Teaching Structure with Molecular Visualization Tools. A Crash Course from Small Molecules to the HIV-1 Protease for Undergraduate Students

Paraskevi E. Mpeza
Department of Agriculture, University of Ioannina, Greece, Kostaki Arta, P.C Gr47100 Arta-Greece

ARTICLE INFO

Article history:
Received: 09 Dec 2020
Revised: 23 March 2021
Accepted: 29 March 2021
Published online: 24 April 2021

Keywords:
Teaching Strategy
Chemistry
Undergraduate
Visualization
Molecular Modeling

ABSTRACT

One of the challenges in the teaching of chemistry to students in higher education is their introduction to the visualization and molecular depiction software, deepening thus their understanding and creating new conceptions in the physical and biological sciences. This paper describes the implementation of a learning unit for 1st and 2nd semester undergraduate students (that were not previously exposed to molecular visualizations) in agricultural technology for chemistry and biochemistry courses respectively. A series of educational activities are presented beginning with the use of simple molecular models as plastic and wooden balls for atoms and sticks for bonds, to enhance the previous knowledge of the 2-D structural conformations. The students are introduced in depth to the theory of chemical bonding, molecular shapes, and polarity. This represents a continuum of increased complexity in inquiry-based learning, starting from simple molecules and ending with the exploration of the shape and function of a protein molecule, such as the human immunodeficiency virus (HIV)-1 protease; Thus easily leads to the study of the interactions between this protease and inhibitory drugs, via freely-available software tools. The gains in the students’ understanding and help for future courses and their career are analyzed.

1. Introduction

One of the most captivating teaching tools in Chemistry is the use of molecular visualization. In particular, biochemistry educators have used it in diverse ways in an effort to enhance the student learning experience (Black, 2020; Craig et al., 2013). After all, Martin Karplus, Michael Levitt and Arieh Warshel shared the 2013 Nobel Prize in Chemistry for the development of multiscale models for complex chemical systems (Sordo, 2014).

* Corresponding author.
E-mail: bmpeza@uoi.gr

Doi: https://doi.org/10.31258/jes.5.2.p.209-223
From the chemical point of view, proteins are by far the most structurally complex and functionally sophisticated molecules known, consisting of thousands of atoms, with diverse shapes and a variety of functions (Craig et al., 2013). To the best of my knowledge students in Greece receive no training in high school on the use of molecular modelling tools in chemistry or biochemistry, but instead their books use the static two-dimensional figures, hindering thus their deeper understanding of structures Biochemistry (Ministry of Education and Religious Affairs/Institute of Educational Policy, Greece).

In many other educational systems teachers in secondary schools have developed a conceptual curriculum for molecular models (Bethel, C., & Liebermann, R., 2014). By contrast, in many other countries the old-style presentation of molecules still persists maintaining a wide gap between high school and university learning material (Moreno et al., 2018; Burgin et al, 2018). To move chemical education from the 2-D depiction of biomolecules to that of three dimensional perception is a demanding task, forcing the students to be able to recognize many symbols and established conventions, to interpret them and create powerful and lasting mental images in their minds. This is, from our experience, very confusing for the students and the teachers themselves (Mungui, 2018; Tsaparlis et al., 2018).

Although there is an exponential growth on 3D literacy and computer graphics in life sciences, it still remains a great challenge to develop teaching materials, practice the necessary skills and to incorporate 3-D models in Life Sciences (Reynolds et al., 2018, Wolle et al., 2018). Teachers can use these illustration tools to create realistic images carrying a high cognitive load. Illustrations, for example, which depict the dynamic nature of an enzyme, revealing the structural transformations during catalysis, or connecting different conformers of a structure with its biological function. “A picture is worth a thousand words”, but the best picture of a chemical-biological structure without the basic tools for understanding the structure is worthless.

