Electron Correlated Ab Initio Study of Amino Group Flexibility for Improvement of Molecular Mechanics Simulations on Nucleic Acid Conformations and Interactions

V. I. Poltev · E. Gonzalez · A. Deriabina · A. Martinez ·
A. Furmanchuk · L. Gorb · J. Leszczynski

Received: 30 October 2007 / Accepted: 26 May 2008 /
Published online: 18 July 2008
© Springer Science + Business Media B.V. 2008

Abstract High level ab initio studies demonstrate substantial conformational flexibility of amino groups of nucleic acid bases. This flexibility is important for biological functions of DNA. Existing force field models of molecular mechanics do not describe this phenomenon due to a lack of quantitative experimental data necessary for an adjustment of empirical parameters. We have performed extensive calculations of nucleic acid bases at the MP2/6-31G(d,p) level of ab initio theory for broad set of amino group configurations. Two-dimensional maps of energy and geometrical characteristics as functions of two amino hydrogen torsions have been constructed. We approximate the maps by polynomial expressions, which can be used in molecular mechanics calculations. Detailed considerations of these maps enable us to propose a method for determination of numerical coefficients in the developed formulae using restricted sets of points obtained via higher-level calculations.

Keywords Molecular mechanics · Correlated ab initio calculations · Nucleic acids · Amino group flexibility

1 Introduction

The genetic functions of nucleic acids are governed by their fine structures. Local variations of double helix geometry with nucleotide sequence are important for expression of genetic information. Special features of certain sequences ensure the specificity of their interactions with proteins regulating transcription. Four DNA bases (adenine, Ade; thymine, Thy; guanine, Gua; and cytosine, Cyt) provide an alphabet for description of all the genetic

V. I. Poltev (✉) · E. Gonzalez · A. Deriabina · A. Martinez
Physics and Mathematics Department, Puebla Autonomous University, Puebla, 72570 Pue, Mexico
e-mail: poltev@cfm.buap.mx

A. Furmanchuk · L. Gorb · J. Leszczynski
Computational Center for Molecular Structure and Interactions,
Jackson State University, Jackson, MS 39217, USA

information necessary for the functioning of living species. A specificity of interactions between complementary bases is a key for correct copying of the information, and interactions between the pairs contribute to the stability of the double helix and to the dependence of its fine structure on nucleotide sequence. Three of the four DNA bases, namely, Ade, Gua, and Cyt, contain amino groups. The structures of these bases (including atoms' numbering) are displayed in Fig. 1.

The amino groups participate in base pairing via one of their hydrogen atoms, and their second hydrogen atom can interact with proteins or other biologically important substances. Such interactions contribute to propeller twisting and buckling of the base pairs, which can contribute to the establishment of specific profiles for different nucleotide sequences. Thus,

Fig. 1 Atom numbering in nucleic acid bases containing amino groups: Gua (*top*), Ade (*middle*), and Cyt (*bottom*). Torsion angles definitions are as follows:
Gua $H_{22}, N_3-C_2-N_2-H_{22}$;
Gua $H_{21}, N_3-C_2-N_2-H_{21}$;
Ade $H_{61}, N_1-C_6-N_6-H_{21}$;
Ade $H_{62}, N_1-C_6-N_6-H_{62}$;
Cyt $H_{41}, N_3-C_4-N_4-H_{41}$;
Cyt $H_{42}, N_3-C_4-N_4-H_{42}$

a correct description of amino groups and their interactions is important for understanding and predictions of nucleic acid structure and functions.

Impressive progress in studying the fine structure of nucleic acid fragments has been achieved during the last two decades with use of such experimental methods as X-ray diffraction (e.g., [1–3] and references therein) and nuclear magnetic resonance (NMR) spectroscopy (for recent references, see, e.g., [4–6]). Systematic X-ray diffraction studies of DNA fragments enables one to reveal regularities in base arrangement for various nucleotide sequences of the DNA duplex, to consider unusual DNA duplex and hairpin motifs (loops, junctions, etc.), and to uncover structures of chemically modified DNA. NMR study of specific nucleotide sequences reveals parallel-stranded DNA structures [4]. Recent investigations, using NMR spectroscopy (along with other physical methods) of DNA duplexes with chemically modified bases [5, 6], contribute to understanding the physical nature of duplex stability. Unfortunately, the accuracy of these methods does not provide atomic level structures of nucleic acids oligomers. Therefore, an application of theoretical methods is inevitable both together with experimental data and also to predict rules which govern the three-dimensional structure of nucleic acids and their complexes.

2 Quantum Mechanics and Molecular Mechanics Methods in Studies of Nucleic Acids

Two groups of theoretical methods can be applied to a study of the structure and properties of biomolecular systems, namely, quantum mechanics (QM) and molecular mechanics (MM). Quantum mechanical consideration implies approximate solution of the Schrödinger equation for a molecule or supramolecular complex. The complexity of the system determines the extent of the approximation. Energy and structure characteristics of rather simple systems, e.g., nucleic acid bases or their pairs, can be predicted by electron correlated ab initio computations with an accuracy of the most precise experiments.

