Utilizing Visual Molecular Dynamics to Visualize FNDC1

Asmita Deb
University of Massachusetts
Amherst, USA
asmitadeb@umass.edu

Oyewole Oyekoya
City University of New York
New York City, USA
oo700@hunter.cuny.edu

Olorunseun Ogunwobi
Michigan State University
East Lansing, USA
ogunwob1@msu.edu

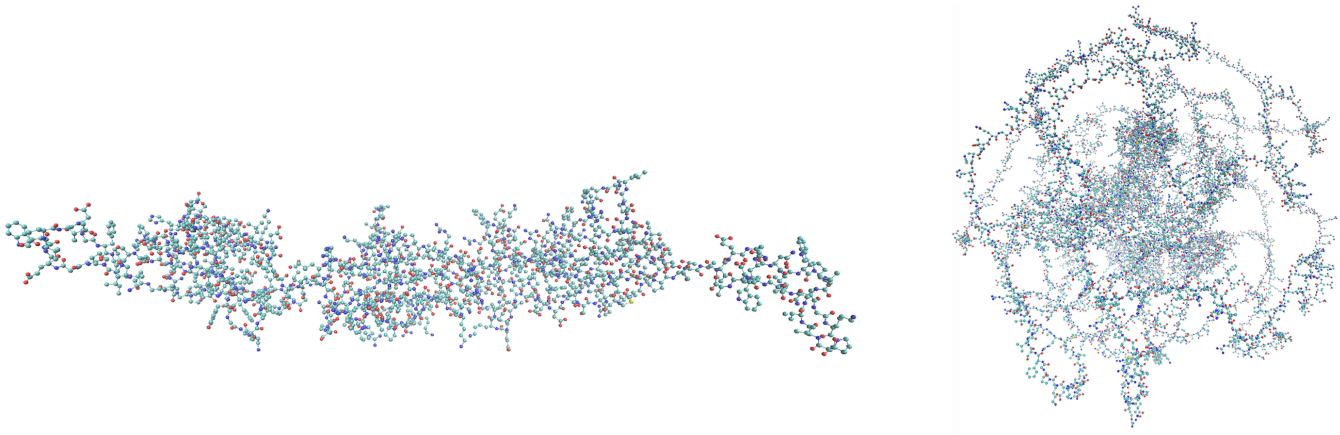

Figure 1: J3KNQ2 and Q4ZHG4 in Corey–Pauling–Koltun Form

ABSTRACT

This paper aims to understand two Protein Data Bank files of the same protein and analyze the structure of the fibronectin type 3 domain containing 1 (FNDC1) based on each file. This is done by using Visual Molecular Dynamics, a visualization computer program to analyze molecules like proteins, lipids, and more. This program visualizes FNDC1 in multiple 3D forms. FNDC1 is very understudied despite its connection to many cancers and other diseases. To understand more about the protein's pathways and progress future research, using VMD for protein structure analysis will be important. Structural information of angles, bonds, residues of FNDC1 were collected. A user study testing VMD as an appropriate tool to analyze complex molecules was also conducted with results showing that as molecular complexity increases, ease and understanding decreases for users.

CCS CONCEPTS

• **Human-centered computing** → **Virtual reality**; • **Computing methodologies** → *Perception*.

KEYWORDS

FNDC1, VMD, visualization, protein, structure, cancer research, complex, molecules

ACM Reference Format:

Asmita Deb, Oyewole Oyekoya, and Olorunseun Ogunwobi. 2024. Utilizing Visual Molecular Dynamics to Visualize FNDC1. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX)*. ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The FNDC1 protein, also known as fibronectin type 3 domain containing 1, is an extracellular matrix protein that participates in the angiogenesis process. Unfortunately, its over-expression is related to many diseases such as cancer as it also promotes tumorigenesis and metastasis. The high expression of FNDC1 also could indicate poor outcomes in patients of cancers such as gastric cancer and lung cancer. The over-expression forms a complex with VEGFR2, which is a regulator of endothelial migration and proliferation. This means the complex is a tumorigenic that is hyperactive in most tumors. The protein also activates G proteins, a family of proteins that can activate their respective signal transmission and energy pathways to help progress tumors, specifically G$\beta\gamma$.

The protein is very understudied, even though it has multiple associations with cancer and other diseases. One way to combat this issue is by studying the structure of the protein and therefore understand its function and know it better for future research. This can be done by molecular visualization, which is a field of computer graphics and has many different representations such as the relationship of structures to one another and the shape of molecules in space. Proteins are typically seen in their ribbon structure. Possible

Conference acronym 'XX, June 03–05, 2024, Woodstock, NY

ACM ISBN xxxx-x-xxxx-xxxx-x/xx/xx
https://doi.org/XXXXXXX.XXXXXXX

molecular visualization software include Visual Molecular Dynamics (VMD) and Paraview. Since understanding protein structure is important to understanding protein function, this research is important. In this paper, this will be done by analyzing a Protein Data Bank file through VMD.

2 RELATED WORK

In a study by Ghosh [1], it was found that FNDC1 interacts with fibronectin, androgen receptor, and c-MYC to progress prostate cancer. There is slight implication that FNDC1 interacts specifically with fibronectin and a change in localization in cancer cells versus normal cells. Unfortunately, the mechanism is still unclear.

