(3S,6S)-3-Isopropyl-1,4-diazabicyclo[4.4.0]decane-2,5-dione, $C_{11}H_{18}N_2O_2$: Bayesian Statistics in Data Analysis

By A. T. H. Lenstra, M. Verbruggen, B. Bracke and F. Vanhouteghem

University of Antwerp (UIA), Department of Chemistry, Universiteitsplein 1, B-2610 Wilrijk, Belgium

and F. Reyniers and F. Borremans

State University of Ghent, Laboratory of Organic Chemistry, Krijgslaan 281-S4, B-9000 Gent, Belgium

Abstract

$C_{11}H_{18}N_2O_2$, $M_r = 210 \cdot 3$, monoclinic, $P2_1$, $a = 10 \cdot 503$ (4), $b = 10 \cdot 464$ (7), $c = 11 \cdot 012$ (4) Å, $\beta = 110 \cdot 74$ (2)°, $V = 1132$ (1) Å³, $Z = 4$, $D_x = 1 \cdot 234$ Mg m⁻³, λ(Mo $K\alpha$) = 0·71073 Å, $\mu = 0 \cdot 10$ mm⁻¹, $F(000) = 456$, room temperature, $wR = 0 \cdot 058$, $R = 0 \cdot 102$ for 2518 observed reflections out of 2538 measured, and 270 variables. The compound, for which the synthesis is also reported, consists of a piperidine ring in the chair conformation and a diketopiperazine ring in a very shallow boat form with C(3) and C(6) as bowsprits. The two rings are *cis* fused. The isopropyl group, pseudo-axially bonded, is folded over the diketopiperazine ring. The unit cell contains two symmetrically independent molecules A and B. Infinite chains, consisting exclusively either of type A or type B molecules, are linked by intermolecular N(4)—H⋯O(1) bonds. To compensate for the systematic weakness of reflections with $h + l =$ odd, a new data analysis scheme was employed, based on Bayesian statistics. It led to an increase in the number of observed reflections from 1155 to 2518. The comparison between a refinement based on 1155 and one based on 2518 reflections showed that the Bayesian re-evaluation of data decreases the final e.s.d.'s considerably and does not create any detectable artefacts.

Introduction

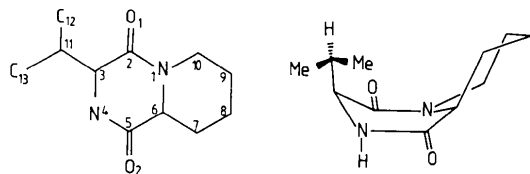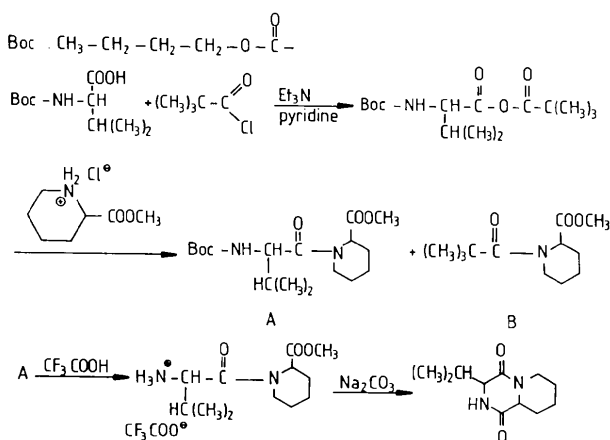The title compound, shown in Fig. 1, contains the 2,5-diketopiperazine moiety (abbreviated DKP) and is a cyclic dipeptide composed of L-valine and L-pipecolic acid. The compound, also known as *cis-cyclo*[-L-Val-L-Pec-], was synthesized in the series of reactions presented below.

It belongs to a group of compounds with interesting biological and medicinal properties. The restrictions brought about by the DKP ring in combination with the rigidity of the piperidine ring make the title compound an attractive model compound in studies of peptide conformational analysis (Ramani, Sasisekharan & Venkatesan, 1977; Anteunis, 1978). Typical aspects are the shape of the DKP ring, the type of fusion of the two rings, the conformation of the valanyl residue and the planarity near N(1). The X-ray determination reported here is one of a series (Dillen & Lenstra, 1983; Van Poucke, Geise & Lenstra, 1983; Van Poucke & Lenstra, 1982*a,b*), the reports of which should be useful in the interpretation of NMR spectra and chemical activity.

Furthermore, the title compound has a pseudo-*B*-centered lattice, causing about half of the intensities to be weak. A standard least-squares analysis based on reflections for which $I(hkl) \geq 3\sigma(I)$ would be dominated by the data with $h + l =$ even, whereas data with $h + l =$ odd tend to be missing. The pseudo

Fig. 1. Structural formula and atomic numbering scheme (left) and conformation (right) of the molecule.

