A Theoretical Study of the Infrared Spectrum of Uracil

Ian R. Gould, Mark A. Vincent and Ian H. Hillier *
Department of Chemistry, University of Manchester, Manchester M13 9PL, UK

The harmonic frequencies, IR intensities and the potential energy distribution of the vibrational modes of the dioxo tautomer of uracil are calculated at the Hartree–Fock 6-31G** level. The use of a common scale factor of 0.9 results in excellent agreement between calculated and observed frequencies except for the C=O stretching modes. The use of a potential energy decomposition scheme is shown to allow a comparison with the atomic character of the modes that are inferred from experiment.

Due to their biological importance[1] there are continuing experimental and theoretical studies of the electronic structure of the pyrimidine bases and of their intermolecular interactions, particularly those involving hydrogen bonds. The technique of IR spectroscopy has been extensively used to study these nucleic acid constituents, particularly in low temperature matrices, where it is generally assumed that intermolecular interactions are minimised. This method has yielded information on the structure and tautomeric equilibria of cytosine,[2] adenine,[3] guanine,[4] thymine[5] and uracil.[5–7] Recently, the study of hydrogen-bonded complexes of uracil and its derivatives by IR spectroscopy has been discussed as a possible way of investigating the mutagenic uracil derivatives.[5]

A prerequisite for the proper interpretation of the IR spectra of such complexes is a complete understanding of the spectra of uracil itself. Work continues to be carried out both experimentally and theoretically in order to achieve this end. A recent paper by Graindourze *et al.*[5] reports experimental data and a more complete assignment and description of the vibrational modes than was hitherto available. On the theoretical side, the use of a relatively small basis set *ab initio* study was inadequate to interpret the vibrational modes, without the application of scale factors to the calculated force field.[8] Of particular concern was the substantial difference in these scale factors for the in-plane and out-of-plane motions. In this paper we use a large basis set to calculate the harmonic frequencies of the dioxo form of uracil to assess the value of a calculation near the Hartree–Fock limit, for predicting the IR spectrum of this molecule. We also use a potential energy decomposition scheme to provide a more quantitative description of the vibrational modes which can be compared to conclusions based upon isotopic substitution and spectral changes upon hydrogen bonding. It is generally agreed that only the dioxo tautomer of uracil is observed in spectral studies, a conclusion borne out by recent high level calculations of the relative energies of a number of tautomeric forms.[9]

Method

We have previously reported the optimised structure of the dioxo tautomer of uracil obtained using a high level (6-31G**) basis.[9] We have used this structure to obtain the harmonic frequencies and intensities by the use of analytic second derivatives as implemented in the program CADPAC.[10] The calculated frequencies of all modes were uniformly scaled by a single factor of 0.9. An analysis of the vibration modes was carried out by calculation of the potential energy distribution (PED) matrix as described by Keresztury and Jalsovszky,[11] utilising the force constant matrix computed within CADPAC.

The choice of the internal coordinates needed for this decompositon is not unique. Harsanyi *et al.*[8] have chosen a set which minimises the contribution of the off-diagonal elements of the force constant matrix. We have chosen our set to be as close to standard chemical concepts as possible, *i.e.* bond lengths, bond angles, torsions and deformations. In fact, we use all six bond stretches and in-plane bends involving non-ring atoms, five bond stretches involving ring atoms and four in-plane bends involving ring angles. There is rather more freedom in choosing the out-of-plane modes. Here we take four out-of-plane deformations involving non-ring atoms, three torsions involving ring atoms only and two torsions involving two ring and two non-ring atoms. In view of this somewhat arbitrary description of the internal coordinates, the associated PED is also not unique. However, we believe that the ones we have chosen are of value in discussing the nature of the normal modes of uracil.