The situation is quite variable at the introductory college level General Chemistry, with several popular textbooks ignoring almost totally the modelled representation of even the smallest of molecules (Taly et al., 2019), while others have admirably woven such representations into the text material, explaining thus molecular shapes, molecular interactions and chemical reactions (Ebbing & Gammon, 2013). We developed a crash course which includes basic chemical principles as seen in small important molecules and expanded it to more complicated molecules such as proteins. I believe that the development of such a course would be helpful to biochemistry education researchers in designing a strong continuum between general chemistry and biochemistry instruction.

The students learn to explore molecular structures, learn how molecules look and interact with one another, and finally to understand a biological process like disease pathogenesis from the molecular structural perspective. The aim of the two laboratory modules discussed here, and the respective lecture courses, is for
molecular shapes to be part and parcel of the student’s science education, not limited to the chemistry classes.

Student Perceptions When Coming into Their First University Chemistry Course

Chemistry students find it difficult to make the connection between the molecular formula, the geometric structure and the physicochemical characteristics of the respective substance. They are more familiar with the 2-D structure on the blackboard, in the textbook as well as on the computer screen. Most students have not developed the skills for conceiving molecular representation into 3-D space. Moreover, there is an inherent difficulty in the interpretation of macroscopic physicochemical properties based on the microscopic world. A lot of chemical information is underlying in biomolecular structure: numerous atoms, unsaturated bonds, different states of carbon atoms, cyclization arrangement of carbon atoms, variety of functional groups, acid and base properties, intramolecular and intermolecular bonds and a complicated tertiary structure. Students show an inability to remember and evaluate prior knowledge with ideas in biochemistry lesson as structure and properties of molecules of biological importance. This emphasizes the need for new strategies in teaching practice with meaningful representations to build their own scientific personality. Step by step guided knowledge needs to be constructed in a creative and meaningful manner.

One of the misleading concepts are electronegativity and polarity of covalent bond. Traditional depictions do not inspire the majority of students to advance their ability in conceiving molecular geometry.

In semesters exam of September 2020 some of the submitted questions are: “In carbon–oxygen covalent bond, common electronic pair is mainly around carbon nucleus. Do you agree with this statement?”

“Polar molecules are main hydrophobic. True or false?”

“In an organic cyclic acid find the number and the position of hydrogen atoms with acid properties” Around 40% of the respondents characterize all the group of hydrogen atoms bonded with carbon and oxygen atoms as responsible for acid behavior

In the question “Which of the following molecules form intermolecular hydrogen bonds? A, CH₄ B.CH₂F₂, CH₃CH₂OH, or CH₃COCH₃”. Eightytwo from one hundred twenty three answered CH₃CH₂OH, but in the question of the same examination test “Determine the type of intermolecular forces between molecular hydrogen in the gaseous phase A) Van der Waals, B) Hydrogen bonds C) covalent bonds, 45 % of the respondents choose as correct answer B (hydrogen bonds), 36 % of them choose covalent bonds, while 39 % of them selected Van der Waals forces. This confirms a common misconception in first year University students on the nature of hydrogen bonds and a difficulty in making distinctions between covalent bonds and intermolecular bonds. All these indicate that the
instructors need to direct student’s views to the detailed description of the molecule’s architecture and also to the weak interactions between molecules. Another interesting topic the meaning of mole entity was inquired. Students seem thinking mole as quantity of mass ignoring this as an vast number of elementary units. More difficulties in the landscape of chemical thinking are arisen from the phonetic similarity of the words molecule and mole.

2. Methodology

A sequence didactic model has been built that evolves student basic chemistry ideas from the material taught in secondary education to enactment of molecular domain. Additive elements as critical thinking and the relevance of science to social issues are also promoted. A set of educational activities oriented to develop representation skills as creation of high quality representations and recognition of molecular interactions were designed. This module has been taught by myself and another colleague to over 240 students for three consecutive years in the first and second semesters of their study.

Molecular visualization for the first semester of General and Inorganic Chemistry (called Agricultural Chemistry in the curriculum) is carried out via simple means like a physical model made of wood or plastic (“ball-and-stick” model or “space-filling models” of connected spheres). Naturally, in the case of large biomolecules, it is difficult to find and hold together the necessary thousands of pieces of balls and sticks to construct the respective structures. Molecular visualization software provides some valuable features with regard to the analysis of the structure. In this class, we use PyMOL for Educational Use-Only by Schrodinger. Protein structures are retrieved from Protein Data Bank (http://www.rcsb.org).

