Reaction Mechanisms of Formaldehyde with Endocyclic Imino Groups of Nucleic Acid Bases

Yan-Tyng Chang* and Gilda H. Loew

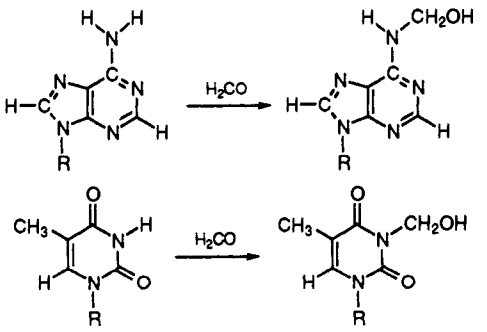
Contribution from the Molecular Research Institute, 845 Page Mill Road, Palo Alto, California 94304

*Received September 7, 1993**

Abstract: Four different possible mechanisms of formaldehyde reacting with the endocyclic imino groups of nucleic acid bases leading to hydroxymethylated nucleic acid adducts were investigated. The potential energy surface for each mechanism was characterized using the techniques of *ab initio* quantum chemistry. A water-assisted mechanism involving concerted interactions among the deprotonated nucleic acid base, formaldehyde, and a water molecule was found to be energetically most favorable. This mechanism also accounts for the experimentally observed features of these reactions, namely, that they are fast and reversible and both the forward and backward reactions are base catalyzed.

Introduction

Formaldehyde occurs naturally in the environment and is used heavily in a wide spectrum of industrial processes.[1] However, its potential carcinogenicity[2-5] and mutagenicity[6-10] evokes great health concerns[1,11] and detailed studies at the molecular level are needed to help the assessment and regulation of its application. Numerous experimental studies have been performed in the past to understand the interactions between formaldehyde and a variety of cellular nucleophiles. For nucleic acids, possible reaction products include DNA–protein cross-links,[12-14] DNA interstrand cross-links,[15] and nucleic acid adducts.[16-30] In addition to its role as a possible carcinogen and mutagen, formaldehyde has also been used frequently to probe the dynamical aspects of DNA and RNA structures.[25-30] Some of the important findings from these studies are the following: (1) DNAs with a high content of A-T base pairs are more susceptible to the attack of formaldehyde, (2) formaldehyde reacts more readily with single-stranded DNA than with double-stranded DNA, (3) the profile of the DNA helix–coil transition changes and the melting temperature decreases, and (4) irreversible denaturation of the DNA double helix may occur. All of these phenomena are related to the formation of nucleic acid–formaldehyde adducts. Although they have not been isolated experimentally, it has been suggested that the major adduct involves a hydroxymethyl ($-CH_2OH$) product, based on the similarity of its UV spectral properties with those of the corresponding methylamino compounds.[26,27] The possible sites of formaldehyde attack on the nucleic acids leading to these products include the exocyclic amino groups $-NH_2$ (such as N_6 of adenine, N_2 of guanine, and N_4 of cytosine) and the endocyclic imino groups >NH (such as N_3 of thymine, N_3 of uracil, N_1 of guanine, and N_1 of cytosine). Using adenine and thymine as examples, these two types of reaction are shown below:



The reason why the A-T rich regions are more susceptible to the attack compared to the G-C base pairs is presumably because the weaker interstrand hydrogen bonding interaction in the A-T base pairs allows local helix openings to occur more often in these regions, providing easier asscess to the amino or imino groups involved in the Watson–Crick base pairing.

Although the formation of the hydroxymethyl products from the reaction between formaldehyde and the exocyclic amino groups of nucleic acids is supported by sound evidence,[16-21,23,24,26] the occurrence of a similar reaction for the endocyclic imino groups is more controversial.[17,18,20,22,27] The reason for the controversy is the very different kinetic and spectral behavior observed between these two types of reaction which makes adduct formation at the endocyclic imino groups much more difficult to verify than that at the exocyclic amino groups. For example, the addition of formaldehyde to the exocyclic amino groups of DNA causes slow and irreversible denaturation of DNA. However, the reaction

* Abstract published in *Advance ACS Abstracts*, March 15, 1994.

(1) (a) *Environmental Health Criteria for Formaldehyde*; World Health Organization: Geneva, 1989. (b) *Formaldehyde: Health and Safety Guide*; World Health Organization: Geneva, 1991.
(2) Burdette, W. J. *Cancer Res.* **1951**, *11*, 555.
(3) Martin, C. N.; McDermid, A. C.; Garner, R. C. *Cancer Res.* **1978**, *38*, 2621.
(4) Swenberg, J. A.; Kerns, W. D.; Mitchell, R. I.; Gralla, E. J.; Pavkov, K. L. *Cancer Res.* **1980**, *40*, 3389.
(5) Ragan, D. L.; Boreiko, C. J. *Cancer Lett.* **1981**, *13*, 325.
(6) Kaplan, W. D. *Science* **1948**, *108*, 43.
(7) Alderson, T. *Nature* **1960**, *185*, 904.
(8) Nishioka, H. *Mutat. Res.* **1973**, *17*, 261.
(9) Auerbach, C.; Moutschen-Dahmen, M.; Moutschen, J. *Mutat. Res.* **1977**, *39*, 317.
(10) Obe, G.; Beek, B. *Drug Alc. Depend.* **1979**, *4*, 91.
(11) Perera, F.; Petito, C. *Science* **1982**, *216*, 1285.
(12) Wilkins, R. J.; Macleod, H. D. *Mutat. Res.* **1976**, *36*, 11.
(13) Obe, G.; Ristow, H. *Mutat. Res.* **1977**, *56*, 211.
(14) Ristow, H.; Obe, G. *Mutat. Res.* **1978**, *58*, 115.
(15) Huang, H.; Solomon, M. S.; Hopkins, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 9240.
(16) Fraenkel-Conrat, H. *Biochim. Biophys. Acta* **1954**, *15*, 307.
(17) Staehelin, M. *Biochim. Biophys. Acta* **1958**, *29*, 410.
(18) Grossman, L.; Levine, S. S.; Allison, W. S. *J. Mol. Biol.* **1961**, *3*, 47.
(19) Haselkorn, R.; Doty, P. *J. Biol. Chem.* **1961**, *236*, 2738.
(20) Lewin, S. *J. Chem. Soc. (B)* **1964**, 792.
(21) Lewin, S.; Barnes, M. A. *J. Chem. Soc. (B)* **1966**, 478.
(22) Eyring, E. J.; Ofengand, J. *Biochemistry* **1967**, *6*, 2500.
(23) Boedtker, H. *Biochemistry* **1967**, *6*, 2718.
(24) Utiyama, H.; Doty, P. *Biochemistry* **1971**, *10*, 1254.
(25) von Hippel, P. H.; Wong, K.-Y. *J. Mol. Biol.* **1971**, *61*, 587.
(26) McGhee, J. D.; von Hippel, P. H. *Biochemistry* **1975**, *14*, 1281.
(27) McGhee, J. D.; von Hippel, P. H. *Biochemistry* **1975**, *14*, 1297.
(28) McGhee, J. D.; von Hippel, P. H. *Biochemistry* **1977**, *16*, 3267.
(29) McGhee, J. D.; von Hippel, P. H. *Biochemistry* **1977**, *16*, 3276.
(30) Lukashin, A. V.; Vologodskii, A. V.; Frank-Kamenetskii, M. D.; Lyubchenko, Y. L. *J. Mol. Biol.* **1976**, *108*, 665.

