# Crystal and Molecular Structure of Cytosine Hemitrichloroacetate

**Maria Gdaniec, Bogumił Brycki, and Mirosław Szafran** *
*Department of Chemistry, A. Mickiewicz University, 60780 Poznań, Poland*

The crystal structure of cytosine hemitrichloroacetate $(C_4H_5N_3O)_2$·$HOOCCCl_3$ has been determined at room temperature by *X*-ray crystallographic analysis. The structure consists of crystallographically independent, centrosymmetric triple hydrogen-bonded cytosine pairs, A and B, and trichloroacetate anions. Base pairing occurs between symmetry-equivalent hemiprotonated cytosines with a proton shared between the N(3) atoms. The observed N(3) ··· N(3′) distances of 2.755(8) and 2.858(8) Å in base pair A and B, respectively, are typical for asymmetric hydrogen bonds described by an asymmetric double minimum-energy curve. The i.r. spectrum is consistent with that predicted for this type of bridge system.

Many oxygen and nitrogen bases may form hydrogen bonds with their conjugated cations to give homoconjugated ion pairs involving two molecules of base $(BHB)^+A^-$.[1-3] Such an interaction occurs between 1-methylcytosine and 1-methylcytosine protonated at the N(3) atom.[4] These complexes are characterized by the triple hydrogen-bonded 1-methylcytosine residue. These interactions lead to a non-Watson–Crick base-pairing scheme,[5] which can be responsible for mutation and the loss of the genetic code.[6] Gray *et al.*[7] have recently shown that helical forms of poly[r(C)], poly[d(C)], and poly[d(C–T)] are stabilized by protonated cytosine base pairs. For each of these polymers, the formation of a helical complex requires the addition of one proton per two cytosine bases. These workers have suggested that the additional proton is shared between the N(3) ring nitrogens of the two cytosine bases, which are thus stabilized by three hydrogen bonds.

We now present the crystal structure of homoconjugated cytosine trichloroacetate. We are aware of one previous report describing the crystal structure of a homoconjugated cytosine complex formed with the aromatic carboxylic acid resorcylic acid.[8] Cytosine hemitrichloroacetate is the first example of such a complex formed with an aliphatic acid. Complexes of aromatic acids differ from those of aliphatic acids.[9]

## Experimental

I.r. spectra were recorded on Perkin-Elmer Model 580 spectrophotometer in Nujol and Fluorolube at room temperature.

Cytosine hemitrichloroacetate was prepared by dissolving cytosine trichloroacetate[10] (1.23 g, 4.5 mmol) and cytosine [4-aminopyrimidin-2(1*H*)-one] (0.5 g, 4.5 mmol) in hot water (50 $cm^3$). The mixture was chilled and allowed to stand at room temperature overnight. The precipitates were filtered off and washed with water. Yield 1.6 g (92.0%), m.p. 140—150 °C (decomp.) (Found: C, 30.9; H, 2.6; N, 21.7. $C_{10}H_{11}Cl_3N_6O_4$ requires C, 31.1; H, 2.9; N, 21.8%).

*Crystal Data.*—$(C_4H_5N_3O)_2$·$Cl_3C_2O_2H$, $M$ = 385.6. Triclinic, space group $P\bar{1}$, $a$ = 5.705(2), $b$ = 10.307(3), $c$ = 13.501(4) Å, $\alpha$ = 105.06(3), $\beta$ = 82.51(3), $\gamma$ = 98.87(3)°, $V$ = 754.0(4) Å$^3$, $Z$ = 2, $D_c$ = 1.70 g cm$^{-3}$, $F(000)$ = 392, $\mu$(Mo-$K_\alpha$) = 5.71 cm$^{-1}$, $\lambda$(Mo-$K_\alpha$) = 0.710 69 Å. Lattice parameters were determined by a least-squares refinement of the diffractometer angles for 15 reflections (2θ in range 18—23°). Crystal of dimensions 0.5 × 0.2 × 0.2 mm was used for the measurement of Bragg intensities on a Syntex $P2_1$ diffractometers with graphite-monochromatized Mo-$K_\alpha$ radiation in the 2θ range 3—48°. Integrated intensities were obtained by peak profile analysis according to Lehmann and Larsen.[11] The data were corrected for Lorentz and polarization effects, but not for absorption. Out of 2 272 measured reflections 1 550 had $I > 2\sigma(I)$ and those were observed. The structure was solved by direct methods with MULTAN 80.[12] Positional and