3. Results and Discussion

Teaching basic chemistry of small molecules with molecular visualization (1st semester, agricultural (general) chemistry)

For the implementation of tasks embodied in our General Chemistry course we have asked the students to re-create by themselves the 3-D molecular geometries (ball-and-stick models) for some of the most important small molecules. Ball and stick models are better conceivable and serve as a good substitute of the real molecular image. Novice chemistry students are able to understand the connectivity between atoms in the molecular building, so we always begin with the construction of such models.

They consider macroscopic properties or chemical concepts such as high melting and boiling points, reactivity, extensive hydrogen bonding on water at the basis of aperceptual particles. Known molecules of real life with agronomical interest as
ethylene, DDT, abscisic acid, anthocyanins are visualized and attract student’s interest.

**Exercise 1: Each atom can make a Defined Number of Covalent Bonds**

Most of the molecules in living systems contain only six different types of atoms: hydrogen, carbon, nitrogen, phosphorus, oxygen, and sulfur. To the relief of the students, these atoms are found in the first three periods of the Periodic Table of Elements, making the understanding of the electronic configuration of each atom and the resulting molecules a lot easier. The students find the position of the elements in the Periodic Table of Elements and use this information to attempt to decipher the physicochemical properties in question (e.g. electronegativity, atomic and ionic radius etc). Students are asked to work in small groups and using these elements to build simple molecules that are most commonly found in biological systems, such as water, ammonia, carbon dioxide, methane, ethylene, sulfur dioxide.

**Hands-on activity, molecular geometry**

Each molecule is represented in three ways, with structural formulas, ball-and-stick and space-filling models. With ball-and-stick the ratio of sphere’s size to bond length is smaller than in reality. However, this allows students to see bond angles clearly. The size of the electronic clouds in the space-filling model is more accurate. We take the opportunity to discuss the advantages and disadvantages of the ball-and-stick over space-filling representations.

Students become acquainted with the contents of molecular kits, “Molecular Model Set for Organic Chemistry” by Prentice Hall, (ISBN:0-205-08136-3) and “Molecular Visions”, Darling (ISBN ISBN:978-0-9648837-1-0) and what each piece in the kit represents. Spheres or polyhedral structures have different colored holes on them. The number of holes represents the number of possible bonds.

Connectors between them, which have scaled lengths, symbolize the bond length between the atoms. Using proper units someone can depict not only the arrangement of atoms in space and their connectivity but also the presence of p-orbital and lone pairs of electrons, especially in the case of Molecular Visions (Figure 1). This ability will greatly enhance the understanding of properties of the molecules we encounter in chemistry, be it general, inorganic, organic or biochemistry.

When building a molecular model, the students may often encounter molecules that are not only composed of different atoms but also of groups of the same atoms with different hybridizations, as is often the case in organic molecules (e.g. acetic acid, CH₃COOH, amino acids).
Figure 1. A. The oxygen atom “with bonds”, red sp$^3$ sticks represent the two bonds of oxygen and the red balls its two pairs of no bonded electrons. B. The nitrogen atom “with bonding orbitals”, blue sp$^3$ sticks represent the three bonding orbitals and one lone-pair of no bonded electrons as a blue ball, C. The tetrahedral carbon “with four bonding orbitals”; D. Carbon-carbon double bonds E. octahedral piece represents a six coordinate metal with octahedral geometry F; ethane molecule, G. Carbon-oxygen double bonds; H. Marker balls to represent p-orbital symmetry. Atomic nucleus at the center of the structure. I. carbon-carbon triple bonds All structures made via tools provided in Molecular Visions”, Darling (ISBN ISBN:978-0-9648837-1-0).

Teaching of biomolecules (2nd semester, biochemistry-introduction to biotechnology)

A lot of research has been carried out on HIV-protease as it is one of the main targets of all antiretroviral therapies HIV Protease, which is used for our study, is an and an iconic example for drug design (Jaskolski et al, 2015). Although a lot of knowledge has accumulated on this molecule, more research is needed in order to explore the molecular details of HIV-1 protease-inhibitor interactions, and resistance mechanisms.

