Tetrahedron Vol 49, No 43, pp 9845-9854, 1993
Printed in Great Britain

0040-4020/93 $6 00+ 00

Cholic Acid as an Architectural Component in Biomimetic/ Molecular Recognition Chemistry; NMR and Molecular Mechanics Study of a "Tetra-acetoxycholaphane".

Richard P. Bonar-Law and Anthony P. Davis*

Department of Chemistry, Trinity College, Dublin 2, Ireland

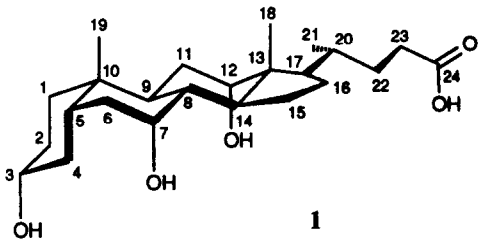
(*Received in UK* 19 *July* 1993)

Abstract *A detailed NMR analysis of cholaphane* **2** *has resulted in the measurement of most of the* ^{1}H *and* ^{13}C *chemical shifts, and many of the* ^{1}H-^{1}H *coupling constants Some of this information has been used to provide "NMR constraints" for a molecular mechanics study of the macrocycle, resulting in a set of four minima which are proposed to dominate the conformational equilibrium of* **2** *in* $CDCl_3$

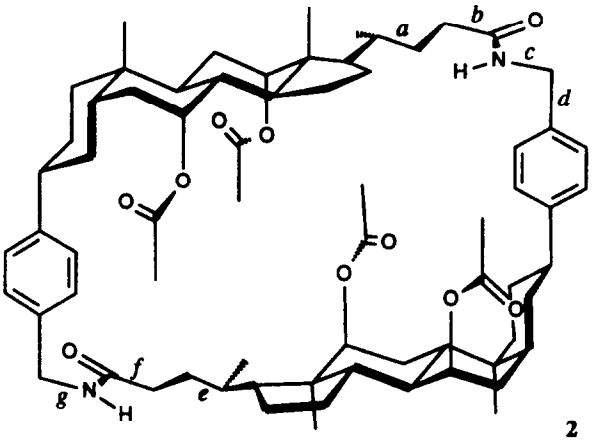
Introduction

In the foregoing paper,[1] and in references cited therein, we and others have noted the potential of cholic acid (**1**) as a building block in biomimetic/molecular recognition chemistry, and have demonstrated its utility in the construction of extended, preorganised molecular architectures Obvious features of **1** are its size, rigidity and degree of functionalisation, but an equally important (though perhaps more subtle) attribute is the irregular, highly asymmetric nature of its structure There are 24 carbons and 40 hydrogens in **1** and, leaving aside the protons on the three methyl groups, no two are in identical environments Moreover, because of the condensed, well-defined structure of the steroid nucleus, there may be quite substantial differences even between chemically similar atoms (e g methylene hydrogens)

This asymmetry is of significance for the assembly of synthetic receptors etc from **1**, because it increases the possibilities for performing regio- and stereo-selective transformations during the syntheses (a good example being the differentiation of the three secondary hydroxyl groups highlighted in the following paper) [2] However, it has a second advantage which is manifested during the later stages of such projects Because of their distinct environments, the protons and carbons derived from **1** give distinct NMR signals, in principle they can be identified and used to give remarkably detailed information on the system in which they are embedded



1



2

The preceding article gave details of the synthesis of "cholaphane" **2**, the first steroidal macrocycle designed with a view to applications in biomimetic/molecular recognition chemistry As an illustration of the potential of NMR in the study of such molecules, we now report (a) an almost complete assignment of the ^{1}H and ^{13}C spectra of **2**, (b) the use of extracted coupling and chemical shift data to infer conformational preferences for certain regions of the macrocycle, and (c) a molecular mechanics study within the foregoing "NMR constraints" which results in what we feel is a reasonable hypothesis as to the conformational preferences of **2** in $CDCl_3$ solution

