Available online at www.sciencedirect.com
ScienceDirect
Chemical Physics 349 (2008) 204–209
Chemical Physics
www.elsevier.com/locate/chemphys

Water-mediated tautomerization of cytosine to the rare imino form: An ab initio dynamics study

Geza Fogarasi *

Institute of Chemistry, Eotvos University, H-1518 Budapest, Pf. 32., Hungary

Received 3 October 2007; accepted 12 February 2008
Available online 15 February 2008

Abstract

Tautomerism in nucleotide bases is one of the possible mechanisms of mutation of DNA. In spite of numerous studies on the structure and energy of cytosine tautomers, little information is available on the process of proton transfer itself. We present here Born–Oppenheimer dynamics calculations, with the potential surface obtained "on the fly" from ab initio quantum chemistry (QC) and the atoms moving classically. In search for water-mediated tautomerization the *monohydrated* complex was studied, running about 300 trajectories each of 3000–5000 points of 1 fs steps. One single trajectory has been found to lead to tautomerization. Although the QC method used in the simulations was inevitably modest (B3LYP/3-21G), higher-level test calculations along the same trajectory suggest that the simulation grasped the basic mechanism of proton transfer: a *concerted*, *synchronous* process characterized by strong coupling between the motions of the two participating hydrogen atoms.

Keywords: Ab initio dynamics; Tautomerism; Nucleobases; Proton transfer; Cytosine–water

1. Introduction

Tautomerization as a possible cause of mutations in DNA has been a recurring question ever since the historic formulation of the double helix model. One recalls Watson and Crick's comment [1]: "spontaneous mutation may be due to ... less likely tautomeric forms". This idea has been advanced by Topal and Fresco [2]. Recently, Williams and co-workers [3] performed a complex experimental study on the nucleoside analogue dP. By chemical synthesis they prepared fixed imino and fixed amino tautomers of *N*-methyl-P and, using UV spectroscopy, determined a tautomer ratio of 11:1 in favor of the imino form. By determining the kinetic parameters of the incorporation of dPTP, they found strong correlation between the tautomeric ratio and the incorporation specificity, leading to the conclusion that "minor tautomeric forms of the natural bases may play an important role in substitution mutagenesis during DNA replication".

The subject of the present study is tautomerization specifically in cytosine. It is now generally accepted that cytosine has three primary, low-energy forms as shown in Fig. 1. The amino-oxo (keto) tautomer **1** is the "canonical" form present in DNA. In the gas phase, however, the amino-hydroxy (enol) structure **2b** dominates. The imino-oxo tautomer **3a** is considered as the "rare" form. Beyond this overall picture, the exact ordering of stabilities is quite difficult because the total range of relative energies of these three tautomers is only 2–3 kcal mol^{-1}, as determined by numerous quantum chemical (QC) calculations [4–27]. Among the latter, highest level calculations [25–27] using the coupled cluster method CCSD(T) and large basis sets give the following energies relative to **2b**: $\Delta E(\mathbf{1}) = 1.2$–$1.6$ kcal mol^{-1}, $\Delta E(\mathbf{3a}) = 1.5$–$2.1$ kcal mol^{-1}. It is remarkable that according to these results the "rare" imino-oxo form should be only little less stable, if at all, than the amino-oxo form. Beyond energies, Gibbs free energies bring the stabilities even closer: at room temperature, ΔG

* Tel.: +36 1 3722540; fax: +36 1 3722909.
E-mail address: fg@chem.elte.hu

0301-0104/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemphys.2008.02.016

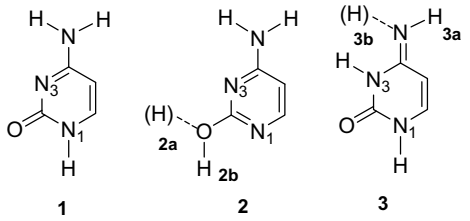


Fig. 1. Three low-energy tautomers of cytosine. Less stable rotamers are indicated by dashed lines, with hydrogens in parentheses.

