

Vibrational spectra of nucleic acid constituents—II. Planar vibrations of cytosine

H. SUSI, J. S. ARD and J. M. PURCELL

Eastern Regional Research Laboratory,* Philadelphia, Pennsylvania 19118

Abstract—Laser-Raman spectra and i.r. spectra of polycrystalline cytosine and cytosine- d_3 have been obtained. In-plane fundamentals have been assigned to absorption bands and Raman lines on the basis of deuteration shifts, comparison with spectra of four isotopic analogs of uracil, and a normal coordinate calculation. A valence force field with thirteen transferred and sixteen refined force constants was employed to calculate a total of forty-three frequencies, to obtain a potential energy distribution and to calculate Cartesian displacements. Several skeletal vibrations appear to be characteristic for the pyrimidine bases of nucleic acids in general.

INTRODUCTION

THE VIBRATIONAL spectra of nucleic acids are extremely difficult to interpret because of the high complexity and low symmetry of these molecules [1]. In an effort to gain additional insight, the i.r. [2, 3] and Raman [4] spectra of various constituents have been investigated in some detail and qualitative group-frequency assignments have been proposed for some characteristic absorption bands [2, 3] and Raman lines [4]. Recently, the i.r. and laser-Raman spectra of uracil and three deuterated analogs were investigated, and assignments based on deuteration shifts and a normal coordinate analyses were proposed for all in-plane fundamentals [5]. The present communication describes the i.r. and laser-Raman spectra of cytosine and cytosine- d_3 . The data are interpreted on the basis of deuteration shifts, a comparison with four isotopic analogs of uracil [5], and a normal coordinate analyses of the in-plane modes. The fundamentals are characterised by the potential energy distribution and Cartesian displacements.

STRUCTURAL CONSIDERATIONS

Cytosine crystals are orthorhombic with the space group $V^1-P2_12_12_1$ [6]. There are four planar molecules (point group C_s) per unit cell. The molecules are arranged into sheets which form a 27.5° angle with the ab plane. Within each sheet there is a net of hydrogen bonds involving the NH_2 , NH , $=N-$ and CO groups of each molecule. In principle, the 23 a' and 10 a'' molecular modes split into i.r. and Raman active B_1 , B_2 , B_3 branches, and a Raman active A branch. The splitting between the branches is expected to be small [4, 5], except for modes associated with groups directly involved in hydrogen bonding. Because of the i.r.

* Eastern Utilization Research and Development Division, Agricultural Research Service, U.S. Department of Agriculture.

[1] G. B. B. M. SUTHERLAND and M. TSUBOI, *Proc. Roy. Soc. (London)* **A239**, 446 (1957).

[2] C. L. ANGELL, *J. Chem. Soc.* 504 (1961).

[3] T. SHIMANOCHI, M. TSUBOI and Y. KYOGOKU, *Advances in Chemical Physics*, Vol. 7, p. 435. Interscience (1964).

[4] R. C. LORD and G. J. THOMAS, JR., *Spectrochim. Acta* **23A**, 2551 (1967).

[5] H. SUSI and J. S. ARD, *Spectrochim. Acta* **27A**, 1549 (1971).

[6] D. L. BARKER and R. E. MARSH, *Acta Cryst.* **17**, 1581 (1964).

inactivity of the A branches, and differences between the relative intensities of the B_1 , B_2 , B_3 branches in i.r. and the Raman effect, the observed apparent frequencies of the mostly unresolved i.r. bands and Raman lines do not agree precisely. In this communication the molecular vibrations are analyzed in terms of C_s symmetry and the modes are classified as in-plane and out-of-plane.

An interesting point regarding molecular structure concerns the bond lengths of the CN bonds. According to two X-ray studies [6, 7] the C—NH₂ bond (1.330Å) is shorter than the CN 'double bond' in the ring (1.337Å), and considerably shorter than the corresponding bonds in aromatic amines (~1.42Å) [8]. The C—NH₂ stretching frequency is therefore expected to be markedly higher than in aromatic amines [9, 10, 11]. The CN 'double bond' in the ring, on the other hand, is not much shorter than neighboring single bonds and is expected to participate in mixed ring-stretching modes, i.e. no separable C=N vibration is expected in the double bond region. The C—NH₂ bond has been described as having *ca.* 30% double bond character and the ring CN 'double bond' *ca.* 60% single bond character [6, 12].

