

Database and Molecular Visualisation for Learning Media of Teaching Computational Chemistry/Pharmacy in Undergraduate Level

Nirwan Syarif¹, Supriatini²

¹ Universitas Sriwijaya, Indralaya, Indonesia; nsyarif@unsri.ac.id

² Universitas Muhamadiyah, Palembang, Indonesia; supriatini.ump@gmail.com

ARTICLE INFO

Keywords:

molecular;
modelling;
computational;

Article history:

Received 2022-12-02

Revised 2023-02-17

Accepted 2023-09-19

ABSTRACT

Teaching chemistry in the information technology era has made it easy to understand topics related to the structure and properties of a molecule. One of the courses that intensively utilises information technology as a learning media is computational chemistry. This study measured student-learning outcomes during computational courses using molecular visualization software and databases. The instruments of this research are the results/answers of final exams and the assignments of the students in the semesters of 2018/2019 and 2019/2020 in the chemistry and pharmacy departments. A hundred students enrolled in these courses per year. The study was conducted for two years. The score results of six assignments for practical and four essay questions tests were analyzed. Students were asked to visualize molecules using the requested application in practical test. The assessment shows that project scores with the topic of using databases and visualization, a 'good' score profile of the results the two years. The student has 'good' score on small molecule database and visualization. The students mastered the Avogadro application more than VMD. Students generally understand to get data (computer files) from the database and then the visual constructions requested by the questions.

This is an open-access article under the [CC BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/) license.



Corresponding Author:

Nirwan Syarif

Universitas Sriwijaya, Indralaya, Indonesia; nsyarif@unsri.ac.id

1 INTRODUCTION

Today's world of education is closely related to information and communication (Fetaji et al., 2018). Information technology has transformed into something very attached to the world of education in today's modern era. Starting from computers, the internet, and other hardware and software, they are continuously being improved so that their utilisation can be perfect in the use of information technology as an educational media, such as telephone, computer, internet, e-mail, social media and databases (Fadhilah et al., 2021).

Media that supports the development of teaching technology is needed to support knowledge to understand concepts and gain skills, especially in chemistry (Gavezzotti & Presti, 2016). Non-physical learning materials such as chemical databases can describe chemical molecules in millions or even

billions. Some of these molecules are natural molecules that function as life support. It is contextual to deliver it in the class. This learning material is also very factual because everything is described based on daily experiences. This factual depiction of chemical-biological molecules is one of the characteristics of molecular modeling. Molecular visualization is built on the ideas of various experts (Kozlíková et al., 2017), such as chemists, biologists, physicists, mathematicians, computers, electronics, and even philosophers. The visuals for molecules are the wonders of thought before computers and measuring instruments were developed. Advances in computers, where computers can work quickly to process images of molecules composed of thousands of atoms with atomic orientation in three-dimension space, computer monitors are precisely matched by increasingly microscopes that can see down to the atomic level. Chemistry / chemical databases are now available online and free on internet with a wide variety. Databases containing molecular structures, such as the Protein Data Bank (PDB), ZINC, Crystallographic Online Database (COD) have been used in academia for many years, as well as the softwares, such as VMD and Avogadro (Rayan & Rayan, 2017).

Databases in education support the idea of "Freedom of Learning" because they are independent, contextual and innovative. In the other hand, visualisation in chemistry is part of the general multimedia explosion in science education, which is characterized by the design of custom visualizations to illustrate particular structures or concepts. Computer-supported databases have become valuable tools for several research laboratories in the industry and academia (Higgins et al., 2022). Applications made separately based on specific needs, are combined to provide ability to visualize molecules so that they can be scientifically accepted by students, teachers, researchers, and practitioners

Computational is an important subject in chemistry teaching and learning activities, because this course has practical values in the form of the ability or skill to make concrete abstract concepts, bring objects that are dangerous or difficult to obtain in the learning environment, can calculate complex molecular properties, can be used as a visualization and animation tool, can also calculate molecular properties and their changes as well as perform simulations on large systems and apply the program to real chemical systems.

One of the most important topics in computational is molecular visualization. This aspect is the interface between humans and computers that fundamentally communicates several other concepts used in computational courses. The visualization has been help student in studying protein molecules and understands course material more deeply in the way of improving their intuitive thinking (Argüello & DempSKI, 2020). Erica et.al (2022) reported students improvement in crystallography understanding in solid-state chemistry courses based on molecular visualization. Maia, et.al (2016) have shown that students' ability to visualize the enantiomer of a simple molecule in three dimensions improves their skill in organic chemistry. In a series of highly detailed studies, Burgin et al. (2018) found significant gains in students' understanding of simple molecular representations after using molecular visualization software.

The results of observations on the final grades of students who attend lectures are still unsatisfactory, so it is necessary to look for the causal factors of this problem. This is due to the students' lack of understanding and systematic approach to complex chemical structures and representation of molecules (Karonen et al., 2021). One attempt to overcome this problem is to apply software that describes molecules in three dimensions. Today's computer visualisation applications have even become interactive and manipulative. Users can use various representations/styles highlighting different structural features as well as zoom, pan, or rotate molecular images.

This paper describes the results of the research that has been carried out to study the application of molecular visualization and chemical databases as learning media. The media are applied in computational courses in chemistry in the 3rd year of the undergraduate in the university. Some facts can be used as consideration for taking the research. That is, the students already have a lot of basic knowledge. They received this basic knowledge as a prerequisite in taking the course and subject being studied, providing the possibility of cognition of the concept to the molecules with a larger structure.

Student learning outcomes were examined, i.e., tests, examination, and assignments related to the use of visualizations and databases in explaining the concept of molecular structures. Students' practical skills in using various presentations to portray molecular structures retrieved from the database were assessed. An analysis of the student's work in their individual projects were carried out from the assignment scores to get a definite picture of how students understand molecular visualization and databases.

The descriptive qualitative method is used in data processing by considering the characteristics of the data used and the information to be explained. The use of this method was carried out with the consideration that the analysis of qualitative descriptive research data is that this method does not use a set of pre-existing rules and does not even use statistics in answering the problem. Presentation of data involves a summary of direct explanations of information related to the data. How the data is organized depends on the researcher and how the data is presented in order to make sense. The data are arranged to have narration according to time of occurrence, category/subcategory, actual or reverse chronological order of events, the most common to the most unusual themes; moving from the broad context of an event to a narrower context or specific cases; or, describe an event from the perspectives of more than one participant. Descriptive summaries of the selected topics are arranged in the most relevant way for the audience they are intended for writing so that they can explain the relevant findings.