was calculated [27] practically the same for **1** and **3a**, and only ~0.8 kcal mol^{-1} above **2b**. From the methodological point of view, it should be noted that the widely used MP2 method seems unsatisfactory in the light of the coupled cluster values: MP2 gives relative energies overestimated by ~0.5 kcal mol^{-1} for **1** and ~1.5 kcal mol^{-1} (!) for **3a**. Density functional theory, DFT, gives a picture [7,10,25,27] *qualitatively* different from standard wave function theory: using various functionals and basis sets, DFT calculations consistently predict the keto form **1**, rather than the enol form **2b**, as the most stable tautomer. This seems to be a failure of the DFT method: a careful analysis by Piacenza and Grimme [28] has shown that the deficiency can be ascribed to the insufficient reproduction of the aromatic character of the ring – the latter being most pronounced in **2b** as indicated by its Lewis structure.

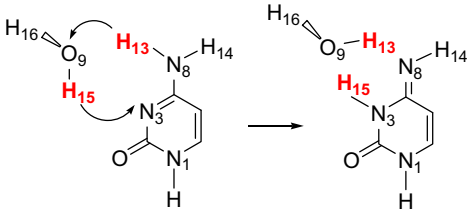
Experimental spectroscopic results [29–33] all agree that the enol form is dominant in the gas state, and the keto form is also present in significant amounts. At the same time, information about the imino-oxo form is very uncertain; normally, the presence of a few percent is suggested, much less than in the theoretical results. In fact, a recent sophisticated study [33] of the infrared spectrum of cytosine in helium nanodroplets does not see **3a** at all, while reports the observation of rotamer **2a**, in addition to the two major components **2b** and **1**. These authors refute the original interpretation of a molecular beam microwave spectrum by Brown et al. [32] which suggested the presence of a small amount of the imino-oxo form. The appearance of the rotamer pair seems to be in accord with theory: virtually independent of the level of quantum chemical method, the computed energy difference between **2a** and **2b** is small, only ~0.75 kcal mol^{-1} [24].

In the presence of water the situation may change qualitatively. Estimates of the interaction with bulk water by electrostatic continuum models show rather large variations [26,34,35]. Explicit consideration of water molecules gives the following picture [36]: water binds to the amino-oxo and imino-oxo form stronger than to the hydroxy tautomer by 1.5–2.0 kcal mol^{-1} and 0.4–0.8 kcal mol^{-1}, respectively. This leads to relative energies for the monohydrates: E(**2b**.H_2O) = 0.0 (by def.), E(**1**.H_2O) = −0.5 to 0.0, E(**3a**.H_2O) = 0.6 to 1.2 kcal mol^{-1}. Complexes of cytosine with up to 14 water molecules have also been investigated by quantum chemistry [35,37,38]. On the experimental side, Nir et al. [39] notice the existence of clusters with up to fifty water molecules in their REMPI spectroscopy study on a supersonic beam.

Concerning the intramolecular proton transfer (PT) between tautomers, several QC studies have calculated the transition state (TS) barrier. Early on, Sobolewski and Adamowicz [19] investigated the monohydrated complexes of tautomers **1** and **3a** and, in conjunction with this, the proton transfer between the two. Defining the PT reaction coordinate simply as the N_8–H_{13} distance (Scheme 1), they determined the potential energy function along it by optimizing all other coordinates. The calculated barrier was ~20 kcal mol^{-1} at the HF/3-21G level. A similar study by Gorb and Leszczynski [40] investigated the other tautomerization, **2b** to **1**. Using more accurate (electron correlation) methods, they emphasized that water lowers the barrier significantly. Similar results were reported by Morpurgo et al. [41] on the **1** to **3a** tautomerization. Podolyan et al. [42] discuss in detail the significance of the rare imino tautomer **3a** for mutation, investigate the equilibrium between **1** and **3a** and estimate the rate constant based on their "instanton model". For the TS barrier they calculate values between 17 and 20 kcal mol^{-1}.