EXPERIMENTAL

The Raman spectra were obtained with a modified Perkin-Elmer Model LR-1 spectrometer.* The instrument was equipped with broad band optics and a Carson Laboratories Argon-Ion laser. The 5145Å line was used for excitation and the laser was operated at *ca.* 400 mW power. The polycrystalline samples were cooled to liquid nitrogen temperature. Cooling resulted in a marked sharpening of the observed lines and improved apparent resolution. It also permitted the measurement of Raman spectra down to *ca.* 50 cm⁻¹ with our relatively simple instrument. Survey spectra at room temperature were also obtained with a modified Spex Ramalog System. The agreement between the two sets of spectra was good, indicating that the modified LR-1 instrument had been properly calibrated and that the observed lines, particularly the strong lines at low frequencies, were real. Infrared spectra were obtained with a Beckman Model IR-7 spectrometer. Mineral oil (Nujol) mulls were used for most of the spectral region, halocarbon mulls for regions where the mineral oil absorbs.

Numerical calculations were carried out with previously used programs [5] on a Univac 1107 computer (Bluebell, Pa.) and an in-house IBM 1130 computer.

The cytosine sample was obtained from the Cyclo Chemical Corporation and used after recrystallization from water. Cytosine-*d*₃ was prepared from cytosine by repeated recrystallization from D₂O.

* Reference to brand or firm name does not constitute endorsement by the U.S. Department of Agriculture over others of a similar nature not mentioned.

- [7] G. A. JEFFREY and Y. KINOSHITA, *Acta Cryst.* **16**, 20 (1963).
[8] H. J. M. BOWEN, J. DONOHUE, D. G. JENKIN, O. KENNARD, P. J. WHEATLEY and D. H. WHIFFEN, *Chem. Soc. (London)*, Spec. Publ. No. 11, pp. 377 (1958).
[9] L. J. BELLAMY, *The Infra-red Spectra of Complex Molecules*, 2nd Edition, Wiley, New York (1958).
[10] Y. KYOGOKU, S. HIGUCHI and M. TSUBOI, *Spectrochim. Acta* **23A**, 969 (1967).
[11] J. C. EVANS, *Spectrochim. Acta* **16**, 428 (1960).
[12] R. E. MARSH, R. BIERSTEDT and E. L. EICHORN, *Acta Cryst.* **15**, 310 (1962).

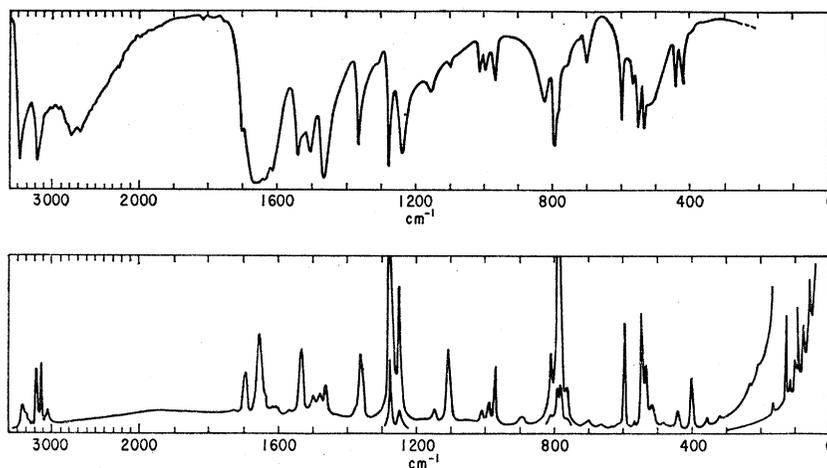


Fig. 1. Infrared and Raman spectra of polycrystalline cytosine. (Raman spectrum below 300 cm^{-1} at $1/4$ intensity.)

