Proc. Indian Acad. Sci. (Chem. Sci.), Vol. 107, No. 3, June 1995, pp. 221–231.


# Proton transfer reactions of nucleic acid bases: A semiempirical molecular orbital study

DIVI VENKATESWARLU and R H DUNCAN LYNGDOH*

Department of Chemistry, North-Eastern Hill University, Bijni Complex, Shillong 793003, India

MS received 17 September 1994; revised 10 April 1995

**Abstract.** The possibilities open for tautomerism, for protonation and for deprotonation in the five nucleic acid bases are examined theoretically using the semiempirical AM1 SCF-MO methodology. The tautomers predicted to be the most stable, other than the usual forms, all involve proton shifts to adjacent sites. The sites predicted to be the most favourable for protonation are the $N^7$–G, $N^3$–A, $N^3$–C, $O^4$–T and $O^4$–U positions of guanine, adenine, cytosine, thymine and uracil respectively. The protons predicted to be the most acidic for each base are the $N^1$–G, $N^9$–A, $N^1$–C, $N^3$–T and $N^3$–U protons. These predictions accord well with the conclusions drawn from experimental work so far as assignments of acidic protons and basic sites for the particular bases are concerned. However, the relative feasibility of these reactions for the different bases is not well predicted by these gas-phase calculations.

**Keywords.** Nucleic acid bases; tautomerism; protic acid–base equilibria; molecular orbital calculations.

## 1. Introduction

Protic acid–base equilibria occupy an important place in physical organic chemistry (Stewart 1985), and are of great relevance for many biological processes as well. The relative feasibility of their theoretical study has led to many quantum chemical investigations on the subject (Pullman and Pullman 1971; Bender *et al* 1984; Kwiatkowski *et al* 1986; Ozment and Schmiedekamp 1992). This study concerns protic shifts and transfers in nucleic acid bases which are of fundamental significance for biological chemistry.

Protic changes in nucleic acid bases would call for three possibilities. The first is that of intramolecular proton shift leading to tautomer formation. The second is that of protonation of basic sites, while the third is that of abstraction of a proton. The question regarding tautomeric shift concerns the identity of the proton involved and the site to which it shifts. For protonation, the question concerns the position of the basic site to be protonated, while for deprotonation the inquiry pertains to the identity of the proton abstracted. For all these possibilities, the question of relative feasibility arises as well. Figure 1 gives the numbering systems for the five nucleic acid bases.

### 1.1 *Tautomerism in nucleic acid bases*

Intramolecular proton shifts in nucleic acid bases and nucleosides are regarded as a possible basis for spontaneous mutations of DNA (Topal and Fresco 1976). A finite

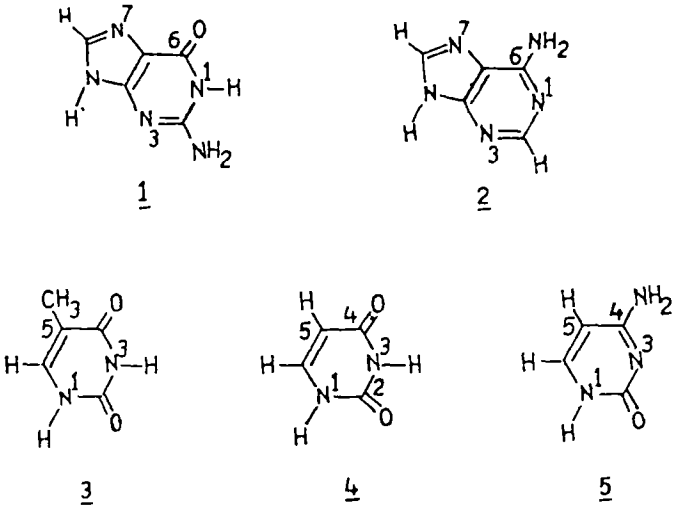

**Figure 1.** Numbering system for the five nucleic acid bases. 1 – Guanine; 2-adenine; 3-thymine; 4-uracil; and 5-cytosine.

possibility exists for tautomer formation (Singer and Grunberger 1983) at biological $\rho$H, the tautomers G*, A*, C* and T* of guanine, adenine, cytosine and thymine respectively, involving shifts in the hydrogen-bonding protons. This theoretical study considers a variety of avenues for tautomer formation in order to establish which tautomeric routes might be predicted as the most feasible.

### 1.2 *Protonation of nucleic acid bases*

Protonation of nucleophilic sites on the DNA base moieties is of relevance for the body of data on experimentally obtained acidic $pK_a$ values for these systems. Assignment of protonation sites has been proceeding along theoretical and mechanistic lines, taking cues from experimental observations as well. The $pK_a$ values quoted here were all taken in aqueous solvents at 25°C. It should be noted that assignments have changed over the years and the following paragraphs present current trends.