B-centering leads to poor e.s.d. values and might even introduce a bias into the molecular model. We therefore decided to explore an alternative approach which allows *all* observed data to be used at any stage of the structure analysis and to test the results against a standard analysis. A detailed description of the alternative approach will be given elsewhere (Lenstra, Geise & Vanhouteghem, 1990). Here we will confine ourselves to the basic features of the newly developed routine.

Experimental

Mixed anhydride between trimethylacetic acid and N-(n-butoxycarbonyl)-L-valine

A solution of 0·5 g (2·3 mmol) *N*-(*n*-butoxy-carbonyl)-L-valine in 25 ml toluene was cooled to 273 K, and 0·18 g (2·3 mmol) pyridine, 0·23 g (2·3 mmol) triethylamine and 0·25 g (2 mmol) trimethylacetyl chloride were added. The mixture was stirred for 15 min after which 0·2 g (2 mmol) triethylamine and 0·36 g (2 mmol) of the hydrochloric acid salt of L-pipecolic acid methyl ester were added. The mixture was stirred at room temperature for 20 h. The organic phase was washed three times with 10 ml of a citric acid solution (10%) and three times with 10 ml of a saturated $NaHCO_3$ solution, subsequently dried over anhydrous $MgSO_4$ and evaporated. The residue was separated by column chromatography on silica gel, using diethyl ether as eluent, into 0·43 g (52%) mixed anhydride (R_f = 0·83; TLC, ether) and 0·16 g (35%) *N*-trimethylacetyl-L-pipecolic acid methyl ester (R_f = 0·75; TLC, ether).

Trifluoroacetic acid salt of L-valyl-L-pipecolic acid methyl ester

0·5 g (1·4 mmol) of the foregoing mixed anhydride was dissolved in 10 ml trifluoroacetic acid and the mixture stirred at room temperature for 1 h. The remaining trifluoroacetic acid was evaporated to give 0·84 g (95%) of the trifluoroacetic acid salt of L-valyl-L-pipecolic acid methyl ester.

(3S,6S)-3-Isopropyl-1,4-diazabicyclo[4.4.0]decane-2,5-dione

A solution of 0·5 g (1·3 mmol) of the foregoing trifluoroacetic acid salt in 20 ml saturated $NaHCO_3$ solution was stirred at room temperature for 2 h and extracted three times with 20 ml $CHCl_3$. After drying the $CHCl_3$ extracts over anhydrous $MgSO_4$ and evaporation, the residue was purified by sublimation and finally recrystallized from diethyl ether. Yield: 0·16 g (60%), R_f = 0·29 (TLC, ether).

A solution of the title compound in $CDCl_3$ used in an NMR analysis was set aside for several weeks and

left unattended. The very slow diffusion of $CDCl_3$ vapour through the plastic stopper of the NMR tube caused formation of a single crystal, which was used in the X-ray diffraction experiment. Unit-cell dimensions deduced from 25 reflections in the range $8 < \theta < 12°$. Enraf–Nonius CAD-4 diffractometer, $\omega/2\theta$ scan mode, Mo radiation monochromatized by pyrolytic graphite. Maximum Bragg angle 27°; 2531 independent measurements in the ranges $0 \le h \le 13$, $0 \le k \le 13$, $-14 \le l \le 14$. The distribution of E values showed a pseudo *B*-centering: for (*hkl*) with $h + l = 2n$ we found $\langle E^2 \rangle = 1\cdot6$, whereas $\langle E^2 \rangle = 0\cdot3$ for $h + l = 2n + 1$. As a result of this accidental additional symmetry many reflections are of low intensity. Indeed only 1155 reflections had values $I > 3\sigma(I)$. Intensity control, every 2 h, showed no drift. No absorption correction applied ($\mu = 0\cdot10$ mm^{-1}, crystal size $0\cdot1 \times 0\cdot1 \times 0\cdot15$ mm). Structure solved using a combination of *MULTAN* (Germain, Main & Woolfson, 1971) and Fourier techniques. H atoms placed at expected positions, keeping their Debye–Waller temperature parameter at $(B_{iso} + 1)$ Å^2, where B_{iso} is the isotropic *B* value of the atom to which H is attached.

