Theoretical investigation of the proton transfer mechanism in guanine-cytosine and adenine-thymine base pairs

ARTICLE *in* THE JOURNAL OF CHEMICAL PHYSICS · NOVEMBER 2012

Impact Factor: 2.95 · DOI: 10.1063/1.4766319 · Source: PubMed

5 AUTHORS, INCLUDING:

Shiyan Xiao
University of Science and Technology of China
16 PUBLICATIONS **57** CITATIONS
SEE PROFILE

Lei Wang
University of Science and Technology of China
10 PUBLICATIONS **71** CITATIONS
SEE PROFILE

Yuan Liu
University of Science and Technology of China
12 PUBLICATIONS **222** CITATIONS
SEE PROFILE

Haojun Liang
University of Science and Technology of China
152 PUBLICATIONS **2,216** CITATIONS
SEE PROFILE

Theoretical investigation of the proton transfer mechanism in guanine-cytosine and adenine-thymine base pairs

Shiyan Xiao, Lei Wang, Yuan Liu, Xiangsong Lin, and Haojun Liang

Citation: J. Chem. Phys. **137**, 195101 (2012); doi: 10.1063/1.4766319
View online: http://dx.doi.org/10.1063/1.4766319
View Table of Contents: http://jcp.aip.org/resource/1/JCPSA6/v137/i19
Published by the American Institute of Physics.

Theoretical investigation of the proton transfer mechanism in guanine-cytosine and adenine-thymine base pairs

Shiyan Xiao,[1] Lei Wang,[1] Yuan Liu,[1] Xiangsong Lin,[2] and Haojun Liang[3,a)]
[1]*CAS Key Laboratory of Soft Matter Chemistry, Department of Polymer Science and Engineering, University of Science and Technology of China, Hefei, Anhui 230026, People's Republic of China*
[2]*Hefei National Laboratory for Physical Sciences at Microscale, University of Science and Technology of China, Hefei, Anhui 230026, People's Republic of China*
[3]*CAS Key Laboratory of Soft Matter Chemistry and Hefei National Laboratory for Physical Sciences at Microscale, University of Science and Technology of China, Hefei, Anhui 230026, People's Republic of China*

(Received 21 August 2012; accepted 21 October 2012; published online 21 November 2012)

Ab initio constrained molecular dynamics and metadynamics were employed to investigate the mechanism of proton transfer in guanine-cytosine (GC) and adenine-thymine (AT) base pairs in the gas phase at room temperature. It is shown that double proton transfer (DPT) in the GC base pair is a concerted and asynchronous mechanism, and three pathways with a similar free energy barrier start from the canonical GC and end up in its "rare" imino-enol tautomer. The activation energy for the route that the DPT starts from the hydrogen atom movement in the O6(G)-N4(C) bridge is approximately 1.0 kcal/mol higher than that which starts in the N1(G)-N3(C) bridge. For the AT base pair, a stable intermediate state is identified in the two-dimensional free energy surface of the DPT event. We found that the movement of the hydrogen atom in the N1(A)-N3(T) bridge occurs before the movement of the hydrogen atom in the N6(A)-O4(T) bridge. Thus, it is demonstrated that the DPT in AT base pairs is a stepwise and an asynchronous mechanism. © *2012 American Institute of Physics*. [http://dx.doi.org/10.1063/1.4766319]

I. INTRODUCTION

Proton transfer in DNA base pairs is of great interest to theoretical and experimental scientists due to its vital role in genetic mutation,[1,2] charge transfer in DNA,[3–6] and ion radical stabilization and migration in radiation-induced DNA damage.[7,8] The intra- or inter-bases hydrogen atom transfer produces the "rare" imino-enol tautomers of Watson-Crick bases, which are thought to be responsible for genetic instability.[9] Indeed, a large number of investigations of proton transfer in DNA bases[10–21] have been performed since the original hypothesis that the fidelity of DNA replication is directly coupled with the proton transferability of DNA bases, proposed in the seminal article by Watson and Crick.[9] Due to the high barrier of single intramolecular proton transition,[22–24] the "rare" forms of the isolated nucleic bases do not reach the equilibrium state within the time necessary for the synthesis of new DNA.[10] In addition, a double proton transfer (DPT) mechanism for producing the "rare" tautomers has been proposed by Löwdin.[1,25]

DPT in guanine-cytosine (GC) and adenine-thymine (AT) base pairs has been extensively studied using a wide range of approaches, not only in the gas phase,[26–33] but also under physiological conditions, such as hydration,[10,34–40] metal ions,[41–44] and base pair stacking.[45,46] However, the detailed mechanism of DPT in base pairs in the gas phase remains controversial. For GC, the calculations of Gorb *et al.*[10] and Cerón-Carrasco *et al.*[47] proved that GC2 (Fig. 1) is the only stable "rare" tautomer at the B3LYP/6-31G(d) and BP86/6-311++G(d,p) levels, respectively; whereas Villani[48] suggested that GC3 (Fig. 1) is also a stable "rare" tautomer state at the B3LYP/cc-pVDZ level. The situation is more ambiguous in the case of AT base pairs. According to the calculations of Cerón-Carrasco *et al.* at the BP86/G-311G(d, p) level,[49] the canonical structure of AT exhibits the global minimum, and neither AT1, nor AT2 (Fig. 2) is a stable structure. This conclusion is consistent with the works of Florián *et al.*[27] and Villani.[30] In contrast, a very shallow minimum representing the AT2 structure and the corresponding transition state have been observed in the potential energy surface of the DPT in AT base pairs.[10] More direct evidence for the stable existence of AT2 comes from the calculations at the level of B3LYP/cc-pVDZ reported by Villani,[50] which showed that there are two diabatic states in the case of the hydrogen movement in the N6(A)-O4(T) bridge while only one diabatic state in the case of hydrogen movement in the N1(A)-N3(T) bridge. The mechanism of the inter-bases movement of the hydrogen atom transfer is also in dispute. The results of Cerón-Carrascon *et al.*[49] indicated that the DPT reaction between AT may not occur in the gas phase, because no tautomer remains stable except for the canonical structure of AT and the DPT between GC must be a concerted and be closer to synchronous process. Gorb *et al.* constructed the two-dimensional adiabatic potential energy surface of proton tansfer at the B3LYP/6-31G(d) level by varying the interatomic distance (O6(G)-H4(C), N3(C)-H1(G), and H2(G)-O2(C) in GC and H6(A)-O4(T) and H3(T)-N1(A) in AT (see Figs. 1 and 2), and optimizing all other geometrical parameters.[10] Based on the potential energy surface, they demonstrated that the DPT process in AT is concerted and asynchronous, and

a)Electronic mail: hjliang@ustc.edu.cn.