Shapiro (1968) has pointed to the evidence for the $N^7$ position as the most likely protonation site for 9-substituted guanine nucleosides based on the results of X-ray diffraction, ir and nmr studies (Jardetzky and Jardetzky 1960; Miles *et al* 1963; Iball and Wilson 1965). The acidic $pK_a$ values reported for guanine, guanosine, deoxyguanosine and for guanine residues in DNA are 3·3, 1·9, 2·27 and 3·45 respectively (Reinert and Weiss 1969; Christensen *et al* 1970; Benoit and Frechette 1985).

Transition state calculations led Pullman (1959) to identify the $N^1$-position as the most likely protonation site in adenine. X-ray crystallographic studies (Cochran 1951), calorimetric studies (Christensen *et al* 1967, 1970), the arguments of Zubay (1958) and the $^1H$ and $^{13}C$ nmr studies of Benoit *et al* (1985) by excess acidity methods have led to the general opinion now that the $N^1$ site is the most likely protonation site for adenine and its derivatives. Acidic $pK_a$ values reported for adenine, adenosine, deoxyadenosine and adenine residues in DNA are respectively 3·835, 3·50, 3·77 and 4·25 (Jordan *et al* 1956; Christensen and Izatt 1962; Reinert and Weiss 1969).

Calorimetric, nmr and absorption spectroscopy studies on cytosine protonation showed ionisation as occurring from the $N^3$ site (Katritzky and Waring 1963; Miles *et al* 1963; Ueda and Fox 1963; Christensen *et al* 1970). Though a tautomeric equilibrium involving microspecies had been proposed (Lewin and Humphreys 1966), proton nmr studies of the $Cu^{2+}$-cytidine system (Berger and Eichhorn 1971) eliminate binding to the $N^4$ site, inferring that the $N^3$–C position is the most probable protonation site. The acidic $pK_a$ values for cytosine, cytidine, 5′–dCMP and cytosine residues in DNA are reported as 4·58, 4·22, 4·44 and 4·50 respectively (Fox *et al* 1953; Cox 1966; Christensen *et al* 1967).

For uracil, thymine and their ribonucleosides, protonation in strongly acidic $\rho$H ($< 0{\cdot}5$) was earlier reported using spectrophotometric data (Cohn 1955). No clearcut assignment of the ionisation sites has been reported so far. Shapiro and Danzig (1972) have confirmed the $pK_a$ of uracil as $-3{\cdot}38$.

### 1.3 *Deprotonation of nucleic acid bases*

Deprotonation of DNA bases and their nucleosides pertains to data on basic $pK_a$ values for these systems. For alkylated DNA bases, it is connected to mechanisms for carcinogenesis and mutagenesis (Saffhill *et al* 1985; Lindahl *et al* 1988; Duncan and Davies 1989; Lyngdoh 1994). The precise identification of the protons involved here is uncertain in some cases. The $pK_a$ values reported here all pertain to aqueous solvent at 25°C.

Basic $pK_a$ values have been reported as 9·42, 9·25 and 9·26 for guanine, guanosine and deoxyguanosine, respectively (Zubay 1958; Katritzky and Waring 1963), where the first and perhaps the last involve the $N^1$ site, not the ribose moiety. The basic $pK_a$ value for the guanine residue in DNA is 11·6 (Suchorukow *et al* 1964), perhaps corresponding to ionisation from the $N^1$–G site.

Deprotonation for free neutral adenine has been assigned to the $N^9$ proton (Lewin 1964). It has been argued that proton loss from deoxyadenosine, although not reported, would have to be from the $N^6$-amino proton. The $pK_a$ values reported for adenine and adenosine are 9·259 and 12·35 respectively (Izatt *et al* 1966; Makar and Williams 1974), corresponding to proton loss from the $N^9$ site and sugar moiety respectively.

The proton ionisation $pK_a$ value of 12·15 (Christensen *et al* 1967) for cytosine has been attributed to the $N^1H$–$C^2O$ grouping, where proton loss could be either from a hydroxyl group or an acidic imino group, depending upon which tautomer is involved. For cytidine, the most facile proton loss has been inferred as occurring from the ribose moiety (Christensen *et al* 1970), for which a value of 12·5 has been reported.

The proton ionisation occurring with $pK_a$ of around 9·6 for uracil, uridine, thymine and thymidine has had various sites assigned to it, including the 2-hydroxyl group for neutral uracil and thymine (Shugar and Fox 1952). Both the $N^1$ and $N^3$ protons of uracil are involved in a simultaneous ionisation from both sites (Lewin 1964). $N^1$–deprotonation would be ruled out, of course, for the nucleosides, for which calorimetric results point to the first ionisation as occurring from the $N^3$ position. The basic $\rho K_a$ values for thymine and thymidine are 9·90 and 9·79 respectively (Christensen *et al* 1967), while those for uracil, uridine and deoxyuridine are 9·28, 9·17 and 9·3 respectively (Christensen *et al* 1970).