Bayesian statistics in data analysis

A net intensity *I* is routinely obtained by subtracting the local background *B* from the raw intensity *R*:

$$I = R - B. \qquad (1)$$

Since the measured quantities *B* and *R* are taken as independent the variance is usually taken as:

$$\sigma^2(I) = \sigma^2(R) + \sigma^2(B). \qquad (2)$$

The application of Bayesian statistics allows the introduction of *experience* into the evaluation of observed data, thus adding new elements to the classic data interpretation. French & Wilson (1978) have proposed a method to avoid negative values of the net intensities, based on Wilson statistics and on prior knowledge that $I(hkl) \ge 0$. Consequently, all observed $F(hkl) = [I(hkl)]^{1/2}$ can be used in the Fourier calculations. We decided to explore how the constraint $I > 0$, together with prior knowledge of background values *B*, allows better estimates of *I* and $\sigma^2(I)$.

X-ray background radiation is known (Keulen, 1969) to decrease slowly with increasing Bragg angle. This suggests that within a suitably small $\sin\theta/\lambda$ interval the distribution of *B* values can be approximated by a Gaussian function, but not necessarily one following counting statistics. To check on this we divided the data set into $\sin\theta/\lambda$ intervals, each containing about 200 independent measurements and, most importantly, we included corrections for diffractometer design and operations (*e.g.* corrected

for the fact that the measuring time is proportional to $a + b\tan\theta$). Then we found practically identical values for $\langle B \rangle_{hkl}$ and $\langle s^2(B) \rangle_{hkl}$. This means that the distribution of B over different hkl is equal to the distribution of B over a time series observed at one single reciprocal lattice point. Thus, in each interval the 200 background values follow a counting statistical distribution, *i.e.* a Gaussian function of the form $\exp[-(B - \langle B \rangle)^2/2\langle B \rangle]$.

To exploit fully the gain in experience we need to summarize at this point some aspects of conditional probability theory. The likehood function, $P(B|b)$, expresses the probability that the experimental value B is observed under the condition that the ideal value is b. Having measured B, one obviously is even more interested in b and $\sigma^2(b)$, or in mathematical language in the first and second moment of $P(b|B)$ which is called the posterior function.

1st moment $\quad b = \langle b|B \rangle = bP(b|B)\mathrm{d}b \quad$ (3a)

2nd moment $\quad \sigma^2(b) = \langle b^2|B \rangle = b^2P(b|B)\mathrm{d}b.$ (3b)

$P(B|b)$ and $P(b|B)$ are related through the theorem of Bayes:

$$P(b|B) = P(B|b)P(b)/P(B) \quad (4)$$

where $P(B)$ and $P(b)$ are the unconstrained probabilities on B and b, respectively. The function $P(b)$ is usually called the 'prior' function. We now set out to find proper analytical forms for the various probability functions occurring in equations (3) and (4).

If an X-ray intensity B follows counting statistics, B is linked to b through:

$$P(B|b) \simeq \exp[-(B - b)^2/2b] \quad (5)$$

provided B is not too small. If B represents a very small number of counts a Poisson function should replace the Gaussian form, but leaving the Gaussian form only results in a small error. Furthermore, $P(B)$ represents one particular measurement, and is single-valued. Thus $P(B)$ is a delta function. It serves in (4) as a normalizing factor and can be left out of the further discussion. $P(b)$ must express the prior knowledge ('experience') one can add to the data evaluation. Its choice is critical for the success of the approach. We begin by analyzing the background situation. When a background value B is observed for a particular reflection in a $\sin\theta/\lambda$ interval with $N + 1$ observations, the total background experience prior to the $(N + 1)$th observation summarizes the available N data. Thus, it must account for $N\langle B \rangle$ counts with an error of $[N\langle B \rangle]^{1/2}$. For a single observation one has to divide by N and the prior distribution becomes:

$$P(b) \simeq \exp[-(b - \langle B \rangle)^2/(2\langle B \rangle/N)]. \quad (6)$$

Substitution of equations (4)–(6) into equation (3) leads to:

$$\langle b|B \rangle \simeq \langle B \rangle$$

and

$$\sigma^2(b) \simeq \langle B \rangle/N.$$

The posterior moments replace the original observations B and $\sigma^2(B) = B$, which are typical for an isolated experiment. Here we allow the prior, *i.e.* our experience, to dominate the posterior moments. Inclusion of the background experience alone already changes (1) and (2) and leads to:

$$I = R - \langle B \rangle \quad (1b)$$

and

$$\sigma^2(I) = \sigma^2(R) + \langle B \rangle/N. \quad (2b)$$

Since $\langle B \rangle/N$ vanishes with respect to $\sigma^2(R) = R$, even for a zero intensity $(R \simeq B)$, the Bayesian approach changes (2b) into $\sigma^2(I) \simeq \sigma^2(R)$ for all practical purposes. So far, we have improved the background values directly, and in doing so the net intensities I indirectly. However, better approximations of I and $\sigma^2(I)$ can be made from a similar Bayesian analysis of R [equations (1) and (2)]. In the classic approach only R is experimentally accessible, so that only R and $\sigma^2(R) = R$ are known. Since we now accept the background to be characteristic of a particular $\sin\theta/\lambda$ interval and to follow counting statistics we can say that the *experimental* value of B in equation (1) – if it were possible to perform the experiment – must lead to a result that follows a distribution centered around $\langle B \rangle$ with a spread $\sigma(B) = \langle B \rangle^{1/2}$. Substitution of this together with $\sigma^2(R) = R$ into (2) leads to:

$$\sigma^2(I) = R - \langle B \rangle = I. \quad (7a)$$

Including the uncertainty on the background average $\langle B \rangle$, which equals $[\langle B \rangle/N]^{1/2}$, the net intensity I has a variance

$$\sigma^2(I) = (I + \langle B \rangle/N). \quad (7b)$$

Thus the Bayesian approach allows values for $\sigma^2(R)$, $\sigma^2(I)$ and $\sigma^2(B)$ to be inferred, as if R, I and B could be measured separately. Ultimately since we require r, i and b (the ideal values of R, I and B, respectively), we write

$$P(r|R) \underset{R}{\simeq} P(B|b)P(b)P(I|i)P(i). \quad (8)$$

We know that $P(B|b)$ and $P(I|i)$ follow counting statistical distributions, while $P(b)$ is given by equation (6). Under ordinary measuring conditions R is only measured once, so that the prior $P(i)$ cannot be constructed in a way similar to $P(b)$. In the absence of a better estimation, we took $P(i)$ to have a uniform distribution in combination with the constraint $P(i) = 0$ for $i < 0$. The latter forces the ideal net intensity i to be non-negative. Using these distri-

butions in (8) and employing only those potential observations *B* and *I* for which $B + I = R$, we calculated numerically $\langle i \rangle$ and $\sigma^2(i)$ as the 'best' estimates of a net intensity and its variance.

Results and discussion

As stated above, two independent analyses were performed. One using the classical values *B*, *I*, $\sigma(I)$, and the other using the Bayesian corrected values *b*, *i*, $\sigma(i)$. Both analyses involved full-matrix least-squares refinements (on *F*) of all positional parameters. Non-H atoms were refined anisotropically, while H atoms were refined isotropically constraining their temperature parameter to $(B_{iso} + 1)$ Å^2, where B_{iso} is the isotropic thermal parameter of the atom to which H is attached. Reflections were given individual weights according to their $\sigma(I)$ or $\sigma(i)$ values. No extinction coefficient was refined. Enraf–Nonius *SDP* (Frenz, 1978) employed. The results, summarized in Table 1, show only a small increase in *wR* (from 0·042 to 0·058) upon introduction of 1363 weak reflections, usually considered poorly determined. The unweighted *R* value doubles, as can be expected. Addition of 1363 very small $|F(hkl)|$ values (average 0·2 on an absolute scale) hardly changes $\Sigma|F_{obs}|$, but will affect $\Sigma||F_{obs}| - |F_{calc}||$. The fact that $\Sigma||F_{obs}| - |F_{calc}||$ changes almost linearly with the number of contributing reflections is another sign that the Bayesian corrections produce rather good intensity estimates. Furthermore, the noise level in the final difference Fourier does not change significantly. The maximum peak height in the difference Fourier based on 1155 reflections is 0·22 e Å^{-3}, compared with a maximum peak height of 0·28 e Å^{-3} when 2518 reflections are included. Atomic scattering factors taken from *International Tables for X-ray Crystallography* (1974, Vol. IV), Enraf–Nonius *SDP* (Frenz, 1978) employed.

For most atoms the positional and anisotropic thermal parameters resulting from the two analyses differ by less than one e.s.d. The maximum parameter shift is three e.s.d.'s. Therefore in Table 2 we give only the refined parameters* emerging from the Bayesian-treated data.

The first impression is that no bias is introduced by the pseudo *B*-centering and so it seems that the Bayesian approach has been superfluous. However, least-squares and Fourier indicators showed (Table 1) that Bayesian re-evaluation of intensities has not created any detectable artefacts, but decreased the

* Lists of structure factors, H-atom positions and anisotropic thermal parameters have been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 53421 (30 pp.). Copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

Table 1. *Comparison of results of analyses performed on classically treated intensity data and on data after a Bayesian treatment*