The molecular mechanics approach implies a consideration of the molecular complex as a system of particles interacting according to the laws of classical physics. These interactions are described by simple algebraic formulae; thus, the method enables one to simulate rather complex systems, e.g., nucleic acid components or their complexes with protein fragments. The energy of the system E is usually expressed via a sum of terms responsible for different types of interactions:

$$E = E_{aa} + E_b + E_a + E_t + E_{ad}, \quad (1)$$

where E_{aa} is the energy of all pair-wise interactions of atoms that are not bonded to each other or to the same third atom; it includes contributions from electrostatic, van der Waals attraction and repulsion terms. The second term, E_b is the sum of energies resulting from deviation of bond lengths from their equilibrium values; E_a is the sum of energies of bond angle distortions; E_t corresponds to torsion energies, which depend on the angles of rotation about bonds; and E_{ad} is an additional term to describe different effects, including special quantum effects, and adjustment for limitations (if they exist), which cannot be taken into account by the other terms.

Several force fields are widely used in MM studies of biopolymers. The most popular are those designed by Kollman's [7] and Karplus's [8] groups and used in AMBER [9] and CHARMM [10] computer programs. Another type of simplified force field, especially adjusted for nucleic acids, was elaborated by one of us ([11, 12] and references therein).

Such simple energy descriptions of rather complicated molecular systems, like any other classical model of a quantum system, cannot be sufficiently rigorous. The molecular mechanics method is essentially semiempirical one, and its force field depends on experimental data used for parameter adjustment and theoretical assumptions. In order to construct more detailed molecular models of nucleic acids and their complexes, it is necessary to refine periodically the method and the force fields using new data both experimental and theoretical.

Thus, both discussed approaches have their own advantages and limitations. An elaboration and refinement of the MM method requires extensive calculations to adjust the force field to a whole set of experimental data. In many cases, there is no sufficient set of reliable and exact experimental results to perform such an adjustment. The results of high level QM methods provide another source for a possible elaboration of the MM force field, namely, the adjustment of force field parameters to reproduce the values obtained by the QM calculations. Calculations for a sufficient number of the system states (mutual positions of the system fragments) provide numerical values which can be used for adjustment of the coefficients in MM energy terms. The first applications of QM computations for deriving the coefficients of MM calculations of nucleic acids took place in the late 1970s. The pioneering works of E. Clementi should be mentioned in this respect (e.g., [13]). The MMFF94 force field [14] can be considered as an example of QM-derived method for MM reproducing well both the computational and experimental data for a variety of molecular systems. Some of the most recent publications on modification of existing force fields to QM results can be mentioned. A need for such modifications becomes evident when well-known and extensively used force fields fail to reproduce new experimental data. In this case, the adjustment of the coefficients to QM results provides a useful tool for the investigations of specific systems. In [15], the AMBER force field was adjusted to reproduce conformational properties of an oxidative DNA lesion, 2,6-diamino-4-hydroxy-5-formamidopyrimidine. The other QM modifications of this force field (e.g., [16]) include parameters of atom types that did not exist in initial publications.

The molecular mechanics computations using modern force fields and correlated *ab initio* quantum mechanics computations provide (when a comparison is possible) nearly equivalent results for simple nucleic acid fragments. Nevertheless, some results of quantum mechanics computations, namely, a phenomenon of nonplanarity and conformational flexibility of base amino groups ([17–20] and references herein) cannot be reproduced by existing MM force fields.

3 The Amino Group Flexibility as Revealed by Quantum Mechanics and a Need for Its Simplified Description

Correlated *ab initio* calculations suggest partial sp^3 hybridization of amino group of isolated nucleic acid bases. Formation of hydrogen-bonded base pairs in crystals of bases, nucleosides, and nucleotides results in nearly planar geometry of the bases including the atoms of amino groups. This planar geometry is used as a starting point for developing force field models. However, there is experimental evidence, including the results of X-ray studies [21, 22], of the conformational flexibility of amino groups. The paper [21] demonstrates the existence of a group of bifurcated H bonds in the region of six dA–dT pairs in a dodecamer crystal. The same hydrogen of the adenine amino group forms H bonds with O₄ atoms of the thymines of both complementary and neighboring pairs. The cross-

strand hydrogen bonds in the duplex major groove were revealed in the G–A mismatch pairs of d(CCAAGATTGG) crystals [22]. Such hydrogen bonds are impossible without out-of-plane displacements of amino group atoms, and proper MM description of such conditions is important for modeling of some nucleotide sequences. Unfortunately, these results cannot provide a basis for the MM description of the nonplanarity phenomenon, as they are rather qualitative.

In view of importance of conformational flexibility of base amino groups and the need for its proper MM description, we started systematic correlated ab initio calculations for all DNA bases with various configurations of amino groups. We suggested that energy dependence of the base energy on amino group configuration should be described by additional terms included in the expression describing the MM total energy. These terms could be obtained by approximation of ab initio results.