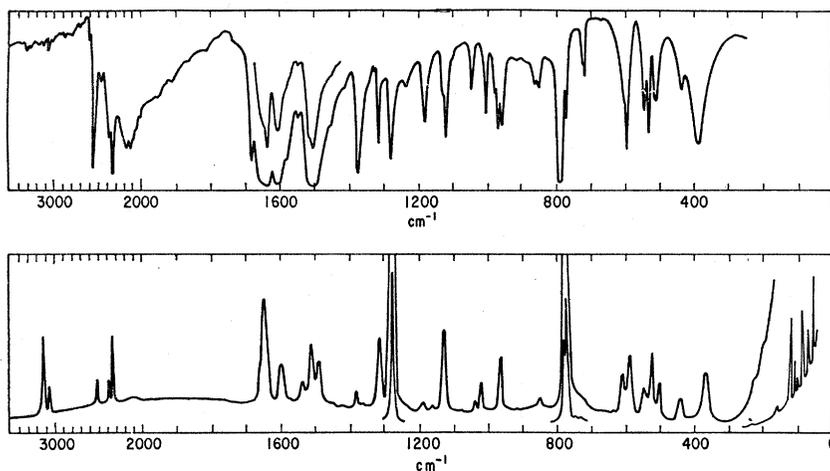


Fig. 2. Infrared and Raman spectra of polycrystalline cytosine- d_3 . (Raman spectrum below 300 cm^{-1} at $1/4$ intensity.)

SPECTRA AND ASSIGNMENTS

The Raman and i.r. spectra of cytosine and cytosine- d_3 are presented in Figs. 1 and 2. Observed and calculated frequencies (see below) of the planar modes are given in Tables 1 and 2. Raman lines and i.r. bands which are not assigned to planar fundamentals are listed in Table 3. A qualitative discussion of the observed spectra follows.

$3500\text{--}2000\text{ cm}^{-1}$. CH, NH, ND *stretching modes*

The high frequency stretching modes are assigned by comparison of the i.r. and Raman spectra of the two homologues. The two CH stretching modes are not resolved. Differentiation between ring NH stretching and —NH_2 stretching

Table 1. Observed and calculated in-plane fundamentals of cytosine (cm⁻¹)

	I.r.	Raman*	Calc.	Assignment†	Designation
1	3380 s	3354(0.3)	3356	ν NH ₂ out-of-phase (100)	
2	—	3230(0)	3240	ν NH ₂ in-phase (99)	
3	3169 s	3176(0.7)	3190	ν NH (99)	
4	—	3117(0.9)	3121	ν CH in-phase (99)	
5	—	3117(0.9)	3114	ν CH out-of-phase (99)	
6	1703 w	1694(0.5)	1689	δ NH ₂ (90)	
7	1662 vs	1653(1.0)	1650	ν C=O (68)	C I
8	1615 s	1612(0.1)	1616	ν C=C (61)	C II
9	1538 m	1533(0.8)	1530	δ NH (50), ν -ring (28)	
10	1505 m	1498(0.2)	1496	ν -ring (56), δ NH (14)	C III
11	1465 s	1462(0.4)	1450	δ CH (40), ν C—N (20)	C IV
12	1364 m	1361(0.8)	1344	ν C—N (38), δ CH (43)	δ CH
13	1277 m	1276(9.0)	1295	δ CH out-of-phase (72)	
14	1236 m	1247(3.0)	1222	ν -ring (98)	C V
15	1100 w	1108(0.9)	1126	ν -ring (52), ρ NH ₂ (28)	C VI
16	1010 w	1011(0.1)	1036	ν -ring (72)	C VII
17	994 w	990(0.2)	1007	ν , δ -ring (81)	C VIII
18	966 w	971(0.7)	955	ν -ring (57), ρ NH ₂ (46)	
19	793 m	792(10.0)	790	ν , δ -ring (52)	C IX
20	600 m	597(1.2)	592	δ -ring (58)	C X
21	549 m	546(1.4)	560	δ CO, δ C—N in-phase (67)	C XI
22	533 m	533(0.8)	533	δ -ring (54)	C XII
23	—	400(0.7)	386	δ CO, δ C—N out-of-phase (70)	C XIII

* Intensity of ν_{19} arbitrarily equal to 10.† ν —stretching mode; δ —bending mode; ρ —rocking mode.Table 2. Observed and calculated in-plane fundamentals of cytosine-*d*₃ (cm⁻¹)