Results and Discussion

NMR investigations. For the purposes of this study, we recorded the following NMR spectra of cholaphane **2**, (a) ^{1}H (400 and 500 MHz) and ^{13}C 1D spectra (including DEPT),[3] (b) several NOE difference spectra (500 MHz), (c) a ^{13}C-^{1}H COSY spectrum, (d) a NOESY spectrum (400 MHz) and (e) two ^{1}H-^{1}H DQF phase-sensitive COSY spectra (500 MHz), one at low resolution covering the full range of the spectrum and one at high resolution covering the region δ 0 6-2 6 The high-resolution ^{1}H-^{1}H COSY, shown in Figure 1, was particularly informative In conjunction with the 1D ^{1}H spectrum it enabled us to determine the chemical shifts of the protons on the steroid nuclei with very few ambiguities, the exceptions being certain pairs of diastereotopic protons in the more flexible portions of the framework In addition, it was possible to extract a substantial quantity of information concerning coupling patterns and constants The results are summarised in Tables 1 and 2, accompanied by further details of the assignment procedure In general the pattern of assignments corresponds to that obtained previously for simple bile acid derivatives,[4-9] although we have attempted to derive more information on ^{1}H-^{1}H coupling than the earlier workers [5-7]

Table 1 ^{1}H NMR Assignments and Coupling Patterns for **2**[a]

Proton	δ	Mult	*J* (Hz)	Proton	δ	Mult	*J* (Hz)
1α	1 80	*dt*	*14, 3*	15'[c]	1 14	qd	12, 6
1β	1 12	td	14 0, 3 4	16[c]	1 905	dtd	13, 9 5, 6
2α	1 24	qd	13, 3 0	16'[c]	1 30	dddd	13, 12, 9 5, 3
2β	1 59	br d	13	17	1 69	br q	9 6
3β	2 44	tt	12, 3 5	18-Me	0 76	s	
4α	2 24	q	13 5	19-Me	0 93	s	
4β[b]	1 48	m		20	1 47	br m	
5[b]	1 50	m		21-Me	0 82	d	6 1
6α	1 645	*dt*	*15 0, 2.5*	22[c]	1 81	*td*	*13, 4*
6β	1 99	ddd	15 0, 4 0, 3 4	22'[bc]	1 45	m	
7	4 92	q	3 3	23[c]	2 21	td	13, 3 5
8	1 65	td	12 0, 4 0	23'[c]	2 045	*td*	*13, 4*
9	2 09	*ddd*	*13, 11,* 4	NH	5 85	dd	7 6, 4 0
11α	1 765	dt	15 0, 3 5	NCH	4 77	dd	15 0, 7 6
11β	1 505	*ddd*	*15, 13, 3*	NCH'	4 07	dd	15 0, 4 0
12	5 095	t	3 0	Ar	7 18	d	8 3
14	1 87	*ddd*	*13, 11, 7*	Ar'	7 15	d	8 3
15[c]	1 38	dddd	13, 10, 7, 3	OAc	2 026	s	
				OAc	2 033	s	

[a] Chemical shifts are expressed in p p m relative to TMS See drawing of **2** for numbering system Most assignments arose directly from connections made by the ^{1}H-^{1}H COSY spectra, and the coupling patterns revealed by the 1D and high-resolution ^{1}H-^{1}H COSY spectra, confirmation of H-C-H connectivities being provided by the ^{13}C-^{1}H COSY Starting points were the 7β,12β and 3β (benzylic) protons 6α and 6β could not be distinguished safely from their coupling patterns, and were assigned by means of an NOE enhancement of 6α on irradiating the 19-CH_3 resonance (which experiment also confirmed the identity of the latter) (*cf* ref 7) The C22 and C23 protons could not be linked to the rest of the spectrum because of the lack of a clear correlation to 20-H (see text) However, an assignment on the basis of chemical shifts should be fairly secure, as the C23 protons are α to C=O Multiplicities and coupling constants were determined by analysis of the 1D spectrum in conjunction with a contour plot of the high resolution ^{1}H-^{1}H COSY In cases where heavy reliance was placed on the COSY the letters and figures are italicised in the Table and are subject to a degree of uncertainty (± ~1 Hz for the coupling constants) No attempt was made to correct for second-order effects [b] A detailed analysis of this coupling was not attempted because of overlap with a neighbouring proton (4β and 5, 20 and 22') [c] One of a diastereotopic pair of protons which are not readily distinguishable The protons resonating at higher field are marked "prime"

