Energy Increment Method Based on Quantum Chemical Results: A General Recipe for Approximative Prediction of Isomerization and Tautomerization Energies of Pyrimidine and Purine Nucleic Acid Bases and Related Compounds

Tae-Kyu Ha, Hans-J. Keller, Rok Gunde, and Hans-H. Gunthard*

Physical Chemistry Laboratory ETH Zentrum, CH-8092 Zurich, Switzerland

Received: December 1, 1998; In Final Form: May 3, 1999

On the basis of HF/6-31G(d,p) and MP2/6-31G(d,p) quantum chemical computations of fully optimized structures of all isomers of cytosine, isocytosine, uracil, thymine, adenine, guanine, xanthine, and many model compounds, a system of additive energy increments has been derived, from which energies of conversion of geometric isomers (conformers) and tautomers of these nucleic acid bases may be reconstructed within approximately 0.5 kcal/mol. The increments are associated with specific structural fragments perceptible in the conventional structural formula. Physically, they correspond to repulsions between H atoms of −OH, $-NH_2$, and ═NH substituents and H atoms bound to C or N ring atoms and to attractions between H atoms of such substituents and sp^2 lone electron pairs localized at N atoms of the ring systems or of imino substituents. Tautomerization energies associated with displacements of H atoms among ring N atoms were included in the estimation process of the increments. Keto−enol and amino−imino tautomerization energies may be estimated from total electronic energies and increments of conformer conversion energies. The estimates also approximately apply for $\Delta U°$ (0) and $\Delta U°$ (T_0) (free molecules). The extended data set allowed some global structural features of the nucleic acid bases (such as formally aromatic 6-rings), which influence certain increments specifically, to be discerned. The set of increments presented may serve as a basis of a rather general recipe for predictive estimation of conformer and tautomer conversion energies of a wide range of hydroxy- and amino-substituted aromatic nitrogen bases and of effects of geometric isomerism on energetics of intermolecular H bonding involving −OH, $-NH_2$, and ═NH substituents and protonated keto groups.

1. Introduction

In previous papers, a system of energy increments based on quantum chemistry has been reported, from which the apparently erratic energies of conversion of geometric isomers of the pyrimidine nucleic acid bases cytosine,[1−3] uracil,[1−3] thymine,[2,3] and isocytosine[4] and of the simplest purine nucleic acid base adenine[5] may be reproduced fairly reliably. The increments may be perceived directly from the structural formula, either as repulsive or attractive interactions. Repulsions were associated with interactions between H atoms of −OH, ═NH, or $-NH_2$ substituents and H atoms bound to C− or N atoms of the ring structure. Attractions on the other hand were assigned to interactions of H atoms of −OH, ═NH, or $-NH_2$ substituents and sp^2 lone electron pairs localized at neighboring N atoms of the ring structure or at imino substituents. Determination of numerical values of increments was based on zeroth order estimates derived from quantum chemical data of simple model compounds.[6,7]

Both types of interactions were observed to be correlated with other features of molecular structure. Typical examples are relaxation of internal structural parameters and systematic alterations of electric field gradients at N atoms of the rings or of imino groups, brought about by interconversion of geometric isomers (conformers) involving −OH and ═NH substituents. Also, nonplanarity of amino substituents appears to depend sensitively on attractions and repulsions entangling NH_2 hydrogen atoms.

In the meantime, analogous studies were devoted to all isomers (conformers and tautomers) of guanine and xanthine. These allowed questions left unresolved in the earlier work to be closed and led to (i) a fairly complete system of energy increments for conversion of geometric isomers associated with −OH and ═NH substituents, (ii) refinements of the classification of both attractions and repulsions, and (iii) new knowledge of regularities followed by conversion energies of tautomers.

The system of energy increments put forward in this work encompasses various degrees of sophistication. In the simplest form, five to six increments are required to coarsely reproduce (predict) conformer conversion energies. In the more elaborate forms, it should allow speculative treatment of stability problems of geometric isomerism of hydroxy and imino (amino) derivatives of 5- and 6-ring nitrogen containig aromatic heterocyclic bases.

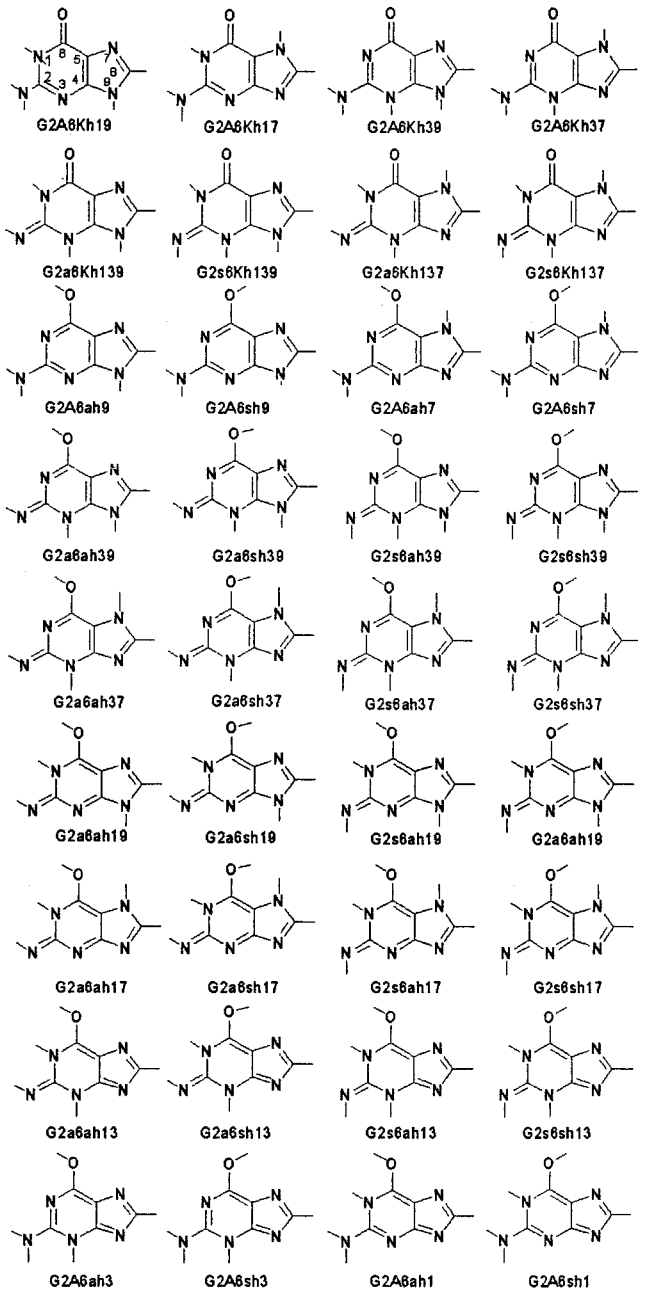
In sections 2 and 3, notation, symbolism, and computational details and results will be outlined. In section 4, systems of energy increments, considered as optimal estimates at this time are presented, and in section 5, relevant aspects of conformer and tautomer conversion processes will be discussed. Finally, in the Appendix, some model compounds will be given (HF/6-31G(d,p) and MP2/6-31G(d,p) approximations).

Quantum chemical studies of guanine isomers have frequently been reported; up to and including 1995, the Quantum Chemistry Library Data Base[8] contains 47 citations. The most complete treatment with regard to this work has been carried out by Sabio et al.,[9] which comprised a study of all 36 isomers at the HF/3-21G level (the three most stable isomers at MP2/6-31G and MP2/6-31G(5d)). Further work at higher quantum chemical approximation mostly concentrated on association by H-bonding,[10] ionization potential, and electron affinity,[11] tautom-

* To whom correspondence should be sent. Phone: +41 01 632 4170. Fax: +41 01 632 1256.

10.1021/jp984564p CCC: $18.00
Published on Web 07/30/1999

SCHEME 1: Guanine[a]



[a] Hydrogen substituents indicated by single bond symbols.

erism,[12] and nonplanarity of NH_2 substituents[13] of selected isomers. More recently work at the MP2/6-31G(d) level was reported by Colominas et al. for six cytosine and five guanine isomers (C2K6Ah1, C2K6Ah3, C2a6A, C2s6A, C2K6ah13, C2K6sh13 and G2A6Kh19, G2A6Kh17, G2A6ah9, G2A6sh9, and G2A6ah7; for notation, see Scheme 1) and by Florian et al.[15] for four cytosine and four guanine isomers (HF/6-31G* and MP2/6-31G**//HF/6-31G* approximations), in connection to the C−G base pairing problem. Four guanine isomers (G2A6Kh19, G2A6Kh17, G2A6ah9, and G2A6ah7) were treated by Gorb et al.[16] by approximations with 6-31G(d) basis (HF and single-point MP2and MP4) in a study of H-bonding to one and two H_2O moleccules.

Xanthine has received much less attention (four citations in Quantum Chemistry Literature Data Base up to 1995). Nonella et al. reported work on geometry optimization and electrostatic potential of two isomers.[17] Work at higher approximation has been published by Leszczynski[18] and by Kassimi et al.,[19] but no study with high-level approximation of all 34 isomers seems hitherto available.

The aims of the present work require a database of uniform high quality. All isomers of the nucleic acid bases considered (cytosine, isocytosine, uracil, thymine, adenine, guanine, and xanthine) and of many model compounds from the pyrimidine and purine group were computed at the levels HF/6-31G(d,p), ..., and MP2/6-31G(d,p), comprising unconstrained structure optimization, total electronic energy, rotational constants, electric dipole moments, and electric field gradients at the positions of the nitrogen nuclei. Local stability in most cases was checked by help of the Hessian of the total electronic energy. For many structures, normal modes, dipole polarizability, and higher electric moments were determined as well.

2. Notation, Increment Model, Computational Details, and Data Analysis

2.1. Notation. In Schemes 1 and 2 structural formulas of all isomers of guanine (G) and xanthine (X) are listed, together with an unequivocal alphameric notation; the isomers are grouped into subsets according to certain structural features relevant for determination of energy increments. Analogous notations will be used to denote isomers of cytosine (C), isocytosine (iC), uracil (U), thymine (T), and adenine (A).

