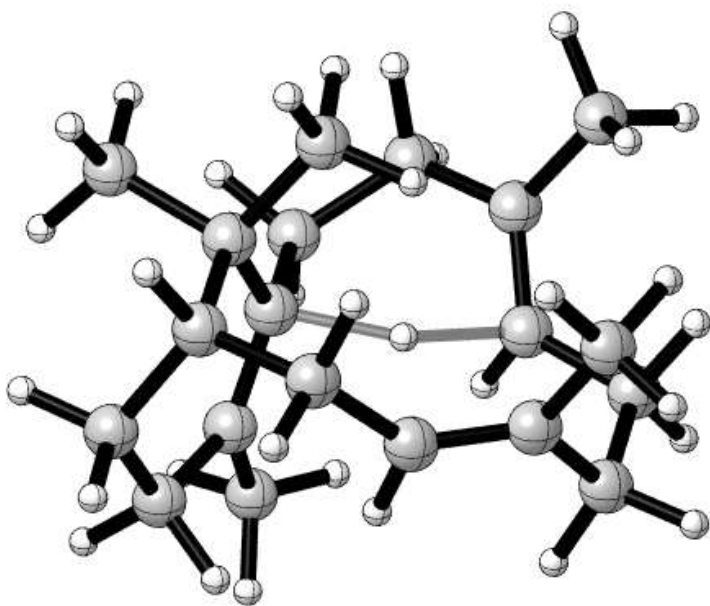
C	-1.413427	0.194109	1.151782
C	-1.630037	0.379592	-1.387885
C	-2.870319	-1.482188	-0.107362

|			
C	-2.459706	-0.869397	-1.448934
C	-1.798720	-1.319463	0.982478
C	-1.047786	0.873757	-0.215710
H	-3.800477	-1.012248	0.227580
H	-1.921249	-1.595777	-2.076643
H	-3.106128	-2.542266	-0.249206
H	-3.348461	-0.624993	-2.051385
H	-2.278226	-1.596685	1.930082
C	-2.643735	0.944082	1.737975
H	-2.388103	1.962013	2.048251
H	-3.003923	0.422459	2.630896
H	-3.473805	1.012357	1.028145
C	-0.288298	0.300900	2.202554
H	-0.084825	1.337145	2.486564
H	0.642706	-0.164042	1.873337
H	-0.606169	-0.213546	3.115939
C	-0.626459	-2.344108	0.835086
H	-1.100116	-3.324093	0.674487
H	-0.105482	-2.422632	1.793401
C	-0.736851	2.391474	-0.216860
H	-1.510365	2.899661	0.364830
C	0.353223	-2.061646	-0.267859
C	1.695753	-2.023569	-0.229714
C	2.853427	0.007236	-1.202067
H	3.371265	0.364161	-2.101837
C	2.782341	1.790694	1.278743
H	3.718018	2.240026	0.914943
H	2.472936	2.340828	2.169665
H	-0.088252	-1.836430	-1.235124
C	2.422500	-1.465842	-1.435109
H	3.328739	-2.041033	-1.662704
H	1.780472	-1.522874	-2.322315
C	2.547674	-2.429304	0.946368
H	1.962915	-2.692527	1.831748
H	3.265143	-1.648721	1.233606
H	3.148014	-3.310202	0.680526
H	3.023819	0.763645	1.567353
H	-0.841950	2.771560	-1.231492
C	-1.492198	1.045154	-2.732243
H	-2.275574	1.807709	-2.851604
H	-1.633987	0.319027	-3.538302
H	-0.534384	1.548679	-2.889417
H	3.597076	0.028044	-0.400958
H	0.379362	0.514198	-0.519315
C	0.644895	2.857595	0.342366
H	0.564507	3.154924	1.390855
H	0.921006	3.772933	-0.201991
C	1.752647	1.861987	0.190662
C	1.738289	0.999749	-0.898241
H	1.226216	1.383822	-1.786085

E = -781.680892 au (B3LYP/6-31G(d)//B3LYP/6-31G(d))

SUPPORTING INFORMATION

This Molecule has 1 Imaginary Frequency (i1260 cm^{-1})
Zero point vibrational energy: 304.304 kcal/mol

