100. *An Infrared Spectroscopic Investigation of Nucleic Acid Constituents.*

By C. L. Angell.

The infrared spectra of cytosine, uracil, adenine, and guanine, of ribose and deoxyribose nucleosides derived from these bases, and of some mononucleotides have been examined in the solid state between 4000 and 650 cm.$^{-1}$. Many of the prominent absorption bands have been assigned to vibrations of structural groups with the help of spectral data from a number of derivatives, including those in which the labile hydrogen atoms have been replaced by deuterium. The infrared evidence confirms the view that in the solid state cytosine and guanine exist in the keto-amino-form, adenine in the amino-, and uracil in the diketo-form. Evidence is presented that the cytidine and adenosine phosphates exist in zwitterionic forms in the crystalline state.

Great interest in the nucleic acids and their constituent molecules—heterocyclic nitrogenous bases, nucleosides and nucleotides—has led to a number of structural investigations by physical methods; infrared spectroscopy has played a prominent rôle in this work. Such spectra have been published previously for derivatives of the pyrimidine [1] and purine bases,[2] and for nucleosides,[3] nucleotides,[4] and nucleic acids [5] in the solid state. In a number of cases infrared methods have also been used to study these compounds in solution in heavy water.[6]

In the present work spectra of substituted derivatives of the relevant bases (where often only one tautomeric form is possible), of all the ribose and deoxyribose nucleosides, and of a number of nucleotides were obtained. In addition, the labile hydrogen atoms of the four bases and of the corresponding ribose nucleosides were exchanged with deuterium and the spectra of these " deuterated " compounds measured.

The spectra were studied with the following principal aims: (1) to confirm the tautomeric structures of the bases; (2) to make assignments of absorption bands to vibrations of structural groupings where this is possible; (3) to note any spectral changes in the series, base ⟶ nucleoside ⟶ nucleotide; and (4) to differentiate if possible between (*a*) ribose and deoxyribose derivatives and (*b*) the 2'- and 3'-phosphates of nucleosides.

Experimental

Materials.—The majority of compounds investigated were obtained from members of the Chemical Laboratory, Cambridge University, or from Professor Sir Alexander Todd's personal collection of chemicals. Others were obtained from the following sources: isocytosine (Dr. R. Markham, Molteno Institute, Cambridge); 1,3-dimethyluracil (Genatosan Ltd., Loughborough); purine, 9-methylpurine, and 2-hydroxypyrimidine (Professor A. Albert, Australian National University); and adenine hydrochloride (L. Light and Company Ltd., Colnbrook, Bucks.).

Cytosine, adenosine, and guanosine crystallise with water of crystallisation; anhydrous samples were obtained by drying them under a vacuum at about 110° for several hours. The hydrochlorides of the bases were obtained as cytosine,HCl,H_2O, adenine,HCl,$\frac{1}{2}H_2O$, and guanine,HCl,H_2O, and were used as such.

The technique used for the deuterium exchange of labile hydrogen atoms was as follows: about 10 mg. of a substance were gently refluxed with 0·4 c.c. of heavy water for 0·5 hr.; the

[1] Short and Thompson, *J.*, 1952, 168; Brown and Short, *J.*, 1953, 331.
[2] Willits, Decius, Dille, and Christensen, *J. Amer. Chem. Soc.*, 1955, **77**, 2569; Lacher, Bitner, Emery, Seffl, and Park, *J. Phys. Chem.*, 1955, **59**, 615.
[3] Blout and Fields, *J. Amer. Chem. Soc.*, 1950, **72**, 479; Fraser, Ph.D. Thesis, London, 1952.
[4] Harris, Orr, Roe, and Thomas, *J.*, 1953, 489; Blout and Fields, *J. Biol. Chem.*, 1949, **178**, 335.
[5] Fraser and Fraser, *Nature*, 1951, **167**, 759; Sutherland, *Rend. Ist. Lombardo Sci. Lettere, Classe Sci.*, 1955, **89**, 67; Sutherland and Tsuboi, *Proc. Roy. Soc.*, 1957, *A*, **239**, 446.
[6] Miles, *Biochim. Biophys. Acta*, 1956, **22**, 247; 1958, **27**, 46; Lenormant and Blout, *Compt. rend.*, 1954, **239**, 1281; Blout and Lenormant, *Biochim. Biophys. Acta*, 1954, **15**, 303; Lenormant and deLozé, *Bull. Soc. chim. France*, 1955, 1501, 1504; Sinsheimer, Nutter, and Hopkins, *Biochim. Biophys. Acta*, 1955, **18**, 13.

