Journal of Molecular Structure 978 (2010) 257–262

Contents lists available at ScienceDirect

Journal of Molecular Structure

journal homepage: www.elsevier.com/locate/molstruc

# Studies on tautomerism: Benchmark quantum chemical calculations on formamide and formamidine

Geza Fogarasi *

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

ARTICLE INFO

*Article history:*
Received 21 December 2009
Received in revised form 23 February 2010
Accepted 23 February 2010
Available online 1 March 2010

*Keywords:*
Tautomerism
Proton transfer
Transition state barrier
Ab initio quantum chemistry
Coupled cluster

ABSTRACT

Tautomerism is a ubiquitous phenomenon in both chemistry and molecular biology and numerous computational studies of variable accuracy are available on it. In the present work, we report on some benchmark quantum chemical calculations on the formamide ↔ formamidic acid (**1**) and formamidine ↔ formamidine (**2**) systems. Some results on the acetaldehyde ↔ vinyl alcohol and acetaldimine ↔ vinyl amine pairs are also presented. A central question of the study is the role of water as a catalyst, investigated in form of the monohydrates of **1** and **2**. Optimized structures have been obtained for the tautomers and the transition states, including their monohydrates. The calculations represent the most comprehensive and highest level study up to now, with the goal of converging the relevant energy differences to an accuracy of 0.5 kcal/mol. To this aim, basis sets cover a range from simple 6-31G(d,p) up to 6-311++G(3df,3pd) and from cc-pVTZ up to aug-cc-pV5Z. The electron correlation treatment has been varied from RHF and DFT up to CCSD(T), plus one case with CCSDT (true triples). The convergence of the results with respect to the computational level is, unfortunately, different. The tautomerization energies of the *free* molecules converge at CCSD/aug-cc-pVTZ or /cc-pVQZ, the triples being not necessarily required. However, in the *general* case (hydrates and transition states) one needs the triples, and the basis set should be of quadruple-zeta quality, to achieve really strict convergence.

© 2010 Elsevier B.V. All rights reserved.

## 1. Introduction

Tautomerism is a special form of isomerism in which a proton is relocated within the molecule, accompanied by appropriate changes in the bonding structure. The process plays an important role in numerous chemical and biological systems. Specifically, one of the possible mechanisms of mutation of DNA may be the tautomerization of nucleotide bases [1].

Tautomerization in various systems has been the subject of numerous quantum chemical (QC) studies. Still, it is difficult to get a clear picture about the *accuracy* of the calculations. This may be a crucial point if the relative energies of tautomers are in a narrow range of a few kcal/mol: for example, the three main tautomers of cytosine differ in energy by 2–3 kcal/mol only [2]. In fact, the present study was largely motivated by the uncertainties concerning cytosine. In cases like that a really good description of relative energies should be accurate within ~0.5 kcal/mol. With this in mind, we want to perform benchmark calculations to check the ultimate accuracy achievable with present day QC methods. Four simple systems, shown in Fig. 1, have been selected for the tests: formamide (FMD) ↔ formamidic acid (FAC), **1**; formamidine (FIM) ↔ formamidine, **2** (in the latter, the two tautomers are identical, the question is the transition state); acetaldehyde ↔ vinylalcohol, **3**; and acetaldimine ↔ vinylamine, **4**. These pairs represent four fundamental cases of tautomerism: amide ↔ imidic acid, amine ↔ imine, keto ↔ enol and imine ↔ enamine. Accordingly, a large number of previous studies are available for comparison, mainly for formamide [3–23] but also for the other, less ubiquitous systems [24–30]. For the pairs in **3** and **4**, tautomerization cannot be expected, the isomers exist as independent molecules. Still, the isomerization energy in these systems can add information about the consistency of the computational methods used. An important part of the present study will be the investigation of the role of water. This will be done on the two smaller systems, in form of monohydrates: first, the change in tautomerization energies will be determined, and then the effect of water on the transition barrier will be studied.

## 2. Computational details

The level of QC calculations was varied systematically, from Hartree–Fock (RHF) through Moller–Plesset 2nd order perturbation (MP2) theory up to Coupled Cluster (CC) theory in the CCSD(T) approximation. (In one single case, even CCSDT – with true triples – was tested.) Beside traditional wave function methods, density

---

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

0022-2860/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.molstruc.2010.02.065

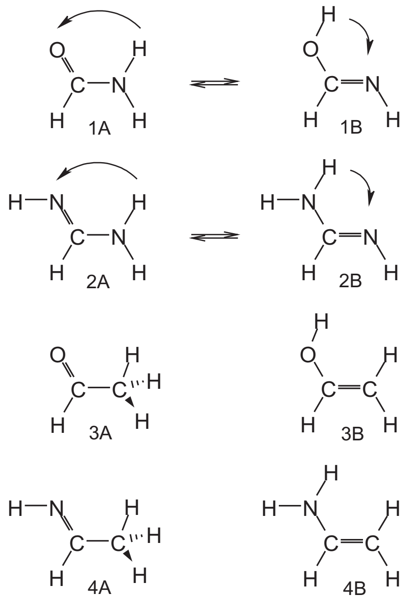

**Fig. 1.** The four tautomer pairs investigated: **1A** – formamide, **1B** – formamidic acid, **2A** – formamidine, **2B** – formamidine (the two tautomer forms are identical), **3A** – acetaldehyde, **3B** – vinylalcohol, **4A** – acetaldimine, **4B** – vinylamine.

functional theory (DFT) was also examined for comparison, applying the B3LYP functional. The quality of basis sets ranged from 6-31G(d,p) to aug-cc-pV5Z. (For the above standard acronyms, see, e.g. [31]). Pople-type sets were used in the lower level, RHF and DFT calculations, while Dunning's correlation consistent (cc) basis sets were chosen for the CC calculations. In between, in the MP2 calculations both types of basis sets were tested. Up to medium levels of theory, optimum geometries and the corresponding energies were calculated at the same level. For the largest calculations, the MP2/aug-cc-pVTZ optimized structures were adopted as our standard geometries and only energies were computed at higher levels. MP2 was used with frozen core. Vibrational frequencies were calculated for the transition states to check the nature of a stationary point. However, frequencies were in general not computed for the energy minima, so ZPE corrections will not be included in the energies (except for one case to check a literature result). This is because our main interest is just in the *internal* accuracy of the methods, and experimental energies are, of course, not available for comparison, anyway. Two computer program packages were used: PQS [32] for all RHF, B3LYP and MP2, and CFOUR [33] for the CC calculations.