Single point E = -781.2909016 au (mPWB1K 6-31+G(d,p))

sobD' → sobD TS
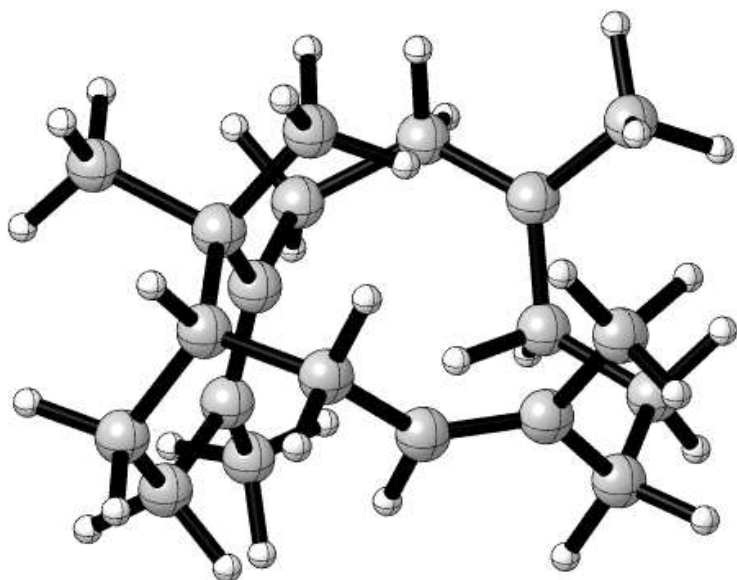
C	-1.676051	0.191165	1.094856
C	-1.613843	0.508153	-1.423979
C	-2.863315	-1.507061	-0.393225
C	-2.381046	-0.775814	-1.652785
C	-1.937184	-1.329271	0.823447
C	-1.257510	0.951064	-0.187143
H	-3.859944	-1.139054	-0.129558
H	-1.756698	-1.434184	-2.275172
H	-2.990875	-2.573404	-0.611858
H	-3.241418	-0.541713	-2.296549
H	-2.503036	-1.684454	1.695542
C	-2.996125	0.837009	1.607127
H	-2.851596	1.892178	1.864677
H	-3.340692	0.327471	2.514508
H	-3.796249	0.788613	0.864223
C	-0.653532	0.323608	2.246310
H	-0.603178	1.346813	2.632416
H	0.352928	0.002224	1.959428
H	-0.966441	-0.307496	3.085215
C	-0.686865	-2.263636	0.792264
H	-1.077739	-3.276732	0.612141
H	-0.242792	-2.301202	1.791857
C	-0.682294	2.354369	-0.039625
H	-1.339034	2.932846	0.621490
C	0.367783	-1.969682	-0.240001
C	1.897366	1.724499	0.272577
C	1.707149	-1.939006	-0.112384

C	1.941571	0.954147	-0.967161
H	1.831187	1.686898	-1.788973
C	3.067160	-0.067354	-1.224037
H	3.574766	0.178951	-2.162250
H	-0.013738	-1.785225	-1.241283
C	2.539987	-1.528482	-1.311991
H	3.411992	-2.184182	-1.430452
H	1.944940	-1.629732	-2.226950
C	2.471413	-2.286099	1.143021
H	2.971037	-3.256870	1.022387
H	1.836220	-2.357998	2.029381
H	3.273454	-1.564997	1.355128
H	0.935403	0.469815	-0.995962
H	-0.725242	2.868255	-0.999338
C	-1.354750	1.249012	-2.721628
H	-2.199834	1.907100	-2.966502
H	-1.266471	0.538286	-3.550367
H	-0.452068	1.866057	-2.726159
H	3.833678	-0.009854	-0.446449
C	0.745261	2.593849	0.572977
H	0.707054	2.743504	1.656499
H	1.079431	3.581123	0.191752
C	2.979225	1.663603	1.276737
H	3.967230	1.717066	0.805006
H	2.940313	0.671793	1.760009
H	2.881009	2.423689	2.054014

E = -781.698355au (B3LYP/6-31G(d)//B3LYP/6-31G(d))
BIGGRID keyword used for geom opt and freq calculation.
This Molecule has 1 Imaginary Frequency (i118 cm^{-1})
Zero point vibrational energy: 306.017 kcal/mol

Single point E = -781.3072581 au (mPWB1K 6-31+G(d,p))