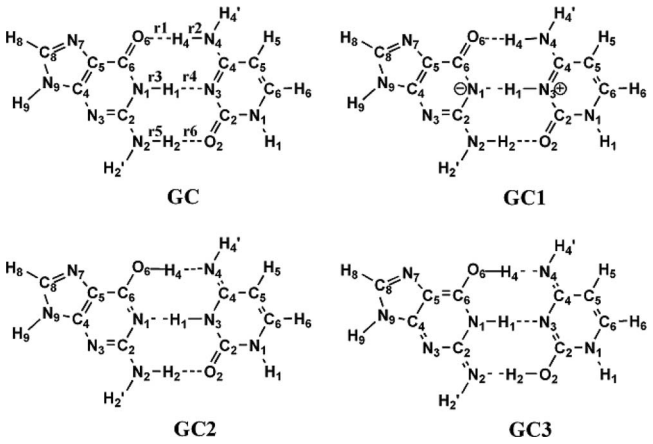

FIG. 1. The structure of a Watson-Crick GC base pair and its possible tautomers. $r1 \sim r6$ in GC comprise the bond length, which is used to describe the structural properties in the proton transfer reactions in GC.

that in GC is concerted, while the movement of the hydrogen atoms depends on the *ab initio* electronic structure methods. Recently, Villani suggested that the concerted double-hydrogen process in DNA base pairs begins with a hydrogen atom of a purine base,[48,50] which implies that only one diabatic state is involved in the hydrogen atom movement from the pyrimidine base to the purine base.

In spite of the extensive studies described above, a deeper understanding of the detailed mechanism of the proton transfer in base pairs in the gas phase is still obviously needed. The chemical reactions are intricate thermodynamics and kinetics processes and are still a challenge for studying in both experiment and theory. The free energy surface is essential in exploring the nature of reactions when temperature and entropy effects are important, so that finding the transition states remains an art for the traditional canonical gradient methods.[51] *Ab initio* molecular dynamics is a promising theoretical method in resolving and identifying the competing chemical pathways because it is a combination of the first principle electronic structure calculations and finite temperature dynamics.[52–54] The goal of the current work is to present a detailed picture of the DPT process in base pairs in the gas phase by considering the thermodynamic effect then constructing the free energy surface of the DPT process. In this paper, we report the DPT mechanism of GC and AT base pairs in the gas phase using *ab initio* constrained molecular dynamics[53] and metadynamics.[54–56]

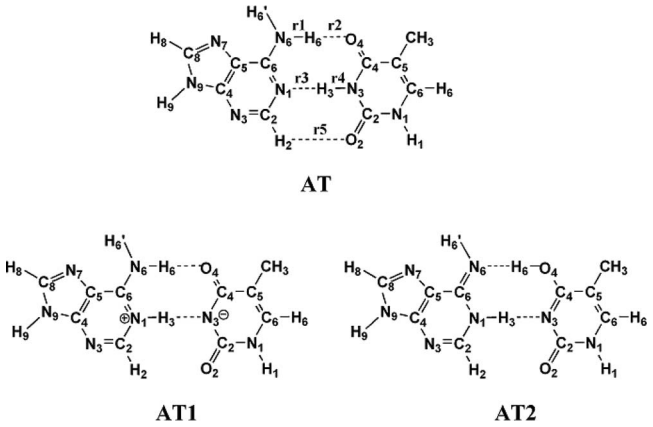

FIG. 2. The structure of a Watson-Crick AT base pair and its possible tautomers. $r1 \sim r5$ in AT comprise the bond length, which is used to describe the structural properties in the proton transfer reactions in AT.

II. SIMULATION AND METHOD DETAILS

All of the calculations reported here were carried out within the Car-Parrinello scheme[52] employing the CPMD software package (version 3.13.2).[57] Car-Parrinello molecular dynamics (MD), which is also called "first principle MD" or "*ab initio* MD," is an efficient simulation technique that has been successfully applied in studies on the hydration structure around peptides and their interactions under aqueous conditions,[58] catalytic mechanisms in metal ions-based enzymes,[59,60] and electronic structure and charge transfer in DNA,[61,62] etc.[63,64] The two nucleic base pairs, the initial topologies of which were generated by PRODRG,[65] were first optimized and then used as the starting conformation for further simulations. The Becke-Lee-Yang-Parr (BLYP) exchange-correlation functional[66,67] was employed and the valence electrons were described by a plane wave set with 70 Ry kinetic energy cutoff, which was shown to be large enough to ensure energy convergence. The BLYP density functional theory has been applied in a broad range of investigations on hydrogen bonding and proton transfer dynamics in DNA base pairs,[10,27,40,46,68–71] and the results demonstrated that the BLYP calculations predict the energetics of these systems comparable to the MP2 theory for the proton-bound dimer.[72] The interaction between the valence electrons and the ionic cores was described by the standard Troullier-Martins norm-conserving pseudopotential.[73] In the present simulations, all of the hydrogen atoms were replaced by deuterium atoms, which allowed the use of a time step of 4.0 a.u. (0.09676 fs), and the fictitious electronic mass was set to 500 a.u. The temperature was kept at 300 K and the massive Nosé-Hoover chain algorithm[74,75] was employed to achieve thermostatting during the simulations. The electronic wave functions were quenched to the Born-Oppenheimer surface at the beginning of the simulations. The nucleic acid base pairs in the gas phase were simulated in an isolated cubic box of length 20.0 Å and the cluster boundary condition method of Martyna and Tuckerman[76] was used.

In the present work, *ab initio* constrained MD[53] and metadynamics[54–56] are used to derive the free energy change along with the predefined reaction coordinate. In the constrained MD, the system achieves the reaction from the initial state A to the product state B as the reaction coordinate ξ propagates from η_A to η_B step by step. MD simulations are used to sample the conformations at each constrained point.[77,78] The relative free energies can be determined by integrating the mean constraint force $\langle f \rangle$,

$$\Delta F(\eta_A \rightarrow \eta_B) = \int_{\eta_A}^{\eta_B} \langle f(\xi) \rangle d\xi. \quad (1)$$

η_A and η_B are the values of ξ at the states A and B, respectively. $\langle f(\xi) \rangle$ is the converged value of f at ξ. For all the reactions investigated in the present paper, the pathway from

the reactant to the product is divided into 15 steps, with a 0.2 Å resolution close to the reactant and product, and 0.1 Å resolution in the vicinity of the intermediate states. The sampling time for each step varies from 2 to 3 ps. As a convergence criterion, the fluctuation of the mean force observed in the last picosecond at each step was demanded to be less than 5%. The free energy surface of the reaction was obtained by the thermodynamic integration equation displayed above (Eq. (1)).