## 2. Theoretical methodology

The semiempirical AM1 SCF-MO method (Dewar *et al* 1985) as implemented in the MOPAC 3·1 package (Stewart 1983) was used to provide the wave function and molecular properties for all species studied, full geometry optimisation being conducted by the Davidon–Fletcher–Powell algorithm (Fletcher and Powell 1963; Davidon 1968). Theoretical indices for the three aspects of tautomer formation, protonation and deprotonation were used as described below.

The only index used for tautomer formation is the thermodynamic stability of the tautomer with respect to the most stable tautomer for each base. This was expressed in terms of the enthalpy change $\Delta H_s$, as well as by the change in free energy $\Delta G_s$ at 25°C, calculated with respect to the most stable tautomer. The corresponding equilibrium constant $K_t$ for each tautomeric transition was also calculated. The free energy of formation for each species was calculated by considering the contribution of the various degrees of freedom of the molecule to calculate the $T\Delta S$ term, which was added to the enthalpy term as follows:

$$\Delta G = \Delta H + T\Delta S. \tag{1}$$

Protonation facility was expressed in terms of the enthalpy of protonation in the gas phase, considering first a bare proton and then a water molecule as proton donor, giving the two enthalpy terms $\Delta H_p(p)$ and $\Delta H_p(w)$ respectively.

Facility of deprotonation was expressed in terms of the enthalpy terms $\Delta H_{dp}(p)$ and $\Delta H_{dp}(w)$ respectively, where the first term involved abstraction of a bare proton, and the second involved removal of the proton by a hydroxide anion.

## 3. Results and discussion

The sites involved in the tautomeric shifts and the corresponding energies reported (Singer and Grunberger 1983) are presented for five nucleic acid bases in table 1. Table 1 also gives the protonation sites and reported acidic $pK_a$ values for the five bases, as well as the deprotonation sites and reported basic $pK_a$ values, as obtained from experiment and from assignment following the results of experiment.

### 3.1 *Tautomerism in nucleic acid bases*

The five nucleic acid bases guanine, adenine, cytosine, thymine and uracil were studied

**Table 1.** Experimental data reported concerning protic shifts in nucleic acid bases. The tautomeric transitions and their energies ($E_t$ in *k*cal/mol); protonation sites (PS) assigned and corresponding acidic $pK_a$ values; proton abstraction sites (PAS) assigned and corresponding $pK_a$ values (references as in text)

| Base system | Tautomeric transition | | $E_t$ | PS | Acidic $pK_a$ | PAS | Basic $pK_a$ |
|---|---|---|---|---|---|---|---|
| Guanine | $N^1$–G | $O^6$–G | 4·27 | $N^7$–G | 3·3 | $N^1$–G | 9·42 |
| Adenine | $N^6$–A | $N^1$–A | 34·06 | $N^1$–A | 3·84 | $N^9$–A | 9·26 |
| Cytosine | $N^4$–C | $N^3$–C | 12·68 | $N^3$–C | 4·58 | $N^1/O^2$ | 12·16 |
| Thymine | $N^3$–T | $O^2/O^4$–T | 7·96 | ? | < 0·5 | $N^3$–T | 9·90 |
| Uracil | $N^3$–T | $O^2/O^4$–U | 7·59 | ? | − 3·38 | $N^3$–U | 9·51 |

with respect to all the possible tautomeric shifts involving the Watson–Crick protons, viz. the $N^1$–G, $N^6$–A, $N^4$–T, $N^3$–T and $N^2$–U protons. These are the protons believed to participate in the tautomeric shifts occurring spontaneously *in vivo*. The following shifts were studied: The $N^1$ proton of guanine to the $N^2$, $N^3$, $N^7$ and $O^6$ sites; the $N^6$ proton of adenine to the $N^1$, $N^3$ and $N^7$ sites; the $N^4$ proton of cytosine to the $N^3$ and $O^2$ sites, and the $N^3$ protons of thymine and uracil to the $O^2$ and $O^4$ sites. This study does not exhaust all tautomeric possibilities, since only shifts involving the Watson–Crick protons are included here. No note is taken of the proton shifts to or from the nitrogen atom bonded to the sugar moiety in the nucleosides (in an attempt to gauge the situation for the corresponding nucleosides).

The AM1 results are presented in table 2, giving the enthalpies $\Delta H_f$ and free energies $\Delta G_f$ of formation for all the tautomers studied, each being denoted in terms of the site on which the shifting proton is situated. The stabilities $\Delta H_s$ and $\Delta G_s$ of each tautomer with respect to the most stable one are also given, this being done to determine how the more easily calculated enthalpy differences compare with the more rigorously obtained free-energy differences.