sobD
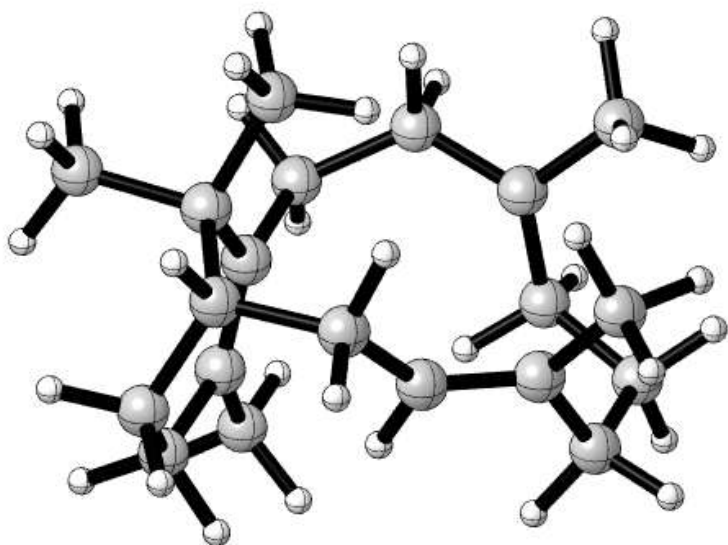
C	-1.968416	0.298796	0.970986
C	-1.549219	0.425713	-1.503414
C	-2.799071	-1.625481	-0.471240
C	-2.349530	-0.845693	-1.724672
C	-2.037123	-1.250347	0.812728
C	-1.297693	0.932655	-0.272854
H	-3.864193	-1.447099	-0.299373
H	-1.749062	-1.495733	-2.379534
H	-2.709010	-2.702515	-0.654517
H	-3.227714	-0.594539	-2.336246
H	-2.636565	-1.621551	1.655116
C	-3.405790	0.883402	1.070129
H	-3.369509	1.964849	1.245991
H	-3.951461	0.432513	1.907875
H	-3.983865	0.723557	0.157279
C	-1.291583	0.624324	2.321213
H	-1.328750	1.694849	2.550913
H	-0.251415	0.290576	2.379287
H	-1.835076	0.115610	3.125288
C	-0.672870	-1.987501	0.953244
H	-0.897056	-3.062510	0.856722
H	-0.284415	-1.871496	1.970482
C	-0.544728	2.240832	-0.100415
H	-1.179426	2.948460	0.449870
C	0.384622	-1.669963	-0.066089
C	1.990181	1.449751	0.352862
C	1.735298	-1.695399	0.062820
C	2.360626	1.085021	-1.028265
H	2.908786	1.995534	-1.355384
C	3.309919	-0.112938	-1.237790
H	3.774944	-0.010857	-2.223551
H	0.015777	-1.513023	-1.073532
C	2.574950	-1.474798	-1.175154
H	3.341546	-2.257423	-1.256714
H	1.943288	-1.563588	-2.065805
C	2.454120	-2.101800	1.325736
H	2.637033	-3.185180	1.319342
H	1.883007	-1.885439	2.234271
H	3.442473	-1.634460	1.412671
H	1.477738	1.025111	-1.669789
H	-0.369573	2.711126	-1.068177
C	-1.086568	1.001709	-2.825348
H	-1.951020	1.175562	-3.478891
H	-0.450070	0.281089	-3.358639
H	-0.540681	1.944443	-2.762714
H	4.129922	-0.085008	-0.512917
C	0.791655	2.247191	0.693249
H	0.626642	2.169063	1.770370
H	1.208857	3.273503	0.584416
C	2.984243	1.314585	1.447126
H	4.015973	1.310144	1.086209

```
H      2.824029    0.347850    1.949482
H      2.850701    2.086242    2.211085
```
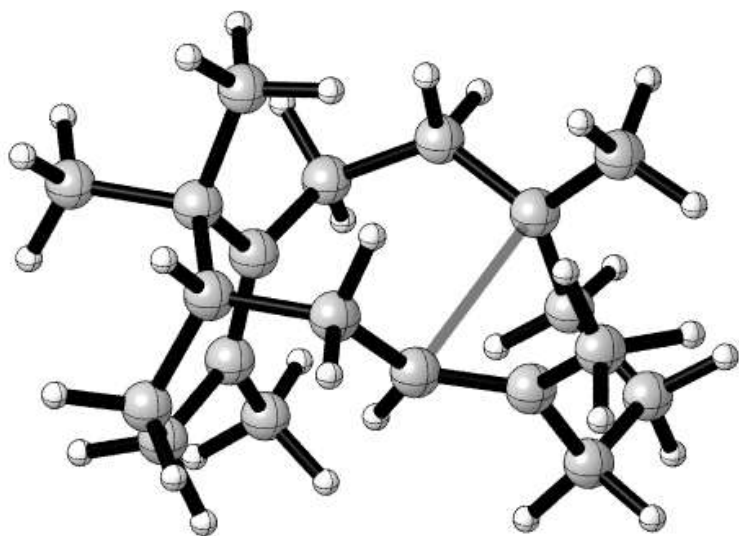
E = -781.701628 au (B3LYP/6-31G(d)//B3LYP/6-31G(d))
BIGGRID keyword used for geom opt and freq calculation.
This Molecule has 0 Imaginary Frequencies
Zero point vibrational energy: 306.623 kcal/mol

