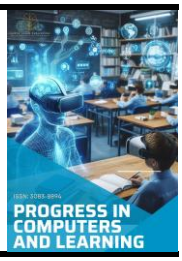
Progress in Computers and Learning

Journal homepage:
https://karyailham.com.my/index.php/picl
ISSN: 3083-8894



MolEdu: Empowering Chemistry Education with Open-Source Cheminformatics for Molecular Drawing

Shafidah Shafian[1,*]

[1] Solar Energy Research Institute, Universiti Kebangsaan Malaysia, Bangi, Selangor 43600, Malaysia

ARTICLE INFO

Article history:
Received 15 April 2025
Received in revised form 4 May 2025
Accepted 8 June 2025
Available online 24 July 2025

Keywords:
RDKit; molecular drawing; 2D and 3D; chemistry

ABSTRACT

Understanding molecular structures is a fundamental aspect of learning chemistry, as it forms the basis for grasping concepts such as functional groups, stereochemistry, and chemical reactivity. Traditional teaching tools often rely on static textbook images or proprietary software that may limit user accessibility and interaction. Many users, particularly beginners, may struggle with drawing complex molecules or interpreting structural representations without guided support. To address this challenge, this research introduces MolEdu, an open-source educational application that enables users to input chemical formulas or SMILES strings and receive immediate visual feedback in both two-dimensional (2D) and three-dimensional (3D) formats. The purpose of this study is to evaluate the effectiveness of MolEdu in supporting molecular visualization and conceptual understanding in chemistry education. The application automatically retrieves molecular data, generates the corresponding structural representation, and displays key molecular properties such as IUPAC name, molecular formula, and molecular weight. Through the integration of interactive visualization and automated chemical analysis, users can explore molecular geometry without the need to manually sketch structures. The results indicate that MolEdu enhances user engagement, improves accuracy in interpreting molecular structures, and supports inquiry-based learning. Additionally, MolEdu provides a scalable, accessible, and interactive platform that empowers users to explore and understand chemical structures, promoting a deeper appreciation for molecular science in educational contexts.

1. Introduction

Understanding molecular structures is central to learning chemistry, as it enables learners to comprehend structure–property relationships, stereochemistry, and reaction mechanisms. Conventional approaches to teaching molecular structure, such as textbook illustrations and physical model kits, while effective, are often limited by static representations and lack of interactivity. In recent years, the use of digital tools has gained traction in chemical education, offering dynamic and visual approaches to support conceptual understanding [1].

* *Corresponding author.*
E-mail address: norshafidah@ukm.edu.my

https://doi.org/10.37934/picl.4.1.4958

To bridge this gap, open-source cheminformatics platforms have emerged as promising solutions that enable the visualization, manipulation, and analysis of molecular structures. One such platform is RDKit, a widely adopted open-source cheminformatics toolkit designed for working with chemical structures and data [2,3]. RDKit allows for the interpretation of chemical identifiers such as SMILES (Simplified Molecular Input Line Entry System) [4,5] and molecular formulas, generation of two-dimensional (2D) and three-dimensional (3D) molecular representations, and computation of chemical descriptors. Its broad functionality and accessibility make it suitable not only for research but also for educational purposes, especially in enhancing students' or users' understanding of molecular concepts through visual and data-driven feedback. Studies in chemical education have shown that interactive 3D visualization can significantly improve students' understanding of molecular geometry and spatial reasoning [6].

This study introduces MolEdu, an interactive, open-access application that allows users to input a chemical formula or SMILES string and instantly view the corresponding 2D and 3D molecular structures, along with basic chemical information. By eliminating the need for manual drawing and offering automated molecular rendering and analysis, MolEdu supports a more intuitive and inclusive approach to learning chemistry. This paper presents the design and functionality of MolEdu, outlines how it leverages open cheminformatics capabilities to facilitate molecular visualization, and discusses its potential to improve engagement and learning outcomes in chemistry education.

2. Methodology

2.1 System Overview

MolEdu is designed as a web-based molecular visualization tool that enables users to input either molecular formulas or SMILES (Simplified Molecular Input Line Entry System) notations [5]. The platform translates these inputs into structural representations and displays both 2D and 3D molecular visualizations along with computed chemical properties. The system is structured to support educational engagement without requiring prior drawing skills or installation of specialized software.

2.2 Input Processing

User input is accepted in the form of either a molecular formula (e.g., C2H6O) or a SMILES string (e.g., "CCO"). Upon submission, the system performs input validation and classification. If the input is recognized as a SMILES string, it is parsed directly into a molecular structure using cheminformatics parsing techniques. If the input is a molecular formula, the system performs a structural lookup against publicly accessible compound databases to identify a corresponding molecule and retrieve its canonical structure.

2.3 Structure Generation and Optimization

Following successful input recognition, MolEdu generates a 2D structural layout using established chemical rendering algorithms. Simultaneously, a 3D molecular conformation is computed. This process involves the addition of hydrogen atoms and conformer generation through distance geometry methods, followed by geometry optimization using a molecular force field model to approximate the most stable conformation.

2.4 Property Calculation

For each valid molecular input, MolEdu automatically calculates a set of fundamental chemical properties to support structural understanding. These include the molecular formula, determined from the atomic composition of the generated structure, and the molecular weight, calculated based on standard atomic masses. Additionally, the system derives canonical SMILES, providing a standardized textual representation of the molecule, and retrieves the systematic IUPAC name from established chemical databases. These properties are presented alongside the molecular visualizations to reinforce the relationship between structure and composition, aiding users in developing a deeper conceptual understanding of chemical identity.

2.5 Visualization Interface

MolEdu presents the results in an interactive user interface. The 2D structure is rendered as a skeletal formula, while the 3D structure is displayed using an embedded viewer with user-controlled rotation and zoom functions. This dual visualization approach supports both planar recognition and spatial understanding of molecular geometry.

2.6 Error Handling

To support learning, the system includes validation checks for improper or unrecognized inputs. In such cases, users receive prompt feedback indicating the error and suggesting corrective action. This mechanism enhances user understanding of proper chemical notation and encourages iterative learning.

3. Results & Discussion

In the main interface of MolEdu, users are introduced to the platform through a clean and intuitive interface, as shown in Figure 1. The sidebar contains two main navigation options which are "About" and "Molecule Draw", while the central content area provides a brief overview of the application's purpose and functionality. The "About" section includes simple step-by-step guidance on how to interact with the system, requiring users to input either a molecular formula (e.g., H_2O, CO_2) or a SMILES string (e.g., C=O or C1CCCCC1). This layout is intentionally designed to lower the barrier to entry for users unfamiliar with cheminformatics software, ensuring that learners can focus on exploring molecular structures without needing prior experience in molecular drawing or computational tools. The visual simplicity and direct instructions support independent use by students in both classroom and remote learning environments.

Fig. 1. MolEdu main interface displaying the "About" section. Users are guided with brief instructions on how to input chemical identifiers such as SMILES strings or molecular formulas. The application interface is designed to be minimal, accessible, and user-friendly.

Upon navigating to the "Molecule Draw" section, as illustrated in Figure 2, users are presented with a minimal and distraction-free workspace. A dedicated input field located at the bottom of the interface prompts users to submit a SMILES string or molecular formula (e.g., "CC" or "C2H6" for ethane). Once submitted, the system processes the input and generates both 2D and 3D visualizations of the molecule, along with relevant chemical data. The seamless transition between input and molecular output enables users to focus entirely on exploring chemical structures without the need for manual drawing. This simple yet functional interface design ensures that even novice users can engage with molecular data quickly and intuitively, reinforcing MolEdu's role as an inclusive and accessible educational tool.

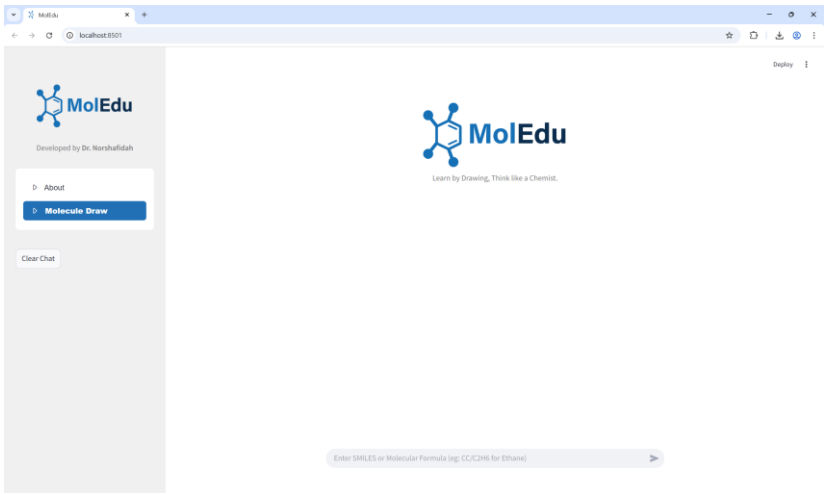


Fig. 2. MolEdu interface highlighting the "Molecule Draw" section. Users can enter either a SMILES string or molecular formula into the input bar at the bottom to initiate real-time structural rendering.

Following an input submission, MolEdu displays a comprehensive molecular analysis, as illustrated in Figure 3. In this example, the input "C2H6" is interpreted as ethane, and the platform responds by presenting the IUPAC name, canonical SMILES string, molecular formula, and molecular

weight. These calculated properties appear above two visual panels: one showing the 2D skeletal structure and the other rendering a manipulable 3D conformation of the molecule. This dual-format output allows users to explore both flat structural notation and realistic molecular geometry in one interface. By reinforcing the relationship between textual input and spatial molecular form, MolEdu enhances user understanding of how chemical information translates into 3D structure, a critical skill in organic chemistry and related disciplines.

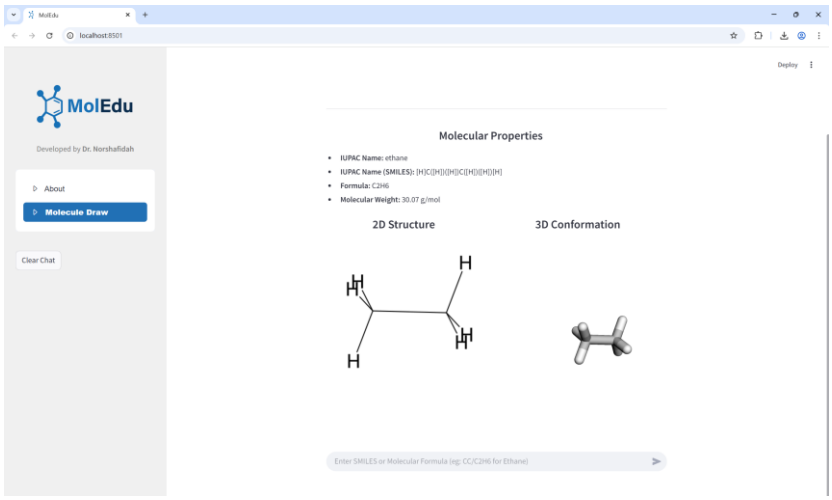


Fig. 3. Example output from MolEdu after submitting the molecular formula “C2H6” (ethane). The system displays key molecular properties along with 2D and 3D structural visualizations.

The examples shown in Table 1 demonstrate MolEdu’s robustness and versatility in handling a wide range of molecular structures commonly introduced in general and organic chemistry curricula. For each compound, users only need to provide a basic chemical formula or a SMILES string as input. MolEdu then retrieves and calculates detailed molecular properties and renders accurate 2D and 3D structures, enabling users to explore both symbolic and spatial representations of molecules.

The table includes classical hydrocarbons such as methane (CH_4), ethylene (C_2H_4), and acetylene (C_2H_2), highlighting MolEdu’s ability to distinguish between different types of carbon–carbon bonding (single, double, and triple bonds). In addition to simple hydrocarbons, aromatic compounds like benzene (C_6H_6) are also correctly depicted with delocalized π-bond systems shown in both 2D and 3D formats. This visual differentiation is important pedagogically, as students often struggle to conceptualize the transition from planar ring diagrams to actual molecular geometry.

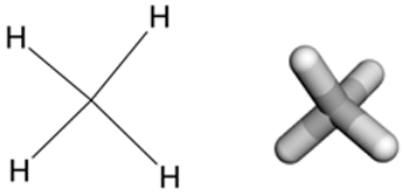
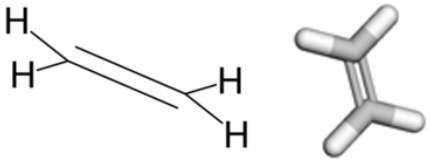
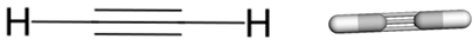
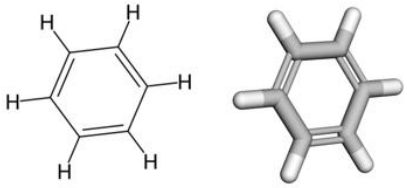
MolEdu also supports functional group recognition in molecules with heteroatoms, including ethanol (alcohol), formaldehyde (aldehyde), acetone (ketone), formic acid (carboxylic acid), methylamine (amine), and acetonitrile (nitrile). Each of these molecules is displayed with precise atom positioning and color-coded 3D visualization to enhance functional group identification—for example, oxygen atoms are rendered in red and nitrogen atoms in blue. The 3D outputs reflect geometry consistent with VSEPR theory, such as trigonal planar structures for aldehydes and tetrahedral arrangements in alcohols and amines. This feature enables users to not only memorize functional groups but also visualize and understand the geometry around heteroatoms, which is often abstract in static textbook images.

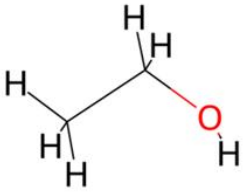
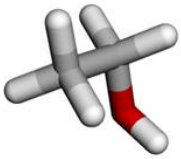
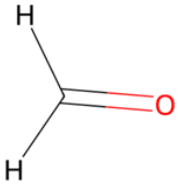
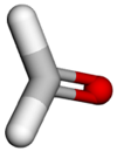
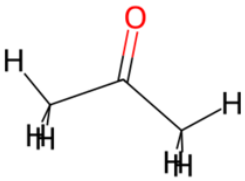
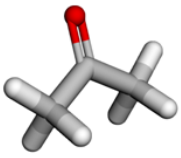
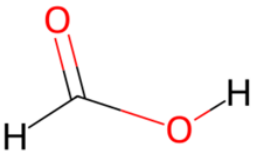
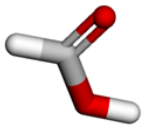
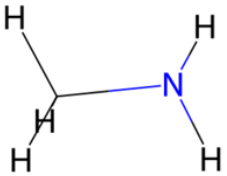
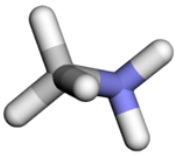
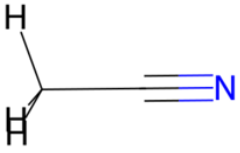
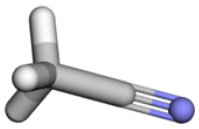
By allowing real-time comparison between input identifiers and rendered outputs, MolEdu promotes active learning and strengthens the user's ability to link molecular formulas with geometry and function. The platform’s consistent accuracy across this diverse set of compounds reinforces its reliability as an instructional tool. Furthermore, the integrated display of IUPAC names and canonical

SMILES alongside structural visuals reinforces nomenclature skills and familiarizes learners with digital chemical notation, both of which are essential competencies in modern chemical education and cheminformatics.

Overall, the dataset in Table 1 exemplifies MolEdu's effectiveness in bridging the gap between theoretical knowledge and visual-spatial understanding. It validates the platform's educational value in reinforcing key chemistry concepts through immediate, accurate, and interactive molecular feedback. This structured exposure to multiple functional groups in one interface supports comprehensive learning, especially when integrated into classroom activities, formative assessments, or self-paced exploration.

Table 1
Examples of user input and MolEdu-generated outputs for common functional groups. For each compound, the platform displays IUPAC-based molecular properties and corresponding 2D and 3D structural representations.

	User Input	Output	
Functional Group	Chemical Formula	Molecular Properties (IUPAC Name / IUPAC Name (SMILES) / Formula / Molecular Weight)	Molecular Structure (2D & 3D)
Alkane	CH4	• Methane • [H]C([H])([H])[H] • CH4 • 16.04 g/mol	
Alkene	C2H4	• Ethylene • [H]C([H])=C([H])[H] • C2H4 • 28.05 g/mol	
Alkyne	C2H2	• Acetylene • [H]C#C[H] • C2H2 • 26.04 g/mol	
Benzene Ring	C6H6	• Benzene • [H]C1C([H]) C([H]) C([H]) C([H]) C1([H]) • C6H6 • 78.11 g/mol	

Alcohol	CH3CH2OH	• Ethanol • [H]OC([H])([H])C([H])([H])[H] • C2H6O • 46.07 g/mol		
Aldehyde	HCHO	• Formaldehyde • [H]C([H])=O • CH2O • 30.03 g/mol		
Ketone	CH3COCH3	• Acetone • [H]C([H])([H])C(=O)C([H])([H])[H] • C3H6O • 58.08 g/mol		
Carboxylic Acid	HCOOH	• Formic Acid • [H]OC([H])=O • CH2O2 • 46.03 g/mol		
Amine	CH3NH2	• Methylamine • [H]N([H])C([H])([H])[H] • CH5N • 31.06 g/mol		
Nitrile	CH3CN	• Acetonitrile • [H]C([H])([H])C#N • C2H3N • 41.05 g/mol		

4. Future Potential in Real-World Reseach Contexts

Although MolEdu was primarily developed as an educational platform, its core features structural parsing, 2D/3D visualization, and molecular property reporting, make it adaptable for wider use in various branches of chemistry and materials science. Here are the several potential application areas.

4.1 Computational Chemistry

In computational chemistry and cheminformatics, researchers frequently work with large libraries of molecular representations in textual form like SMILES. MolEdu allows rapid visualization and inspection of these representations, offering a streamlined interface for verifying structural correctness prior to input into more advanced simulation tools such as density functional theory (DFT) or molecular dynamics packages [7]. It serves as a reliable pre-processing and educational tool for interpreting digital chemical structures.

4.2 Materials Chemistry

Polymers are essential in many real-life materials, and understanding their molecular structure helps explain why they behave the way they do. MolEdu can be used to explore these structures, showing how changes in the chemical makeup affect properties like flexibility, strength, or conductivity. Some common examples include biodegradable plastics [8]. Polymers like polylactic acid and polyhydroxyalkanoates are used for eco-friendly packaging. Their structures break down more easily in the environment. Users can study how their atoms are arranged to understand why they degrade faster. Other examples include electronic devices. Small molecules such as non-fullerene acceptors (e.g., Y6 derivatives) [9,10] or polymers such as P3HT [11], PTB7-Th [12], PM6 [13] often feature fused-ring systems and side chains that influence molecular packing and charge transport. Using MolEdu, researchers can input candidate SMILES structures to immediately visualize the spatial orientation of aromatic cores and alkyl substituents, aiding in the preliminary screening of molecular geometry and symmetry. This visual inspection can guide rational design decisions prior to quantum chemical calculations or device fabrication. In the field of perovskite solar cell research, MolEdu holds future potential as a tool for visualizing and cataloging organic and hybrid organic–inorganic compounds used as charge transport materials or interface modifiers [14-17]. Organic cations such as formamidinium, methylammonium, or more complex π-conjugated ligands can be represented in 3D to assess their molecular geometry, size, and interactions with inorganic frameworks. MolEdu also holds potential for expansion into inorganic and hybrid materials research. Compounds such as titanium dioxide (TiO_2) [18-20], tin oxide (SnO_2)[21], and zinc oxide (ZnO) are widely studied in applications including photocatalysis, solar energy conversion, gas sensing, and transparent conducting films. Although these materials are often represented and analyzed using crystallographic data, simplified molecular representations or empirical formulas (e.g., TiO_2, SnO_2, ZnO) are frequently used during early-stage screening, formulation design, or educational demonstrations.

4.3 Drug Discovery and Medicinal Chemistry

MolEdu can serve as a visualization and ideation tool during the design of lead compounds, analogs, and bioisosteres [22]. Medicinal chemists can inspect molecular scaffolds, functional group positioning, and hydrogen bonding potential through both 2D skeletal formulas and spatially

accurate 3D conformations. The platform is also well-suited for training pharmacy and biomedical students in structure-based drug design by enabling quick comparisons between similar compounds. For instance, students could investigate how small changes to a core scaffold alter the compound's shape, polarity, or interaction sites.

5. Conclusions

In conclusion, MolEdu is a step toward modernizing chemistry education through interactive, computer-assisted learning. It capitalizes on the strengths of open-source cheminformatics which is accuracy, flexibility, and community support and aligns them with pedagogical needs in chemistry. By providing an engaging platform for molecular drawing and visualization, MolEdu helps users to visualize, experiment, and learn chemistry in a more impactful way. This continued development and use of such tools will not only improve individual learner outcomes but also inspire a greater appreciation for the role of computational methods in understanding chemistry.

Acknowledgement
Authors acknowledges the Geran Penyelidikan Translasi UKM, grant number UKM-TR2024-09 and Geran Galakan Penyelidik Muda (GGPM), grant number GGPM-2023-048, funded by Universiti Kebangsaan Malaysia.

Conflict of Interest Statement
The author declares that there is no conflict of interest.

Author Contributions Statement
S.S. responsible for the conceptualization, validation, formal analysis, investigation, resources, visualization, original draft preparation, review and editing of the manuscript, and project administration.

Data Availability Statement
All data generated or analysed during this study are included in this published article. Additional datasets are available from the corresponding author upon reasonable request.

References
[1] Jones, Loretta L, Jordan, Kenneth D, Stillings, Neil A, Molecular visualization in chemistry education: the role of multidisciplinary collaboration. *Chemistry Education Research and Practice* 6 (2005): 136-149. https://doi.org/10.1039/B5RP90005K
[2] Landrum, Greg, Rdkit documentation. *Release* 1 (2013): 4.
[3] Landrum, Greg, RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling. *Greg Landrum* 8 (2013): 5281.
[4] O'Boyle, Noel M, Towards a Universal SMILES representation-A standard method to generate canonical SMILES based on the InChI. *Journal of cheminformatics* 4 (2012): 1-14. https://doi.org/10.1186/1758-2946-4-22
[5] 5. Quirós, Miguel, Gražulis, Saulius, Girdzijauskaitė, Saulė, Merkys, Andrius, Vaitkus, Antanas, Using SMILES strings for the description of chemical connectivity in the Crystallography Open Database. *Journal of cheminformatics* 10 (2018): 23. https://doi.org/10.1186/s13321-018-0279-6
[6] Seshadri, Keshavan, Liu, Peng, Koes, David Ryan, The 3dmol. js learning environment: a classroom response system for 3d chemical structures. *Journal of Chemical Education* 97 (2020): 3872-3876. https://doi.org/10.1021/acs.jchemed.0c00579
[7] Shafian, Shafidah, Mohd Salehin, Fitri Norizatie, Lee, Sojeong, Ismail, Azlan, Mohamed Shuhidan, Shuhaida, Xie, Lin, Kim, Kyungkon, Development of Organic Semiconductor Materials for Organic Solar Cells via the Integration of

Computational Quantum Chemistry and AI-Powered Machine Learning. *ACS Applied Energy Materials* 8 (2025): 699-722. https://doi.org/10.1021/acsaem.4c02937

[8] Gross, Richard A, Kalra, Bhanu, Biodegradable polymers for the environment. *Science* 297 (2002): 803-807. https://doi.org/10.1126/science.297.5582.803

[9] Hong, Minjeong, Youn, Jiyae, Ryu, Ka Yeon, Shafian, Shafidah, Kim, Kyungkon, Improving the Stability of Non-fullerene-Based Organic Photovoltaics through Sequential Deposition and Utilization of a Quasi-orthogonal Solvent. *ACS Applied Materials & Interfaces* 15 (2023): 20151-20158. https://doi.org/10.1021/acsami.3c02071

[10] Kim, Hyunkyoung, Kong, Ye-Jin, Kim, Won-Suk, Shafian, Shafidah, Kim, Kyungkon, Enhancing Reproducibility in Organic Solar Cell Fabrication via Static Sequential Deposition with Cross-Linked Polymer Donor and Nonfullerene Acceptor. *ACS Applied Polymer Materials* 6 (2024): 5814-5821. https://doi.org/10.1021/acsapm.4c00477

[11] Shafian, Shafidah, Jang, Yoonhee, Kim, Kyungkon, Solution processed organic photodetector utilizing an interdiffused polymer/fullerene bilayer. *Optics Express* 23 (2015): A936-A946. https://doi.org/10.1364/OE.23.00A936

[12] Ryu, Ka Yeon, Shafian, Shafidah, Shin, Jongchan, Lee, Yu Jin, Lee, Minjae, Kim, Kyungkon, Linear polyurethane ionenes for stable interlayer of organic photovoltaics. *Journal of Power Sources* 542 (2022): 231772. https://doi.org/10.1016/j.jpowsour.2022.231772

[13] Kim, Hyunkyoung, Heo, Yuchan, Na, Yeji, Shafian, Shafidah, Kim, BongSoo, Kim, Kyungkon, Cross-Linking-Integrated Sequential Deposition: A Method for Efficient and Reproducible Bulk Heterojunctions in Organic Solar Cells. *ACS Applied Materials & Interfaces* 16 (2024): 55873-55880. https://doi.org/10.1021/acsami.4c13237

[14] Lee, Sojeong, Yoon, Young Seon, Shafian, Shafidah, Kim, Jin Young, Kim, Kyungkon, Sequential Co-Deposition of Perovskite Film: An Effective Way of Tailoring Bandgap in All Vacuum Processed Perovskite Solar Cells. *Small Methods* (2025): 2500104.

[15] Zhang, Xiao, Wang, Linqing, Shafian, Shafidah, Wang, Peng, Zhao, Yuping, Wang, Pengcheng, Wu, Bin, Zhai, Jiaxue, Chen, Jiayan, Sun, Licheng, Crosslinking-Driven Chemical Homogeneity Enhances Performance of Pre-Seeded Perovskite Solar Cells. *Small* 21 (2025): 2408362. https://doi.org/10.1002/smll.202408362

[16] Wang, Peng, Shafian, Shafidah, Qiu, Feng, Zhang, Xiao, Zhao, Yuping, Wu, Bin, Kim, Kyungkon, Hua, Yong, Xie, Lin, Improving redox reactions of Spiro-OMeTAD via p-type molecular scaffold to reduce energy loss at Ag-electrode in perovskite solar cells. *Journal of Energy Chemistry* 102 (2025): 151-160. https://doi.org/10.1016/j.jechem.2024.10.027

[17] Qi, Shuwen, Ge, Chenghao, Wang, Peng, Wu, Bin, Zhao, Yuping, Zhao, Rongjun, Shafian, Shafidah, Hua, Yong, Xie, Lin, Improving Perovskite Solar Cell Performance and Stability via Thermal Imprinting-Assisted Ion Exchange Passivation. *ACS Applied Materials & Interfaces* 16 (2024): 51037-51045.

[18] Shafian, Shafidah, Kim, Kyungkon, Panchromatically responsive organic photodiodes utilizing a noninvasive narrowband color electrode. *ACS Applied Materials & Interfaces* 12 (2020): 53012-53020. https://doi.org/10.1021/acsami.0c17183

[19] Shin, Solbi, Shafian, Shafidah, Ryu, Ka Yeon, Jeon, Young-Kyo, Kim, Won-Suk, Kim, Kyungkon, Solution-Processed TiO2 Nanoparticles Functionalized with Catechol Derivatives as Electron Transporting Layer Materials for Organic Photovoltaics. *Advanced Materials Interfaces* 9 (2022): 2200118. https://doi.org/10.1002/admi.202200118

[20] Kim, Bo Youn, Shafian, Shafidah, Kim, Kyungkon, High-Performance Semitransparent Color Organic Photodiodes Enabled by Integrating Fabry–Perot and Solution-Processed Distributed Bragg Reflectors. *Advanced Materials Interfaces* 10 (2023): 2300421. https://doi.org/10.1002/admi.202300421

[21] Salehin, Fitri Norizatie Mohd, Chelvanathan, Puvaneswaran, Goje, Adamu Ahmed, Ludin, Norasikin Ahmad, Ibrahim, Mohd Adib, Shafian, Shafidah, Design of blue, green and red colorful semitransparent films using Ag/SnO2/Ag color filter for integrated into solar cells. *Results in Physics* 70 (2025): 108172.

[22] Adelusi, Temitope Isaac, Oyedele, Abdul-Quddus Kehinde, Boyenle, Ibrahim Damilare, Ogunlana, Abdeen Tunde, Adeyemi, Rofiat Oluwabusola, Ukachi, Chiamaka Divine, Idris, Mukhtar Oluwaseun, Olaoba, Olamide Tosin, Adedotun, Ibrahim Olaide, Kolawole, Oladipo Elijah, Molecular modeling in drug discovery. *Informatics in Medicine Unlocked* 29 (2022): 100880. https://doi.org/10.1016/j.imu.2022.100880