Journal of Molecular Graphics and Modelling 29 (2010) 246–255

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

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

Hierarchical approach to conformational search and selection of computational method in modeling the mechanism of ester ammonolysis

Miroslav A. Rangelov[a,*], Galina P. Petrova[b], Vihra M. Yomtova[a], Georgi N. Vayssilov[b,*]

[a] Laboratory of BioCatalysis, Institute of Organic Chemistry, Bulgarian Academy of Sciences, Sofia 1113, Bulgaria
[b] Faculty of Chemistry, University of Sofia "St. Kliment Ohridski"; 1, James Bourchier Ave., Sofia 1126, Bulgaria

ARTICLE INFO

Article history:
Received 24 April 2010
Received in revised form 17 July 2010
Accepted 20 July 2010
Available online 30 July 2010

Keywords:
Reaction mechanisms
Computational chemistry
Conformational search
Transition states

ABSTRACT

We describe automated procedures for the first stages of a systematic computational investigation of reaction mechanisms. They include (i) selection of computational method and basis set based on statistical analysis of structural and energy data relating to experimental values, (ii) determination of all distinct conformations of transition states with large conformational freedom, and (iii) generation of unknown geometry of the transition states, based on pre-defined connectivity of the atoms involved in the reaction. For the conformational search we employed an efficient procedure for exploration of various possible conformations of the transition states and elimination of the equivalent structures in several steps using molecular-mechanical and quantum-mechanical methods. The procedure was applied to the determination of the structures of transition states and intermediates in the ammonolysis of monoformylated 1,2-ethanediol, which were subsequently used for identification of the lowest energy reaction paths. For the same reaction system we also used the approach for generation of the initial structures of transition states with unknown geometry. The reported procedures are implemented in the MolRan program suite.

1. Introduction

The computational modeling of reaction mechanisms and energetics has become an important step in the investigation of different processes, occurring in biological systems [1–4], catalytic systems [5–8], organic synthesis [9–11], etc. The selection of the computational method [12] and construction of the initial structures of the reactants, transition states, intermediates and products are often carried out after comparison with previous knowledge and chemical intuition. This approach becomes, however, inappropriate when the rigorous study of the reaction mechanism requires calculation of a large number of structures of the stationary points because of the size or flexibility of the modeled system. Here, we propose automated procedures for these first stages of a systematic computational investigation of reaction mechanisms that provide both computational efficiency and reliability.

The initial step in a computational modeling is the selection of the computational method and basis set that are appropriate for the explored chemical system. In order to determine which method is reliable and at the same time computationally efficient, we performed an exhaustive test of different computational method/basis set combinations and statistical analysis of the obtained structural and energetic data, which were compared with experimental values (Section 3).

Each rigorous theoretical investigation requires systematic inspection of all possible transition state structures and intermediates, belonging to different reaction paths, and identification of the structures corresponding to the lowest barriers in order to define the most energetically favorable reaction route. This search is particularly complicated in systems with large conformational freedom where it is important to check various possible conformations of the transition states and intermediates, in order to select among them a set of distinct structures with low energy, to be computed at the final stage by a chemically adequate (but more time consuming) computational method. Hereby, we describe a smooth and automated procedure for conformational search of transition state structures. The procedure is applied to the ammonolysis of 1-O-formyl 1,2-ethanediol (Section 4.1) and it helped us to find all lowest energy structures of the stationary points along different reaction paths. In addition, in Section 4.2 we report an approach to the automatic construction of the initial structures of transition states with desired atomic connectivity but unknown geometry using the same model reaction. We selected this reaction since it is a suitable model for elucidation of the mechanism of biosynthesis of peptide bonds in the ribosome, in connection with recent theoretical [13–17] and experimental [18–20] investigations.

* Corresponding authors.
E-mail addresses: mran@orgchm.bas.bg (M.A. Rangelov), gnv@chem.uni-sofia.bg (G.N. Vayssilov).

1093-3263/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2010.07.006

All described procedures are implemented and performed with the program MolRan [21] (see Section 2.1 for more details).

2. Computational details

2.1. Method

All quantum-chemical calculations described in the current work were performed with the program package Gaussian03 [22]. The search of appropriate level of theory for modeling the studied processes included different methods—HF, MP2 [23,24], B3LYP [25–29], MP4(SDQ) [30,31], and CCD [32]. The basis sets were varied from 3-21G [33] to 6-311G [34] as polarization and diffuse functions were also considered.

Preliminary optimization of the structures was performed at molecular-mechanical level with a modified version of MM force field [35,36] as the calculations were carried out with the MM optimizer module within MolRan program [21] by applying a combined conjugate-gradient quasi-Newton minimization algorithm [37,38]. Since in many processes the structures with inter- or intramolecular hydrogen bonds are of special interest, the program allows the addition of a small exponential term in the MM field in order to account for the contribution of hydrogen bonds to the energy of the system in a more precise way [39]. This term depends on the distances between the atoms in the proton-donor group and between the proton and the potential proton-acceptor, which are calculated at each step of the relaxation procedure.

The conformational search of the transition states is performed with the help of different modules implemented in the MolRan program suite [21].

2.2. MolRan program

MolRan [21] is a graphical tool for visualization and generation of molecular geometry and analysis of the chemical properties and electronic structure information obtained by means of quantum-chemical calculations with specialized computational software, e.g., Gaussian03 [22]. Various pre-defined z-matrix, Cartesian and fractional file formats (e.g., xyz, pdb, cif) are supported as in addition a fully customized file reader can extract and present structural data from text files, i.e., output files generated by computing software. Various graphical representation tools are implemented which allow the user to visualize the Hessian, gradients and other vectors, Newman projections along selected atoms, MD trajectories, RMS (root mean square) structures, etc.

The program also includes a set of tools for generating, loading, and displaying meshes generated on the base of different data, e.g., Van der Waals spheres, solvent accessible surfaces, meshes from loaded orbital coefficients and primitives producing AO or MO surface. In the last case the mesh could be also loaded from Gaussian03 cube file. The implemented tool allows mathematical modification of the generated/visualized meshes.

A system of tools is available for constructing job files for conformational analysis (see Section 4) and for analysis of the results obtained. The implemented file constructor can generate multiple job file with linear and nonlinear structures varied by selected structural parameters (interatomic distances, valence angles or dihedral angles) as the chosen variables can be altered synchronously or independently.

3. Selection of computational method and basis set

One of the important decisions in each quantum-chemical study is the selection of the method and basis set appropriate for simulation of the chosen system. In order to choose the theory level for our modeling, we considered a simple model system—formic acid and its dimer, which allows comparison of the theoretically obtained results with the available experimental data for the structure in the gas phase of both species [40,41]. The Gibbs free energy of the reaction of ammonolysis of formic acid in the gas phase was also estimated and compared to the experimental data [42]. In addition to the availability of experimental and theoretical [43–45] data for this process, another reason for choosing this model for testing different combinations of methods and basis sets is that the interaction between formic acid and ammonia is the minimal model for the aminolysis reaction, related to peptide bond formation.