0002-7863/94/1516-3548$04.50/0

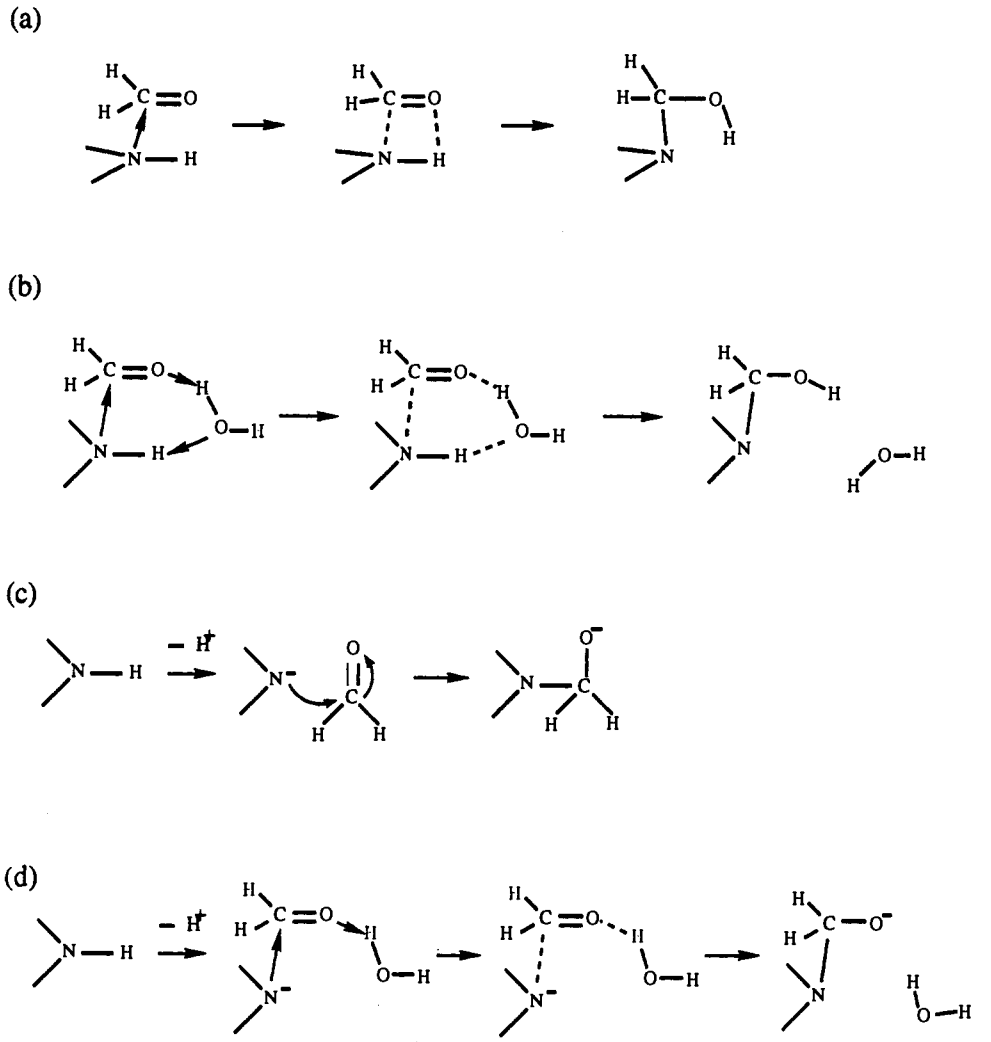
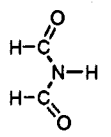


Figure 1. The four possible reaction mechanisms for the hydroxymethylation of the model thymine by formaldehyde. Mechanisms I–IV are shown in parts a–d, respectively.

with the endocyclic imino groups is instantaneous and reversible. Moreover, the imino reactions are characterized by a much smaller spectral change (i.e. the UV difference spectra between the unreacted and reacted nucleic acids) and a smaller equilibrium constant.[26,27] Furthermore, McGhee and von Hippel have shown that over the pH range from 5 to 9, the rate of reaction of formaldehyde with the exocyclic amino groups of adenine, cytosine, and guanine is pH independent, whereas both the forward and backward reactions involving the endocyclic imino groups of thymine and guanine are specific base catalyzed.[26,27]

While, as just described, a number of key features of the reaction of formaldehyde with endocyclic imino groups of nucleic acid bases have been experimentally determined, the mechanism of the reaction leading to these observations and the type of adduct formed are still unresolved. In this study, the techniques of theoretical chemistry are used to address these questions. Specifically, in this work, we have used these techniques to characterize and compare a number of plausible mechanisms in order to obtain an energetically favorable pathway that can also explain the experimental observations. To this end, four different mechanisms shown in Figure 1a–d were examined. The first two mechanisms involve neutral imino groups reacting with formaldehyde. Hence they are plausible candidates under neutral conditions and could also be pH independent. Mechanism I involves a direct nucleophilic attack of the imino nitrogen on the carbon atom of formaldehyde, with a sequential or concerted proton transfer from the imino nitrogen to the oxygen atom of formaldehyde. This mechanism is similar to that proposed for the reaction between formaldehyde and the exocyclic amino groups.[26] In the second mechanism, the same reaction is considered but with the intervention of a catalytic water molecule. The third mechanism, proposed by McGhee and von Hippel,[27] is one in which an anionic imino nitrogen, obtained by a fast deprotonation at higher pH, attacks the positive charged carbonyl carbon atom forming an anionic adduct. The last mechanism is similar to mechanism III, except that a water-assisted variation of it was examined. In these studies, a model compound as shown below was used. It contains an imino moiety as in thymine and preserves its nearest neighbors in the thymine pyrimidine ring which should be most important in the reaction.