	I.r.	Raman*	Calc.	Assignment†	Designation
1	3115 w	3115(1.0)	3121	ν CH in-phase (99)	
2	3115 w	3115(1.0)	3114	ν CH out-of-phase (99)	
3	2545 s	2511(0.3)	2500	ν ND ₂ out-of-phase (78)	
4	2376 sh	2376(0.3)	2351	ν ND ₂ in-phase (97)	
5	2337 s	2338(0.9)	2346	ν ND(97)	
6	1637 vs	1648(1.3)	1646	ν C=O (74)	C I
7	1605 s	1598(0.5)	1609	ν C=C (58)	C II
8	1517 sh	1514(0.8)	1501	ν -ring (78)	C III
9	1505 vs	1491(0.5)	1477	ν C—N (42), δ CH (33)	C IV
10	1377 s	1383(0.2)	1372	δ CH (53), ν C—N (18)	δ CH
11	1316 m	1316(0.9)	1328	δ CH out-of-phase (52)	
12	1281 s	1282(8.0)	1270	ν -ring (68)	C V
13	1182 m	1190(0.1)	1199	δ ND ₂ (57)	
14	1120 s	1130(1.0)	1165	ν -ring (90)	C VI
15	1046 m	1038(0.1)	1025	ν , δ -ring (67)	
16	1005 m	1021(0.4)	987	ν -ring (67)	
17	970 m	965(0.7)	967	δ ND (43), ν -ring (32)	
18	788 s	783(4.0)	794	δ , ν -ring (40), ρ ND ₂ (34)	
19	771 m	777(10.0)	759	δ , ν -ring (43), ρ ND ₂ (34)	
20	596 s	590(0.9)	584	δ -ring (62)	C X
21	531 m	524(0.9)	525	δ CO, δ C—N (48), δ -ring (26)	
22	511 m	504(0.5)	516	δ CO, δ C—N (39), δ -ring (28)	
23	—	369(0.7)	363	δ CO, δ C—N out-of-phase (68)	C XIII

*, † As in Table 1.

modes is accomplished by comparison with uracil spectra. The missing i.r. bands of cytosine (see Table 1) must be very weak and appear as barely detectable shoulders on the two strong observed bands.

Table 3. Out-of-plane modes and unassigned bands and Raman lines*

Cytosine			Cytosine- d_3		
R	I.r.		R	I.r.	
894(0.1)	—			863 w	
811(0.8)	823 m	γ NH	849(0.1)	850 w	
782(3.0)	782 m			724 w	
764(0.4)	760 sh			720 w	
704(0.1)	701 w		611(0.4)	605 sh	γ ND
568(0.1)	566 w	sk	550(0.3)	545 m	sk
517(0.2)	520 sh	τ NH ₂			
442(0.2)	442 w	sk	438(0.3)	437	sk
—	421 w			388	τ ND ₂
356(0.4)					
162(0.5)			162(0.5)		
123(5.0)			122(5.0)		
112(1.0)			110(2.0)		
100(2.1)			100(1.7)		
90(4.5)			91(4.1)		
74(1.8)			72(2.5)		
56(3.0)			57(4.3)		

* γ —out-of-plane bending. τ —out-of-plane wagging. sk—skeletal.

1700–1550 cm^{-1} . Double bond stretching and NH₂ bending modes

The Raman spectrum of cytosine shows lines of medium intensity at 1694 and 1653 cm^{-1} , and a weak line at 1612 cm^{-1} . Cytosine- d_3 exhibits medium intensity lines at 1648 and 1598 cm^{-1} . The three cytosine lines are assigned to NH₂ bending, C=O stretching and C=C stretching, the two lines of the deuterated sample to C=O and C=C stretching modes. The assignment is in agreement with calculations reported further below, and with the notion (cf. STRUCTURAL CONSIDERATIONS) that no N=C stretching mode is expected in this spectral region. The i.r. spectra in this region are complex and exhibit overtones and/or combinations in possible Fermi resonance with the fundamentals. They neither contradict nor positively support the proposed assignment.

1550–1200 cm^{-1} . Skeletal modes. CH, NH in plane deformations

Two CH bending, one NH bending, and two ring stretching modes [5] as well as the C—NH₂ stretching mode [2, 9] are expected in this region. In the deuterated sample the ND mode moves to a lower frequency. A number of Raman lines and absorption bands appear in this region of the observed spectra, but no simple group-frequency assignment seems possible. The two very strong Raman lines, and corresponding i.r. bands, in the 1200–1350 cm^{-1} range are probably associated with a ring stretching mode [4, 5] and CH out-of-phase deformation [5, 9]. The remaining four modes of cytosine must involve NH bending, CH bending, skeletal stretching and C—NH₂ stretching. None of the frequencies corresponds to an established group frequency, in particular to the CH in-phase bending frequency (\sim 1400 cm^{-1}) [5, 9]. N-deuteration affects all bands in this region and suggests mixed vibrations.