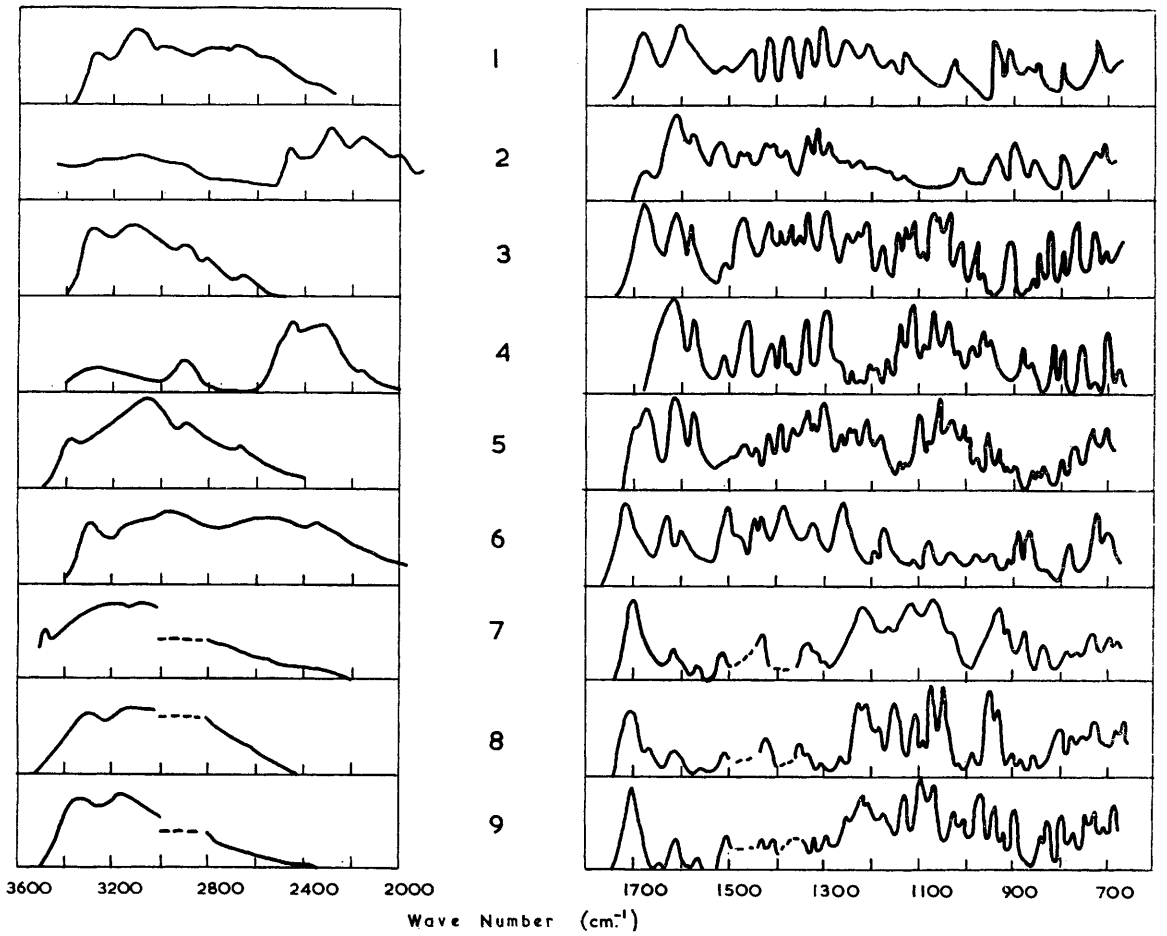
solution then was evaporated to dryness and the residue used to make up a mull in Nujol by the usual technique. All these operations were carried out in a dry-box. The heavy water used was a Norsk Hydro-Electrisk (Norway) product and contained 99·70% of D_2O.

TABLE 1. *List of additional spectra studied.*

D-Ribose	9-Methylpurine	Thymidine diphosphate Ba salt
Deoxy-1-ribose	9-Methyladenine	Thymidine-3′ phosphate Ba salt
2-Hydroxypyrimidine	Guanine hydrochloride	Thymidine-5′ phosphate Ba salt
6-Methoxy-3-methyl-2-pyrimidone	Deoxy-5-methylcytidine hydrochloride	5′-Acetyladenosine
6-Dimethylamino-3-methyl-2-pyrimidone	Adenine picrate	5′-Acetyldeoxyadenosine
3-Methylcytosine	Adenosine picrate	3′-Acetyldeoxyadenosine
1,3-Dimethylcytosine	Deoxyadenosine picrate	3′,5′-Diacetyladenosine
Cytosine-*N* benzyl phosphate	Cytidine-2′ benzyl phosphate	3′,5′-Diacetyldeoxyadenosine
Isocytosine	Cytidine-3′ benzyl phosphate	2′,3′,5′-Triacetyladenosine
4-Methyluracil	Cytidine-3′ methyl phosphate	3′-Acetylthymidine
3-Methyluracil	Adenosine-5′ benzyl phosphate	5′-Acetylthymidine
Purine	Adenosine diphosphate Ba salt	3′-Acetyldeoxyguanosine
	Uridine-2′,3′ phosphate Ba salt	3′,5′-Diacetyldeoxyguanosine
	Guanosine-2′,3′ phosphate Ba salt	

Spectra.—These were obtained on a Perkin–Elmer Model 21 double-beam spectrometer with a sodium chloride prism and a Hilger D209 spectrometer and calcium fluoride and sodium

Infrared spectra of: (1) *adenine*, (2) *deuterioadenine*, (3) *adenosine*, (4) *deuterioadenosine*, (5) *deoxyadenosine*, (6) *adenine hydrochloride*, (7) *adenylic acid* a, (8) *adenylic acid* b, (9) *adenosine-5′ phosphate.*



chloride prisms. The majority of the spectra were taken as Nujol and hexachlorobutadiene mulls and some in KBr discs. A number of the spectra were obtained from the collection of infrared spectra at the Chemistry Laboratories, Cambridge, and some of these (five adenosine nucleotides [7] and six cytidine nucleotides [8]) have already been published. These known spectra were used in their original (full-size) form in the present work for the assignment of bands.

The spectra of a representative series of compounds are presented in the Figure for the region 4000—650 cm.$^{-1}$; the frequencies of these and other compounds are listed in Tables 2—6. A list of related compounds for which spectra were also available is given in Table 1. Data for these compounds were used in assignments and comparisons.

TABLE 2. *Cytosine derivatives.*

Cytosine	Deuterio-cytosine	Cytosine HCl	Cytidine	Deuterio-cytidine	Deoxy-cytidine
700w		697m		694w	
	718w	733w	714m	720w	717m
		755s	757w	745w	752w
	771w			775w	
784w	787s	782w			786w
794s		787w	791s	790s	792s
823m		825s	818s	810m	
	851w	850w	855m	845m	860w
	864w	885s	872m	855m	899w
				917m	931w
966w	968w		944m	960m	960m
995w	978w	976m	985m	980w	993m
1011w	1006w	1005m		1003m	
			1036w	1040m	1047s
	1046w		1054s	1063s	1057w
		1105w	1098m	1080m	1096w
	1118m		1136w	1115m	1115m
1156w	1178w		1153w	1170w	
			1192w	1183w	1176w
1236s	1235w	1228s	1212m		1213w
1275s	1279m		1247w	1255m	1279s
			1290s	1295s	
	1312m		1309w		
1362s	1374s	1370w	1342w		1356w
	1416w	1414s	1400w	1396m	1397w
1458s					
1506w	1502s	1497m	1504s	1500s	
1538m	1548w	1543s	1543m	1515w	1527w
1621w	1605s		1607s	1608s	1618s
1655w	1637s			1640s	1642w
1667s	1680w		1653s		1667m
		1682s			
		1736s			
	2123s				
	2155s			2427s	
	2320s			2564m	
1667m	2525s	2630w	2725w	2725w	
2778m		2899m	2915m		2940m
			3085m	3077w	
3155s		3105m	3225s	3205w	3106s
3367s	3278w	3355w	3333s	3333w	3370s
			3448m		