Activity 1: Interesting features of the HIV-1 protease

In a brief presentation of the features of molecule, the instructor talks about the protease highlighting key structures elements (Palese, 2017). HIV-1 protease is a small enzyme, composed of two identical protein chains, each only 99 amino acids long. The two chains assemble to form a long tunnel, covered by two flexible “flaps”, with a catalytic Asp at position 25 located in the interior. Thus the active site is composed of two Asp25 residues, each from a different subunit. The flaps need to “open” to allow the substrates to access the active site. The
HIV-1 protease enzyme activity can be inhibited by blocking this symmetric active site (Figure 2). The residue numbers in the second chain are primed.


Each Thr26O1 accepts a hydrogen bond from Thr26’NH from the other subunit. Thr26 also donates a hydrogen bond to the O atom of carbonyl group of Ile24 in the other loop. This creates a rigid structure which is called “fireman grip” and stabilizes catalytic active site of protease. Gly 27 and Gly27’ tie to the substrate and Asp25 and Asp25’ attack to it (Jaskolski et al., 2015). Students select a specific amino acid and locate it in the polypeptide chain. They measure distances between neighborhoods amino and intramolecular forces as hydrogen bond, salt bridges and Van der Waals forces in the structure (Figure 3).

Activity 2: Molecular recognition

Since the discovery of HIV, 26 anti-HIV compounds have been approved by the US Food and Drug Administration (FDA). Ten protease inhibitors have been approved to date for treatment of HIV infection: saquinavir (SQV), indinavir (IDV), nelfinavir (NFV), ritonavir (RTV), fosamprenavir (FPV), atazanavir (ATV), lopinavir (LPV), amprenavir (AMP), tipranavir (TPV), and darunavir (DRV) (Zhengtong et al., 2015).

Visualization via PyMOL provides the opportunity to study drug-protease molecular recognition. This is an example of how the knowledge of basic
molecular mechanisms can rapidly translate into the development of clinically effective molecules.

![Figure 3](image)

**Figure 3.** A. Polar contacts between Thr26 and Thr26’ stabilizing the active site named as fireman grip schematic B. The dotted lines between R87, D29, and R80 indicate hydrogen bond and salt bridges. C. Inter chain hydrogen bonds between green and blue chain (1–4 and 96–99” and 96–99 1–4’). D. CH-CH London dispersion forces between Ile50 and Ile47, Ile84 as dotted lines (PDB code: 1HIV).

All these intramolecular forces delineate the overall protein shape.

A research question is:

*Find amino acids interactions in the interface of two chain of the homodimeric protein.*

The drugs all mimic a protein chain, binding to the enzyme as protein substrate chains do. Such drug-HIV-1 protease complexes are more stable than a native protease-protein substrate complex. The HIV-1 protease cannot cleave the former, so they stay lodged to the active site, blocking the function of the enzyme (Pietrucci et al, 2015).

The protease active site comprises the residues: Arg 8, Leu23, Ap25, Gly27, Ala28, Asp29, Asp30, Val32, Lys45. Ile47, Met46, Gly48, Gly49, Ile50, PHe53, Leu76, Thr 80, Pro81, Val82, Ile84. Hydrogen bond interactions are mainly situated in the floor of active site while hydrophobic interactions from the inhibitor is targeted to the residues of I50 (I50’), I84 (I84’), and V82 (V82’) which create hydrophobic core clusters to further stabilize the flexible flaps. All inhibitors have carbon-rich groups arrayed along either side, interacting with the sides of the active site tunnel (Palese, 2017)

Two phenyl rings and one isobutyl group in DRV’s structure, a third generation inhibitor, favors strong hydrophobic contacts into the protease cavity. More over
high resolution crystal structures reveal a second active site, a surface pocket formed by one of the flaps (Zhengtong et al, 2015). DRV atoms of oxygen in a bis-tetrahydrofuran moiety, forms strong hydrogen bonding interactions with the main chain of Asp29 and Asp30 in the protease (figure 6A). DRV has two oxygen atoms at the center, that interact with a special water molecule, known as catalytic water, connecting with the flaps. It seems like a lid in the structure the tetrahedrally coordinated water molecule forms two more hydrogen bonds via the Ile50 backbone nitrogen from each subunit (Figure 4A).