**Table 1.** Final fractional atomic co-ordinates

| Atom | *x* | *y* | *z* |
|---|---|---|---|
| Cl(1) | 0.150 5(3) | 0.217 8(2) | 0.229 1(1) |
| Cl(2) | 0.277 9(3) | 0.426 9(2) | 0.405 7(1) |
| Cl(3) | −0.106 6(3′) | 0.450 1(2) | 0.296 8(2) |
| C(1P) | 0.177(1) | 0.395 1(6) | 0.281 6(5) |
| C(2P) | 0.352(1) | 0.477 7(8) | 0.211 7(5) |
| O(1P) | 0.405(1) | 0.410 9(5) | 0.123 1(4) |
| O(2P) | 0.405 1(9) | 0.596 8(5) | 0.251 9(3) |
| N(1A) | −0.488(1) | 0.237 2(6) | 0.553 8(5) |
| C(2A) | −0.710(1) | 0.169 2(7) | 0.562 0(6) |
| O(2A) | −0.841 0(9) | 0.191 5(5) | 0.642 0(4) |
| N(3A) | −0.784 5(9) | 0.071 0(5) | 0.475 7(4) |
| C(4A) | −0.640(1) | 0.044 9(6) | 0.386 6(5) |
| N(4A) | −0.715(5) | −0.051 5(6) | 0.309 3(5) |
| C(5A) | −0.407(1) | 0.121 3(7) | 0.380 9(5) |
| C(6A) | −0.340(1) | 0.213 5(7) | 0.464 9(5) |
| N(1B) | 1.220(1) | 0.266 1(6) | −0.132 2(4) |
| C(2B) | 1.204(1) | 0.146 6(7) | −0.103 1(5) |
| O(2B) | 1.354 8(9) | 0.067 0(5) | −0.134 6(4) |
| N(3B) | 1.009(1) | 0.117 8(7) | −0.038 2(4) |
| C(4B) | 0.840(1) | 0.201 4(7) | −0.006 3(5) |
| N(4B) | 0.661(1) | 0.171 8(7) | 0.061 5(4) |
| C(5B) | 0.853(1) | 0.319 9(7) | −0.042 8(5) |
| C(6B) | 1.048(2) | 0.347 1(8) | −0.104 9(5) |
| H(5A) | −0.289 | 0.1043 | 0.3105 |
| H(6A) | −0.165 | 0.2707 | 0.4625 |
| H(5B) | 0.713 | 0.3847 | −0.0218 |
| H(6B) | 1.067 | 0.4370 | −0.1336 |
| H(1A) | −0.46(1) | 0.304(6) | 0.614(5) |
| H(3A) | −1.000 0(0) | 0.0(0) | 0.500 0(0) |
| H(41A) | −0.87(1) | −0.101(5) | 0.318(4) |
| H(42A) | −0.62(1) | −0.068(7) | 0.240(5) |
| H(1B) | 1.36(2) | 0.30(1) | −0.172(8) |
| H(3B)[a] | 0.99(2) | 0.04(1) | −0.031(8) |
| H(41B) | 0.64(1) | 0.081(7) | 0.073(5) |
| H(42B) | 0.57(1) | 0.230(7) | 0.080(5) |

[a] Disordered with a site occupancy factor of 0.5.