Metadynamics[54–56] is a powerful method that can be proficiently used to reconstruct the free energy surface and accelerate the sampling of rare events. Based on a set of preliminarily selected variables (CVs) that are assumed to be able to describe the reaction process, such as angle, bond distance, coordinate numbers, and so on, an extended Lagrangian can be constructed by adding the harmonic potential of the freedom of S_α and the history-dependent Gaussian potential to the CP Lagrangian L_{CPMD}.

$$L_{MTD} = L_{CPMD} + \sum_\alpha \frac{1}{2}\mu_\alpha \dot{S}_\alpha^2 - \sum_\alpha \frac{1}{2}K_\alpha (S_\alpha(r) - S_\alpha)^2 - V(t,s), \quad (2)$$

where the variable can be seen as a set of atomic coordinates or electronic functions and K_α is the force constant of the corresponding harmonic potential. The last term $V(t,s)$ is the history-dependent Gaussian potential, which reads as

$$V(t,s) = \sum_{t_i<t} H exp\left[-\frac{(s-s_i)^2}{2(\Delta W^\perp)^2}\right] \times exp\left[-\frac{((s^{i+1}-s^i)\cdot(s-s^i))^2}{2\left(\Delta W_i^\parallel\right)^4}\right]. \quad (3)$$

In the equation above, H is the height of Gaussian biasing potential and the first spherical Gaussian, with width $\Delta W^\perp$, is multiplied with a second Gaussian, with width $\Delta W_i^\parallel$, the value of which is dependent on the displacement of between the potentials added at t_i and t_{i+1}. The slowly building potential forces the dynamics to escape from the local energy minima and span the entire phase area. $V(t,s)$ is used to reconstruct the free energy surface after the simulation. The system involved with metadynamics is able to escape from any free energy minimum and walk around the whole dimensional space defined by prior selected CVs as the history-dependent potential iteratively compensates the underlying free energy. This feature makes metadynamics a rather flexible tool in exploring new reaction pathways.[79] The CVs used in this work are the length difference of two consecutive bonds. More details on the CVs will be presented in greater detail in Sec. III. In the metadynamics simulations of the DPT in both the GC and AT base pairs, similar parameters were used. The mass of the CVs, μ_α, was set to 10 amu, and the force constant of the spring, K_α, was set to 0.6 a.u. The temperature T was set to 300 K. Initially, the width ($\Delta W^\perp$) and height (H) of the Gaussian-shaped potential hills were equal to 0.20 a.u. and 0.005 a.u. (0.314 kcal/mol), respectively. The distance criteria for the minimal displacement from the previous hill before adding the next hill, ΔS^{sim}, was set to $\Delta S^{sim} = \frac{3}{2}W = 0.30$ a.u. The minimum and maximum MD steps (Δt^{min} and Δt^{max}, respectively) for the system to relax before the next hill can be added were 40 and 500, respectively. After these hills had filled up the reagent and the product wells, medium type Gaussian-shaped potential hills were added. The width and height of the medium type hills were 0.10 a.u. and 0.002 a.u., respectively. During the second stage, the distance criteria ΔS^{sim} was set to 0.15 a.u. The smaller type Gaussian-shaped potential hills were added after the medium type hills filled up the reagent and the product wells. The width and height of the small type hills were 0.05 a.u. and 0.001 a.u., and ΔS^{min} was set to 0.075 a.u. For the DPT in the GC and AT base pairs, a total of more than 20 000 Gaussian-shaped potential hills were added for each during the course of metadynamics simulation.

III. RESULTS AND DISCUSSION

The main purpose of the present work is to explore the mechanism underlying the movement of protons in the proton transfer reaction in base pair from the most stable canonical Watson-Crick structure to its possible tautomers (see Figs. 1 and 2). Before presenting the results and discussion, we would like to illustrate several keywords, which are important in describing the reaction mechanism. The concerted and stepwise (step by step) mechanisms are two different pathways in the proton transfer reaction of a base pair. The difference between the concerted and stepwise pathways determined by the existence of the intermediate in the potential energy surface. The concerted pathway is such that multiple hydrogen transfers occur in a single reaction with no intermediate, while there is a stable intermediate state in a stepwise reaction. Asynchronous and synchronous are two additional terms corresponding to the movement of protons in the proton transfer reaction. They refer to the fact that the proton movements in both the concerted and stepwise DPT are not always synchronous, they are sometimes asynchronous.

A. The GC base pair

There are three possible types of tautomers produced by DPT in GC base pairs. The calculations of Cerón-Carrasco *et al.*[47] showed that only the tautomer produced by DPT in the O6(G)-N4(C) and N1(G)-N3(C) bridge is possible, so we only take GC2 (Fig. 1) into consideration in the present work. Two mechanisms of the hydrogen movement of DPT have been investigated by *ab initio* constrained MD: (a) the movement of H1(G) in the N1(G)-N3(C) bridge; (b) first, H1(G) moves from G to C in the N1(G)-N3(C) bridge, while the bond length of H4(C)-N4(C) is restrained to 1.04 Å with a force constant of 1.0 kcalmol^{-1}Å^{-2} (SPT1); then the H4(C) atom moves from C to G, while the bond length of H1(G)-N3(C) is restrained to 1.04 Å with a force constant of 1.0 kcalmol^{-1}Å^{-2} (SPT2). The mechanism (b) is proposed based on static *ab initio* studies by Gorb *et al.*[10] and Villani.[48] Additionally, in both mechanism (a) and (b), although no restraint has been applied to the distance of N1(G)-N3(C) or

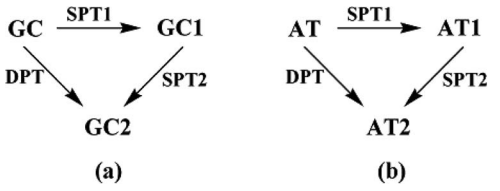

FIG. 3. Schematic diagrams of possible proton transfer reactions in GC (a) and AT (b) base pairs in the gas phase.

O4(G)-N4(C), no big change has been observed for the value of $r1 + r2$ or $r3 + r4$ in these studies (see Figs. 4 and 5); thus, it is reasonable to use $R_{[r2-r1]}$ or $R_{[r3-r4]}$ as the reaction coordinate. The reaction diagram is presented in (a) of Fig. 3.