1200–350 cm^{-1} . Skeletal modes. NH₂, ND, ND₂ rocking and bending modes

A comparison of the spectra of cytosine and uracil [5] leads to an assignment of the eight in-plane skeletal modes expected in this region. The frequencies of

the skeletal modes of uracil and cytosine appear to be surprisingly close. A cytosine line at 971 cm^{-1} and i.r. band at 966 cm^{-1} which have no counterpart in uracil spectra are assigned to a fundamental involving NH_2 rocking. In the deuterated sample ND_2 and ND deformation modes fall into the range where skeletal stretching and stretch-bend modes appear. The result is a pattern of highly mixed in-plane modes in the discussed frequency range. The calculated and observed frequencies are in reasonable agreement; skeletal modes which retain their essential characteristics (see below) are indicated in the last column of Table 2.

Unassigned i.r. bands and Raman lines

The vibrational spectra of cytosine should exhibit ten out-of-plane modes of which five could be classified as skeletal modes, four as out-of-plane CH or NH bending, and one as NH_2 torsion. Table 3 lists i.r. bands and Raman lines which are not assigned to planar modes. Unfortunately no concise assignment appears possible (cf. Ref. [5]). Near 440 and 550 cm^{-1} absorption bands and Raman lines are observed in both analogs which are very close to skeletal out-of-plane modes of other pyrimidine bases [5, 10] and appear to be characteristic for the ring. NH and ND bending modes, and NH_2 and ND_2 wagging modes are tentatively assigned as indicated in Table 3. The latter appear at lower frequencies than in aromatic amines [11], probably because of different hybridization within the NH_2 group (the C— NH_2 grouping in aromatic amines is non-planar [11, 13]). Below 170 cm^{-1} a series of intense Raman lines is observed. The higher ones could be associated with molecular out-of-plane modes, the lower ones must be caused by lattice modes [14]. The listed low frequency Raman lines were observed with two different instruments, equipped with different lasers, and are believed to be real.

FORCE CONSTANTS AND CALCULATED FREQUENCIES

A valence force field (VFF) analogous to the one used for uracil [5] was constructed in the following way: (a) A force constant was assigned to each chemical bond and each bond angle. (b) Bending force constants on either side of outer atoms and on either side of the NH_2 group were assumed to be equal. (c) In general, only stretch-stretch interactions with one common atom and stretch-bend interactions with two common atoms were considered. (These are the VFF interactions resulting from a Urey-Bradley force field with *gem* interactions [15].) (d) The following exceptions were made for interaction constants:

- (i) *meta* and *para* ring stretch-stretch interactions were added [5];
- (ii) interactions involving CH and NH stretching coordinates could not be determined and were neglected [5];
- (iii) a CH bending, C— NH_2 stretching interaction was added to obtain a reasonable frequency agreement in the $1200\text{--}1500\text{ cm}^{-1}$ range (cf. SPECTRA AND ASSIGNMENTS).

The numbering of diagonal force constants is given in Fig. 3. The *G*-matrix was based on the structure given by BARKER and MARSH [6]. Refinements were

[13] M. TSUBOI, *Spectrochim. Acta* **16**, 505 (1960).

[14] I. HARADA and R. C. LORD, *Spectrochim. Acta* **26A**, 2305 (1970).

[15] J. OVEREND and J. R. SCHERER, *J. Chem. Phys.* **32**, 1289 (1960).

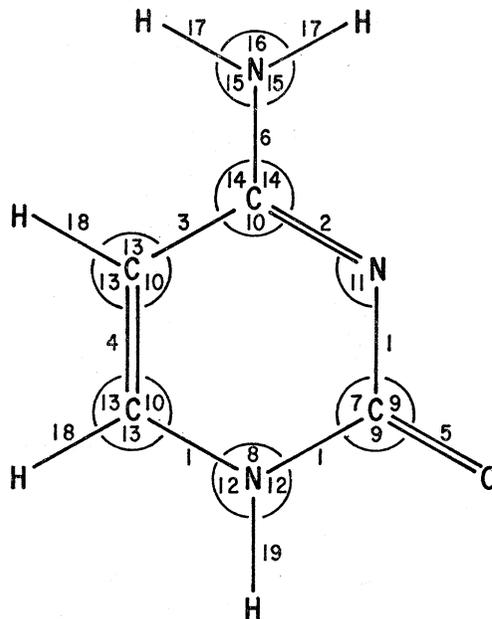


Fig. 3. Numbering of diagonal force constants.