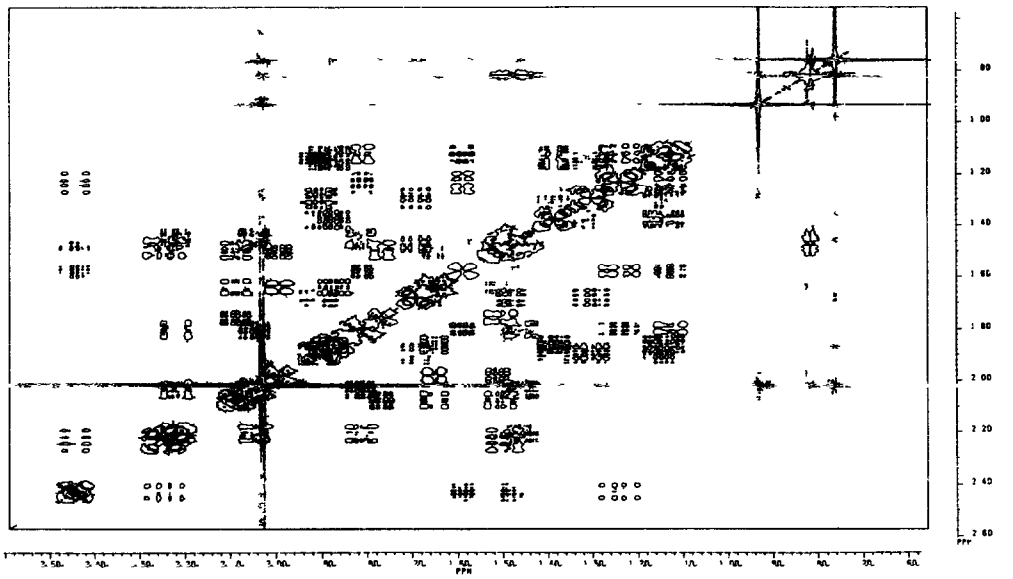


Figure 1: High resolution DQF ^{1}H-^{1}H COSY of **2** at 500 MHz, covering the region of the spectrum from δ 0 6 to 2 6 p p m

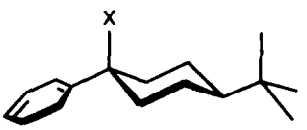
Table 2 ^{13}C NMR Assignments for **2**[a]

Carbon	δ	Carbon	δ	Carbon	δ
1	37 0	12	75 8	23	33 2
2	29 8	13	45 0	24	173 3
3	44 0	14	44 4	7-OCO*Me*	21 2
4	35 8	15	22 8	7-OCOMe	169 8
5	42 5	16	27 6	12-OCO*Me*	21 6
6	31 6	17	46 5	12-OCOMe	170 0
7	71 0	18	12.1	Ar	146 8
8	37.5	19	23 0	Ar	136 2
9	29.5	20	35 0	2Ar	127 4
10	34 1	21	17 4	2Ar	126 7
11	25 7	22	32 4	CH_2N	42 8

[a] Chemical shifts expressed in p p m relative to TMS See drawing of **2** for numbering system Most of the assignments emerged from the analysis of the ^{1}H spectrum (see Table 1), the expected multiplicities being being confirmed by DEPT The quaternary carbons 10 and 13 were distinguished by comparison with the literature.[9] The small degree of uncertainty regarding the C22 *vs* C23 ^{1}H assignments translates to the corresponding ^{13}C signals.

A priori it was not clear that this study could do more than complete the characterisation of **2** The examination of a CPK model of the macrocycle indicated that, despite the rigidity of the steroid nucleus and the aromatic spacer group, the framework might have considerable flexibility Unless it transpired that a very few conformations (preferably a single one) were dominant in $CDCl_3$ at room temperature, averaging effects would diminish the value of the data to the point where little could be said However, consideration of the ^1H spectrum encouraged us to suspect that the molecule might indeed have surprisingly little conformational freedom Our attention was drawn to the following points

(1) The chemical shift difference between the 2α and 2β protons is 0 35 p p m , similar to that expected for a 3α-hydroxy cholanoic acid derivative [4] This suggested that the aromatic ring spends most of its time in the orientation specified in the drawing of **2**, i e eclipsing the C3-H bond, and roughly perpendicular to the C_2 axis of the macrocycle In compounds of the general form **3**, in which the conformation shown is enforced by a bulky geminal substituent X, the neighbouring equatorial protons are quite strongly deshielded relative to their axial counterparts [10]



3

(2) The coupling pattern for 17-H implies that $^3J_{17,20}$ is *ca* 9 5 Hz, and thus that the predominant conformation for the C17-C20 bond places these hydrogens *anti* rather than *gauche* (i e as shown in the drawing of **2**)