The calculations consisted of geometry optimization of the structures at each combination of method and basis set (Section 3.1), followed by computation of the matrix of the analytical force constants and determination of the thermochemical characteristics of the system (Section 3.2). The calculations were performed at 60 different levels representing a combination of method and basis set as the applied quantum-chemical methods were HF, MP2, B3LYP, MP4(SDQ), and CCD. All obtained data were statistically scrutinized by means of cluster and canonical analyses. The hierarchal complete-linkage approach was applied for the cluster analysis as the data were standardized and Euclidean distance was used as similarity measure.

3.1. Structural parameters

The dendrogram obtained from the analysis of the errors in the geometry characteristics of formic acid and its dimer is presented in Fig. 1a. The bold line in the dendrogram (Fig. 1a; at distance 3.5) corresponds to the distance defined by the experimental error. As can be seen, the experimental data are connected with two clusters, {**4**} and {**5**} (see Fig. 1a), containing mainly data at B3LYP, MP2, and MP4 levels with basis sets with polarization functions as the main difference between the two clusters is the presence (cluster {**5**}) or absence (cluster {**4**}) of diffuse functions. The data included in the other three clusters differ more significantly from the experimental data. Cluster {**2**} consists mainly of methods which take into account the electron correlation, combined with basis sets without polarization functions or with basis sets including polarization functions but with a few primitives (3–21). The other two clusters contain the HF method with or without polarization function in the basis sets (clusters {**1**} and {**3**}, respectively). Thus, if we consider the computational time necessary for modeling the system at the different levels of theory, as well as the necessity of polarization function at least for the heavy atoms, the optimal method for the correct description of the geometry of the modeled structures and eventually formed hydrogen bonds should be B3LYP with basis set at least 6-31G*.

In order to quantify the influence of the different factors (i.e., method, basis set, presence or absence of polarization and diffuse functions), we also performed a canonical analysis. The canonical correlations of the elements of the computational procedure compared to their error relating to the experimental results are provided in Table 1. The negative values of the correlation coefficients result from the fact that the error is correlated. As can be seen, the factor with strongest influence on the accuracy in the simulation of the single bonds length is the presence of polarization functions at the heavy atoms. This factor is also important in the case of the double bonds, however for the correct simulation of the double bonds the computational method is more important. For the accurate description of the hydrogen bonds, the presence of polarization functions for the H atoms forming the bond is essential, while the inclusion of diffuse functions for these H atoms plays a modest role, even smaller with respect to the diffuse functions for the heavy atoms. The overall size of the basis set is important for the correct simulation of the valence angles in the modeled structures.

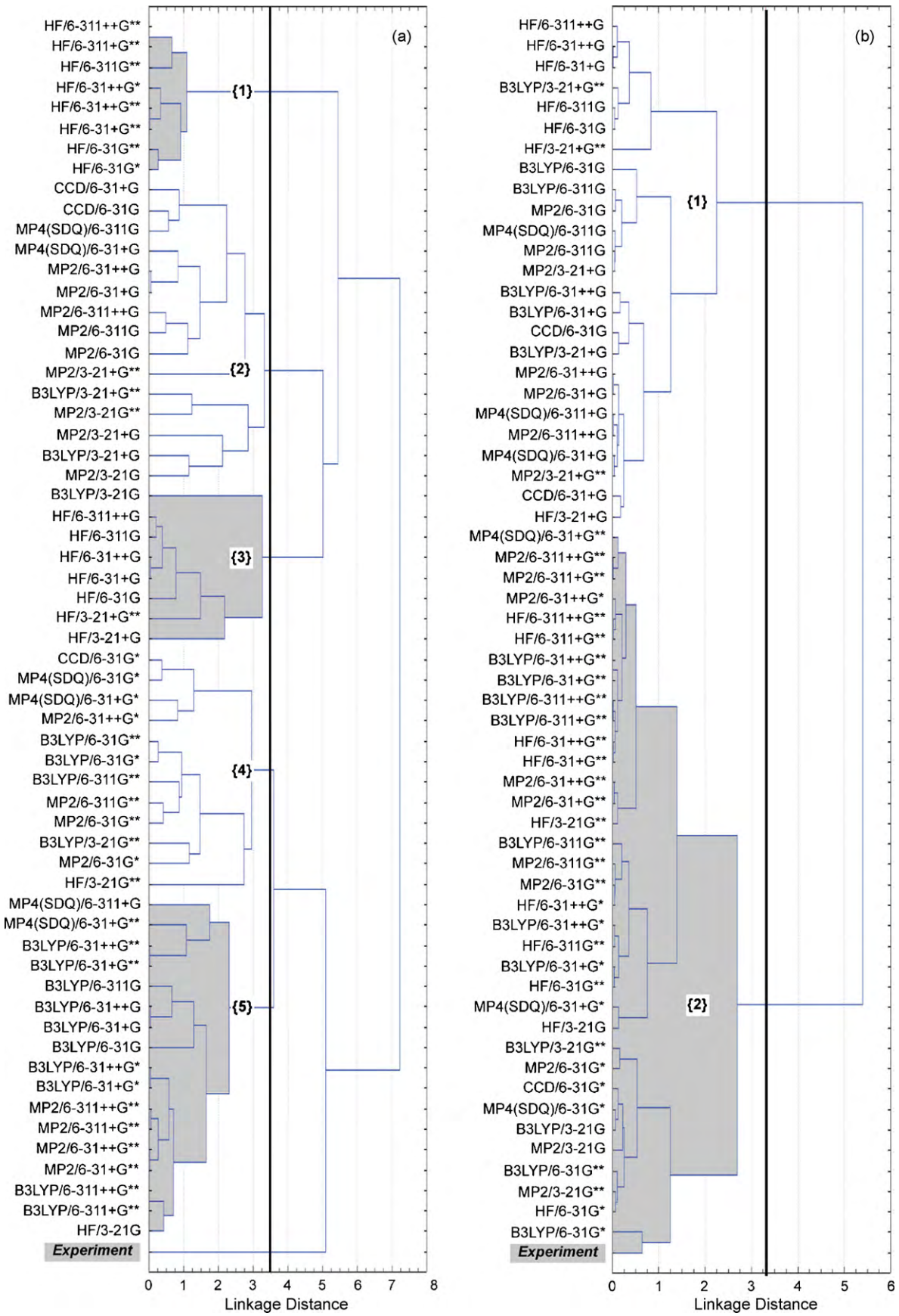


Fig. 1. Dendrogram obtained by cluster analysis of the computational data at the different tested levels of theory: (a) the clustering of the available data is performed on the base of the errors in the description of the geometry of formic acid and its dimer; (b) the clustering of the available data is performed on the base of the errors in the description of the thermochemistry of formic acid ammonolysis.

Table 1
Correlation coefficients between the computational procedure and the error relating to the experimental data. The numbers in bold show the characteristic of the computational procedure that affects the most strongly the selected structural parameter.

