An Introductory Classroom Exercise on Molecular Model Building and Energy Minimization: Students' Approaches and Difficulties

Andrés Alberto Poeylaut-Palena, Luciana Méndez and María de los Ángeles Laborde*

Instituto de Química Rosario. Facultad de Ciencias Bioquímicas y Farmacéuticas – Universidad Nacional de Rosario, Argentina, laborde@iquir-conicet.gov.ar

Received October 5, 2015. Accepted November 6, 2015.

Abstract: In an increasingly digital world, computing technology plays a key role in the understanding, teaching and development of science disciplines. Modern molecular modeling involves the use of computers that can calculate physical properties of molecules and manipulate images in three dimensions. In this study, we present a learning module for the understanding of molecular modeling essentials. It is an introductory exercise for students who are not familiar with the field of computational chemistry and which can be delivered by professors who are not necessarily experts in such field. This classroom exercise is designed to be carried out following detailed instructions that make software handling straightforward. Students will be able to understand issues related to molecular model building and energy minimization. In addition, this learning module not only deals with molecular modeling but also helps students improve the learning processes of basic concepts such as resonance structures or conformational energy of structures with different dihedral angles. This is accomplished by providing new and modern alternatives for knowledge acquisition and result assessment.

Introduction

The use of the computer as a learning tool is ubiquitous nowadays. Technology is playing a key role both in the understanding and the development of science disciplines. Modern molecular modeling involves the use of computers that can calculate physical properties of molecules and manipulate images in three dimensions. The use of molecular modeling as an educational tool in the undergraduate curriculum has been implemented only within the past two decades [1]. The integration of molecular modeling into the Chemistry lab encourages students to process, revise and reinterpret key chemical concepts in a modern way [2]. Molecular modeling softwares are becoming increasingly user-friendly and, in addition, several articles explaining calculation methods have been published [3]. Combined, these factors set the ground for a widespread implementation of computational chemistry as an educational tool in the undergraduate curriculum [4].

Like any other area of knowledge, computational chemistry has different levels of complexity [5]. Herein, we present a learning module which is intended to be introductory to the subject. The operations carried out through this learning module can be performed on ordinary personal computers. Simple exercises are performed in order to teach key issues regarding molecular modeling. Students will learn the basis of molecular modeling in order to minimize the energy of given molecular models. During the learning module, exercises performed with *n*–butane and phenol molecular models are used to point out key issues about local and global minima in the energy surface or fundamental concepts about resonance structures. In this way, students revise basic chemical concepts through the eyes of molecular modeling, which provides a great opportunity to improve the learning process. Building a more complex molecular model, such as estradiol, students get closer to a biological interesting compound with many stereocenters and more complicated features regarding the calculation process.

Materials

To carry out this work, an ordinary personal computer is used individually or every two students. The molecular modeling software recommended is Hyperchem 7.5. Software minimal requirements in order to run HyperChem on PCs are Windows NT, 98, ME, 2000, XP, or Vista. 128 MB of RAM and 50–150 MB of hard disk space. The requirements for running Windows are generally more severe than for running HyperChem. As an example, this work was conducted on an AMD sempron // 798 Hz // 512 MB RAM, Windows XP Home Edition SP2. Introductory information, a discussion text and software instructions are included in a Study Guide which is presented as Supplementary material to this article. Each student is provided with a Study Guide. Crystallographic structure of estradiol was extracted from the published crystallographic structure of estradiol in complex with its biological receptor [6]. A file of the structure is also provided as supporting information.

Experimental Methodology

Information Organization. Our main efforts were focused on the study guide design. All software related instructions were removed from the general introduction. A detailed explanation of this procedure was published elsewhere [4]. The Study Guide is composed of two main sections: on the one hand, a discussion about molecular modeling and, on the other hand, a detailed list of commands to follow on the Software interface (see Supporting Information). Commands to be executed were included in the general introduction section so that students can read the text and execute them on the computer every time they find a call in the text. Each numbered command is fully explained and accompanied with the 'expected effect'. The list of commands is extremely detailed, which allows

Table 1. *n*-butane geometry optimization starting from varying C1-C2-C3-C4 dihedral angles (β)

Entry	Initial Angle	Final Angle	Energy
1	2	75	–1235.1841 kcal/mol
2	45	75	–1235.1841 kcal/mol
3	118	75	–1235.1841 kcal/mol
4	122	180	–1235.8908 kcal/mol
5	180	180	–1235.8908 kcal/mol