The structural and free energy properties of the mechanisms (a) and (b) are displayed in Figs. 4 and 5, respectively. In mechanism (a), as $R_{[r3-r4]}$ increases from negative to positive, the hydrogen atom H1(G) achieves the transfer from N1(G) to N3(C), and a well-defined barrier can be identified at $R_{[r3-r4]} = 0.25$ Å. The free energy surface shows that two minima, the reagent and product, are separated by a transition state. The free energy of the product is 9.4 kcal/mol higher than that of the reagent, and the activation energy is 13.6 kcal/mol. The decrease of the O6(G)-H4(C) distance is slight and the change of the H4(C)-N4(C) bond length is very small before H1(G) achieves the transfer from N1(G) to N3(C). It is interesting to note that there is a step of $R_{[r2-r1]}$ around the transition state. Obviously, as H1(G) arrives at the critical point in the N1(G)-N3(C) bridge ($R_{[r3-r4]} = 0.25$ Å), H4(C) "jumps" from C to G and achieves the transition. This phenomenon implies that the hydrogen atom H4(C) in the O6(G)-H4(C) bridge moves asynchronously with the movement of H1(G) in the N1(G)-N3(C) bridge. The behavior of H4(C) observed here is consistent with the work reported by Villani,[48] which presented that the transfer of H4(C) in the

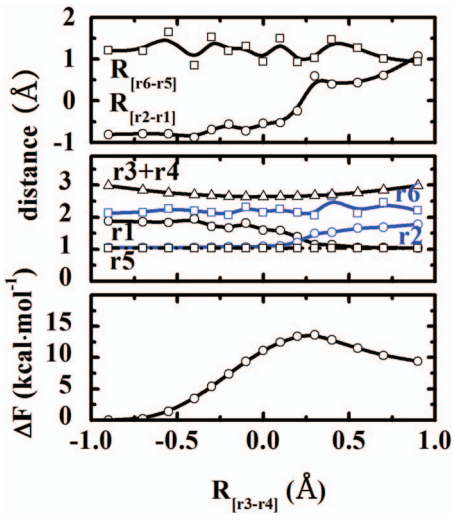

FIG. 4. The structural and free energy properties of mechanism (a) (see Sec. III A) of the DPT in the GC base pair in the gas phase derived using the *ab initio* constrained MD. The bond length difference between N1(G)-H1(G) and H1(G)-N3(C) ($R_{[r3-r4]}$) was chosen as the reaction coordinate. The free energy profile, average length of selected bonds along the pathway, and the bond length difference between N4(C)-H4(C) and H4(C)-O6(G) ($R_{[r2-r1]}$), and O2(C)-H2(G) and H2(G)-N2(G) ($R_{[r6-r5]}$) are presented.

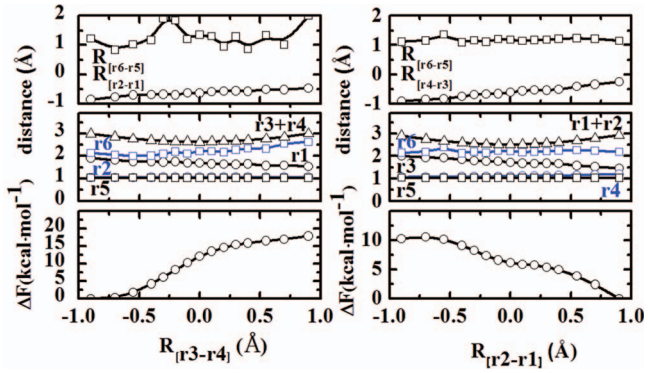

FIG. 5. The structural and free energy properties of mechanism (b) of the DPT in the GC base pair in the gas phase derived using the *ab initio* constrained MD. The left graph is the first step (SPT1), and the bond length difference between N1(G)-H1(G) and H1(G)-N3(C) ($R_{[r3-r4]}$) was chosen as its reaction coordinate; the right graph is the second step (SPT2), and the bond length difference between N4(C)-H4(C) and H4(C)-O6(G) ($R_{[r2-r1]}$) was chosen as its reaction coordinate.

O6(G)-H4(C) bridge happens suddenly until H1(G) reaches a particular position. In mechanism (b), the DPT process is artificially divided into two steps. From the free energy profile, we know that no stable intermediate state exits. The free energy increases and decreases in SPT1 and SPT2, respectively. This behavior indicates that GC1, the product of the SPT1 reaction, is unlikely to remain stable. The barrier of the entire reaction (SPT1 and SPT2) is 18.1 kcal/mol, which is 4.5 kcal/mol higher than that of mechanism (a). For the atoms not directly involved in the reactions, a structural change may be seen during the process of SPT1 and SPT2. In SPT1, as H1(G) moves from N1(G) to N3(C), the distance between H2(G) and O2(C) ($r6$) increases, while the distance between O6(G) and H4(C) decreases. The opening induced by the single proton transfer in the N1(G)-N3(C) bridge implies the tendency towards the movement of H4(C) in the O6(G)-N4(C) bridge, and again implies that the created zwitterionic GC1 is not able to remain stable. The structure of GC1 is adjusted slightly in SPT2 and achieves GC2. To summarize, the free energy profile and structural properties of mechanisms (a) and (b) calculated by *ab initio* constrained MD demonstrate that the DPT reaction in GC is a concerted and asynchronous process.

We employed metadynamics to construct a two dimensional free energy surface for the DPT in GC (Fig. 6). During the simulation of metadynamics, the distance of N1(G)-N3(C) is restrained to 2.96 Å (its value at equilibrium state for canonical GC) with a force constant of 0.01 kcalmol^{-1}Å^{-2}. For this reaction, however, there exist, several pathways starting at GC that all end up at the local minimum GC2, which is approximately 11 kcal/mol higher in energy than the former. We found that three representative pathways from GC to GC2, labeled 1, 2, and 3, all contribute to the same reaction channel. The free energy difference between GC2 and GC, as estimated by metadynamics is approximately 2 kcal/mol higher than that estimated by constrained MD. These two values are slightly smaller than the value calculated by the DFT approach at the level of B3LYP/cc-pVDZ (approximately 13 kcal/mol),[48] but slightly larger than the values calculated at