	Single bonds	Double bonds	Hydrogen bonds	Angles
Computational method	0.36	**−0.41**	−0.25	−0.08
Basis set	−0.25	−0.21	−0.13	**−0.35**
Diffuse functions (heavy atoms)	−0.06	−0.21	−0.21	−0.26
Diffuse functions (H atoms)	−0.20	−0.14	−0.14	−0.21
Polarization functions (heavy atoms)	**−0.61**	−0.37	−0.32	−0.13
Polarization functions (H atoms)	−0.36	−0.19	**−0.53**	−0.11

3.2. Thermochemistry

The dendrogram based on the errors in the description of the thermochemistry of formic acid ammonolysis is presented in Fig. 1b (the bold line at distance about 3.3 corresponds to the distance defined by the experimental error). As can be seen in the figure, the methods and basis sets could be divided into two groups as the main difference between the two groups is again the presence or absence of polarization functions in the basis set (cluster {**2**} and cluster {**1**}, respectively). Since cluster {**2**}, defined by the basis sets with polarization functions, include the experimental data, it is necessary to consider basis sets with polarization functions for the correct description of the thermochemistry of the modeled reaction. Thus, the method B3LYP with basis set at least 6-31G**, proposed on the base of the previous cluster analysis of the geometry data, should provide results comparable with the experimental data.

Due to the necessity to compare the energies between reactions with neutral molecules and anions requiring the presence of diffuse functions in the basis set, we chose to perform our computational study at B3LYP/6-31 + G* level of theory. The structures optimized at B3LYP/6-31 + G* and B3LYP/6-31 + + G** levels were analyzed for geometric similarities by means of the procedure described in Section 4 and, in consistence with the cluster analysis (Fig. 1), they were found to be similar since the RMS differences of the interatomic distances between equivalent atoms of the equivalent structures optimized at the two levels of theory were estimated at 0.025 Å.

In summary of this section, the density functional approach with the hybrid functional B3LYP was found the best compromise allowing accuracy of the calculated structural parameters and reaction energies for the modeled type of systems at reasonable computational time. One should note, however, that this may not be the case for other systems, for example for systems, in which the dispersion interactions plays a crucial role [46]. For such more "difficult" systems and problems one could employ new types of density functionals [12,47].

4. Search of transition state structures

The formation of the peptide bond is the key reaction step in protein synthesis which in living cells is catalyzed by the ribosome [48]. During protein synthesis, the amino group of the amino acid in the ribosome's A-site plays the role of the nucleophile, while the carbonyl belongs to the P-site amino acid, esterified to the 3′-ribose of tRNA. Thus, the reaction is actually an aminolysis of esters, which is often considered as a model reaction for the formation of peptide bonds during translation in both experimental and theoretical investigations [13–18,43]. Despite the numerous investigations of the reaction [18,49–51], its mechanism in the ribosome remains unresolved. For our model study we investigated the interaction between monoformylated ethanediol and ammonia. This selection of the reagents is not accidental but determined by the presence in the formylethanediol molecule of a vicinal hydroxyl next to the reaction center analogously to the ribose ring in tRNA. As shown in our theoretical investigations of the mechanism for this model reaction (Scheme 1a and b) [17,52], the latter one based on the transition state search reported here, the catalytic role of this vicinal hydroxyl group could be explained by the stabilization of the transition state structures via formation of intramolecular hydrogen bonds (tetragonal transition states in Scheme 1a) or by a direct participation of this vicinal hydroxyl in the proton transfer process (hexagonal transition states in Scheme 1b). Both possible reaction paths are presented in Scheme 1a and b, while the applied notation of the atoms is presented in Scheme 1c on the example of the tetrahedral intermediate I.

In the case of the reaction path through tetragonal transition states the geometry of the reaction ring is well known from previous theoretical data [43] and the search of transition states was oriented towards defining their most stable conformers by means of the conformational search algorithm described in Section 4.1. On the other hand, the geometry of the reaction ring for the proposed hexagonal transition states was unknown and the structural search consisted of generation of relevant initial structures based on pre-defined bonding between atoms from the reaction ring (see Section 4.2). In Section 4.3 we summarize the calculated results for the different modeled types of transition states and analyze the preference for the alternative reaction paths.

4.1. Conformational search algorithm for the tetragonal transition states

The optimized structures of the tetragonal transition states for the stepwise and concerted mechanisms are presented in Fig. 2 as the first structure in each line corresponds to a system without intramolecular hydrogen bond, while the second one represents the most stable conformer with hydrogen bond found by means of the proposed procedure. The applied search algorithm is schematically presented in Scheme 2 and consists of the following steps:

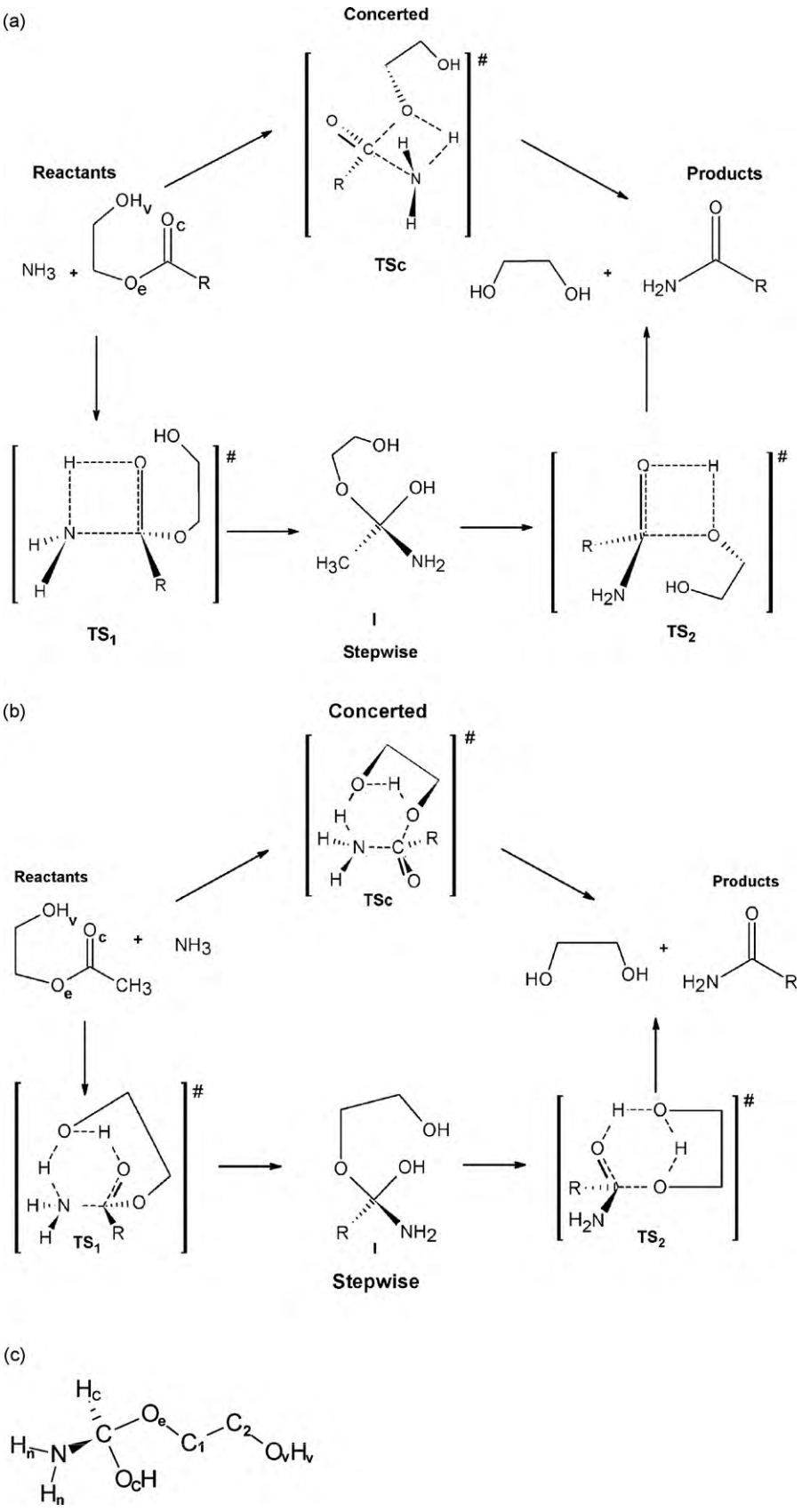
1. Generation of transition state initial structures.