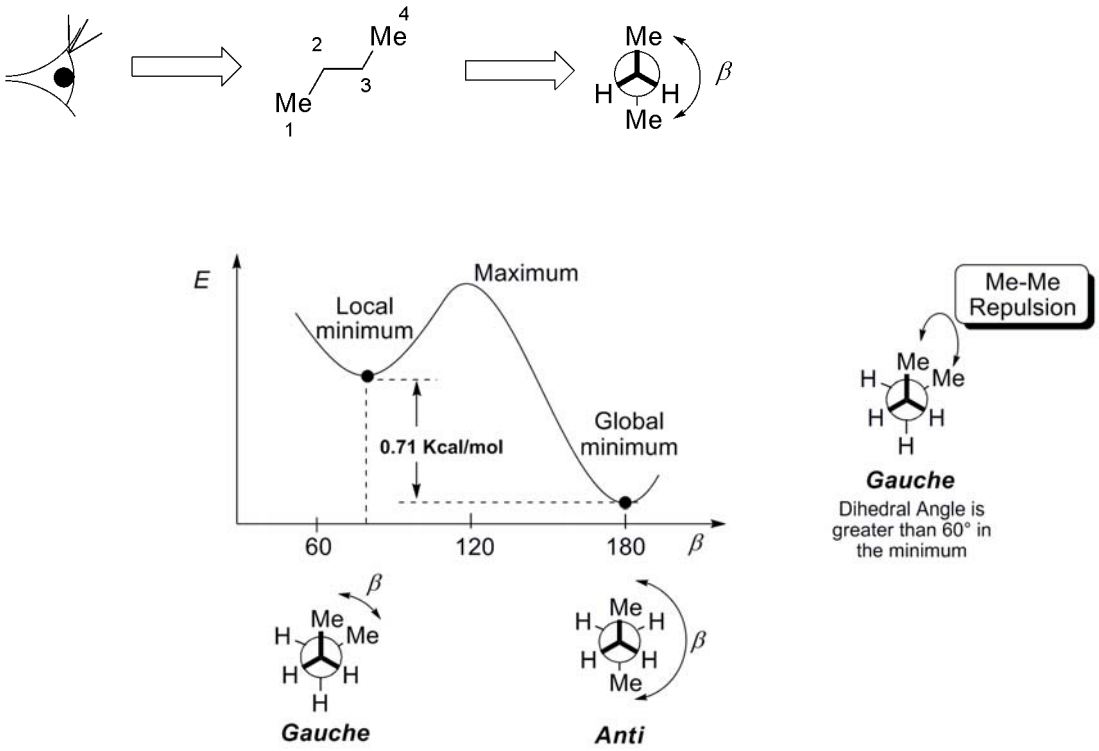


Figure 1. Energy surface potential drawn as a function of the collected data (see Table 1). A local minimum is located at 75° and a global minimum at 180°. There is a difference in 0.71 kcal/mol between the minima. Maximum is located at 120° but its energy could not be estimated from the data. Local minimum is not located exactly at 60° because of the distortion caused by Me-Me repulsion.

most students to perform the exercise almost with no help. The 'expected effect' for each action allows students to corroborate that each command has been properly executed. Students should be asked to read the Study Guide and then to proceed with the task, avoiding to go through the list of commands, and ignoring the Study Guide discussion.

The concepts related to geometry optimization are discussed within the main text, which is so simple that students can read it as an introduction prior to the class. This worked very well and became the essence of the Study Guide, allowing us to reduce the complexity associated with program handling. Students then learn how to use the program by following simple commands. There are strategically placed stop points (*ca.* every ten commands), which enables all students to complete the exercises stepwise and simultaneously. When all the students reach a stop point, a group discussion and a theoretical evaluation of the results are carried out, allowing a better control of the learning process.

First Steps. The first step in the exercise is a general introduction to the programs. Graphical interface is presented and a brief explanation on menu operations is given. Modern programs for molecular modeling can be used intuitively by a regular computer user. Students quickly learn how to draw and optimize the structure of *n*-butane. To perform calculations, students employ AM1[7] a semi-empirical method, which is fast and accurate for our purposes and which has been previously employed in classroom exercises [8]. Before starting the first calculation, students spend a great deal of time thinking about the requirements needed. Students realize that the intervention of the operator during energy minimization is very limited. The iterative process of the calculation is a concept that calls their attention: they want to know what an iteration is and how it works. Most students understand the process when explained for the first time. In order to make it simpler, energy surface can be explained as the first derivative of one-dimensional function [9]. After these initial movements, data analysis and theorization become a central part of the exercise.