Additive Increment Model. The basic assumptions of the model were discussed earlier, but will be recapitulated very briefly.

(i) For each geometric isomer S_m associated with a tautomer S endowed with geometric isomers (conformers) S_1, S_2, ..., S_m, ..., the total electronic energy is assumed to be built from a term E_S characteristic for the tautomer (described by its constitution) and a sum (index set J_{S_m}) of increments E_{S_mk} characteristic for geometric isomer S_m:

$$E^{\mathrm{tot}}_{S_m} = E_S + \sum_{J_{S_m}} E_{S_mk} \tag{2-1}$$

(ii) Conversion $S_m \rightarrow S_{m'}$ of two geometric isomers S_m and $S_{m'}$ of tautomer S implies the energy alteration

$$\Delta E^{\mathrm{tot}}_{S_m \rightarrow S_{m'}} = \sum_{J_{S_{m'}}} E_{S_mk} - \sum_{J_{S_m}} E_{S_mk} \tag{2-2}$$

(iii) Conversion of geometric isomer S_m of tautomer S to geometric isomer $S'_{m'}$ of tautomer S' implies a total energy change of

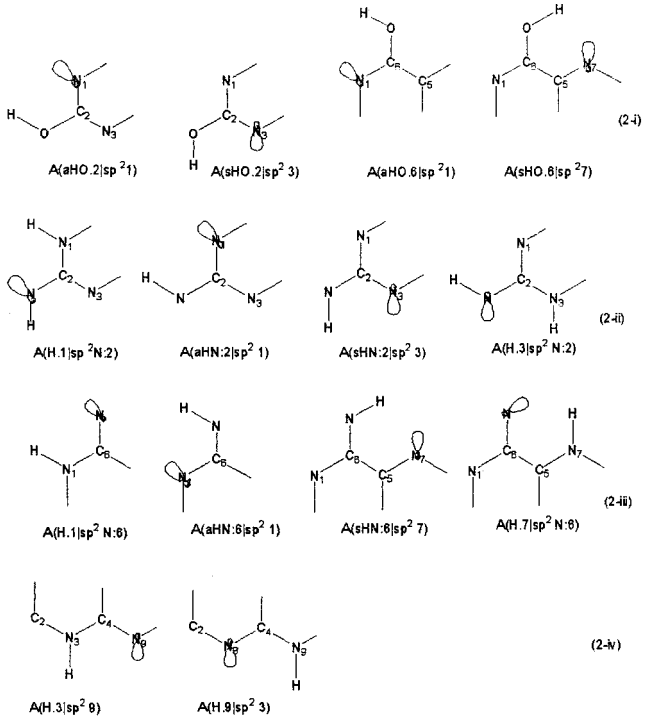
$$\Delta E^{\mathrm{tot}}_{S_m \rightarrow S'_{m'}} = E^{\mathrm{tot}}_{S'_{m'}} - E^{\mathrm{tot}}_{S_m} = \Delta E^{\mathrm{tauto}}_{S \rightarrow S'} + \sum_{J_{S'_{m'}}} E_{S'_mk} - \sum_{J_{S_m}} E_{S_mk} \tag{2-3}$$

the quantity $\Delta E^{\mathrm{tauto}}_{S \rightarrow S'}$ will be termed tautomerization energy. It might be perceived as electronic energy change associated with the conversion of the constitution of S into the constitution of S', each of these taken regardless of geometric isomerism. The assumptions of the model imply that $\Delta E^{\mathrm{tauto}}_{S \rightarrow S'}$ must be independent of the choice of particular geometric isomers $S_{m'}$, $S'_{m'}$ from within the sets $\{S_1, S_2, ...\}$, $\{S_1', S_2', ...\}$ of isomers of S and S'.

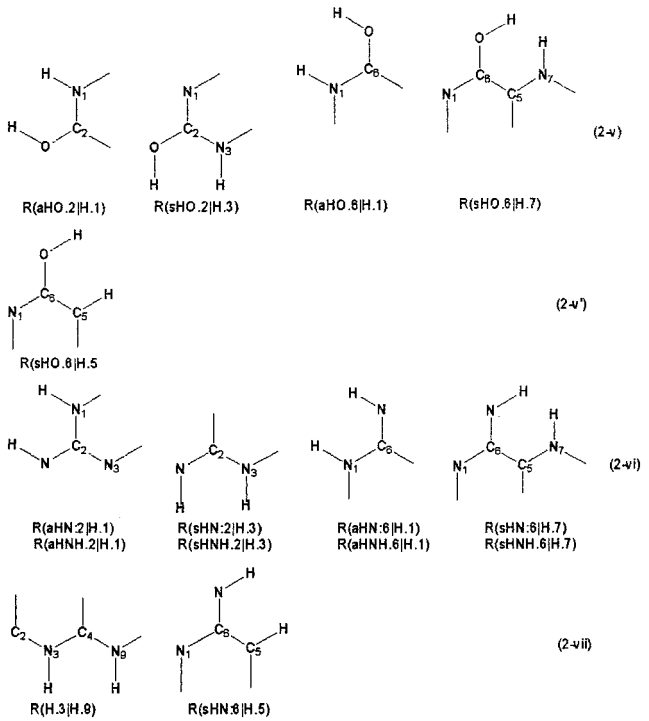
2.2. Structural Fragments: Attractive/Repulsive Interactions Correlations. In this paragraph, a set of structural fragments associated with attractive and repulsive interactions will be presented. This set was found to be required to describe the energetics of conversion of geometric isomers and tautomers according to the rules of the linear increment model. It

encompasses fragments encountered in the common pyrimidine and purine nucleic acid bases as well as in simpler hydroxy/keto and amino/imino derivatives of pyrimidine and purine. In the following compilation, a and s denote the anti and syn position of NH and OH substituents and lobes symbolize sp^2 lone electron pairs localized at pyridinic and imino N atoms. Numbering of atoms relates to conventions for pyrimidine and purine systems (See Schemes 1 and 2).

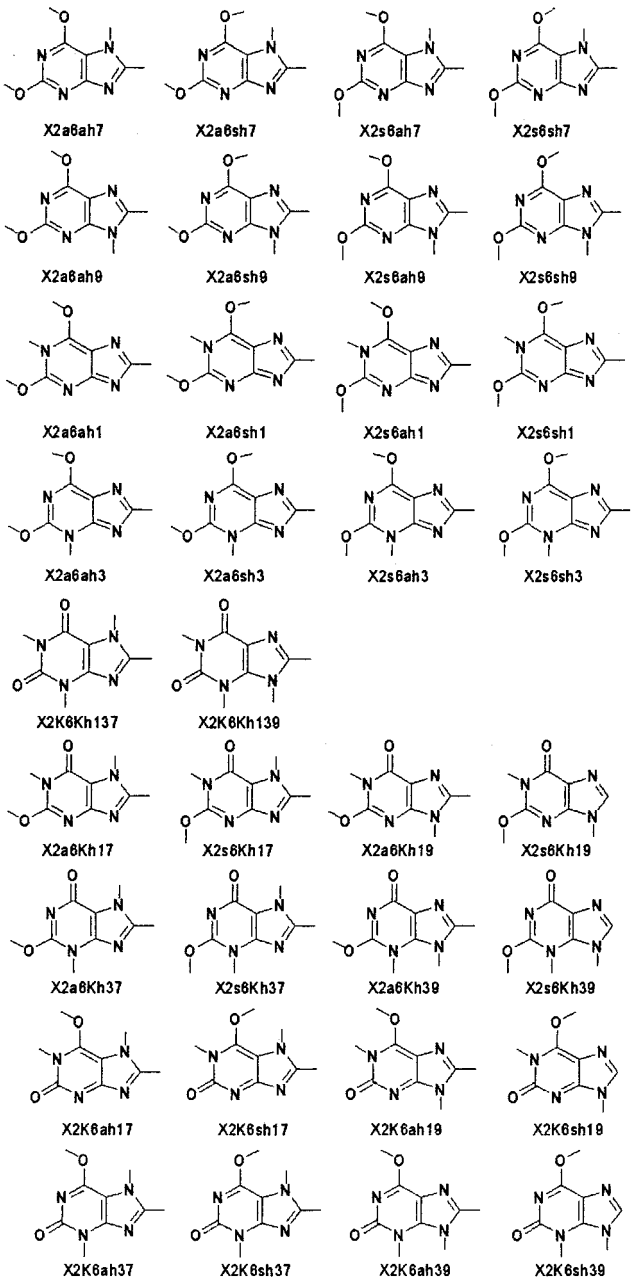
Attractions and associated structural fragments, OH and NH/NH_2 substituents:



Structural fragments associated with repulsions will be symbolized as follows for hydroxyl and imino/amino groups, respectively:

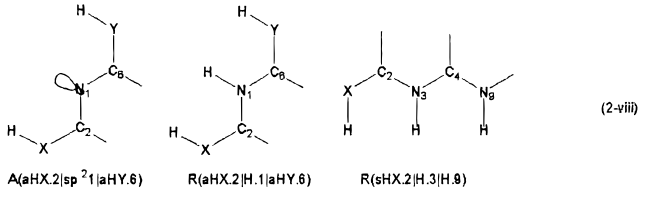


SCHEME 2: Xanthine[a]



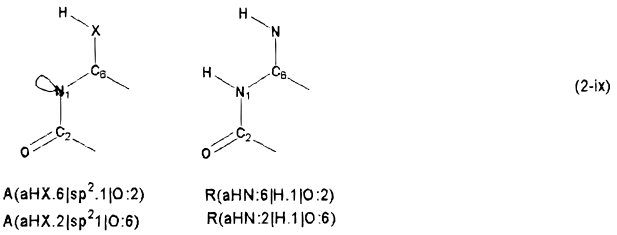
[a] Hydrogen substituents indicated by single bond symbols.