Single point E = -781.310138 au (mPWB1K 6-31+G(d,p))

sobD → sobE TS


```
C     -2.146056    0.085785    0.920689
C     -1.588449    0.568476   -1.462292
C     -2.473273   -1.796329   -0.758632
C     -2.268608   -0.728123   -1.865478
C     -1.881082   -1.421276    0.616672
C     -1.438285    0.931138   -0.167086
H     -3.544292   -1.974521   -0.629400
H     -1.680922   -1.156051   -2.690813
H     -2.055583   -2.754486   -1.088974
H     -3.236238   -0.484070   -2.326720
H     -2.430379   -1.999756    1.371620
C     -3.669522    0.387286    0.842088
H     -3.863750    1.425718    1.134086
H     -4.229872   -0.260040    1.527544
H     -4.075000    0.252878   -0.162207
C     -1.755401    0.388778    2.384676
H     -1.896873    1.445034    2.637155
H     -0.731188    0.107089    2.644321
H     -2.412536   -0.181980    3.049746
C     -0.402496   -1.870796    0.782884
H     -0.393329   -2.966680    0.658453
H     -0.069801   -1.714935    1.815757
C     -0.670425    2.175958    0.223492
```

H	-1.300078	2.829457	0.841514
C	1.839858	1.238327	0.632100
C	1.967698	-1.469192	-0.154686
C	2.393984	1.403176	-0.737903
H	2.908185	2.384124	-0.687284
C	3.417106	0.361990	-1.208778
H	3.823585	0.679078	-2.174359
C	2.737307	0.865288	1.766292
H	3.664772	0.380404	1.472023
H	2.995726	1.795096	2.298989
C	2.766684	-1.035068	-1.363275
H	3.559949	-1.763548	-1.570299
H	2.115648	-1.015276	-2.244000
C	2.688815	-2.220412	0.932800
H	2.176265	-2.181708	1.898199
H	3.720099	-1.877413	1.065330
H	2.751690	-3.280685	0.648767
H	1.585296	1.532250	-1.463828
H	2.214322	0.243161	2.499892
H	-0.420133	2.766638	-0.658429
C	-1.100090	1.334172	-2.672930
H	-1.909613	1.410139	-3.409996
H	-0.285996	0.796086	-3.181906
H	-0.759084	2.351494	-2.471867
H	4.268511	0.319187	-0.521835
C	0.608780	-1.325025	-0.190473
H	0.210143	-0.920207	-1.111422
C	0.620078	1.982033	1.071472
H	0.375513	1.676399	2.088278
H	1.045224	3.000010	1.191881

E = -781.701404 au (B3LYP/6-31G(d)//B3LYP/6-31G(d))
BIGGRID keyword used for geom opt and freq calculation.
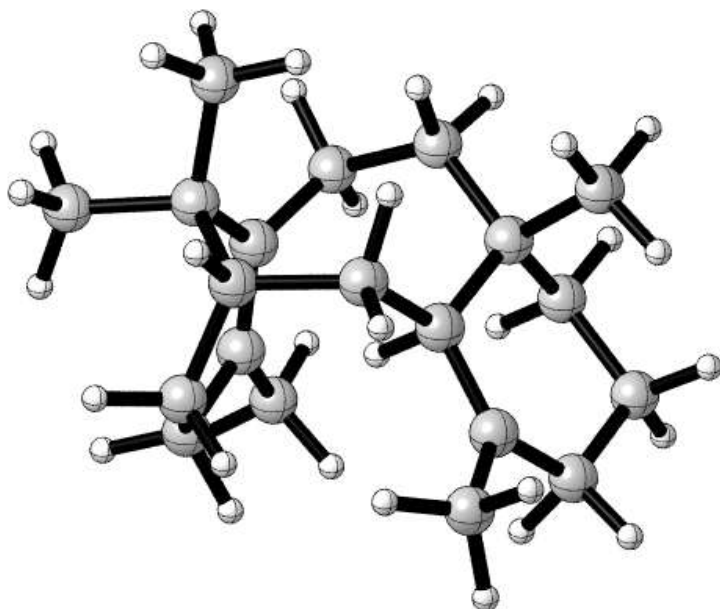This Molecule has 1 Imaginary Frequency (i45 cm^{-1})
Zero point vibrational energy: 306.840 kcal/mol