2 METHODS

The descriptive qualitative research was subjected to the students' learning outcomes from the Department of Chemistry and the Department of Pharmacy of Universitas Sriwijaya. This study analyzes and describes data in the form of numbers, namely test scores and assignment scores which are then used to explain student's understanding on the learning topic of course, i.e. visualization and databases. The data were the results/answers of the assignments in the project and the final test of chemistry and pharmacy students in the semesters of 2018/2019 and 2019/2020 in taking computational course. Numbers of respondents in this study was 200 students.

The results of the examination and assignments/projects in the courses were used. The project consists of two parts, namely (1) using online databases, namely protein data bank (rcsb.org), ZINC20 (zinc20.docking.org) and COD (crystallography.net). The students download the files containing the structures and (2) visualize the structures/presentations emulating the reference images. The exam consisted of four essay questions that can be used to provide an overview of the theoretical understanding of student on databases and visualization applications. The questions are (1) what is meant by modeling. What does this have to do with molecular visualization? (2) what kind of molecular model can be visualized? (3) Describe molecular databases and their application? (4) What is the role of molecular visualization in databases?

The Avogadro provides the styles/presentations for the molecule so that the appearance/visualization of the molecules is the same as the reference images (Figures 1). Full marks are given if all molecule presentations constructed by student and the reference images are similar. Total number of presentations to be done by students were six. The students' practical understanding of databases and molecular visualization are assessed from the assignment.

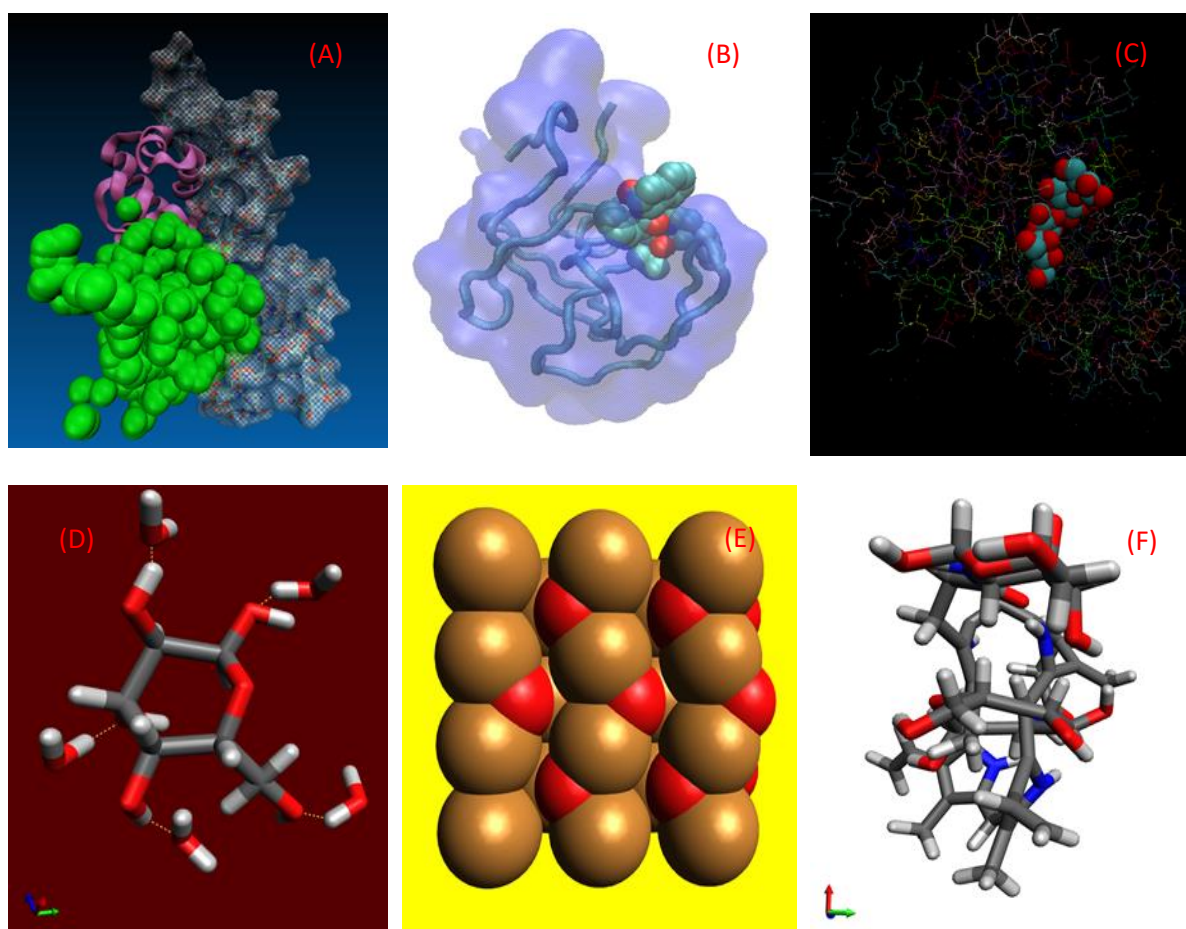


Figure 1. Reference image of the downloaded files (A – C) from the PDB and was completed using VMD; the files (D – F) from ZINC20 and COD were performed with the Avogadro. The images from three PDB files are (A) 3cro, (B) 1hsg, and (C) 1fqa respectively. The three images are (A) Porphyrin from ZINC169366333, (B) Galactose from ZINC901155 and (C) water molecules from COD 1000063 CIF.

The files were downloaded from PDB (rcsb.org), three files, from ZINC20 (zinc.docking.org), two files and from COD (crystallography.net), one file. In the first part to assess students' practical skills in using databases and molecular visualisation. There were six styles that students have to construct copying the reference images. Figures 1 (A – C) show structures from PDB files with visualisation / style varied. Figures 1 (D – F) are provided with screenshots of the application that are used as references for constructing structures of relatively small molecules taken from ZINC20 and COD. The reference images were given as a guide in doing the task. The students were expected to be able to do using visualisation application. i.e., the VMD and Avogadro.