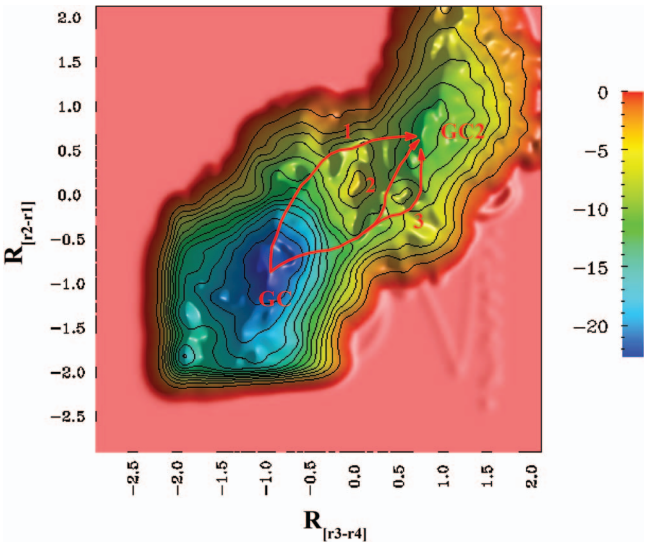

FIG. 6. Free energy surface of the DPT in the GC base pair in the gas phase derived from two-dimensional metadynamics. The bond length differences between N1(G)-H1(G) and H1(G)-N3(C) ($R_{[r3-r4]}$), N4(C)-H4(C) and H4(C)-O6(G) ($R_{[r2-r1]}$) were chosen as the reaction coordinates.

the level of MP2/BP86/6-311++G(d,p) (8.28 kcal/mol)[47] and MP2/cc-pVTZ//MP2/6-31G(d) (7.6 kcal/mol, its thermal correction to enthalpy and entropy values is taken from MP2/6-31G(d) calculations[10]). From the free energy surface, it was determined that there is a very shallow local minimum (∼1 kcal/mol) around $R_{[r3-r4]} = 0.0$ Å in the three pathways. The free energy required for the system to surmount the barrier separating GC and GC2 is approximately 14 kcal/mol, which is quite close to the value calculated by constrained MD (13.6 kcal/mol) and slightly larger than the values calculated at the level of MP2/BP86/6-311++G(d,p) (9.72 kcal/mol)[47] and MP2/cc-pVTZ//MP2/6-31G(d) (10.6 kcal/mol).[10] Besides, we again know that the energetics of the studied systems (such as the reactant, product, and other states on the free energy surface) presented here is comparable to that calculated at the level of MP2/cc-PVDZ.[10] The active barrier of route 1 is higher by approximately 1 kcal/mol than that of route 2 or 3. Although the activation free energy is similar for the three pathways, the routes are different. In route 1, the transition of H4(C) in the O6(G)-N4(C) bridge occurs before the transition of H1(G) in the N1(G)-N3(C) bridge for the reaction from GC to GC2. The transition states are located at $R_{[r2-r1]} = 0.5$ Å, $R_{[r3-r4]} = 0.0$ Å. The paths in routes 2 and 3 are very close to each other. In routes 2 and 3, the progress of the movement of H1(G) in the N1(G)-N3(C) bridge is faster than that of H4(C) in the O6(G)-N4(C) bridge before the system arrives at the transition state at $R_{[r2-r1]} = 0.0$ Å, $R_{[r3-r4]} = 0.5$ Å for route 2, and $R_{[r2-r1]} = -0.2$ Å, $R_{[r3-r4]} = 0.6$ Å for route 3. Thus, the movements of H1(G) and H4(C) are rather asynchronous processes based on our observation of the two-dimensional free energy surface constructed using metadynamics after considering the dynamics and temperature effects, although they are coupling with each other at some stages during the DPT reaction. This result is different from the conclusion derived from the static calculations using the DFT theories at the level of B3LYP, which state that the pathway for DPT between GC in the gas phase is a synchronous mechanism,[47] but more likely in line with the MP2 results.[10,47] Therefore, we believe that the hydrogen transfer begins with either the H1 atom of G, or the H4 atom of C will induce a concerted and asynchronous DPT and generate the imino-enol form of GC base pair. The movement of the hydrogen atom in the O6(G)-N4(C) bridge is slightly harder, by approximately 1.0 kcal/mol, than that which occurs in the N1(G)-N3(C) bridge.

B. The AT base pair

Just as in the case of GC, two mechanisms of hydrogen movement in DPT in the AT base pair were investigated by *ab initio* constrained MD: (a) the movement of H3(T) in the N1(A)-N3(T) bridge; (b) H3(T) first moves from T to A in the N1(A)-N3(T) bridge while the bond length of N6(A)-H6(A) is restrained to 1.03 Å with a force constant of 1.0 kcalmol^{-1}Å^{-2} (SPT1); then H6(A) moves from A to T while the bond length of H3(T)-N1(A) is restrained to 1.03 Å with a force constant of 1.0 kcalmol^{-1}Å^{-2} (SPT2). Similar as GC, no restraint has been applied to the distance of N1(A)-N3(T) or N6(A)-O4(T) in both mechanism (a) and (b). The reaction diagram can be found in (b) of Fig. 3.

The structural and free energy properties of the mechanisms (a) and (b) are displayed in Figs. 7 and 8, respectively. In mechanism (a), H3(T) transfers from T to A as the $R_{[r4-r3]}$ value increases from negative to positive. From the free energy profile, we know that a region ($R_{[r4-r3]} > 0.0$ Å) exists, where H3(T) is attached to A and a relatively stable state of the AT base pair is observed, but there is no stable configuration (local minimum on the free energy surface) at the region $R_{[r4-r3]} > 0.8$ Å. The movement of H3(T) induces the distance between H2(A) and O2(T) to decrease, then increase in the region of $R_{[r4-r3]} \sim 0.0$ Å. Moreover, although the bond length of N6(A)-H6(A) increases and that of H6(A)-O4(T) deceases

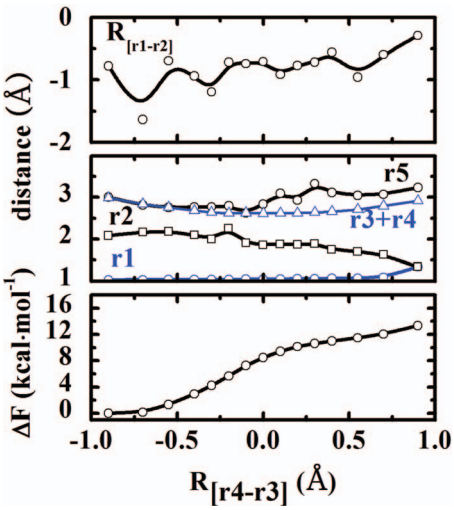

FIG. 7. The structural and free energy properties of mechanism (a) (see Sec. III B) of proton transfer in the AT base pair in the gas phase derived using the *ab initio* constrained MD. The movement of hydrogen atom H3(T) starts at the N3(T)-N1(A) bridge. The bond length difference between N3(T)-H3(T) and H3(T)-N1(A) ($R_{[r4-r3]}$) was chosen as the reaction coordinate.