In spite of numerous studies like the above, little is known about the *mechanism* of proton transfer in cytosine. Our aim here is to catch the process in detail, on the basis of explicit ab initio dynamics calculations. Because the H-atom on N_1 is blocked by the glycoside bond in DNA, only the amino-oxo → imino-oxo tautomerization, sketched in Scheme 1, will be investigated.

The chemist's notion of reaction mechanisms is based on the Born–Oppenheimer (B–O) approximation: the atomic nuclei move and rearrange on a multidimensional potential energy surface (PES), on which the energy in each point is the electronic energy (plus the additive term of fixed nuclear repulsion), as determined by solving the electronic Schrödinger equation. Normally it is assumed that atomic movements can satisfactorily be described by classical Newtonian mechanics. Furthermore, the simplest models assume that reactions follow closely the minimum energy pathway (MEP) going from reactants to products through a TS. (The MEP expressed in mass-weighted Cartesian coordinates is referred to as the internal reaction coordinate, IRC). Studying only the MEP is, of course, not yet a full dynamics study. More importantly, recent theoretical investigations in Hase's group [43] and in Dupuis' group [44] have come to the remarkable observation that



Scheme 1. Tautomerization of cytosine by water-mediation from the amino-oxo to the imino-oxo isomer.

chemical reactions may follow a route totally *different* from the IRC!

A true dynamics calculation requires knowledge of the complete multidimensional PES. In very small dimensions the PES can be determined in advance, with dynamics done separately on this surface. This approach is of course impractical for larger systems and recent methods generate the surface "on the fly": at each point of the trajectory, energy and forces (negative first derivatives of the energy with respect to the nuclear coordinates) are determined by a QC method and the atomic displacements are calculated from these by classical mechanics. The most successful method has been developed by Car and Parrinello [45]. This method is computationally efficient because the electronic wave function – rather than being optimized at the trajectory points – is "propagated". As a consequence, the system is moving close to, but not exactly on the B–O surface.

In B–O dynamics, the electronic wave function is *fully optimized in each step* along the trajectory. Perhaps the first study of this type was reported by Field [46], and such studies are becoming increasingly viable; it should be realized, however, that tens of thousands of full QC computations are needed to obtain reasonable trajectories. This is the approach adopted here.

2. Computational details

All dynamics calculations were done by a private version of the PQS program system [47]. The dynamics part of the package uses direct Newtonian molecular dynamics in conjunction with the Verlet algorithm [48]. For accelerating SCF convergence, the recently suggested "Fock dynamics" method [49] was used: at each trajectory point, a starting Fock matrix is determined by extrapolation of the last few Fock matrices. No thermostat is used at present, the temperature serves only as an orientation value: the initial velocities are randomly set consistent with a kinetic energy of kT. Thus, if the dynamics run was started at the energy minimum, the kinetic energy will develop toward $kT/2$. Somewhat arbitrarily, we have run the simulations at $T \sim 350$ K. (In part by chance, we started the calculations at this temperature, and after having numerous trajectories completed, we stayed with that; room temperature would have been more sensible.) Overall translation and rotation are removed from the molecule before a dynamics run is started. Several quantum chemical methods can be combined with the dynamics program and we used here the density functional theory (DFT) method with the B3LYP potential [50,51]. To be able to run the huge number of required calculations, the small 3-21G basis set was used as a compromise. Some exploratory runs were done using larger basis sets, and also including electron correlation at the MP2 level. To avoid any built-in bias, trajectories were started at the equilibrium configuration. All internal degrees of freedom were allowed to vary. Individual runs differ in the random starting velocities.

3. Results and discussion

For isolated cytosine no tautomerization is expected due to the high barrier of 30–40 kcal mol^{-1} [41]. Still, just to check the behaviour of the system, we have run ~50 trajectories at extreme high temperatures. It was interesting to see, for example, that at ~2000 K (!) the amino group still stays stable, while the CH bonds and the ring start breaking up.

