Eur. Phys. J. D **20**, 421–430 (2002)
DOI: 10.1140/epjd/e2002-00144-9

THE EUROPEAN PHYSICAL JOURNAL D

Modeling of recognition sites of nucleic acid bases and amide side chains of amino acids. Combination of experimental and theoretical approaches

V.S. Shelkovsky[1], S.G. Stepanian[1], I.K. Galetich[1], M.V. Kosevich[1], and L. Adamowicz[2,a]

[1] Department of Molecular Biophysics, B. Verkin Institute for Low Temperature Physics and Engineering of the National Academy of Sciences of Ukraine, 47 Lenin avenue, Kharkov 61103, Ukraine
[2] IRSAMC, Laboratoire Physique Quantique[b], Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex, France and
Department of Chemistry, University of Arizona, Tucson, Arizona 85721, USA

Received 5 February 2002 and Received in final form 14 March 2002
Published online 13 September 2002 –

Abstract. A combined experimental-theoretical approach to modeling of building blocks of recognition complexes formed by nucleic acid bases and the amino-acids side-chain amino group is reviewed. The approach includes the temperature dependent field-ionization mass spectrometry and *ab initio* quantum chemical calculations. The mass spectrometric technique allows determination of interaction enthalpies of biomolecules in the gas phase, and the results it produces are directly comparable to the results obtained through theoretical modeling. In our works we have analyzed both thermodynamic and structural aspects of the recognition complexes of four canonical nucleic acid bases and acrylamide, which models the side chain of asparagine and glutamine. It has been shown that all bases can interact with amide group of the amino acids *via* their Watson-Crick sites when being incorporated into a single strand DNA or RNA. Stability of the complexes studied, expressed as $-\Delta H$ (kJ mole^{-1}) decreases as: m^9Gua (-59.5) > m^1Cyt (-57.0) > m^9Ade (-52.0) ≫ m^1Ura (-40.6). We have determined that in the double stranded DNA only purine bases can be recognized.

PACS. 87.15.Aa Theory and modeling; computer simulation – 87.15.Kg Molecular interactions; membrane-protein interactions – 87.14.Gg DNA, RNA – 33.15.Ta Mass spectra

1 Introduction

The importance of the molecular recognition and especially the protein-nucleic acid recognition governing many biological processes has resulted in many investigations [1,2]. The main problem encountered in studies of the bio-recognition phenomenon has been to distinguish the events directly involved in the recognition from all other non-specific non-covalent interactions occurring in biological systems and in biochemical reactions. To overcome this problem, model systems that allow reconstructing of particular steps involved in specific interactions have been usually considered in the laboratory experiments. These model systems have different levels of complexity and range from native and synthetic biopolymers to smaller parts of biopolymers, which reproduce particular molecular fragments participating in the recognition processes, and to simple monomers active in the recognition cites. At the level of biopolymers complexes, various techniques (X-ray crystallography, NMR, IR and UV spectroscopy, mass spectrometry) have been routinely utilized to acquire structural data. However energy characteristics of the specific recognition sites have been more difficult to obtain because the energy measurements usually bear integral character and do not permit to isolate the yields corresponding to separate sites. The obtain site-specific energy effects one need to resort to studying interactions at the level of monomers, which in protein-nucleic acid complexes are nitrogen bases and amino acids.

An approach to effectively model building blocks of the recognition complexes at the level of monomers has to combine experimental and theoretical methods. For computational modeling of medium size complexes correlated level *ab initio* methods such as DFT or MP2 are the most useful. To directly compare the results obtained from theoretical calculations with the experimental results, it is desirable that the experimental technique allows studying associates of isolated biomolecules without presence of solvents or other environmental factors. This implies the use of gas phase or vacuum conditions in the experiment. The method of the temperature-dependent (TD) field-ionization (FI) mass spectrometry has these advantageous and it has been used in our investigations. The method allows studying of the thermodynamics of

[a] e-mail: ludwik@u.arizona.edu
[b] UMR 5626 du CNRS

intermolecular interactions of biomolecules in isolation. In this capacity the method was employed to determine enthalpies of formation of the Watson-Crick nucleic acid base pairs [3,4]. Those early studies were later followed by other groups (for example [5,6]), who have continued investigating DNA base pairs in the gas-phase. A description of the procedure and, in particular, its application to determine enthalpies of site-specific interactions in bio-complexes is given below in Section 2.1.

The aim of the study reviewed here is to investigate the structural and energetic parameters of particular types of complexes that involve nucleic acid bases and the amide group of the side chains of the amino acids asparagine (Asn) and glutamine (Gln). In the study we have applied a combination of the experimental TD FI mass spectrometric technique and theoretical quantum-chemical calculations. Analysis of the available X-ray structural data of protein-nucleic acid complexes reveals a number of recognition sites which involve direct contacts of amide group of Asn and Gln with nitrogen bases adenine (Ade) [7–13], uracil (Ura) [14–16], guanine (Gua) [10,17,18] and cytosine (Cyt) [19–21]. In general, asparagine-adenine contact is one of the most frequently encountered in the recognition motifs [22,23].

Selection of model complexes to study recognition processes in the present study was predetermined by the constraints of the experimental technique. It requires that a sufficient concentration of the monomer mixture in the gas phase can be generated to allow study of the intermolecular complexes that the monomers form in the gas phase. Four canonical nucleic acid bases, Ura, Cyt, Ade and Gua, have been considered. Their volatility was additionally increased by methylation. Acrylamide (Acr) was chosen to mimic the amide group of the side chains of Asn and Gln amino acids; this compound has the advantages of a high volatility and the absence of any other H-bond-forming groups apart from the amide group. The question about the correspondence between the interaction of a base and a separate side chain of the amino acid and the interaction of the base with the whole amino-acid complex has to be considered. Unfortunately, thermal instability and low volatility of the Asn and Gln amino acids hampers a direct experimental measurement of complexes involving these systems. Under these circumstances a comparison of theoretical data calculated for the acrylamide-base complexes with the data obtained for the amino acid-base complexes become the only source of validation that the choice of acrylamide as a model of the amino acids side chains is adequate.