Water molecules are conserved between in protease cavity and mediate distance making contacts between inhibitor carbonyl atoms and protease amino acid side chains which could not be possibly form hydrogen bonds. Other water molecules simply contribute to the integrity of architecture active cavity by hydrogen bonding to the polar atoms of Thr26, Asp29 and Arg87 (Liu et al, 2013).

The hydroxyl group of the inhibitor interacts with the carboxyl group of the protease active site residues, Asp25 and Asp25’, by hydrogen bonds (Figure 4A). Notably non polar interactions as unconventionally hydrogen bonds CH…O, CH-π and C…C interactions at distances less than 4Å are portrayed (figure 4B). These interactions are essential for enzyme inhibition. Inhibitors are differentiated through these nonpolar interactions with protease.

**Figure 4.** Active site wild protease complexed A. with DRV, PDB code: 4LL3. A.The amino acids of the protease and the ligands are in stick form with color by atom. Color code for protease: C, green; N, blue; O, red, while for ligands: C, gray; N, blue; O, red; and S, orange. Water molecules are shown as red spheres and hydrogen bonds are indicated as yellow dashed lines B. CH-π, CH-CH and CH…O interactions as lines with magenta dashed lines. Interatomic distances are shown in Å. For the sake of clarity only one conformation of DRV is shown. Figures were prepared with PyMOL.

The students determine via PyMOL the hydrophilic interactions in the active site-embedded residues and the drugs IDV, AMP and DRV. It seems that some hydrophilic interactions are conserved in all three drugs, while others are not
present in all of them. The students write down the hydrophilic interactions which are formed in the active site for inhibitor-enzyme complex. Moreover, they recognize aliphatic, aromatic carbon atoms and localize unconventional hydrogen bonds such as CH…O that participate in intermolecular interactions in protease’s cavity.

**Activity 3: Mutation**

The AIDS pandemic in the early 1980s killed almost 30 million individuals. Highly effective drugs against virus mechanism replication have been a great success for treatment. One of the main targets is the enzyme of HIV protease. However, there is a continuous quest for the design of new drugs combating virus resistance. A plethora of mutant structures have been systematically aggregated in the HIV-resistant drug database of Stanford University in order to analyze mutation mechanisms Exploring bonding interactions between inhibitors and enzyme from a molecular stand point gives insight how to treat resistance against HIV mutants (Stanford database).

Drug resistance occurs when in the presence of an inhibitor, protease may cleave the HIV polyprotein. In other words, molecular recognition between natural substrate and protease still functions, while mutation occurs at the residues which are more important for inhibitor binding.

Different pathways of protease resistance have been revealed. By thoroughly studying the wealth of crystal structures of virus protease that have been deposited in the PDB, new insights in the architecture of active cavity have been discovered (Agniswamy et al. 2016; Weber et al. 2015). Small changes in the size and shape of side chain amino acids located in active cavity leads to fewer hydrophobic Van der Waals contacts (Figure 5). The hydrophobic interactions between flap and 80s (80’s) loop residues (mainly 150-184’ and 150’-184) while play an important role in maintaining local environment of HIV-1 protease they do not participate in natural substrate recognition but rather in inhibitor binding.