The role of water was studied by adding just one water molecule to the system. For the cytosine–water complex the geometry was optimized to obtain a *local* minimum which holds the water in the neighbourhood between the amino- and the C=O group. (In the absolute minimum the water is between the C=O and the N1–H group [36]). Starting with random velocities from this minimum, about 300 trajectories were run, each of 3000–5000 points of 1 fs steps, thus covering time ranges of 3–5 ps. The total number of wave function calculations, including forces, was about one million. (For one point along the trajectory the QC calculation takes about 1 min on a dual-processor (Athlon 2600+) PC and a cluster of 10 PCs was used.)

In typical cases only vibrations are seen along the trajectory. In one single case, however, the trajectory did lead to tautomerization! To show how this takes place, the critical steps of the tautomerization process are reproduced in snapshots in Fig. 2. It is interesting to observe that once the system is close to the TS, proton transfer occurs remarkably fast, in 15–20 fs (steps 2–6). (For comparison, the time period of an infrared X–H stretching vibration is ~10 fs.) Before the points shown in Fig. 2, there are only vibrations in the system; after tautomerization is completed, the imino form stays preserved. More details are available in form of moving pictures from the author.

Variation of the potential energy and selected atom-pair distances during the simulation are followed in Fig. 3. Note about the potential energy that, in independent calculations, we determined the transition state energy E_{TS} by standard methods (stationary point on the PES with one imaginary frequency). At this level of theory $E_{TS} = -468.7535$ a.u. and the graph was constructed with this value as the ground level. (The minimum energy is $E_e = -468.7627$ a.u., yielding a barrier of 5.9 kcal mol^{-1}; see below.) As seen in the graph, the system is above the TS all along and the success of hydrogen transfer depends on the *position of the water molecule*: it must be close enough to the relevant nitrogen atoms, and correctly *oriented*. In Fig. 3b, the two O–N distances indicate the overall distance of water from cytosine, showing quite regular oscillations. Incidentally, at about 1480 fs, both N_3–O_9 and N_8–O_9 are decreasing: the water molecule is approaching *both* nitrogen atoms. As a consequence, H_{15} of water is getting closer to N_3. Note, however, that no structural changes are taking place at this point yet; specifically, the *bonded* N_8–H_{13} and O_9–H_{15} stay stable. It happens first at ~1530 fs that both O–N distances have decreased to about 250 pm thus offering possibility for the formation

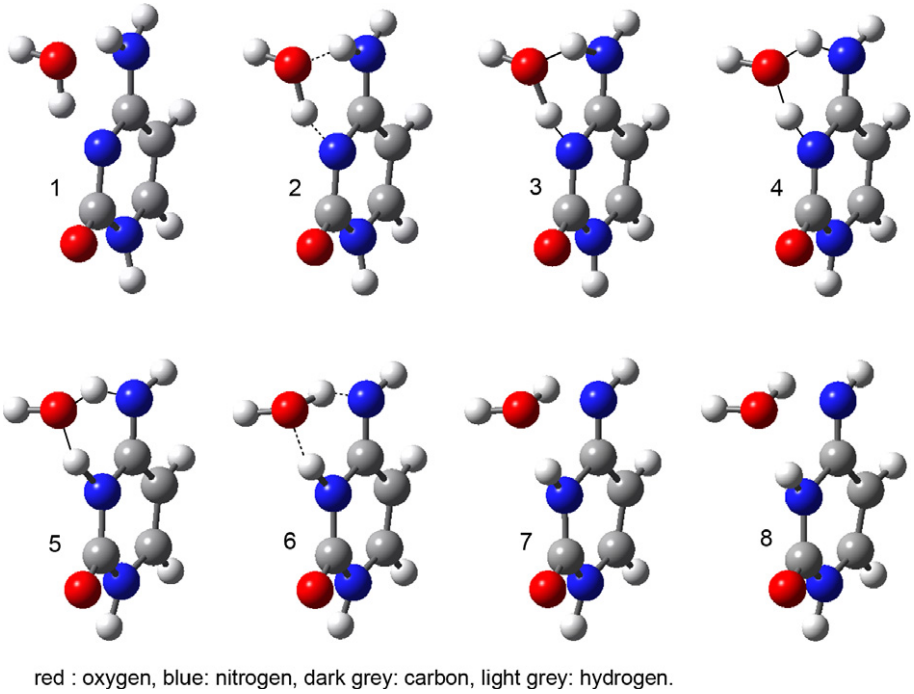


Fig. 2. Snapshots from the tautomerization trajectory. Total time span 5 ps, with resolution of 1 fs (5000 steps). Shown are steps from 1535 to 1570 fs taken from the original trajectory, every 5th geometry reproduced.