carried out by the usual iterative methods [15, 16]. Initial zero order calculations were performed with the C=O, C=N and C—NH₂ stretching force constants calculated from bond lengths by previously described procedures [5] and the rest transferred from uracil [5]. The frequency agreement as well as the potential energy distribution was quite reasonable except for the NH₂ bending mode (which was too high) and the (mixed) fundamentals in the 1200–1550 cm⁻¹ range. For a least squares refinement [15, 16], the following scheme was adopted. All skeletal stretching force constants and all force constants involving the C=O group were transferred from uracil or calculated from bond lengths [5] as indicated in Table 4, and were not refined. The CNC bending constant $F(8)$ and the *meta* interaction constant could not be determined and were also transferred from uracil. The remaining force constants (involving primarily CH, NH and NH₂ modes as well as skeletal interactions) were refined.

Calculated frequencies are listed in the fourth column of Tables 1 and 2. Figure 4 gives an overall view of the frequency agreement in the critical 300–1700 cm⁻¹ range. Noticeable differences between observed and calculated frequencies occur in the 900–1400 cm⁻¹ region where skeletal stretching and stretch-bend modes predominate. Considering the simple nature of the employed force field and the complexity of the molecules, such discrepancies are to be expected. The agreement appears sufficient to support the assignments, and to estimate the nature of skeletal modes by potential energy distribution and Cartesian displacements. The overall average frequency error was *ca.* 10 cm⁻¹ (~1.0%). The force constants are listed in Table 4.

[16] J. H. SCHACHTSCHNEIDER and R. G. SNYDER, *Spectrochim. Acta* **19**, 117 (1963).

Table 4. Valence force constants

Description*	Value	Dispersion	Description	Value	Dispersion
$F(1)$	6.380†		$F(16)$	0.394	0.018
$F(2)$	7.340‡		$F(17)$	6.002	0.043
$F(3)$	6.202†		$F(18)$	5.280	0.029
$F(4)$	8.702†		$F(19)$	5.604	0.049
$F(5)$	10.550‡		$f(\nu_{C=O}, \nu_{skel})$	1.397†	
$F(6)$	7.570‡		$f(\nu_{C=O}, \delta_{C=O})$	0.857†	
$F(7)$	1.620†		$f(\nu_{skel}, \delta_{C=O})$	0.304†	
$F(8)$	1.166†		$f(\nu, \nu \text{ meta})$	-0.204†	
$F(9)$	1.034†		$f(\nu, \nu \text{ para})$	0.880	0.108
$F(10)$	0.620	0.099	$f(\nu, \nu \text{ skel})$	0.851	0.069
$F(11)$	1.890	0.374	$f(\delta, \nu \text{ skel})$	0.599	0.065
$F(12)$	0.457	0.020	$f(\delta_{NH}, \nu_{skel})$	0.193	0.041
$F(13)$	0.369	0.008	$f(\delta_{CH}, \nu_{skel})$	0.316	0.037
$F(14)$	1.412	0.136	$f(\nu_{N-C}, \delta_{CH})$	0.058	0.015
$F(15)$	0.413	0.027			

Stretching: m dyn/Å; bending: m dyn Å/rad²; stretch bend interactions: m dyn/rad.

* See Fig. 3.

† Transferred from uracil [5].

‡ Calculated from bond length.

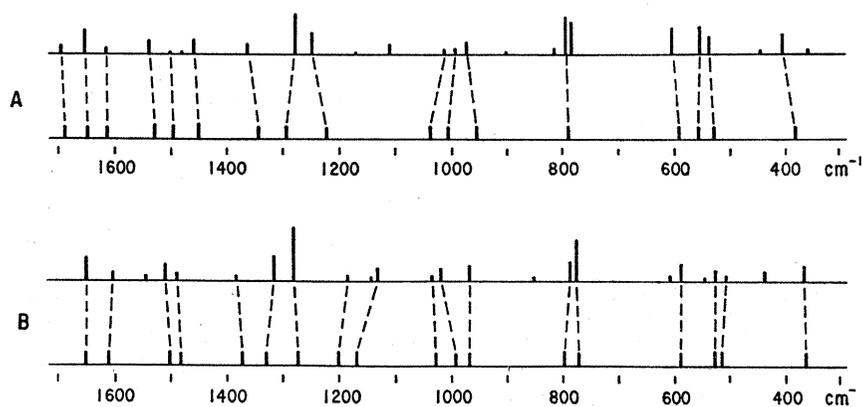


Fig. 4. Observed Raman frequencies and calculated frequencies of A, cytosine and B, cytosine- d_3 , 300–1700 cm^{-1} .