In a few isomers, "double" interactions arise; structural fragments of this type are (X, Y stand for O and/or N)



It has earlier been found that such interactions may be approximated by a sum of two single interactions and a defect (ΔA or ΔR), e. g., A(aHX.2|sp^{2}1|aHY.6) $\approx$ A(aHX.2|sp^{2}1) + A(aHY.6|sp^{2}1) + ΔA_{216} or R(aHX.2|H.1|aHY.6) $\approx$ R(aHX.2|H.1) + R(aHY.6|H.1) + ΔR_{216}, where $\Delta A_{216} \approx \Delta R_{216} \approx +1.1-1.4$ kcal/mol, with the single interactions falling into the usual intervals.

Some rarely encountered structural fragments were found to feature exceptionally large (small) attractions (repulsions).



(2-ix)

Interactions of the types introduced previously may be qualified as local in nature. In the course of the study, dependence of certain attractions and repulsions on more global structural features was found, e. g., formal aromatic structure of the six-membered ring.

For illustration, equation 2-2 will explicitly be given for conversion of geometric isomers within the guanine quadruple G2x6yh39 (x,y denote anti (a) or syn (s) conformation of substituents =NH and −OH in positions 2 and 6, respectively). From Scheme 1 and Table 5 (HF/6-31G(d,p) and MP2/6-31G-(d,p) data, kcal/mol) one obtains, writing shortly, aa → as for the conversion G2a6ah39 → G2a6sh39:

$$\Delta E^{\mathrm{tot}}_{\mathrm{aa}\rightarrow\mathrm{as}} = +2.59/+1.65 \approx -A(\mathrm{aHO.6|sp^2 1}) - \Delta A_{216} + A(\mathrm{sHO.6|sp^2 7}) \quad (2\text{-}4)$$

$$\Delta E^{\mathrm{tot}}_{\mathrm{aa}\rightarrow\mathrm{sa}} = +7.98/+6.85 \approx -A(\mathrm{aHN:2|sp^2 1}) - \Delta A_{216} - A(\mathrm{H.3|sp^2 N:2}) + R(\mathrm{sHN:2|H.3}) + \Delta R_{239} \quad (2\text{-}5)$$

$$\Delta E^{\mathrm{tot}}_{\mathrm{as}\rightarrow\mathrm{ss}} = +9.07/+7.85 \approx -A(\mathrm{aHN:2|sp^2 1}) - A(\mathrm{H.3|sp^2 N:2}) + R(\mathrm{sHN:2|H.3}) + \Delta R_{239} \quad (2\text{-}6)$$

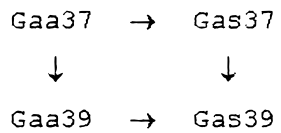
$$\Delta E^{\mathrm{tot}}_{\mathrm{sa}\rightarrow\mathrm{ss}} = +3.68/+2.64 \approx -A(\mathrm{aHO.6|sp^2 1}) + A(\mathrm{sHO.6|sp^2 7}) \quad (2\text{-}7)$$

Equations 2-4 through 2-7 are linearly dependent (Hess' law).

For the tautomer conversion G2a6ah37 → G2a6ah39 (h7 → h9 hydrogen shift), eq 2-3 takes the form (cf. Table 5, kcal/mol)

$$\Delta E^{\mathrm{tot}}_{\mathrm{h37}\rightarrow\mathrm{h39}} = +3.73/+4.50 \approx \Delta E^{\mathrm{tauto}}_{\mathrm{h37}\rightarrow\mathrm{h39}} + R(\mathrm{H.3|H.9}) - A(\mathrm{H.3|sp^2 9}) - \Delta A_{\mathrm{23h9}}. \quad (2\text{-}8)$$

Analogous equations hold for the tautomerization processes G2a6sh37 → G2a6sh39, G2s6ah37 → G2s6ah39, and G2s6sh37 → G2s6sh39. Equations 2-4 through 2−7 and the set of equations of type 2-8 contain linear dependencies. A typical relation of this sort is illustrated by the following diagram which demonstrates the existence of one relation between the four processes marked by arrows. Analogous diagrams hold for any other choice of geometric isomers.



2.3. Quantum Chemical Computations. Ab initio quantum chemical calculations of all isomers of guanine and xanthine have been carried out at the HF/6-31G(d,p) and MP2full/6-31G-(d,p)[20] levels of accuracy. This comprises 36 and 34 isomers of guanine and xanthine, respectively. Furthermore, eight isomers of 6-hydroxypurine and two acethylenic molecules have been included as model compounds. Geometrical structure and energy were optimized at both levels of theory.

TABLE 1: Guanine (Cytosine, Isocytosine, Adenine): Data Subsets Selected for Least-Squares Determination of Energy Increments[a]

subsets				
GLS1[b] (norm)	GLS2[c] (arom)	GLS3[h] (exot)	GLS4[i] (path)	GLS5[j] (noco)
G2x6yh17, G2x6yh19	Py,[d] Pym[e]	G2A6yh1	C2K6yh13	C2K6Ah1, iC2A6Kh1
G2x6yh37, G2x6yh39	C2x6A, iC2A6y[f]	G2A6yh3	iC2x6Kh13	C2K6Ah3, iC2A6Kh3
	G2A6yh7, G2A6yh9	G2x6yh13	G2x6Kh137	G2A6Kh17, G2A6Kh37
	P6xh7, P6xh9[g]	P6xh1[g]	G2x6Kh139	
		P6xh3[g]		

[a] See Scheme 1 for notation of guanine isomers, x and y symbolize geometric isomerism (anti, syn) of imino and hydroxyl substituents in 2- and/or 6-position. [b] Least-squares (LS) determination of 20 increments and tautomerization energies from 37 equations for conformer and tautomer conversion processes and zeroth order estimates; for structural features see Scheme 1. [c] LS computation of 10 increments and tautomerization energies from 17−21 equations for conversion processes and zeroth order estimates; data set includes isomers of hydroxy pyridines (Py), -pyrimidines (Pym), and -purines (P) and adenine. [d] See ref 2 for hydroxypyridine data. [e] See ref 4 for hydroxypyrimidine data. [f] See refs 1−4 for cytosine (C) and isocytosine (iC) data. [g] See Table 7 and ref 5 for purine (P) and adenine data. [h] LS determination of 18−20 increments and tautomerization energies from 24−32 conversion equations and zeroth order estimates. [i] LS estimation of 12 increments and tautomerization energies from 17 conversion equations and zeroth order estimates; subset "path" features unusually low repulsions *R*(aHN:2|H.1|O:6), etc. [j] Subset comprising cytosine, isocytosine, and guanine tautomers possessing no geometric isomer ("noco"); estimation of six tautomerization energies from 19 conversion equations and zeroth order estimates.

Comparison of energies obtained from HF and MP2 approximations shows the respective order of relative stability to differ slightly, predominantly in regions of nearly degenerated electronic energy. To keep the increment approach open for application to complex molecules, it was based on HF/6-31G-(d,p) data uniformly for all nucleic acid bases and model compounds considered.

The Gaussian 94 program package[21] has been used for the ab initio computations.

2.4. Data analysis. Unique determination of a sufficient set of increments and tautomerization energies for the complete set of isomers considered in this work encounters the difficulty that the number of these quantities exceeds the number of linearly independent equations of types 2-2 and 2-3 available from quantum chemistry. To determine (preferably nonarbitrary and uncorrelated) estimates, a number of assumptions were made.

(i) A and R type interactions associated with the 6,1,2,3-region were considered transferable between pyrimidine and purine systems with formally analogous electronic structure of the 6-ring entity.

(ii) Wherever possible, zeroth order estimates for specific interactions were determined from suitably chosen model compounds. Choice of the model compounds and their quantum chemical data were described in earlier papers;[1−7] see also the Appendix.

(iii) In the course of the analysis including guanine and xanthine, several observations emerged. First, systems with formally aromatic 6-ring systems showed systematic differences

TABLE 2: Xanthine (Uracil, Thymine): Data Subsets Selected for Least-Squares Determination of Energy Increments[a]

subsets				
UXLS1[b,c]	XLS2[d]	XLS3[e]	UXLS5[b,f]	XLS6[g]
		(exot)	(arom)	(keto)
U2x6Kh1, U2x6Kh3	X2x6Kh17, X2x6Kh19	X2x6yh1	U2x6y	X2x6Kh17, X2x6Kh19
U2K6yh1, U2K6yh3	X2x6Kh37, X2x6Kh39	X2x6yh3	X2x6yh7	X2K6yh17, X2K6yh19
X2x6Kh17, X2x6Kh19	X2K6yh17, X2K6yh19		X2x6yh9	X2x6Kh37, X2x6Kh39
X2x6Kh37, X2x6Kh39	X2K6yh37, X2K6yh39			X2K6yh37, X2K6yh39
X2K6yh17, X2K6yh19	X2x6yh7, X2x6yh9			
X2K6yh37, X2K6yh39				
X2x6yh1, X2x6yh3				
X2x6yh7, X2x6yh9				

[a] See Scheme 2 for notation of xanthine isomers; x, y symbolize conformation (anti, syn) of OH groups in 2- and 6-position. [b] See refs 1−3 for uracil (U) data. [c] Least-squares analysis (LS) for 14 unknown increments with 44 equations comprising zeroth order estimates and (solely) geometric isomer conversions. [d] LS for 22 increments and tautomerization energies based on 44 equations for conversion processes and zeroth order estimates; only xanthine species without types arom and exot. [e] LS for 11 increments and tautomerization energies, based on 21 equations for conversion processes and zeroth order estimates; only type "exot" isomers. [f] LS for 10 increments and tautomerization energies, based on 23 equations for uracil and xanthine isomers with formally aromatic 6-ring [g] LS for 21 increments/tautomerization energies and 38 equations for conversion processes/zeroth order estimates; only xanthine isomers with keto group in 2- or 6-position admitted.

with respect to specific interactions when compared with isomers not having this feature. Second, the transfer of interactions from simple pyrimidine and purine model compounds revealed significantly different values for a very few particular increments. Third, to obtain robust estimates, it proved appropriate to subdivide the full set of isomers into subsets of analogous compounds and to determine optimal increment values for each subset separately. This approach provided some insight into robustness and correlation of the numerical values. Tables 1 and 2 list some of the selected subsets, together with the numbers of unknowns and available equations on which the least-squares analysis has been based. Since equations 2-2 and 2-3 are linear in the increments of conformer and tautomer conversion processes, linear least-squares may be used for optimal estimation of these quantities. The use of zeroth order estimates and the transfer of increments between data subsets introduce a certain degree of bias in the estimates. For some subsets, condition numbers of the normal matrices turned out rather high, therefore the least-squares *error* estimates for increments and tautomerization energies appear to be rather too low.