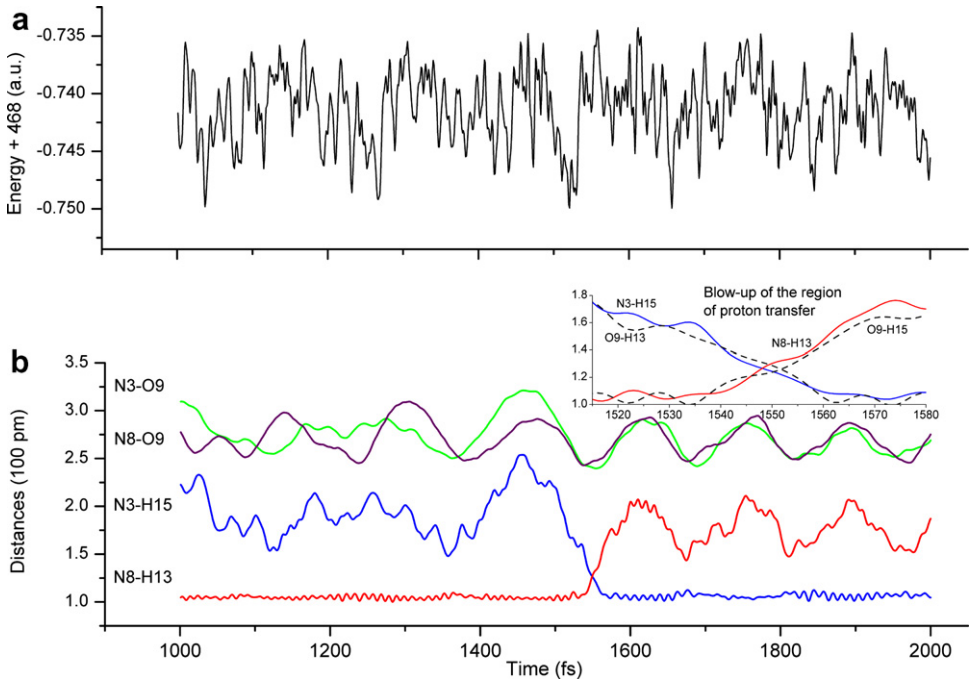


Fig. 3. Variation of (a) the potential energy, (b) selected bond distances along the tautomerization trajectory; 1 ps window.

of hydrogen bonds. The critical region is shown in more detail in the insert to Fig. 3b, including two additional O–H distances. Here, the stretching of the two relevant X–H bonds, O_9–H_{15} and N_8–H_{13} starts simultaneously. Within 30–40 fs, H_{15} has got attached to N_3 and H_{13} to O_9 and the proton transfer is completed. It is remarkable that the two hydrogen motions have been running parallel during the whole process, indicating strong coupling between the two.

To get more information about this coupling we have run two sets of *constrained* geometry optimizations: if O_9–H_{15} in water is stretched, with all other parameters optimized, N_8–H_{13} becomes stretched, and *vice versa* (Fig. 4), proving directly the coupling. By contrast, one