For each of the four mechanisms, the energetics of the reaction was calculated and comparisons were made among them in order to determine which mechanism is energetically most feasible and can best explain the experimental observations.

Method of Calculation

The Gaussian 92 program[31] was used for all calculations. For each mechanism, the energy and geometry of all species involved in the reaction were optimized using standard energy gradient techniques. The potential energy surface was explored at the Hartree–Fock level of theory with the 3-21G basis set. Transition states for each mechanism were identified

Table 1. Absolute Energy (in hartrees) of All Species Involved in Mechanisms I–IV[a]

species	HF/3-21G	species	HF/3-21G
OH^-	−74.868 63	B1′	−468.919 94
H_2O	−75.585 96	TSB	−468.906 74
H_2CO	−113.221 82	B2	−468.958 21
R^-	−279.487 92[b]	C1	−392.738 10
RH	−280.084 80[b]	D1	−468.343 46[b]
R′H	−280.047 81	TSD	−468.338 27[b]
TST	−279.977 48	D2	−468.339 32[b]
RCH_2OH	−393.343 28	E1	−468.339 08[b]
RCH_2OH	−393.332 65[b]	TSE	−468.337 19[b]
TSP	−393.209 11[b]	E2	−468.346 41[b]
A1	−468.927 14[b]	F1	−468.358 12
TSA	−468.865 05[b]	TSF[c]	−468.348 72
A2	−468.937 68[b]	F2	−468.348 55
B1	−468.941 86		

[a] R represents the HCO–N–(HCO) moiety of the model thymine and R′H is the enol tautomer HCO–N–(HCOH) of the model thymine. Intermediate stable complexes are denoted as A1–F2, whose structures are shown in figures in this paper. Complexes that were constrained with C_s symmetry, with the plane of symmetry bisecting the "ring" of the model thymine are noted. [b] C_s symmetry constraint imposed. [c] TSF is a strucrture in the "inflection region" of the lower potential energy curve shown in Figure 11.

as those with only one negative eigenvalue of the force constant matrix of the structure.

Results and Discussion

The energies for all species involved in mechanisms I–IV are listed in Table 1. Only structural features that are important in the context of the reactions are presented in this paper. A complete structural description of all species can be obtained by request to the authors. As expected, both the optimized neutral (RH) and anionic model thymine (R^-) reactants shown in Figure 2, parts a and b, respectively, have planar geometries. For the optimized hydroxymethylated reaction product as shown in Figure 2c, an internal hydrogen bond is formed between the hydroxyl hydrogen and one of the two carbonyl oxygens. The strength of this H-bonding interaction was calculated to be 6.7 kcal/mol, from the energy difference between the fully optimized structure and a non-hydrogen-bonded structure constrained with C_s symmetry where the plane of symmetry is perpendicular to the plane of the model thymine. For comparison, the structures of the hydroxymethyl adducts of the full thymine and adenine bases were also calculated. As in the model, both products were found to contain an internal H bond. The energy of the reaction between the model thymine and formaldehyde to form this adduct was found to be −23.0 kcal/mol.

Mechanism I. The optimized geometry of the transition state (TSP) for the nucleophilic attack of the neutral imino nitrogen on the carbon atom of formaldehyde is shown in Figure 3. This transition state structure is a four-membered ring with a tetrahedral-like imino nitrogen environment. The relative position of the imino nitrogen is roughly perpendicular to the plane of the carbonyl group of formaldehyde with an N–C distance of 2.027 Å indicating incipient bond formation. The N–H bond is significantly bent from the plane of the model thymine with its distance enlongating from 1.003 Å in the reactant to 1.335 Å in the transition state indicating bond breaking while the H···O distance of 1.227 Å indicates the formation of an O–H bond. Only one imaginary frequency (i2200.0 cm^{-1}) was obtained, thus verifying that the structure found was a true transition state. Inspection of the eigenvector with the imaginary frequency further

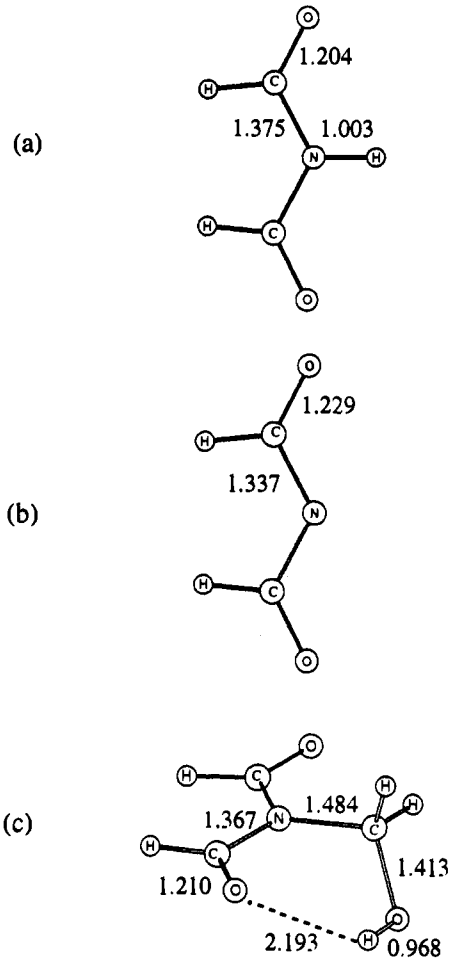


Figure 2. Optimized geometry of (a) neutral model thymine RH, (b) anionic model thymine R^-, and (c) hydroxymethylated product of the model thymine, RCH_2OH.