	Classical data	Bayesian data
Number of observations	1155	2518
Number of variables	270	270
$\Sigma| F_{obs}| - |F_{calc}||$	569	1483
$\Sigma|F_{obs}|$	12398	15186
$\Sigma w| F_{obs}| - |F_{calc}||^2$	3767	9410
$\Sigma w\, F_{obs}|^2$	2132080	2751960
$\Sigma||F_{obs}| - F_{calc}||/\Sigma|F_{obs}|$	0·046	0·102
$\Sigma w||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$	0·042	0·058
|Maximum density in difference Fourier|	0·22 e Å^{-3}	0·28 e Å^{-3}

Table 2. *Positional parameters (fractional) and equivalent isotropic temperature factors* (Å^2)

See Fig. 1 for the numbering of the atoms. The e.s.d.'s given in parentheses refer to the least significant digit. Equivalent isotropic temperature factors are calculated from anisotropic temperature parameters using $B_{eq} = (4/3)[a^2B(1,1) + b^2B(2,2) + c^2B(3,3) + ab(\cos\gamma)B(1,2) + ac(\cos\beta)B(1,3) + bc(\cos\alpha)B(2,3)]$. All anisotropic temperature parameters were physically acceptable.

	x	y	z	B_{eq}
O(1,A)	0·3459 (3)	0·6957 (3)	0·4754 (3)	4·8
O(2,A)	0·2900 (3)	0·1919 (3)	0·4911 (4)	6·3
N(1,A)	0·2124 (3)	0·5274 (4)	0·4753 (3)	3·3
N(4,A)	0·4156 (3)	0·3602 (4)	0·4768 (3)	3·5
C(2,A)	0·3264 (4)	0·5790 (5)	0·4706 (4)	3·2
C(3,A)	0·4352 (4)	0·4943 (4)	0·4543 (4)	2·6
C(5,A)	0·3028 (5)	0·3068 (5)	0·4802 (4)	3·8
C(6,A)	0·1829 (4)	0·3911 (5)	0·4697 (4)	3·2
C(7,A)	0·0592 (5)	0·3612 (5)	0·3534 (5)	4·0
C(8,A)	−0·0611 (5)	0·4398 (6)	0·3533 (5)	5·1
C(9,A)	−0·0270 (5)	0·5808 (6)	0·3620 (6)	6·0
C(10,A)	0·0999 (5)	0·6090 (5)	0·4795 (5)	5·3
C(11,A)	0·4477 (4)	0·5206 (5)	0·3218 (4)	3·4
C(12,A)	0·5711 (5)	0·4474 (6)	0·3119 (5)	5·2
C(13,A)	0·3191 (5)	0·4901 (6)	0·2109 (5)	5·6
O(1,B)	0·8415 (3)	0·7010 (3)	−0·0690 (3)	4·9
O(2,B)	0·7978 (3)	0·2087 (4)	0·0385 (3)	5·8
N(1,B)	0·7149 (3)	0·5397 (4)	−0·0374 (3)	3·4
N(4,B)	0·9202 (3)	0·3660 (4)	−0·0047 (3)	3·3
C(2,B)	0·8272 (4)	0·5859 (5)	−0·0510 (4)	3·2
C(3,B)	0·9388 (4)	0·4945 (4)	−0·0488 (4)	3·0
C(5,B)	0·8064 (4)	0·3194 (5)	0·0054 (4)	3·6
C(6,B)	0·6863 (4)	0·4061 (5)	−0·0228 (4)	3·4
C(7,B)	0·5642 (4)	0·3609 (6)	−0·1386 (5)	4·2
C(8,B)	0·4415 (5)	0·4444 (6)	−0·1550 (5)	5·6
C(9,B)	0·4749 (5)	0·5864 (6)	−0·1648 (5)	5·6
C(10,B)	0·5992 (5)	0·6251 (5)	−0·0496 (5)	4·7
C(11,B)	0·9543 (5)	0·4923 (5)	−0·1826 (4)	3·6
C(12,B)	1·0822 (5)	0·4188 (6)	−0·1745 (5)	5·0
C(13,B)	0·8289 (5)	0·4411 (6)	−0·2890 (5)	5·5

e.s.d.'s substantially (in this case by about $2^{1/2}$). It is the (here 30%) reduction in error margins which makes the Bayesian data handling attractive.