In the example of ammonolysis of formylethanediol at this first stage all structures in completely periplanar conformation for the concerted and the stepwise mechanism were optimized at HF/6-31G level. The obtained structures of the tetragonal transition states and the intermediates correspond to those described in the literature for the ammonolysis of methylformiate [43,53–55].

2. Generation of series of modified structures.

These structures were based on the structures of transition states and intermediates obtained in the first step. The new sets of geometries were generated by stepwise variation of all dihedral angles not involved in the tetragonal reaction ring by 10°, i.e., 36 positions were initially considered, while the positions of the atoms participating in the reaction ring were kept fixed. The generation of the different structures is performed with the module for conformational search implemented in the program MolRan [21].

For the transition states and intermediates in the ammonolysis of formylethanediol for TSc and TS2 transition states the angles



Scheme 1. Mechanism of ammonolysis of 1-O-formyl 1,2-ethanediol (R≡H) via tetragonal (a) and hexagonal (b) transition states, and (c) notation of the atoms on the example of intermediate I. In panels (a) and (b) the upper reaction route corresponds to the concerted mechanism, while the lower one represents the stepwise mechanism.

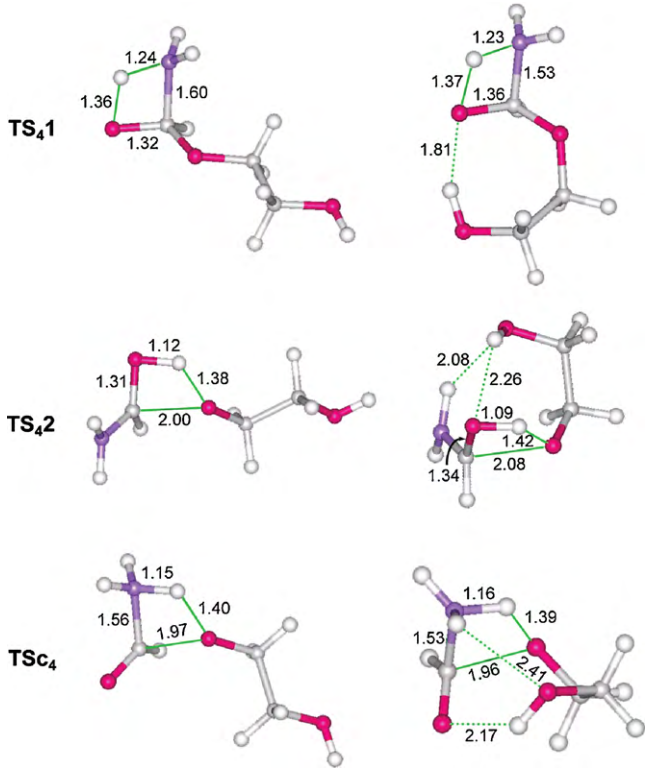
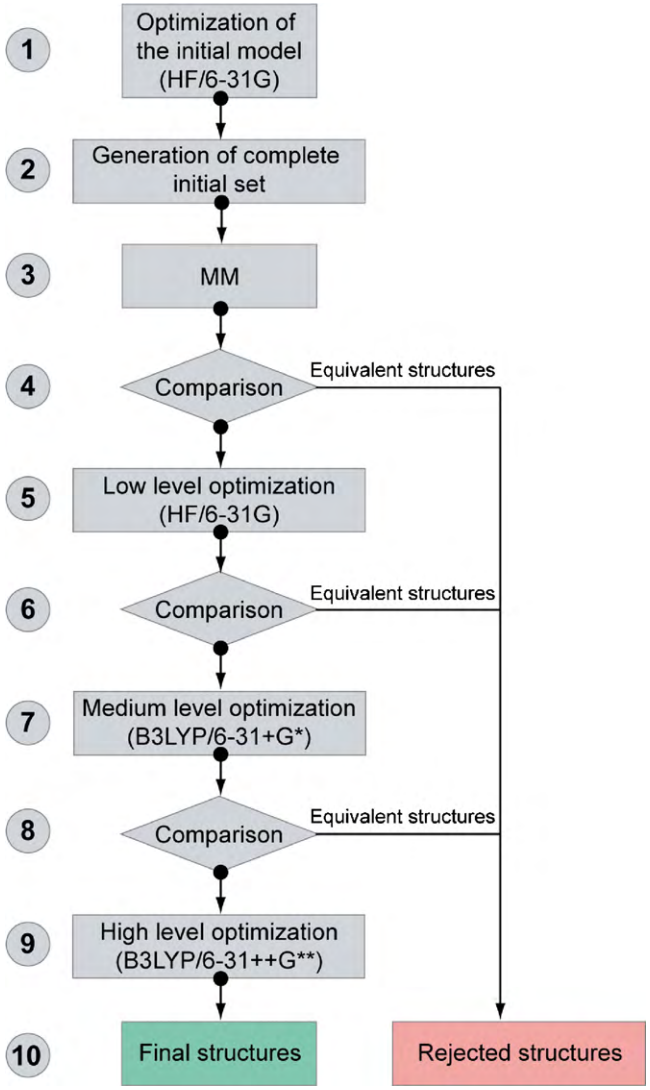


Fig. 2. Optimized structures of the tetragonal transition states of the stepwise ($TS_4$1 and $TS_4$2) and of the concerted (TSc_4) mechanism without (left-hand side) and with intramolecular (right-hand side) hydrogen bond. The bond lengths and some important interatomic distances are given in Å.



Scheme 2. Conformational search algorithm applied for defining the final set of structures in the case of formylethanediol model.