DISCUSSION

Cytosine Derivatives.—In the spectrum of cytosine itself (6-amino-2-hydroxypyrimidine or 6-amino-2-pyrimidone) there is a broad strong band in the double-bond region centred at 1667 cm.$^{-1}$ with several shoulders on the low-frequency side. 3-Methylcytosine, however, gives two distinct bands, at 1626 and 1667 cm.$^{-1}$. It is supposed that the complex band of cytosine itself results from the superposition of two absorptions caused by the NH_2 deformation and the conjugated C=O stretching vibration. In derivatives

[7] Brown and Todd, *J.*, 1952, 44.
[8] Michelson and Todd, *J.*, 1954, 34.

where the amino-group is removed or substituted, this band moves to a lower frequency: *e.g.*, deuteriocytosine 1637 cm.$^{-1}$, 2-hydroxypyrimidine (2-pyrimidone) 1644 cm.$^{-1}$, and 6-dimethylamino-3-methyl-2-pyrimidone 1645 cm.$^{-1}$. This shows that the carbonyl frequency is at about 1640 cm.$^{-1}$ and that the NH_2 vibration in cytosine is at a somewhat higher frequency. In 1,3-dimethylcytosine the conjugation of the carbonyl group is removed, and the carbonyl band moves to 1681 cm.$^{-1}$. These results and those in the 3000 cm.$^{-1}$ region discussed below confirm the structure of cytosine as that of a 2-pyrimidone derivative, as has been suggested previously on infrared evidence. Another weaker band in the 1600—1640 cm.$^{-1}$ region usually appearing with derivatives of cytosine and other 2-pyrimidones is assigned to a ring vibration (probably mainly coupled C=N stretching modes). This band is characteristic of the pyrimidone ring system as a whole, for it is not present in 1,3-dimethylcytosine or in uracil derivatives which have no ring C=N bonds.

In the region 1535—1550 cm.$^{-1}$ many cytosine derivatives show a strong band which seems to be characteristic of the pyrimidone structure: 2-hydroxypyrimidine (2-pyrimidone) gives the frequency 1545 cm.$^{-1}$, and 4-hydroxypyrimidine (4-pyrimidone) 1541 cm.$^{-1}$. This band is most reasonably assigned to an N–H in-plane deformation vibration. The band at 1538 cm.$^{-1}$ in the spectrum of cytosine is removed on deuteration, and becomes

TABLE 3. *Cytidylic acids.*

Cytidine HCl	Deoxy-cytidine HCl	Cytidylic acid *a*	Cytidylic acid *b*	Cytidine-5′ phosphate	Deoxy-cytidine-3′ phosphate	Deoxy-cytidine-5′ phosphate
703w	711w	671w	666m	681w		
723m	721m	728m	725w	717m	730w	727w
752w	746m	755w	743w	755w		752w
		762w	755w			770w
784w	772s	779w	777m	773w	778w	787w
816w	788w	797m	790w	786w	800w	
830s	832w	844w		808m	830w	806m
864w	847s	874w	854m	837w	857w	866w
873w		881w	896w	870w	897w	888w
915m		914m	914w	905w		
		926m	939w	931m	935w	927s
983m	957m	973w	969s	988m	966m	980m
993m	1026w	1002w		1016w	1013m	1030m
1045m	1042w	1044w	1042m	1048s		1038m
1081m		1069s	1075s	1069m	1058m	1060s
1093w	1087s			1093m	1093m	1093w
1117s	1129m	1108m	1108w	1111m		
	1143w	1124s	1150s	1161m	1156m	1156w
1188m	1183w	1181w	1190w		1195w	1196w
1239m	1228m	1209m	1227w	1222m	1235w	1227m
1278s	1271s	1266m	1261w	1255w	1258w	1259w
1293w	1285m	1277m	1280m	1279m	1282s	1288m
1333w	1311w	1309w	1307w	1333w	1324w	1333w
	1349w	1351w	1330w	1351w	1351w	1355w
1396w	1420w	1418w	1414w		1408m	
1538s	1538s	1536m	1538m	1545m	1538s	1553m
				1605w		1600w
			1647m	1626w	1647w	
1681s	1675s	1721s	1695s	1704s	1686m	1698m
1724s	1715s	1748m	1724s	1724s	1730m	1733s
3096s	3086m	3090m				
	3215m	3225m		3125w		3120w
3300s	3333s	3300m	3246m	3300m	3320s	3333w
3448m				3570w		

very weak in the spectra of 3-methylcytosine, cytidine, and deoxycytidine, all of which have the hydrogen atom of $N_{(3)}$ substituted. In the spectrum of deuteriocytosine a new band appears at 1118 cm.$^{-1}$, giving an isotopic ratio of 1·375. Strong evidence in favour of the origin of this band in a principally hydrogen deformation vibration is found in the fact that the hydrochlorides of cytidine and deoxycytidine give a strong band in this region. In these compounds the additional proton is added to $N_{(1)}$ and thereby re-establishes a

secondary amide group. This band also appears in the same region in the spectra of all the guanine derivatives.