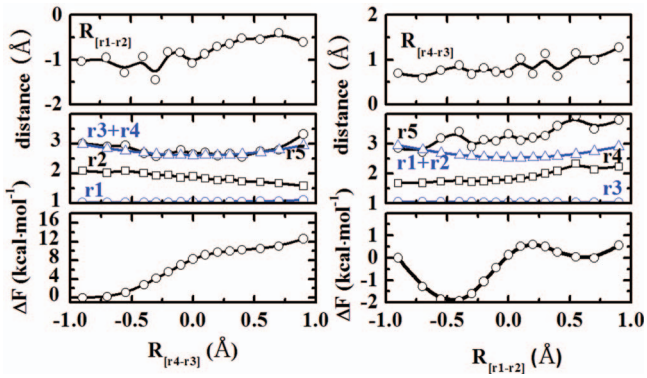

FIG. 8. The structural and free energy properties of mechanism (b) of the DPT in the AT base pair in the gas phase derived using the *ab initio* constrained MD. The left graph is the first step (SPT1), and the bond length difference between N3(T)-H3(T) and H3(T)-N1(A) ($R_{[r4-r3]}$) was chosen as its reaction coordinate; the right graph is the second step (SPT2), and the bond length difference between N6(A)-H6(A) and H6(A)-O4(T) ($R_{[r1-r2]}$) was chosen as its reaction coordinate.

in the reaction, H6(A) maintains a stable attachment to N6(A) and there is no complete transfer in the N6(A)-O4(T) bridge observed. Thus, in mechanism (a), a single proton transfer occurs and a zwitterionic AT1 is formed. In order to study the potential existence of AT2, mechanism (b) was investigated. In SPT1, the free energy increases and only one stable configuration corresponding to the canonical AT is observed. The free energy profile of SPT2 exhibits two local minima. One that corresponds to an intermediate is located at $R_{[r1-r2]}$ ~ -0.5 Å, and the other one, corresponding to the product AT2, is located at $R_{[r1-r2]} \sim 0.8$ Å. The free energy of the intermediate state is 12.7 kcal/mol higher than that of AT and is 1.9 kcal/mol lower than the product AT2. Thus, we know that the DPT in AT base pair is a stepwise process but there is no conclusion regarding the coupling of the movements of the hydrogen atoms between A and T yet. The existence of AT2 proved in our studies is also certified by the potential energy surfaces calculated by the DFT approach at the B3LYP/cc-pVDZ level[50] and *ab intio* approach at the MP2/cc-pVDZ//MP2/6-31G(d) level.[10] However, the local minimum of AT2 is disappeared on the Gibbs free energy surface calculated by Gorb *et al.* at the level of MP2/cc-pVDZ//MP2/6-31G(d).[10] The free energy surfaces in the present paper are calculated using MD simulations at room temperature, while the Gibbs free energies reported by Gorb *et al.*[10] are calculated based on structures achieved by geometry optimizations at 0 K. The difference of methodologies in these two cases would induce conformation and energy of AT2 differ slightly, and thus result in different results. At the structural level, in SPT1, the movement of H3(T) in the N1(A)-N3(T) bridge induces the distance between H2(A) and O2(T) to first decrease and then increase, while the bond length of H6(A) and O4(T) decreases as the the reaction proceeds. The distance between H2(A) and O2(T) keeps increasing during SPT2. We observed that the length of H2(A)-O2(T) of AT2 is approximately 0.7 Å larger than that in the original canonical AT. In addition to the small barrier between AT2 and the intermediate (less than 0.5 kcal/mol for AT2), the structural twisting of the base pair induced by the opening of H2(A)-O2(T) would make AT2 unstable and also greatly increase the possibility of the reverse proton transfer of SPT2, so we think that the configuration of AT2 would not be stable in the gas phase although the DPT may occur in AT. This is why the transfer of H6(A) in the N6(A)-O4(T) bridge was not observed in mechanism (a), in which we only push the movement of H3(T) atom in the N1(A)-N3(T) bridge.

In the above studies using the *ab initio* constrained MD, only the mechanism, in which hydrogen atom H3(T) moves in the N1(A)-N3(T) bridge first and then H6(A) atom in the N6(A)-O4(T) bridge, has been considered. However, Villani proposed that hydrogen atoms H6(A) and H3(T) perform in different ways if we only push H6(A) in the N6(A)-O4(T) bridge.[50] It was shown that H3(T) would achieve the transfer from T to A as $R_{[r1-r2]}$ arrives at ~ -0.4 Å. To capture the complete DPT process in AT, we took the two CVs $R_{[r1-r2]}$ and $R_{[r4-r3]}$ simultaneously into account and span a two dimensional free energy subspace to be mapped by metadynamics sampling. During the simulation, the distance of N1(A)-N3(T) is restrained to 2.88 Å (its value at equilibrium state for canonical AT) with a force constant of 0.01 kcalmol^{-1}Å^{-2}. The reconstructed free energy surface is shown in Fig. 9. Two minima beyond the one that defines the starting structure, AT, were readily identified. The DPT process is described by the motion of the system on the free energy surface from the reagent (AT) to the product (AT2) and is found to consist of two distinct steps. In the first step, $R_{[r4-r3]}$ increases from -1.0 Å to -0.5 Å. Here, the H3(T)-N3(T) bond is broken, while the bond N6(A)-H6(T) is maintained. H3(T) transfers from T to A and achieves the intermediate AT1. The active barrier is approximately 9 kcal/mol, and the free energy difference between AT1 and AT is approximately 8 kcal/mol. The second step of the DPT is that of the transition H6(A) from A to T, which is connected with an increase of the $R_{[r1-r2]}$ value from ~ -1.0 Å to ~ 1.0 Å. The free energy required for the system to surmount the barrier

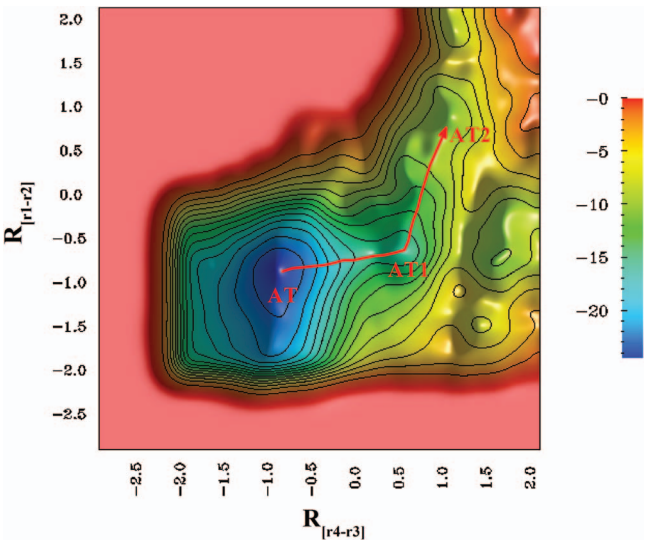

FIG. 9. Free energy surface of the DPT in the AT base pair in the gas phase derived from two-dimensional metadynamics. The bond length differences between N3(T)-H3(T) and H3(T)-N1(A) ($R_{[r4-r3]}$), and N6(A)-H6(A) and H6(A)-O4(T) ($R_{[r1-r2]}$) were chosen as the reaction coordinates.