Influence of the Initial Model Geometry on the Final Result. Most minimization algorithms do not move uphill in the energy surface and are designed to go only downhill, searching for an energy minimum. Potential energy surfaces are usually intricate, having several minima. Thus, different results for each starting point can be obtained. Therefore, after a single geometry optimization calculation, the operator is not able to determine whether the structure obtained is already at the global energy minimum. This concept can be understood by anyone familiar with computational chemistry. However, most students are unable to understand it unless they are provided with a good example. Interestingly, a simple molecule like *n*-butane can thus be used as an example to introduce most of these fundamental concepts. In a conformational analysis, dihedral angle β, which is defined by carbon atoms C1-C2-C3-C4, is modified five times in the starting molecular model. Results show two different energy minima depending on the initial dihedral angle. According to the energies of each conformation, the global minimum is located at 180° while there is a local minimum at 75° (Table 1, entries 1 and 5).

Students find out that the result obtained is a final structure with a dihedral angle of 75° for values within 2 and 120° (entries 1–3) while for values within 122 and 180°, the final structure corresponds to a

Table 2. Interatomic distances for the phenol molecular model before and after AM1 based geometry optimization

Entry	Bond	Initial Model (Å)	Post-Calculation (Å)	Ranking C=C
1	C1-C2 | C1-C6	1.400 | 1.400	1.402 | 1.405	Longest
2	C2-C3 | C5-C6	1.400 | 1.400	1.393 | 1.391	Shortest
3	C3-C4 | C4-C5	1.400 | 1.400	1.394 | 1.396	Intermediate
4	C-H2 | C-H6	1.080 | 1.080	1.099 | 1.098	
5	C-H3 | C-H5	1.080 | 1.080	1.100 | 1.100	
6	C-H4	1.080	1.099	
7	C-O	1.360	1.377	

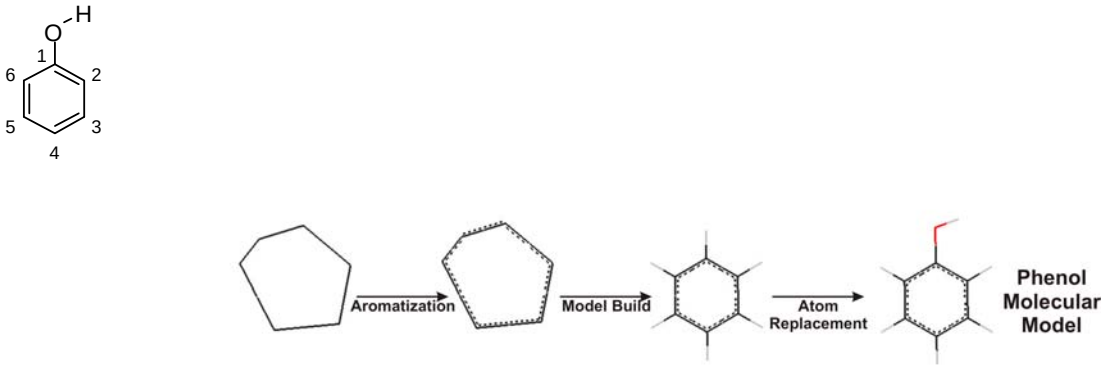


Figure 2. Phenol molecular model drawing, carbon-carbon bonds are drawn as an irregular hexagon. After aromatization, the bond distances are adjusted to standard values (model build) and finally a hydrogen atom is replaced by oxygen.

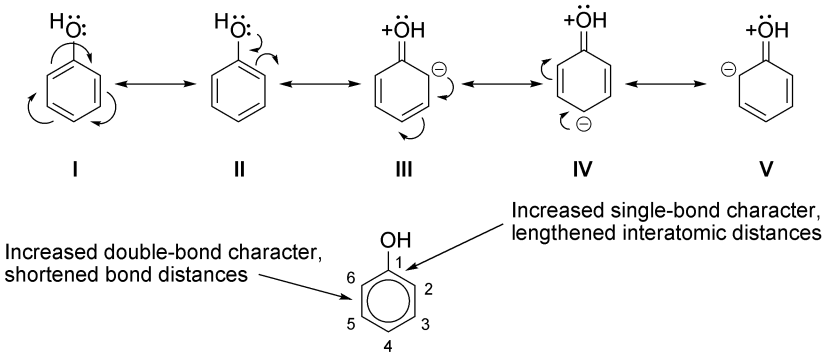


Figure 3. Resonance structures of phenol. Predominantly, C1-C2 bond is represented as a single-bond (3 out of 5 canonical structures), which leads to lengthened bond distance. The opposite occurs with C2-C3 bond, its double-bond character is reflected in the shortened interatomic distance.

dihedral angle of 180° (entries 4–5). This provides data to draw an approximate energy surface as a function of the dihedral angle: a global minimum at 180°, a maximum within the region 118–122° and a local minimum at 75° (Figure 1). Note that the local minimum corresponds to a dihedral angle of 75° which is 15° away from the theoretical value of 60°. This is related to the repulsion between methyl groups, which distorts geometry.

Phenol Geometry Optimization. Once the introduction to molecular modeling and conformational analysis is finished, the exercise continues by drawing the phenol molecular model. Students draw an irregular hexagon and ask the program to transform it into an aromatic ring (Figure 2). It is well known that aromaticity implies that double bonds are conjugated and the ring is a hybrid structure arising from canonical forms. If we do not make it explicit that those double bonds are conjugated, the software calculates the structure as a 1,3,5–cyclohexatriene [10]. The process ends with the addition of hydrogens, assignment of standard bond distances and finally with the addition of the oxygen atom (Figure 2).

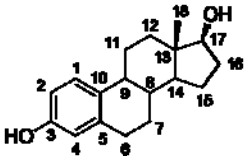
Once the drawing is finished, geometry optimization is performed. Students will measure bond distances before and after performing the calculation and collect data in a table. Before calculation, each bond type has a standard distance (Table 2, Initial Model). After calculation, physical parameters reach their optimized value (Post-Calculation) and each bond distance has a particular value which reflects an intrinsic molecular property.

In the optimized model, C1–C2 and C1–C6 bonds are the longest while C2–C3 and C5–C6 are the shortest (Table 2, Ranking). Besides, variations in bond distances are small, which reflects the stability of benzene structure. Most of the observed bond distances can be rationalized on the basis of the delocalization of oxygen lone pair electrons on the aromatic ring. When drawing resonance structures for phenol, we can observe that C1–C2 and C1–C6 bonds are represented in 4 out of 5 canonical forms as a C-C bond, which decreases the double bond character (Figure 3). At this point, a Basic Organic Chemistry theoretical concept, such as resonance hybrids, is correlated with extremely sophisticated calculations.

Another fact that students immediately find out while filling the table is the asymmetry in the model. Bond distances are not exactly the same in both sides of the plane perpendicular to C1–C4 (see Table 1, Post-Calculation). This is because the O-H bond is located on the same plane as the aromatic ring, at one side of the C1–C4 plane, turning the model asymmetric. It is worth mentioning that energy minimization results in a minimum energy structure which does not necessarily reflect the actual physical properties of the molecule. Experimental physical parameters arise from a weighted average distribution of all possible conformations. Thus, the phenol molecule

Table 3. Physical parameters of AM1 geometry optimized phenol molecular model compared with a crystallographic model

Entry	Parameter	AM1 Optimized Structure	Crystallographic Structure	Disagreement
1	*d* O3–O17	10.7779 Å	10.8149 Å	0.3%
2	Dihedral Angle C7–C8–C9–C11	176.818	178.505	0.9%
3	Dihedral Angle C6–C7–C8–C14	175.436	175.944	0.3%
4	Dihedral Angle C8–C14–C15–C16	158.488	180.000	11.9%
5	Dihedral Angle C13–C17–C16–C15	16.7274	6.1625	63.1%



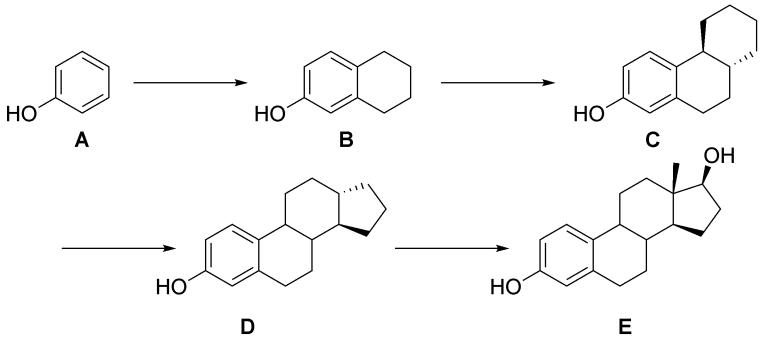


Figure 4. Stepwise construction of estradiol (**E**), rings are added starting from phenol (**A**). Intermediates **C**, **D** and **E** possess stereogenic centers which must be properly constructed.