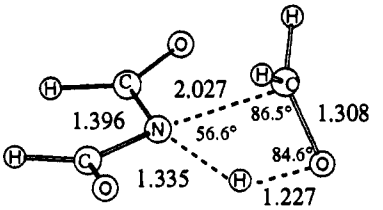


Figure 3. Optimized geometry of TSP, the true transition state of a direct nucleophilic addition–proton transfer mechanism between the model thymine and formaldehyde. This structure has a plane of symmetry bisecting the "ring" of the model thymine.

reinforced that the transition state contains characteristics of both N–C bond formation and proton transfer from the imino nitrogen to the carbonyl oxygen of formaldehyde. However, the activation energy is very high for both the forward and backward reactions (61.2 and 84.2 kcal/mol, respectively), and hence this mechanism does not appear to be a favorable pathway for adduct formation of formaldehyde with the endocyclic imino groups of nucleic acids.

Mechanism II. The possible participation of a catalytic water molecule in the formation of the hydroxymethylation adduct of the endocyclic imino group of nucleic acid bases by formaldehyde was investigated through a direct interaction among the neutral model thymine, formaldehyde, and a water molecule. As a first step, the three-component complexes along the reaction pathway were constrained to have C_s symmetry. Figure 4 shows the reactant (a), transition state (b), and the product (c) complexes obtained for this pathway. As shown in Figure 4a, in the energy minimized reactant geometry A1, the catalytic water forms a H-bonded bridge between the two reactants, with one H bond

(31) Frisch, M. J.; Trucks, G. W.; Head-Gordon, M.; Gill, P. M. W.; Wong, M. W.; Foresman, J. B.; Johnson, B. G.; Schlegel, H. B.; Robb, M. A.; Replogle, E. S.; Gomperts, R.; Andres, J. L.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Martin, R. L.; Fox, D. J.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Pople, J. *Gaussian 92, Revision A*; Gaussian, Inc.: Pittsburgh PA, 1992.

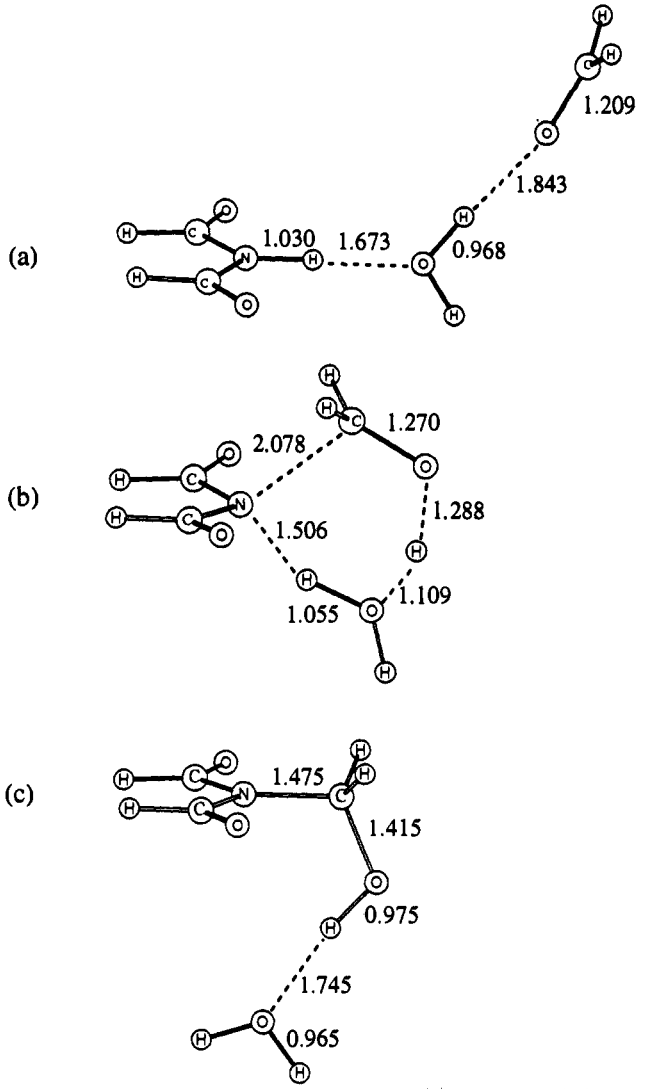


Figure 4. The progression of the structure of the neutral model thymine–formaldehyde–water complex from (a) optimized reactant structure, A1, to (b) a transition state, TSA, to (c) optimized adduct complex, A2. All three structures were constrained with C_s symmetry.

formed between the model thymine and the water modelcule and the other formed between water and formaldehyde. The transition state complex (TSA in Figure 4b) has a six-membered ring structure with the imino nitrogen in a tetrahedral-like environment. This state is similar to the four-membered-ring transition state found for the model thymine–formaldehyde interaction without the intervention of a water molecule (i.e., mechanism I). The eigenvector with the imiginary frequency (i965.4 cm^{-1}) clearly shows the characteristic of N–C addition with the water serving as both a proton acceptor from the imino nitrogen and a donor to the formaldehyde carbonyl oxygen. In the stable C_s constrained adduct complex, A2, shown in Figure 4c, the water molecule is still attached to the hydroxymethyl group with a hydrogen bond, but the N–C and C–O distances are very similar to the isolated adduct shown in Figure 2c.

The energetics of this C_s constrained mechanism is shown in Figure 5a. The energy of the reactant complex A1, the transition state complex TSA, and the product complex A2 is –21.7, +17.3, and –28.3 kcal/mol, respectively, relative to the energy of the three isolated reactants. Compared to the unassisted model thymine–formaldehyde reaction, mechanism I, the participation of the water molecule in this reaction leads to a much lower transition state energy and hence its participation as a catalyst in the reaction is highly supported. However, the forward activation energy of 39.0 kcal/mol relative to A1 and the backward activation energy of 45.6 kcal/mol relative to A2 of this water-