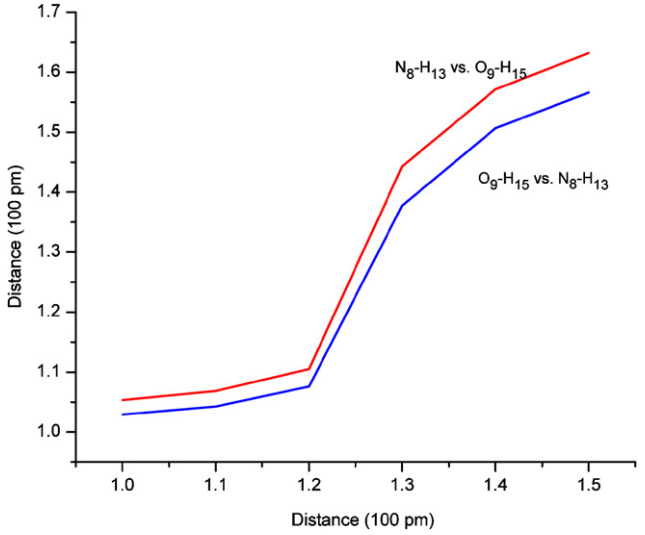


Fig. 4. The coupling between two hydrogen motions as indicated by constrained geometry optimizations (B3LYP/3-21G). Blue (lower curve): N_8–H_{13} in the amino group varied, all other parameters optimized. Red (upper curve): O_9–H_{15} of the water molecule varied, all others optimized. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

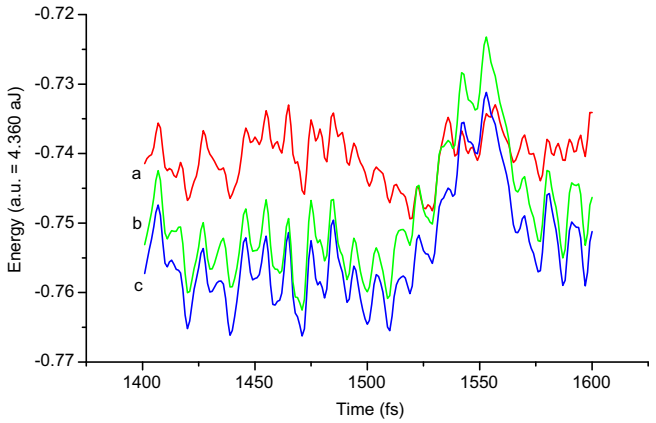


Fig. 5. Following the tautomerization trajectory at various levels of theory. To bring the three graphs together, energies are shifted. **a** (red) - the original B3LYP/3-21G results, E + 468; **b** (green) - B3LYP/6-31G**; E + 470.606; **c** (blue) - MP2/aug-cc-pVDZ, E + 469.416. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

could imagine a stepwise process with the intermediate formation of a protonated cytosine molecule. Obviously, however, this would require higher energy.

Because only one successful event was found we have no statistics for estimating the rate of the process. What we have seen is the *mechanism* of proton transfer, as it occurred in one single case. This mechanism can be easily reproduced, however, if one biases the system towards crossing the TS by stretching the relevant OH or NH bond in the starting configuration. We have run several such tests: tautomerization then occurs more frequently and confirms the above picture.

The cytosine–water complex may be one of the largest systems studied up to now by extensive ab initio dynamics but this could only be achieved by serious compromises. Besides limitations of the DFT method itself, the basis set is, of course, far too small. As a result, the barrier of ~6 kcal mol^{-1} is grossly underestimated; more accurate calculations all give a barrier of ~15–20 kcal mol^{-1} [19,41,42]. As part of the present study, we have also redetermined the transition state: at the level of CCSD(T)/cc-pVTZ//MP2/cc-pVTZ [52] the barrier is 18.8 kcal mol^{-1}.

We are fully aware that the compromises in the dynamics calculation are severe and may even question the reasonableness of the study. However, the following test gives some reassurance: 200 configurations have been taken from the tautomerization trajectory around the TS and the energy in each point was recalculated at higher levels of theory, using larger basis sets and including electron correlation in form of the second-order Moller–Plesset perturbation theory (MP2). Two of these results are compared with the original one in Fig. 5. The various plots, while being obviously different, agree in their basic features. This proves that the trajectory obtained in the lower-level simulation is basically correct and reflects a realistic mechanism of the proton transfer.