3. Results

In Tables 3 and 4, optimal estimates of energy increments for conversion of geometric isomers and tautomers of guanine and xanthine are collected. Total electronic energies for guanine and xanthine are listed in Tables 5 and 6; naming of the isomers in these tables is illustrated in Schemes 1 and 2.

The information given in Table 3 also holds for cytosine, isocytosine,[22] and adenine,[23] where applicable. Specific results for the latter three bases were reported earlier.[2−5] The increments for xanthine as listed in Table 4 also hold for uracil and thymine, where applicable. For the latter nucleic acid bases, the results reported in previous work should be consulted.[1−3]

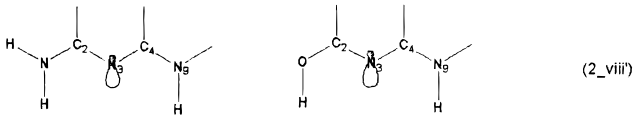
Finally, quantum chemical data of model compounds are put forward in Table 7, together with the notation for hydroxypurines used in Tables 1−4.

4. Discussion

The following discussion is centered around Tables 3 and 4. Error estimates of the quantities listed are given in parentheses following the least-squares estimates.

4.1. Guanine (Adenine, Cytosine, Isocytosine) Set. Table 3 shows that estimates obtained from the various subsets listed in Table 1 mostly have equal values within errors. However, a few deviations stand out. Increments with nearly equal values in all subsets are, e.g., attractive interactions involving amino/imino hydrogen and sp^2N lone pairs (fragments 2-ii and 2-iii, which amount to −2.7 to −3.2 kcal/mol. The attractions A(H.3|$sp^2$9) and A(H.9|$sp^2$3), associated with fragments 2-iv, lie near the first bound of this interval.

The quantities ΔA_{216}, ΔA_{23h9} range mostly in the interval 1.1−1.4 kcal/mol. In contrast, the quantity ΔA_{23s9} (fragment 2-iii) amounts for nearly 0.8 kcal/mol for subsets GLS1("norm") and GLS5("noco") and to 0.0(1) for subset GLS2("arom"). The latter value resulted from a variety of least-squares calculations with different zeroth order estimates and transfer of uncritical increments. It also should be opposed to its analogue in the xanthine subset UXLS5 ("arom"), where a value in the usual range (1.3 kcal/mol) has been found. Possibly the difference is to be traced to the difference of the associated structural fragments (type 2-viii and 2-viii′).



(2_viii')

Double attractions (fragment (2-viii)) involving NH or NH_2 follow the sum ansatz proposed in section 2; for the two quantities A(aHNH.2|$sp^2$1|O:6) and A(aHNH.6|$sp^2$1|O:2), lack of conformer conversion data prevents checking of the ansatz.

Attractions involving OH substituents in position 2 or 6 and sp^2N.1 or sp^2N.3 lone pairs fall into the range −5(5) kcal/mol for subsets GLS1, GLS2, and GLS3. However, attraction A(sHO.6|$sp^2$7) is (absolutely) extraordinarily low for subsets GLS1 and GLS2 (−1.8 and −2.9 kcal/mol, respectively), whereas it ranges near −5.4 kcal/mol for subset GLS3. The first low value is apparently well determined by least squares, the the second and third seem characteristic for subsets "arom" and "exot", respectively, in analogy to xanthine. The isomers of subset GLS3 possess (abs.) large A(aHO.6|$sp^2$1) and A(sHO.6|$sp^2$7) interactions. The latter feature is also born out by model compounds (see Appendix) and is apparently linked to the particular electronic structure of the imidazole region of "exotic" isomers.

The diversity of the values of A(sHO.6|$sp^2$7) between subsets GLS1 and GLS2 on one hand and GLS3 on the other hand motivates the following reflections. On the basis of the rule that interactions in the 6,1,2,3-region should be transferable (cf. section 2.4) and on the processes (cf. Tables 5 and 6, conversion

TABLE 3: Guanine Isomers: Electronic Energy Increments for Conversion of Geometric Isomers (Conformers) and Tautomers[a]

	data subsets[b]				
increments	GLS1	GLS2	GLS3	GLS4	GLS5
	(norm)	(arom)	(exot)	(path)	(noco)
A(aHN:2|sp^{2}1)	−3.03(22)		−3.19(32)		
A(sHN:2|sp^{2}3)	−3.20(25)		−3.01(31)		
A(H.1|sp^{2}N:2)	−3.12(21)		−3.00(23)	−2.99(6)	
A(H.3|sp^{2}N:2)	−3.16(25)		−3.08(46)	−2.89(6)	
A(H.1|sp^{2}N:6)				−2.96(6)	
A(H.3|sp^{2}9)	−2.70(27)		−2.70(33)	−2.70(7)	−3.01(37)
A(H.9|sp^{2}3)	−2.70(39)	−2.70(15)			−2.70(43)
ΔA_{216}	+1.08(15)	+0.71(8)	+1.28(26)		
ΔA_{23h9}	+1.41(46)		+1.32(29)	+1.42(5)	
ΔA_{23s9}	+0.76(19)	−0.0(1)			+0.76(43)
A(aHNH.2|sp^{2}1)			−3.38(42)		
A(sHNH.2|sp^{2}3)			−2.81(42)		−3.20(43)
A(aHNH.2|sp^{2}1|O:6)					−4.30(43)
A(aHNH.6|sp^{2}1|O:2)					−4.30(61)
A(aHO.2|sp^{2}1)		−4.75(14)			
A(sHO.2|sp^{2}3)		−4.75(15)			
A(aHO.6|sp^{2}1)	−5.72(21)	−4.63(13)	−5.11(27)		
A(sHO.6|sp^{2}7)(G)	−1.75(23)	−2.93(15)	−5.37(30)		
A(sHO.6|sp^{2}7)(P)			−6.33(20)		
R(aHN:2|H.1)	+2.44(26)		+2.50(28)		
R(sHN:2|H.3)	+2.44(38)		+2.29(28)	+2.49(6)	
R(aHNH.2|H.1)			+2.70(41)		
R(sHNH.2|H.3)			+2.10(41)		+2.44(43)
R(aHN:2|H.1|O:6)				+0.89(5)	
R(aHN:6|H.1|O:2)				+0.87(7)	
R(aHNH.2|H.1|O:6)					+0.88(43)
R(aHNH.6|H.1|O:2)					+0.88(61)
R(sHN:6|H.5)		+1.29(7)		+2.09(6)	
R(H.3|H.9)	+2.30(27)			+2.30(7)	+2.27(43)
ΔR_{216}	+1.33(15)		+1.31(14)		
ΔR_{239}	+0.43(46)			+0.52(5)	+0.50(43)
R(aHO.6|H.1)	+5.45(21)		+4.82(27)		
R(sHO.6|H.7)	+6.17(23)	+4.95(15)			
$\Delta E^{tauto}_{h17\rightarrow h19}$	−7.32(42)				
$\Delta E^{tauto}_{h37\rightarrow h39}$	−0.15(61)				
$\Delta E^{tauto}_{h7\rightarrow h9}$		−1.69(11)			
$\Delta E^{tauto}_{h1\rightarrow h3}$(G)			−4.49(62)		
$\Delta E^{tauto}_{h1\rightarrow h3}$(P)			−3.40(29)		
$\Delta E^{tauto}_{h137\rightarrow h139}$				+4.11(11)	
$\Delta E^{tauto}_{h17\rightarrow h19}$					+1.05(71)
$\Delta E^{tauto}_{h37\rightarrow h39}$					+7.80(79)
$\Delta E^{tauto}_{h17\rightarrow h37}$					+9.80(95)
$\Delta E^{tauto}_{h19\rightarrow h39}$					+16.48(128)
$\Delta E^{tauto}_{h1\rightarrow h3}$(C)					−2.04(96)
$\Delta E^{tauto}_{h1\rightarrow h3}$(iC)					+10.11(93)

[a] HF/6-31G(d,p) data, energy in kcal/mol. [b] See data subsets listed in Table 1.

energies in kcal/mol, superscipts a, e, and n stand for arom, exot, norm, HF results):

$$\text{G2x6yh19} \in \text{GLS1}: \quad \Delta E_{aa\rightarrow as} = +8.21, \quad \Delta E_{sa\rightarrow ss} = -6.94$$

$$\text{G2x6yh13} \in \text{GLS3}: \quad \Delta E_{aa\rightarrow as} = -10.80, \quad \Delta E_{sa\rightarrow ss} = -9.51$$

$$\text{G2A6yh1} \in \text{GLS3}: \quad \Delta E_{a\rightarrow s} = -11.49 \qquad (4\text{-}1)$$

$$\text{G2A6yh3} \in \text{GLS3}: \quad \Delta E_{a\rightarrow s} = -1.80$$

$$\text{G2A6xh3} \in \text{GLS2}: \quad \Delta E_{a\rightarrow s} = +1.04$$

the corresponding equations of type 2-2 may be combined to yield

$$A^e(\text{sHO.6}|\text{sp}^2\text{7}) - A^n(\text{sHO.6}|\text{sp}^2\text{7}) \approx -2.59$$

$$A^e(\text{sHO.6}|\text{sp}^2\text{7}) - A^a(\text{sHO.6}|\text{sp}^2\text{7}) \approx -2.84 \qquad (4\text{-}2)$$

The second of equations 4-2 gives, using the data for A^e(sHO.6|sp^{2}7) from Table 3, A^a(sHO.6|sp^{2}7) ≈ −2.53, in satisfactory agreement with the least squares estimate in Table 3. Similar arguments applied to the first of equations 4-2 predict A^n(sHO.6|sp^{2}7) ≈ −2.78, deviating significantly from the least-squares estimate (−1.75). Hence, the estimate of A(sHO.6|sp^{2}7) of "normal" isomers (subset GLS1) shown in Table 3 appears as too low. In predictions of geometric isomer conversions of GLS1 type molecules, one should consider a range −1.8 to −2.8 for A(sHO.6|(sp^{2}7). Similar relationships hold for analogous isomers in the xanthine group.