2 Methods

2.1 Temperature dependent field-ionization mass spectrometry

The general idea of an experiment aimed at obtaining interaction energies of molecules in the gas phase, which main steps are schematically shown in Figure 1, is rather simple. It is necessary to create a gas-phase mixture of the

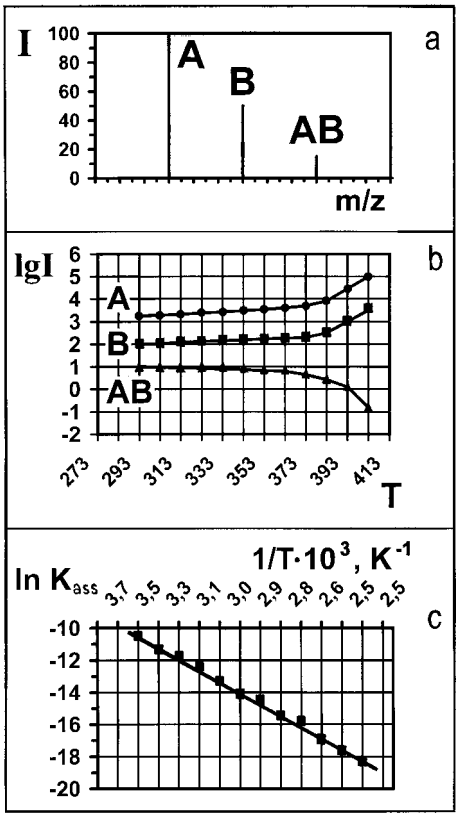

Fig. 1. A scheme of any, including TD FI mass spectrometric, experiment for determination of enthalpies of interaction of biomolecules in the gas phase: (a) determination of the composition (type and concentration of the components) of the reactive system; (b) recording of the changes in the system with variation of its temperature; (c) construction of van't-Hoff's plots as $\ln K_{\rm ass}$ *vs.* $1/T$ and calculation of ΔH; the total T range of measurements in the present mass spectrometric experiments is shown on the schemes, the ranges for individual systems fall within this range.

two components (A and B) whose molecules form dimers. The dimer formation reaction, $\mathrm{A} + \mathrm{B} \Leftrightarrow \mathrm{AB}$, is characterized by the association constant $K_{\rm ass}$:

$$K_{\rm ass} = \frac{[\mathrm{AB}]}{[\mathrm{A}][\mathrm{B}]} \tag{1}$$

(see Fig. 1a) which can be determined by measuring the concentrations [A], [B], and [AB]. Then it is necessary to affect the equilibrium by changing the temperature T and to remeasure the concentrations of the components [A], [B], and [AB] (see Fig. 1b) to determine the new association constant $K_{\rm ass}$. Using these data van't-Hoff's plot, that describes the dependence of $K_{\rm ass}$ on the inverse of the absolute temperature, $1/T$, can be constructed (Fig. 1c) according to the equation:

$$\ln K_{\rm ass} = -(\Delta H/RT) + const. \tag{2}$$

From this relation the enthalpy of the interaction, ΔH, can be determined:

$$\Delta H = -R\left[\delta\left(\ln K_{\rm ass}\right)/\delta(1/T)\right]. \tag{3}$$

However, a practical realization of this scheme, widely used in thermodynamic studies of gaseous and highly volatile compounds, encounters significant difficulties when applied to biological molecules. The main problem is to generate gas-phase complexes of thermally unstable molecules at equilibrium with a concentration sufficient for the experimental analysis. It is also important to use a method of study that is non-destructive and would least affect the investigated system. Some spectroscopic techniques could be considered for such studies [24], but as the literature data show, it would be very difficult to study the systems being the subject of the present investigations with such methods because of their low volatility. For example, IR spectroscopic experiments concerning associations of nucleic acid bases gave reliable results only for the most volatile m^1Ura base [3,25]. The main problem in IR spectroscopic studies can also be to distinguish in the IR spectra the bands of monomers, homo and hetero associates.

The advantage of the mass spectrometric technique is its high sensitivity and its sensitivity in distinguishing components of the reactive system by their mass (although a problem of non-destructive so-called "soft ionization" exists). The transfer of intact biomolecules to the gas phase has usually involved desorption from the condensed (solid and liquid) state [26,27]. Among the gas-phase ionization techniques most frequently used, there are some that cannot be applied to the systems studied here. For example, the electron impact technique causes strong fragmentation, the photo ionization is only applicable to a certain group of compounds, and in the chemical ionization, ion-molecule reactions often interfere with the interaction of the neutral monomers. The latter method was, however, successfully used in thermodynamic studies of proton affinities and the neutral-ion binding [28,29]. Among the numerous ionization methods, the soft ionization technique, FI [30], appears to be the most efficient in studies of interactions of simple biomolecules, and the TD FI mass spectrometry technique has been refined through the years of application and has gained popularity [3,4,31–33].

The TD FI mass spectrometry method allows us to avoid most of the problems listed above. First of all, it causes very little distraction of the studied system. Moreover, the soft mode of the ionization *via* tunnel loss of electron (in high electric field with strength of the order of 10^{10} V m^{-1}) has very small effect on the weakly H-bonded associates. On the other hand, field extraction of rapidly formed positive ions deletes them fast from the reactive zone, thus preventing the ion-neutral interactions to form. The most important feature of FI is the possibility of creation of an equilibrium gas-phase system at a very well controlled temperature and at a pressure sufficient for multiple collisions between volatile biomolecules to occur and for efficient formation of van-der-Waals and hydrogen-bonded complexes.