Bond distances and valence angles are given in Table 3. They are similar to values found in analogous derivatives (Ramani, Sasisekhara & Venkatesan, 1977; Van Poucke & Lenstra, 1982*a*,*b*; Van Poucke, Geise & Lenstra, 1983; Dillen & Lenstra, 1983). Endocyclic torsion angles and ring puckering parameters (Cremer & Pople, 1975) are presented in Table 4. It follows that the piperidine ring has a chair conformation with a puckering very close to that observed in *trans-cyclo*(-D-Phe-L-Pec-) (Van Poucke & Lenstra, 1982*b*). The DKP ring approaches a boat form with C(3) and C(6)

Table 3. *Bond lengths (Å) and valence angles (°) of the two independent molecules (A and B) in the title compound*

	A	B		A	B
O(1)—C(2)	1·236 (7)	1·238 (5)	C(3)—C(11)	1·535 (7)	1·539 (7)
O(2)—C(5)	1·220 (7)	1·227 (7)	C(11)—C(12)	1·542 (7)	1·523 (7)
N(1)—C(2)	1·331 (7)	1·331 (7)	C(11)—C(13)	1·500 (5)	1·517 (5)
C(2)—C(3)	1·506 (7)	1·506 (7)	C(6)—C(7)	1·499 (5)	1·529 (5)
C(3)—N(4)	1·452 (7)	1·466 (7)	C(7)—C(8)	1·507 (7)	1·514 (7)
N(4)—C(5)	1·322 (7)	1·331 (7)	C(8)—C(9)	1·513 (7)	1·539 (7)
C(5)—C(6)	1·508 (7)	1·494 (6)	C(9)—C(10)	1·522 (5)	1·518 (5)
C(6)—N(1)	1·456 (7)	1·451 (7)	C(10)—N(1)	1·472 (7)	1·476 (7)
O(1)—C(2)—N(1)	122·5 (5)	122·3 (5)	N(1)—C(6)—C(7)	110·3 (4)	110·6 (4)
O(1)—C(2)—C(3)	117·7 (4)	118·8 (4)	C(6)—N(1)—C(2)	125·2 (5)	126·1 (5)
N(1)—C(2)—C(3)	119·8 (5)	118·9 (5)	C(2)—N(1)—C(10)	120·6 (5)	120·2 (5)
C(11)—C(3)—C(2)	109·4 (4)	109·9 (4)	C(6)—N(1)—C(10)	114·1 (4)	113·4 (4)
C(11)—C(3)—N(4)	113·6 (4)	112·0 (4)	C(6)—C(7)—C(8)	111·1 (4)	110·4 (4)
C(2)—C(3)—N(4)	112·9 (4)	113·2 (4)	C(7)—C(8)—C(9)	110·6 (4)	111·1 (4)
C(3)—N(4)—C(5)	126·6 (4)	126·3 (4)	C(8)—C(9)—C(10)	111·1 (4)	110·4 (4)
O(2)—C(5)—N(4)	123·4 (5)	122·2 (4)	C(9)—C(10)—N(1)	109·4 (4)	110·6 (4)
O(2)—C(5)—C(6)	117·8 (4)	119·3 (4)	C(3)—C(11)—C(12)	109·6 (4)	110·0 (3)
N(4)—C(5)—C(6)	118·9 (5)	118·5 (5)	C(3)—C(11)—C(13)	112·4 (4)	112·8 (4)
C(5)—C(6)—N(1)	114·2 (4)	114·5 (4)	C(12)—C(11)—C(13)	112·0 (4)	111·9 (4)
C(5)—C(6)—C(7)	113·0 (4)	112·5 (4)			

Table 4. *Endocyclic torsion angles (°) and Cremer-Pople ring-puckering parameters*

Torsion angles
E.s.d.'s are about 1°. The sign convention of IUPAC (1974) is used.

DKP ring	Mol. A	Mol. B	Piperidine ring	Mol. A	Mol. B
N(1)—C(2)—C(3)—N(4)	−14·1	−11·9	N(1)—C(6)—C(7)—C(8)	−55·7	−56·0
C(2)—C(3)—N(4)—C(5)	16·0	14·9	C(6)—C(7)—C(8)—C(9)	55·2	54·3
C(3)—N(4)—C(5)—C(6)	−4·8	−3·8	C(7)—C(8)—C(9)—C(10)	−54·8	−53·9
N(4)—C(5)—C(6)—N(1)	−8·4	−10·0	C(8)—C(9)—C(10)—N(1)	54·4	54·2
C(5)—C(6)—N(1)—C(2)	9·7	12·8	C(9)—C(10)—N(1)—C(6)	−56·9	−58·1
C(6)—N(1)—C(2)—C(3)	1·9	−1·4	C(10)—N(1)—C(6)—C(7)	57·8	58·5

Ring-puckering parameters
E.s.d.'s are according to Norrestam (1981). The sequences N(1), C(2), C(3), N(4), C(5), C(6) and N(1), C(6), C(7), C(8), C(9), C(10) were taken.