C–O_e–C1–C2, O_e–C1–C2–O_v and C1–C2–O_v–H_v were varied while for TS1 and intermediate I the dihedral angle N–C–O_e–C1 was also included (for the used notation see Scheme 1c). As could be seen from Scheme 1a, the chosen dihedral angles do not participate in the reaction ring. By means of this methodology we prepared ca. 46 000 initial structures for the transition states TSc and TS2 and ca. 160 000 for TS1 and intermediate I due to additional variation of the fourth angle N–C–O_e–C1 with the participation of the nitrogen atom from the attacking nucleophile.

3. Relaxation of the obtained structures at molecular-mechanical (MM) level.

This optimization is performed with the program MolRan [21] with conjugated gradient optimization procedure using analytical first derivatives. As described in Section 2, for this preliminary optimization we used modified version of MM force field supplemented by an exponential term for description of hydrogen bonds. During the geometry optimization of the structures at MM level the atoms from the cleaving/forming bonds are kept fixed with respect to each other.

4. Removal of equivalent structures.

After the MM relaxation of the structures, the equivalent ones were removed from the modeled series by calculating the root mean square deviation (RMSD) of the distances between equivalent atoms of each pair of structures. This procedure is performed as one of the structures is translated at the center of the coordinate system while the other structure of the pair is set over it by completing one translation and one rotation. The geometrical similarity of the structures is further estimated by numerical optimization of the geometry position of one molecule with respect to the other one by conjugated gradient quasi-Newton method. The gradients are calculated by means of Ridder polynomial extrapolation method and the RMSD distance between equivalent atoms from the two molecules in the pair serves as minimization function. The procedure applied for the elimination of the equivalent structures as described here is implemented in the program MolRan [21]. This step of the procedure allows reduction of the number of structures to be treated at higher computational level.

In tracing the reaction path of formylethanediol we also excluded the structures which do not contain any intramolecular hydrogen bonds except for the structures with O atoms of the ethanediol in *trans* position, which we used as referent structures.

5. Optimization at higher computational level.

The nonidentical structures of transition states and intermediates remaining from the previous step were optimized, respectively as saddle points and minima, at the next computational level (HF/6-31G) without any geometry restriction. All transition state structures were confirmed by calculating the matrixes of the analytical force constants and checking of the transition vector components. The optimized structures were again subjected to the elimination procedure described in step 4.

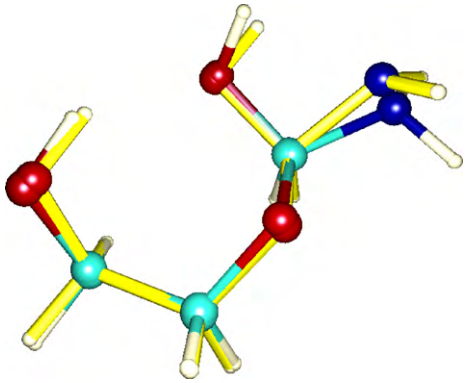


Fig. 3. Graphical result of the procedure for search of an intermediate structure which geometry is closest to a defined transition state structure. The transition state structure, which is varied relatively to the position and the transition vector, is presented with yellow bonds between the atoms.

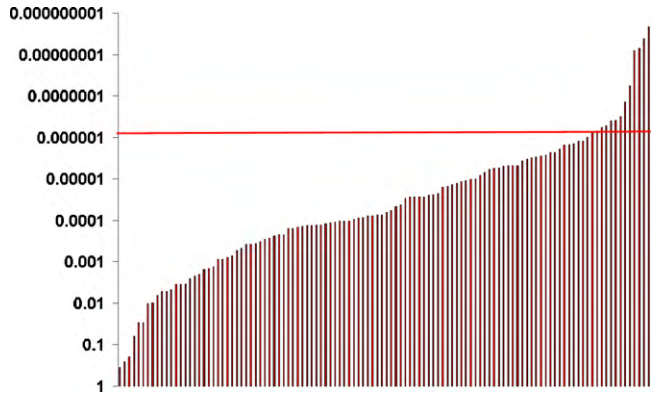


Fig. 4. Population of the conformers with intramolecular hydrogen bonds of the intermediate in the stepwise mechanism of formylethanediol ammonolysis relating to the most stable intermediate structure. The population of the referent structure of the intermediate without hydrogen bond is presented with a horizontal line.

The following steps (see Scheme 2, steps 6–10) include successive optimization of the obtained structures and further elimination of the equivalent ones at medium and higher level of theory. In the particular example the medium optimization level was B3LYP/6-31+G* level and the higher level for final optimization of the remaining structures was B3LYP/6-31++G** with additional polarization and diffuse functions at H atoms.

Due to the complexity of the transition vector of the modeled transition structures and the variation of the atoms, over which the main components of the vector are positioned along the reaction coordinate, the IRC search cannot reach the stable structures. For this reason, in order to find the intermediates, to which the different transition structures of TS1 will lead, and the intermediates, from which the transition structures of TS2 could be obtained, a technique similar to the procedure applied for elimination of the equivalent structures described above is used (see Fig. 3). In the current procedure each couple of intermediate and transition state structures are positioned over each other as the structure of the intermediate is translated into the centre of the coordination system and the position of the transition state structure is set via translation and rotation. The next step consists of variation of the structure of the transition state by enforcing over the atoms the transition vector of the structure multiplied by a variation coefficient. Analogously to the equivalent structures search, the optimization is performed by conjugated gradient quasi-Newton method and the RMSD distance between equivalent atoms from the two molecules in the pair serves as minimization function. The graphical result of the described procedure is presented in Fig. 3.

As a result of the performed conformational search we obtained 14, 9, and 6 stable structures for TS_41, TS_42 and TSc transition states, respectively, and more than 100 conformers of the intermediate structure, I. In Fig. 4 the population of the different conformers of I (assuming Boltzman distribution) with respect to the most stable structure is presented.

The common methods for transition state search focus on obtaining the local structures in the region in which the bonds are cleaved or formed. In various cases, however, the conformation of the rest of the molecule can affect strongly the energy of the obtained transition state, as it is shown here for ammonolysis of formylethanediol. Thus, our extensive conformational search approach complements the common TS search algorithms by accounting for the contribution of the mobility of the parts of the reactants that are not directly involved in the reaction center.

Frequently the transition state conformer with lowest energy does not correspond to the most stable conformers of the reactants or intermediates. Therefore, most of the methods for transition state search as linear synchronous transit (LST) cannot find transition state conformer with lowest energy starting from reactants or intermediates in their most stable conformers. Our approach systematically generates all possible conformers of the transition states and intermediates and compares all structures in the pool of the intermediates with the pool of the transition state structures. Thus, one can find the transition state structure with lowest energy and the intermediate conformer that corresponds to it.