At the first step, the relaxed scans of MP2/6-31G(d,p) interaction energy and the geometrical parameters of the DNA bases were performed as functions of three amino group torsions, namely, $N_3-C_4-N_4-H_{41}$ (H_{41}), $N_3-C_4-N_4-H_{42}$ (H_{42}), and $C_2-N_3-C_4-N_4$ (N_4). These preliminary calculations demonstrate a complex character of the dependences of base energy on amino group configuration. We revealed a nonadditivity of contributions of the changes of three amino group torsions into the base energy. It forced us to perform extensive calculations for two-dimensional maps of the energy and geometry parameters as functions of two torsions of the amino group hydrogen atoms, i.e., nearly a thousand energy minimizations for each base at each level of theory.

When this work was in progress, a paper of Ryjáček et al. [23] was published on improvement of the AMBER force field using the resolution-of-identity second-order Møller–Plesset perturbation/split-valence plus polarization level of theory. The authors optimized bond angle and torsion angle constants of the force field using the ab initio results. The changes of parameters were made without modifications of analytical expressions for terms of energy and without the addition of new terms. In view of our conclusions on the nonadditivity of three amino group torsion contributions to the base energy, it is not surprising that rather imperfect reproduction of ab initio barrier (e.g., 0.5 kcal/mol instead of 1.05 kcal/mol for Gua) and torsion (9° instead of 17° for Ade H_{61} ($N_1-C_6-N_6-H_{61}$) or 32° instead of 39° for Gua H_{22} , $N_3-C_2-N_2-H_{22}$) values were obtained in [23] by use of such a reparametrized AMBER force field. We believe that this modification can be considered as an approximate description of amino group flexibilities. Another method, suggested by us, for modification of any force field by approximating energy dependence on amino group torsions by analytical expressions and application of these formulae as additional terms in molecular mechanics computations appears more promising.

Our preliminary results for the Cyt amino group flexibility as revealed by MP2/6-31G(d,p) studies approximated by polynomial expression were published earlier [24]. The results of calculations demonstrate, in accord with previous publications, that two symmetrically arranged energy minima exist. They could be described by a displacement of amino group hydrogens out of the ring plane and by a smaller displacement of the amino nitrogen in the opposite direction. Two main conclusions were obtained from these results, which were confirmed by subsequent considerations of purine bases and the use of more extended basis sets. The first conclusion refers to a rather wide range of amino group configurations within rather narrow variations of the base internal energy (not greater than contributions of ambient temperature fluctuations for Cyt [24] and for Ade). This range includes not only the region between two minima but it extends up to dozens of degrees of consistent changes of amino group torsions. Rather, small energy

changes take place within a variation of two amino group hydrogen torsions along the line connecting the two energy minima in a region between them. In other words, the surface of potential energy is extremely flat in this region. The variations of amino group torsions along this line in opposite directions cause a much faster increase of the energy. Less pronounced changes of the energy take place with variations of torsions in directions perpendicular to this line. Thus, one can conclude that not only nonplanar minimum energy configurations and planarization barriers are important for nucleic acid structures and functions but also, rather, extended sets of amino group configurations should be taken into consideration.

The recent paper of Wang and Schaefer [25] demonstrates that the applications of more extended basis sets (up to MP2/aug-cc-pVQZ level of theory) resulted in pronounced decrease of amino group inversion barriers and less pronounced decrease of amino group nonplanarity for minimal energy structures. Nevertheless, amino group flexibility remains important for nucleic acid interactions at any level of theory.

The second important conclusion of our previous study [24], already mentioned above, refers to nonadditive contributions of amino group torsions to the internal DNA base energy. It means that the contribution of one of the torsions depends not only on its value but also on the values of the other torsions. Thus, the description of amino group flexibility in the framework of molecular mechanics force fields requires nonadditivity of some contributions to the total energy of a molecular system.

4 Amino Group Flexibility and Introduction of Nonadditivity to Molecular Mechanics

The principal assumption of molecular the mechanics method is the additivity of energy terms describing interactions of one type (e.g., the change of the electrostatic Coulomb interaction of two molecules or two fragments is the sum of hundreds of changes of atomic charge interactions) and of the terms describing different interactions (e.g., due to bond length, bond angle, and torsion variations). Cross terms used in some MM force fields usually contribute to some extent, and they can be ignored in most of the calculations carried out for biologically important molecules. However, this is not the case for amino group torsion angle contributions.

In the previous study [24], we approximated Cyt internal energy by the polynomial function of the two amino group hydrogen torsions. The fitting of the coefficients was performed using a standard mathematical procedure and computer program. The polynomial function obtained contains some terms depending on each of torsion angles and other terms depending on both the angles. The introduction of nonadditive terms in such a way does not significantly increase the complexity of molecular mechanics force fields or the computer resources required. The function with nine coefficients enables us to reproduce *ab initio* map with rather good accuracy (an agreement of energy values being better than 99.9%). We obtained, on a basis of MP2/6-31G(d,p) computations and two-dimensional map construction, such accurate functions also for Ade and Gua. We do not present these expressions here due to certain restrictions of their validity. These functions are valid for the selected level of *ab initio* theory (MP2/6-31G(d,p)) and cannot be improved using more rigorous data. Thus, being applied to the molecular mechanics method, the functions would restrict their validity to this level. Here, we describe a new, improved method of analytical approximation of two-dimensional maps which needs only limited *ab initio* data and can be applied to any level of the theory.