(3) Both protons on C23 experience large and small vicinal couplings (13 and *ca* 4 Hz respectively) Considering the conformations available to the C22-C23 bond (Figure 2) it is clear that only conformation **A**, with C20 and C24 *anti* to each other, is consistent with this pattern Moreover, the *difference* between the vicinal constants may be used to argue that conformations **B** and **C** are relatively unimportant, a substantial population of either should perturb the values from those observed

(4) The two HN-CH couplings are significantly different from each other, indicating a preferred local conformation in which the N-H is angled towards a particular benzylic proton

(5) The 20-H resonance, which is almost superimposed on 22'-H (see Table 1), shows at most a very weak correlation with 22-H in the high resolution ^1H-^1H COSY This suggests that one of the protons on C22 has a very small coupling constant to 20-H (≤ 2 Hz), placing quite firm restrictions on the possible conformations and conformational freedom of the C20-C22 bond

The recurring theme in the above analysis is the implication that individual parts of the macrocyclic framework may have clear preferences for particular local conformations In some cases this was not unexpected A molecular mechanics study on monomeric unit **4** established that the C3-aryl group did have a substantial preference for the "C-H eclipsed" conformation, and that the C17-C20 conformation is effectively fixed to be as shown in the diagrams [11] The latter point finds support in the crystallography of cholic acid and its derivatives, where the illustrated C17-C20 conformation seems to be ubiquitous [12,13] However the molecular mechanics study implied that the C20-C22 and C22-C23 bonds should not have strong intrinsic

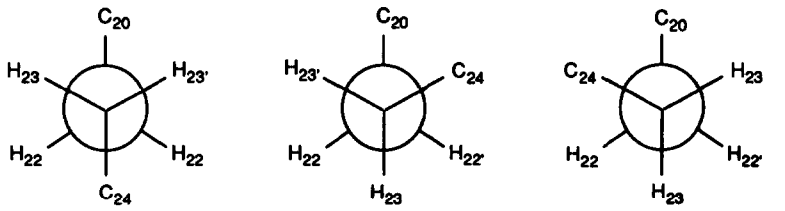
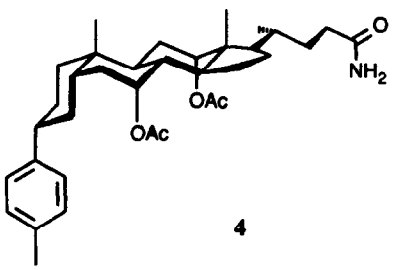


Figure 2: Conformations about the C22-C23 bonds in **2** The H-22/22' and 23/23' distinctions could not be made from the NMR spectra (see Table 1), and are therefore arbitrary in the diagram

preferences,[14] and one would expect the NH-CH_2 bond to be capable of essentially free rotation It therefore seemed that the macrocyclic framework may be imposing a fairly high degree of order on the region between C20 and the aromatic spacer, raising the possibility that rather few conformations may be important in $CDCl_3$ solution



4

In order to explore this question further, it was necessary to resort to molecular mechanics calculations Although aware of the limitations of this technique when dealing with large, potentially flexible molecules in solution, we felt that by operating within the constraints of the NMR data it might be possible to obtain a meaningful result A conformational search was performed on **2** using the MacroModel[15] molecular modelling package In order to simplify the calculations we decided that, during the initial search, the C3-Ar, C17-C20 and C22-C23 bonds should be kept in the positions implied by the NMR analysis and (in the first two cases) the preliminary calculations on **4** The procedure used for the search can be understood by reference to the drawing of **2** Starting from a partially-minimised structure (of which the drawing is a fair representation) a benzylic C-N bond (g) was broken and conformations were generated by nested rotations of bonds (b), (c) and (d) at 10° resolution, (f) at 60° resolution, and (a) and (e) at 120° resolution The criterion for ring closure was a C N distance of 1 - 2 Å 323 conformations were generated, and passed through a series of minimisations and eliminations to give 35 distinct minima spanning an energy range of 29 KJ mol^{-1} During the later stages of the process the GB/SA solvation treatment ($CHCl_3$)[16] was included in the energy calculations

The energies of the first 12 conformers are summarised in Table 3, along with calculated percentages at 297 K according to the Boltzmann distribution (ignoring the effect of higher energy species) The Boltzmann calculation takes account of the fact that the C_2-symmetric conformers are unique while the non-C_2-symmetric minima exist as degenerate pairs It can be seen that, according to the computations, the conformational equilibrium should be dominated by four low-energy minima within *ca* 1 KJ mol^{-1} of baseline, with smaller contributions from a fifth and sixth Structures (i)-(iv), of which (ii) and (iii) are C_2-symmetric, are shown in Figure 3