4.2. Structure search for hexagonal transition states using the complex method

The search of the hexagonal transition states was based on completely different approach. Since these structures were not known in advance, we applied the *complex* algorithm. The *complex* algorithm was first proposed by Box [56] as extension of the simplex method for constrained optimization and in different modifications it is widely applied for identifying global extrema. The method is helpful for search of stable conformers of structures with predefined (intuitively guessed) topology but with unknown geometry and its main advantage is that it does not require any derivatives of the minimizing function and thus, is computationally very simple and efficient. Since the iterative procedure of this approach is described in details elsewhere [37,56], we are not going to discuss it here and will present only the basic idea of the method. With this approach one constructs a sequence (complex) of points in the phase space of geometry coordinates, which allows identification of the point corresponding to the global minimum eventually obeying imposed constrains. In the concrete application the procedure results in locating of the lowest energy transition state with desired connectivity by pre-definition of the forming and cleaving bonds in the hexagonal reaction ring (see Scheme 1b) and the covalent bonds in the reacting species. The minimizing function was defined from the modified MM field as implemented in the MolRan program suite [21] as the terms, taking into account the formation of hydrogen bonds, were excluded. The initial complex consisted of $2N/5$ (N is the number of all degrees of freedom in the system) random geometries as up to 1×10^{-4} of the volume of the complex (in the space of Cartesian coordinates) was optimized and the variation of the volume in the complex was 1×10^{-3}. The obtained structures were further optimized at the B3LYP/6-31G** level and the corresponding hexagonal transition states of the stepwise and concerted mechanisms are presented in Fig. 5. Due to the formation of a large reaction cycle with participation of the vicinal OH group in these transition states, their conformational freedom is strongly restricted and exhaustive conformational search (as described in Section 4.1) is not necessary.

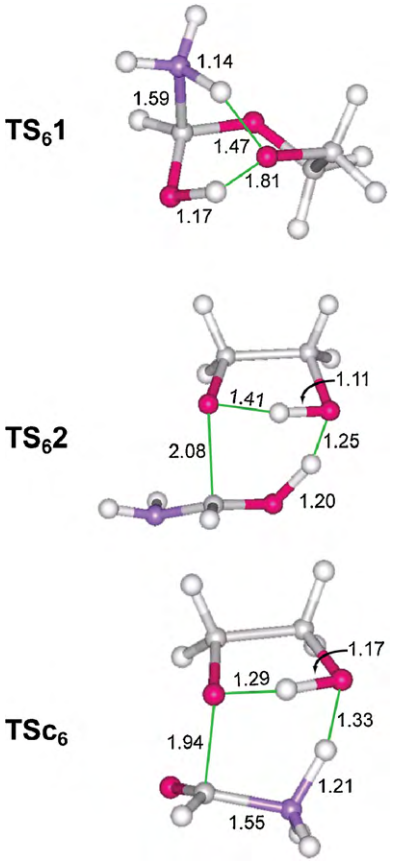


Fig. 5. Optimized structures of the hexagonal transition states of the stepwise ($TS_6$1 and $TS_6$2) and of the concerted (TSc_6) mechanism. The bond lengths and some important interatomic distances are given in Å.

Unlike other methods used for finding of transition states structures, this approach is capable to find lowest energy transition state structure with pre-defined topology without initial guess of the structure. This method is particularly useful when the reaction system undergoes large unpredictable changes from reactants to products, or when the elementary reaction step includes cleavage/formation of more than one bond where the applicability of most of the traditional approaches for transition state search is limited.

4.3. Comparison of different reaction paths

The detailed search of the tetragonal transitions state structures with various conformations (Section 4.1) and of the hexagonal transition states (Section 4.2) allowed us to identify the transition states with the lowest energy in the alternative stepwise and concerted reaction paths (see Scheme 1a and b). The calculated electronic energies with B3LYP functional and two types of basis sets, 6-31+G* and 6-31++G**, are reported in Table 2 (only the values obtained with the larger basis set are discussed in the following). In order to evaluate the influence of the vicinal OH group in formylethanediol on the reaction paths for ester ammonolysis as a reference we consider the reaction with the structure, in which the two O atoms of the ethanediol are in *trans* position and an intramolecular hydrogen bond cannot be formed. In this case the reaction can be accomplished only via tetragonal transition states, $TS_4$1, $TS_4$2, or TSc. The rate-determining step of the stepwise mechanism is the first transition state with ΔE = 52.7 kcal/mol, but the transition state of the concerted mechanism has slightly lower energy, 51.5 kcal/mol.

In the most stable conformers of the tetragonal transition states with intramolecular hydrogen bonds from the vicinal hydroxyl group the electronic energies of $TS_4$1, $TS_4$2, or TSc are reduced to 44.9, 41.2, and 43.8 kcal/mol. Thus, similarly to the reference process, the reaction occurs preferably via concerted mechanism with catalytic effect of the vicinal OH group (reduction of the activation energy barrier) estimated at 7.7 kcal/mol. The reason for this weak catalytic effect is the preservation of the tetragonal structure of transition states, as it is in the case of non-catalyzed process, since this structure imposes rather small angles for proton transfer in the transition states, around 125–130°, much smaller than the optimal angle of 180°.

The other mechanism of catalytic activation of the ester ammonolysis of formylethanediol—by participation of the vicinal OH in the proton transfer via hexagonal transition states, stabilizes considerably the first transition state of the stepwise mechanism, by 21.5 kcal/mol with respect to the referent value. The main reason for this effect is the replacement of single unfavorable proton transfer in $TS_4$1 by two favorable proton transfers via a six-atomic ring in $TS_6$1 that decreases the angular distortions for the proton transfers (the angles for proton transfers are larger, 149° and 155°). However, the energy of the other two transition states TSc_6 and $TS_6$2 is reduced by only 2.2 and 4.9 kcal/mol compared to the reference reaction since their structures are partially constrained from five-atomic ring for proton transfer (see Scheme 1b and Fig. 5). Comparing the activation energy barriers for the hexagonal transition states, 31.2 kcal/mol, 25.2 kcal/mol (with respect to the intermediate), and 49.3 kcal/mol for $TS_6$1, $TS_6$2 and TSc_6, one finds that the reaction occurs preferably by stepwise mechanism with its first step as rate-determining. Since this barrier, 31.2 kcal/mol, is significantly lower than the energy barriers for the tetragonal transition states with or without hydrogen bond from the vicinal hydroxyl, 43.8 and 51.5 kcal/mol, we can conclude that the process via hexagonal transition states is preferable.

The analysis of the activation energies of the lowest energy transition states of the different reaction pathways for ammonolysis of acylated diol suggests that the total catalytic effect of the vicinal OH group can be estimated at 21.5 kcal/mol when the reaction occurs via hexagonal transition states. As reported elsewhere [52], the analysis of the Gibbs free energy of activation of different transition states resulted in essentially the same conclusion.

Table 2
Electronic energy ΔE (in kcal/mol) of the intermediate, product, and transition state structures with lowest energy for each type of transition state for 1-O-formyl 1,2-ethanediol ammonolysis[a]. The energies are calculated with respect to the initial state.

Energy	Basis set	I	Product[b]	TSc	TS1	TS2
Reference (no H-bond)	6-31+G*	16.7	−2.6	53.5	54.7	48.3
Tetragonal TS		11.2	−2.6	45.0	46.4	44.1
Hexagonal TS		11.2	−2.6	50.9	32.5	42.3
Reference (no H-bond)	6-31++G**	16.6	−4.5	51.5	52.7	45.2
Tetragonal TS		9.9	−4.5	43.8	44.9	41.2
Hexagonal TS		9.9	−4.5	49.3	31.2	40.3

[a] The reaction mechanism is shown in Scheme 1 and the structures corresponding to the stationary points along the reaction paths are provided in Figs. 2 and 5.
[b] The reported energy values correspond to a complex formed by the species produced during the reaction.