Students' theoretical understanding was assessed from the answers to the final semester exam. Students' answers were categorized into three, namely: 'good', 'fair' and 'low'. The 'low' category was given if the student gives an answer, such as, "I don't know" or it is empty. The calculation of test scores was carried out as in the molecules.

The assessment was carried out according to the scoring method developed by Pianta et al. (2008). The score of each question were calculated from the visual components appear on the molecule presentation generated by the students. The scores for practical test, i.e., six molecule representation as stated in Figure 6, were calculated using the formula.

$$S = \frac{V_i}{V_T + D} \times 100$$

$$\bar{S} = \frac{\sum S}{6} \quad (1)$$

Where S, V, V_T , D, and \bar{S} were score per file, visual component appeared by student, total visual component provided by reference, database retrieving, and average score, respectively in one reference image. The score for the theoretical tests were calculated by the formula, where the total questions were four.

$$S = \frac{C}{C_T} \times 100$$

$$\bar{S} = \frac{\sum S}{4} \quad (2)$$

Where C and C_T were correct answer by student and reference total correct answer, respectively in one question.

3 FINDINGS AND DISCUSSION

This study focuses on two problems: first, how was the students' understanding of molecular visualization, especially its definition, types, and uses; second, how was the students' knowledge of the database, its usefulness, and its relation to molecular visualization, theoretical and practically. We examined the effect of the variables, i.e., the knowledge and time of course enrollment on students' level of understanding. Detailed data from students' answers in both semesters, practically and theoretically are presented below. The results of analysis for students' assessment of each of questions are discussed individually as sub-chapters.

3.1 Practical Tests

Students were asked to represent molecules using the requested application to measure students' perceptions of the usefulness of visualization software. The project is considered 100% complete if the presentation made follows the reference image provided. Figure 2 shows (A) the assessment results on project scores with the topic of using databases and visualization, (B) a 'good' score profile of the results of the two years. We can see from the plot (Figure 2A) that 53% and 67% of student has a 'good' score on ZINC COD database and Avogadro visualisation. These scores have not changed much for two years (Figure 3B). The profiles for VMD and PDB are relatively lower, indicating that students are more familiar with small molecule databases with more straightforward visualization. The interface on the search page and results look more straightforward and intuitive, showing that mastery of the application is followed by knowledge of the database. Databases for small molecules, such as those of ZINC20 and COD, do not use the complex interfaces of VMD. Asking students to use a ZINC20 or COD database and then displaying the structure required some additional instructions beside written instructions from the manual. This is different from PDB which is relatively easier and straightforward. Avogadro's interface looks simpler than VMD's. The features in the VMD interface are far more numerous with more diverse explanations, so it is not easy to understand. The students only need to master the application's practical materials to understand basic database concepts and molecular visualisations.

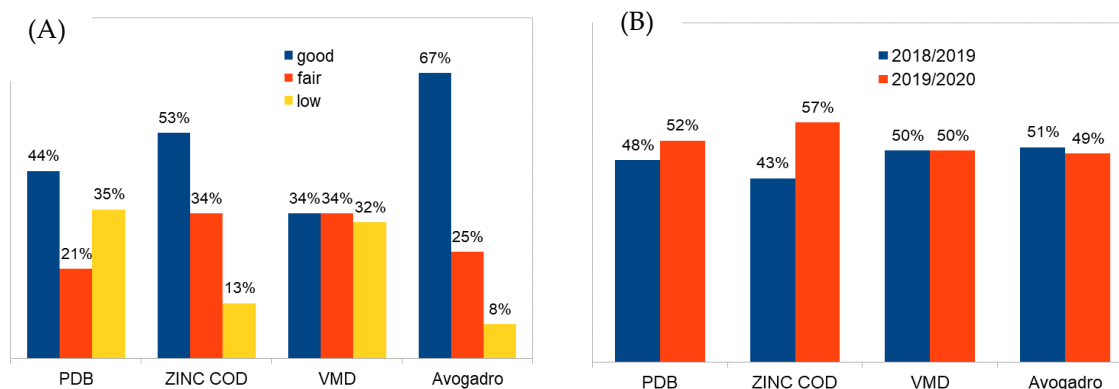


Figure 2. (A) The results of data processing for the tests with the topic of using databases and visualizations, (B) a 'good' score profile for the exams for two years

Students mastered the Avogadro application more than VMD. Students generally understand how to get data (computer files) from a database and then the visual constructions requested by the questions. Molecular structure construction involves building a physical model from a crystal structure model kit (Simmons et al., 2016). The instructions also guide students to use a web-based unit cell visualization tool and intersperse the activity with short segments of hands-on instruction using several drawn unit cell representations.

This course was delivered offline and conducted in a computer laboratory with 50 students facilitated by computers, networks, and the internet. Lectures and labwork were integrated, where the theory was delivered before the labwork was carried out. Access to the database can be done directly from the available browser because the databases used are open-access, namely PDB, ZINC, and COD. Each student is guided to download the applications used, namely VMD and Avogadro, for molecular visualization tools. The teaching is carried out after the application is installed on the computer used by students. The project consists of several steps, namely: searching for files in the database, downloading files, visualizing the molecular structure on a local computer, and styling parts of the molecule to match the given molecular model.

A visualisation is an approach that uses multiple representations in developing visuo-spatial thinking in chemistry (Kiernan et al., 2021), especially for the student. The teachers are expected to be able to use molecular visualization tools as well as using physical molecular kits. This visualization application provides an interactive way to view and dynamically manipulate multiple representations of molecular structures. The student can also use this application in other courses to adjust the existing curriculum and activities. For this reason, The assignments of students taking part in computational chemistry/pharmacy course to work on the final project were assessed. The assessment results were broken down into three categories: 'good', 'fair', and 'low'.

3.2 What is meant by modeling?

Figure 3 shows the assessment results of the theoretical test. The questions related to the databases and visualization were conducted in essays (first question in Figure 3). We can see in Figure 3A that the students have a 'fair' answer on the definition of modeling and databases. Exact solutions were obtained from students answering more or less the same as what was described previously, accompanied by examples of images that visualize molecular models both in 2D and 3D. This question then relates to the type of molecular model displayed in software such as Avogadro and VMD.

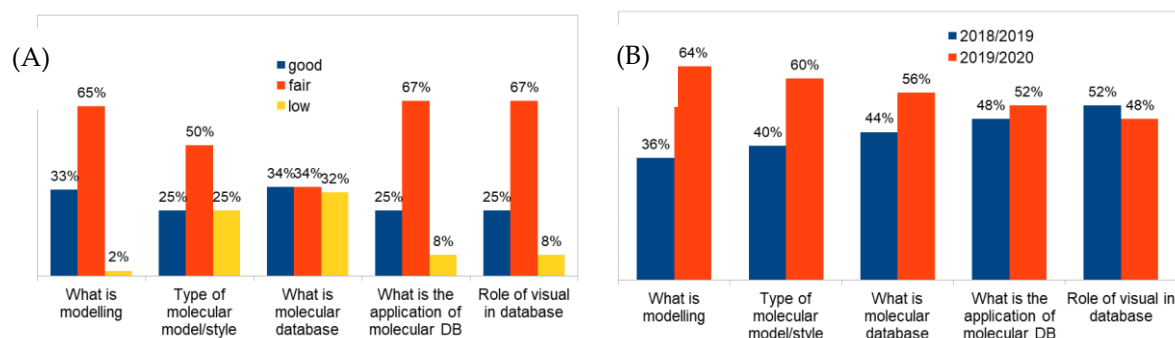


Figure 3. A) The results of data processing of practical assignments on the topic of using databases and visualization, (B) a 'good' score profile for the exams for two years

The results of data analyses show that most of students could not answer the questions that only provide general answers to molecular models that are often exemplified in lectures for both 2D and 3D. As previously explained, molecular models are not only for visualizing molecules but can also be used to describe specific molecule characteristics. Structural characters of molecules such as enzymes can be explained not only by using this general visualization but also by using cartoon visualization. Visualization with cartoon style for enzymes makes it easier for students to determine the dominant structure of enzymes. Students can also understand the concept of extended structure in crystals, determine the formula for ionic compounds from the unit cell, and determine coordination numbers.

When the complete answer to the question What is meant by modelling? What does this have to do with molecular visualization? The most frequent answer given by students in the two years was 'fair', 65%, and the 'good' answer, 33%. For student results in 2018, the 'good' answer is around 36%, while in 2019 is 64%. In particular, the emphasis on 'fair' answers in 2019 was reflected in the complexity and diversity of student answers, as measured by the average number of categories in each student's responses, which increased significantly by 32%.

As an exact answer to the question above that, modeling is a way to represent or describe or explain an object, actual situation (reality), process, system, or concept, simplified by only paying attention to the essential and ignoring factors that are deemed unimportant in the study carried out or idealization. Molecular modeling is an idealization of chemical compound molecules. It took more than a hundred years to describe the molecule. The molecule of a chemical compound itself was originally an abstract entity. Most of the students gave precise answers regarding the definition of molecular structures. Students' responses are interfered with by the definition of mathematical models, which are also studied in other courses, such as mathematical models and physical models. Mathematical models, as used in teaching kinetics, are formulas that contain numbers and symbols so that they become equations and inequalities. While the physical model is photos, pictures, paintings, and others that are physical.

3.3 What kind of molecular model can be visualised?

This question is related to second question in Figure 3, i.e., type of molecular model/style. Figure 3A shows the students have fair understanding about type of model/style can be modelled to visualise molecule. This question demands more knowledge of the scope of molecular visualization. The results of processing test score data for the question "what kind of molecular model can be visualized?" shows that the answer most often given by students in two academic years is "fair." There is an increasing in 'good' answer in second year, i.e., for student results in 2018, the 'fair' answer is 40%, while in 2019 is 60%. With the growing understanding of molecular structures, the a need for new visualizations. Molecular visualisation supports knowledge for students for material characteristics, and interactions by creating molecular structures. The benefit is that it helps the 'rational' design of new molecules, for

example, for catalysts or drugs. Visualising biological molecules, such as enzymes, focuses on graphical depictions of these biomolecules' structure, interactions, and functions. Students in practice have to select a particular representation for good predictive performance call. The existing representations encode different information.

Visualization of molecular structures, in some ways, shows particular characteristics that are physically more obvious than atomistic representations. In other words, the visual styles complement each other depending on the details of the character of the molecule to be explained (Tate, 2005). Multiple representations were needed to link to a single molecular structure as shown in reference image (Figure 1). For example, molecular weight or polarity can correlate well with other properties, such as boiling point, but are not very precise when associated with complexly bonded molecules. Thus, most molecular representations encode information optimized for a particular use.

3.4 What is molecular database?

This question is related to the third question in Figure 3, i.e., "What is the molecular database?" The result of data analysis in Figure 3A show that all answer categories are relatively the same". For student results in 2018, the 'good' answer is 44%, while in 2019 is 56%. The explanation for the molecular database is relatively hard to describe to the student. This profile is caused by the broad scope or range of the database. The databases in chemistry are very diverse for such educational resources supporting molecular explorations (Zardecki et al., 2022). This question also support practical test (Figure 1 (A – B)) in giving description about student competence on database.

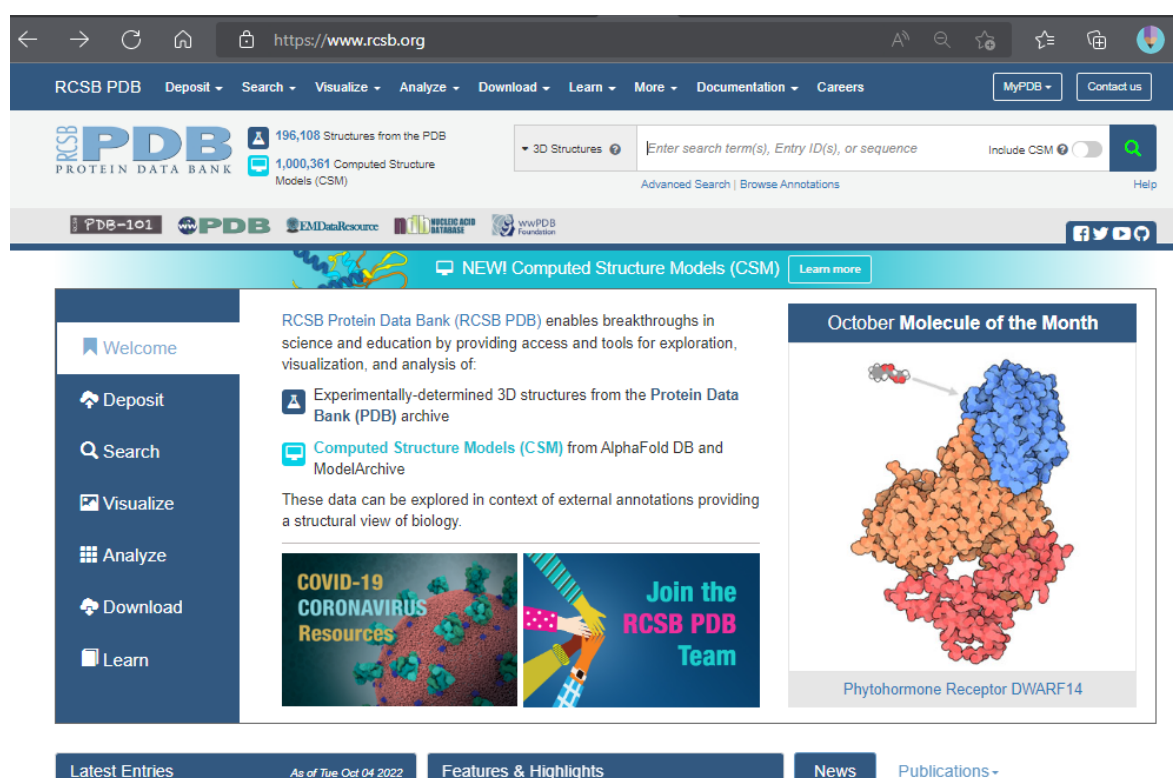


Figure 4. PDB main page on rcsb.org; accessed October 7, 2022

The result shows that the student has average ability and cognition on the database. Database for large molecules such as PDB (rcsb.org) (Figure 4) provides information for future study of the student. The search on the database is carried out based on potential requirements for chemical, physical and biological activity, namely by identifying compounds/molecules that are suitable to be continued in the in-silico research stage. The database stores molecule structures such as enzymes and DNA, RNA and

biomembranes. Enzymes are hosts as molecular targets for ligands other than natural substrates. There are known in-vitro and in-vivo stages within the scope of drug or medical research or the use of natural ingredients that is very important for the student in doing drug design (Edwards et al., 2021).

ZINC20

Welcome to ZINC, a free database of commercially-available compounds for virtual screening. ZINC contains over 230 million purchasable compounds in ready-to-dock, 3D formats. ZINC also contains over 750 million purchasable compounds you can search for analogs in under a minute.

ZINC is provided by the [Irwin](#) and [Shoichet](#) Laboratories in the Department of Pharmaceutical Chemistry at the University of California, San Francisco (UCSF). We thank [NIGMS](#) for financial support (GM71896).

To cite ZINC, please reference: Irwin, Tang, Young, Dandarchuluun, Wong, Khurelbaatar, Moroz, Mayfield, Sayle, *J. Chem. Inf. Model* 2020, in press. <https://pubs.acs.org/doi/10.1021/acs.jcim.0c00675>. You may also wish to cite our previous papers: Sterling and Irwin, *J. Chem. Inf. Model*, 2015 <http://pubs.acs.org/doi/abs/10.1021/acs.jcim.5b00559>. Irwin, Sterling, Mysinger, Bolstad and Coleman, *J. Chem. Inf. Model*, 2012 DOI: [10.1021/ci3001277](https://doi.org/10.1021/ci3001277) or Irwin and Shoichet, *J. Chem. Inf. Model*. 2005;45(1):177-82 PDF, DOI.

Getting Started

- [Getting Started](#)
- [What's New](#)
- [About ZINC 20 Resources](#)
- [Current Status / In Progress](#)
- [Why are ZINC results "estimates"?](#)

Explore Resources

Chemistry
[Tranches, Substances, 3D Representations](#),
[Rings, Patterns](#)
And More
[Catalogs, Genes, ATC Codes](#)

Ask Questions

You can use ZINC for **general** questions such as

- How many substances in current clinical trials have PAINS patterns? (150)
- How many natural products have names in ZINC and are not for sale? (9296) [get them as SMILES, names and calculated logP](#)
- How many endogenous human metabolites are there? (47319) and how many of these can I buy? (8271) How many are FDA approved drugs? (94)
- How many compounds known to aggregate are in current clinical trials? (60)

ZINC20 News

- ZINC20 has been released

Caveat Emptor: We do not guarantee the quality of any molecule for any purpose and take no responsibility for errors arising from the use of this database. ZINC is provided in the hope that it will be useful, but you must use it at your own risk.

Figure 5. ZINC main page currently version 20 at zinc20.docking.org; accessed October 7, 2022

In contrast with PDB, the databases for small molecules such as ZINC20 in zinc20.docking.org (Figure 5) provide templates for designing ligands or drug molecules that are substrates for enzymes (Irwin et al., 2020). The molecular structure of ZINC20 is treated as a template for the design of other molecules. A hundreds of different molecules are formed from one template molecule, distinguished from their functional groups.

Student relatively has little knowledge about ZINC20 compared to PDB. ZINC's original focus was primarily to serve data needs on molecular docking with an analog-by-catalog approach and molecular similarity principles. The database found to be helpful to search for analogs in real time, where "what if" questions about compound availability can be explored. Students can use the database to identify analogs of compounds with a structure-activity relationship approach.

Student, in advance, used these of molecules from both database with in-silico experiment. In their parent molecules, enzymes. The interaction between the ligands, i.e. small molecule from ZINC20 and the enzyme, i.e. macromolecule from PDB was assessed using several parameters to conclude that some

of the best ligand structures were said to be 'prospective' drugs. These best molecular structures are then feasible to be synthesized and enter the next testing stage, in-vitro and in-vivo tests.

Learning key patterns from complex observations is practical to use in visualising molecules. Students can develop many skills related to visualise the molecule. This ability is developed from the bond character, bond form, bond distance, bond angle, and the molecular plane of a molecule. Their values predict its interaction with large molecules and relate it to its biological function, such as hydrophobicity/lipophilicity, three-dimensional geometry, surface electric charge, molecular contours, reaction sites, polarity, and others. The student can use it within the range of teaching to explain the relationship between molecular structure and physical phenomena or even experimental measurement results.

ZINC provides downloadable 2D and 3D versions and a website that enables fast molecular and analogue searches (Figure 6). ZINC already holds data on about two billion compounds on its servers. Users use ZINC by users is handled an interface that is always dynamic from time to time to respond to advanced user needs, storage growth, and device developments. Thousands of users use the ZINC website every month, including researchers, developers, students, and teachers. The exchange of data between users and servers can be seen through the number of terabytes of data downloaded every week. The students can use the statistics of "Usage" via the "About" menu followed by "Usage" at the top of each ZINC page. Searching the database based on potential requirements for biological activity (Figure 6A) or chemical structures (Figure 6B) makes it easy to identify specific compounds/groups of compounds. Accessing chemical information on ZINC by utilising computer facilities has become more significant. The dimensions of all databases in terms of structure or reaction proliferate yearly. Therefore, computer-supported databases have become valuable tools for several research laboratories in the industry and academia.

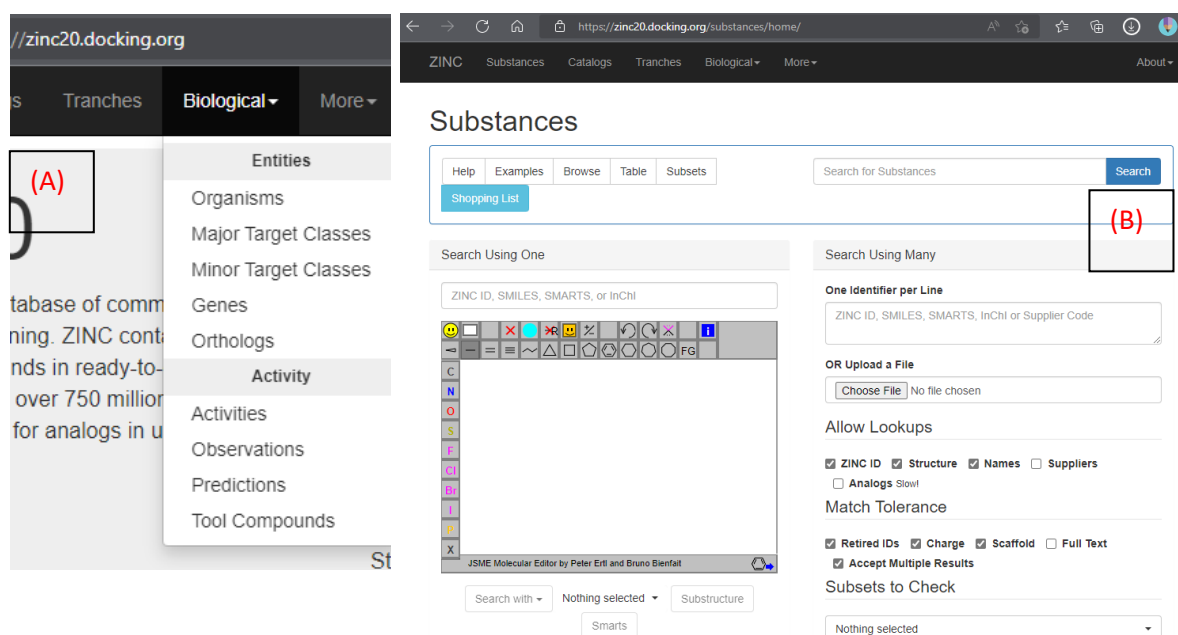
The image shows a screenshot of the ZINC20 docking.org website. On the left, a navigation menu is open, with 'Biological' selected. A red box labeled '(A)' highlights the 'Biological' menu item. The main content area is titled 'Substances' and features a search bar with a 'Search' button. A red box labeled '(B)' highlights the search bar. Below the search bar, there are two search options: 'Search Using One' and 'Search Using Many'. The 'Search Using One' option has a text input field for 'ZINC ID, SMILES, SMARTS, or InChI'. Below this is a JSME Molecular Editor. The 'Search Using Many' option has a text input field for 'ZINC ID, SMILES, SMARTS, InChI or Supplier Code'. There are also checkboxes for 'Allow Lookups' (ZINC ID, Structure, Names, Suppliers) and 'Match Tolerance' (Retired IDs, Charge, Scaffold, Full Text). At the bottom, there are buttons for 'Search with', 'Nothing selected', and 'Substructure', along with a 'Smarts' button.

Figure 6. Search facilities available on ZINC

Figure 7. COD main page on crystallography.net; accessed October 7, 2022

Relatively unfamiliar database for the student is COD (crystallography open database) on www.crystallography.net (Figure 7). The results of test and assessment show that the students have lowest cognition and practical ability in using COD. This database provides the coordinates of the atoms that make up the crystal structure of organic, inorganic, metal-organic, and mineral materials. The molecular structure stored in this database results from X-ray diffraction measurements of single crystals or powders, gamma photons, neutrons, electrons, or other particles.

3.5 What is the application of molecular database?

The fourth question is “What is the application of molecular database?” I have “fair” answer from the student score (Figure 3) assessment and have slightly increasing in the second year. Data in Figure 3 shows that all answer categories in relatively the same. For student results in 2018, the 'good' answer is 48%, while in 2019, it is 52%. This data infers that more concrete examples of database applications have to provided so that students do not find it too difficult to provide examples for the application of molecular databases.

The databases provide real-life and high-volume examples from scientific databases to formulate challenging data mining and processing tasks. The open nature database allows downloading this database or parts of it to any number of computers, thereby making use of many parallel resources available in grid/volunteer computing environments. The database contains information about the experimental conditions (temperature, pressure, or sample history) for each molecular structure. The unit cell is one of the input data when the user searches the database for comparison with the unit cell, which is also the identity of the molecular data stored in the database. Chemical formulas are checked for exact matches, and both formulas provided by users and calculated from the contents of the crystal structure (atomic coordinates, crystal density and molecular weight) are compared with each other (resulting in four comparisons if the formula provided by the author is in both files). The range of compounds included in COD allows the use of COD for material identification using powder

diffraction patterns and powder patterns calculated theoretically from COD entries (Sutrisno et al., 2018). Even without 100% coverage of all known compounds (and elements).

There are fundamental differences between the search results pages of the three databases, primarily related to the visualization displayed. The page on PDB displays a 3D molecular structure, while COD does not display any structure at all. The difference in the appearance of the database is related to the application of the database, the value of the information contained in the database, and the effectiveness of the presentation of the data itself. 3-dimensional molecular visualization on PDB because the data displayed is macromolecule which will not be effective when displayed in 2-dimensional format.

3.6 What is the role of molecular visualisation in databases?

The answer for the role of molecular visualisation in databases is a 'fair', 67% given by students as the fifth question. Data in Figure 3 on the question " shows that all answer categories in relatively the same in consecutive two years. For student results in 2018, the 'good' answer is 52%, while in 2019, it is 48%. This data infers that the teacher provides more concrete examples of database applications so that students do not find it too difficult to provide examples for the application of molecular databases.

Visualisation in chemistry is part of the general multimedia explosion in science education, which is characterized by the design of custom visualizations to illustrate particular structures or concepts. However, the linkage between scientific concepts and visual representation is often not profoundly principled or algorithmic. General principles inform better illustrations of good visual representation and graphic design. In chemistry, however, a unified theory is used to make quantitative spatial predictions, which allows highly generalized rules to link theoretical concepts to visual displays. Expanded structural formulas that show the connectivity within a molecule by lines connecting the elemental symbols of the atoms are an example of effective visualisation in which a set of diagrammatic rules produces qualitatively informative diagrams of an unlimited range of molecules. The visual variety of the database is associated with the application of the database (Hossain, 2019).

Crystallographic data combined with computational power drive the design of representations and the visualization of three-dimensional structures. The visualisation provides a general-purpose representation designed for molecule properties prediction. The development of visualisation prompted the re-optimization of the representation of molecular on accessing the databases. This feature creates new opportunities to utilise spatial information for molecular studies, such as quantitative structure – activity relationship and similarity analysis (Baskin & Zhokhova, 2019). Several accessible small molecule collections are available and provide start-up for the studies. Despite containing the three-dimensional structure, there also provide data, such as tautomeric, protonate, and partial charging in molecules. Virtual filtering libraries for molecules can be constructed from the combined information of 2D – 3D structures.

4 CONCLUSION

Visualization in chemistry is part of the general multimedia explosion in science education, which is characterized by the design of custom visualizations to illustrate particular structures or concepts. The better illustrations are informed by general principles of good visual representation and graphic design, but the linkage between scientific concepts and visual representation is often not deeply principled or algorithmic. Molecule visualisation and databases greatly influence the learning process. The course at the undergraduate level generally do not discuss these two aspects, so this research was conducted to see the extent to which the results/success of special learning regarding two topics that have just been introduced and have been applied for the past 2 years. In line with technological advances and the needs of the world of work, computing courses must accommodate these two aspects. Information technology related to the two aspects used in teaching computing courses has never been studied for its usefulness as a learning medium. The examination of students' cognition of the concept

of molecules, and also the effects of expanding various representations/styles of the structure originating from the database and their practical skill, is reported.

The results, in general, show us the application of molecular visualization at higher education levels. Where learning media is applied at the undergraduate level, namely in their second - third year of education, where students have received a lot of basic knowledge. They received this basic knowledge as a prerequisite in taking the subject being studied, namely computational. The existence of higher levels of education provides the possibility of applying the concept to more complex molecules. The project scores with the topic of using databases and visualization, a 'good' score profile of the results the two years. The student has good score on ZINC, COD database and Avogadro. These scores have not changed much for two years in a row. Most of the students gave precise answers regarding the definition of molecular structures and database. There are fundamental differences between the search results pages of the three databases, primarily related to the visualization displayed. The interface on the search page and results look more straightforward and intuitive, showing that mastery of the application is followed by knowledge of the database. Databases for small molecules, such as those of ZINC20 and COD, do not use the complex interfaces of VMD. Students mastered the Avogadro application more than VMD. Students generally understand how to get data (computer files) from a database and then the visual constructions requested by the questions.

The results of the study showed that the students' understanding of the topics varied. This is due to the short exposure of the two topics and the lack of training applied to students. As a consideration for these two topics, each is divided into 2 subjects. Students must be able to master each of the two software and two databases. This provides a basis for considering the topics given in this lectures, for example, reducing the topic of discussion or the difficulty level for each topic / subject.

Acknowledgements: DIPA Badan Layanan Umum Universitas Sriwijaya 2020 funded the writing of this article with SP DIPA-023.17.2.677515 / 2020, revision 01, date March 16 2020. In accordance with the Rector's Decree No: 0685 / UN9 / SK. BUK.KP/2020/ on date July 15 2020.

Conflicts of Interest: The authors declare no conflict of interest..

REFERENCES

- Argüello, J. M., & Dempski, R. E. (2020). Fast, Simple, Student Generated Augmented Reality Approach for Protein Visualization in the Classroom and Home Study. *Journal of Chemical Education*, 97(8), 2327–2331. <https://doi.org/10.1021/acs.jchemed.0c00323>
- Baskin, I. I., & Zhokhova, N. I. (2019). Continuous molecular fields and the concept of molecular co-fields in structure–activity studies. *Future Medicinal Chemistry*, 11(20), 2701–2713. <https://doi.org/10.4155/fmc-2018-0360>
- Bird, C. L., & Frey, J. G. (2016). From Big Data to Chemical Information. *Chemistry International*, 38(3–4). <https://doi.org/10.1515/ci-2016-3-424>
- Burgin, S. R., Oramous, J., Kaminski, M., Stocker, L., & Moradi, M. (2018). High school biology students use of visual molecular dynamics as an authentic tool for learning about modeling as a professional scientific practice: Authentic Tool for Learning About Modeling as a Professional Scientific Practice. *Biochemistry and Molecular Biology Education*, 46(3), 230–236. <https://doi.org/10.1002/bmb.21113>
- Edwards, T., Foloppe, N., Harris, S. A., & Wells, G. (2021). The future of biomolecular simulation in the pharmaceutical industry: What we can learn from aerodynamics modelling and weather prediction. Part 1. understanding the physical and computational complexity of *in silico* drug design. *Acta Crystallographica Section D Structural Biology*, 77(11), 1348–1356. <https://doi.org/10.1107/S2059798321009712>
- Fadhilah, F. D., Harahap, F. H., Nur Zarith Sofia, Ihsan, M. T., & Prayoga, S. (2021). The Utilization of Information Technology as Learning Media. *Jurnal Riset Dan Inovasi Pembelajaran*, 1(2), 164–173. <https://doi.org/10.51574/jrip.v1i2.48>

- Fetaji, B., Fetaji, M., Ebibi, M., & Kera, S. (2018). Analyses of Impacting Factors of ICT in Education Management: Case Study. *International Journal of Modern Education and Computer Science*, 10(2), 26–34. <https://doi.org/10.5815/ijmecs.2018.02.03>
- Gavezzotti, A., & Presti, L. L. (2016). Building Blocks of Crystal Engineering: A Large-Database Study of the Intermolecular Approach between C–H Donor Groups and O, N, Cl, or F Acceptors in Organic Crystals. *Crystal Growth & Design*, 16(5), 2952–2962. <https://doi.org/10.1021/acs.cgd.6b00305>
- Gražulis, S., Chateigner, D., Downs, R. T., Yokochi, A. F. T., Quirós, M., Lutterotti, L., Manakova, E., Butkus, J., Moeck, P., & Le Bail, A. (2009). Crystallography Open Database – an open-access collection of crystal structures. *Journal of Applied Crystallography*, 42(4), 726–729. <https://doi.org/10.1107/S0021889809016690>
- Heller, S. R., McNaught, A., Pletnev, I., Stein, S., & Tchekhovskoi, D. (2015). InChI, the IUPAC International Chemical Identifier. *Journal of Cheminformatics*, 7(1), 23. <https://doi.org/10.1186/s13321-015-0068-4>
- Higgins, S. G., Nogiwa-Valdez, A. A., & Stevens, M. M. (2022). Considerations for implementing electronic laboratory notebooks in an academic research environment. *Nature Protocols*, 17(2), 179–189. <https://doi.org/10.1038/s41596-021-00645-8>
- Hossain, S. (2019). *Visualization of Bioinformatics Data with Dash Bio*. 126–133. <https://doi.org/10.25080/Majora-7ddc1dd1-012>
- Irwin, J. J., Tang, K. G., Young, J., Dandarchuluun, C., Wong, B. R., Khurelbaatar, M., Moroz, Y. S., Mayfield, J., & Sayle, R. A. (2020). ZINC20—A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *Journal of Chemical Information and Modeling*, 60(12), 6065–6073. <https://doi.org/10.1021/acs.jcim.0c00675>
- Karonen, M., Murtonen, M., Södervik, I., Manninen, M., & Salomäki, M. (2021). Heuristics Hindering the Development of Understanding of Molecular Structures in University Level Chemistry Education: The Lewis Structure as an Example. *Education Sciences*, 11(6), 258. <https://doi.org/10.3390/educsci11060258>
- Kiernan, N. A., Manches, A., & Seery, M. K. (2021). The role of visuospatial thinking in students' predictions of molecular geometry. *Chemistry Education Research and Practice*, 22(3), 626–639. <https://doi.org/10.1039/D0RP00354A>
- Kozlíková, B., Krone, M., Falk, M., Lindow, N., Baaden, M., Baum, D., Viola, I., Parulek, J., & Hege, H.-C. (2017). Visualization of Biomolecular Structures: State of the Art Revisited: Visualization of Biomolecular Structures. *Computer Graphics Forum*, 36(8), 178–204. <https://doi.org/10.1111/cgf.13072>
- Llanos, E. J., Leal, W., Luu, D. H., Jost, J., Stadler, P. F., & Restrepo, G. (2019). Exploration of the chemical space and its three historical regimes. *Proceedings of the National Academy of Sciences*, 116(26), 12660–12665. <https://doi.org/10.1073/pnas.1816039116>
- Pianta, R. C., La Paro, K. M., & Hamre, B. L. (2008). *Classroom Assessment Scoring System™: Manual K-3*. Paul H. Brookes Publishing Co.
- Popova, M., Bretz, S. L., & Hartley, C. S. (2016). Visualizing Molecular Chirality in the Organic Chemistry Laboratory Using Cholesteric Liquid Crystals. *Journal of Chemical Education*, 93(6), 1096–1099. <https://doi.org/10.1021/acs.jchemed.5b00704>
- Rayan, B., & Rayan, A. (2017). Avogadro Program for Chemistry Education: To What Extent can Molecular Visualization and Three-dimensional Simulations Enhance Meaningful Chemistry Learning? *World Journal of Chemical Education*, 5(4), 136–141. <https://doi.org/10.12691/wjce-5-4-4>
- Simmons, C. R., Zhang, F., Birktoft, J. J., Qi, X., Han, D., Liu, Y., Sha, R., Abdallah, H. O., Hernandez, C., Ohayon, Y. P., Seeman, N. C., & Yan, H. (2016). Construction and Structure Determination of a Three-Dimensional DNA Crystal. *Journal of the American Chemical Society*, 138(31), 10047–10054. <https://doi.org/10.1021/jacs.6b06508>

- Stella, E., Agosti, I., Di Blas, N., Finazzi, M., Lanzi, P. L., & Loiacono, D. (2022). A virtual reality classroom to teach and explore crystal solid state structures. *Multimedia Tools and Applications*. <https://doi.org/10.1007/s11042-022-13410-0>
- Sutrisno, H., Ariswan, A., & Purwaningsih, D. (2018). Qualitative and Quantitative Phase-Analysis of Undoped Titanium Dioxide and Chromium Doped Titanium Dioxide from Powder X-Ray Diffraction Data. *Indonesian Journal of Chemistry*, 18(3), 486. <https://doi.org/10.22146/ijc.28892>
- Tate, J. (2005). Molecular Visualization. In P. E. Bourne & H. Weissig (Eds.), *Methods of Biochemical Analysis* (pp. 135–158). John Wiley & Sons, Inc. <https://doi.org/10.1002/0471721204.ch7>
- Zardecki, C., Dutta, S., Goodsell, D. S., Lowe, R., Voigt, M., & Burley, S. K. (2022). PDB -101: Educational resources supporting molecular explorations through biology and medicine. *Protein Science*, 31(1), 129–140. <https://doi.org/10.1002/pro.4200>