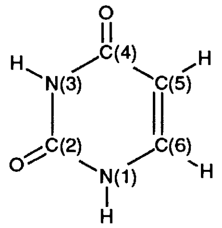

Fig. 1 Atom labelling of dioxo tautomer of uracil

Results and Discussion

The low temperature IR matrix spectrum of uracil has been reported by Szczesniak *et al.*,[7] Barnes *et al.*,[6] and more recently by Graindourze *et al.*[5] The latter discuss the description of the vibrational modes with the aid of data from isotopic substitution and from shifts observed upon hydrogen bonding. In Table 1 we compare the results of our calculations with these experimental data. Uracil has 30 fundamental modes, 21 belong to a′ and 9 to the a″ species. These may be formally classified as 6 N–H modes (2 stretching, 2 bending and 2 out-of-plane modes), 6 C=O modes, 6 C–H modes and 12 ring modes (6 stretching, 3 bending and 3 torsion). It is expected that substantial coupling between these various modes will occur, and will be revealed in our PED analysis. Although there is little doubt as to the actual assignment of the vibrational modes, we wish to study the quantitative agreement between the calculated and observed frequencies and between the predicted and experimentally inferred description of these modes. In the following discussion we use the atom labelling shown in Fig. 1.

In the high energy region of the spectrum the first two absorptions at 3485 and 3435 cm^{-1} are assigned to N(1)H and

Table 1 The experimental and calculated frequencies (cm^{-1}), calculated intensities (km mol^{-1}) and potential energy distribution (PED) of the absorption bands of uracil

Mode	Expt.[a]	Calculated[c] 6-31G**		PED(%)[d]
		$\tilde{\nu}$	A	
ν_1	3485 (2587)[e]	3511 (2577)[f]	133	ν N(1)H (100)
ν_2	3435 (2549)	3477 (2550)	99	ν N(3)H (100)
ν_3	— —	3087 (3087)	1	ν C(5)H(92)
ν_4	— —	3055 (3055)	4	ν C(6)H (93)
ν_5	1764 (1767)	1811 (1805)	451	ν C(2)O (53), ν C(4)O (26)
ν_6	1706 (1713)	1795 (1778)	1226	ν C(4)O (54), ν C(2)O (20)
ν_7	1643 (1634)	1663 (1656)	127	ν C=C (59), β C(6)H (15)
ν_8	1472 (1440)	1483 (1443)	147	β N(1)H (35), ring st (33), ν C(2)O (11)
ν_9	1400 (1380)	1401 (1379)	75	β N(3)H (35), ring st (19)
ν_{10}	1389 (1318)	1399 (1318)	110	ring bend (23), β N(3)H (21), β N(1)H (17)
ν_{11}	1359 (1255)	1377 (1266)	18	β C(6)H (28), β C(5)H (20), β N(1)H (10), ring st (21)
ν_{12}	1217 —	1213 (1129)	17	ring bend (13), β C(5)H (19), ring st (56)
ν_{13}	1185 (1113)	1176 (1109)	114	β C(6)H (40), β C(5)H (18), ring st (18), β N(1)H (14)
ν_{14}	1075 (970)	1058 (952)	20	ring str. (50), β C(5)H (26)
ν_{15}	980 (915)	969 (905)	7	ring bend (77), ring str. (14)
ν_{16}	958 (777)	948 (815)	7	ring str. (36), β N(3)H (12), ring bend (19), ν C(4)O (14)
ν_{17}	759 (724)	747 (742)	4	ring str. (61), ring bend (28)
ν_{18}	562 (543)	548 (537)	5	ring bend (72), β C(4)O (15)
ν_{19}	537 (524)	534 (524)	7	ring bend (38), β C(2)O (37), β C(40) (24)
ν_{20}	516 (508)	505 (494)	24	ring bend (90)
ν_{21}	391[b] (388)	383 (380)	32	β C(4)O (41), β C(2)O (31), ring bend (23)
ν_{22}	987 —	999 (999)	0.1	C(6)H oop def. (90)
ν_{23}	804 (803)	814 (814)	103	C(4)O oop def. (60), C(5)H oop def. (30)
ν_{24}	757 (758)	777 (774)	106	C(2)O oop def. (76)
ν_{25}	718	723 (719)	5	C(4)O oop def. (69), C(5)H oop def. (30)
ν_{26}	662	658 (501)	96	N(3)H oop def. (91)
ν_{27}	551	530 (403)	74	N(1)H oop def. (92)
ν_{28}	411[b]	387 (362)	31	Ring torsion (95)
ν_{29}	185[b]	161 (160)	1	Ring torsion (80)
ν_{30}	—	151 (147)	1	Ring torsion (84)

[a] Experimental data are from ref. 5 except where indicated otherwise. [b] Ref. 6. [c] All calculated frequencies have been scaled by 0.9. [d] Only PED contributions > 10% are quoted. [e] Experimental values for *N,N'*-dideuteriouracil, from ref. 7. [f] Calculated values for *N,N'*-dideuteriouracil.