A problem arising from the use of model compounds in the guanine set should be mentioned. On the basis of the conformer conversion processes of "exotic" 6-hydroxypurines (P6xh1, P6xh3, cf Appendix) and "exotic" guanine isomers (subset GLS3), it may be shown by similar arguments, as in the foregoing paragraph, that (superscripts P and G relate to purine

TABLE 4: Xanthine Isomers: Electronic Energy Increments for Conversion of Geometric Isomers (Conformers) and Tautomers[a]

	data subsets[b]				
increments	UXLS1	XLS2	XLS3	UXLS5	XLS6
			(exot)	(arom)	(keto)
A(aHO.2|sp^{2}1)	−4.63(53)	−4.66(24)	−4.73(31)	−4.64(20)	
A(aHO.2|sp^{2}1|O:6)		−5.49(30)			−5.89(19)
	−6.40(54)				
A(aHO.6|sp^{2}1|O:2)		−5.99(23)			−5.96(22)
ΔA_{216}	+1.35(40)	+1.45(19)	+1.48(30)	+1.26(13)	
A(sHO.2|sp^{2}3)	−5.02(57)	−5.03(24)	−5.54(39)	−4.93(20)	−5.00(20)
A(aHO.6|sp^{2}1)	−4.30(63)	−4.52(27)	−4.59(47)	−4.94(18)	
A(sHO.6|sp^{2}7)	−3.73(62)	−2.36(25)	−5.49(50)	−2.99(23)	−2.24(25)
ΔA_{23s9}	+1.58(52)	+1.54(17)	+1.54(17)	+1.39(18)	+1.62(18)
A(H.3|sp^{2}9)		−2.75(34)	−2.75(46)		−2.75(28)
A(H.9|sp^{2}3)		−2.75(49)		−2.70(27)	−2.75(28)
R(aHO.2|H.1)	+5.23(56)	+4.91(22)	+5.58(37)		+5.10(19)
R(sHO.2|H.3)	+5.55(54)	+6.20(28)	+5.60(34)		+5.75(20)
R(aHO.6|H.1)	+5.42(57)	+5.07(23)	+5.44(47)		+5.07(22)
R(sHO.6|H.5)	+1.82(69)			+1.33(19)	
ΔR_{216}	+2.31(57)	+1.25(34)	+1.42(30)		
R(sHO.6|H.7)	+5.65(69)	+5.58(25)		+4.93(23)	+5.70(21)
ΔR_{239}	+1.27(70)	+1.39(25)			+1.37(38)
R(H.3|H.9)		+2.27(34)			+2.27(28)
$\Delta E^{tauto}_{Xh17\rightarrow h19}$ (6K)					+1.36(34)
$\Delta E^{tauto}_{Xh17\rightarrow h37}$ (6K)		+9.17(53)			+9.70(41)
$\Delta E^{tauto}_{Xh19\rightarrow h39}$ (6K)		+15.37(73)			+15.96(50)
$\Delta E^{tauto}_{Xh37\rightarrow h39}$ (6K)					+7.62(44)
$\Delta E^{tauto}_{Xh17\rightarrow h19}$ (2K)		−7.11(55)			−7.11(34)
$\Delta E^{tauto}_{Xh17\rightarrow h37}$ (2K)					−6.85(34)
$\Delta E^{tauto}_{Xh19\rightarrow h39}$ (2K)					−0.52(44)
$\Delta E^{tauto}_{Xh37\rightarrow h39}$ (2K)		−0.78(55)			−0.79(44)
$\Delta E^{tauto}_{Xh1\rightarrow h3}$			−5.14(65)		
$\Delta E^{tauto}_{Xh7\rightarrow h9}$		−2.33(53)		−2.32(33)	

[a] HF/6-31G(d,p) data, energy in kcal/mol [b] see data subsets listed in Table 2.

and guanine, HF/6-31G(d,p) data, kcal/mol)

$$A^{P}(\text{sHO.6}|\text{sp}^2\text{7}) \approx A^{G}(\text{sHO.6}|\text{sp}^2\text{7}) - \Delta A_{216}$$

$$A^{P}(\text{sHO.6}|\text{sp}^2\text{7}) \approx A^{G}(\text{sHO.6}|\text{sp}^2\text{7}) - \Delta R_{216} \quad (4\text{-}3)$$

Since consistently $\Delta A_{216} \approx \Delta R_{216} \approx 1.1-1.4$ for all sets, eqs 4-3 independently demonstrate that A(sHO.6|sp^{2}7) for "exotic" 6-hydroxypurines and guanines differ systematically. This is one of the very few exceptional cases encountered in this work where corresponding increments of model compounds and nucleic acid bases differ significantly.

Repulsions associated with structural fragments 2-ix, e.g., R(aHN:6|H.1|O:2), and analogues with NH_2 substituents were found to be "pathologically" low (0.9). In earlier work on cytosine and isocytosine, this was conjectured; inclusion of guanine data now allowed these types of interactions to be estimated.

The repulsions R(aHO.6|H.1) and R(sHO.6|H.7) turn out systematically higher for subset GLS1 (+5.5 and 6.2 kcal/mol, respectively) than for subsets GLS2 and GLS3 (5.0 and 4.8 kcal/mol).

The two quantities ΔR_{216} and ΔR_{239} associated with fragments of type 2-iii behave uniformly in the subsets GLS1, GLS4, and GLS5; the first lies in the coventional range (1.3 kcal/mol), the second is much smaller (0.5 kcal/mol) with rather high error estimates. Since its statistical basis is small (only three occurrences in the full guanine set), the significance of ΔR_{239} might be questioned.

4.2. Xanthine (Uracil, Thymine) Set. The results of the least-squares evaluation of energy increments for geometric isomer and tautomer conversions representative for all isomers of xanthine, uracil, and thymine are shown in Table 4. In analogy with the procedure followed for the guanine set, increments were determined separately for the subsets listed in Table 2.

It first should be mentioned, that inclusion of thymine data in subsets UXLS1 and UXLS5 on one hand confirmed within estimated errors increments reported earlier for this base[2,3] and, on the other hand, did not reveal significant alterations of the increments found from the xanthine/uracil data. Therefore, the information presented in Table 4 also holds for thymine, where applicable. For repulsions involving CH_3 substituents, the reference cited above might be inspected.

Overviewing Table 4 reveals the majority of increments associated with conformer conversions to have approximately equal values for all subsets. Deviation from this general behavior will subsequently be discussed in some detail.

Attraction A(sHO.6|sp^{2}7), whose value for subset XLS3 (type "exot") ranges near −5.5 kcal/mol, assumes significantly lower values for all other subsets. For subset UXLS1, the value −3.7 kcal/mol results from admission of type "exot" isomers in the least-squares process; for subsets XLS2 and XLS5 (comprising solely isomers of type "arom") and XLS6 (comprising solely isomers of type "keto"), respectively, A(sHO.6|sp^{2}7) values between −2.2 and −3.0 kcal/mol were found. The value mentioned above for subset UXLS1 should therefore be considered to be an artifact caused by the selection of the molecules forming this subset.

By comparison of conversion energies of the processes X2a6ah3 → X2a6sh3, X2s6ah3 → X2s6sh3, X2a6ah9 → X2a6sh9, X2s6ah9 → X2s6sh9 it may be shown, that attraction A(sHO.6|sp^{2}7) is different for isomers of type "exot" and "arom", in close analogy to the behavior of this quantity outlined above for guanine.

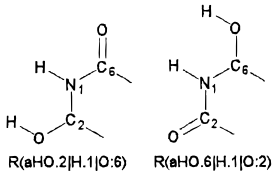
TABLE 5: Guanine Isomers: Quantum Chemical Data