Although disappointed that a single predominant conformation had not been predicted, we proceeded to calculate the expected NH,CH and 20-H,22-H vicinal couplings for each of the first six conformers, according to modified versions of the Karplus equation [17] The results are summarised in Table 4, along with average values weighted according to the Boltzmann calculation The experimental data is included for comparison We were pleased to find a fair degree of agreement between the last two lines, suggesting that the picture derived by theory may be quite close to the truth However, it is notable that the experimental coupling constants lie "outside" the calculated average values (i e are closer to the extremes of their possible ranges) Considering the values for the individual conformers, a reasonable interpretation is that conformation (ii) has been under-emphasized by the molecular mechanics calculation and may in fact be the single most important structure

The analysis presented in this paper suffers from a number of uncertainties, and we would not suggest that our conclusion is more than a good hypothesis However, the exercise does highlight a major advantage of the steroid nucleus as a structural component in biomimetic/molecular recognition chemistry Its inherent asymmetry means that each proton and carbon can act as an NMR probe, signalling information about a potential receptor's structural and complexation properties With regard to the latter, the possibility of intracomplex (intermolecular) NOE's is particularly striking Although **2** itself has not shown the ability to complex organic molecules, other cholaphanes (notably the tetraol derived by *O*-deacetylation) have been found to be effective receptors for carbohydrate derivatives in organic solvents [18] We hope to find related systems in which the host-guest complex has a single, well-defined structure In such cases, intermolecular NOE's involving the array of protons on the internal surface of the host should reveal the guest's location with unusual precision

Experimental

NMR spectra were run in $CDCl_3$ solution on Bruker WH 400 or AM 500 instruments Standard methods were used for the 2D experiments [19] The molecular mechanics studies employed MacroModel V3 1X,[15] running on a Silicon Graphics IRIS 4D25TG workstation Energy calculations used the MM2 force field, and conformational searches were performed using MULTIC sub-mode The minimisation sequence used to for the conformers of **2** was as follows, Polak-Ribiere Conjugate Gradiant (PRCG) 100 steps, PRCG 500 steps, PRCG 500 steps, PRCG 1000 steps, Full Matrix Newton Raphson (FMNR) 120 steps Duplicate conformations were rejected at the end of each stage, allowance being made for the constitutional C_2 symmetry The last two stages included the GB/SA solvation treatment ($CHCl_3$) [13] With a few exceptions (of relatively high energy) all structures minimised to an RMS gradient of < 0 01 KJ/Å Figure 3 was produced using QUANTA 3 3 (Molecular SimulationsCorp) after translation of the relevant structures from the MacroModel format