There is a strong band near 1275 cm.$^{-1}$ in the spectrum of cytosine and all cytosine derivatives. But it is not present in those of 2-hydroxypyrimidine (2-pyrimidone) or 6-methoxy-3-methyl-2-pyrimidone where the external amino-group is missing. This band is therefore attributed to the external C–N bond, as assignment in good agreement

TABLE 4. *Uracil derivatives.*

Uracil	Deuterio-uracil	5-Methyl-uracil	1,3-Di-methyl-uracil	Uridine	Deuterio-uridine	Deoxy-uridine	Thymidine	Uridylic acid
			683m	715w	720w		723w	673w
	733w	740m	728w	752w	743w	734w	734w	732w
760m	765s	762s	762s	766s	766s	766s	758s	760w
781w				775w	786w		766m	
807w			800w	807w	806s		792w	810w
822s	820s	815s	816s	831s	828s	830w		833m
851m		845s		853w	853w	863m	851m	869w
	882m			877w	897w		870w	891w
	965w	935m	931w	953m	969m	934w	957m	
994s	980s	983m	966w	982m		960m	971m	973m
1004s			1004w	1005w	1010s	998m	1008m	997s
		1029m		1035w		1042m	1025w	1020m
		1048w		1054s	1056s	1056m		1042m
				1075m	1075w	1073w	1064s	1075w
1099w				1099s	1105s	1100m	1096m	1101w
	1140s		1136m	1136w	1130w	1112w	1120m	1124m
	1165w		1172m	1172w	1183w	1167m	1170w	
		1205s						1203w
1238s	1230w	1244s	1230m	1211m	1239m	1221w	1221m	1247w
			1261w	1269s	1264w	1259s	1270s	1260m
	1310w			1323w	1287m	1288w	1316m	1282s
	1330s	1366w	1340s	1366m	1326w		1362m	1314w
1389m	1390w	1381m	1377m	1398m	1393m		1400m	1379s
1418s	1422w	1422s	1400s	1424m	1416w		1435s	1412w
1451s	1475s	1447s	1437s	1475m	1460m		1475s	1466s
		1484m	1480m					
1508m		1495m					1515w	
	1565w		1587w		1610m	1618w		
1666m	1640s		1655w		1647s	1664m	1661s	
1681m		1681s	1664s	1686s	1686s	1681s		1686s
1724s	1712s	1739s	1709w			1695w	1709s	1724s
1745w	1740w	1785w		1788w				
	2135w							
	2268s				2326m			
	2400w				2475s			
		2800w		2810w			2833m	
2924m		2920m	2940w	2933m	2915m		2967m	2915m
				2985m			3000m	
3080s	3065m	3020s	3077w		3086m		3077w	3077w
		3170m		3125m		3125m	3155w	3185s
				3340s	3333s		3300s	3415m

with that given by Bellamy [9] for the C–N stretching vibration in primary aromatic amines.

In the region 780—800 cm.$^{-1}$ all cytosine derivatives exhibit a strong and sharp band. These bands are very similar to those due to out-of-plane hydrogen deformation vibrations in the case of substituted benzene derivatives and olefins, and they probably arise from this type of vibration. In the spectrum of cytosine itself there is also a broad band at 823 cm.$^{-1}$ which may be assigned to an out-of-plane vibration of the hydrogen atom on $N_{(3)}$ because it is removed both by deuteration and by replacement of this hydrogen by a methyl group. It is well known that out-of-plane N–H or O–H vibrations usually give characteristically broad bands in the low-frequency region.

Deuteration causes significant changes in the 1400—1600 cm.$^{-1}$ region of the spectrum

[9] Bellamy, " The Infra-red Spectra of Complex Molecules," Methuen, London, 1954.

of cytosine, indicating that vibrations usually considered to involve doubly-bonded and conjugated heavy atoms include considerable contributions from hydrogen vibrations. Actually the spectra of deuteriocytosine and cytidine closely resemble each other in the absence of the strong band found at 1458 cm.$^{-1}$ in cytosine, and the appearance of others at 1503 and 1605 cm.$^{-1}$. These changes are probably associated with the removal of the hydrogen atom from $N_{(3)}$ and hence very little further change is observed in this region on deuteration of cytidine.

An examination of the 3000 cm.$^{-1}$ region provides confirmation that cytosine and its

TABLE 5. *Adenine derivatives.*

Adenine	Deuterio-adenine	Adenosine	Deuterio-adenosine	Deoxy-adenosine	Adenine HCl	Adenylic acid *a*	Adenylic acid *b*	Adenosine-5′ phosphate
	708m	705w	704s	703m	702w	696w	677w	683m
722m	720w	722m	730w	728m	718m	724w	723m	725m
		767m	756s	770m	776m	778w	775m	737m
797m	796m	795w	796m	796w			796m	787m
		823m	812s	837w		825w	849w	822m
847m	855w	844m	858w	858w	858m		869w	831w
870w		860w	877s	897w	888m	868w	879w	893w
911m	896m	903s		906w	904w	904w	926m	913w
938m	935m		948m	930w	937w	923m	941s	930m
		978m	962s	950m				966s
			980m	977w	974w		975m	
1022m	1010w	1011m	1020w	1006m	1024w	1020w		1020w
		1037s	1036s	1028m	1036w	1059m	1042s	
		1054m	1067s	1054s	1072w		1070s	1064s
		1071m	1085w	1099s				1089s
1124m	1130w	1107m	1111s	1123w	1105w	1108w	1100m	1114m
1155w	1156w	1142w	1134m	1158w		1154w	1143m	
	1183w	1176w	1163w	1180w	1171m		1186w	1170w
1206m		1208m	1200w	1209m		1213m	1218s	1219m
	1226w	1225w	1235w	1233w				
1250m	1250w	1250w	1261w	1264w	1258s		1259w	1250w
	1285m		1290s					
1305m	1310s	1300s		1300s			1304w	1294w
1333m	1335s	1333m	1332s	1324m	1319w	1324w	1328w	1317w
		1351w	1370w	1339m			1340m	1345w
1368w	1375m	1370w	1383m	1366w	1381m			
	1410m	1387w	1402w	1390w				1402w
1418m	1420m	1414s		1418w	1425w	1422w	1416m	1422w
1451w	1460m		1460m	1447w	1441w			
	1475m	1470s		1473w				
1510w	1510s		1510m	1515w	1500m	1507w	1504w	1500m
	1575s	1575m	1572s	1574s	1597m	1555w	1553w	1562w
1605m	1610s	1608m	1623s	1612s	1623s	1607w	1615s	1613m
1672s	1673w	1672s		1672s	1712s	1691s	1706s	1700s
	2155m							
	2288s		2320s					
	2463m		2457s		2565w			
2680w		2667w	2500w	2680w				
2780w		2825w						
		2900m	2920w	2915m				
3000w					2958m			
3105s	3096w	3125s		3067s		3075s	3125m	3155s
3280m		3300s	3300m	3378w	3280m	3225s	3300m	3350m
						3480w		