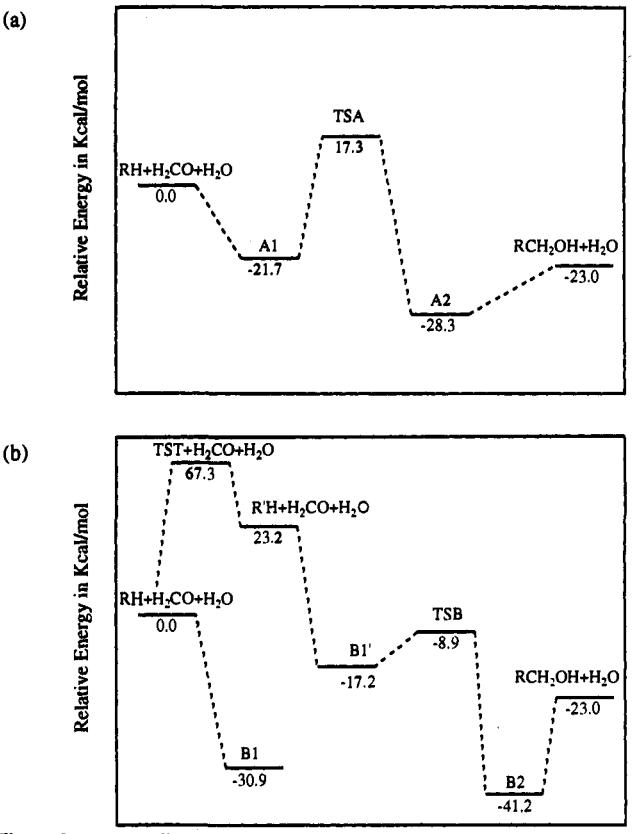


Figure 5. Energy diagrams for the species involved in mechanism II. (a) Pathway A describes a C_s constrained reaction. (b) In pathway B, all coordinates of the complexes were completely optimized. A description of some structural features for complexes involved in pathway A can be found in Figure 4, while Figure 6 shows those of complexes involved in pathway B.

assisted pathway to adduct formation are still too high to be consistent with the experimental observations.

The water-assisted mechanism just described was further explored by releasing the C_s symmetry constraint. This refinement was necessary for a more accurate description of the pathway since, although each of the three complexes was found to be a stationary point on the C_s constrained potential energy surface, one or two small imaginary frequencies corresponding to an out-of-plane motion (A″) were found in each of these complexes. Both the unconstrained optimized reactant (B1 as in Figure 6a) and product (B2 as in Figure 6d) complexes have one extra internal H bond compared to their constrained forms. Furthermore, as shown in Figure 5b, each complex is more stable than its constrained counterpart, by 9.2 kcal/mol for the reactant and by 12.9 kcal/mol for the product. However, while an optimized reactant (B1) and a product (B2) were obtained with the C_s constraint released, no transition state leading directly from one to the other could be found.

Therefore, a third water-assisted pathway involving the enol tautomer of the model thymine was explored. The energies of the optimized isolated enol reactant (R′H) and of the transition state (TST) between the keto and the enol forms are given in Table 1. The three-component enol reaction pathway was then characterized. The optimized geometry of the enol-like reactant complex (B1′) is shown in Figure 6b. It is composed of a planar 11-membered ring with three hydrogen bonds. A valid transition state for the N–C addition reaction, TSB (Figure 6c), was found. It has a nonplanar 8-membered ring and clearly indicates the N–C bond formation. An inspection of the reaction coordinate, i.e., the eigenvector with the imaginary frequency (i735.7 cm^{-1}), shows characteristics not only of N–C addition but also of one proton transfer from the "enol" hydrogen to the water molecule and another proton transfer from the water molecule to formaldehyde to form the adduct complex B2.

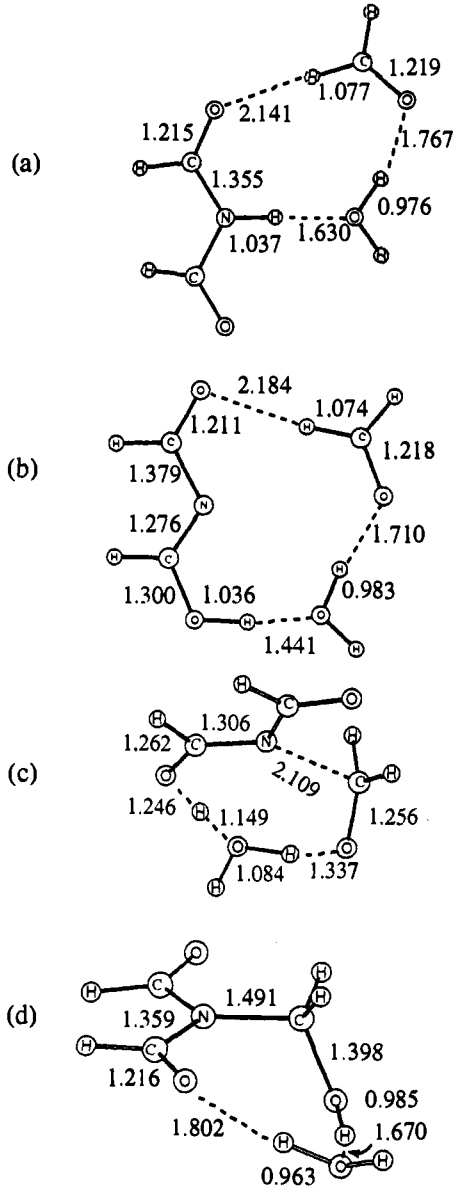


Figure 6. The structures of the unconstrained neutral model thymine–formaldehyde–water complexes: (a) a planar "keto" reactant complex B1, (b) a planar "enol" reactant complex B1′, (3) the transition state from the enol reactant to the final product, TSB, and (4) the final hydroxymethylated adduct complex, B2.

The energetics of the tautomerization transition state TST, the enol tautomer R′H and the complexes involved in this enol pathway relative to the isolated keto reactant (RH), formaldehyde, and the water reactants is also presented in Figure 5b. The small activation energy (8.3 kcal/mol) from B1′ to TSB makes this enol pathway more promising than the C_s constrained keto reaction with a 39.0 kcal/mol forward activation energy. However, the high activation energy (67.3 kcal/mol) involved in the tautomerization of the isolated model thymine makes the probability of the existence of the enol tautomer very low. Furthermore, the activation energy for the backward reaction in the enol pathway from B2 to TSB is 32.3 kcal/mol, much higher than that for the forward reaction. This result is not consistent with the observed reversibility of the reaction. These two factors taken together greatly reduce the probability of hydroxymethylation through the enol pathway.

Mechanism III. In this mechanism, possible at high pH, direct adduct formation between the deprotonated endocyclic imino group and formaldehyde was examined. According to the proposed mechanism of McGhee and von Hippel,[27] the anionic nitrogen attacks the partially positive charged carbon atom.

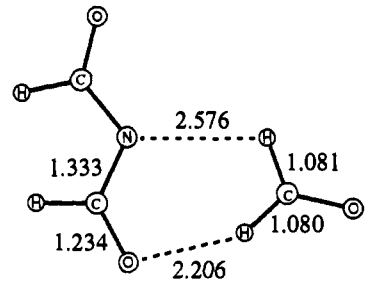


Figure 7. An optimized complex, C1, composed of the anion of model thymine and formaldehyde.

However, in our search for the anionic reaction product $>NCH_2O^-$, no stable species with R(N–C) ~ 1.5 Å could be found. Moreover, the potential energy surface between $>N^-$ and CH_2O is repulsive in the region of N–C distances ≤3.0 Å. Instead of a hydroxymethyl adduct formation, a H-bonded complex, C1, as shown in Figure 7 is formed, with two H bonds between the anionic model thymine (R^-) and formaldehyde. This complex is more stable than the reactants by 17.8 kcal/mol. Therefore, contrary to the proposed mechanism, direct interaction between an anionic imino group of nucleic acids and formaldehyde does not lead to a hydroxymethyl adduct.

Mechanism IV. While direct interaction between the anionic model thymine and formaldehyde does not lead to a stable adduct, it is possible that a water could catalyze the reaction, as in the corresponding reaction involving the neutral model thymine (mechanism II). The species involved in this reaction were also initially constrained to have C_s symmetry. As shown in parts a and d of Figure 8, two different reactant complexes, one cyclic and one linear, were considered. The cyclic reactant in Figure 8a with the water molecule close to both formaldehyde and the anionic model thymine is a six-membered ring with an N–C distance of only 2.660 Å. Figure 8 shows the progression from this optimized reactant complex D1 (a), to the transition state TSD (b), and finally to the adduct complex D2 (c). In the transition state TSD, the N–C distance is dramatically reduced to 1.884 Å, a clear indication of an incipient N–C bond formation, and the transition state is no longer cyclic. In the adduct complex D2, the N–C bond is further reduced to 1.643 Å and a H-bond interaction with the water molecule is maintained. This N–C distance is ~0.16 Å longer in D2 than in the isolated neutral hydroxymethylated product (Figure 2c). However, the potential energy becomes repulsive at smaller r(N–C) and no further stable complex could be found.

A second linear reactant complex, E1 (Figure 8d), with the water interacting only with the formaldehyde was also considered to form an adduct in this water-assisted anionic model thymine reaction with formaldehyde. Parts e and f of Figure 8 show the optimized transition state, TSE, and adduct E2 obtained from this reaction complex. As seen in these figures, N–C bond formation occurs via this pathway as well, with decreasing N–C distance of 2.467, 1.968, and 1.558 Å from reactant to transition state to the adduct. In adduct E2, there is a somewhat stronger interaction of the water with the carbonyl oxygen of formaldehyde than in D2, and the C–O distance is a little longer.

The energetics for these two pathways is shown in Figure 9. As indicated, the formation of the anionic model thymine from its neutral species in basic condition is very exothermic (–75.6 kcal/mol), consistent with the experimental observation that the forward reaction is base catalyzed. The energies of the two stable anionic reactant complexes, D1 and E1, are 30.0 and 27.2 kcal/mol relative to that of the three isolated reactant species. The forward activation energies for reaction paths D and E are only 3.3 and 1.2 kcal/mol, respectively. The activation energies for the backward reactions are also quite small, only 0.7 kcal/mol from D2 to TSD and 5.8 kcal/mol from E2 to TSE. These results

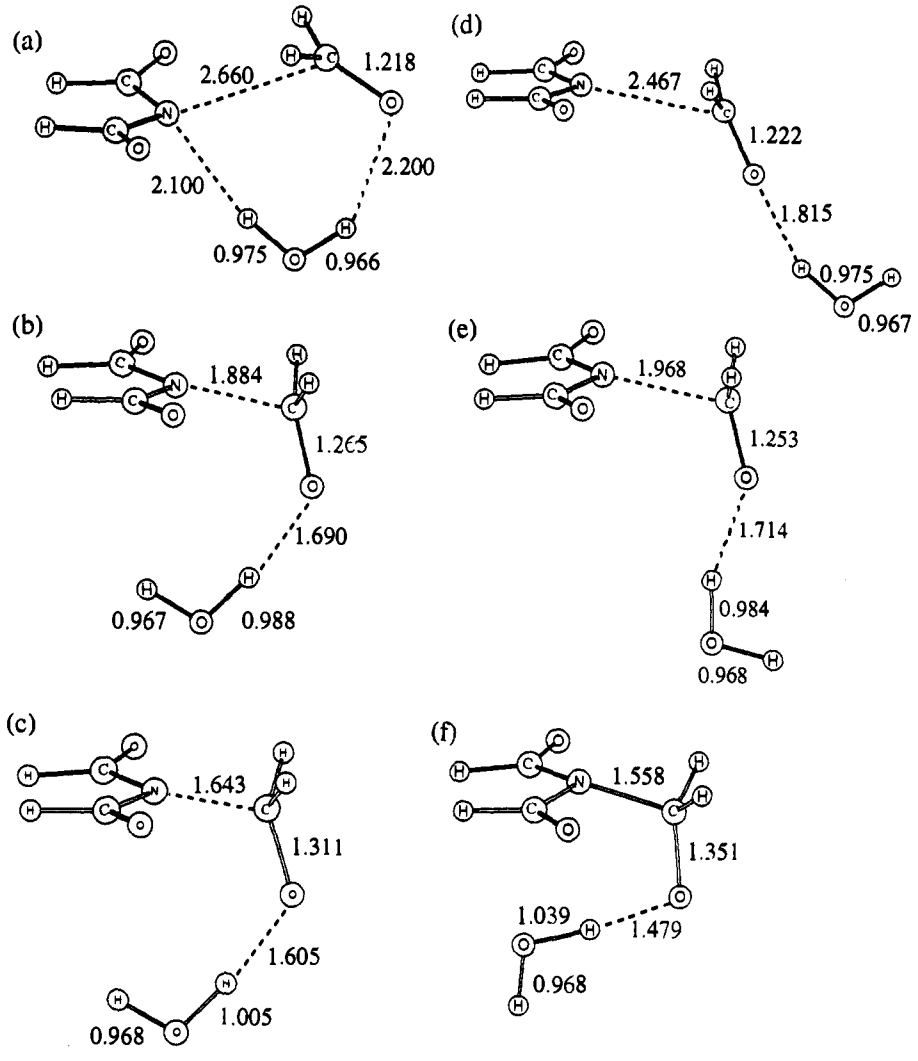


Figure 8. The progression of the structure of the anionic model thymine–formaldehyde–water complex along the reaction pathways D and E. C_s symmetry constraint was imposed. From a–c: D1, TSD, and D2. From d–f: E1, TSE, and E2.

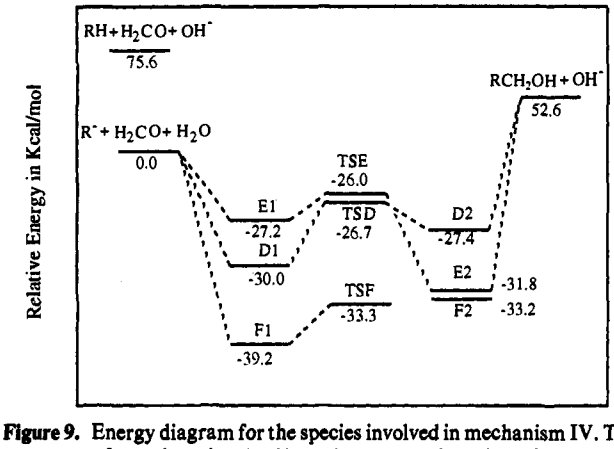


Figure 9. Energy diagram for the species involved in mechanism IV. The structures of complexes involved in pathways D and E, where C_s symmetry constraint was imposed, can be found in Figure 8. Those of pathway F, where all geometries were optimized without constraint, can be found in Figure 10.

are consistent with and can explain the observation that the endocyclic imino group hydroxymethylation is very fast and reversible.

Releasing the C_s symmetry constraint, only one stable reactant like complex F1, Figure 10a, and one product like complex F2, Figure 10c, were found. Complex F1 is, similar to complex D1, cyclic with two H bonds. However, the H bond between the water molecule and the model thymine imino nitrogen atom in D1 is replaced with one between water and one of the model thymine carbonyl oxygens. The N–C distance of 2.525 Å is a value between that of D1 (2.660 Å) and E1 (2.467 Å). On the other hand, the adduct complex F2 has a structure very similar to that of E2. Its N–C distance is 1.573 Å, a little longer than the 1.484-Å value of an isolated hydroxymethylated model thymine.

Unlike the C_s constrained pathways D and E, no true transition state could be found when the N–C addition was allowed to proceed from the reactant complex F1. Rather, using r(N–C) as a reaction coordinate, with all other geometric parameters optimized at each point, two potential energy curves were found as shown in Figure 11. The lower one proceeds from the reactant complex F1 and represents the forward reaction to adduct formation. The upper curve proceeds from the product complex F2 and represents the backward reaction to the reactants. As shown in the lower curve of this figure, the very small barrier obtained from the C_s constrained cyclic reactant D1 becomes an "inflection region" for $\sim 1.65 < r(\text{N–C}) < \sim 1.8$Å for its unconstrained counterpart. A structure, denoted as TSF, chosen from this region is shown in Figure 10b, with an energy only 5.9 kcal/mol higher than that of F1. Similar to pathways D and E, the potential curve becomes repulsive again when proceeding further for smaller N–C distances. As shown in the upper potential curve of Figure 11, adduct complex F2, (Figure 10c) is a local minimum with an energy only 6.0 kcal/mol higher than that of F1. With little or no energy barrier, as the N–C distance in the adduct F2 increases from 1.573 to >1.73 Å, the motion of the water molecule in F2 becomes very floppy and transformation to

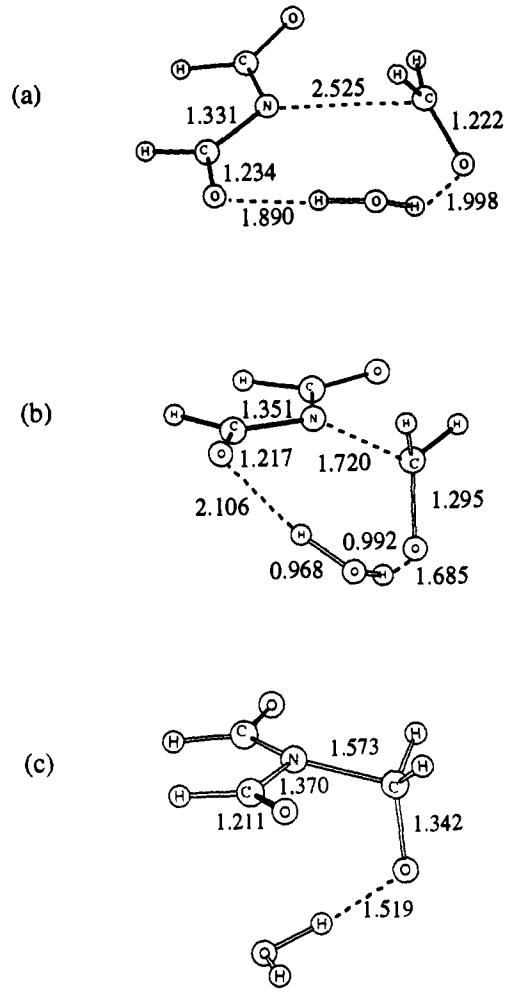


Figure 10. The unconstrained optimized structures of (a) the reactant complex F1, (b) an intermediate complex TSF in the "inflection region" of the potential curve shown in Figure 11, and (c) an adduct complex F2.

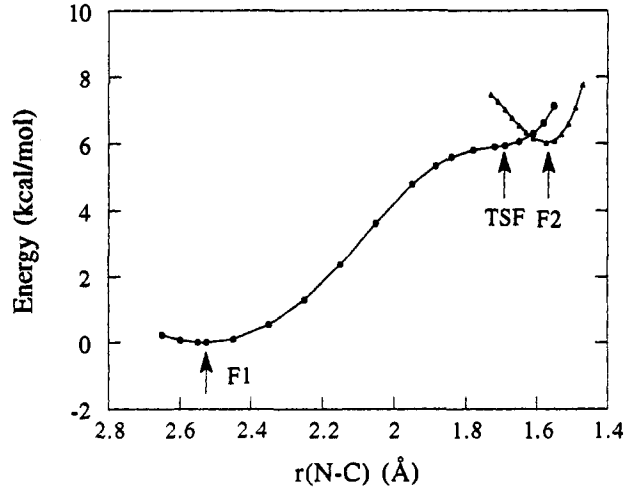


Figure 11. Potential energy curves along the N–C addition coordinate. The lower curve is obtained by starting with the stable reactant complex F1. No transition state exists along this curve. Instead, this curve was found to contain an "inflection region". The upper curve is obtained from starting with the geometry of the stable adduct complex F2.

complex TSF occurs, thus provided a facile pathway for the backward reaction.

The energies of these three complexes relative to the isolated reactants are also shown schematically in Figure 9. Each complex is more stable than their counterpart from the C_s constrained pathways D and E. The general features of this unconstrained reaction pathway are also consistent with and can explain the experimental observations.

The product complexes D2, E2, and F2 formed by the anionic model thymine with formaldehyde and water are best described as anionic adducts weakly H bonded to a water molecule. The possible formation of a neutral hydroxymethylation adduct by proton transfer from the water molecule to the oxygen atom of the formaldehyde moiety in these product complexes was thus subsequently examined. This step was probed by starting with structures D2, E2, and F2 and using the O···H distance as a reaction coordinate, while relaxing all other coordinates during geometry optimization. The result obtained was that no stable neutral hydroxymethylation adduct with O···H distance ~1.0 Å could be found. In fact, the potential curve for proton transfer to the anionic adduct from the water molecule is repulsive along this coordinate. As shown in Figure 9, the heat of reaction from D2, E2, and F2 to the isolated neutral hydroxymethylation adduct and a hydroxide ion is about 80 to 90 kcal/mol endothermic. These results strongly indicate that, in the presence of OH^-, a neutral hydroxymethylated product would form a deprotonated hydroxymethylation adduct weakly H bonded to a water molecule which would then quickly proceed in the reverse reaction to form the reactant complexes, thus accounting for the reversibility and base catalysis of the backward reaction.

Studies of all of the mechanisms described reveal a few important findings: (1) There does not appear to be a viable pathway to adduct formation in the reaction of neutral thymine nucleic acid bases with formaldehyde. (2) Water plays an important catalytic role in the reaction of the anionic imino group of the model thymine with formaldehyde in three ways, by (a) stabilizing the reactant complex, (b) allowing adduct formation, and (c) reducing the activation energies for both the forward and backward reactions. With regard to the first role, recall in mechanism III, without water as a catalyst, the stability of the reactant complex C1 is only 17.8 kcal/mol relative to its isolated reactants. By contrast, much more stable reactant complexes D1, E1, and F1 are formed, with water as an intrinsic part of the complexes. They are 30.0, 27.2, and 39.2 kcal/mol, respectively, more stable than the three isolated reactant species. With regard to the second role, remember that there is no indication of N–C addition in the unassisted complex C1. However, in the water-assisted complexes D1, E1, and F1, the proceeding of the formation of an N–C bond between the formaldehyde carbon atom and the anionic imino nitrogen is clear. The catalytic effect of the water molecule that aids product formation and reduces activation energy is to reduce the repulsion between the anionic thymine and formaldehyde and bring the nitrogen and carbon atom closer. The small forward activation energies and small or zero backward activation energies found in pathways D, E, and F are consistent with the experimental observations that the hydroxymethylation of endocyclic imino groups is very fast and reversible. (3) The reason why both the forward and backward reactions are base catalyzed can be understood from the energy diagram of Figure 9. It can be seen that the model thymine is readily deprotonated at high pH condition, since this reaction is exothermic by 75.6 kcal/mol. The anionic thymine can then quickly react with formaldehyde and a water molecule and form a stable reactant complex (i.e., D1, E1, or F1) with a small N–C distance (~2.5 Å), which leads to the anionic adduct (D2, E2, or F2) formation with a very low energy barrier. For the backward reaction, a neutral hydroxymethylation adduct will readily react with a nearby OH^- with large exothermicity (~80 to 90 kcal/mol) and hence lead to the formation of a complex with the characteristics of a deprotonated hydroxymethylated product, H bonded to a water molecule. This product can then quickly proceed in the reverse direction to form the reactant complexes.

Conclusion

In this work, four different mechanisms for the reaction of the imino nitrogen in a model thymine with formaldehyde were investigated in order to identify the most plausible one that could explain all of the experimental observations for the hydroxymethylation of the endocyclic imino groups of nucleic acid bases. None of the neutral or pH-independent mechanisms appear viable because of the large activation energies involved. Of the two base-catalyzed mechanisms, the direct unassisted N–C addition reaction between the anionic nucleic acid and formaldehyde is not likely because there is a ~25 kcal/mol repulsive potential when the formaldehyde carbon atom moves toward the anionic imino nitrogen atom of the nucleic acid and therefore no hydroxymethyl adduct can be formed. However, a base-catalyzed mechanism involving concerted interactions among the anionic nucleic acid, formaldehyde, and a water molecule is energetically favorable and also explains the fast, reversible, and base-catalyzed features for the covalent addition of formaldehyde to the endocyclic imino groups of nucleic acid bases. The results obtained from this theoretical work thus provide insights into the mechanism of the hydroxymethylation of the endocyclic imino group of nucleic acid bases by formaldehyde. They also indicate that under physiological conditions such as at pH ~7.0, this reaction is negligible and should not contribute to the carcinogenic and mutagenic potency of formaldehyde.

Acknowledgment. We thank Prof. Joshua Lederberg for stimulating our interest in this project. This work was supported by Environmental Protection Agency Grant No. CR818677-01-0. We also wish to acknowledge the EPA for the access of its Cray-YMP machine on which all calculations for this work were performed and to thank Dr. Krish Namboodiri for his technical help. The input from Prof. George Pack in the early stage of this project is also gratefully acknowledged.