In order to evaluate the influence of the polarization and diffuse functions in the basis set of hydrogen atoms, we can compare the calculated energies of different stationary points along the reaction paths obtained with 6-31 + G* and 6-31 + + G** basis sets (Table 2). With the larger basis set all energies are reduced compared to the initial state by up to 3.1 kcal/mol. The largest reduction is found for TS2 transition states by 2.0–3.1 kcal/mol, while the intermediates are stabilized by at most 1.4 kcal/mol. This difference likely reflects the importance of the proton transfer in the transition states, which is described better with diffuse and polarization functions added for hydrogen atoms.

The results reported here can be related to an experimental study of the intramolecular aminolysis reactions in presence and absence of a vicinal OH group [57]. In agreement with the computational results for acceleration of the reaction via the hexagonal transition states, the lactamization of ornithinyl adenosine is orders of magnitude faster as compared to the lactamization of the corresponding deoxy compound.

5. Conclusions

By modeling the structure of formic acid and its dimer and of the ammonolysis of formic acid at different levels of theory including about sixty combinations of 6 quantum-chemical methods and different basis sets and further statistical analysis of the obtained structural and energetic data, we defined B3LYP/6-31G** level as the least time consuming, yet still reliable method for correct description of the geometry of the modeled structures and the reaction energies.

The conformational search procedure described in the current paper and implemented in the MolRan program suite allows identification of all distinguishable conformers of the transition states and selection of the lowest energy structures thus, leading to prediction of the energy favorable reaction route. As demonstrated here, the procedure is appropriate for tuning the geometry of transition states with known structure of the reaction center and with large conformational freedom of the parts of the reacting molecules not involved in the reaction center. On the other hand, we have also showed the successful application of the complex method for generation of initial structures of unknown transition states based on pre-defined bonding (topology) between atoms participating in bond-breaking and forming.

Acknowledgement

The work was supported by Bulgarian National Fund for Scientific Research (contracts DOO2-133 and DOO2-210).

References

[1] M.R.A. Blomberg, P.E.M. Siegbahn, Quantum chemistry as a tool in bioenergetics, Biochim. Biophys. Acta 1797 (2010) 129–142.
[2] P. Banáš, P. Jurečka, N.G. Walter, J. Šponer, M. Otyepka, Theoretical studies of RNA catalysis: hybrid QM/MM methods and their comparison with MD and QM, Methods 49 (2009) 202–216.
[3] T. Dudev, C. Lim, Metal binding affinity and selectivity in metalloproteins: insights from computational studies, Ann. Rev. Biophys. 37 (2008) 97–116.
[4] J. Gao, S. Ma, D.T. Major, K. Nam, J. Pu, D.G. Truhlar, Mechanisms and free energies of enzymatic reactions, Chem. Rev. 106 (2006) 3188–3209.
[5] N. Rösch, V.A. Nasluzov, K.M. Neyman, G. Pacchioni, G.N. Vayssilov, Supported metal species and adsorption complexes on metal oxides and in zeolites: density functional cluster model studies, in: J. Leszczynski (Ed.), Computational Material Science. Theoretical and Computational Chemistry Series, vol. 15, Elsevier, Amsterdam, 2004, pp. 367–450.
[6] K. Morokuma, Theoretical studies of structure, function and reactivity of molecules—a personal account, Proc. Jpn. Acad. B 85 (2009) 167–182.
[7] L.V. Moskaleva, Z.-X. Chen, H.A. Aleksandrov, A.B. Mohammed, Q. Sun, N. Rösch, Ethylene conversion to ethylidyne over Pd(1 1 1): revisiting the mechanism with first-principles calculations, J. Phys. Chem. C 113 (2009) 2512–2520.
[8] D. Loffreda, F. Delbecq, F. Vigné, P. Sautet, Fast prediction of selectivity in heterogeneous catalysis from extended brønsted-evans-polanyi relations: a theoretical insight, Angew. Chem., Int. Ed. 48 (2009) 8978–8980.
[9] T. Mineva, E. Sicilia, N. Russo, Density-functional approach to hardness evaluation and its use in the study of the maximum hardness principle, J. Am. Chem. Soc. 120 (1998) 9053–9058.
[10] F.P. Rotzinger, Treatment of substitution and rearrangement mechanisms of transition metal complexes with quantum chemical methods, Chem. Rev. 105 (2005) 2003–2037.
[11] B. Galabov, V. Nikolova, J.J. Wilke, H.F. Schaefer, W.D. Allen, Origin of the S_N2 benzylic effect, J. Am. Chem. Soc. 130 (2008) 9887–9896.
[12] E. Proynov, J. Kong, Assessment of a new hybrid functional on thermochemistry, metal–ligand bonding and reaction barriers, in: G.N. Vayssolov, T. Mineva (Eds.), Theoretical Aspects of Catalysis, vol. 3, Heron Press, Sofia, 2009, pp. 1–9.
[13] M.A. Rangelov, G.N. Vayssilov, V.M. Yomtova, D.D. Petkov, Theoretical study of the *o*-OH participation in catechol ester ammonolysis, Org. Biomol. Chem. 3 (2005) 737–744.
[14] M.A. Rangelov, G.N. Vayssilov, D.D. Petkov, Quantum chemical model study of the acyl migration in 2′(3′)-formylnucleosides, Int. J. Quant. Chem. 106 (2006) 1346–1356.
[15] A. Gindulyte, A. Bashan, I. Agmon, L. Massa, A. Yonath, J. Karel, The transition state for formation of the peptide bond in the ribosome, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 13327–13332.
[16] G. Wallin, J. Åqvist, The transition state for peptide bond formation reveals the ribosome as a water trap, Proc. Natl. Acad. Sci. U.S.A. 107 (2010) 1888–1893.
[17] M.A. Rangelov, G.N. Vayssilov, V.M. Yomtova, D.D. Petkov, The syn-oriented 2-OH provides a favorable proton transfer geometry in 1,2-diol monoester aminolysis: implications for the ribosome mechanism, J. Am. Chem. Soc. 128 (2006) 4964–4965.
[18] M.V. Rodnina, M. Beringer, W. Wintermeyer, Mechanism of peptide bond formation on the ribosome, Quart. Rev. Biophys. 39 (2006) 203–225.
[19] M.M. Changalov, G.D. Ivanova, M.A. Rangelov, P. Acharia, S.N. Acharya, Minakawa, A. Foldesi, I.B. Stoineva, V.M. Yomtova, C.D. Roussev, A. Matsuda, J. Chattopadhyaya, D.D. Petkov, 2′/3′-O-peptidyl adenosine as a general base catalyst of its own external peptidyl transfer: implications for the ribosome catalytic mechanism, ChemBioChem 6 (2005) 992–996.
[20] M.H. Zhong, S.A. Strobel, Synthesis of isotopically labeled P-site substrates for the ribosomal peptidyl transferase reaction, J. Org. Chem. 73 (2008) 603–611.
[21] M.A. Rangelov, *MolRan*, Version 1.8, 2008. More information about the program can be obtained from the author via e-mail: mran@orgchm.bas.bg.
[22] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery Jr., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople, Gaussian 03, Revision E.01, Gaussian, Inc., Wallingford, CT, 2004.
[23] J.A. Pople, J.S. Binkley, R. Seeger, Theoretical models incorporating electron correlation, Int. J. Quant. Chem. 10 (1976) 1–19.
[24] M.J. Frisch, M. Head-Gordon, J.A. Pople, A direct MP2 gradient method, Chem. Phys. Lett. 166 (1990) 275–280.
[25] R. Ditchfield, W.J. Hehre, J.A. Pople, Self-consistent molecular-orbital methods. IX. An extended gaussian-type basis for molecular-orbital studies of organic molecules, J. Chem. Phys. 54 (1971) 724–728.
[26] W.J. Hehre, R. Ditchfield, J.A. Pople, Self-consistent molecular orbital methods. XII. Further extensions of gaussian-type basis sets for use in molecular orbital studies of organic molecules, J. Chem. Phys. 56 (1972) 2257–2261.
[27] P.C. Hariharan, J.A. Pople, Accuracy of AH_n equilibrium geometries by single determinant molecular orbital theory, Mol. Phys. 27 (1974) 209–214.
[28] M.S. Gordon, The isomers of silacyclopropane, Chem. Phys. Lett. 76 (1980) 163–168.
[29] P.C. Hariharan, J.A. Pople, The influence of polarization functions on molecular orbital hydrogenation energies, Theor. Chim. Acta 28 (1973) 213–222.
[30] R. Krishnan, J.A. Pople, Approximate fourth-order perturbation theory of the electron correlation energy, Int. J. Quant. Chem. 14 (1978) 91–100.
[31] R. Krishnan, M.J. Frisch, J.A. Pople, Contribution of triple substitutions to the electron correlation energy in fourth order perturbation theory, J. Chem. Phys. 72 (1980) 4244–4245.
[32] J.A. Pople, R. Krishnan, H.B. Schlegel, J.S. Binkley, Electron correlation theories and their application to the study of simple reaction potential surfaces, Int. J. Quant. Chem. 14 (1978) 545–560.
[33] J.S. Binkley, J.A. Pople, W.J. Hehre, Self-consistent molecular orbital methods. 21. Small split-valence basis sets for first-row elements, J. Am. Chem. Soc. 102 (1980) 939–947.
[34] A.D. McLean, G.S. Chandler, Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, $Z = 11–18$, J. Chem. Phys. 72 (1980) 5639–5648.
[35] N.L. Allinger, Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V_1 and V_2 torsional terms, J. Am. Chem. Soc. 99 (1977) 8127–8134.