4. Conclusion

Tautomerization of cytosine, as mediated by one water molecule, has been studied by extensive Born–Oppenheimer dynamics calculations. In the model used, the potential surface is determined on the fly by ab initio (DFT) quantum chemistry, while the movement of the atoms is described classically. Of several hundred randomly started trajectories, one trajectory was found to lead to tautomerization. Details of the atomic movements show a *concerted* process, with no intermediate state. The primitive changes are the rupture of the N_8–H_{13} bond and the formation of N_3–H_{15}. These motions are strongly coupled so that the changes occur practically simultaneously, in a fast, *synchronous* process.

Acknowledgements

The author thanks Prof. Peter Pulay for a private version of the PQS program package. This work was supported by the National Scientific Research Foundation of Hungary (OTKA), Grants T043628 and K68427.

References

[1] J.D. Watson, F.H.C. Crick, Nature 171 (1953) 964–967.
[2] M.D. Topal, J.R. Fresco, Nature 263 (1976) 285–289.
[3] V.H. Harris, C.L. Smith, W.J. Cummins, A.L. Hamilton, H. Adams, M. Dickman, D.P. Hornby, D.M. Williams, J. Mol. Biol. 326 (2003) 1389–1401.
[4] T.-K. Ha, H.H. Gunthard, J. Mol. Struct. 300 (1993) 619–635.
[5] J. Sponer, P. Hobza, J. Phys. Chem. 98 (1994) 3161–3164.
[6] J. Sponer, P. Hobza, J. Am. Chem. Soc. 116 (1994) 709–714.