**Table 2.** AM1 values for enthalpies $\Delta H_f$ and free energies $\Delta G_f$ of formation of nucleic acid bases and their tautomers, including their stabilities $\Delta H_s$ and $\Delta G_s$ with respect to the most stable tautomer in each case (all values in kcal/mol), and values of the equilibrium constants $K_t$ for the tautomeric transitions based on calculated free energy differences at 25°C.

| Tautomer* | $\Delta H_f$ | $\Delta H_s$ | $\Delta G_f$ | $\Delta G_s$ | $K_t$ |
|---|---|---|---|---|---|
| *Guanine* | | | | | |
| $N^1$–G | 48·82 | 0·00 | 22·16 | 0·00 | — |
| $N^2$–G | 91·48 | 42·65 | 64·21 | 42·05 | $1{\cdot}44 \times 10^{-31}$ |
| $N^3$–G | 62·98 | 14·15 | 35·88 | 13·72 | $8{\cdot}65 \times 10^{-11}$ |
| $N^7$–G | 73·15 | 24·33 | 46·61 | 24·25 | $1{\cdot}64 \times 10^{-18}$ |
| $O^6$–G | 53·47 | 4·64 | 26·75 | 4·59 | $4{\cdot}30 \times 10^{-4}$ |
| *Adenine* | | | | | |
| $N^6$–A | 86·79 | 0·00 | 61·22 | 0·00 | — |
| $N^1$–A | 100·37 | 13·58 | 75·44 | 14·22 | $3{\cdot}72 \times 10^{-11}$ |
| $N^7$–A | 129·35 | 42·56 | 104·40 | 43·18 | $2{\cdot}14 \times 10^{-32}$ |
| $N^3$–A | 110·77 | 23·98 | 85·36 | 24·14 | $1{\cdot}97 \times 10^{-18}$ |
| *Cytosine* | | | | | |
| $N^4$–C | 2·72 | 0·00 | −21·59 | 0·00 | — |
| $N^3$–C | 6·19 | 3·47 | −18·09 | 3·50 | $2{\cdot}71 \times 10^{-3}$ |
| $O^2$–C | 24·41 | 21·69 | 0·13 | 21·46 | $1{\cdot}82 \times 10^{-16}$ |
| *Thymine* | | | | | |
| $N^3$–T | −61·01 | 0·00 | −87·52 | 0·00 | — |
| $O^2$–T | −40·79 | 20·22 | −67·39 | 20·13 | $1{\cdot}72 \times 10^{-15}$ |
| $O^4$–T | −43·34 | 17·67 | −69·30 | 18·22 | $4{\cdot}33 \times 10^{-14}$ |
| *Uracil* | | | | | |
| $N^3$–U | −53·86 | 0·00 | −77·52 | 0·00 | — |
| $O^2$–U | −33·29 | 20·57 | −57·12 | 20·40 | $1{\cdot}09 \times 10^{-15}$ |
| $O^4$–U | −35·76 | 18·10 | −59·42 | 18·10 | $5{\cdot}30 \times 10^{-14}$ |

*Each tautomer represented by site of attachment of proton

Formally, the equilibrium constant $K_t$ for a tautomeric transition from the stablest tautomer B to another tautomer B* in terms of the partition functions and the free energy difference (for the transition) may be expressed as.

$$K_t = Q/Q^* \times \exp(-\Delta G_t/RT), \tag{2}$$

where $Q$ and $Q^*$ are the partition functions for the tautomers B and B* respectively, $\Delta G_t$ being the free energy difference for the tautomeric transition. The AM1 partition functions were calculated for a number of representative cases using MOPAC version 3·1. It was found that, although the individual contributions to the partition functions varied from tautomer to tautomer, the net partition functions had identical values for any series of tautomers, so that (2) above reduces to

$$K_t = \exp(-\Delta G_t/RT). \tag{3}$$

Values of the equilibrium constants calculated this way are given in the last column of table 2, the temperatures being 25°C.

Using the relative values of the enthalpies and free energies of formation for the various tautomers, as well as the values of the equilibrium constants, the relative stability of the tautomers can be obtained. It is seen that the most stable tautomers predicted for each case are the ones conventionally depicted as naturally occurring under normal conditions of temperature, pressure and *p*H. These are the $N^1$–G, $N^6$–A, $N^4$–C, $N^4$–T and $N^3$–U tautomers of guanine, adenine, cytosine, thymine and uracil respectively, each tautomer being designated in terms of the basic site to which the concerned proton is attached. The stability ordering for the different tautomers in each case is given below, where the results obtained are the same whether using the $\Delta H_s$ or the $\Delta G_s$ index:

Guanine: $N^1$–G > $O^6$–G > $N^3$–G > $N^7$–G > $N^2$–G,
Adenine : $N^6$–A > $N^1$–A > $N^3$–A > $N^7$–A,
Cytosine: $N^4$–C > $N^3$–C > $O^2$–C,
Thymine: $N^3$–T > $O^4$–T > $O^2$–T,
Uracil : $N^3$–U > $O^4$–U > $O^2$–U.