derivatives exist as amino-forms. Cytosine, 3-methylcytosine, cytidine, and deoxycytidine all give two strong bands in this region which on deuteration shift to lower frequencies as follows:

cytosine	3155 and 3367 cm.$^{-1}$ to 2320 and 2525 cm.$^{-1}$
3-methylcytosine	3130 and 3333 cm.$^{-1}$
cytidine	3225 and 3333 cm.$^{-1}$ to 2427 and 2564 cm.$^{-1}$
deoxycytidine	3106 and 3370 cm.$^{-1}$

These frequencies are in good agreement with those of aromatic amino-compounds where

the presence of the amino-group is certain *e.g.*, α-naphthylamine [10] (3230 and 3330) and 5-aminopyrimidine (3180 and 3345). 1,3-Dimethylcytosine gives only one strong band in this region, at 3268 cm.$^{-1}$; this can be assigned to the imino(NH)-group.[11]

If the transition from cytosine to the two nucleosides is considered, very little change occurs in those parts of the spectra due to the cytosine residue, but a number of strong bands appear between 1000 and 1200 cm.$^{-1}$ which are due to the sugar molecules. When, however, the transition is followed from the nucleosides to the nucleotides, quite considerable changes in the spectra are evident; these changes closely parallel those occurring on passing from cytosine or the nucleosides to their hydrochlorides. Instead of the 1667 cm.$^{-1}$ band of cytosine, the hydrochlorides show two strong bands in the ranges 1675—1682 cm.$^{-1}$ and 1715—1736 cm.$^{-1}$, while the cytidylic acids have two bands at 1686—1720 cm.$^{-1}$ and 1724—1748 cm.$^{-1}$. These bands are assigned as NH_2 deformation and C=O stretching vibrations respectively. Other similarities between the spectra of the hydrochlorides

TABLE 6. *Guanine derivatives.*

Guanine	Deuterio-guanine	Guanosine	Deuterio-guanosine	Deoxy-guanosine	Guanylic acid
689m	687w	678m	665w	678m	
703m	702w		697br		692w
727w	721w	733w	726w	723w	
776s	776m	778s	777m	778s	780m
789w		800w	807w	800w	
	825w	827w			839w
851s	850m			870w	848w
877m	876w	882w	870w	890w	
	912w	910m	900w		917sh
949s	948w	935w		932m	930w
		978w	980s	993w	976s
				1029w	1030m
1042w		1036m	1042s	1058m	1057m
		1078s	1093s	1095m	1078s
1120m	1120w	1118m			1127s
1172s	1167m	1167w		1171m	1176m
1215w	1215w	1227w			
1263s	1261w		1263w	1245w	1235w
	1307w	1317m	1302m	1324sh	
	1337w	1337m	1348w	1368sh	1355w
1375s	1364s	1393m		1393m	1380w
1418m	1416w	1422m	1404w		1416w
1464w	1460m		1456w		1455w
1477m	1475m	1486s		1488m	1481w
1555s		1538m	1534s	1538m	1538w
1567s	1565m	1575w	1575s	1595sh	
	1616s	1613sh		1608s	1608m
1639w		1631m		1637s	
1681s	1678s		1680m		
1701s	1695s	1736m	1704m	1733s	1698s
	2114m		2075m		
	2278s		2370m		
	2494s		2500s		
2680s	2688m	2703m		2703w	
2840s	2872m	2833w	2900m	2857sh	
3077s	3086m			2924m	
		3175m		3185s	3150m
3310s	3300s	3280m	3330w	3310s	3425s

and the cytidylic acids show that the same changes in the cytosine nucleus occur in the two types of compound. Since in the hydrochlorides a proton is added to the $N_{(1)}$ atom, this must also occur in the case of the cytidylic acids. It is concluded, therefore, that all the cytidylic acids exist in a zwitterion form in the solid state, one of the hydrogen atoms from the phosphate group having moved to the cytosine residue. Such zwitterion

[10] Flett, *J.*, 1948, 1441.
[11] Angyal and Werner, *J.*, 1952, 2911.

formation has already been suggested to exist in solution,[12] but the present infrared results appear to be the first to present strong evidence for the existence of zwitterions for nucleic acids in the solid state.

Uracil Derivatives.—In the case of uracil the bond distances for the two C–O bonds obtained from *X*-ray evidence [13] clearly indicate that the molecule exists in the diketo-form, *i.e.*, as 1,2,3,4-tetrahydro-2,6-dioxopyrimidine; strong supporting evidence for this structure for uracil derivatives comes from the examination of the infrared spectra in the double-bond region.[1] Uracil and each of its methyl-substituted derivatives show two strong bands in this region, and these have been assigned by Randall *et al.*[14] to the 4- and 2-keto-groups. While these bands are rather ill-defined for the solid, in the case of 1,3-dimethyluracil they could be satisfactorily resolved in solution (in $CHCl_3$ at 1660 and 1703 cm.$^{-1}$, in CCl_4 at 1670 and 1710 cm.$^{-1}$, and in C_6H_6 at 1670 and 1707 cm.$^{-1}$). The position of these bands seems to be characteristic for the grouping –NH·CO·NH·CO– in a six-membered ring containing a carbon–carbon double bond (see also xanthine and caffeine [14]). The band of moderate intensity at about 1500 cm.$^{-1}$ in uracil and its 5- and 4-methyl derivatives is considered to correspond to the band in the same region in the cytosine derivatives, *i.e.*, an N–H in-plane deformation vibration. On deuteration of uracil this band is removed and a new band appears at 1140 cm.$^{-1}$ (ratio 1·32).