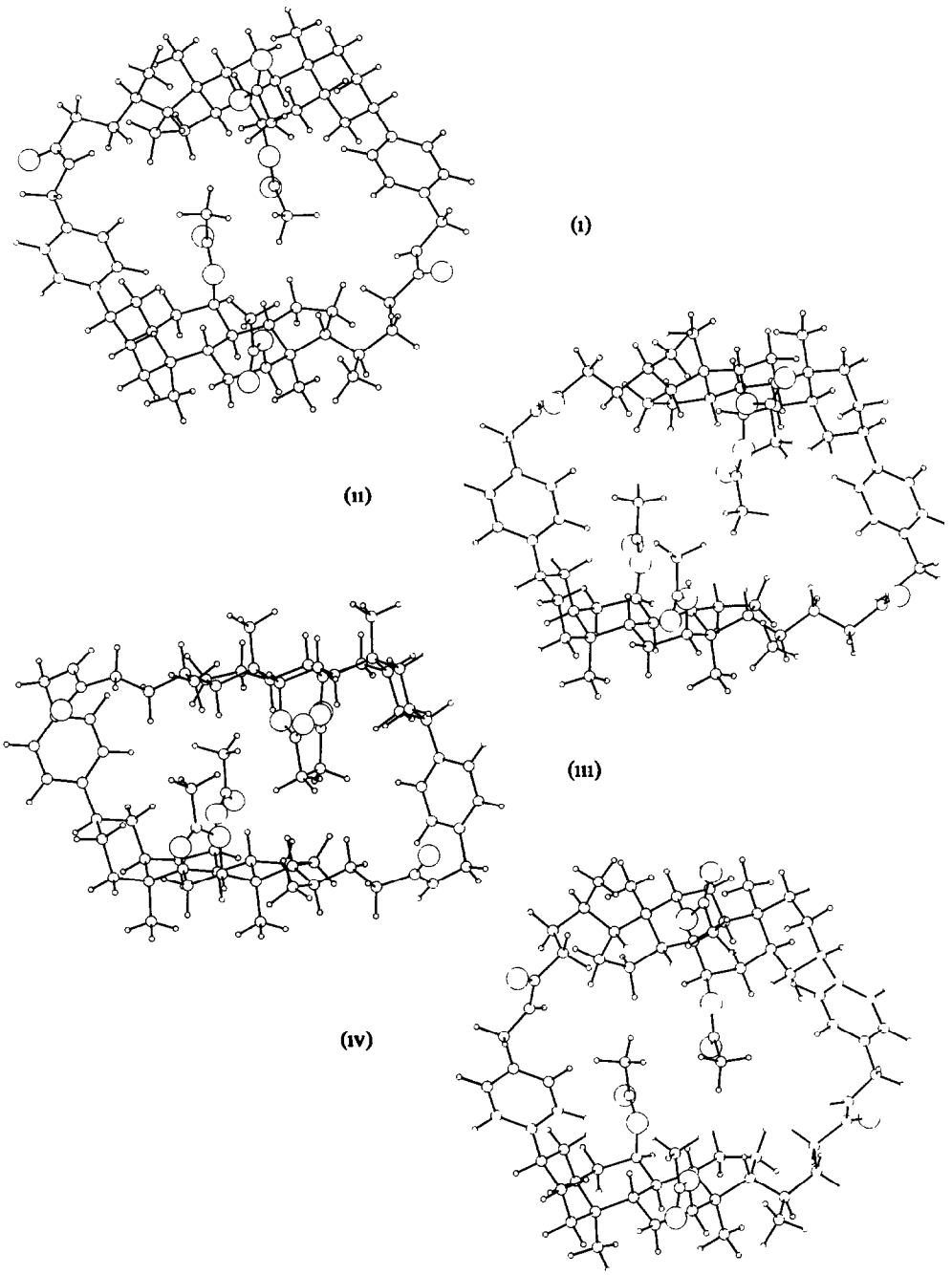


Figure 3: Conformations (I) - (IV) (the four lowest-energy minima) from the molecular mechanics study on **2** The largest circles represent oxygen atoms, the smallest hydrogens, and the remainder represent carbons and nitrogens

Table 3 Molecular Mechanics Energies and Calculated Occupancies for the Lowest-Energy Minima Located During the Conformational Search of **2**

Conf	Energy ($KJmol^{-1}$)	C_2-symmetry	Occupancy[a] (%)	Conf	Energy ($KJmol^{-1}$)	C_2-symmetry	Occupancy[a] (%)
(i)	645 56	No	38 7	(vii)	654 04	Yes	0 63
(ii)	646 51	Yes	13 2	(viii)	654 18	Yes	0 59
(iii)	646 57	Yes	12 9	(ix)	657 43	No	0 32
(iv)	646 59	No	25 6	(x)	657 49	No	0 31
(v)	648 78	Yes	5 3	(xi)	657 50	No	0 31
(vi)	652 98	No	1 9	(xii)	657 72	No	0 28

[a] Calculated according to a Boltzmann distribution, see text

Table 4 Estimated Coupling Constants, Including Weighted Averages, for Conformers (i)-(vi) (see Table 3) Observed Values are Included for Comparison

Conformer	$^3J_{NH,CH}$ (Hz)		$^3J_{20,22}$ (Hz)	
	pro-*R* CH	pro-*S* CH	pro-*R* 22-H	pro-*S* 22-H
(i)	7 0, 5 3	2 3, 6 8	3 9, 1 7	3 0, 12 0
(ii)	8 8	3 1	1 7	12 1
(iii)	6 3	5 7	2 3	12 3
(iv)	6 5, 5 3	2 5, 6 9	3 8, 4 4	3 0, 2 4
(v)	5 4	6 7	4 1	2 6
(vi)	8 8, 2 7	3 1, 8 8	1 5, 2 5	11 9, 12 3
Weighted Avg [a]	*6 4*	*4 7*	*3 0*	*7 3*
Observed	7 6	4 0	≤ 2	