	HF/6-31G(d,p)			MP2/6-31G(d,p)	
isomer[a]	$-E_{tot}$ −539.000000 (au)	E_{rel} (kcal/mol)	isomer	$-E_{tot}$ −541.000000 (au)	E_{rel} (kcal/mol)
G2A6ah9	0.412780	0.00	G2A6Kh17	0.088920	0.00
G2A6Kh19	0.412510	0.17	G2A6Kh19	0.088844	0.05
G2A6sh9	0.411120	1.04	G2A6ah9	0.087143	1.12
G2A6Kh17	0.410994	1.12	G2A6sh9	0.086155	1.74
G2A6ah7	0.405272	4.71	G2A6ah7	0.081780	4.48
G2a6Kh137	0.401769	6.91	G2A6Kh37	0.077771	7.00
G2s6Kh137	0.401609	7.01	G2s6Kh137	0.077302	7.29
G2A6Kh37	0.399843	8.12	G2a6Kh137	0.077114	7.41
G2A6sh7	0.390927	13.71	G2A6sh7	0.067897	13.19
G2a6Kh139	0.389553	14.58	G2A6sh3	0.066858	13.84
G2A6sh3	0.389314	14.73	G2a6Kh139	0.064860	15.10
G2A6ah3	0.386441	16.53	G2A6ah3	0.063413	16.01
G2s6Kh139	0.386226	16.66	G2s6Kh139	0.061829	17.00
G2a6ah37	0.380014	20.56	G2A6Kh39	0.056978	20.04
G2A6Kh39	0.378109	21.76	G2a6ah37	0.056732	20.20
G2A6sh1	0.375766	23.23	G2A6sh1	0.055146	21.19
G2a6ah39	0.374078	24.29	G2a6ah39	0.049567	24.69
G2s6sh19	0.374050	24.30	G2s6ah37	0.048672	25.26
G2s6sh13	0.372412	25.33	G2s6sh19	0.048114	25.61
G2a6sh13	0.370294	26.66	G2a6sh39	0.046936	26.35
G2s6ah37	0.370205	26.72	G2s6sh13	0.046913	26.36
G2a6sh39	0.369958	26.87	G2a6sh13	0.044207	28.06
G2s6ah19	0.362986	31.25	G2a6sh37	0.040188	30.58
G2a6sh37	0.362353	31.64	G2s6ah39	0.038644	31.55
G2s6ah39	0.361356	32.27	G2A6ah1	0.037756	32.11
G2a6sh19	0.361236	32.34	G2s6ah19	0.036891	32.65
G2A6ah1	0.357462	34.71	G2a6sh19	0.036287	33.03
G2s6ah13	0.357263	34.84	G2s6sh39	0.034431	34.19
G2s6sh39	0.355497	35.95	G2s6ah13	0.031762	35.87
G2a6ah13	0.353082	37.46	G2s6sh37	0.030660	36.56
G2s6sh37	0.350880	38.84	G2a6ah13	0.027178	38.74
G2a6ah19	0.348158	40.55	G2a6ah19	0.023319	41.17
G2s6ah17	0.347861	40.74	G2s6ah17	0.022035	41.97
G2s6sh17	0.347059	41.24	G2s6sh17	0.021037	42.60
G2a6sh17	0.333139	49.98	G2a6sh17	0.008329	50.57
G2a6ah17	0.331686	50.89	G2a6ah17	0.007226	51.26

[a] See Scheme 1 for structures of isomers.

Attractions Associated with Structural Fragments 2-ix. From Table 4, such attractions (−6 kcal/mol) are seen to be noticeably higher than interactions of type 2-i, a finding that was made already in earlier work. The value found with subset UXLS1 (−6.4 kcal/mol) again seems to be strongly biased by the composition of this subset.

Repulsions in the xanthine set associated with the following structural fragments



(2-ix″)

which are analogues of structural fragments 2-ix, possess values typical for single interactions (fragment type 2-v), ∼5.0−5.5 kcal/mol. This contrasts findings with the guanine set, where repulsions associated with structural fragments of type 2-ix turned out exceptionally low (subset GLS4 "path").

Repulsion *R*(sHO.6|H7) (fragment 2-v) tends to be somewhat smaller for subset UXLS5 ("arom") than that found for other subsets. However, the difference is not distinct.

As a whole, the xanthine set behaves markedly more uniformly than the guanine set. In the latter, the imino/amino substituents imply a larger variety of local interactions; furthermore, amino substituents feature nonplanar structures of considerable diversity. Quantum chemically predicted structures of all isomers of the xanthine set are planar. Analogous behavior has been observed with the sets cytosine−isocytosine vs uracil−thymine.

TABLE 6: Xanthine Isomers: Quantum Chemical Data

	HF/6-31G(d,p)			MP2/6-31G(d,p)	
isomer[a]	$-E_{tot}$ −559.000000 (au)	E_{rel} (kcal/mol)	isomer	$-E_{tot}$ −560.000000 (au)	E_{rel} (kcal/mol)
X2K6Kh137	0.262738	0.00	X2K6Kh137	0.944208	0.00
X2K6Kh139	0.248985	8.63	X2K6Kh139	0.930432	8.64
X2s6Kh17	0.243016	12.38	X2s6Kh17	0.926861	10.89
X2s6Kh19	0.242535	12.68	X2s6Kh19	0.925514	11.73
X2s6ah9	0.237790	15.66	X2s6ah9	0.919198	15.69
X2a6ah9	0.237346	15.93	X2a6ah9	0.918732	15.99
X2a6sh9	0.236247	16.62	X2a6sh9	0.918378	16.21
X2s6sh9	0.234691	17.60	X2s6sh9	0.917015	17.06
X2K6ah37	0.233365	18.43	X2a6Kh37	0.916574	17.34
X2a6Kh37	0.233025	18.65	X2K6ah37	0.915936	17.74
X2s6ah7	0.232206	19.16	X2s6ah7	0.915363	18.10
X2a6ah7	0.229138	21.08	X2a6Kh19	0.912961	19.61
X2a6Kh19	0.229004	21.17	X2a6ah7	0.912248	20.06
X2K6sh19	0.227400	22.17	X2a6Kh17	0.911436	20.56
X2a6Kh17	0.226697	22.62	X2K6sh19	0.908050	22.69
X2K6ah39	0.226179	22.94	X2K6ah39	0.907608	22.97
X2K6sh39	0.220640	26.42	X2K6sh39	0.903615	25.47
X2a6sh3	0.219104	27.38	X2a6sh3	0.903170	25.75
X2s6sh7	0.216441	29.05	X2s6sh7	0.900305	27.55
X2K6ah19	0.216179	29.22	X2s6Kh37	0.899539	28.03
X2a6sh7	0.215451	29.67	X2a6sh7	0.899106	28.30
X2a6ah3	0.215384	29.72	X2a6ah3	0.898966	28.39
X2s6Kh37	0.214901	30.02	X2K6sh37	0.898078	28.95
X2K6sh37	0.214301	30.39	X2K6ah19	0.896922	29.67
X2a6Kh39	0.213033	31.19	X2a6Kh39	0.896251	30.09
X2s6sh1	0.208447	34.07	X2s6sh1	0.892801	32.26
X2s6sh3	0.203301	37.30	X2s6sh3	0.888536	34.93
X2s6ah3	0.201679	38.32	X2s6ah3	0.886324	36.32
X2K6ah17	0.200069	39.33	X2K6ah17	0.881355	39.44
X2K6sh17	0.199425	39.73	X2K6sh17	0.880196	40.17
X2s6Kh39	0.192421	44.12	X2s6Kh39	0.876785	42.31
X2s6ah1	0.190953	45.05	X2s6ah1	0.875634	43.03
X2a6sh1	0.190000	45.64	X2a6sh1	0.875116	43.36
X2a6ah1	0.170503	57.88	X2a6ah1	0.856124	55.27

[a] See Scheme 2 for structures of isomers.

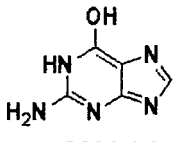
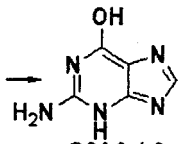
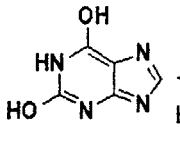
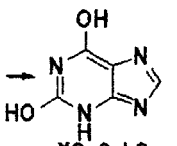
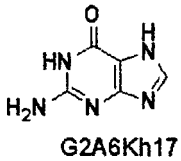
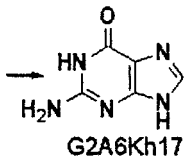
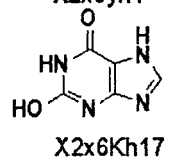
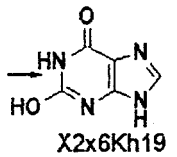
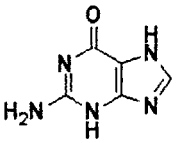
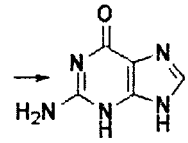
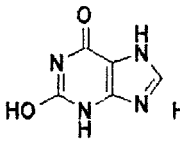
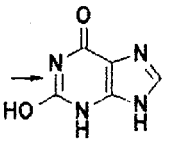
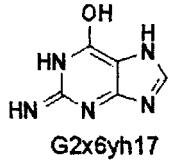
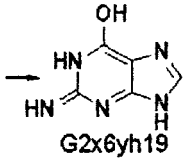
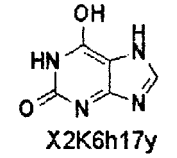
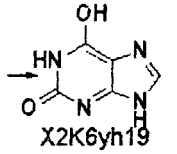
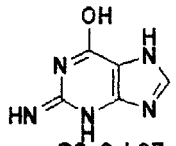
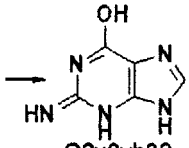
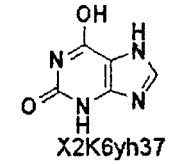
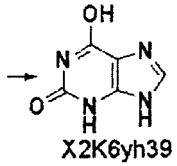
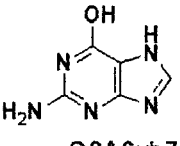
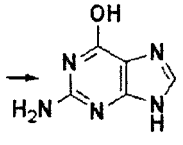
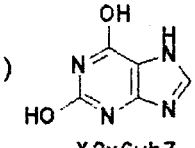
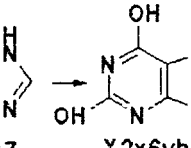
TABLE 7: 6-Hydroxypurines: Quantum Chemical Data

	HF/6-31G(d,p)		MP2/6-31G(d,p)	
isomer[a]	$-E_{tot}$ −484.000000 (au)	E_{rel} (kcal/mol)	$-E_{tot}$ −485.000000 (au)	E_{rel} (kcal/mol)
P6ah7	0.352747	+3.69	0.860979	+2.77
P6sh7	0.337490	+13.26	0.846211	+12.04
P6ah9	0.358631	0.00	0.865388	0.00
P6sh9	0.356077	+1.60	0.863391	+1.25
P6ah1	0.309563	+17.40	0.848215	+16.72
P6sh1	0.327624	+6.07	0.821566	+5.73
P6ah3	0.334971	+1.46	0.845845	+1.49
P6sh3	0.337303	0.00	0.848215	0.00

[a] See Appendix; P6ah7−P6sh9 and P6ah1−P6sh3 correspond to guanine isomers of type "arom" and "exot", respectively.