[7] I.R. Gould, N.A. Burton, R.J. Hall, I.H. Hillier, J. Mol. Struct. (Theochem.) 331 (1995) 147–154.
[8] C. Colominas, F.J. Luque, M. Orozco, J. Am. Chem. Soc. 118 (1996) 6811–6821.
[9] J. Florian, V. Baumruk, J. Leszczynski, J. Phys. Chem. 100 (1996) 5578–5589.
[10] J.S. Kwiatkowski, J. Leszczynski, J. Phys. Chem. 100 (1996) 941–953.
[11] P. Hobza, J. Sponer, M. Polasek, J. Am. Chem. Soc. 117 (1995) 792–798.
[12] J. Sponer, J. Leszczynski, P. Hobza, J. Comp. Chem. 17 (1996) 841–850.
[13] J. Florian, J. Leszczynski, B.G. Johnson, J. Mol. Struct. 349 (1995) 421–426.
[14] J. Florian, J. Leszczynski, S. Scheiner, Mol. Phys. 84 (1995) 469–480.
[15] G. Bakalarski, P. Grochowski, J.S. Kwiatkowski, B. Lesyng, J. Leszczynski, Chem. Phys. 204 (1996) 301–311.
[16] J. Florian, J. Leszczynski, J. Am. Chem. Soc. 118 (1996) 3010–3017.
[17] J. Florian, J. Leszczynski, Int. J. Quant. Chem. Suppl. 22 (1995) 207–225.
[18] G. Paglieri, G. Corongiu, D.A. Estrin, Int. J. Quant. Chem. 56 (1995) 615–625.
[19] A.L. Sobolewski, L. Adamowicz, J. Chem. Phys. 102 (1995) 5708–5718.
[20] M.P. Fulscher, B.O. Roos, J. Am. Chem. Soc. 117 (1995) 2089.
[21] J. Smets, L. Adamowicz, G. Maes, J. Phys. Chem. 100 (1996) 6434–6444.
[22] T.K. Ha, H.J. Keller, R. Gunde, H.H. Gunthard, J. Mol. Struct. 376 (1996) 375–397.
[23] R. Santamaria, A. Quirozgutierrez, C. Juarez, J. Mol. Struct. (Theochem.) 357 (1995) 161–170.
[24] G. Fogarasi, J. Mol. Struct. 413–414 (1997) 271–278.
[25] R. Kobayashi, J. Phys. Chem. 102 (1998) 10813–10817.
[26] S.A. Trygubenko, T.V. Bogdan, M. Rueda, M. Orosco, F.J. Luque, J. Sponer, P. Slavicek, P. Hobza, Phys. Chem. Chem. Phys. 4 (2002) 4192–4203.
[27] G. Fogarasi, J. Phys. Chem. A 106 (2002) 1381–1390.
[28] M. Piacenza, S. Grimme, J. Comput. Chem. 25 (2004) 83–98.
[29] E.D. Radchenko, G.G. Sheina, N.A. Smorygo, Yu.P. Blagoi, J. Mol. Struct. 116 (1984) 387–396.
[30] M. Szczesniak, K. Szczepaniak, J.S. Kwiatkowski, K. Kubulat, W.B. Person, J. Am. Chem. Soc. 110 (1988) 8319–8330.
[31] M.J. Nowak, L. Lapinski, J. Fulara, Spectrochim. Acta A 45A (1989) 229–242.
[32] R.D. Brown, P.D. Godfrey, D. McNaughton, A.P. Pierlot, J. Am. Chem. Soc. 111 (1989) 2308–2310.
[33] M.Y. Choi, F. Dong, R.E. Miller, Phil. Trans. R. Soc. A 363 (2005) 393–413.
[34] J.R. Sambrano, A.R. de Souza, J.J. Queralt, J. Andres, Chem. Phys. 253 (2000) 13–19.
[35] C. Aleman, Chem. Phys. 253 (2000) 13–19.
[36] G. Fogarasi, P.G. Szalay, Chem. Phys. Lett. 356 (2002) 383–390.
[37] C. Aleman, Chem. Phys. 244 (1999) 151–162.
[38] O.V. Shishkin, L. Gorb, J. Leszczynski, J. Phys. Chem. B 104 (2000) 5357–5361.
[39] E. Nir, K. Kleinermanns, M.S. de Vries, Nature 408 (2000) 949–951.
[40] L. Gorb, J. Leszczynski, Int. J. Quant. Chem. 70 (1988) 855–862.
[41] S. Morpurgo, M. Bossa, G.O. Morpurgo, Adv. Quant. Chem. 36 (2000) 169–183.
[42] Y. Podolyan, L. Gorb, J. Leszczynski, Int. J. Mol. Sci. 4 (2003) 410–421.
[43] X.L. Sun, K. Song, W.L. Hase, Science 296 (2002) 875–878.
[44] S.C. Ammal, H. Yamataka, M. Aida, M. Dupuis, Science 299 (2003) 1555–1557.
[45] R. Car, M. Parrinello, Phys. Rev. Lett. 55 (1985) 2471–2474.
[46] M.J. Field, J. Phys. Chem. 95 (1991) 5104–5108.
[47] PQS version 2.4, Parallel Quantum Solutions, 2013 Green Acres Road, Fayetteville, AR 72703. Private version from Peter Pulay.
[48] L. Verlet, Phys. Rev. 159 (1967) 98–103.
[49] P. Pulay, G. Fogarasi, Chem. Phys. Lett. 386 (2004) 272–278.
[50] A.D. Becke, J. Chem. Phys. 98 (1993) 5648–5652.
[51] C.T. Lee, W.T. Yang, R.G. Parr, Phys. Rev. B 37 (1988) 785–789.
[52] ACES II, Mainz-Austin-Budapest version, J.F. Stanton, J. Gauss, J.D. Watts, P.G. Szalay, R.J. Bartlett; with contributions from A.A. Auer, D.B. Bernholdt, O. Christiansen, M.E. Harding, M. Heckert, O. Heun, C. Huber, D. Jonsson, J. Jusélius, W.J. Lauderdale, T. Metzroth, C. Michauk, D.P. O'Neill, D.R. Price, K. Ruud, F. Schiffmann, A. Tajti, M.E. Varner, J. Vázquez and the integral packages MOLECULE (J. Almlöf and P.R. Taylor), PROPS (P.R. Taylor), and ABACUS (T. Helgaker, H.J. Aa. Jensen, P. Jorgensen, J. Olsen), see also http://www.aces2.de.