There are two strong bands in the spectra of all uracil derivatives in the regions 1415—1440 cm.$^{-1}$ and 1450—1490 cm.$^{-1}$, respectively. These are assigned principally to vibrations of the uracil nucleus but shifts on deuteration suggest that they are partially coupled with some N–H vibration. The second of these bands shifts by about 20 cm.$^{-1}$ in the same direction on deuteration as when the $N_{(3)}$ atom is otherwise substituted (3-methyluracil 1486 cm.$^{-1}$, uridine 1475 cm.$^{-1}$).

In all the uracil derivatives there are two strong, sharp bands at about 760 and 820 cm.$^{-1}$. They are not affected by deuteration or by substitution of both N–H groups. They may correspond to out-of-plane C–H vibrations and seem to be very characteristic of the tetrahydro-2,6-dioxopyrimidine structure. There are two additional broad bands in the spectrum of uracil, at 807 and 851 cm.$^{-1}$. These are assigned to the two out-of-plane NH deformation vibrations because they are absent in the spectrum of deuterio-uracil; they are present in 4-methyluracil, but one of them is removed in 3-methyluracil and both of them are removed from the spectrum of 1,3-dimethyluracil.

The 3000 cm.$^{-1}$ region shows a strong band for uracil and its methyl-substituted derivatives between 3050 and 3150 cm.$^{-1}$. By comparison with spectra for phthalimide (3205 cm.$^{-1}$) and succinimide (3145 cm.$^{-1}$) these bands can be satisfactorily assigned as N–H stretching vibrations of the –CO·NH·CO– grouping. The other NH fundamental probably contributes to this absorption band. On deuteration of uracil this band shifts to 2268 cm.$^{-1}$. 1,3-Dimethyluracil gives only two weak bands, at 3077 and 2940 cm.$^{-1}$, which are probably due to unsaturated and methyl C–H stretching vibrations, respectively.

There is very little change in the spectrum due to the uracil nucleus when a sugar group is added. Uridine and deoxyuridine show only one broad band in the 1700—1600 cm.$^{-1}$ region, but it seems probable that this would be resolvable if the compounds could be examined in solution. A number of strong bands appear in the region 900—1100 cm.$^{-1}$ which are characteristic of the sugar residue.

For uridylic acid *b*, again very little change is found on the addition of the phosphate group to uridine. This is in agreement with expectation, since in this case zwitterion formation is much less likely. The similarity of the spectra of uracil and its nucleosides and nucleotide suggests that in these more complex compounds the uracil is also present in the diketo-form.

[12] Cavalieri, *J. Amer. Chem. Soc.*, 1952, **74**, 5804.
[13] Parry, *Acta Cryst.*, 1954, **7**, 313.
[14] Randall, Fowler, Fuson, and Dangl, " Infra-red Determination of Organic Structures," D. van Nostrand Co., New York, 1949.

Adenine Derivatives.—Since infrared evidence strongly suggests that adenine (6-amino-purine) exists in the amino-form and it would be expected that the structure of the purine nucleus would not be significantly affected by the addition of the amino-group, an attempt was made to correlate bands in the spectra of adenine and purine. However, it was found that there was little resemblance between the two spectra except for the presence of two strong bands at 1575 and 1610 cm.$^{-1}$. These two bands are very characteristic of the purine nucleus (or the pyrimidine ring; pyrimidine [1] has bands at 1570 and 1610 cm.$^{-1}$) and are given by all adenine derivatives (including the six acetyl-substituted adenosines) in the narrow frequency ranges 1575 $\pm$ 5 and 1610 $\pm$ 10 cm.$^{-1}$. The first of these bands appears only as an unresolved shoulder in the spectrum of adenine itself but is clearly present in those of deuterated adenine and 9-methyladenine.

Evidence for the amino-form can be found in the bands at 1672 cm.$^{-1}$ in the spectra of adenine, adenosine, and deoxyadenosine, which can without doubt be assigned to an NH_2 deformation vibration. They are removed on deuteration and are in approximately the same position as the corresponding bands in the spectra of fourteen aminopyrimidine derivatives [1] (including 5-aminopyrimidine, which cannot have any other tautomeric form). In contrast to the cases of cytosine and uracil, deuteration causes few changes in the spectra of adenine and adenosine, except for removal of the 1672 cm.$^{-1}$ band and appearance of a new one at 1510 cm.$^{-1}$ with both compounds, and another at 1285 cm.$^{-1}$ for adenine and at 1290 cm.$^{-1}$ for adenosine. The first of these is probably a shifted skeletal vibration and the second is assigned to the ND_2 deformation vibration (ratio 1·30), even although its intensity in deuterioadenine is less than would be expected.

In the region 1300—1500 cm.$^{-1}$ the spectrum of adenine resembles that of 9-methyl-purine more than that of purine itself. A series of five fairly strong bands of adenine in this region can also be found in the spectrum of 9-methylpurine and of all the adenine derivatives in approximately the same positions. The strongest of these, near 1305 cm.$^{-1}$, is very characteristic of these compounds. None of these bands is affected by deuteration. The broad band at 870 cm.$^{-1}$ in the spectrum of adenine is assigned to the out-of-plane deformation vibration of the $N_{(9)}$–H; it is removed on deuteration and by $N_{(9)}$-methylation. A broad band in the same position can also be found in the spectrum of purine but is absent from that of 9-methylpurine.

Addition of the ribose or deoxyribose unit to the adenine molecule does not cause any changes in the characteristic features of the spectrum due to the adenine part, but once again a number of strong bands appear in the region 950—1150 cm.$^{-1}$ which are due to the sugar residues. Very considerable changes, however, occur in the spectra when a phosphate group is added to the adenine nucleosides; these changes are closely parallel to those which occur when a proton is added to the adenine nucleus.