POTENTIAL ENERGY DISTRIBUTION AND CARTESIAN DISPLACEMENTS. CHARACTERISTIC CYTOSINE MODES

The predominant term(s) in the potential energy distribution (PED) [17] of each vibrational mode are given in parentheses in the fourth column of Tables 1 and 2. For modes which necessarily involve more than one internal displacement coordinate, the value represents the sum of the fractions of potential energy associated with the primarily involved force constants, i.e. for ring stretching modes, the sum over all ring stretching terms; for ring bending modes the sum over all bending terms.

Table 1 reveals that for cytosine a surprising number of in-plane modes can be classified as 'group vibrations' by PED criteria. This holds for all modes above

[17] Y. MORINO and K. KUCHITSU, *J. Chem. Phys.* **20**, 1809 (1952).

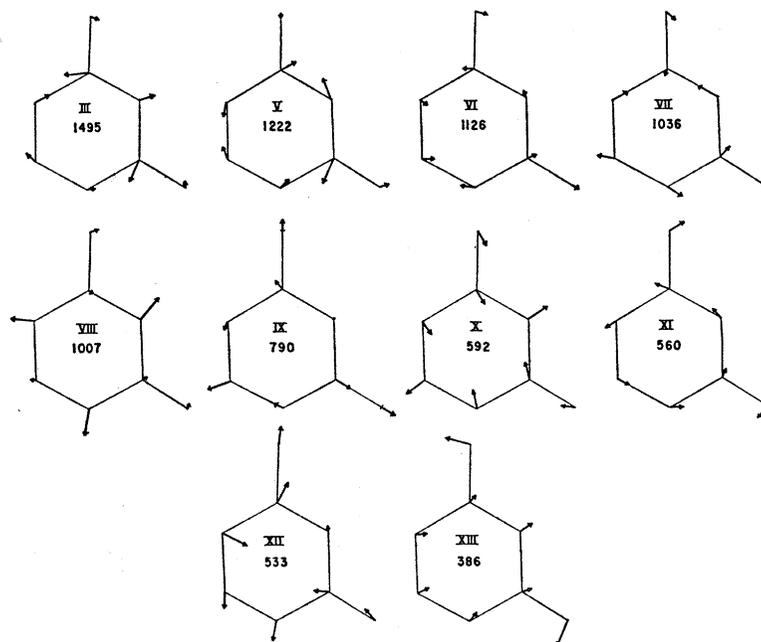


Fig. 5. Calculated Cartesian displacements for skeletal modes of cytosine. (Hydrogen atoms are not shown; see text.) Position of ring as in Fig. 3.

1600 cm^{-1} and for a number of skeletal modes at lower frequencies. In the 1300–1600 cm^{-1} region excessive mixing takes place (cf. SPECTRA AND ASSIGNMENTS). The modes which can be classified as essentially skeletal vibrations are designated C I to C XIII, in analogy with corresponding modes of uracil [5], in order to facilitate comparison with characteristic vibrations of related molecules.

Figure 5 shows the calculated Cartesian displacements for the C III, and C V to C XIII modes (C I, C II and C IV are essentially associated with C=O, C=C and C—NH₂ stretching). The displacements are drawn to the same scale as the equilibrium geometry and represent four times the unit changes in the normal coordinates. The hydrogen atoms are not shown. (Exaggerating the displacements of heavy atoms, and omitting the hydrogen atoms seems to be the only way to depict the movements of the heavy atoms in small-scale drawings; cf. Ref. [5].) A comparison of Fig. 5 with corresponding data of uracil [5] reveals that the C V to C XIII modes are quite similar to the U V to U XIII modes of uracil. These vibrations thus appear to be characteristic for the pyrimidine bases of nucleic acids in general, by frequency criteria, PED criteria and Cartesian displacements criteria.

In the spectra of cytosine-*d*₃, ND and ND₂ vibrations interact with and distort several skeletal modes below *ca.* 1300 cm^{-1} . The ones which retain their essential characteristics are indicated in the last column of Table 2.

Acknowledgements—The authors are grateful to Dr. JAMES R. SCHERER of the Western Regional Research Laboratory (USDA) for supplying supplementary Raman data.