N(3)H stretches respectively. The observed intensity ratio (1.31)[5] is close to our predicted value of 1.34 (Table 1). The experimental frequencies of the next two vibrations [C(5)H and C(6)H stretches respectively] are not definite owing to their very low intensity. The first (ν_3) has been reported[7] at 3130 cm^{-1}, and ν_4 at 2970 cm^{-1}, to be compared with our predicted values of 3087 and 3055 cm^{-1} respectively.

The next two vibrations (ν_5, ν_6) are assigned to C=O stretches and are of particular interest in view of their potential value to probe hydrogen-bonding. We predict substantial mixing of C(2)–O and C(4)–O vibrations in these modes with C(2)–O and C(4)–O being dominant in ν_5 and ν_6 respectively. These conclusions agree with those of Graindourze *et al.*[5] based on ^{18}O substitution and hydrogen bonding studies. However, two areas of discrepancy between the theoretical and experimental description of ν_5 and ν_6 should be noted. Firstly, ν_6 is predicted to be substantially more intense than ν_5. Although intensity measurements of ν_5 to ν_6 are difficult, the ratio is larger than one. Secondly, there are substantial differences between the calculated and observed frequencies for these absorptions, particularly for ν_6 (89 cm^{-1}), a value to be compared to the mean square deviation of the in-plane modes of 27 cm^{-1} (15 cm^{-1} without ν_5 and ν_6). In the case of formaldehyde, a calculation at the 6-31G** level predicts a C–O stretching frequency of 1823 cm^{-1}, compared with an experimental[12] (gas phase) value of 1744 cm^{-1}, giving an error similar to that found here. A possible resolution of the discrepancy between the calculated and observed frequency of ν_6, is to assign the absorption[6] at 1741 cm^{-1} to ν_6 as suggested by Harsanyi *et al.*[8] thus reducing this difference to 54 cm^{-1} and giving an observed ν_5–ν_6 splitting of 23 cm^{-1} compared to the predicted value of 16 cm^{-1}.

It is generally assumed that the noble gas environment interacts only very weakly with the matrix isolated molecules so that it is usual to discuss the spectra of the latter in terms of free molecule calculations. However, we note in passing that examination of the crystal structure of uracil[13] reveals that the oxygen atom of C(4)–O is involved in *two* intermolecular hydrogen bonds with neighbouring nitrogen atoms, which are absent for the oxygen of C(2)–O. We would thus expect that ν_6 would be more greatly perturbed than ν_5 in the solid state.

The third vibration between 1800 and 1600 cm^{-1} (ν_7) is predicted (1663 cm^{-1}) close to the experimental value (1643 cm^{-1}). It is predominantly a C=C stretching mode. However, we do not find the substantial C(4)–O stretching contribution suggested by Graindourze *et al.*[5]

The positions of the next three bands (ν_8–ν_{10}) are accurately predicted by our calculations, and all are predicted to have a major NH bending contribution in line with experiment. The prediction of the frequencies of the remaining bands (ν_{11}–ν_{14}) above 1000 cm^{-1} is in excellent agreement with experiment, the only discrepancy between theory and experiment being a lack of N(3)H bending character in ν_{13} suggested by Graindourze *et al.*[5]