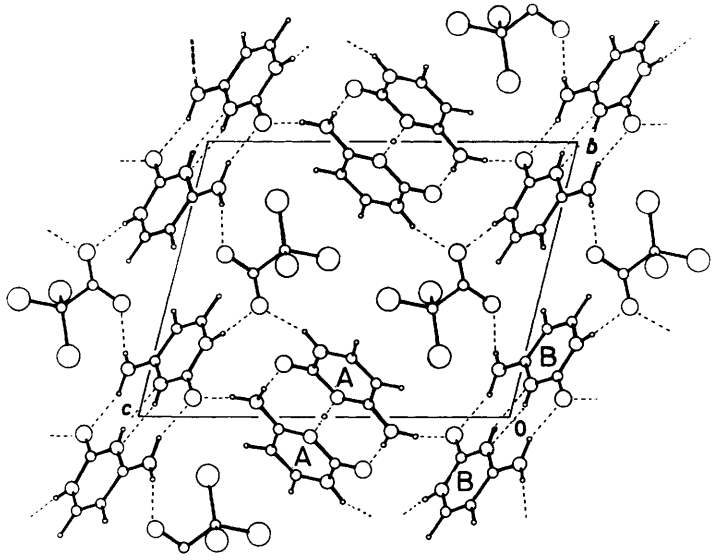


**Figure 1.** Crystal packing of the cytosine hemitrichloroacetate viewed along the *a* axis

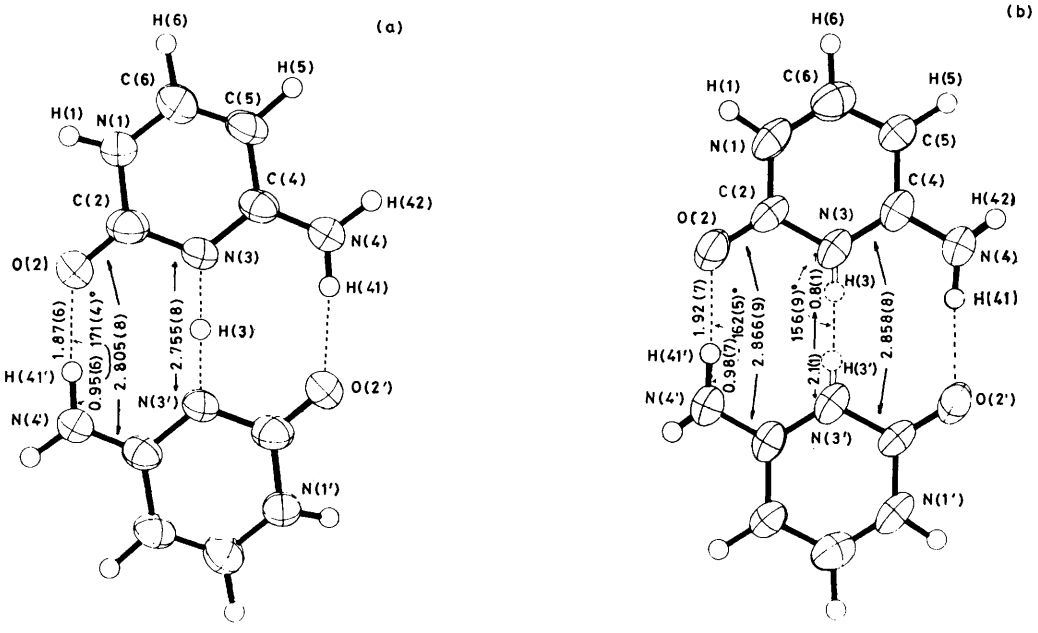


**Figure 2.** Hydrogen bond geometries in homoconjugated cations of cytosine hemitrichloroacetate: (a) base pair A; (b) base pair B

**Table 2.** Geometry of the hydrogen bonds

| D | H | A | D–H (Å) | D · · · A (Å) | H · · · A (Å) | D–H · · · A (°) |
|---|---|---|---|---|---|---|
| N(1A) | H(1A) | O(2P$^{i}$) | 0.94(6) | 2.787(8) | 1.87(6) | 166(4) |
| N(4A) | H(42A) | O(2P$^{ii}$) | 1.02(7) | 2.907(8) | 1.92(7) | 162(4) |
| N(4B) | H(42B) | O(1P) | 0.82(6) | 2.940(9) | 2.13(7) | 167(5) |
| N(1B) | H(1B) | O(2P$^{iii}$) | 0.95(10) | 2.921(8) | 1.99(10) | 166(6) |
| N(4A) | H(41A) | O(2A$^{iv}$) | 0.95(6) | 2.805(8) | 1.87(6) | 171(5) |
| N(4B) | H(41B) | O(2B$^{v}$) | 0.98(7) | 2.866(9) | 1.92(7) | 162(5) |
| N(3A) | H(3A) | N(3A$^{iv}$) | | 2.755(8) | | |
| N(3B) | H(3B) | N(3B$^{v}$) | 0.78(10) | 2.858(8) | 2.13(12) | 156(9) |

Symmetry codes:

| | | | |
|---|---|---|---|
| i | $-x$, | $1-y$, | $-z$ |
| ii | $1-x$, | $-y$, | $-z$ |
| iii | $2-x$, | $1-y$, | $1-z$ |
| iv | $-2-x$, | $-y$, | $1-z$ |
| v | $2-x$, | $-y$, | $-z$ |

**Table 3.** Bond lengths (Å)

| Cytosine | A | B |
|---|---|---|
| N(1)–C(2) | 1.354(9) | 1.373(10) |
| N(1)–C(6) | 1.364(10) | 1.342(11) |
| C(2)–O(2) | 1.222(9) | 1.240(9) |
| C(2)–N(3) | 1.396(9) | 1.369(9) |
| N(3)–C(4) | 1.357(9) | 1.347(9) |
| C(4)–N(4) | 1.306(9) | 1.330(9) |
| C(4)–C(5) | 1.441(9) | 1.420(10) |
| C(5)–C(6) | 1.332(10) | 1.344(11) |
| Trichloroacetate | | |
| C(1P)–Cl(1) | 1.770(7) | |
| C(1P)–Cl(2) | 1.775(6) | |
| C(1P)–Cl(1) | 1.768(6) | |
| C(1P)–C(2P) | 1.586(9) | |
| C(2P)–O(2P) | 1.216(9) | |
| C(2P)–O(1P) | 1.240(9) | |

**Table 4.** Bond angles (°)

| Cytosine | A | B |
|---|---|---|
| C(6)–N(1)–C(2) | 122.3(6) | 122.4(6) |
| O(2)–C(2)–N(1) | 121.8(6) | 122.3(6) |
| N(3)–C(2)–N(1) | 117.8(5) | 116.7(6) |
| N(3)–C(2)–O(2) | 120.5(6) | 121.0(5) |
| C(4)–N(3)–C(2) | 121.0(5) | 121.5(5) |
| N(4)–C(4)–N(3) | 118.5(5) | 118.6(6) |
| C(5)–C(4)–N(3) | 119.4(5) | 120.8(6) |
| C(5)–C(4)–N(4) | 121.1(6) | 120.6(6) |
| C(6)–C(5)–C(4) | 118.2(6) | 116.4(6) |
| C(5)–C(6)–N(1) | 121.4(6) | 122.0(6) |
| Trichloroacetate | | |
| Cl(2)–C(1P)–Cl(1) | 107.3(3) | |
| Cl(3)–C(1P)–Cl(1) | 109.2(3) | |
| Cl(3)–C(1P)–Cl(2) | 108.0(3) | |
| C(2P)–C(1P)–Cl(1) | 113.1(4) | |
| C(2P)–C(1P)–Cl(2) | 110.9(4) | |
| C(2P)–C(1P)–Cl(3) | 108.2(4) | |
| O(1P)–C(2P)–C(1P) | 114.4(6) | |
| O(2P)–C(2P)–C(1P) | 115.2(6) | |
| O(2P)–C(2P)–O(1P) | 130.3(6) | |