4.3. Tautomerization Energies. In this subsection, tautomerization energies observed for isomer conversion within the guanine and xanthine sets will be discussed. Investigation of these quantities will be restricted to simple shifts of H atoms bound to ring nitrogen atoms; a large variety of tautomerization processes involving H shifts between ring N atoms and O and N substituents will be disregarded. It should be remarked, however, that any electronic tautomerization energies $\Delta E^{tauto}_{S\rightarrow S'}$ within the guanine and xanthine sets may be estimated by the aid of equations of type 2-3, quantum chemical data for the isomers $S_{m'}$ and $S'_{m'}$ concerned (as listed in Tables 5 and 6) and

TABLE 8: Guanine and Xanthine Isomers: Tautomerization Energies (kcal/mol, HF/6-31G(d,p) Data)

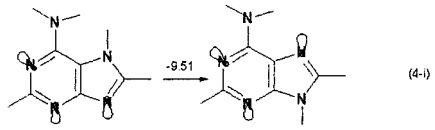
Guanine		ΔE^{tauto}	Xanthine		ΔE^{tauto}
 G2A6yh1	 G2A6yh3	-4.49(62) exot	 X2x6yh1	 X2x6yh3	-5.14(65) exot
 G2A6Kh17	 G2A6Kh17	+1.05(71) noco	 X2x6Kh17	 X2x6Kh19	+1.36(34) keto
 G2A6Kh37	 G2A6Kh39	+7.80(71) noco	 X2x6Kh37	 X2x6Kh39	+7.62(44) keto
 G2x6yh17	 G2x6yh19	-7.32(42) norm	 X2K6h17y	 X2K6yh19	-7.11(55) keto
 G2x6yh37	 G2x6yh39	-0.15(61) norm	 X2K6yh37	 X2K6yh39	-0.78(55) keto
 G2A6yh7	 G2A6yh9	-1.69(11) arom	 X2x6yh7	 X2x6yh9	-2.32(33) arom

estimates for local interactions (A, R, ΔA, ΔR, as listed in Tables 3 and 4).

In Tables 8 and 9, a selection of tautomerization energies relating to displacements of H atoms between positions 1 and 3 and 1, 3, 7, and 9 of the pyrimidine and purine skeleton, respectively, is listed.[22] Entries in the tables have been ordered such that the following rule should be illustrated: "tautomerization energies associated with shifts of single H atoms bound to ring N atoms of formally equal chemical structures have coarsely equal values". Chemical structure is understood here as defined by the conventional graph expressing constitution by vertexes (atom symbols) connected by edges (single bonds, double bonds, ...).

A few further examples relating to adenine may serve to expose this rule further (energy in kcal/mol).

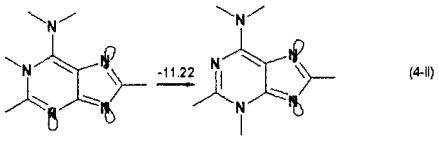
(i) Conversion A6Ah7 → A6Ah9[5,23].



(4-i)

Least-squares determination yielded $\Delta E^{tauto}_{Ah7\rightarrow h9} \approx -2.06(60)$, use of eq 2-3 and Table 3 gives $-1.79(125)$. Both isomers belong to type "arom"; hence, comparison (Table 8) should be made with $\Delta E^{tauto}_{Gh7\rightarrow h9} \approx -1.69(11)$ and $\Delta E^{tauto}_{Xh7\rightarrow h9} \approx -2.32(33)$.

(ii) Conversion A6Ah1 → A6Ah3 (Type "exot").



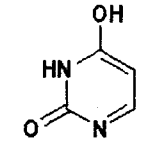
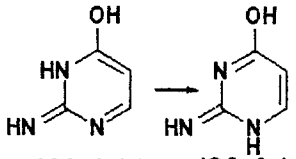
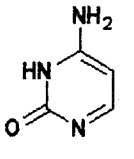
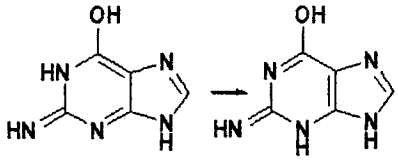
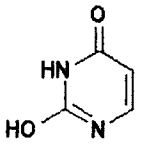
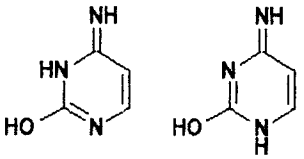
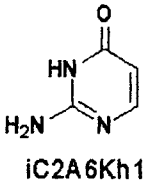
(4-ii)

Least-squares computation and direct estimation via eq. (2−3) produce $\Delta E^{tauto}_{Ah1\rightarrow h3} \approx -2.17(60)$ and $\approx -2.66(125)$, respectively. Comparable guanine and xanthine conversions (of type "exot") yield $\Delta E^{tauto}_{Gh1\rightarrow h3} \approx -4.49(62)$ and $\Delta E^{tauto}_{Xh1\rightarrow h3} \approx -5.14(65)$.

(iii) Conversion A6xh17 → A6xh19 (Adenine Isomers with Imino Substituent). Least-squares treatment gave the estimate $\Delta E^{tauto}_{Ah17\rightarrow h19} \approx +1.87(7)$, which might be related to the following guanine and xanthine conversions (type "noco"): $\Delta E^{tauto}_{Gh17\rightarrow h19} \approx +1.05(71)$ and $\Delta E^{tauto}_{Xh17\rightarrow h19} \approx +1.36(34)$ (cf. Table 8).

Examples (i) and (iii) rather closely obey the rule put forward above: though the substituents in the 2- and 6-position of the adenine isomers (H−, NH═), of the guanine isomers (H_2N−, O═), and of the xanthine isomers (HO−, O═) have different chemical nature, connectivity to and within the ring moiety is the same in the three cases. In example (ii), analogous

TABLE 9: Pyrimidine Nucleic Acid Bases: Tautomerization Energies (kcal/mol, HF/6-31G(d,p) Data)

conversion	ΔE^{tauto}		conversion	ΔE^{tauto}	
 U2K6yh1 U2K6yh3	-1.43(20)	keto	 iC2x6yh3 iC2x6yh3	-0.96(35)	norm
 C2K6Ah1 C2K6Ah3	-2.04(35)	noco	 G2x6yh19 G2x6yh39	-1.51(150)	norm
 U2x6Kh1 U2x6Kh3	+10.34(29)	keto	 C2x6yh1 C2x6yh3	+7.33(15)	norm
 iC2A6Kh1 iC2A6Kh3	+10.11(100)	noco			

situations with regard to the chemical graphs exist, but the respective tautomerization energies (adenine vs guanine and xanthine) show differences exceeding error estimates. No reasons for the larger discrepancy may be offered at this time.

The apparent (energetic) equivalence of single bonds H−C_2 (adenine) with bonds H_2N−C_2 (guanine) and HO−C_2 (xanthine) appears surprising. Table 8 also shows the latter two fragments to be approximately equivalent, likewise the pair HN=C_2, O=C_2. The data listed in Table 9 confronting tautomerization energies of "comparable" pyrimidine and purine nucleic acid bases supports this view: tautomerization energies seem to be approximately equal for isomers possessing graphs with analogous connectivity; the chemical nature of the substituents commonly occurring with these bases appears relatively immaterial.

5. Conclusions and Final Remarks

5.1. Additive Increment Model. The inclusion of guanine, xanthine, and a number of pyrimidine and purine model compounds allowed noticeable extensions and refinements of the additive increment model. On the side of local interactions (attractions and repulsions), some new structural fragments associated with specific interactions emerged. Furthermore, data analyses embracing the complete set of available data lead to recognition of a number of more global structural features (type arom, exot, etc.), to which local interactions with specific, distinctly different values should be associated. Though these pecularities create a slightly more complex energetic behavior of the nucleic acid bases, the principal aspects of the linear model and its use remain simple. On the other side, the necessity to refine the original model reflects that even in a set of apparently closely related organic compounds there persists a domain of individual behavior.

The limits of the additive increments model are demonstrated by a number of findings. First, this follows from the fact, that least-squares error estimates were but slightly reduced upon taking into account purine bases. Second, it was observed that transfer of model compound data requires careful examination with regard to systematic discrepancies. It therefore seems not possible to improve precision, error estimates, and reliability of the present results by further enlargement of the database.

Intimately connected with limited accuracy and transferability is the question of possible domains of applications. On the basis of presently available results, the model may be used for estimation of geometrical isomer stability of a wide variety of derivatives of nucleic acid bases (e.g., hypoxanthine, alloxanthine, ureic acid, etc.) and tautomeric forms of amino and hydroxy derivatives of 6-ring aromatic N-heterocycles (azines, diazines, triazines, mono- or polycyclic).

In all applications, it should be kept in mind, that the present estimations hold for free molecules. They also hold within error estimates for determination of $\Delta U^0(0)$ and $\Delta U^0(298)$ values.

5.2. Rules for Use. The following recipe for applications might be followed.

(i) Examine the graphs of the chemical structures to be studied for structural fragments associated with local interactions (attractions, repulsions); see subsection 2.2 and section 4 for a catalog of fragments incorporated in the present model.

(ii) Inspect the graphs for eventual global structural features (type norm, arom, etc.).

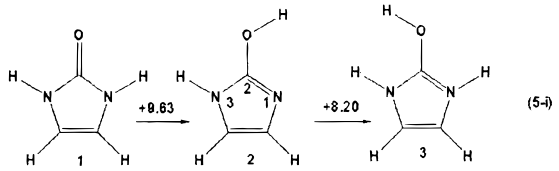
(iii) Estimate energies of conversion (including errors) of geometric isomers using eqs 2-2 and estimates for increments collected in Tables 3 and 4.

(iv) To estimate tautomer conversion energies, use eqs 2−3,

provided the graphs of the structures involved are comparable (see section 4) and total electronic energies are available.