The structure of adenine hydrochloride has been determined by *X*-ray crystallography,[15] and the extra proton was found to be located on the $N_{(1)}$ atom. It would be expected that the charge distribution and bond orders for the $(adenine,H)^+$ ion would be somewhat different from those existing in the adenine molecule and that changes might well occur in the infrared spectrum. The most outstanding of these changes observed is the shift of the NH_2 deformation vibration from 1672 cm.$^{-1}$ in adenine to 1712 cm.$^{-1}$ in the hydrochloride (the picrate has this band at 1704 cm.$^{-1}$ and the sulphate [14] at 1698 cm.$^{-1}$). From the direction of the shift it is clear that we are dealing, not with an NH_3^+ unit, but with a NH_2 group affected by the addition of a neighbouring positive charge. In the spectra of all the adenine nucleotides the corresponding band is in the region 1690—1710 cm.$^{-1}$. It is, therefore, suggested that in the adenylic acids the adenine part has virtually the same molecular and electronic structure as in adenine hydrochloride. This implies that there is probably a proton on the $N_{(1)}$ atom and that the adenylic acids exist in a zwitterion form.

[15] Cochran, *Acta Cryst.*, 1951, **4**, 92.

Other marked changes can also be observed in the spectra of adenine,HCl and the adenylic acids in the region associated with the aminopurine unit. For example, the band at 1605 cm.$^{-1}$ in adenine has lost intensity considerably, and the band at 1572 cm.$^{-1}$ has disappeared completely, giving place to a very weak band at 1550—1560 cm.$^{-1}$ in the adenylic acids. On the other hand, a new band appears near 1500 cm.$^{-1}$, strong in the spectrum of adenine,HCl and weaker in the adenylic acids. Gone also is the outstanding band at 1305 cm.$^{-1}$, and the region 1300—1500 cm.$^{-1}$ shows only weak bands in the spectrum of adenylic acid *b* hydrate (the only adenine nucleotide examined in hexachlorobutadiene also). Since most of these bands were probably associated with stretching vibrations of the adenine derivatives, the above changes indicate considerable changes in the electronic distribution within the adenine nucleus. It should be mentioned that in the case of the six acetyl-substituted adenine nucleosides (where the acetyl groups are in corresponding positions to the phosphate groups) no such changes in the spectra occurred. All six of these compounds have strong bands at 1667—1682 and 1605—1613 cm.$^{-1}$ and a band of medium intensity at 1574—1584 cm.$^{-1}$ as for adenine itself. Since the introduction of one or more acetyl groups into the sugar residue of adenosine cannot lead to zwitterion formation, these data lend additional support to the hypothesis of zwitterion formation in the case of the adenylic acids.

In the 3000 cm.$^{-1}$ region adenine and the two nucleosides each show two strong bands at about 3100 and 3300 cm.$^{-1}$ which are shifted on deuteration to lower frequencies. These bands are characteristic of aromatic amines, and their presence strengthens the evidence for the existence of these compounds in the amino-form. The broad absorption region between 2900 and 1500 cm.$^{-1}$ in adenine is absent from the spectrum of 9-methyladenine and is most probably due to the $N_{(9)}$–H vibration. A similar set of bands for purine between 2540 and 2750 cm.$^{-1}$ is removed on methylation at position 9. For nucleotides the spectra in the 3000 cm.$^{-1}$ region are not very satisfactory and hence no evidence could be obtained for or against zwitterion formation. However, for the hydrochlorides of adenine, cytosine, and guanine the main bands are broader but still in approximately the same position as those of the parent bases, again indicating that the proton is added to a ring-nitrogen atom and does not form an NH_3^+ group.

Guanine Derivatives.—Guanine (2-amino-6-hydroxypurine or a tautomeric form) bears a close structural relation to cytosine, except for the reversal of the positions of the amino- and hydroxyl-groups. It would, therefore, be expected that it exists in the same tautomeric form, *i.e.*, as the amino-keto-derivative; the infrared spectra can be satisfactorily interpreted along these lines. However, there are two important differences between guanine and cytosine which might show in the spectra of guanine derivatives. The first is that in the nucleosides the sugar units are attached to the $N_{(9)}$ atom and not to the pyrimidine ring. The second is that, according to the *X*-ray evidence on guanine hydrochloride,[16] the proton is added to the $N_{(7)}$ atom and hence the effect on the pyrimidine ring structure would be considerably less if analogous zwitterion formation occurs in guanylic acid.

The spectrum of guanine has two strong bands, at 1701 and 1681 cm.$^{-1}$, which are assigned to the C=O stretching and NH_2 deformation vibrations, respectively. The considerable shift of these bands to about 1735 and 1635 cm.$^{-1}$ in the spectra of guanosine and deoxyguanosine is difficult to explain, but may be due to differences in intermolecular hydrogen bonding between guanine and its nucleosides. In the case of deuterioguanine the band at 1695 cm.$^{-1}$ is ascribed to the C=O stretching vibration; the appearance of a new strong band at 1616 cm.$^{-1}$ closely parallels the case of deuteriocytosine.

The removal of one of the bands near 1600 cm.$^{-1}$ in guanine on deuteration supports the assignment of the NH_2 deformation vibration to this region. In the spectrum of deuterioguanosine the strong doublet at 1704 and 1680 cm.$^{-1}$ is assigned to the C=O vibration and resembles the spectrum of 6-hydroxy-2-methylpurine; [2] the reason for this doubling, however, is not clear in either case.

[16] Broomhead, *Acta Cryst.*, 1951, **4**, 81.