## 3. Relative energies of free tautomer pairs

Results for three tautomer pairs are listed in Table 1. In very broad terms, the general picture is the same at all levels of theory: the keto form and its imine analogue is the more stable tautomer within a pair in all calculations and quantitatively, the energy differences do not change more than a few kcal/mol with the method. It is notable that even the simplest RHF method performs well in this respect.

However, as pointed out above, we are interested in an accuracy better than 0.5 kcal/mol. Therefore, in the discussion below, convergence will be considered safe if the last changes do not exceed 0.2 kcal/mol. We discuss first the role of basis sets. In the LCAO-MO-type RHF and B3LYP calculations, as seen in Table 1, convergence within the group of Pople-type basis sets has likely been achieved for system **1**, while systems **3** and **4** seem a little more sensitive to the quality of basis set. About the larger, correlation consistent (cc) basis sets we have information in the group of MP2 results as, of course, the SCF wave function is a necessary first step in these calculations. (The geometry is fixed in these calculations – see the footnotes to Table 1 – but the changes in geometry are already negligible energetically at this point.) Within the MP2 calculations the RHF energies (not listed in Table 1) show already perfect convergence: for system **1**, the maximum change in $\Delta E$ is only 0.02(!) kcal/mol from the cc-pVTZ basis set up to aug-cc-pV5Z; for the other two systems the corresponding value is larger but still below 0.2 kcal/mol. Note that the MP2 energy itself is more sensitive to the basis set in **3** and **4** than in **1**, see next below. So, from the one-electron point of view, the aug-cc-pVTZ basis set can be considered as complete.

The size of a basis set plays, of course, another and more significant role in correlation calculations, as the volume of the configuration space increases rapidly with increasing basis sets. This can be seen explicitly in the MP2 results on **3** and **4**: in Table 1, when going from aug-cc-pVTZ to (aug)-cc-pV5Z the MP2 energies change by 0.5 kcal/mol in both systems; at the same time, the corresponding SCF energies change only 0.16 and 0.17 kcal/mol for **3** and **4**, respectively (see also above). Thus, in the total change of 0.5 kcal/mol about one third comes from the incompleteness of the one-electron basis while the role of the configuration space expansion is twice larger.

Electron correlation effects in general can be judged by comparing the perturbation theory results with the higher level coupled cluster results. For **1**, the best MP2 result is $\Delta E = 11.5–11.6$ kcal/mol, which should be compared with the best CCSD and CCSD(T) values of 10.6–10.7 kcal/mol (the latter seem to have converged, being stable within 0.2 kcal/mol). Thus, electron correlation beyond MP2 has an effect of about 1 kcal/mol. In the treatment of correlation one goes normally up to the CCSD(T) level, where the triple substitutions (excitations) are only approximated (by fourth-order perturbation theory). System **1** is the smallest one among those investigated here and we were able to check even the effect of *true* triples. As seen in Table 1, the CCSDT/cc-pVTZ result of 10.6 kcal/mol confirms the CCSD(T) approximation.

As already seen above, the acetaldehyde ↔ vinylalcohol (**3**) and acetaldimine ↔ vinylamine (**4**) systems are more sensitive to the computational method. For **3**, the best CCSD(T) result is $\Delta E = 9.3$ kcal/mol and seems very well converged with basis sets. For **4**, the aug-cc-pV5Z calculations would have been too expensive because of the lack of any symmetry (non-planar amino group). Nevertheless, the last two points in Table 1, obtained with basis sets aug-cc-pVQZ and cc-pV5Z suggest good convergence: they are listed with one decimal as 3.0–3.1 kcal/mol, but are in fact even closer, the more precise difference being 0.06 kcal/mol.

It is important to check the importance of triple substitutions. Comparing the CCSD and CCSD(T) results, it is reassuring to see that for the purpose of the present study, interested in *relative* energies within a tautomer pair, they are very close to each other: for **1** there is a difference of 0.1 kcal/mol only, for **3**, the difference is 0.2 kcal/mol, both very systematic. For **4** the two sets of data agree within 0.1 kcal/mol. Thus, one may easily jump to the conclusion that in analogous future calculations a significant part of computer time could be saved by neglecting triple excitations and using just the CCSD method. See, however, the hydrates below.

Finally, we discuss the DFT results. According to Table 1, DFT (with the popular and well established B3LYP functional) performs well, but seems less consistent than MP2. For system **1**, B3LYP is – somewhat surprizingly – practically the same as RHF, and with $\Delta E = 12.2–12.3$ kcal/mol *over*estimates the best CC-result by one and a half kcal/mol. At the same time, for system **3** DFT brings a significant improvement over RHF, with the result of $\Delta E = \sim 9.6$ kcal/mol (may be not quite converged with basis set) being close to the CC-result of 9.3 kcal/mol. Then, however, for system **4** DFT

Download English Version:

https://daneshyari.com/en/article/1409864

Download Persian Version:

https://daneshyari.com/article/1409864

Daneshyari.com