From the above orderings, the most feasible tautomeric shifts predicted are the $N^1 \rightarrow O^6$ shift for guanine, the $N^6 \rightarrow N^1$ shift for adenine, the $N^4 \rightarrow N^3$ shift for cytosine, and the $N^3 \rightarrow O^4$ shifts for thymine and uracil. Thus the most favourable shifts predicted all involve shift of the concerned proton to the most adjacent basic site, which is in fact what is generally believed to be the situation in reality.

The stability ordering for the tautomers of both the purines follows the order $N^1 > N^3 > N^7$, where each tautomer is represented by the site of location of the concerned proton. The stability ordering for the tautomers of thymine and uracil both follow the order $N^3 > O^4 > O^2$, indicating that the $C^5$–methyl group of thymine would have no drastic effect on the relative stabilities of the tautomers. The charge-separated tautomers $N^7$–G and $N^2$–G of guanine and the $N^7$–A tautomer of adenine are the least stable for each case, as would be expected from elementary considerations.

The relative feasibility of the most probable tautomeric shift with respect to the base species is as follows: $N^4 \rightarrow N^3$ for C > $N^1 \rightarrow O^6$ for G > $N^6 \rightarrow N^1$ for A > $N^3 \rightarrow O^4$ for T > $N^3 \rightarrow O^4$ for U. The corresponding AM1 calculated free energies of tautomeric transition are 3·50, 4·59, 14·22, 18·22 and 18·10 kcal/mol for cytosine, guanine, adenine, thymine and uracil respectively. These values compare only broadly with those

obtained experimentally, which are 12·68, 34·27, 4·06, 7·96 and 7·59 Kcal/mol for cytosine, guanine, adenine, thymine and uracil respectively (Singer and Grunberger 1983). So while these gas-phase AM1 calculations do not accurately reproduce the experimental values of the tautomeric transition, correct predictions are made with regard to identification of the most stable tautomer as well as the most feasible tautomeric transition for each base system.

It may be noted that the relative stabilities from the view-point of free energy compare very well with those from that of enthalpy, even though the individual free energies and heats of formation differ greatly. This is presumably because the entropy terms cancel out approximately. This observation justifies the convenience introduced by substituting free energy change by enthalpy change for these cases as well as for the rest of the cases covered in this study.

### 3.2 *Protonation of nucleic acid bases*

The protonation at the following basic sites in the five nucleic acid bases was studied: $N^2$, $N^3$, $N^7$ and $O^6$ of guanine; $N^1$, $N^3$, $N^6$ and $N^7$ of adenine; $O^2$, $N^3$ and $N^4$ of cytosine; and $O^2$ and $O^4$ of thymine and uracil. The atoms attached to the sugar in the corresponding nucleosides were not considered. Table 3 furnishes the AM1 calculated values of the various theoretical indices for protonation facility, viz. the $\Delta H_p(p)$ and $\Delta H_p(w)$ indices for protonation at various sites on the five nucleic acid bases.

Basicity being primarily a thermodynamic concept, chief recourse was made to the enthalpy index $\Delta H_p(w)$. For each base, the ordering for basicity of the sites as predicted

**Table 3.** AM1 values for indices of ease of protonation of nucleic acid bases ($\Delta H_p(p)$ and $\Delta H_p(w)$ in kcal/mol) together with reported values of molecular electrostatic minima (MEP).

| Species | $\Delta H_p(p)$ | $\Delta H_p(w)$ | *MEP* |
|---|---|---|---|
| *Guanine* | | | |
| $N^2$–G | −189·53 | −26·53 | −2·0 |
| $N^3$–G | −211·29 | −48·30 | −64·0 |
| $N^7$–G | −222·08 | −59·08 | −91·0 |
| $O^6$–G | −210·35 | −47·35 | −75·0 |
| *Thymine* | | | |
| $O^2$–T | −196·24 | −33·24 | −50·2 |
| $O^4$–T | −203·59 | −40·59 | −54·6 |
| *Adenine* | | | |
| $N^6$–A | −203·65 | −40·65 | −24·8 |
| $N^1$–A | −220·70 | −57·70 | −77·0 |
| $N^7$–A | −212·78 | −49·78 | −69·4 |
| $N^3$–A | −221·31 | −58·31 | −78·3 |
| *Cytosine* | | | |
| $N^4$–C | −195·63 | −32·63 | −13·7 |
| $O^2$–C | −214·46 | −51·46 | −84·8 |
| $N^3$–C | −225·67 | −62·67 | −93·5 |
| *Uracil* | | | |
| $O^2$–U | −195·32 | −32·32 | — |
| $O^4$–U | −201·97 | −38·97 | — |

by this index is given as follows:

Guanine: $N^7$–G > $N^3$–G ≃ $O^6$–G > $N^2$–G,
Adenine : $N^3$–A ≃ $N^1$–A > $N^7$–A > $N^6$–A,
Cytosine: $N^3$–C > $O^2$–C > $N^4$–C,
Thymine: $O^4$–T > $O^2$–T,
Uracil : $O^4$–U > $O^2$–U.

The most basic sites in each base as predicted by these gas-phase AM1 calculations are the $N^7$–G, $N^3$–A (close to $N^1$–A), $N^3$–C, $O^4$–T and $O^4$–U sites of guanine, adenine, cytosine, thymine and uracil respectively. These results follow the general trend from experiment for guanine and cytosine, as discussed above, although not quite so in the case of adenine. However, the closeness (less than 1 kcal/mol difference) between the $\Delta H_p(w)$ values for the $N^1$–A and $N^3$–A sites, suggests the possibility of competition. For thymine and uracil, with no experimental confirmation of protonation site, the results predict the $O^4$ sites as the most likely. Note that the molecular electrostatic potential minima (MEP) values as obtained by Bonnacorsi *et al* (1972, 1975) predict almost the same relative orderings for basicity as these AM1 SCF-MO $\Delta H_p(w)$ index values.

The ring nitrogens are predicted to be more basic than the exocyclic amino nitrogens, as would be expected from the presence of two already-bonded hydrogens in the latter. Protonation at an amino nitrogen is predicted as being least favourable among all sites for guanine, adenine and cytosine. The environment provided by the $O^6$ atom is guanine makes the $N^7$–G site particularly basic as compared to the $N^7$–A site in adenine. Likewise, the $O^6$–G site is rendered more basic by the imidazole nitrogen than the other oxygen sites on cytosine, thymine and uracil, as seen from a comparison of the enthalpy indices.

Table 4 summarises these findings by presenting the most basic site in each of the 5 bases as predicted here, coupled with the values of the $\Delta H_p(w)$ index for ease of protonation and the experimentally obtained $pK_a$ values (along with their assignments of ionisation sites). The predicted order of basicity is thus seen to be C > G > A > T > U, the reverse of which would point to the predicted ordering for the acidic $pK_a$ values of the bases. The experimental ordering for the $pK_a$ values is U > T > G > A > C, indicating some degree of positive correlation between the theoretical gas-phase proton acidities and the solvent-phase $pK_a$ values. The correlation might have been quantitatively better if recourse was made incorporating models for the solvent phase, as has been done for deoxyguanosine (Ford and Wang 1993). Table 4 also compares the $\Delta H_p(w)$ index with calculated values of the molecular

**Table 4.** Theoretical data for relative basicity of basic sites in nucleic acid bases compared with $pK_a$ (acidic) values and their protonation site assignments ($\Delta H_p(w)$ and MEP values in kcal/mol).

| Base | Site | $\Delta H_p(w)$ | MEP | $pK_a$ | Assignment |
|---|---|---|---|---|---|
| G | $N^7$–G | −59·08 | −91·0 | 3·30 | $N^7$–G |
| A | $N^3$–A | −58·31 | −78·3 | 3·84 | $N^1$–A |
| C | $N^3$–C | −62·67 | −93·5 | 4·58 | $N^3$–C |
| T | $O^2$–T | −40·59 | −50·2 | < 0·50 | $O^2/N^1$–T |
| U | $O^2$–U | −38·97 | — | −3·38 | $O^2/N^1$–U |

**Table 5.** AM1 calculated indices for facility of deprotonation of nucleic acid bases ($\Delta H_{dp}(p)$ and $\Delta H_{dp}(w)$ values in kcal/mol).

| Proton | $\Delta H_{dp}(p)$ | $\Delta H_{dp}(w)$ |
|---|---|---|
| *Guanine* | | |
| $N^1$ G | 335·84 | −74·98 |
| $N^9$ G | 337·28 | −72·54 |
| $N^2$ G | 346·03 | −64·79 |
| *Adenine* | | |
| $N^6$ A | 350·00 | −60·82 |
| $N^9$ A | 332·89 | −77·93 |
| *Cytosine* | | |
| $N^1$ C | 338·87 | −71·95 |
| $O^2$–C* | 337·14 | −73·68 |
| $N^4$ C | 351·49 | −59·33 |
| *Thymine* | | |
| $N^1$ T | 327·94 | −82·86 |
| $O^2$ T* | 312·62 | −98·20 |
| $N^3$ T | 344·09 | −66·73 |
| *Uracil* | | |
| $N^1$ U | 328·39 | −82·43 |
| $O^2$ U* | 319·07 | −91·75 |
| $N^3$ U | 344·48 | −66·34 |

*The $O^2$ protons here for the pyrimidines pertain to the less stable tautomer in each case

electrostatic minima (MEP) for each site in the solitary base obtained by Bonnaccorsi *et al* (1972, 1975). These are seen to furnish the same predictions concerning relative basicity as the $\Delta H_p(w)$ index.