The "hart" of the FI ion source is shown in Figure 2. Molecular flows of two substances are supplied to a relatively small volume within a cylindrical counter-

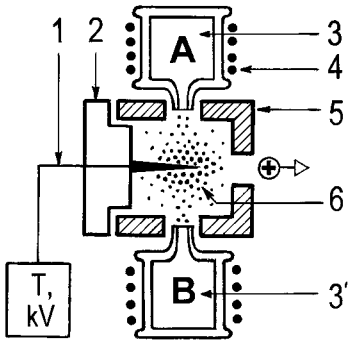

Fig. 2. A scheme of the creation of the reaction zone in the FI ion source: 1 - needle type FI emitter connected to voltage supply and resistive heating system; 2 - Teflon packing; 3 - glass evaporators with the studied substances; 4 - heaters of the evaporators; 5 - cylindrical counterelectrode; 6 - reaction zone.

electrode (5) from two heated glass evaporators (3). Careful balancing of all experimental parameters (such as control of flows of compounds of significantly different volatility by separate heating of evaporators, prevention of electrical discharges and thermal destruction or condensation, etc.) is necessary to create an equilibrium gas phase system (6). It is important, that the interactions in this system occur among the neutral molecules. The temperature of the gas phase is easily controlled by changing the temperature of the FI emitter (1) by its resistive heating. The molecules entering the gas phase acquire the temperature of the FI emitter due to multiple collisions with its surface. The relation between the three characteristic times, namely the lifetime of the molecules in the neutral state prior to their ionization, τ_1, the average time span between the collisions of the molecules, τ_2, and the time for removing the ions formed from the reactive zone, τ_3:

$$\tau_1(10^{-6}\text{ s}) \gg \tau_2(10^{-11}\text{ s}) > \tau_3(10^{-12}{-}10^{-13}\text{ s}) \qquad (4)$$

shows that, on the one hand, there is more than enough time for neutral molecule interactions and efficient thermalization to occur and, on the other hand, the probability of the ion-molecule interactions is vanishingly small. Whether the thermal equilibrium has been establish in the reaction zone can be easily verified by checking how constant are the values of the ion currents of the components of the mixture and, consequently, by how constant K_{ass} has become. It is assumed, that the ion currents of the compounds, I_{A}, I_{B}, I_{AB}, recorded in the FI mass spectra of the compounds are proportional to their concentrations [A], [B], [AB], if the compounds have similar physical parameters and ionization potentials (Fig. 1a) in the gas phase mixture [3,34].

The so-called field attraction effect experienced by polar molecules in a high electric field is responsible for the trapping effect which results in formation of a microscopic zone of enhanced density near the FI emitter surface [30,34] with the gas pressure sufficient for formation of molecular associates. The pressure expressed in terms

of the number of molecules, n, per cm^3 can be determined using the barometric equation:

$$n(F) = n_0(T) \exp[(\mu F + (\alpha/2)F^2)/kT], \qquad (5)$$

where μ is the dipole moment, α the polarizability, F potential at the FI emitter, $n_0(T)$ the pressure at zero field at a given temperature T. For the nucleic acid bases, $n(F)$, calculated for the temperature in the range of 300–400 K, is equal to 10^{19}–10^{20} cm^{-3} in average.

The values of the enthalpies of the associate formation determined using FI technique could be dependent on the field strength [35]. It is known that polar molecules tend to align along the field direction, and the "stretching" effect of a nonhomogeneous electric field can influence association of such molecules. The alignment of the monomer dipoles along the direction of the field can cause either stabilization or destabilization of the dimer: the destabilization effect is the strongest for dimers with antiparallel orientation of the dipole vectors of the monomers. In the previous studies on the homodimerization of nitrogen bases, noticeable field dependencies were observed only for cytosine derivatives [36]. This was not unexpected because these compounds have rather large dipole moments. A more detailed analysis of the field stretching effect and its relation to the TD FI technique can be found elsewhere [3,4,31].

The ion beam formed in the FI ion source is subjected to a standard mass spectrometric analysis *via* separation of the particles in accord with their mass-to-charge ratio, m/z. Mass spectrometric measurements were performed using the sector magnetic mass spectrometer MI-1201 ("Electron" Works, presently "SELMI" JSC, Sumy, Ukraine), equipped with a FI ion source. Needle-type tungsten FI emitters were used. Resistive heating of the emitter by the electric current of 0–1.2 A allowed variation of the temperature of the reactive zone in the range of 293–423 K. Emitter was usually kept at the potential of 5 keV; field dependencies were obtained at 2.5, 3, 4, 5 keV.

Methylderivatives of nucleic acid bases listed in Table 1 were used to increase the volatility of the compounds. To avoid the distortion of the recognition sites, the methylation was performed at the $N1$ atom in pyrimidines and at the $N9$ atom in purines. These sites of the bases are involved in glycoside bond in the corresponding nucleotides. In Table 1 the enthalpies of sublimation ($\Delta H_{\rm subl}$), which reflect the volatility of the bases [3], and the temperature ranges for the heating which provided sufficient molecular flows without thermal destruction of the bases are presented. Acrylamide is characterized by high volatility at the room temperature. Thus in the experiments, the acrylamide crystals were placed in the evaporator only in small quantities and they were not heated.

The methylated nucleic acid bases were synthesized at the Chemistry Department of the Kharkov State University (Ukraine); acrylamide was purchased from Reanal (Hungary).

Table 1. Methylderivatives of nucleic acid bases under study: their label, monoisotopic mass M, enthalpy of sublimation $\Delta H_{\rm subl}$ and temperature range of optimal volatility.

Nitrogen base	M, Da	$\Delta H_{\rm subl}$ [3]	T range, K
m^1Ura	126	112.2 ± 1.7	378–418
m^1Cyt	125	142.7 ± 1.5	418–460
$m_2^{1,N}$Cyt	139	140.8 ± 2.1	395–445
$m_2^{1,N,N}$Cyt	153	110.9 ± 1.7	373–420
m^9Ade	149	121.0 ± 1.3	403–468
m^9Gua	165	173.3 ± 1.7	453–532

2.2 Quantum chemical calculations

All possible dimers of acrylamide with the nucleic acid bases stabilized by at least two intermolecular H-bonds were analyzed in the calculations. For the Acr-m^9Gua complex we have taken into account the possible coexistence in the gas phase of three different m^9Gua tautomers: the keto tautomer (**Gk**), the enol tautomer (**Ge**) and the enol tautomer with the trans orientation of the hydroxy group (**Get**). These tautomers were found to exist for m^9Gua in the gas phase frozen in the low temperature inert gas matrices [37–39]. In total two Acr-m^1Ura, one Acr-m^1Cyt, one Acr-$m_2^{1,N}$Cyt, two Acr-$m_3^{1,N,N}$Cyt, two Acr-m^9Ade and eight Acr-m^9Gua dimers were considered.