	DKP ring		Piperidine ring	
q_2(Å)	0·16 (1)	0·17 (1)	0·00 (1)	0·01 (1)
q_3(Å)	−0·03 (1)	−0·01 (1)	0·56 (1)	0·56 (1)
Q (Å)	0·17 (1)	0·17 (1)	0·56 (1)	0·56 (1)
φ_2(°)	302 (2)	306 (2)	353 (2)	234 (2)
θ_2(°)	99 (2)	94 (2)	180 (2)	179 (2)

Distance (Å) of C(13) from some intramolecular atoms

	N(1)	C(2)	C(3)	N(4)	C(5)	C(6)	C(7)
Mol. A	3·52	2·98	2·56	3·08	3·59	3·79	3·96
Mol. B	3·56	3·03	2·58	3·05	3·57	3·79	3·93

acting as bowsprits. In the notation introduced by Boeyens (1978) the DKP ring approaches the $B_{3,6}$ form. Judging the puckering by the Q parameter the ring is very shallow, comparable with the puckering of *cyclo*(-Gly-L-Bai-) (Dillen & Lenstra, 1983). A parameter of interest to NMR spectroscopists, the angle between the C(3)C(2)N(1)C(6) and C(3)N(4)-C(5)C(6) planes, is 13·0 (5) and 13·3 (5)° for the two crystallographically independent molecules.

A pyramidal configuration around the peptide N(1) atom has been observed in a number of dipeptides (Sletten, 1970; Pattabhi, Venkatesan & Hall, 1973, 1974; Van Poucke & Lenstra, 1982a,b), whereas the planar configuration seems to occur less frequently (Van Poucke, Geise & Lenstra, 1983; Dillen & Lenstra, 1983). In the title compound N(1)

is insignificantly [0·024 (4) and 0·042 (4) Å for the two molecules] out of the C(2)C(6)C(10) plane. Nor do we observe in the surrounding valence and torsion angles (Tables 3 and 4 and Fig. 2) signs of a deviation from sp^2 character.

Another factor of interest is the way the two rings are fused. In a *trans* fusion the two endocyclic torsion angles involved have different signs, whereas they have the same sign in a *cis* fusion. If we accept this as a definition, then the fusion (see Fig. 2a) in *cyclo*(-L-Val-L-Pec-) is *cis*. However, the most prominent feature connected with the side chain/ backbone interaction is the way the isopropyl group is situated with respect to the DKP ring. Judged by the polar angle (Cremer & Pople, 1975) and torsion angles (Figs. 1 and 2b,c) the isopropyl group is pseudoaxially bonded, while the methyl group C(13) is folded over the DKP ring.

The folding is characteristic for cyclic dipeptides containing tyrosyl and phenylalanyl residues, and now also appears with a valanyl residue. The torsion angle C(13)—C(11)—C(3)—N(4), *i.e.* χ'_{val} in the IUPAC–IUB Commission on Biochemical Nomenclature (1970) notation, has a value of about 63° (see Fig. 2c). It shows that the methyl group C(13) sits perfectly symmetrically over the ring. The same conclusion follows from the distances of C(13) to nearby nonbonded atoms (Table 4). Some of these distances, being considerably shorter than the sum of the van der Waals radii, cast some doubt on the conclusions of Chandrasekaran, Lakshminarayan, Mohanakrishnan & Ramachandran (1973) and Ramani, Sasisekhara & Venkatesan (1977) who attributed the folded conformation essentially to van der Waals interactions.

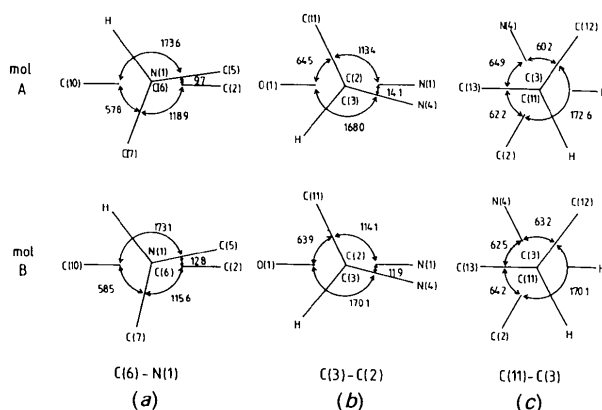The packing consists of two infinite chains of molecules. One chain consists of molecules A (inter-

Fig. 2. Newman projections (*a*) along C(6)—N(1) showing *cis*-like fusion of the two rings, (*b*) along C(3)—C(2) showing pseudo-axial positioning of isopropyl group and (*c*) along C(11)—C(3) showing the folding of the isopropyl group over the DKP ring.

related *via* the twofold screw axis) while the other consists of molecules *B* (also interrelated by the twofold screw axis). Within each chain the molecules are linked through N(4)—H···O(1) bridges with an N(4)···O(1) distance of 2·93 (chain *A*) and of 2·91 Å (chain *B*). Between the chains only van der Waals forces are present. The fact that O(1) is engaged in hydrogen bonding, whereas O(2) is not, also shows itself in a small, but distinct elongation of C(2)=O(1) compared to C(5)=O(2).