### 3.3 *Deprotonation of nucleic acid bases*

The protons considered for deprotonation are as follows: guanine ($N^1$, $N^2$ and $N^9$), adenine ($N^6$ and $N^9$), cytosine ($N^1$, $O^2$ and $N^4$), thymine and uracil ($N^1$, $O^2$ and $N^3$). The pyrimidine systems incorporate the $O^2$–C, $O^2$–T and $O^2$–U tautomers for study as well, since there is some question of involvement of these species as microspecies in the equilibria. Table 5 presents the AM1 calculated indices for the feasibility of abstraction of these protons in the base systems, viz. the $\Delta H_{dp}(p)$ and $\Delta H_{dp}(w)$.

The order of proton acidities is predicted as follows by the $\Delta H_{dp}(p)$ and $\Delta H_{dp}(w)$ indices, each proton being identified by the site of attachment:

Guanine: $N^1 > N^9 > N^2$,
Adenine : $N^9 > N^6$,
Cytosine: $O^2 > N^1 > N^4$,
Thymine: $O^2 > N^1 > N^3$,
Uracil : $O^2 > N^1 > N^3$.

The above order of acidities could be interpreted as furnishing predictions for successive

proton ionisations in each case, where the protons attached to ring nitrogens are predicted to be more acidic than those attached to the exocyclic amino groups. The most acidic protons for each system (in the normal tautomeric form) are thus predicted to be as follows: $N^1$ (guanine), $N^9$ (adenine), $N^1$ (cytosine), and $N^3$ (thymine and uracil). This compares well with the assignments made for bases guanine, adenine, thymine and uracil, as discussed above, and lends further theoretical confirmation to these assignments. For tautomers of cytosine, thymine and uracil, the $O^2$ proton of the less stable tautomer is appreciably more acidic than the $N^1$ protons of the normal tautomers. This greater acidity of the $O^2$ protons is, however, balanced by the relatively small abundance of the less stable tautomer. Thus, this may be taken to confirm the involvement of proton ionisation from tautomers participating as microspecies as has been discussed earlier.

With respect to the base systems in their normal forms, the predicted order of proton acidities is $T \simeq U > A > G > C$. If these gas-phase calculations may be assumed to be applicable to the solvent phase, then the order of basic $pK_a$ values expected from these results would be $T \simeq U < A < G < C$. The order of reported aqueous phase $pK_a$ values (table 1) is, however, given by $A < G < U < T < C$, so that there is poor agreement between these gas-phase calculations and experimental results. Here, the use of gas-phase proton affinities for the experimental data set might lead to better correlations, but there are no data available. Calculations incorporating models for the solvent phase might also lead to better correlations with the $pK_a$ data set. So it emerges that these gas-phase AM1 calculations fare well for identification of proton ionisation sites, but are not as reliable for predicting quantitatively the ordering for solvent phase $pK_a$ values.

## 4. Conclusions

The most stable form in which the 5 nucleic acid bases are predicted to exist are the ones commonly known, while their most stable tautomeric counterparts involve shift of the hydrogen-bonding protons to adjacent basic sites. The sites predicted to be most favourable for protonation of the neutral species are the $N^7$–guanine, $N^3$–adenine (close to $N^1$), $N^3$–cytosine and the $O^4$ sites of thymine and uracil, which is corroborated to an extent by experimental findings. The sites predicted as most favourable for deprotonation in the neutral species are the $N^1$–guanine, $N^9$–adenine, $N^1$–cytosine and the $N^3$ sites of thymine and uracil, which compare well with assignments. While these gas-phase AM1 calculations perform well insofar as identification of sites for tautomerism, protonation and deprotonation go, they do not fare as well in furnishing quantitative predictions for the relative energetics of the processes in the solvent phase.

## Acknowledgements

D V is grateful to the University Grants Commission, Government of India, New Delhi and the North-Eastern Hill University, Shillong for the award of a fellowship, while RHDL is grateful to the Board of Research in Nuclear Sciences, Department of Atomic Energy (Government of India) for financial assistance. Thanks are also due to University Computer Centre, for allocating generous computer time, all calculations having been done on the VAX 3140 system.