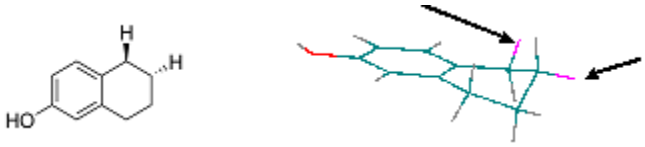


Figure 5. Fixing stereochemistry in the construction of **C** from **B**. Hydrogen atoms to be replaced are pointed with arrows, one above the plane, the other below.

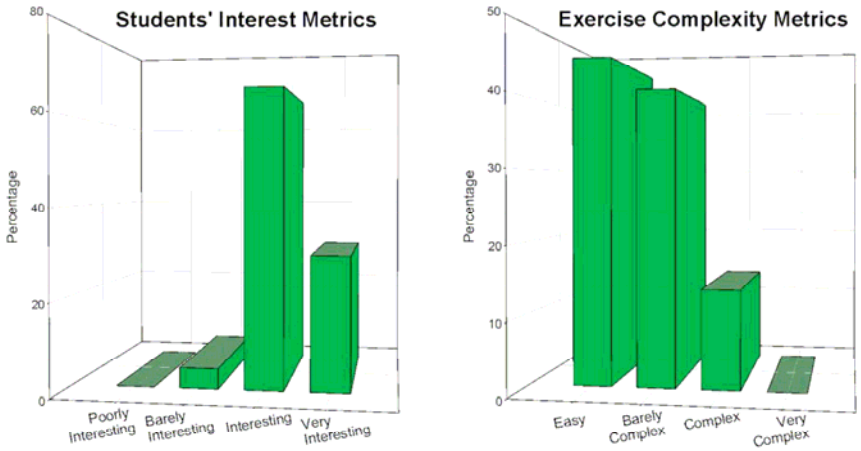


Figure 6. Results of the anonymous survey asking for interestingness and complexity of the exercise.

is completely symmetrical since the O-H bond is distributed on both sides of the C1–C4 plane.

Estradiol Construction. Once the introduction to calculation is concluded, the next step of this exercise is to build a molecular model of greater complexity. The molecule chosen is estradiol (**E**), a female sex hormone that consists of four fused rings and five stereocenters (Figure 4) [11]. Construction starts from the previously drawn phenol molecular model to which the remaining rings are added stepwise with the proper stereochemistry. Even though this part of the drawing can be quite attractive for the students due to the complexity of the molecule, it is also highly error prone. Structures drawn by the students need a strict corroboration by the teaching staff. There is a 'stop point' within the command list that serves this purpose. The first ring fusion to obtain structure **B** is generally easily achieved whereas

the construction of the molecular model of structure **C** is highly demanding for beginners. In order to draw the molecular model with the proper stereochemistry, we need the ability to observe objects in three dimensions to recognize the concepts "above the plane" and "below the plane" so that we can decide which hydrogen atom should be replaced to set the stereochemistry (Figure 5). Once the structure **C** is obtained with proper stereochemistry, students have acquired the ability required in order to easily build the remaining ring fusions with stereogenic centers to obtain **D** and **E**.

Data Analysis. After model building and energy minimization, the final part of the exercise is reached. To complete the task, the obtained results are compared with data from a crystallographic structure (Table 3). Numerical comparison reveals that both structures are very similar (entries 1–3), which indicates that calculations can predict data from reality. However, it can be observed that in the region of the cyclopentane ring there is a slightly higher discrepancy (entries 4–5), giving rise to debate about its potential origin. At least, three main origins can be mentioned. First, calculations could be wrong, which would show that theoretical approaches and calculation methods can fail and there is always a need for new methods to overcome errors. Second, we may have calculated a structure which is in a different minimum from that of the crystallographic model. Finally, another fact to highlight is that geometry optimization was performed in vacuum while the molecule in the crystal was in a completely different environment.

Metrics of Success

This exercise was carried out by 200 undergraduate Pharmacy students in two semesters. Taking into account that this is an introductory exercise, our main purpose was to keep it simple and our main goal to achieve a high understanding rate of the given concepts. In this sense, less than 5% of the students failed when tested and only one student failed when retested. Exams covered two main issues: (*i*) knowledge acquired by reading the study guide prior to the classroom exercise and (*ii*) knowledge about general topics on molecular modeling learnt during the exercise. Some of the representative questions of each issue are:

Study Guide Knowledge. 1.- Which database were the structures used during this classroom exercise obtained from? 2.- Which technique was employed to determine the structure which has been used?