The drug resistant mutations are distributed also at the dimer interface 1-4 N-terminus and 94–99 C-terminus or in the flap regions (residues 47-56). Synergestic action of many mutated amino acids remote from active cavity induce small conformational changes and the expansion of the cavity). Inhibitory molecules share structural features, so many mutations are characterized as cross resistance against several of these drugs (Zhentong et al, 2015).
Figure 5. A. Structures of wild type and mutant HIV-1 PR complex with Darunavir. Wild type (pdb code:2IEN), mutant Ile84V (pdb code:2IEO) red sticks.
for Mutant Val84, Darunavir is shown as Cyan sticks for carbon, red for Oxygen and yellow for Sulfur B. Wild-type (pdb code:3NU3) and L90M mutant protease (pdb code: 3NUO). Leu is shown with cyan sticks, splitpea dashed line symbolize C-H…O interactions. C. Wild-type (pdb code:1SDT) Ile shown with green sticks and mutant Ile50Val protease (pdb code: 2AVS), Val is shown with cyan sticks, CH…CH interactions Interatomic and intratomic interactions distances shown in Å.

Protease mutant Ile84Val is one of the major mutation which leads to a decrease in Van der Vaals contacts in inhibitor complexes (Figure 5A). Single substitutions in active site cavity residues can result in altered drug interactions while protease still maintains catalytic activity. Shortening of amino acid’s side chain leads to fewer Van der Waal’s contacts in protease–DRV interactions. This mutation confers high degree of resistance to one or more inhibitors (Weber et al, 2015).

The L90M is an example of a distal mutation with no direct contact with the substrate. The longer side chain of Met90 forms unconventional hydrogen bond C-H...O interactions with carbonyl oxygen of Asp25 that cannot form with the shorter Leu90 in wild type enzyme (Figure 5B). This creates a disorder in the active site’s hydrogen bond net which is likely to influence catalytic activity (Agniswamy et al, 2016).

Ile50 sits at the tip of the flap and interacts with the second flap in the protease dimer and also with the inhibitor. In mutant I50V makes the flap more flexible and prone to increased opening and release of inhibitor (Figure 5C). Ile50 binds via a molecule of water with the inhibitor in the active site but also contributes to the dimer stability. The I50V variant has been studied with indinavir, darunavir and saquinavir (Weber et al, 2015).

A research question is: Explore the structure with pdb-code 1RQ9. Are there any mutations? Draw in PyMOL the structure and highlight the positions of the mutations. Find multiple resistant drug HIV-protease structures in PDB. Which are the most common mutations?

4. Conclusion

This article reports an approach for teaching students in an undergraduate General Chemistry course and subsequent Biochemistry course to actively use structural information starting from small molecules to macromolecules. The students learn various visualization tools from simple molecular kits for small molecules to molecular depiction software for showing and exploring the structures of proteins, and are provided with a detailed guide in the use of these tools.

The students through this window gain access to the current research in biochemistry and drug design that centers around the study of proteins. As more and more structures become available, the knowledge gained thus results in leaps
in the understanding of other disciplines of the biomedical sciences, from animal and plant physiology to pharmacology, nutrition and extending to therapeutics. The students are encouraged to think in molecular terms about how structures enable chemical reactivity and biological function, and to experiment with different ways to convey information using structural representations.

Students literate in molecular representations certainly will benefit from such training. The students consider themselves as a part of the scientific community, by gathering and interpreting data, making measurements or observations, and interpreting these in the light of the questions asked. Modern biochemistry education must be in pace and powered with, cutting-edge research. Real time experimental data are being integrated in fundamental chemistry concepts.

Acknowledgments

The author, thanks Dr. George K. Papadopoulos, Professor of Biochemistry and Biophysics of Agriculture Department in Technological Educational Institute of Epirus for his substantial contribution in curriculum design of the Agricultural Chemistry and Biochemistry courses and his continuous guidance. Also many thanks are due to the students who attended the courses of Agricultural (General) Chemistry and Biochemistry in the academic years 2015-2019. Their involvement gave us great inspiration.

References


Chemistry for the 10th-grade general education, (in Greek), Athens, Greece: Greek Ministry of Education and Religious Affairs/Institute of Educational Policy/Diophantos, Institute of Technology, Computers, and Publications.

Chemistry for the 11th-grade general education, (in Greek), Athens, Greece: Greek Ministry of Education and Religious Affairs/Institute of Educational
Policy /Diophantos, Institute of Technology, Computers, and Publications, pp. 52–61 and 74–77

How to cite this article: