Automated structure elucidation of two unexpected products in a reaction of anα,β-unsaturated pyruvate

ARTICLE *in* MAGNETIC RESONANCE IN CHEMISTRY · JULY 2004

Impact Factor: 1.18 · DOI: 10.1002/mrc.1396 · Source: PubMed

MAGNETIC RESONANCE IN CHEMISTRY
Magn. Reson. Chem. (in press)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/mrc.1396

MRC

Automated structure elucidation of two unexpected products in a reaction of an α,β-unsaturated pyruvate

Gary J. Sharman,[1]* Ian C. Jones,[1] Mark P. Parnell,[2] Michael C. Willis,[2] Mary F. Mahon,[2] Dean V. Carlson,[3] Antony Williams,[3] Mikhail Elyashberg,[4] Kirill Blinov[4] and Sergey G. Molodtsov[5]

[1] AstraZeneca, Silk Road Business Park, Macclesfield, Cheshire SK10 2NA, UK
[2] Department of Chemistry, University of Bath, Bath BA2 7AY, UK
[3] Advanced Chemistry Development, 90 Adelaide Street West, Suite 600, Toronto, Canada, M5H 3V9
[4] Advanced Chemistry Development, Bakuleva 6, str. 1, 117513 Moscow, Russian Federation
[5] Novosibirsk Institute of Organic Chemistry, Siberian Branch of the Russian Academy of Science, Lavrentiev Avenue 9, Novosibirsk 630090, Russian Federation

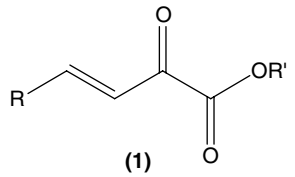
Received 4 November 2003; Revised 24 January 2004; Accepted 7 February 2004

The reaction between an α,β-unsaturated pyruvate and ethyl diazoacetate (EDA) yielded two unexpected products. The structures of these products were determined by automated elucidation of the chemical structures using spectroscopic inputs of a series of 1D and 2D NMR data using the computer program ACD/Structure Elucidator, StrucEluc. The formation of these products is rationalised. Their structures were also confirmed by x-ray crystallography. Copyright © 2004 John Wiley & Sons, Ltd.

KEYWORDS: NMR; 2D NMR; computer-aided structure elucidation; unsaturated pyruvate

INTRODUCTION

In recent work we have been investigating the reactivity of α,β-unsaturated pyruvates (**1**), focusing on the reaction with various carbene sources. Generally, cyclopropanation or epoxidation products are the expected result of the reaction of a carbene or carbene-type reagent with an alkene or a carbonyl, respectively.[1] However, α,β-unsaturated pyruvates have been shown to display remarkable reactivity in a range of cycloaddition reactions.[2]



(1)

The reaction between one of these pyruvates (R = naphthyl, R′ = ethyl) and ethyl diazoacetate (EDA), using $ZrCl_4$ as the Lewis acid, yielded two products. Initial evaluation of the ^{1}H NMR (Fig. 1) and mass spectrometric (MS) data indicated that neither of these products were among those expected from this reaction. To investigate the structure further, a number of two-dimensional NMR experiments were performed. A double-quantum-filtered correlation spectroscopy (DQF-COSY) experiment was used to correlate proton resonances based on proton–proton couplings. Heteronuclear multiple-quantum coherence (HMQC) and heteronuclear multiple bond (HMBC) experiments gave direct and indirect (long-range) proton–carbon correlations, respectively.

An initial survey of the two-dimensional data did not reveal a trivial solution to the structural problem. In such situations, the possibility of applying computer-aided structure elucidation (CASE) techniques becomes attractive. Such techniques have two potential advantages over the more traditional expert interpretation approach: first, the potential to save time by utilizing computer technology to generate a solution more quickly, thereby freeing up the expert to focus on other tasks, and second, and perhaps more important, the exhaustive nature of the computer search and application of the elucidation algorithms remove human bias. The process of manual structure elucidation is biased by prior knowledge and expectations regarding potential target structures. An automated process is biased solely by the implemented algorithms and, with appropriate design, the system is unbiased and can better produce a diverse range of result structures which are consistent with the data and which may include novel molecules that might not otherwise be considered. Following the generation of a potential result set, it must be carefully filtered and ranked for evaluation. At the same time, if the chemist has some structural hypotheses, the program should provide an opportunity to generate all conceivable corollaries, i.e. a full set of structures meeting the input suggestions.

In this work, we utilized the commercially available software program ACD/Structure Elucidator from Advanced Chemistry Development (ACD/Labs). This system can

*Correspondence to: Gary J. Sharman, AstraZeneca, Silk Road Business Park, Macclesfield, Cheshire SK10 2NA, UK.
E-mail: gary.sharman@astrazeneca.com

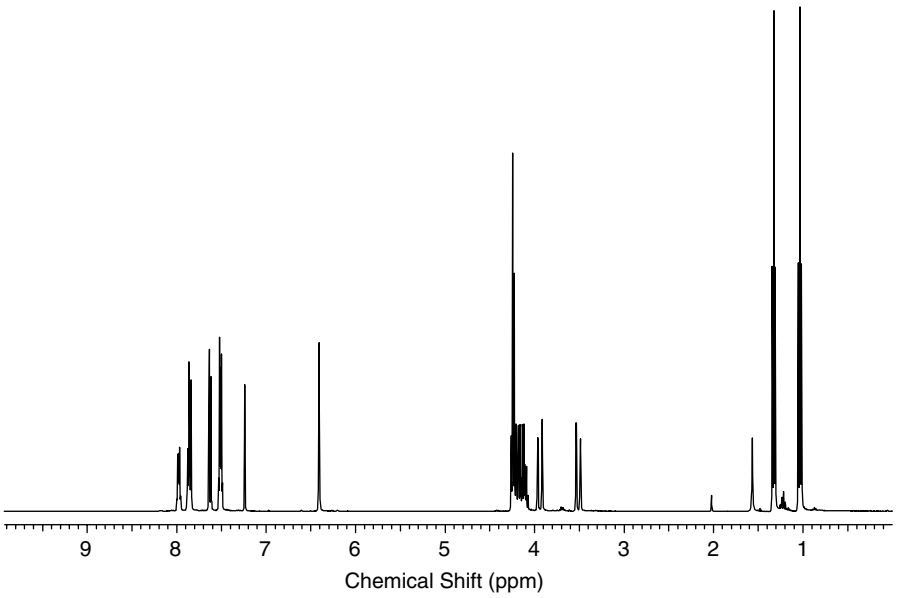


Figure 1. ^{1}H NMR spectrum of the first reaction product.

synergistically analyse one-dimensional (1D) and two-dimensional (2D) NMR spectra, IR spectra and mass spectra to elucidate a chemical structure. StrucEluc has been discussed in detail elsewhere.[3–5] The core of StrucEluc is a set of three software engines, or generators, that deduce structural information from spectroscopic data by different means. The generator used in this study was the Correlation Spectroscopy Based (CSB) generator, which uses 2D NMR data and does not require a 1D ^{13}C NMR spectrum. The advantage of the CSB approach is that two-dimensional correlation data provide far more reliable evidence of molecular connectivity than chemical shift data alone, and can also differentiate structures where chemical shifts would be similar. This is commonly the case in larger molecules containing more than 20 skeletal atoms. In addition, when sample amounts are limited, low concentrations may preclude the acquisition of a conventional 1D carbon spectrum. However, indirectly detected heteronuclear correlation experiments may still be feasible. The program makes use of the ^{13}C dimension of short- and long-range correlation experiments to obtain most of the information that it could obtain from a conventional 1D carbon experiment. Furthermore, a multiplicity-edited HSQC experiment can give much of the information available in a DEPT experiment.

The second important feature of StrucEluc is that it provides the ability to rank structures according to how well they explain the available experimental data. Deviations (*d*-statistics) are calculated between experimental ^{1}H, ^{13}C, ^{15}N, ^{19}F and ^{31}P NMR chemical shift data and those predicted from the structure using the ACD/Labs NMR predictor modules. This is crucial as a large number of structures can in principal be created, and some objective measure of closeness of match to the experimental shifts is therefore very useful. Structures can also be ranked according to how well they agree with the experimental MS fragmentation data (if available).

RESULTS AND DISCUSSION

The NMR data utilised by the StrucEluc program in this study were extracted from the NMR spectra of the first unknown, and are summarised in Table 1. StrucEluc can suggest a potential molecular formula or formulae based on an input mass and associated NMR spectra. In this case, the program suggested a single molecular formula of $C_{20}H_{20}O_5$ from the measured low-resolution molecular mass and the carbon count extracted from the HMBC and HMQC spectra.

The 2D NMR correlations were used to generate molecular connectivity diagrams (MCDs), from which the resultant structure set may be derived (Fig. 2). However, inspection of many of the hypothetical structures so generated revealed highly improbable ring junctions. Moreover, few of these structures contained the naphthalene carbon skeleton present in the starting material, which seemed improbable on chemical grounds. When a consistency check of the MCD was performed by the program, it was found that there were a number of apparent contradictions in the 2D NMR data. The origin of these contradictions can be explained in the following way.

Deriving structures from the MCD is based on a number of optional settings representing the number of bonds over which correlations are observed. The defaults for strong correlations were used in the initial run, which are 2–3 bonds both for HMBC and for COSY. Although these limits are generally applicable, they are by no means always appropriate. In conjugated or aromatic systems, for example, it is fairly common to see four-bond HMBC correlations, and in certain systems five-bond proton–proton couplings can be observed. By increasing the range of correlations used by the program, the likelihood of inappropriately constrained linkages is reduced, but the number of possible structures generated can increase dramatically, as can the associated computational time. A useful compromise, therefore, is to increase the range of constraints for definite atoms on the basis of some heuristics and empirical rules.[4,5] One such

Table 1. Summary of 2D NMR data used for **2**

Chemical shift	COSY correlations	HMQC correlations	HMBC correlations
7.98	7.51	124.2	127.3 and/or 127.6, 129.3 and/or 128.9, 135.5, 143.4
7.87	7.51	128.9	127.3 and/or 127.6, 131.0
7.85	7.62	131.0	143.4, 138.4, 135.5, 129.3, 128.9, 118.4
7.62	7.85	118.4	131.0, 135.5, 138.4, 143.4, 157.3
7.51	7.87, 7.98	127.3, 127.6	124.2, 128.9 and/or 129.3, 135.5
6.40	3.94, 3.51	109.5	47.4, 138.4, 167.1 (weak)
4.24	None	None	81.7, 47.4, 143.4, 176.1
4.23	1.32	60.1	167.1, 14.4
4.17	1.03, 4.12	62.8	13.9, 176.1
4.12	1.03, 4.17	62.8	13.9, 176.1
3.94	3.51, 6.40	47.4	109.5, 138.4, 143.4, 157.3, 167.1, 176.1, 81.7 (weak)
3.51	3.94, 6.40	47.4	81.7, 109.5, 138.4, 157.3, 167.1, 143.4 (weak), 176.1 (weak)
1.32	4.23	14.4	60.1
1.03	4.12, 4.17	13.9	62.8

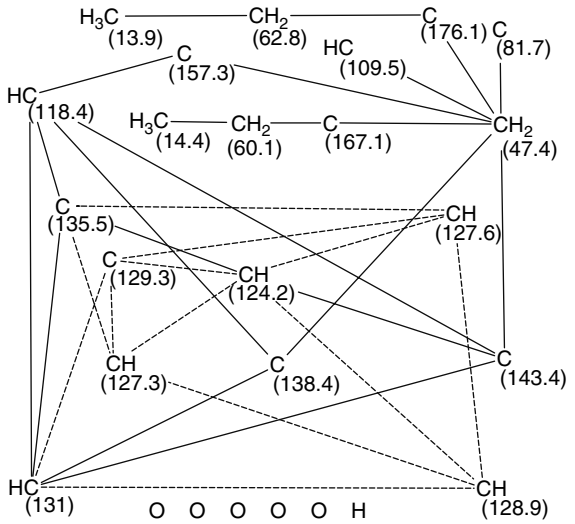


Figure 2. Molecular connectivity diagram obtained from the CSB generator.

approach is to increase the default mappings, but only for those atoms showing the highest number of correlations. The greater the number of correlations, the more likely they are to include some longer range ones, as the more bonds out from an atom we go, the more possibilities for correlations there are. The default mappings were therefore increased to 2–4 and 2–5 bonds for HMBC and COSY, respectively, but only for the two protons showing the highest number of correlations. Structure generation was then repeated, resulting in an increase in computation time to 4.5 h and the generation of ca 12 000 structures. These were ranked based on the deviation of their ^{13}C chemical shifts from calculated values. In the StrucEluc system, the ranking is carried out in two stages. In the first stage, the program predicts the ^{13}C NMR chemical shifts by a fast method,[5] and all the structures are ranked in order of increasing deviation between experimental and calculated spectra (d_F). Then more accurate methods of ^{13}C and ^{1}H spectrum calculation are applied to the first 10–25 structures of the previously ranked file and these structures are ranked again by new ^{13}C NMR deviations (d_A). The highest ranked four structures of the final file are shown in Fig. 3. The top-ranked structure, demonstrating a deviation of $d(1)_A = 3.71$ ppm, is **2**. The top structure is reliably considered as the most probable one as the difference $d(2)_A - d(1)_A = 3.0$ ppm is rather large. The structure of **2** was later confirmed to be correct by x-ray crystallography.

The possibility of dramatically reducing the computation time by the introduction of a 'user fragment' *before* the generation of structures has been demonstrated in previous

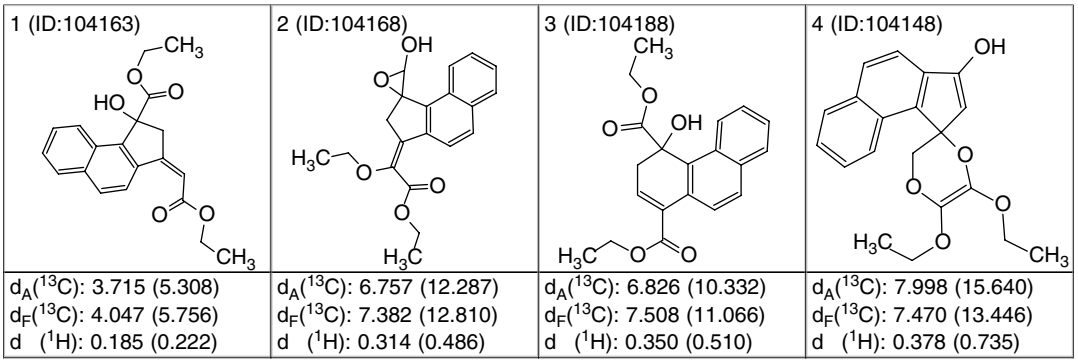
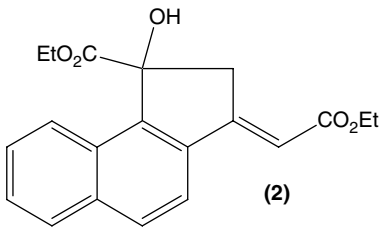


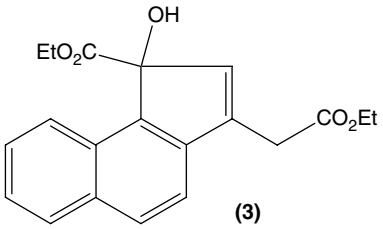
Figure 3. Four structures from the start of the output file. The highest ranked structure is the first one (compound **2**), which was confirmed as the correct structure by x-ray crystallography.



(2)

studies. The naphthyl carbon skeleton found in the starting material could provide such a fragment. This assumption is reasonable based on the nature of the reaction and a survey of the ^{1}H NMR spectrum. Such an approach is common in most structural elucidations, where prior information from starting materials is available and is used to influence manual spectral interpretation. It should be noted that this approach does reduce the 'unbiased' nature of the computer-aided structure elucidation process, but is nevertheless pragmatic and reasonable in most cases. The CSB generator was therefore run again using this user fragment. The reduction in time and the number of resultant structures was substantial: 14 structures were generated with seven remaining after filtration and removal of duplicates, in a total time of 1 s. Again, the correct structure was ranked number one. This clearly demonstrates that large gains in efficiency which can be achieved by using all of the available data, here in the form of knowledge of the starting materials.

With the structure of **2** solved, attention turned to the second product of the reaction. Considering the assigned structure of **2**, together with inspection of the NMR data for the second impurity, it seemed highly likely that the structure was **3**, in which the double bond is *endo* rather than *exo*. An NOE measurement was performed for both isomers, which gave the expected confirmatory result—a strong NOE enhancement between one of the naphthyl protons and the olefinic proton for the first compound, but between naphthyl and methylene for the second compound. This was subsequently confirmed by x-ray crystallography. However, for completeness, we also tested the StrucEluc program on the second dataset. Simply using the available data as input (Table 2), and without biasing with the data from structure **2**, we obtained four candidates, three of which contained an unlikely nine-membered bridged fragment, whereas the fourth was the structure **3**. Of these, structure **3** was ranked highest.



(3)

The similarity of the two structures provided us with an excellent way to test the effect of 'system training' on the elucidation. From the experimental spectra, it was clear that both compounds had common features. Therefore, the elucidation results of one (the assigned molecule) should be helpful in speeding the elucidation and improving the ranking predictions of the other. For this process, a database of one or more assigned structures is processed to yield two new databases: a fragments database, used for construction of new molecules, and a ^{13}C NMR predictor database, used to improve the accuracy of the predictions for the evaluation of candidate structures. If the unknown is sufficiently similar (for example, by visual comparison of the spectra) to a previously elucidated molecule, or if there is a collection of somewhat similar molecules with assignments available, a third advantage of training can be realized: these system training databases can be used in place of the internal databases. This can result in tremendous time savings (particularly for complex natural products), although it should be noted that this also can tend to bias the results by excluding grossly dissimilar structures from consideration.

In the present case, solving for structure **3** using the assignments and structure of **2** produced only a single candidate molecule, **3**. The reverse, solving for **2** using the data of **3**, also resulted in a single candidate, **2**. In both cases, the prediction accuracy was not affected significantly, but the

Table 2. Summary of 2D NMR data for **3**

Chemical shift	COSY correlations	HMQC correlations	HMBC correlations
7.88	7.46	122.7	125.0, 132.9, 139.7
7.84	7.43	129.6	126.7, 128.5, 129.6, 132.9, 141.3 (interchangeable with 7.83)
7.83	7.39	128.5	126.7, 128.5, 129.6, 132.9, 141.3 (interchangeable with 7.84)
7.46	7.39, 7.88	126.7	128.4 (w), 128.5 (w)
7.43	7.84	118.2	132.9, 139.7, 140.3
7.39	7.46, 7.83	125.0	122.7, 132.9
6.26	3.61 (weak)	134.5	33.7, 83.9, 139.7, 141.3
4.17	1.23	60.8	13.7, 169.8
4.09	0.98	62.3	13.5, 174.3
3.97	None	None	83.9, 139.7(w), 174.3
3.61	6.26 (weak)	33.7	134.5, 140.3, 141.3, 169.8
1.23	4.17	13.7	60.8
0.98	4.09	13.5	62.3

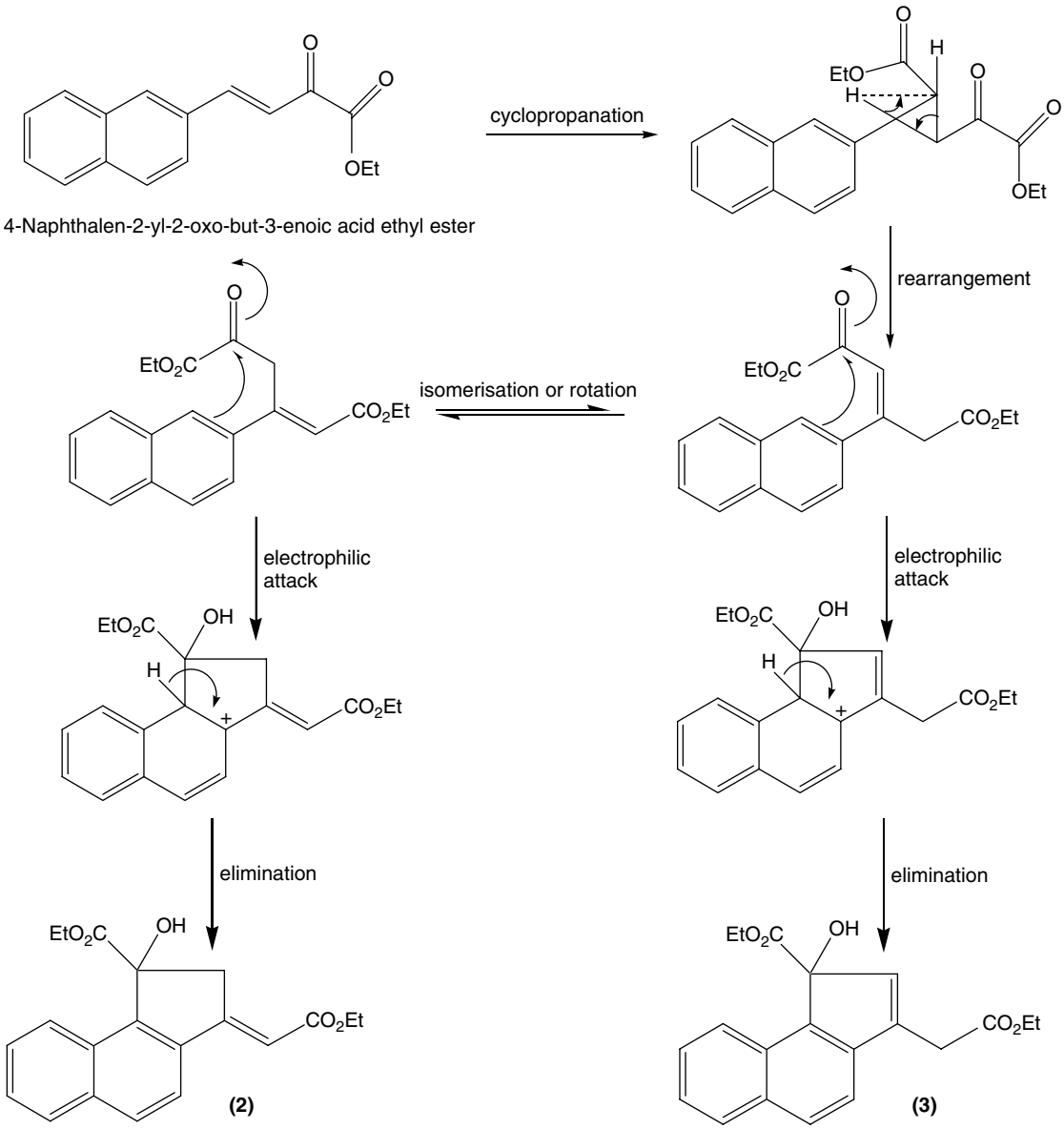
generation time was <1 s. Because the molecules were small, the gains due to training were not particularly significant. However, for much larger molecules, it could mean the difference between days and minutes. Furthermore, this is typical of many structure elucidation problems—often there are several related compounds to be solved, so these efficiency gains are particularly relevant.

A possible mechanism for formation of these products is shown in Scheme 1. The initial step is the expected cyclopropanation of the α,β-unsaturated pyruvate **1** by the carbene derived form ethyl diazoacetate. This is followed by a rearrangement to give an olefin. Ring closure then occurs, through electrophilic addition to the keto moiety, followed by elimination. Ring closure on to either of the two carbonyls, or migration of the double bond after rearrangement of the cyclopropane, can account for the double bond being *endo* or *exo* in the final products, resulting in structures **2** and **3**, respectively. Although we cannot be certain that this is the mechanism, it is certainly plausible and rationalizes the formation of the two products.

In conclusion, a computer-aided approach allowed us to identify rapidly the unexpected products of this reaction. Furthermore, basing the search on the connectivity information present in 2D NMR data proved essential for finding the correct structures.

EXPERIMENTAL

All NMR experiments were performed in $CDCl_3$ using a Bruker DRX500 spectrometer and standard library pulse sequences. For ^1H spectra, 32 K complex points were collected. FIDs were multiplied by an exponential line broadening function of 0.2 Hz prior to Fourier transformation. For 2D experiments, 2 K points (4 K for DQF COSY) were collected in f_2, with 512 f_1 increments. Spectral widths were 10 ppm in f_2, and in f_1 10 ppm for COSY, 160 ppm for HMQC and 220 ppm for HMBC experiments. The f_2 data were not zero-filled and f_1 data were zero-filled to 1 K. Data were



Scheme 1. Possible mechanism for the formation of the two products.

processed using a shifted sine-squared bell window function. Structure calculations were conducted using Version 6.00 of the ACD/Structure Elucidator program on a Pentium III 1 GHz system equipped with 512 MB RAM and the Microsoft Windows 2000 operating system.

Preparation of 3-ethoxycarbonylmethylene-1-hydroxy-2,3-dihydro-1*H*-cyclopenta[*α*]naphthalene-1-carboxylic acid ethyl ester (2) and 3-ethoxycarbonylmethyl-1-hydroxy-1*H*-cyclopenta[*α*]naphthalene-1-carboxylic acid ethyl ester (3)

To a stirred solution of pyruvate **1** (R = naphthyl, R′ = ethyl) (1.0 g, 4 mmol), in freshly distilled DCM (150 ml), under a nitrogen atmosphere at room temperature, was added zirconium tetrachloride (0.932 g, 4 mmol). On addition of the Lewis acid the yellow solution of pyruvate turned dark red–brown. Ethyl diazoacetate (0.63 ml, 6 mmol) was then added dropwise over $\sim$ 5 min (vigorous gas evolution was observed). The reaction was monitored for complete disappearance of starting material by TLC (25% EtOAc–hexane). After 12 h no starting material remained and the reaction was quenched by addition of saturated brine solution (150 ml). The organic layer was separated and washed with distilled water (2 × 100 ml). This solution was then dried over magnesium sulfate before concentration *in vacuo*, followed by crystallization (EtOAc–hexane), to yield 203 mg of an off-white solid. TLC analysis showed this to be a mixture of two products, which were then separated by flash chromatography to give, in order of elution, **2** [0.122 g, 0.36 mmol, 9%; ν_{max} (film), 1733 (C=O), 1672 (C=O), 1638 (C=C) and 1376 (C–OH); ^{1}H NMR ($CDCl_3$), δ 7.98 (m, 1H), 7.87 (m, 1H), 7.85 (d, J = 8.8 Hz, 1H), 7.62 (d, J = 8.8 Hz, 1H), 7.51 (m, 2H), 6.40 (t, J = 2.4 Hz, 1H), 4.24 (s, 1H, exchangeable), 4.23 (q, J = 7.1 Hz, 2H), 4.17 (m, 1H), 4.12 (m, 1H), 3.94 (dd, J = 19.1, 2.4 Hz, 1H), 3.51 (dd, J = 19.1, 2.4 Hz, 1H), 1.32 (t, J = 7.1 Hz, 3H) and 1.03 (t, J = 7.1 Hz, 3H); MS, m/z (EI+) 363 ($M + Na^+$), 358 ($M + NH_4$), 341 ($M + H$), 323 ($M - OH$)], followed by **3** [0.065 g, 0.191 mmol, 4.8%; ν_{max} (Nujol), 1731 (C=O), 1712 (C=O) and 1377 (C–OH); ^{1}H NMR ($CDCl_3$), δ 7.88 (d, J = 8.4 Hz, 1H), 7.84 (d, J = 8.2 Hz, 1H), 7.83 (d, J = 8.4 Hz, 1H), 7.46 (t, J = 8.4 Hz, 1H), 7.43 (d, J = 8.2 Hz, 1H), 7.39 (t, J = 8.4 Hz, 1H), 6.26 (t, J = 1.0 Hz, 1H), 4.17 (q, J = 7.1 Hz, 2H), 4.09 (m, 2H), 3.97 (s, 1H, exchangeable), 3.61 (d, J = 1.0 Hz, 2H), 1.23 (t, J = 7.1 Hz, 3H) and 0.98 (t, J = 7.1 Hz, 3H); MS, m/z (EI+) 363 ($M + Na^+$), 358 ($M + NH_4$), 341 ($M + H$), 323 ($M - OH$)].

REFERENCES

1. Li A, Dai L, Aggarwal VK. *Chem. Rev.* 1997; **97**: 2341.
2. (a) Boger DL, Robarge KD. *J. Org. Chem.*, 1983; **53**: 3373; (b) Evans DA, Olhava EJ, Johnson JS, Janey JM. *Angew Chem., Int. Ed.* 1998; **37**: 3372; (c) Thorhauge J, Johannsen M, Jorgensen KA. *Angew Chem., Int. Ed.* 1998; **37**: 2404.
3. Blinov KA, Elyashberg ME, Molodtsov SG, Williams AJ, Martirosian ER. *Fresenius' J. Anal. Chem.* 2001; **369**: 709.
4. Elyashberg ME, Blinov KA, Williams AJ, Martirosian ER, Molodtsov SG. *J. Nat. Prod.* 2002; **65**: 693.
5. Elyashberg ME, Blinov KA, Williams AJ, Martirosian ER, Molodtsov SG. *J. Chem. Inf. Comput. Sci.* accepted for publication.