For coarse estimation of relative energies of geometric isomers which feature structural fragments similar to but not identical with those listed above, it is suggested to proceed as follows. Ascribe attractions $A = -4.2$ to -5.2 kcal/mol and -2.7 to -3.2 kcal/mol to structural fragments comparable to 2-i through 2-iv. For repulsions similar to 2-v and 2-v′ use R values of 5.0(5) and 1.3−1.8 kcal/mol, respectively. For repulsions comparable to 2-vi and 2-vii use R values = 2.4−2.7 kcal/mol, respectively. Examine the structures to be compared for features such as "arom", "path", etc. and eventually adapt A and R values accordingly.

A number of examples have been reported earlier.[2,3] As a further application, we may consider the following conversions (HF/6-31G(d,p) results, kcal/mol)



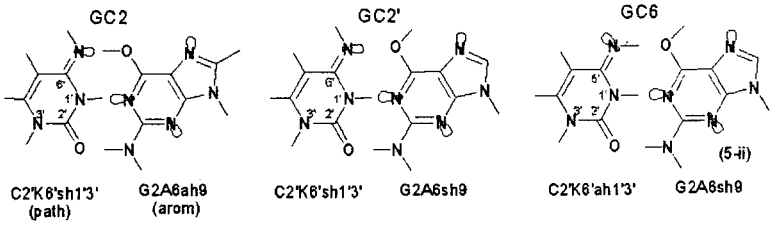
which involve geometric isomer conversion of a hydroxyl substituent attached to an imidazole system. None of the structural fragments so far considered relevant in this work apply in a straightforward manner. Coarse estimation yields

$$\Delta E_{2\rightarrow 3} \approx -A(\text{aHO.2}|\text{sp}^2\text{1}) + R(\text{sHO.2}|\text{H.3}) \approx -(-4.7(5)) + 5.0(5) = +9.7(10)$$

The discrepancy to the quantum chemical value (+8.2) exceeds the error estimate, but this appears tolerable, keeping circumstances in mind. Alternatively, the example illustrates the limitations immanent in coarse estimations.

It earlier has been found with suitably chosen model compounds that protonation of keto groups leads to $=\text{Op}^{+1}$ groups which experience repulsive and attractive interactions of similar magnitude as −OH groups.[1] This observation originally provided a number of zeroth order estimates for increments related to −OH substituents in nucleic acid bases. It might be useful for examination of conversion of conformations of protonated keto groups of nucleic acid bases and related compounds.

As a final example, predictive estimation of energy of conformers and single intermolecular H-bridges will be considered based on quantum chemical data for guanine−cytosine base pairs published by Florian et al[15] (E denotes energy increments of intermolecular H-bridges, increments assigned to each pair ordered into columns, primed numbers relate to cytosine[3]). From HF/6-31G** data, $\Delta E_{\text{GS2}\rightarrow\text{GS6}} = +6.5$ kcal/mol. From the equations above and entries for appropriate increments in Table 3, one obtains $-E(\text{aHO.6}|\text{sp}^2\text{N:6′}|\text{O:2′}) + 1.0(4) + 1.7(2) = +6.5$ kcal/mol; that is, the binding energy of the single H-bond (sp^2N:6′|aHO.6) is coarsely estimated to −3.8 kcal/mol. This estimate encompasses systematic errors, e.g., increments for a planar guanine moiety in GC6 were assumed. Furthermore, the intermolecular H-bridges in the pairs must compete with local (intramolecular) interactions, implying systematic alterations of both external and internal interactions (analogous to the situation prevailing in type 2-viii double interactions). At present, no estimates of such saturation effects



A(H.1′|sp²N:6′|O:2′)	A(H.1′|sp²N:6′|O:2′)	R(aHN:6′|H.1′|O:2′)
R(sHN:6′|H.5′)	R(sHN:6′|H.5′)	-
A(aHNH.2|sp²1)	A(aHNH.2|sp²1)	A(aHNH.2|sp²1)
A(sHNH.2|sp²3)	A(sHNH.2|sp²3)	A(sHNH.2|sp²3)
A(H.9|sp²3)	A(H.9|sp²3)	A(H.9|sp²3)
$\Delta A_{23\text{s}9}$	$\Delta A_{23\text{s}9}$	$\Delta A_{23\text{s}9}$
A(aHO.6|sp²1)	A(sHO.6|sp²7)	A(sHO.6|sp²7)
ΔA_{216}	-	-
E(H.1′|sp²1)	E(H.1′|sp²1)	E(H.1′|sp²1)
E(aHNH.2|O:2′)	E(aHNH.2|O:2′)	E(aHNH.2|O.2′)
E(aHO.6|sp²:6′|O:2′)	-	-

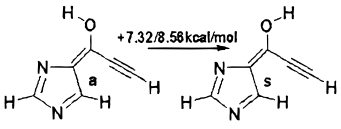
$\Delta E_{GC2\rightarrow GC2'} \approx$
-A(aHO.6|sp²1)
$-\Delta A_{216}$
-E(aHO.6|sp²N:6′|O:2′)
+A(sHO.6|sp²7).

$\Delta E_{GC2'\rightarrow GC6}$ (planar) ≈
-A(H.1′|sp²N:6′|O:2′)
+R(aHN:6′|H.1′|O:2′)
-R(sHN:6′|H.5′)

are known; presumably these will be of the same magnitude as for local effects. The energy alteration of the process GC2′ → GC6(planar) coarsely amounts to +1.7(3)kcal/mol, reflecting the delicate interplay with local interactions. Both by quantum chemistry[15] and by the Legon−Millen rule,[24] GC6 is predicted to be nonplanar.

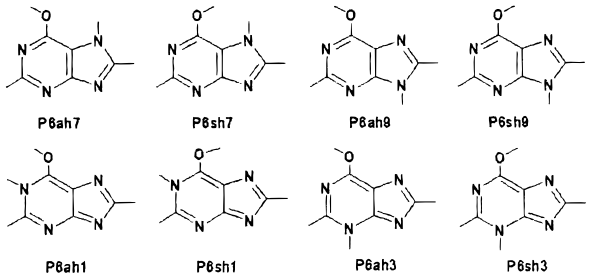
Appendix: Model Compounds

1. Acethylenes.



Etot(au)	Etot(au)
-413.203962	-413.192299(HF/6-31G(d,p))
-414.514940	-414.501302(MP2/6-31G(d,p))

2. Scheme Hydroxypurines[a].



[a]Hydrogen substituents indicated by single bond symbol.

Acknowledgment. The authors wish to thank the ETHZ administration for generously granting computing time and Mrs. K. Gunthard for typing the manuscript.

References and Notes

(1) Ha, T.-K.; Gunthard, H. H. *J. Mol. Struct. (THEOCHEM)* **1992**, *276*, 209.

(2) Ha, T.-K.; Gunthard, H. H. *J. Am. Chem. Soc.* **1993**, *115*, 11939.

(3) Ha, T.-K.; Gunthard, H. H. *J. Mol. Struct. (THEOCHEM)* **1993**, *300*, 619.

(4) Ha, T.-K.; Keller, H. J.; Gunde, R.; Gunthard, H. H. *J. Mol. Struct. (THEOCHEM)* **1996**, *376*, 375.

(5) Ha, T.-K.; Keller, H. J.; Gunde, R.; Gunthard, H. H. *J. Mol. Struct. (THEOCHEM)* **1996**, *364*, 161.

(6) Ha, T.-K.; Gunthard, H. H. *J. Mol. Struct. (THEOCHEM)* **1992**, *259*, 229.

(7) Ha, T.-K.; Gunthard, H. H. *J. Mol. Struct. (THEOCHEM)* **1992**, *276*, 187.

(8) Quantum Chemistry Literature Data Base, Bibliography of ab initio calculations 1978−1980. Supplements to *J. Mol. Struct. (THEOCHEM)* 1982−1995; Elsevier: Amsterdam. The reader might consult this compilation for nearly complete citations.

(9) Sabio, M.; Topiol, S.; Lumma, W. C. *J. Phys. Chem.* **1994**, *94*, 1366.

(10) Florian, J.; Leszczynski, J.; Johnson, B. G. *J. Mol. Struct. (THEOCHEM)* **1995**, *349*, 421.

(11) Sevilla, M. D.; Besler, B.; Colson, A. D. *J. Phys. Chem.* **1995**, *95*, 1060.

(12) Gould, I. R.; Burton, N. A.; Hass, R. J.; Hillier, I. H. *J. Mol. Struct. (THEOCHEM)* **1995**, *331*, 147.

(13) Sponer, H.; Hobza, P. *J. Am. Chem. Soc.* **1994**, *116*, 709.

(14) Colominas, C.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **1996**, *118*, 6811.

(15) Florian, J.; Leszczynski, J. *J. Am. Chem. Soc.* **1996**, *118*, 3010.

(16) Gorb, L.; Leszczynski, J. *J. Am. Chem. Soc.* **1998**, *120*, 5024.

(17) Nonella, N.; Haenggi, B.; Dubler, E. *J. Mol. Struct. (THEOCHEM)* **1993**, *279*, 173.

(18) Leszczynski, J. *Chem. Phys. Lett.* **1991**, *181*, 123.

(19) El-Bakali Kassimi, N.; Thakker, A. J. *J. Mol. Struct. (THEOCHEM)* **1996**, *336*, 185.

(20) Moeller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618. Pople, J. A.; Binkley, J. S.; Seeger, R. *Int. J. Quantum Chem. Symp.* **1976**, *10*, 1.

(21) Frisch, M. J., et al. *Gaussian 94*, revision C. 3; Gaussian Inc.: Pittsburgh, 1995.

(22) The value of $\Delta E^{\text{tauto}}_{\text{i6}\rightarrow 11}$ in Table 7 of ref 4 should read −0.96 kal/mol.

(23) In Table 8 of ref 5, the value of $\Delta E^{\text{tauto}}_{9\rightarrow 7}$ should read +2.06 kcal/mol and the tautomerization energy in the last line is $\Delta E^{\text{tauto}}_{9\rightarrow 1}$.

(24) Legon, A. C.; Millen, D. J. *Faraday Discuss. Chem. Soc.* **1982**, *73*, 71−87, 127, 128.