The lower-frequency band of the doublet at 1567 and 1555 cm.$^{-1}$ in guanine is assigned to the N–H deformation; it is removed on deuteration while the other band stays (at 1565 cm.$^{-1}$). This band persists for the other derivatives and is also given by the two acetyldeoxyguanosine derivatives, at 1545 and 1538 cm.$^{-1}$ severally. As mentioned previously, the addition of the sugar residue to guanine does not affect the hydrogen atom on $N_{(1)}$ and, therefore, the band at about 1540 cm.$^{-1}$ is expected to persist through the complete series, in agreement with experiment. Bands of considerable intensity in the spectrum of guanine at 1477 and 1375 cm.$^{-1}$ can be found in the spectra of all the other guanine derivatives in the same regions and are assigned to vibrations of the ring systems. The only other band common to the spectra of all the guanine derivatives (and the two acetyldeoxyguanosines) is the strong band at 785—775 cm.$^{-1}$. This is most probably due to the out-of-plane vibration of the only hydrogen atom attached to a carbon atom (at position 8). The strong band at 1698 cm.$^{-1}$ in guanylic acid is undoubtedly due to the C=O stretching vibration. The next band at 1608 cm.$^{-1}$ has a considerably lower frequency than the corresponding bands in the spectra of guanosine and deoxyguanosine. In the 3000 cm.$^{-1}$ region the spectrum of deuterioguanosine shows that deuteration in this case was not very satisfactory. However, the positions of the bands in guanine and the two nucleosides, and their shifts on deuteration, confirm the amino-structure for these compounds.

Sugar and Phosphate Groups.—All the nucleosides and nucleotides, and the corresponding deoxy-compounds examined, exhibited a number of strong bands in the region 900—1200 cm.$^{-1}$ which were not present in the spectra of the corresponding parent bases. A search was made for characteristic vibration frequencies of the ribose and deoxyribose units, but a careful examination of the spectra of all the sugar derivatives in the present work (20 ribose and 19 deoxyribose derivatives) has shown that no regularities could be found among the bands in this region. Ten such pairs of compounds were available for comparison. Barker *et al.*[17] found for sugar pyranosides that all the deoxy-compounds exhibited an additional band at 867 cm.$^{-1}$, which they have assigned to a CH_2 rocking mode. No such additional bands for the deoxy-derivatives could be found in the present series of compounds, although it should be realised that the spectra are more complex. An examination of the region 950—1200 cm.$^{-1}$, where the bands are due to C–O and C–C vibrations, showed that nearly all of these compounds gave five bands in this region. A statistical analysis of the frequencies of these bands indicated that only for one of these bands is there a significant difference in frequency (ribose 1076 $\pm$ 6·6 cm.$^{-1}$, deoxyribose 1087 $\pm$ 8·7 cm.$^{-1}$).

Fraser [18] has suggested that in the spectra of ribose nucleic acid and deoxyribose nucleic acid the bands at about 1087 cm.$^{-1}$ are due to the sugar ring and those at 1070 and 1052 cm.$^{-1}$, respectively, due to the phosphate group. Our results show that this assignment cannot be applied to the smaller units because a band at this position is given by 36 out of the 39 compounds examined and only 13 of these contain a phosphate group.

The spectra of the available nucleoside phosphates were carefully examined for any additional bands beyond those in the spectra of the corresponding nucleosides in the regions 1300—1200 (P=O band), and 1150—900 cm.$^{-1}$ (phosphate ion or P–O–C band), but none of these investigations yielded any bands characteristic of the phosphate grouping. It had, therefore, to be concluded that, because of the large number of strong bands already present in the spectra of the nucleosides, no characteristic bands of the phosphate group could be picked out in the spectra of the nucleoside phosphates. Tsuboi [19] has shown that a band in the region 1220—1240 cm.$^{-1}$ in the spectra of the sodium salts of nucleic acids can be assigned to a vibration of the PO_2^- group. He also ascribes bands at 980 and 1100 cm.$^{-1}$ in the spectra of nucleotides in aqueous solutions at high pH's to

[17] Barker and Stephens, *J.*, 1954, 4550.
[18] Fraser, " Progress in Biophysics," 1953, Vol. III, p. 47.
[19] Tsuboi, *J. Amer. Chem. Soc.*, 1957, **79**, 1351.

vibrations of the PO_3^{2-} group, but observes that these bands disappear at lower pH's. This absence of bands is consistent with our findings on the solid nucleotides where the phosphate group presumably exists as PO_3H_2, or as PO_3H^- when zwitterion formation has occurred.

The important problem of the position of the phosphate group on the sugar entity in the case of the *a* and *b* isomers of the nucleotides was also studied by the examination of the 950—1100 cm.$^{-1}$ region in the spectra of the three pairs of isomers (the adenylic acids, the cytidylic acids, and the cytidine benzyl phosphates). It was found that in each case the *a* isomers gave only one strong band (at 1059, 1069, and 1075 cm.$^{-1}$, respectively), while the *b* isomers gave two strong bands (at 1042, 1070; 1042, 1075, and at 1047, 1074 cm.$^{-1}$, respectively). The guanylic and uridylic acids, which were only available as the *b* isomers, also gave two pairs of strong bands (1057, 1078; and 1042, 1075 cm.$^{-1}$, respectively) in this region. However, the spectra of a larger number of isomeric pairs will have to be examined to find out whether this correlation is of general applicability.

Cytidine, uridine, adenosine, and thymidine were also examined in neutral water and deuterium oxide solutions. By comparing these spectra with those of the same compounds in the solid state, it was found that only the stronger bands appeared in the solution spectra but that every band corresponded exactly to a band in the solid state spectrum, within the experimental error. These observations, that the spectra do not change when the compounds are dissolved in water, are of importance when attempts are made to extend the results for the simple compounds obtained in the solid state to the spectra of the nucleic acids of high molecular weight that are always examined in a highly hydrated state.

The author expresses his gratitude to Dr. Norman Sheppard for supervision, advice, and kind encouragement, and to Sir Alexander Todd, Drs. G. W. Kenner, A. M. Michelson, and D. M. Brown for a number of helpful discussions, for most of the compounds examined, and for permission to use some of their spectra. The author was supported by an Oliver Gatty Studentship of the University of Cambridge during this research.

Department of Colloid Science,
Free School Lane, Cambridge.
[Present address: Union Carbide Research Institute,
P.O. Box 324, Tuxedo, New York, U.S.A.]

[*Received, June 23rd,* 1960.]