separating AT1 and AT2 is approximately 7 kcal/mol, and the free energy of AT2 is approximately 6 kcal/mol higher than that of AT1. We observed that the free energy difference between AT2 and the intermediate state estimated by metadynamics is much larger than that in mechanism (b), which is calculated by constrained MD. We attribute the variance to the artificial design of the reaction steps in mechanism (b). The bond length of N6(A)-H6(T) and N1(A)-H3(T) is restrained in SPT1 and SPT2 in mechanism (b), respectively. In contrast, the movements of H3(T) and H6(A) are coupled with each other in the second step on the free energy surface. The free energy difference between AT2 and AT calculated here is approximately 14 kcal/mol, which is close to the value calculated at the B3LYP/cc-pVDZ level by Villani (approximately 13 kcal/mol), but no intermediate state was observed in their calculations.[50] In summary, from this investigation, it is concluded that the DPT in AT base pair is a stepwise and, most likely, an asynchronous process.

IV. CONCLUSIONS

The mechanism of proton transfer in GC and AT base pairs was investigated by *ab intio* MD techniques. It is shown that only two minima (the reagent and product), separated by a transition state, are present in the free energy profile calculated by the *ab initio* constrained MD. H4(C) in the O6(G)-N4(C) bridge moves asynchronously with the movement of H1(G) in the N1(G)-N3(C) bridge. The calculations for mechanisms (a) and (b) indicate that the DPT in the GC base pair is a concerted and asynchronous process. The two-dimensional free energy surface constructed by the sampling on the $R_{[r3-r4]}$ and $R_{[r2-r1]}$ coordinates using the metadynamics Car-Parrinello technique shows that there are three pathways starting from the canonical GC to GC2. The free energy barriers of the three pathways are close to each other, but they are nevertheless significantly different, especially in terms of the movements of the hydrogen atoms in the two bridges.

The transition of H6(A) in the N6(A)-O4(T) bridge in the AT base pair was not observed when we only push H3(T) from T to A (mechanism (a)). In mechanism (b), a stable intermediate state was observed in the DPT based on the free energy profile. More details about the DPT in the AT base pair can be derived from the two-dimensional free energy surface whose reaction coordinates are $R_{[r1-r2]}$ and $R_{[r4-r3]}$. It is demonstrated that the DPT in AT consists of two distinct steps: first, H3(T) transfers from T to A, but the bond N6(A)-H6(A) is maintained, and the intermediate state AT1 is generated; second, H6(A) transfers from A to T and the whole reaction is completed. Thus, we conclude that the DPT in the AT base pair is a stepwise and most likely asynchronous process. Based on our observations and analyses, we understand that the real mechanism of the DPT in base pairs is complex and may mix with different characteristics. Studies from different angles are necessary for revealing the truth.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (NNSFC) (Grant Nos. 20934004 and 91127046); the NBRPC (Grant Nos. 2012CB821504 and 2010CB934500), and the "Bairen" fund of the CAS.

[1] P.-O. Löwdin, Rev. Mod. Phys. **35**, 724 (1963).
[2] P.-O. Löwdin, "Quantum genetics and the aperiodic solid: Some aspects on the biological problems of heredity, mutations, aging, and tumors in view of the quantum theory of the DNA molecule," Adv. Quantum Chem. **2**, 213–360 (1966).
[3] H.-A. Wagenknecht, Angew. Chem., Int. Ed. **42**, 2454 (2003).
[4] T. Ito and S. E. Rokita, Angew. Chem., Int. Ed. **43**, 1839 (2004).
[5] A. K. Ghosh and G. B. Schuster, J. Am. Chem. Soc. **128**, 4172 (2006).
[6] F. L. Gervasio, M. Boero, and M. Parrinello, Angew. Chem., Int. Ed. **45**, 5606 (2006).
[7] D. H. Johnston, K. C. Glasgow, and H. H. Thorp, J. Am. Chem. Soc. **117**, 8933 (1995).
[8] S. Steenken, J. P. Telo, H. M. Novais, and L. P. Candeias, J. Am. Chem. Soc. **114**, 4701 (1992).
[9] J. D. Watson and F. H. C. Crick, Nature (London) **171**, 964 (1953).
[10] L. Gorb, Y. Podolyan, P. Dziekonski, W. A. Sokalski, and J. Leszczynski, J. Am. Chem. Soc. **126**, 10119 (2004).
[11] J. S. Kwiatkowski and B. Pullman, Adv. Heterocycl. Chem. **18**, 199–335 (1975).
[12] M. Nowak, K. Szczepaniak, A. Barski, and D. Shugar, J. Mol. Struct. **62**, 47 (1980).
[13] G. M. Blackburn and M. J. Gait, *Nucleic Acids in Chemistry and Biology* (Oxford University Press, New York, 1990), p. 446.
[14] J. S. Kwiatkowski, T. J. Zielinski, and R. Rein, Adv. Quantum Chem. **18**, 85–130 (1986).
[15] W. B. Person, K. Szczepaniak, M. Szczesniak, J. Kwiatkowski, L. Hernandez, and R. Czerminski, J. Mol. Struct. **194**, 239 (1989).
[16] G. Maroulis, M. Sana, and G. Leroy, Int. J. Quantum Chem. **19**, 43 (1981).
[17] I. R. Gould, N. A. Burton, R. J. Hall, and I. H. Hillier, J. Mol. Struct.: THEOCHEM **331**, 147 (1995).
[18] J. Šponer and P. Hobza, J. Phys. Chem. **98**, 3161 (1994).
[19] C. Colominas, F. J. Luque, and M. Orozco, J. Am. Chem. Soc. **118**, 6811 (1996).
[20] E. L. Stewart, C. K. Foley, N. L. Allinger, and J. P. Bowen, J. Am. Chem. Soc. **116**, 7282 (1994).
[21] J. Leszczynski, J. Phys. Chem. A **102**, 2357 (1998).
[22] L. Gorb and J. Leszczynski, J. Am. Chem. Soc. **120**, 5024 (1998).
[23] L. Gorb, Y. Podolyan, and J. Leszczynski, J. Mol. Struct.: THEOCHEM **487**, 47 (1999).
[24] Y. Podolyan, L. Gorb, and J. Leszczynski, Int. J. Quantum Chem. **4**, 410 (2003).
[25] P.-O. Löwdin, *Electronic Aspects of Biochemistry* (Academic, New York, 1964).
[26] J. Florián, V. Hrouda, and P. Hobza, J. Am. Chem. Soc. **116**, 1457 (1994).
[27] J. Florián and J. Leszczyński, J. Am. Chem. Soc. **118**, 3010 (1996).
[28] J. Marãnón, A. Fantoni, and J. Grigera, J. Theor. Biol. **201**, 93 (1999).
[29] V. Guallar, A. Douhal, M. Moreno, and J. M. Lluch, J. Phys. Chem. A **103**, 6251 (1999).
[30] G. Villani, Chem. Phys. **316**, 1 (2005).
[31] G. Villani, Chem. Phys. **324**, 438 (2006).
[32] S. Perun, A. L. Sobolewski, and W. Domcke, J. Phys. Chem. A **110**, 9031 (2006).
[33] F. Liu, P. Qian, S. Yan, and Y. Bu, J. Mol. Struct.: THEOCHEM **760**, 209 (2006).
[34] O. V. Shishkin, L. Gorb, and J. Leszczynski, J. Phys. Chem. B **104**, 5357 (2000).
[35] H. E. Herbert, M. D. Halls, H. P. Hratchian, and K. Raghavachari, J. Phys. Chem. B **110**, 3336 (2006).
[36] M. Kabeláč and P. Hobza, Phys. Chem. Chem. Phys. **9**, 903 (2007).
[37] H. S. Park, S. H. Nam, J. K. Song, S. M. Park, and S. Ryu, J. Phys. Chem. A **112**, 9023 (2008).
[38] A. Kumar, P. C. Mishra, and S. Suhai, J. Phys. Chem. A **109**, 3971 (2005).
[39] A. Kumar, M. D. Sevilla, and S. Suhai, J. Phys. Chem. B **112**, 5189 (2008).
[40] A. Kumar and M. D. Sevilla, J. Phys. Chem. B **113**, 11359 (2009).
[41] J. Šponer, M. Sabat, L. Gorb, J. Leszczynski, B. Lippert, and P. Hobza, J. Phys. Chem. B **104**, 7535 (2000).
[42] N. Gresh and J. Šponer, J. Phys. Chem. B **103**, 11415 (1999).
[43] J. Muñoz, J. Šponer, P. Hobza, M. Orozco, and F. J. Luque, J. Phys. Chem. B **105**, 6051 (2001).
[44] M. Noguera, J. Bertran, and M. Sodupe, J. Phys. Chem. B **112**, 4817 (2008).