[a] Calculated using the occupancies listed in Table 3, with double weighting for values derived from the C_2-symmetric structures (ii), (iii) and (v)

Acknowledgements. Financial support was provided by EOLAS, the Irish Science and Technology agency The TCD Molecular Graphics Facility was funded in part by Loctite Corporation We are grateful to Prof M T Reetz and Dr A E Derome for assistance in obtaining high-field NMR spectra and advice concerning their interpretation, and Dr M E Bridge for the computer programme used to calculate the Boltzmann distribution

References and Footnotes

1 Bonar-Law, R P , Davis, A P , preceding paper in this journal

2 Bonar-Law, R P , Davis, A P , Dorgan, B J , following paper in this journal

3 Abbreviations DEPT, distortionless enhancement by polarisation transfer; NOE(SY) nuclear Overhauser effect (spectroscopy), DQF, double quantum filtered, COSY, correlated spectroscopy

4 Kirk, D N , Toms, H C , Christal, D , White, K A , Smith, K E , Latif, S , Hubbard, R W P *J Chem Soc Perkin Trans II* **1990**, 1567

5 Campredon, M , Quiroa, V , Thevand, A , Allouche, A , Pouzard, G *Magn Reson Chem* **1986**, *24*, 624

6 Waterhous, D V , Barnes, S , Nuccio, D D *J Lipid Res* **1985**, *26*, 1068

7 Barnes, S , Geckle J M *J Lipid Res* **1982**, *23*, 161

8 Blossey, E C , Cannon, R G , Ford, W T , Periyasamy, M , Mohauraj, S , *J Org Chem* **1990**, *55*, 4664

9 Blunt, J W , Stothers, J B *Org Magn Res* **1977**, *9*, 439

10 (a) Geneste, P , Herrmann, P , Kamenka, J M , Pons, A *Bull Soc Chim France* **1975**, 1619 (b) Davis, A P , Egan, T J , Orchard, M G , Cunningham, D , McArdle, P *Tetrahedron* **1992**, *48*, 8725 The latter reference reports a number of 1-substituted 4-t-butyl-1-phenylcyclohexanes **3** In cases with bulky 1-substituents, chemical shift differences of 0 9 - 1 5 p p m were observed between the axial and equatorial protons on carbons 2 and 6

11 MacroModel[15] calculations revealed that the C3-aryl bond would revert to the "C-H eclipsed" conformation from any starting position on PRCG minimisation, and that the illustrated C17-C20 conformation (i e with the protons in an *anti* relationship) was preferred over the two alternatives by ≥8 KJmol^{-1}

12 Leading references (a) Miyata, M , Shibakami, M , Chirachanchai, S , Takemoto, K , Kasai, N , Miki, K *Nature (London)* **1990**, *343*, 446 (b) Hogan, A , Ealick, S E , Bugg, C E , Barnes, S *J Lipid Res* **1984**, *25*, 791 (c) Giglio, E , Quagliata, C *Acta Cryst* **1975**, *B31*, 743

13 Crystal structure of an alternative framework Davis, A P , Orchard, M G , Slawin, A M Z , Williams, D J *J Chem Soc , Chem Commun* **1991**, 612

14 Support is again provided by crystallographic determinations in which various side-chain conformations may be observed See e g ref 12c

15 Mohamadi, F , Richards, N G J , Guida, W C , Liskamp, R , Lipton, M , Caulfield, C , Chang, G , Hendrickson, T , Still, W C *J Comput Chem* **1990**, *11*, 440 We are grateful to Prof Still for providing this software on very reasonable terms

16 Still, W C , Tempczyk, A , Hawley, R C , Hendrickson, T *J Am Chem Soc* **1990**, *112*, 6127

17 (a) $^3J_{NH\,CH}$ de Marco, A , Llinás, M *J Mag Res* **1980**, *39*, 253 (b) $^3J_{20,22}$ Haasnoot, C , de Leeux, F , Altona, C *Tetrahedron* **1980**, *36*, 2783

18 (a) Bonar-Law, R P , Davis, A P , Murray, B A *Angew Chem , Int Ed Eng* **1990**, *29*, 1407 (b) Bhattarai, K M , Bonar-Law, R P , Davis, A P , Murray, B A *J Chem Soc , Chem Commun* **1992**, 752

19 Derome, A E *Modern NMR Techniques for Chemistry Research*, Pergamon Press Oxford, 1987