Single point E = -781.3091423 au (mPWB1K 6-31+G(d,p))

sobE

C	-2.329907	0.440456	0.422671
C	-1.180224	-0.591840	-1.529399
C	-1.777243	-2.015525	0.569605
C	-1.550835	-1.953914	-0.969067
C	-1.640321	-0.656694	1.292416
C	-1.447936	0.561099	-0.835354
H	-2.784564	-2.392788	0.762856
H	-0.747489	-2.647262	-1.257911
H	-1.107212	-2.757874	1.015536
H	-2.434491	-2.340173	-1.494795
H	-2.199058	-0.729538	2.234942
C	-3.760969	0.009385	-0.012564
H	-4.261013	0.847002	-0.510225
H	-4.357903	-0.255575	0.867587
H	-3.779865	-0.834305	-0.703043
C	-2.562153	1.726651	1.250410
H	-2.880876	2.565657	0.624421
H	-1.706721	2.052860	1.841586
H	-3.374467	1.535699	1.959171
C	-0.184212	-0.281500	1.729299
H	0.119637	-0.995726	2.499454
H	-0.216364	0.681302	2.245162
C	-0.778141	1.874559	-1.151080
H	-1.507642	2.692698	-1.166286
C	1.452640	1.255843	0.227134
C	1.913241	-1.239627	0.583414
C	2.376079	1.143329	-1.007429
H	2.755669	2.143239	-1.246937
C	3.550889	0.187027	-0.813309
H	4.120391	0.085994	-1.742636
C	2.234098	1.831462	1.430273
H	3.093629	1.209565	1.703034
H	2.610192	2.832344	1.191141

C	3.070160	-1.197454	-0.365432
H	3.887351	-1.803426	0.046410
H	2.724100	-1.781248	-1.238988
C	1.830146	-2.488872	1.388484
H	2.370663	-2.293549	2.332116
H	2.347192	-3.320191	0.900517
H	0.819180	-2.786289	1.663512
H	1.788922	0.822468	-1.877580
H	1.602124	1.924554	2.318815
H	-0.326105	1.854039	-2.144562
C	-0.545744	-0.663897	-2.900626
H	-1.231623	-1.188904	-3.578919
H	0.379138	-1.255665	-2.894300
H	-0.332080	0.307242	-3.348312
H	4.250196	0.582600	-0.069206
C	0.923569	-0.196245	0.627317
H	0.353716	-0.465686	-0.342410
C	0.308136	2.254655	-0.096788
H	-0.175261	2.549976	0.834583
H	0.792998	3.167559	-0.464576

E = -781.714033 au (B3LYP/6-31G(d)//B3LYP/6-31G(d))
This Molecule has 0 Imaginary Frequencies
Zero point vibrational energy: 308.791 kcal/mol

Single point E = -781.3323331 au (mPWB1K 6-31+G(d,p))

Comparison of MPWB1K/6-31+G(d,p) energies obtained for B3LYP/6-31+G(d,p) and B3LYP/6-31G(d) geometries.

Entry	Cation	MPWB1K/6-31+G(d,p)// B3LYP/6-31+G(d,p)	MPWB1K/6-31+G(d,p)// B3LYP/6-31G(d)	ΔE (kcal/mol)	Δzpe (kcal/mol)
1	^{vert}C	-781.32043	-781.32074	0.19	-0.34
2	^{sob}C	-781.31198	-781.31173	-0.16	0.60
3	^{vert}D	-781.31233	-781.31184	-0.31	-0.43
4	^{sob}D	-781.30911	-781.310138	0.64	0.68
5	^{vert}E	-781.31775	-781.31768	-0.04	0.20
6	^{sob}E	-781.33262	-781.33233	-0.18	0.22

Table S1: Comparison of MPWB1K/6-31+G(d,p)//B3LYP/6-31+G(d,p) and MPWB1K/6-31+G(d,p)//B3LYP/6-31G(d) energies for carbocations **C**, **D** and **E** in the verticillene-like and sobralene-like conformations. B3LYP/6-31G(d) zero-point energies were scaled by a factor of 0.993 prior to subtraction from the B3LYP/6-31+G(d,p) values.

IRC-like Calculations

IRC-like calculations were performed to confirm that starting materials and products were connected *via* the transition states identified. As described by Thiel *et al.* (van Rijn, J. P. M.; Escorcia, A. M.; Thiel, W., QM/MM study of the taxadiene synthase mechanism. *J. Comput. Chem.* **2019,** *40*, 1902-1910):

"An IRC-like calculation makes use of an approximate IRC procedure. In an IRC-like calculation, a fraction of the normal mode eigenvector corresponding to the imaginary frequency of the transition state is added to (or subtracted from) the structure of the transition state. The new structure is subjected to an unconstrained geometry optimization, and the resulting structure is visually inspected to confirm that it is the reactant or product."