5 Analysis of MP2/6-31G(d,p) Maps of Potential Energy and Deriving their Analytical Expressions

A potential energy map for Gua constructed from MP2/6-31G(d,p) ab initio results is displayed in Fig. 2a. We consider guanine here in more detail as this base has the most pronounced nonplanarity and the highest energy barriers calculated for any level of ab initio theory, as compared to the two other bases. The displayed map demonstrates that two sets of amino group torsions H_{21} and H_{22} correspond to the most favorable configuration of the base. This map and the corresponding maps for two other bases considered possess a center of symmetry corresponding to the position of two amino hydrogens in the plane of the base ring (i.e., to an energy barrier between the two most favorable structures). We label H_{22} and H_{21} guanine torsions (as well as H_{61} and H_{62} of Ade and H_{41} and H_{42} of Cyt) via x_2 and y_2 . The values of energy minima are designated via $\pm x_{m2}$ and $\pi \pm y_{m2}$.

After an extensive analysis of the maps, we found that scaling one of the axes and shifting the center of symmetry to the coordinate center leads to a more symmetrical map (Fig. 2b), with the axis of symmetry passing through two energy minima and the center of symmetry. This is done via the transformation.

$$x_1 = gx_2 \quad (2)$$

$$y_1 = y_2 - \pi, \quad (3)$$

where the scaling factor g is an adjustable parameter, which is different for different bases. The location of minima in these coordinates is given by:

$$x_{m1} = gx_{m2}$$

$$y_{m1} = y_{m2}$$

Fig. 2 Two-dimensional MP2/6-31G(d,p) energy map for guanine displayed in various coordinate systems. Numbers at isoenergy contour lines present values (in kcal/mol) above energy minima. Angles are shown in units of degrees

The rotation of the coordinate system around the z -axis by angle α results in the map displayed in Fig. 2c via the transformation.

$$x = x_1 \cdot \cos \alpha - y_1 \cdot \sin \alpha \quad (4)$$

$$y = x_1 \cdot \sin \alpha + y_1 \cdot \cos \alpha \quad (5)$$

where $\alpha = \arctg(x_{m1}/y_{m1})$.

As a result of these transformations, we converted the initial map into a map with axes x and y as two axes of symmetry. Cross-sections of the map along x and y axes are displayed in Fig. 3. The function approximating this map can be represented as a sum of the functions describing the two cross-sections separately. The final expression represents a polynomial containing even degrees of x and y with adjustable coefficients. The simplest polynomial describing ab initio energy maps is given by:

$$E = ay^6 + by^4 - cx^2y^2 + dx^2 - ey^2 + f \quad (6)$$

where f is the energy barrier, the coefficients a , b , and e can be derived from the evolving curve of energy on y (Fig. 3 left), while c and d can be derived from the two energy dependences on x (solid lines on Fig. 3 right). The adjustment of coefficients has been performed manually by comparison of the MP2 curves displayed in Fig. 3 with analytical approximations.

Fig. 3 Cross-sections of energy map for Gua (Fig. 2c). Solid continuous curves are used for coefficient adjustment. Units of energy, kcal/mol; units of angles, radians

Table 1 Numerical values of coefficients for description of MP2/6-31G(d,p) energy maps

Base	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	x_m	y_m
Ade	3.02	6.78	5.5	7.8	2.87	0.28	0.983	0.32	2.84
Gua	0.75	2.91	0.2	1.7	4.48	1.48	1.85	0.21	2.46
Cyt	1.93	3.22	2.0	5.5	2.20	0.32	1.33	0.24	2.73

Units of energy are kcal/mol; units of angles are radians.

Numerical coefficients for approximation of the MP2/6-31G(d,p) energy for the three bases are presented in Table 1. The comparison of calculated via the map (2)–(6) with ab initio results for Gua is presented in Fig. 4. Figure 4 demonstrates that the two maps are nearly identical. The central region of the map, which is the most important region for the base interactions, is reproduced with very high accuracy.

After analysis of the ab initio maps and those obtained via analytical expressions described above, we concluded that the numerical coefficients can be extracted from a restricted set of points in the maps (placed along the selected lines on Fig. 3). Not all of the optimized points (thousands calculated for each base) are necessary for determination of such coefficients.

This conclusion has led us to much more efficient option of construction of the maps using only a few dozen points which are calculated at higher level of theory (obtained using more extended basis sets). Such an option results in inclusion of more rigorous levels of the ab initio theory into modified molecular mechanics force fields.

Fig. 4 The comparison of approximated (*continuous lines*) and ab initio MP2/6-31G(d,p) (*dots*) maps for guanine. Numerical values are the same as in Fig. 2a

6 Analytical Construction of Two-Dimensional Maps Using Restricted Results of MP2/aug-cc-pVDZ Calculations

Preliminary calculations using MP2/aug-cc-pVDZ, MP2/aug-cc-pVTZ, and other selected levels of *ab initio* theory performed for selected points on two-dimensional maps demonstrated a substantial dependence of the results on the basis set. The utilization of extended basis sets results in a considerable change of energy barrier, a moderate displacement of the positions of energy minima, and a deviation of the energy values in various points of the maps from those for the 6-31G(d,p) basis set. These deviations depend on the position of the point, and direct use of expressions derived from MP2/6-31G(d,p) map is unfeasible. Thus, it is important to have the map for higher level of correlated *ab initio* computations. On the other hand, the use of extended basis sets for the calculations of optimized structures of nucleic acid bases is much more time consuming, and the direct construction of two-dimensional maps at the MP2/aug-cc-pVDZ (or more exact) level of theory is practically prohibited.

Our considerations, and experience in the adjustment of analytical expressions to the MP2/6-31G(d,p) maps, enable us to construct two-dimensional maps for the three nucleic acid bases using restricted sets of optimized points at the MP2/aug-cc-pVDZ level of theory. These sets include a few points on the x and y axes of the rotated maps (see Fig. 2c) and a few points in a line passing through the energy minimum parallel to the x axis (see evolving curves in Fig. 3). The manual adjustment of the dependences on x and y to *ab initio* values results in the coefficients presented in the Table 2. Some additional (control) points have been calculated for each of the bases, and comparison between MP2/aug-cc-pVDZ values and those calculated using analytical expressions has been performed. It appears that the mean differences for the control points are nearly the same as for the points used for the coefficient adjustments. The mean deviation of calculated energies from MP2/aug-cc-pVDZ values for 36 guanine points is 0.036 kcal/mol; most of the considered points have a difference less than 0.1 kcal/mol; somewhat greater differences appear for a few points with energy values greater than 3 kcal/mol. Nearly the same level of accuracy has been obtained for adenine and cytosine. It is interesting that rather a greater difference was obtained in comparison between calculated and MP2/6-31G(d,p) values for the same set of the torsion angles.

The comparison of the approximated maps for MP2/6-31G(d,p) and MP2/aug-cc-pVDZ levels of theory is displayed in Fig. 5. It is worth noting that the maps differ substantially in their center region, most important for the description of base interactions (between two minima and a region around position corresponding to the barrier), while the regions with large energy are nearly equally prohibited for both levels of *ab initio* theory.

Table 2 Numerical values of coefficients for description of MP2/aug-cc-pVDZ energy maps

Base	a	b	c	d	e	f	g	x_m	y_m
Ade	4.32	4.44	1.43	6.5	7.5	0.102	0.983	0.27	2.89
Gua	1.165	1.42	0.7	1.85	3.07	1.08	1.85	0.22	2.47
Cyt	2.13	2.22	3.0	5.58	1.0	0.097	1.33	0.21	2.82

Units of energy are kcal/mol; units of angles are radians.

Fig. 5 The approximated MP2/6-31G(d,p) (*continuous lines*) and MP2/aug-cc-pVDZ (*dots*) maps for Gua. Other designations are the same as in Fig. 2a

7 The Description of Changes of Base Geometrical Parameters Resulting from Variations of Amino Group Torsions

The construction of two-dimensional maps for geometrical parameters of the bases demonstrates some regularity in changes of bond lengths, bond angles, and torsion angles with variations of the two torsions of amino group hydrogens. The geometrical parameters of the base rings, of extra-ring carbonyl groups, and of C–H groups vary slightly. For the MP2/6-31G(d,p) level of theory, all the changes of bond lengths and bond angles fall in 0.003 Å and 0.5° limits, respectively, for the energy variations within 2 kcal/mol. The deviations of these parameters from those for equilibrium structure become less pronounced with basis set extension. This observation of the bond length dependence agrees with the conclusions of the paper [25] (up to MP2/aug-cc-pVQZ level of theory) that these bond lengths become practically the same for the planar and equilibrium structures. The variations of these parameters are smaller than the differences between ab initio values and corresponding values generally accepted in the force field approximation as the equilibrium ones. Thus, such variations can be estimated by the standard MM procedures or can be ignored. The variations of extra-ring N–C–N and C–C–N bond angles are more pronounced, but generally, they do not exceed 1–2°.

On the other hand, several other variables considered for each amino group, namely, C–N bond length and C–N–H and H–N–H bond angles, depend substantially on amino group torsions. It is difficult to expect that the angle variations of 10° (and more) with characteristic ellipsoidal dependence on the two torsion angles would be properly described by the standard MM formulae. The same conclusion refers to C–NH₂ bond length variations, which can be as great as 0.04 Å with energy changes of 2 kcal/mol. We suggest evaluating “equilibrium” values of these parameters from two-dimensional maps and then

Table 3 Numerical values of the constants in (7) for C–NH₂ bond length and the three bond angles of base amino groups

Base	Bond length or bond angle	a_x	a_y	z_0
Gua	C ₂ –N ₂	0.005	0.033	1.363
	C ₂ –N ₂ –H ₂₁	–0.013	–0.22	2.146
	C ₂ –N ₂ –H ₂₂	0.015	–0.22	2.049
	H ₂₁ –N ₂ –H ₂₂	–0.003	–0.28	2.091
Ade	C ₆ –N ₆	0.014	0.058	1.353
	C ₆ –N ₆ –H ₆₁	0.009	–0.45	2.096
	C ₆ –N ₆ –H ₆₂	0.03	–0.45	2.079
	H ₆₁ –C ₆ –H ₆₂	–0.023	–0.45	2.112
Cyt	C ₄ –N ₄	0.043	0.011	1.358
	C ₄ –N ₄ –H ₄₁	0.03	–0.28	2.056
	C ₄ –N ₄ –H ₄₂	–0.01	–0.24	2.135
	H ₄₁ –N ₄ –H ₄₂	–0.01	–0.29	2.095

Units of bond length, Å; units of angles, radians

calculating energy changes related to distortion of these parameters by standard expressions of MM. We have derived such analytical expression for the values of the parameters:

$$B = a_x x^2 + a_y y^2 + z_0, \quad (7)$$

where B is bond angle or bond length for torsion angles x and y defined in (2–5).

The numerical values of the constants obtained from MP2/6-31G(d,p) two-dimensional maps are listed in Table 3. Rather, small corrections arise from use of MP2/aug-cc-pVDZ data; thus, the same values could be used for the approximation of higher levels of ab initio results. It is interesting to note that changes of C–N bond length for MP2/aug-cc-pVDZ level of theory as compared to the MP2/6-31G(d,p) level are nearly the same (0.007 ± 0.001) Å for all the points on the two-dimensional maps considered. One more parameter of amino group configurations, the torsion angle responsible for nitrogen displacement from the ring plane (N₂, N₆, or N₄), varies in the range of a few degrees for any point of the two-dimensional maps. Its variations can also be described by a rather simple expression.

8 Further Improvements and Simplifications of the Description of Amino Group Flexibility and Incorporation of this Description into the Force Fields of Molecular Mechanics

The algebraic expressions for descriptions of conformational flexibility of amino groups of nucleic acid bases proposed above can be directly incorporated into the MM software without substantial increase of computational resources necessary for biopolymer simulations. The following discussion details possible developments and their consequences.

The most obvious improvement results from the applications of higher levels of ab initio theory for adjustment of analytical expressions. The use of MP2/aug-cc-pVTZ and MP2/aug-cc-pVQZ methods for deriving enhanced coefficients, instead of those given in Table 2, would result in a more accurate description of the phenomenon. After careful choice of the points on two-dimensional maps, such optimization can be performed during a reasonable time when using high-speed computers. We plan to consider such an improvement later, but it is not obvious that an influence of such enhancement will be greater than the accuracy of molecular mechanics methods in general.

One more improvement needed for more accurate MM calculations with flexible amino groups of bases is the introduction of a special term responsible for planarity (or near planarity) of most of the nucleic acid base pairs. The nearly planar structure of these pairs is partially quantum in nature. As already mentioned [23], this phenomenon is connected with electron density withdrawal from the amino group nitrogen upon complexation. Molecular mechanics reproduces well ab initio results when using equilibrium structures with planar amino groups. An inclusion of amino group flexibility will result in overestimation of the nonplanarity of base pairs. The special term for accounting for this effect can be derived from ab initio study of the dependence of the base pair energies on propeller twist and buckle values.

A more important current task includes a simplification of ab initio derived terms for MM treatment of nucleic acid bases. For this purpose, we performed selected ab initio calculations for dependence of base energy and amino group geometry on amino hydrogen torsions, using a fixed geometry for the rest of the base. The calculations at the MP2/6-31G(d,p) level of theory (numerical values not shown) demonstrate minor changes of the dependences on ring geometry fixations. These changes would not be greater for extended basis sets as can be concluded from the previous section. Thus, the use of rigid base rings can be justified by the results of ab initio calculations.

An additional simplification can arise from more careful consideration of the two-dimensional maps and their cross-sections along the x and y axes. Such considerations could result in the derivation of interdependency parameters (Tables 1–3) and, consequently, in a decrease in the number of points of high-level ab initio calculations needed for the adjustment of MM expressions. Such studies are currently in progress.

Fig. 6 Base arrangements in two guanine–adenine base pairs, GA1 (a); GA3 (b)

9 Preliminary Estimations of the Contribution of Amino Group Flexibility to the Base–Base Interactions and to Base Pair Structure and Energy

To evaluate quantitatively the contribution of the flexibility of amino groups to nucleic acid structure, we have chosen four intrinsically nonplanar base pairs connected by two H bonds. These pairs, along with some other related DNA base complexes, were considered in [23] and referred to as GA1, GA3, GG3, and AA2. The two Gua–Ade pairs are displayed in Fig. 6. We chose these pairs since they can contribute to mismatch formations during DNA biosynthesis and to the secondary structure of RNA. The nonplanarity of Gua–Ade pairs is well-pronounced and is due to steric repulsion of hydrogen atoms in planar structures with two H bonds. The Gua–Gua and Ade–Ade pairs can adopt planar structures, but these structures are less favorable than nonplanar ones.

Energy and geometry characteristics for the four base pairs are presented in Table 4. The values of interaction energy and mutual base arrangements calculated by the most rigorous *ab initio* method feasible for such systems differ substantially from those obtained with AMBER [7, 9] force field [23]. Namely, the difference in interaction energies between those the obtained by these two approaches can be as much as 4 kcal/mol (about 25%) for GA1 pair, and the value of the angle between the base planes is twice as big at the MM level for the GA3 compared to *ab initio* results (see Table 4). Virtually, rigid and planar amino groups assumed by MM calculations prevent the bases from adopting the arrangements of these pairs revealed by the QM *ab initio* method.

The comparison of *ab initio* and AMBER results demonstrates that the AMBER force field substantially underestimates the interaction energy and greatly overestimates the

Table 4 Interaction energies, E , H-bond lengths, and angles between base-planes, α , of four nonplanar base pairs as calculated using the *ab initio* method [23], two MM force fields [7, 9, 11, 12], and using the force field of [11, 12] with amino group flexibility corrections

Base pairs and characteristics	<i>Ab initio</i> , adapted from [23]	MM, AMBER, adapted from [23]	MM [11, 12]	<i>This work</i>
GA1				
E	−16.5	−12.4	−14.2	−15.1
$N_1 \dots N_1$	2.87	2.94	2.89	2.88
$O_6 \dots N_6$	2.87	2.88	2.90	2.92
α	20	42	45	33
GA3				
E	−15.8	−13.0	−12.9	−15.0
$N_1 \dots N_1$	2.86	2.94	2.89	2.85
$O_6 \dots N_6$	2.90	2.88	2.88	2.91
α	22	58	56	35
GG3				
E	−19.0	−16.7	−18.5	−18.7
$N_1 \dots N_7$	2.84	2.91	2.89	2.85
$N_2 \dots O_6$	3.20	2.96	2.88	2.91
α	8	30	11	12
AA2				
E	−11.5	−10.1	−11.2	−11.4
$N_1 \dots N_6$	2.97	2.97	2.86	2.89
$N_6 \dots N_7$	2.98	2.99	2.88	2.90
α	28	52	37	32

Units of energy are kcal/mol, units of H-bond lengths are Å, and units of angles are degrees.

nonplanarity of these pairs (second and the third columns of Table 4). The computations with the method developed by one of us [11, 12] (fourth column) diminish some of these discrepancies; nevertheless, planar amino groups make relatively unfavorable positions of the bases that correspond to ab initio results. The results of computations using the MM force field [11, 12] and ab initio geometry as a starting point with corrections suggested here are presented in the fifth column. A considerable improvement corresponding to ab initio data can be seen. It is worth mentioning that in spite of the positive contribution of the additional energy term, the total energy becomes more negative for all the pairs considered. This is a result of more favorable mutual base arrangement due to the amino group flexibility. The contribution of the additional energy term as calculated via the transformations (2)–(6) expression is rather small (about 0.1 kcal/mol for GA1 and GA3 at the MP2/aug-cc-pVDZ level of theory) as a result of rather favorable amino group configurations. Nevertheless, the contribution of amino group flexibility to the energy and geometry of the base pairs is substantial (as Table 4 demonstrates), and such flexibility must be properly described.

Besides, as already mentioned here and in [23], an additional negative term of essentially quantum nature should be added to the total energy in order to avoid overestimation of nonplanarity and for better agreement with ab initio results. This term is a function of the angle between the interacting molecules that may depend on structures and pairing. Extensive QM and MM computations of base–base interactions will be performed to derive a rather simple expression for this term, but its contribution is not great, as can be seen from energy values of Table 4.

For one of the selected base pairs (GG3) with $N_1\text{--H}\cdots N_7$ and $N_2\text{--H}\cdots O_6\text{--H}$ bonds, the contribution of amino group nonplanarity is of minor importance. It appears, after MM calculations using AMBER [7, 9] and our [11, 12] force fields, that dependence of the interaction energy on the angle between the bases is rather flat. The energy difference between the ab initio level minimum and that from the MM calculations is less than 0.1 kcal/mol. Besides, the calculations with our force field [11, 12] without corrections for amino group flexibility demonstrate rather good agreement with ab initio [23] data (18.5 vs. 19.0 kcal/mol for the energy minimum and 10° vs. 8° for the interbase angle). Thus, the amino group flexibility plays a secondary role in the characteristics of separate GG3 pairs. The same is true for some other pairs, including Watson–Crick types, when the quantitative estimation of the contribution of amino group flexibility is not important for description of base arrangement in the region of base–base interaction around the energy minimum. Nevertheless, when interactions with other species (or other DNA fragments) force bases to shift substantially from the minimum corresponding to the pair configuration (large values of propeller and buckle parameters of a duplex region), the contributions of amino group flexibility become important factors that might affect such displacements. DNA duplexes with bifurcated H bonds (e.g., [21, 22]) provide examples of such cases.

Acknowledgements This work has been supported by the projects of Autonomous University of Puebla (29/EXC/06G, 29/EXC/07G), by NSF CREST (HRD-0318519), ONR (N00034-03-1-0116) grants, and Mississippi Center for Supercomputer Research (MCSR).

References

1. Hays, F.A., Teegarden, A., Jones, Z.J.R., Harms, M., Raup, D., Watson, J., Cavaliere, E., Ho, P.S.: How sequence defines structure: a crystallographic map of DNA structure and conformation. *Proc. Natl. Acad. Sci.* **102**, 7157 (2005)

2. Chou, S.-H., Chin, K.-H., Wang, A.H.-J.: Unusual DNA duplex and hairpin motifs. *Nucleic Acids Res.* **31**, 2461 (2003)
3. Pallan, P.S., Lubini, P., Bolli, M., Egli, M.: Backbone-base inclination as a fundamental determinant of nucleic acid self- and cross-pairing. *Nucleic Acids Res.* **35**, 6611 (2007)
4. Parvathy, V.R., Bhaumik, S.R., Chary, K.V.R., Govil, G., Liu, K., Howard, F.B., Miles, H.T.: NMR structure of a parallel-stranded DNA duplex at atomic resolution. *Nucleic Acids Res.* **30**, 1500 (2002)
5. Pfaff, D.A., Clarke, K.M., Parr, T.A., Cole, J.M., Geierstanger, B.H., Tahmassebi, D.C., Dwyer, T.J.: Solution structure of a DNA duplex containing a guanine-difluorotoluene pair: a wobble pair without hydrogen bonding? *J. Am. Chem. Soc.* **130**, 4869 (2008)
6. Matsugami, A., Ohyama, T., Inada, M., Inoue, N., Minakawa, N., Matsuda, A., Katahira, M.: Unexpected A-form formation of 4'-thioDNA in solution, revealed by NMR, and the implications as to the mechanism of nuclease resistance. *Nucleic Acids Res.* **36**, 1805 (2008)
7. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M. Jr., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., Kollman, P.A.: A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179 (1995)
8. MacKerell, A.D. Jr., Bashford, D., Bellott, R.L., Dunbrack, R.L. Jr., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, W.E., III, Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., Karplus, M.: All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586 (1998)
9. Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.S., Cheatham, T.E. III, DeBolt, S., Ferguson, D., Seibel, G., Kollman, P.: Amber, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **91**, 1 (1995)
10. Brooks, B.R., Brucoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M.: CHARMM. A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* **4**, 187 (1983)
11. Poltev, V.I., Shulyupina, N.V.: Simulation of interactions between nucleic acid bases by refined atomatom potential functions. *J. Biomol. Struct. Dyn.* **4**, 739 (1986)
12. Poltev, V.I., Deriabina, A.S., Gonzalez, E., Grokhlina, T.I.: Interactions between nucleic acid bases: new parameters of potential functions and new energy minima. *Biophysics* **47**, 972 (2002)
13. Scordamaglia, R., Cavallone, F., Clementi, E.: Analytical potentials from "ab initio" computations for the interaction between biomolecules. 2. Water with the four bases of DNA. *J. Am. Chem. Soc.* **99**, 5545 (1977)
14. Halgren, T.A.: Merck molecular force field. I. Basis, form, scope, parameterization and performance of MMFF94. *J. Comput. Chem.* **17**, 490 (1996)
15. Song, K., Hornak, V., De Los Santos, C., Grollman, A.P., Simmerling, C.: Molecular mechanics parameters for the FapydG DNA lesion. *J. Comput. Chem.* **29**, 17 (2008)
16. Oda, A., Yamaotsu, N., Hirono, S.: New AMBER force field parameters of heme iron for cytochrome P450s determined by quantum chemical calculations of simplified models. *J. Comput. Chem.* **26**, 818 (2005)
17. Leszczynski, J.: Are the amino groups in the nucleic acid bases coplanar with the molecular rings? Ab initio HF/6-31G* and MP2/6-31G* studies. *Int. J. Quantum Chem. Quantum Biol. Symp.* **19**, 43 (1992)
18. Spomer, J., Florian, J., Leszczynski, J., Hobza, P.: Nonplanar DNA base pairs. *J. Biomol. Struct. Dyn.* **13**, 827 (1996)
19. Hovorun, D.M., Gorb, L., Leszczynski, J.: From the nonplanarity of the amino group to the structural nonrigidity of the molecule: A post-hartree-fock ab initio study of 2-aminoimidazole. *Int. J. Quantum Chem.* **75**, 245 (1999)
20. Spomer, J., Leszczynski, J., Hobza, P.: Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases. *Biopolymers* **61**, 3 (2002)
21. Nelson, H.C.M., Finch, J.T., Luisi, B.F., Klug, A.: The structure of an oligo(dA) oligo(dT) tract and its biological implications. *Nature* **330**, 221 (1987)
22. Prive, G.G., Heinemann, U., Chandrasegaran, S., Kan, L.-S., Kopka, M.L., Dickerson, R.E.: Helix geometry, hydration, and G.A mismatch in a B-DNA decamer. *Science* **38**, 498 (1987)
23. Ryjáček, F., Kubar, T., Hobza, P.: New Parameterization of the Cornell et al. empirical force field covering amino group nonplanarity in nucleic acid bases. *J. Comput. Chem.* **24**, 1891 (2003)
24. Poltev, V.I., Gonzalez, E., Deriabina, A.S., Lozano, L., Martinez, A., Robinson, T., Gorb, L., Leszczynski, J.: *J. Mol. Struct. (Theochem)* **729**, 59 (2005)
25. Wang, S., Schaefer, H.F. III: The small planarization barriers for the amino group in the nucleic acid bases. *J. Chem. Phys.* **124**, 044303 (2006)