In the low frequency region, the present calculation avoids the substantial effects of scaling of the out-of-plane frequencies necessitated by smaller basis set calculations,[8] and thus hopefully we can propose a more definite assignment of these modes. The predicted position (814 cm^{-1}) of the first out-of-plane mode having significant intensity (ν_{23}) is closer to the experimental value (804 cm^{-1}) than the value given previously by scaled calculations (832 cm^{-1}), and our predicted character, [γC(4)O], is in line with experiment.[5] We also predict that ν_{24} and ν_{25} arise from C–O out-of-plane deformations, [C(2)O and C(4)O respectively], although experiment does not distinguish between these two motions. Again, the frequencies of these two modes are accurately predicted by our calculation. The frequencies of the remaining in-plane modes (ν_{15}–ν_{21}) are all accurately predicted by our calculation, and their description is in accord with previous suggestions.[5,8] The remaining out-of-plane motions assigned to out-of-plane deformations (ν_{24}–ν_{27}) are also accurately predicted as far as frequencies are concerned and our PED gives a description of these modes in agreement with that of previous workers.[8] The lowest energy out-of-plane motions (ν_{28}–ν_{30}) are assigned to ring torsions, and where comparison with experiment is possible, good agreement is found.

In Table 1 we also show our calculated frequencies for *N,N′*-dideuteriouracil, together with the experimental values from Szczesniak *et al.*[7] It should be noted that the PED in this table is for the unsubstituted species. Thus, where there is substantial mixing of N–H with other motions (ν_8–ν_{16}) a direct correlation between the modes in the substituted and unsubstituted species is not possible. We find a generally satisfactory agreement, between the observed and calculated frequencies although we are unable to assign two weak absorptions reported[7] at 829 and 768 cm^{-1}.

Conclusions

The IR spectrum of the dioxo tautomer of uracil has been reproduced with good accuracy using a 6-31G** basis. A uniform scale factor of 0.9 is appropriate for both the in-plane and out-of-plane modes, and avoids the different scale factors for these two motions needed when a 4-21 basis is employed. Our mean square deviations between calculated and observed frequencies, 27 and 17 cm^{-1} for the in-plane and out-of-plane modes respectively are comparable with those for the 4-21 basis, using a number of optimised scaling factors. The value of a potential energy decomposition giving a more detailed description of the vibrational modes has been illustrated. Remaining problems associated with the assignment of the C=O stretching modes are highlighted. These may be due at least in part, to the problem of interpreting the spectra of matrix isolated species in terms of free molecule calculations.

Acknowledgements

We thank SERC for support of this research under grant No. GR/E26921.

References

1 P. O. Löwdin, *Rev. Mod. Phys.*, 1963, **35**, 724; J. K. Landquist, *Applications as Pharmaceuticals*, in *Comprehensive Heterocyclic Chemistry*, ed. A. R. Katritzky and C. W. Rees, Pergamon Press, Oxford, 1984, p. 159.
2 L. Lapinski, M. J. Nowak, J. Fulara, A. Lés and L. Adamowicz, *J. Phys. Chem.*, 1990, **94**, 6555.
3 M. J. Nowak, L. Lapinski and J. S. Kwiatkowski, *Chem. Phys. Lett.*, 1989, **157**, 14.
4 K. Szczepaniak, M. Szczesniak and W. B. Person, *Chem. Phys. Lett.*, 1988, **153**, 39.
5 M. Graindourze, J. Smets, Th. Zeegers-Huyskens and G. Maes, *J. Mol. Struct.*, 1990, **222**, 345.
6 A. J. Barnes, M. A. Stuckey and L. LeGall, *Spectrochim. Acta, Part A*, 1984, **40**, 419.
7 M. Szczesniak, M. J. Nowak, H. Rostkowska, K. Szczepaniak, W. B. Person and D. Shugar, *J. Am. Chem. Soc.*, 1983, **105**, 5969.
8 L. Harsanyi, P. Csaszar, A. Csaszar and J. E. Boggs, *Int. J. Quantum Chem.*, 1986, **29**, 799.
9 I. R. Gould and I. H. Hillier, *J. Chem. Soc., Perkin Trans. 2*, 1990, 329.
10 R. D. Amos and J. E. Rice, 'CADPAC: The Cambridge Analytic Derivatives Package,' Issue 4.0, Cambridge, 1987.
11 G. Keresztury and G. Jalsovszky, *J. Mol. Struct.*, 1971, **10**, 304.
12 H. H. Nielsen, *Phys. Rev.*, 1934, **46**, 117.
13 R. F. Stewart and L. H. Jensen, *Acta Crystallogr.*, 1967, **23**, 1102.

Paper 1/04791D
Received 16*th September* 1991
Accepted 15*th October* 1991