isotropic thermal parameters of non-hydrogen atoms were refined by full-matrix least-squares method to give an $R$ value of 0.13.* Further refinement, employing anisotropic thermal parameters, gave an $R$ value of 0.074. A difference Fourier map was calculated at this stage to obtain the positions of the hydrogen atoms. On this map the hydrogen atoms bonded to C(5), C(6), N(1), and N(4) of cytosines A and B appeared as pronounced maxima near their anticipated positions. Protons taking part in the N(3)···N(3′) hydrogen bond were located at the centre of symmetry in the cytosine base pair A and in a general position (with occupancy factor 0.5) in the cytosine base pair B. In the following cycles of refinement the positional and isotropic thermal parameters of the hydrogen atoms bonded to the nitrogen atoms were allowed to vary. The positions of the hydrogen atoms bonded to C(5) and C(6) of the cytosines were calculated from stereochemical considerations after each cycle, assuming C–H = 1.08 Å, and their temperature factors refined. Final values of $R$ and $R_w$ are 0.063 and 0.077, respectively. Throughout the refinement the quantity $\Sigma w(|F_o| - |F_c|)^2$ was minimized where $w^{-1} = \sigma^2(F) + 0.0003\,F^2$. Goodness-of-fit is 2.7. All shift/error values were less than 0.2 in the last cycle. The final difference Fourier map showed a maximum and minimum of 0.31 and −0.37 e Å$^{-3}$. The scattering factors used in the calculations were those included in SHELX 76.[13] Calculations were carried out using SHELX 76,[13] local programs written by Jaskólski[14] and programs PLUTO[15] and ORTEP[16] for drawings. Final positional parameters are collected in Table 1.

* *Supplementary data* (see section 5.6.7. of Instructions for Authors in the January issue). Thermal parameters have been deposited at the Cambridge Crystallographic Data Centre.

## Results and Discussion

The crystal packing shown down the $a$ axis is illustrated in Figure 1. Two types of centrosymmetric triply hydrogen-bonded cytosine base pairs can be distinguished. The base pairs of type B lie near $z = 0$ while those of type A lie near $z = 1/2$. The anion occupies a position between them. The structural elements are connected *via* hydrogen bonds into a three-dimensional network. In this structure the trichloroacetate anion is involved in three hydrogen bonds to cytosine. The geometry of all the hydrogen bonds is given in Table 2.

Figure 2 illustrates the interbase hydrogen bonding in $(\text{cytosine})_2\text{H}^+$ cations A and B. Owing to crystal symmetry requirements these cations have to be centrosymmetric. Base pairing between symmetry-equivalent hemiprotonated cytosines involves the N(3)···N(3′) hydrogen bond and two symmetry-equivalent hydrogen bonds between the exocyclic amino groups and the carbonyl oxygen atoms. The proton participating in the N(3)···N(3′) hydrogen bond can lie either at the centre of symmetry or be statistically distributed between the two positions, closer to one or the other N(3) atom in each pair. The Fourier difference map indicates that in the base pair B the proton is statistically distributed and therefore H(3B) was placed in a general position with an occupancy factor of 0.5. In the base pair A a rather sharp maximum of 0.5 e Å$^{-3}$ was located at the centre of symmetry and H(3A) was placed in a special position. The observed N(3)···N(3′) distances in both base pairs differ by 0.1 Å but lie in a typical range for asymmetric hydrogen bonds described by an asymmetric double minimum energy curve.[1,4b]

The bond lengths and valence angles of the hemiprotonated cytosine molecules A and B, and the trichloroacetate anion are given in Tables 3 and 4. All differences in dimensions between hemiprotonated cytosines are less than 3σ. Values of the Taylor–Kennard discriminant function[17] calculated for the cytosine rings from the C(2)–N(3)–C(4) and N(1)–C(2)–O(2) bond angles are −8 and −4 for base pairs A and B, respectively. Those values confirm that the geometry of the rings in the cytosine hemitrichloroacetate is intermediate between the geometry of the neutral cytosine (−54) and the monoprotonated one (54).

A further six structurally characterized systems contain triply hydrogen-bonded base pairs of cytosine derivatives.[4b] In (cytosin-5-yl)acetic acid no base stacking is observed and the proton in the N(3)–H···N(3′) hydrogen bond is asymmetrically located and equally distributed between the two sites near each of the symmetry-related N(3) atoms. This situation is analogous to that in base pair B in cytosine hemitrichloroacetate [Figure 3(b)]. In five other systems, including a 2:1 complex of cytosine with resorcylic acid, base-base overlap appears; the mean separation between the essentially planar base pairs is 3.22—3.36 Å. In those five systems the hydrogen bond N(3)–H···N(3′) is asymmetric and the proton is located near one of the cytosine residues. On the basis of this observation Kistenmacher *et al.*[4b] concluded that the asymmetry of cytosine interbase hydrogen bonding is stimulated to a large part by base-base stacking interactions. Our results deny this conclusion. In the cytosine hemitrichloroacetate structure base pairs A show a large overlap (the mean

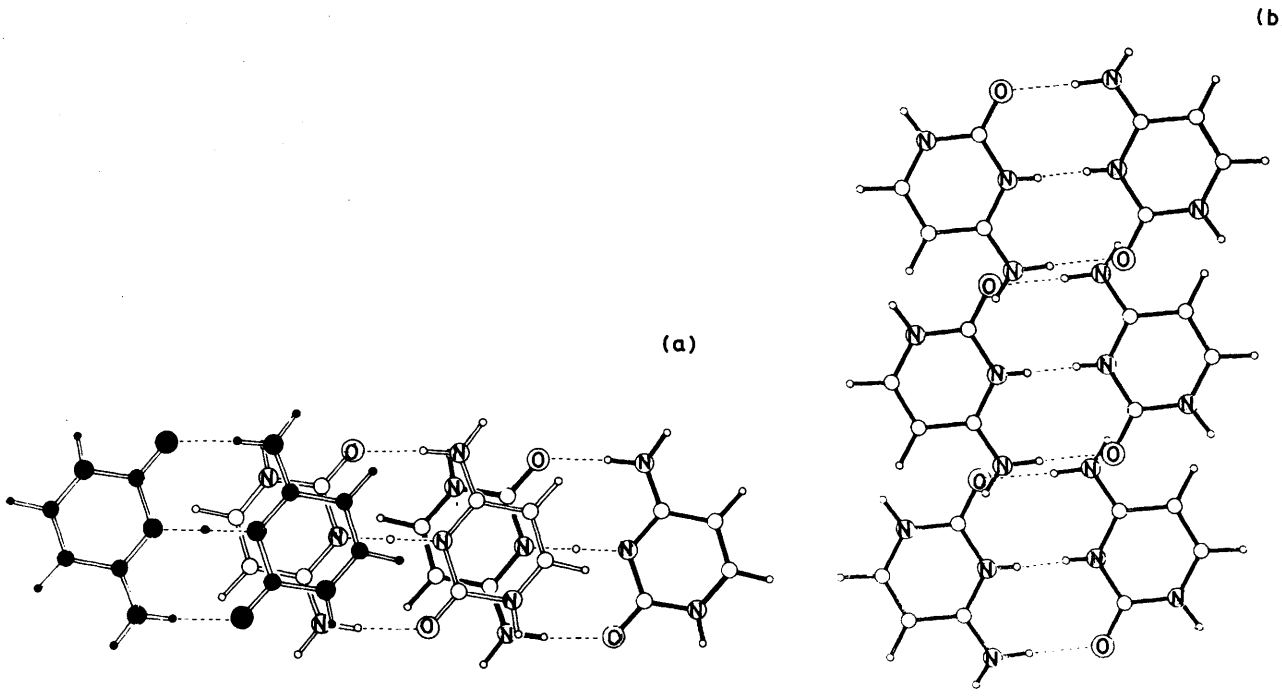


**Figure 3.** Base pairs overlap in cytosine hemitrichloroacetate: (a) base pair A; (b) base pair B

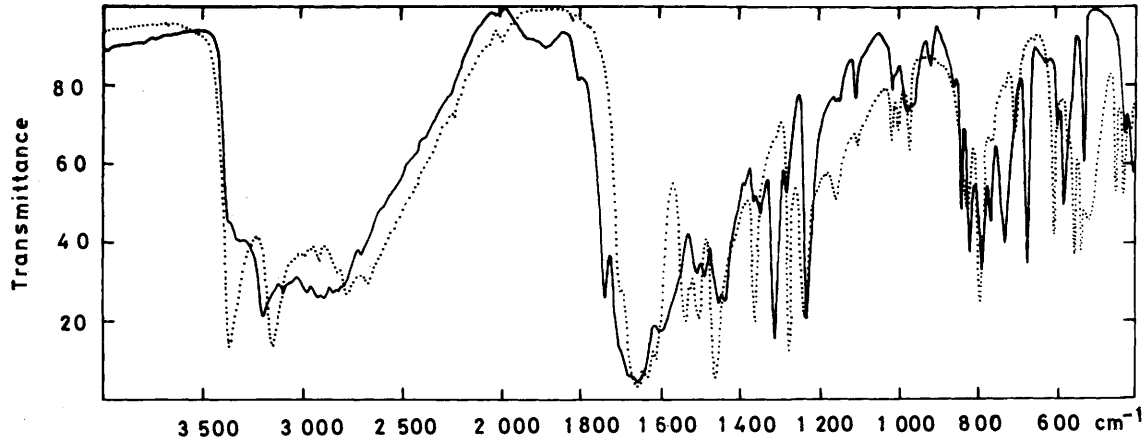


**Figure 4.** I.r. spectra of cytosine (· · · ·) and cytosine hemitrichloroacetate (——) in Nujol and Fluorolube

separation between base pairs is 3.35 Å) while the proton in the N(3)···N(3′) bridge is equally shared between the two cytosine molecules [Figure 3(a)].

*I.r. Spectra.*—The i.r. spectrum of homoconjugated cytosine trichloroacetate shows marked differences from that of pure cytosine (Figure 4). In particular the ring NH stretching band, ν(NH), which occurs as a broad band at *ca.* 2 750 $cm^{-1}$ in pure cytosine, is shifted to *ca.* 2 900 $cm^{-1}$ in the complex. The amine stretching modes, $\nu(NH_2)$, in cytosine (3 380 and 3 170 $cm^{-1}$) are less pronounced and slightly shifted in the hemitrichloroacetate.

The ν(C=O) and ν(C=C) band is also sensitive to changes in hydrogen bonding and the complex band indicates a pronounced structure. The shoulder at *ca.* 1 680 $cm^{-1}$ may be attributed to $\nu_{as}COO^-$ vibration.[18] In the spectrum of cytosine trichloroacetate the $\nu_{as}COO^-$ vibration appears as a resolved band. The i.r. spectrum of cytosine hemitrichloroacetate also shows a broad band at *ca.* 1 900 $cm^{-1}$; this is similar to that observed for homoconjugated complexes with aromatic nitrogen bases.[19] This absorption is not observed in cytosine trichloroacetate. The second characteristic band for the homoconjugated complex occurs at *ca.* 2 500 $cm^{-1}$. In cytosine hemitrichloroacetate this region is masked by broad absorption with its centre at 2 900 $cm^{-1}$. The i.r. spectrum confirms the structure determined by *X*-ray crystallographic analysis.

## Acknowledgements

We acknowledge financial support from the Polish Academy of Sciences (Project CPBP 01.12).

## References

1 S. N. Vinogradov, in 'Molecular Interactions,' eds. H. Ratajczak and W. J. Orville-Thomas, J. Wiley and Sons Ltd., Chichester, 1980, p. 179.
2 P. Barczyński, J. Koput, and M. Szafran, *J. Mol. Liq.*, 1983, **26**, 1, and references therein.
3 M. Szafran and Z. Dega-Szafran, *J. Mol. Struct.*, 1983, **99**, 189, and references therein.
4 (*a*) T. J. Kistenmacher, M. Rossi, J. P. Caradonna, and L. G. Marzilli, *Adv. Mol. Relaxation Interact. Processes*, 1979, **15**, 119; (*b*) T. J.

Kistenmacher, M. Rossi, C. C. Chiang, J. P. Caradonna, and L. G. Marzilli, *ibid.*, 1980, **17**, 113.
5 J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737.
6 (*a*) P. O. Löwdin, *Rev. Mod. Phys.*, 1964, **35**, 724; (*b*) P. O. Löwdin, *Biopolym. Symp.*, 1964, **1**, 293; (*c*) P. Romby, E. Westhof, D. Morae, R. Giege, C. Houssier, and H. Grosjean, *J. Biomol. Struct. Dyn.*, 1986, **4**, 193.
7 (*a*) D. M. Brown, D. M. Gray, M. H. Patrick, and R. L. Ratliff, *Biochemistry*, 1985, **24**, 1673; (*b*) D. M. Gray, T. Cui, and R. L. Ratliff, *Nucleic Acid Res.*, 1984, **12**, 7565.
8 C. Tamura, S. Sato, and T. Hata, *Bull. Chem. Soc. Jpn.*, 1973, **46**, 2388.
9 L. Sobczyk and Z. Pawełka, *J. Chem. Soc., Faraday Trans. 1*, 1974, **70**, 832.
10 B. Nogaj, B. Brycki, Z. Dega-Szafran, M. Szafran, and M. Maćkowiak, *J. Chem. Soc., Faraday Trans. 1*, 1987, **83**, 2541.
11 M. S. Lehmann and F. K. Larsen, *Acta Crystallogr., Sect. A*, 1974, **30**, 580.
12 P. Main, S. J. Fiske, S. E. Hull, L. Lessinger, G. Germain, J. P. Declercq, and M. M. Woolfson, MULTAN 80, A System of Computer Programs for the Automatic Solution of Crystal Structures from *X*-Ray Diffraction Data, University of York, England and Louvain, Belgium, 1980.
13 G. M. Sheldrick, SHELX 76, A Program for *X*-Ray Crystal Structure Determination, University of Cambridge, 1976.
14 M. Jaskólski, Collected Abstracts of the Fourth Symposium on Organic Crystal Chemistry, Poznań, September 1982, pp. 70—71; ed. Z. Kałuski, A. Mickiewicz University, Poznań, Poland.
15 W. D. S. Motherwell and W. Clegg, PLUTO, A Program for Plotting Molecular and Crystal Structures, University of Cambridge, England, 1978.
16 C. K. Johnson, ORTEP, Report ORNL-5138, Oak Ridge National Laboratory, Tennessee, 1976.
17 R. Taylor and O. Kennard, *J. Mol. Struct.*, 1982, **78**, 1.
18 E. Spinner, *J. Chem. Soc.*, 1964, 4217.
19 B. Brzeziński and G. Zundel, *J. Chem. Soc., Faraday Trans. 2*, 1976, **72**, 2127.

*Received* 23*rd July* 1987; *Paper* 7/1338