[45]M. Żabicki, E. Gudowska-Nowak, and S. F. Fischer, Phys. Lett. A **374**, 50 (2009).
[46]H.-Y. Chen, C.-L. Kao, and S. C. N. Hsu, J. Am. Chem. Soc. **131**, 15930 (2009).
[47]J. P. Cerón-Carrasco, A. Requena, J. Zúñiga, C. Michaux, E. A. Perpète, and D. Jacquemin, J. Phys. Chem. A **113**, 10549 (2009).
[48]G. Villani, J. Phys. Chem. B **114**, 9653 (2010).
[49]J. P. Cerón-Carrasco, A. Requena, C. Michaux, E. A. Perpète, and D. Jacquemin, J. Phys. Chem. A **113**, 7892 (2009).
[50]G. Villani, Phys. Chem. Chem. Phys. **12**, 2664 (2010).
[51]B. Ensing, A. Laio, M. Parrinello, and M. L. Klein, J. Phys. Chem. B **109**, 6676 (2005).
[52]R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[53]A. Curioni, M. Sprik, W. Andreoni, H. Schiffer, J. Hutter, and M. Parrinello, J. Am. Chem. Soc. **119**, 7218 (1997).
[54]M. Iannuzzi, A. Laio, and M. Parrinello, Phys. Rev. Lett. **90**, 238302 (2003).
[55]A. Laio and M. Parrinello, Proc. Natl. Acad. Sci. U.S.A. **99**, 12562 (2002).
[56]C. Micheletti, A. Laio, and M. Parrinello, Phys. Rev. Lett. **92**, 170601 (2004).
[57]See http://www.cpmd.org/ for CPMD; copyright IBM corp. 1990–2008; copyright MPI für Festkörperforschung Stuttgart 1997–2001.
[58]T. Liang and T. R. Walsh, Mol. Simul. **33**, 337 (2007).
[59]M. D. Vivon, M. D. Per, and M. L. Klein, J. Am. Chem. Soc. **130**, 10955 (2008).
[60]M. D. Peraro, L. I. Llarrull, U. Rothlisberger, A. J. Vila, and P. Carloni, J. Am. Chem. Soc. **126**, 12661 (2004).
[61]F. L. Gervasio, P. Carloni, and M. Parrinello, Phys. Rev. Lett. **89**, 108102 (2002).
[62]F. L. Gervasio, A. Laio, M. Parrinello, and M. Boero, Phys. Rev. Lett. **94**, 158103 (2005).
[63]J. Thar, W. Reckien, and B. Kirchner, Top. Curr. Chem. **268**, 133 (2007).
[64]M. D. Perarol, P. Ruggerone, S. Raugei, S. Raugei, F. L. Gervasio, and P. Carloni, Curr. Opin. Struct. Biol. **17**, 149 (2007).
[65]A. W. Schüttelkopf and D. M. F. van Aalten, Acta Crystallogr., Sect. D: Biol. Crystallogr. **60**, 1355 (2004).
[66]A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[67]C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[68]J. J. Dannenberg and M. Tomasz, J. Am. Chem. Soc. **122**, 2062 (2000).
[69]A. Müller, F. Talbot, and S. Leutwyler, J. Am. Chem. Soc. **124**, 14486 (2002).
[70]V. Zoete and M. Meuwly, J. Chem. Phys. **121**, 4377 (2004).
[71]X. Li, Z. Cai, and M. D. Sevilla, J. Phys. Chem. B **105**, 10115 (2001).
[72]S. Y. Han and H. B. Oh, Chem. Phys. Lett. **432**, 269 (2006).
[73]N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).
[74]S. Nosé, J. Chem. Phys. **81**, 511 (1984).
[75]W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).
[76]G. J. Martyna and M. E. Tuckerman, J. Chem. Phys. **110**, 2810 (1999).
[77]W. K. den Otter, J. Chem. Phys. **112**, 7283 (2000).
[78]M. Sprik and G. Ciccotti, J. Chem. Phys. **109**, 7737 (1998).
[79]A. Laio and F. L. Gervasio, Rep. Prog. Phys. **71**, 126601 (2008).