General Molecular Modeling. 3.- Describe the consequences of the existence of local and global energy minima in the energy minimization process. 4.- Which could the origin of differences among calculated and experimental structures be?

An anonymous survey was carried out to see how interesting students had found the exercise. Results were very motivating: about 96% of the students said the exercise was either interesting or very interesting (Figure 6). Moreover, 86% claimed they had minor difficulties in understanding the exercise or even no difficulty at all.

Conclusions

This article presents a learning module in molecular modeling. In this exercise, students learn how to build molecular models and perform elementary calculations in order to determine their most stable conformation. They also learn basic notions about calculations and how to extract information from their minimized structures. Most of the teaching was carried out with simple molecular models such as *n*-butane or phenol. Finally, students build a molecular model of estradiol and perform energy minimization to compare this information with a molecular model obtained from experimental data.

There is no need for sophisticated computers or highly qualified professors to teach this exercise. In a simple way, our Medicinal Chemistry course has been upgraded.

Through this activity, students not only learn about molecular modeling but also revise basic chemical concepts (*e.g.* resonance structures, dihedral angle conformations, etc.), improving their learning process by the use of alternative ways of knowledge acquisition. Furthermore, students realize how helpful computational chemistry can be for chemists and biochemists in the development of new active principles.

Acknowledgements. This study was supported by the Universidad Nacional de Rosario. Authors acknowledge María Robson, Mariana de Sanctis, Geraldine Raimundo, Carolina Perret, and Romina Ricardo for kindly reviewing the manuscript in search of English language corrections.

Supporting Materials. The student Study Guide Computational Chemistry I and crystallographic structure of estradiol is available.

References and Notes

1. Ealy, J. B. *J. Sci. Educ. Technol.,* **1999**, *8*(4), 309–321, DOI: 10.1023/A:1009444711570.
2. a) Ealy, J. B. *J. Sci. Educ. Technol.,* **2004**, *13*(4), 461–471, DOI 10.1007/s10956-004-1467-x. b) Dori, Y. J.; Barak, M. *Educ. Technol. Soc.*, **2001**, *4* (1), 61–74, DOI: 10.1021/ed059p269.
3. Box, V. G. S. a) *J. Chem. Educ.*, **1988**, *68* (8), 662–664. b) Boyd, D. B.; Lipkowitz, K. B. *J. Chem. Educ.,* **1982**, *59*(4): 269–274, DOI:10.1021/ed059p269.
4. Poeylaut-Palena Andrés A. and Laborde M. de los Ángeles, *Molecular Biology Education*, **2013**, *41*(2), 118–124, DOI: 10.1002/bmb.20685.
5. Leach, A. (2001). Molecular Modelling: Principles and Applications. 2nd ed.; Prentice Hall.
6. Brzozowski, A. M.; Pike, A. C. W.; Dauter, Z.; Hubbard, R. E.; Bonn, T.; Engström, O.; Öhman, L.; Greene, G. L.; Gustafsson, J.-A.; Carlquist, M. *Nature*, **1997**, *389*, 753–758, DOI: 10.1038/39645.
7. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.*, **1985**, *107*, 3902–3909, DOI: 10.1021/ja00299a024.
8. a) Montes, I.; Prieto, J. A.; García, M. *Chem. Educ*ator, **2002**, *7*, 293–296, DOI:10.1007/s00897020593a. b) Hessley, R. K. *J. Chem. Educ.,* **2000**, *77*(2), 203, DOI: 10.1021/ed077p203. c) Starkey, R.; Feller, S. E.; Dallinger, R. F.; Mckinney, P. C. *J. Chem. Educ.,* **2004**, *81*(2), 283–287, DOI: 10.1021/ed081p1819.
9. Hinchliffe, A. (2000). Modelling Molecular Structures. John Wiley & Sons, LTD: Chichester, UK.
10. Lenthe, J. H. Van; Havenith, R. W. A.; Dijkstra, F.; Jenneskens, L. W. *Chem. Phys. Lett.,* **2002**, *361*(3–4), 203–208, DOI:10.1016/S0009-2614(02)00873-4.
11. Anstead, G. M.; Carlson, K. E.; Katzenellenbogen, J. A. *Steroids*, **1997**, *62*(3), 268–303, DOI:10.1016/S0039-128X(96)00242-5.