The interaction energies in the dimers have been calculated using the DFT (Density Functional Theory) and MP2 (Møller-Plesset Second-Order Perturbation Theory) methods [40,41]. The DFT method was also employed in the calculations of the harmonic frequencies, which were used to determine the ZPVE (Zero-Point Vibrational Energy) corrections. The DFT calculations were carried out with the three-parameter density functional called B3LYP, which includes Becke's gradient exchange correction [42], the Lee, Yang, Parr correlation functional [43] and the Vosko, Wilk, and Nusair correlation functional [44]. The interaction energies of the dimers at both the DFT and MP2 levels were calculated with the account of the basis set superposition error (BSSE) by the counterpoise method of Boys and Bernardi [45]. The method involves a calculation of the total energy of the dimer and two calculations of the monomer energies performed with the basis set of the dimer. The BSSE-corrected interaction energy was calculated for each dimer as the difference of the dimer energy and the BSSE-corrected monomer energies. The BSSE correction for each monomer was determined as the difference of the monomer energy calculated with the basis set of the monomer and the energy calculated with the basis set of the dimer.

Initially the geometries of all the dimers were fully optimized at the DFT/B3LYP/6-31++G** level of theory and used to calculate the DFT harmonic frequencies. For the Acr dimers with m^9Ade and m^9Gua, the DFT/B3LYP/6-31++G** geometries were used to calculate the MP2/6-31++G** dimer interaction energies. Geometries of the Acr dimers with pyrimidine bases were also calculated at the MP2/6-31+G* level of theory.

Table 2. Experimentally determined enthalpies of formation (ΔH, kJ mole^{-1}) of recognition complexes of methylated nucleic acid bases with acrylamide and glutamine and calculated interaction energies (IE, kJ mole^{-1}) of these complexes[a].

	Parameters of base-acrylamide dimer				Base-glutamine dimers		
Base	Experiment	Theory			Theory		
	$-\Delta H$	Structure label	IE	Ref.	Structure label	IE	Ref.
m^1Ura	40.6 ± 4.2	I	51.4	[47]	-	-	
		II	47.8				
m^1Cyt	57.0[b]	III	54.8	[48]	-	-	
$m_2^{1,N}$Cyt	58.7[b]	IV	58.3	[48]	-	-	
$m_3^{1,N,N}$Cyt	45.7[b]	V	33.2	[c]	-	-	
		VI	32.6		-	-	
m^9Ade	52.0 ± 5.0	VII	40.9	[49]	VII′	48.2	[49]
		VIII	38.8		VIII′	45.1	
m^9Gua	59.5 ± 3.8	IX (**Gk**)	65.0	[50]	IX′	65.1	[50]
		X (**Gk**)	36.7		-	-	
		XI (**Ge**)	43.5		XI′	43.3	
		XII (**Ge**)	30.8		-	-	
		XIII (**Ge**)	29.0		-		
		XIV (**Get**)	44.5		XIV′	44.6	
		XV (**Get**)	29.9		-	-	
		XVI (**Get**)	41.2		-	-	

[a]The calculated interaction energies of the pyrimidine bases were obtained at the MP2/6-31+G* level for the equilibrium geometries fully optimized at the same level of theory. The interaction energies of the purine bases were calculated at the MP2/6-31G++G** level for the DFT/B3LYP/6-31++G** geometries. All energies were calculated with accounting for the BSSE and ZPVE corrections. [b] Extrapolated to zero field. [c] Data of the present work.

The comparison of the MP2 and DFT interaction energies between acrylamide and pyrimidine bases demonstrated that the DFT method is capable of producing the interaction energies that are very similar to the ones obtained at the MP2 level of theory. Finally, DFT/B3LYP/6-31++G** calculations were also performed for the Glutamine (Gln)-m^9Ade and Gln-m^9Gua systems.

All calculations were done on the IBM Power 3 and SGI ORIGIN 2000 computers using the Gaussian98 quantum-chemical program package [46].

3 Results and discussion

The enthalpies of formation (ΔH) of the acrylamide-base dimers obtained experimentally using the TD FI mass spectrometry technique, as well as the interaction energies calculated at the DFT and MP2 levels of theory, are summarized in Table 2. Theoretical values of the complexes formation energies are given for a number of equilibrium dimer structures labeled I–XVI shown in Figures 3–5. As the calculated values of the interaction energies indicate, several types of H-bonded dimers could be formed in the gas phase. The close interaction energies of the recognition complexes make it difficult to estimate their relative yields for some of the studied systems.

In relation to the question about the correspondence between the interaction of a base with a separate side chain to the interaction of the base with the whole amino acid, the following can be concluded. Good agreement between the experimental and the theoretical data for most base-acrylamide systems (Tab. 2) allows the use the theoretical data to determine whether the base-acrylamide dimers are good models for the interactions between the bases and the amino acids. The calculated results used in this determination concern complexes of the purine bases, m^9Ade and m^9Gua, with Gln. These results are shown in Table 2.

Let us analyze now the results for interaction complexes of acrylamide with the different bases. For the Acr-m^1Ura system we found two possible dimer structures (denoted as I and II) which are shown in Figure 3. Both complexes are planar and stabilized by two pairs of N–H···O H-bonds, which form a cyclic structure. The O4 and N3H groups of the base in the dimer I and the O2 and N3H groups in dimer II are involved in the H-bonding. The two associates have comparable interaction energies.

In the case of the complexes of the cytosine methylderivatives, the analysis of the experimental results was complicated by the need to account for the dependence of the enthalpy of the formation of the complex on the field strength. This effect was more pronounced in this case because of a larger dipole moment of cytosine (see Sect. 2.1). Similar field effect on the interaction enthalpy was previously observed for homodimers of some