## References


Bender M L, Bergeron R J and Komiyama M 1984 *The bioorganic chemistry of enzymatic catalysis* (New York: Wiley)
Benoit R L and Frechette M 1985 *Can. J. Chem.* **63** 3053
Benoit R L, Boulet D, Seguin L and Frechette M 1985 *Can. J. Chem.* **63** 1228
Berger N A and Eichhorn G L 1971 *Biochemistry* **10** 1857
Bonnaccorsi R, Scrocco E, Tomasi J and Pullman A 1972 *Theor. Chim. Acta (Berlin)* **24** 51
Bonnaccorsi R, Scrocco E, Tomasi J and Pullman A 1975 *Theor. Chim. Acta (Berlin)* **36** 339
Christensen J J and Izatt R M 1962 *J. Phys. Chem.* **66** 1030
Christensen J J, Rytting J H and Izatt R M 1967 *J. Phys. Chem.* **71** 2700
Christensen J J, Rytting J H and Izatt R M 1970a *Biochemistry* **9** 4907
Christensen J J, Rytting J H and Izatt R M 1970b *J. Chem. Soc. B* 1643
Cochran W 1951 *Acta Crystallogr.* **4** 81
Cohn W E 1955 In *The nucleic acids* (eds) E Chargraff and J N Davidson (New York: Academic Press) vol. 1 p. 217
Cox R A 1966 *Biochem. J.* **100** 146
Davidon W C 1968 *Comput. J.* **11** 406
Dewar M J S, Zoebisch E G, Healy E F and Stewart J J P 1985 *J. Am. Chem. Soc.* **107** 3902
Duncan R H and Davies G S 1989 *J. Theor. Biol.* **145** 345
Fletcher R and Powell M J D 1963 *Comput. J.* **6** 163
Ford G P and Wang B 1993 *J. Mol. Struct. (Theochem)* **283** 49
Fox J J, Cavalieri L F and Chang N 1953 *J. Am. Chem. Soc.* **75** 4315
Gerson D 1989 *QCPE Bull.* **9** 99
Iball J and Wilson H R 1965 *Proc. R. Soc.* **A288** 418
Izatt R M, Rytting J H, Hansen L D and Christensen J J 1966 *J. Am. Chem. Soc.* **88** 2641
Jardetzky C D and Jardetzky O 1960 *J. Am. Chem. Soc.* **82** 222
Jordan D O , Mathieson A R and Matty S 1956 *J. Chem. Soc.* 158
Katritzky A R and Waring A J 1963 *J. Chem. Soc.* 3046
Kwiatkowski J S, Zielinski T J and Rein R 1986 *Adv. Quantum Chem.* **18** 85
Lewin S 1964 *J. Chem. Soc.* 792
Lewin S and Humphreys D A 1966 *J. Chem. Soc. B* 210
Lindahl T, Sedgwick B, Sekiguchi M and Nakabeppu Y 1988 *Annu. Rev. Biochem.* **57** 133
Lyngdoh R H D 1994 *J. Biosci.* **19** 131
Makar G K R and Wiliams D R J 1974 *Inorg. Nucl. Chem.* **36** 1675
Miles H T, Bradley R B and Becker E D 1963a *Science* **142** 1569
Miles H T, Howard F B and Frazier J 1963b *Science* **142** 1458
Ozment J L and Schmiedekamp A M 1992 *Int. J. Quantum Chem.* **43** 783
Pullman B 1959 *J. Chem. Soc.* 1621
Pullman B and Pullman A 1971 *Adv. Heterocycl. Chem.* **13** 77
Reinert H and Weiss R 1969 *Z. Physiol. Chem.* **350** 1321
Saffhill R, Margison G P and Connor P J O' 1985 *Biochim. Biophys. Acta* **823** 111
Shapiro R 1968 *Prog. Nucleic Acid Res.* **8** 73
Shapiro R and Danzing M 1972 *Biochemistry* **11** 23
Shugar D and Fox J J 1952 *Biochim. Biophys. Acta* **9** 199
Singer B and Grunberger D 1983 *Molecular biology of mutagens and carcinogens* (New York: Plenum) p. 23
Stewart J J P 1983 *QCPE Bull.* **3** 43 (Quantum Chemistry Program Exchange, Program No. 455)
Stewart R 1985 *The proton: Applications to organic chemistry* (Orlando, FL: Academic Press)
Suchorukow B T, Poltew V I and Blumenfeld L 1964 *Abh. Deut. Akad. Wiss. Berl. Kl. Med.* 381
Topal M D and Fresco J R 1976a *Nature (London)* **263** 285
Topal M D and Fresco J R 1976b *Nature (London)* **263** 289
Ueda T and Fox J J 1963 *J. Am. Chem. Soc.* **85** 4024
Zubay G 1958 *Biochim. Biophys. Acta* **28** 644