References

ANTEUNIS, M. J. O. (1978). *Bull. Soc. Chim. Belg.* **87**, 627–650.
BOEYENS, J. C. A. (1978). *J. Cryst. Mol. Struct.* **6**, 317–320.
CHANDRASEKARAN, R., LAKSHMINARAYAN, A. V., MOHANA-KRISHNAN, P. & RAMACHANDRAN, G. R. (1973). *Biopolymers*, **12**, 1421–1426.
CREMER, D. & POPLE, J. A. (1975). *J. Am. Chem. Soc.* **97**, 1354–1358.
DILLEN, J. & LENSTRA, A. T. H. (1983). *Acta Cryst.* C**39**, 905–907.
FRENCH, S. & WILSON, K. (1978). *Acta Cryst.* A**34**, 517–525.
FRENZ, B. A. (1978). *Computing in Crystallography*, edited by H. SCHENK, R. OLTHOF-HAZEKAMP, H. VAN KONINGSVELD & G. C. BASSI, pp. 64–71. Delft Univ. Press.
GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1971). *Acta Cryst.* A**27**, 368–376.
IUPAC (1974). *Rules for the Nomenclature of Organic Chemistry, Section E: Stereochemistry, Recommendations* 1974. Oxford: Pergamon Press.
IUPAC-IUB COMMISSION ON BIOCHEMICAL NOMENCLATURE (1970). *J. Mol. Biol.* **52**, 1–17.
KEULEN, E. (1969). PhD Thesis, Groningen, The Netherlands.
LENSTRA, A. T. H., GEISE, H. J. & VANHOUTEGHEM, F. (1990). To be published.
NORRESTAM, P. (1981). *Acta Cryst.* A**37**, 764–765.
PATTABHI, V., VENKATESAN, K. & HALL, S. R. (1973). *Cryst. Struct. Commun.* **2**, 223–226.
PATTABHI, V., VENKATESAN, K. & HALL, S. R. (1974). *J. Chem. Soc. Perkin Trans.* 2, pp. 1722–1727.
RAMANI, R., SASISEKHARAN, V. & VENKATESAN, K. (1977). *Int. J. Pept. Protein Res.* **9**, 277–292.
SLETTEN, E. (1970). *J. Am. Chem. Soc.* **92**, 172–177.
VAN POUCKE, M., GEISE, H. J. & LENSTRA, A. T. H. (1983). *Acta Cryst.* C**39**, 227–230.
VAN POUCKE, M. & LENSTRA, A. T. H. (1982*a*). *Cryst. Struct. Commun.* **11**, 853–859.
VAN POUCKE, M. & LENSTRA, A. T. H. (1982*b*). *Bull. Soc. Chim. Belg.* **91**, 213–218.

Structure and Molecular Orbital Studies of Potentially Mutagenic Methylchrysenes and their π–π* Electron Donor–Acceptor Molecular Complexes

BY DAVID E. ZACHARIAS, KEITH PROUT,† CYNTHIA B. MYERS AND JENNY P. GLUSKER‡

The Institute for Cancer Research, The Fox Chase Cancer Center, Philadelphia, PA 19111, *USA*

Abstract

The mutagenic and carcinogenic potency of 5-methylchrysene contrasts strongly with the lack of such activity in any other monomethylchrysene. In order to improve our understanding of the biochemical properties of these compounds, their electronic and molecular structures and π–π* electron donor–acceptor complexes have been examined by X-ray diffraction and molecular orbital methods. The crystal structures of the hydrocarbons chrysene (redetermination), 1-methylchrysene and 6-methylchrysene, and of the 1:1 complexes of 1,3,5-trinitrobenzene with chrysene, 2-methylchrysene, 3-methylchrysene, 4-methylchrysene, 5-methylchrysene, 6-methylchrysene, the 2:1 complex with 1-methylchrysene and the 1:1 complex of 5-methylchrysene with pyromellitic dianhydride have been determined. 5-Methylchrysene, the carcinogenic hydrocarbon, shows considerable disorder alone and in complexes. In the complexes the stacking of molecules involves an alternation of hydrocarbon with complexing agent, with the aromatic ring of 1,3,5-trinitrobenzene lying over a hydrocarbon bond involved in ring fusion, as

† Visiting scientist on leave from the Chemical Crystallography Laboratory, University of Oxford, 9 Parks Road, Oxford OX1 3PD, England.
‡ Address correspondence to this author.