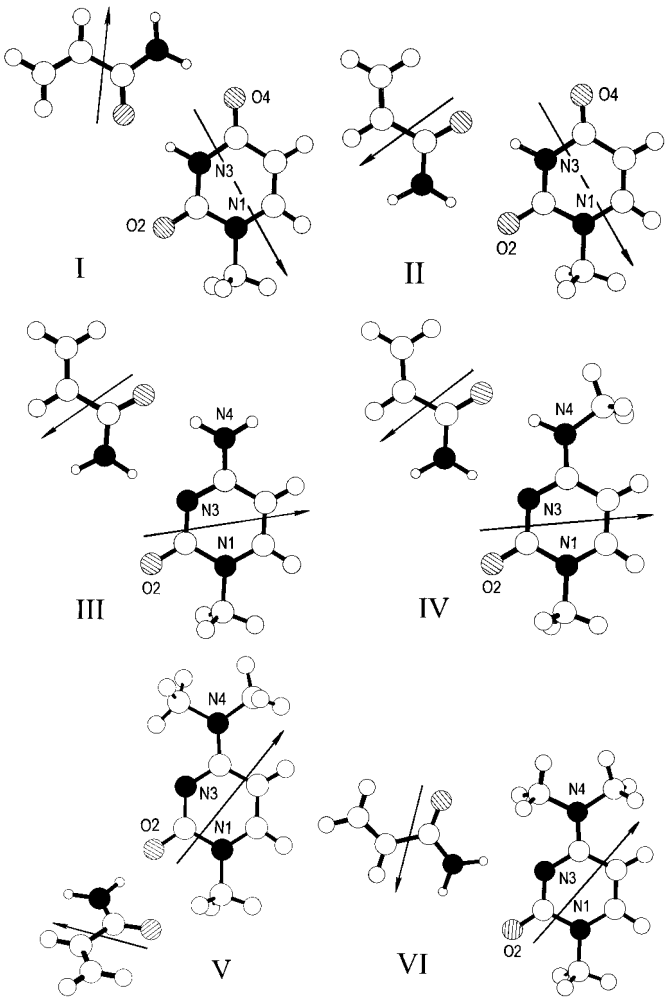

Fig. 3. Structures for acrylamide dimers with methylderivatives of pyrimidine bases m^1Ura (I, II), m^1Cyt (III), $m_2^{1,N}$Cyt (IV), $m_3^{1,N,N}$Cyt (V, VI). The equilibrium geometries were calculated at MP2/6-31+G* level of theory for structures I-IV and on DFT/B3LYP/6-31++G** level for structures V, VI [45,46]. The vectors of dipole moments of the monomers are shown for Acr-m^1Cyt and Acr-$m_2^{1,N}$Cyt complexes. Nitrogen atoms are shown in black, oxygen (hatched), carbon and hydrogen (smaller circles, blank); the atoms numbering is in accord with a standard nucleic acids nomenclature [1].

nitrogen nucleic acid bases [36]. We should add that the cytosine-containing acrylamide-base complexes were the only complexes where field-dependence of ΔH was observed. For other bases the effect was negligibly small. The dependence of ΔH on the potential of the FI emitter were measured for the three Acr-Cyt methylderivatives complexes. The curves describing the dependencies were extrapolated to the zero field [48]. The ΔH values resulting from this extrapolation are shown in Table 2.

The equilibrium structures for the Acr complexes with three methylated cytosines (denoted as III-VI) are presented in Figure 3. Single, almost planar structures were found both for the Acr-m^1Cyt (III) and Acr-$m_2^{1,N}$Cyt (IV)

Table 3. Dipole moments (μ) of some pyrimidine bases and their dimers with acrylamide calculated at the DFT/B3LYP/6-31++G** level of theory.

Base	μ, D		
	Base	Base-acrylamide pair	
m^1Ura	5.14	2.78 (I)	5.72 (II)
m^1Cyt	6.44	3.51	
$m_2^{1,N}$Cyt	6.61	3.49	
$m_3^{1,N,N}$Cyt	6.12	4.70 (V)	3.76 (VI)

dimers. In these complexes, the deviation from the planarity of about 15° occurs only at the amino group. The dimers are connected by N–H···O and N–H···N H-bonds. The N3 atom and the N4H site of the exocyclic amino group of the bases are involved in the binding.

Two Acr-m$_3^{1,N,N}$Cyt dimers (V and VI in Fig. 3) with close formation energies (see Tab. 2) were found. Dimer V is stabilized by a O2···H–N hydrogen bond and by a weak bonding interaction between acrylamide oxygen and the methyl group at N1 of the base; in dimer VI there is a N3···H–N H-bond and a weak attractive contact between the acrylamide oxygen and the methyl group at N4 of the base. Since steric constrains permit formation of only one H-bond in Acr-$m_3^{1,N,N}$Cyt dimer, the interaction energy in both complexes V and VI is lower, than that for the Acr-m^1Cyt (III) and Acr-$m_2^{1,N}$Cyt (IV) dimers stabilized by two H-bonds. At the same time, the experimental enthalpies of formation of the Acr-m^1Cyt and Acr-$m_2^{1,N}$Cyt dimers are noticeably, by about 17–19 kJ mole^{-1}, higher than that in the Acr-m^1Ura dimer. This is unexpected, since the latter has two H-bonds similar to the H-bonds in the Acr-m^1Cyt and Acr-$m_2^{1,N}$Cyt dimers. One can explain the additional stabilization of the Acr-m^1Cyt and Acr-$m_2^{1,N}$Cyt complexes by the proximity of the carbonyl oxygen, O2, of the bases to the acrylamide amino group.

As already mentioned, the orientation of the dipole moments of the molecules under study determines the magnitude of the field effect on the intermolecular interactions. In Table 3 the calculated values of the dipole moments of the methylated cytosines and their complexes with acrylamide are shown. The corresponding data for m^1Ura are also given for comparison. As shown in Figure 3, the dipole moment vectors of the individual molecules in the Acr-Cyt III, IV, and VI dimers, appear to be almost antiparallel and the resulting dipole moments of the dimers are rather low. This is consistent with the observed lowering of ΔH in strong fields, since the lower dipole moments of the dimers than of the corresponding monomers lead to higher stabilization of the monomers in the high field, which results in lowering of the dimerization enthalpies. In the case of m^1Ura, the dipole moment of the base monomer is similar to that of dimer II. In result, this structure under strong field conditions becomes more thermodynamically stable and it is produced in higher quantities. Although the structural differences cannot be directly distinguished on the basis of the FI mass spectra,

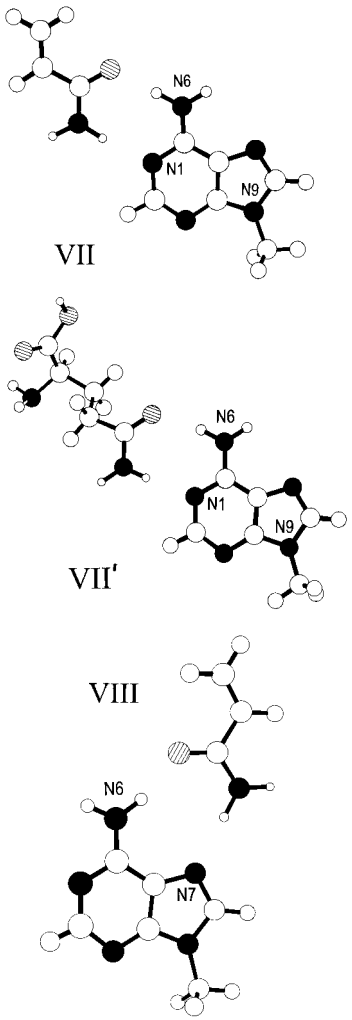

Fig. 4. Equilibrium structures of the Acr-m^9Ade dimers VII and VIII and Gln-m^9Ade dimer VIII′ calculated at the DFT/6-31++G** level of theory [49].

the absence of field dependence of ΔH for the Acr-m^1Ura system points tentatively to the dominance of formation of the dimer with the higher dipole moment. It is worth noticing, however, that the relative change of ΔH with the field strength, which constitutes the field effect, the smallest for $m_3^{1,N,N}$Cyt [48] which has the smallest dipole moment among the three cytosine derivatives (see Tab. 3).

For the Acr-m^9Ade complex, two almost planar equilibrium structures VII and VIII (Fig. 4) were found in the calculations. Similar structures (VII′ and VIII′) were found for the Gln-m^9Ade complex. In the dimers VII, VII′, the side chain of the amino acid interacts with the Watson-Crick sites of the base *via* N–H···O and N–H···N H-bonds which involve the N1 and N6A atoms of m^9Ade. In the dimers VIII, VIII′, the Hoogsteen sites are involved in the hydrogen bonding with the amide group of either Acr or Gln with the N7 and N6A atoms of m^9Ade. The calculations predict rather short H-bonds in all the dimers. Changes of the bond lengths of the bases due to formation of the complexes with acrylamide or glutamine are only 0.004–0.006 Å. The interaction energies are very close for the VII and VIII types of dimers, what indicates that both of them could exist in the gas phase. At the same time the interaction energies in the Acr-m^9Ade and Gln-m^9Ade complexes are almost identical and differ by only 1 kJ mole^{-1}. This allows us to conclude that acrylamide can indeed be used as a model of the amide group of the amino acid side chain.

It should be noted that for the Acr-m^9Ade complex the largest deviation between the experimental and theoretical values of the interaction energies among the studied systems is observed. Possible reasons causing such a deviation in the computations were discussed in reference [49].

The present results concerning the interaction sites in Acr-m^9Ade and Gln-m^9Ade complexes agree well with the results of modeling of the Asn-Ade interactions obtained from the interaction free energy surface [22]. The same preferential paring of Asn at the N6H and N7 sites of adenine (for Ade involved in the Ade-Thy pair corresponding to the double-strand DNA) was described in [22].

The calculations revealed the largest isomeric diversity for the Acr-m^9Gua complex, since guanine can exist in the gas phase in the three tautomeric forms: keto (**Gk**), enol (**Ge**) and enol tautomer with the *trans* orientation of the hydroxy group (**Get**) (see Sect. 2.2). All these tautomers have both proton donor and proton acceptor groups allowing multiple acrylamide connections *via* two H-bonds. In total eight H-bonded dimers formed by the three lowest-energy m^9Gua tautomers were found in the calculations (Tab. 2). Geometries of the most stable of these complexes for each of m^9Gua tautomer are depicted in Figure 5. All dimers are almost planar except for the methyl groups. Similarly to associates with the purine base adenine, acrylamide can interact with guanine at the Watson-Crick (IX, XI, XIII, XVI) and Hoogsteen (XIV) sites and the site exposed to the DNA minor groove (X, XII, XV).

The most important feature of these data is the significant difference in the interaction energy of the Acr-Gk IX dimer from the ones found for all other eight dimers. The lowest energy complex IX (Fig. 5) involves the keto tautomer of m^9Gua and is stabilized by two intermolecular H-bonds, O6···H–N and N1–H···O. It is more stable than the second lowest energy dimer by ≈25 kJ mole^{-1}, and the calculated energy of its formation, −65.0 kJ mole^{-1}, is very close to that obtained in the experiment, -59.5 ± 3.8 kJ mole^{-1}. Moreover, the interaction energy in the dimer IX is the only one that matches the experimental value. This leads to the conclusion that the dimer with the structure IX is preferentially formed in the experiment.

The reason of the highest stability of the structure IX in comparison with all other structures can be found in the specific distinct features of its geometry (Fig. 5). In the structure of this dimer, in addition to the two O6···H–N and N1–H···O H-bonds, there is another short (2.513 Å) bonding contact involving the N4–H···O interaction. This additional contact is similar to the additional contact enhancing the stability of the acrylamide-cytosine dimers III and IV, which we described above.

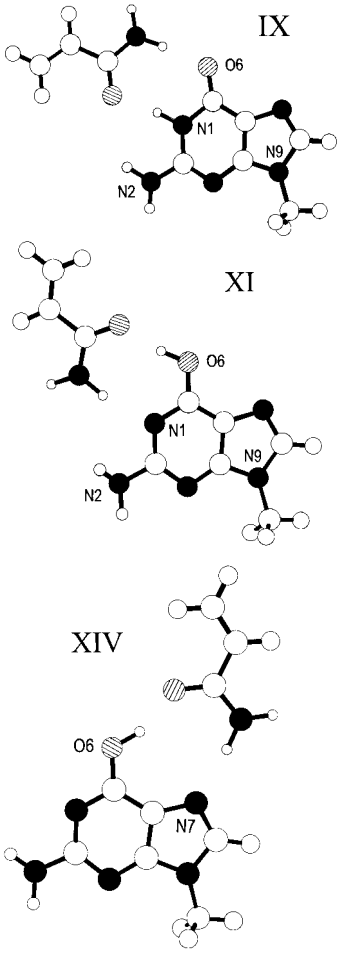

Fig. 5. Equilibrium geometries of the most favorable for each of three tautomers of guanine complexes with acrylamide: Acr-Gk (IX), Acr-Ge (XI), Acr-Get(XIV). Calculations performed at the DFT/6-32++G** level of theory [50].

For each guanine tautomer, possible structures of the complex with the amino-acid glutamine were determined in the calculations. The interaction energies in these complexes (IX′, XI′, XIV′), like in the adenine case, are similar to those in the corresponding Acr-m^9Gua complexes (see Tab. 2).

In relation to the recognition process, it is interesting to compare the experimental and the theoretical interaction energies (kJ mol^{-1}) in the acrylamide-base and Watson-Crick base pairs. Let us include in this comparison the Acr-m^9Ade (−52.0) and Acr-m^1Ura (−40.6) dimers and the m^9Ade-m^1Ura (−54.6) [3,36,51] base pair, as well as for the Acr-m^9Gua dimer (−59.5) and m^9Gua-m^1Cyt (−87.3) [36,51] base pair. Comparing the energy values for the pairs involving adenine and uracil and the dimers of each of them with acrylamide indicate that the amide group can compete with uracil in pairing with adenine. On the other hand, the enthalpy of the m^9Gua-m^1Cyt Watson-Crick pair formation is much higher than that of the Acr-m^9Gua dimer, which means that in this case the amide group cannot compete with cytosine in pairing with guanine. This explains the sensitivity of the transition temperature of the polynucleotides rich in Ade-Ura pairs[1] to the presence of acetamide [52].

It may be concluded that, in accord with the experimentally determined values of the enthalpies, $-\Delta H$ (kJ mol^{-1}), of the formation of the recognition complex, the Acr associates with different bases can be arranged in the following stability row:

$$m^9\text{Gua } (-59.5) > m^1\text{Cyt } (-57.0) > m^9\text{Ade } (-52.0) \gg m^1\text{Ura } (-40.6). \quad (6)$$

In most cases the measurements of the enthalpies of the dimer formation between acrylamide and the bases were carried out at the emitter temperature of 298 K. One may expect that at this relatively high temperature the relative stability of the dimers might be affected by the entropy factors. To elucidate this effect, we calculated the thermal corrections to the energy, to the enthalpy, and to the Gibbs free energy at the temperature of 298 K for two Acr-m^9Ade dimers (VII and VIII). These systems were chosen because they have very close calculated interaction energies (Tab. 2) and the thermal corrections might differentiate them in terms of other thermodynamic functions. The calculations were performed at the DFT/B3LYP/6-31++G** level of theory. We found that the thermal corrections to the energy and to the enthalpy were almost identical for the two dimers. The differences were less than 0.01 kJ mol^{-1}. The difference between the thermal corrections to the Gibbs free energy was larger but still very small (0.04 kJ mol^{-1}). Thus we can conclude that the thermal corrections do not significantly influence the calculated relative stability of the dimers.

4 Conclusions

Summarizing the above-described results of the combined experimental and theoretical investigations of the nucleic acid bases-acrylamide systems, it can be concluded that all studied bases can interact with the amide group of the side chains of the amino acids Asn and Gln. The most favorable equilibrium geometries found in the calculations agree with the "point contact model" which requires formation of at least two H-bonds in a recognition process.

Higher stabilization for some recognition complexes was found and explained as due to formation of a third additional acrylamide-base interaction contact. In the Acr-m^1Cyt (III) complex, this additional bonding contact involves the O2 atom of m^1Cyt (in addition to the O4 and N3H H-bond forming sites), and in the Acr-m^9Gua complex (IX) the additional contact involves N4H of m^9Gua (in addition to the O6 and N1H H-bond forming sites).

If only the energetic relations are considered, the stability of the complexes of acrylamide with nucleic acid

[1] In the native DNA uracil can be formed in the result of cytosine desamination; although this mutation is rapidly repaired by the reparation system, a uracil residue remains in DNA for some finite time.

bases decreases as: Gua > Cyt > Ade > Ura. This stability trend indicates that, based on the thermodynamics, the amide group can distinguish between the bases in the single stranded DNA.

The interaction energies of the amino acid Gln with the purine bases are similar to those found for the base-acrylamide complexes, which justifies the use of acrylamide as a model of the amino-acids side chain.

On the basis of comparing the structures of the model complexes described in this work with the literature data on the recognition complexes of the nucleic acids and proteins biopolymers, the site preferences of the glutamine and asparagine side chains in the interactions with nucleic acid bases can be outline. All bases can interact with amide group of the amino acids *via* their Watson-Crick sites when being incorporated into a single strand DNA or RNA.

The purine bases can be recognized by the amide group in the double- and single-strand DNA, while pyrimidine bases cannot be recognized in the double stranded DNA. m^9Ade is the only base which can be recognized in the major groove of the double-strand DNA. These results are important, because the interaction with the amide group in the major groove lowers the stability of the adenine-uracil base pair [52,53]. The keto tautomer of m^9Gua, which is the dominant form of guanine in DNA, is the only structure that can be recognized in the minor groove of the double-strand DNA. The above information may be useful in considerations of more complex models and eventually of the real biological systems.

This work was partially supported by INTAS (Grant No. 99-00478).

References

1. (a) W. Saenger, U. Heinemann, *Protein-Nucleic Acid Interactions* (Macmillan Press, Bashingstoke, 1989); (b) W. Saenger, G. Jeffrey, *Hydrogen Bonding in Biological Structures* (Springer-Verlag, Berlin, 1991)
2. T.A. Steitz, *Structural Studies of Protein-Nucleic Acid Interactions: the Sources of Sequence-specific binding* (Cambridge University Press, 1993)
3. B.I. Verkin, I.K. Yanson, L.F. Sukhodub, A.B. Teplitsky, *Interactions of Biomolecules: New Experimental Approaches and Methods* (Naukova Dumka, Kiev, 1985) [*in Russian*]
4. M. Dey, J. Grotemeyer, E.W. Schlag, Z. Naturforsch. A **49**, 776 (1994)
5. C. Desfrançois, H. Abdul-Carime, C.P. Schulz, J.P. Schermann, Science **269**, 1707 (1995)
6. L.F. Sukhodub, Chem. Rev. **87**, 598 (1987)
7. S.Y. Wodak, M.Y. Lin, H.W. Wykoff, J. Molec. Biol. **116**, 855 (1977)
8. R.S. Jordan, C.O. Pabo, Science **242**, 893 (1988)
9. A.T. O'Neil, R.N. Hoess, W.F. DeGrado, Science **249**, 774 (1990)
10. L. Fairall, J.W.R. Schwabe, L. Chapman, J.N. Finch, D. Rhodes, Nature **366**, 483 (1993)
11. K. Ogata, S. Morikawa, H. Nakamura, A. Sekikawa, T. Inoue, H. Kanai, A. Sarai, S. Ishii, Y. Nishimura, Cell **79**, 639 (1994)
12. J.D. Klemm, M.A. Rould, R. Aurora, W. Herr, C. Pabo, Cell **77**, 21 (1994)
13. H.B. Houbaviy, A. Usheva, T. Shenk, S.K. Burley, Proc. Natl. Acad. Sci. USA **93**, 13577 (1996)
14. M.A. Rould, J.J. Perona, D. Söll, T.A. Steitz, Science **246**, 1135 (1989)
15. Y. Kim, J.H. Geiger, S. Hahn, P.B. Sigler, Nature **365**, 512 (1993)
16. J.L. Kim, D.B. Novikov, S.K. Burley, Nature **365**, 520 (1993)
17. R. Arni, U. Heinemann, R. Tokuoka, W. Saenger, J. Biol. Chem. **263**, 15358 (1988)
18. A.K. Aggarwal, D.W. Rodgers, M. Drottar, M. Ptashne, S.C. Harrison, Science **242**, 99 (1988)
19. C. Helene, G. Lancelot, Progr. Biophys. Molec. Biol. **39**, 1 (1982)
20. J.E. Anderson, M. Ptashne, S.C. Harrison, Nature **326**, 846 (1987)
21. R.S. Hedge, S.R. Grossman, L.A. Lainins, P.B. Sigler, Nature **359**, 505 (1992)
22. F. Pichierri, M. Aida, M.M. Gromiha, A. Sarai, J. Am. Chem. Soc. **121**, 6152 (1999)
23. H. Kono, A. Sarai, Proteins **35**, 114 (1999)
24. C. Desfrançois, S. Carles, J.P. Scherman, Chem. Rev. **100**, 3943 (2000)
25. B.I. Verkin, I.K. Yanson, A.B. Teplitsky, G.D. Strizhko, Doklady Acad. Sci. USSS (Russ.) **255**, 219 (1980)
26. G. Suizdak, *Mass Spectrometry for Biotechnology* (Academic Press, San Diego, 1996)
27. E. DeHoffmann, J. Charette, V. Stroobant, *Mass Spectrometry. Principles and Applications* (John Wiley and Sons, Chichester, 1996)
28. P. Kebarle, Int. J. Mass Spectrom. **1–3**, 313 (2000)
29. M. Peshke, A.T. Blades, P. Kebarle, J. Am. Chem. Soc. **122**, 1492 (2000)
30. H.D. Beckey, *Principles of Field Ionization and Field Desorption Mass Spectrometry* (Pergamon Press, Oxford, 1977)
31. L.F. Sukhodub, V.S. Shelkovsky, K.L. Wierzchowski, Biophys. Chem. **19**, 191 (1984)
32. V.S. Shelkovsky, Ph.D. thesis, Institute for Low Temperature Physics and Engineering, Kharkov, 1985
33. I.K. Galetich, Ph.D. thesis, Kharkov State University, 1999
34. E.N. Korol, V.V. Lobanov, V.A. Nazarenko, V.A. Pokrovsky, *Physical Principles of Field Mass Spectrometry* (Naukova Dumka, Kiev, 1978) [*in Russian*]
35. A.B. Teplitsky, I.K. Galetich, L.F. Sukhodub, Mol. Biol. (Russ.) **25**, 709 (1990)
36. I.K. Yanson, A.B. Teplitsky, L.F. Sukhodub, Biopolymers **18**, 1149 (1979)
37. G.G. Sheina, S.G. Stepanian, E.G. Radchenko, Yu.P. Blagoi, J. Mol. Struct. **158**, 275 (1987)
38. K. Szczepaniak, M. Szczesniak, J. Mol. Struct. **156**, 29 (1987)
39. K. Szczepaniak, M. Szczesniak, W.B. Person, Chem. Phys. Lett. **153**, 39 (1988)
40. J.S. Binkley, J.A. Pople, Int. J. Quant. Chem. **9**, 229 (1975)
41. J.A. Pople, J.S. Binkley, R. Seeger, Int. J. Quant. Chem., Quant. Chem. Symp. **10**, 1 (1976)
42. A.D. Becke, Phys. Rev. B **38**, 3098 (1988)

43. C. Lee, W. Yang, R.G. Parr, Phys. Rev. B **37**, 785 (1988)
44. S.H. Vosko, L. Wilk, M. Nusair, Can. J. Phys. **58**, 1200 (1980)
45. S.F. Boys, F. Bernardi, Molec. Phys. **19**, 533 (1970)
46. M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A. Montgomery Jr, R.E. Stratmann, J.C. Burant, S. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, A.G. Baboul, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, J.L. Andres, C. Gonzalez, M. Head-Gordon, E.S. Replogle, J.A. Pople, *Gaussian 98, Revision A.7*, Gaussian, Inc., Pittsburgh PA, 1998
47. I. Galetich, M. Kosevich, V. Shelkovksy, S.G. Stepanian, Yu.P. Blagoy, L. Adamowicz, J. Mol. Struct. **478**, 155 (1999)
48. I. Galetich, S.G. Stepanian, V. Shelkovsky, M. Kosevich, Yu.P. Blagoy, L. Adamowicz, J. Phys. Chem. B (Biophys. Chem.) **103**, 11211 (1999)
49. I. Galetich, S.G. Stepanian, V. Shelkovsky, M. Kosevich, Yu.P. Blagoi, L. Adamowicz, J. Phys. Chem. A. **104**, 8965 (2000)
50. I.K. Galetich, S.G. Stepanian, V.S. Shelkovsky, M.V. Kosevich, L. Adamowicz, J. Phys. Chem. (2002, accepted)
51. L.F. Sukhodub, I.K. Yanson, Nature **264**, 245 (1976)
52. M. Molina, P. Carmona, J. Mol. Struct. **175**, 283 (1988)
53. A. Sarai, M. Saito, Int. J. Quant. Chem. **25**, 527 (1984)