[36] A. Hocquet, M. Langgård, An evaluation of the MM + force field, J. Mol. Model. 4 (1998) 94–112.
[37] S.S. Rao, Engineering Optimization: Theory and Practice, Wiley–Interscience, 1996.
[38] S. Stoyanov, Optimization of Technological Processes, Tehnika, Sofia, Bulgaria, 1993 (in Bulgarian).
[39] J. Hermans, Hydrogen bonds in molecular mechanics force fields, Adv. Protein Chem. 72 (2005) 105–119.
[40] A. Almenningen, O. Bastiansen, T. Motifeldt, A reinvestigation of the structure of monomer and dimer formic acid by gas electron diffraction technique, Acta Chem. Scand. 23 (1969) 2848–2864.
[41] A.F. Wells, Structural Inorganic Chemistry, Oxford University Press, Oxford, UK, 1984.
[42] J.D. Cox, G. Pilcher, Thermochemistry of Organic and Organometallic Compounds, Academic Press, New York, 1970.
[43] T. Oie, G.H. Loew, S.K. Burt, J.S. Binkley, R.D. MacElroy, Quantum chemical studies of a model for peptide bond formation: formation of formamide and water from ammonia and formic acid, J. Am. Chem. Soc. 104 (1982) 6169–6174.
[44] J.H. Jensen, K.K. Baldridge, M.S. Gordon, Uncatalyzed peptide bond formation in the gas phase, J. Phys. Chem. 96 (1992) 8340–8351.
[45] S. Chalmet, W. Harb, M.F. Ruiz-Lopez, Computer simulation of amide bond formation in aqueous solution, J. Phys. Chem. A 105 (2001) 11574–11581.
[46] O. Takahashi, Y. Kohno, M. Nishio, Relevance of weak hydrogen bonds in the conformation of organic compounds and bioconjugates: evidence from recent experimental data and high-level *ab initio* MO calculations, Chem. Rev. (2010), doi:10.1021/cr100072x.
[47] Y. Zhao, D. Truhlar, The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals, Theor. Chem. Acc. 120 (2008) 215–241.
[48] M. Simonović, T.A. Steitz, A structural view on the mechanism of the ribosome-catalyzed peptide bond formation, Biochim. Biophys. Acta 1789 (2009) 612–623.
[49] M.V. Rodnina, W. Wintermeyer, Peptide bond formation on the ribosome: structure and mechanism, Curr. Opin. Struct. Biol. 13 (2003) 334–340.
[50] M. Beringer, M.V. Rodnina, The ribosomal peptidyl transferase, Mol. Cell. 26 (2007) 311–321.
[51] I. Wekselman, C. Davidovich, I. Agmon, E. Zimmerman, H. Rozenberg, A. Bashan, R. Berisio, A. Yonath, Ribosome's mode of function: myths, facts and recent results, J. Pept. Sci. 15 (2009) 122–130.
[52] M.A. Rangelov, G.P. Petrova, V.M. Yomtova, G.N. Vayssilov, Catalytic Role of Vicinal OH in Ester Aminolysis - Proton Shuttle versus Hydrogen-Bonds Stabilization, submitted for publication.
[53] G.M. Blackburn, W.P. Jencks, The mechanism of the aminolysis of methyl formate, J. Am. Chem. Soc. 90 (1968) 2638–2645.
[54] K. Takashima, J.M. Riveros, Gas-phase pathways for ester hydrolysis, J. Am. Chem. Soc. 100 (1978) 6128–6132.
[55] S. Ilieva, B. Galabov, D.G. Musaev, K. Morokuma, H.F. Schaefer, Computational study of the aminolysis of esters. The reaction of methylformate with ammonia, J. Org. Chem. 68 (2003) 1496–1502.
[56] M.J. Box, A new method of constrained optimization and a comparison with other methods, Comput. J. 8 (1965) 42–52.
[57] S.G. Bayryamov, M.A. Rangelov, A.P. Mladjova, V. Yomtova, D.D. Petkov, Unambiguous evidence for efficient chemical catalysis of adenosine ester aminolysis by its 2′/3′-OH, J. Am. Chem. Soc. 129 (2007) 5790–5791.