An overall study done by Jameel et al. [2] is of the P53 protein was conducted because this protein is essential for tumor suppression. The researchers aimed to study the structure by using VMD and therefore understand the protein better. The paper had several steps and images describing the way they visualized and manipulated the protein.

Jiang et al. [3] presented research on the FNDC protein family as a biomarker for cancer research because of these proteins' roles in tumorigenesis. The fibronectin type 3 domain is important as it is involved in cell adhesion and migration, both relevant processes in cancer progression. They assessed expression levels of mRNA and patient survival in cancers ranging from breast to gastric. It was found that FNDC1 was overexpressed in most cancers compared to average tissues and that was associated negatively in breast and colorectal cancers.

In research conducted by Lu et al. [4], they found that the FNDC1-VEGFR2-G$\beta\gamma$ complex was enhancing proliferation of gastric cancer cells. They also concluded that arsenite targets FNDC1 and inhibits this function. Also, they noted that FNDC1 was upregulated by TWIST1 in human gastric cancer cells.

The research Marion et al. [5] conducted was a web application based on WebGL and ParaViewWeb. The application combines local and remote rendering of small and massive 3D molecules on multiple devices. Their research was a presentation of a proposed system and the initial experiment as well as a comparison with the current technologies.

In the paper written by Song et al. [6], the researchers concluded that FNDC1 could be used as a therapeutic target because of its connection to lung cancer. FNDC1 was found to be upregulate lung adenocarcinoma cells through the JAK2-STAT3 pathway. This means that these cells would invade and migrate into healthy cells, so possibly inhibiting this protein could decrease development of metastatic cancers.

A summarizing paper [7] about VMD for preparation of visualization for quantum mechanical calculations and molecular mechanism simulations. The researchers discuss a number of tools and plugins that would work with these smaller datasets. They mention that VMD is a good basic to use for the foundation of these associated tools.

Venkatesan et al. [8] analyzed how extended reality could be used in the medical field for education because it provides a 3D visualization alternative to a constrained 2D field. Extended reality could be useful in biomedical applications, surgical guidance, and molecular data visualization.

In 2023, Vos [9] studied miRNA-1207-3p, which targets mRNA coding. This microRNA has an implication with prostate cancer and so the study revolved around visualizing FNDC3 through VMD. There was a user study with this research that assessed the effectiveness of a VMD-based video tutorial in finding protein mutations. The results showed there could be improvements with better interface design to make analysis better.

3 METHODS

The software used for this specific project is Visual Molecular Dynamics (VMD). A Protein Data Bank file is a textual file that showcases the 3D structure of the molecules held in the worldwide Protein Data Bank. This is a data bank that holds structural data obtained by biocurists and biochemists and then shared for public access. The two files used to analyze FNDC1 both come from Homo sapiens. One of the files has a primary accession Q4ZHG4, which is the identifier number that has been assigned to the entry when it first entered UniprotKB, which is a resource for protein sequence and functional information of protein. The second file has a primary accession of J3KNQ2.

In VMD there are various analysis and graphics tools to manipulate the protein of interest and therefore understand its structure better. The main toolbars used in VMD for this project were from Graphics and Analysis. The Graphics section is used to manipulate the appearance of the protein. In this, we can change the representation, which is essentially a method of showing the protein. There is a coloring method and a drawing method. The drawing method pertains to the rendering method, so different models of the protein. This could be Licorice, CPK, Ribbons, and more. The coloring method is used to isolate certain aspects of the protein in specific color codes. The residues, elements, secondary structure, and more come from this section and are colored specifically. The protein can also be broken down slightly by showing the backbone, waters, and nucleic acids by boolean keywords.

The most important section, which provided the most data and information about the proteins, is the Extensions toolbar. This toolbar contained analysis forms, modeling forms, simulation forms, data forms, and visualization forms. In the analysis subsection, the Ramachandran Plot provides information about the protein's secondary structure and the sterically allowed and disallowed regions. Another important piece of information comes from the Sequence Viewer window. This provides the residue number, chain letter, B-factor, amino acid, and structure of the protein. The terminal of VMD also contains valuable data such as the number of bonds, angles, atoms, fragments, and more such information about the protein.

Another possible way to analyze the protein is by obtaining a topology file and a simulation of the protein through a Gromacs file. The topology file was generated successfully for both protein accessions through Gromos43a1 force field and the spc water model. This data is just more thorough structural data after a "cleaning" through cmap torsions.

The second part of the project revolves around the user study. The user study aims to test the usability of VMD and if the audience was able to identify structures and data of the given molecule. The hypothesis was that as molecular complexity increases, there would

be less usability and ease. Each question had a textual description of what the participant must identify, similar to the official VMD User Guide. Along with this text, there would be videos of the rotating molecules. The three questions were as following: state how easy or difficult it was to find sulfur atoms in the molecules, state how easy or difficult it was to find the alpha helices and beta sheets in the molecule, and finally open-ended questions asking the audience if they one molecule was harder than the other or if they felt that VMD was an appropriate tool for extremely complex molecules.

4 RESULTS AND ANALYSIS

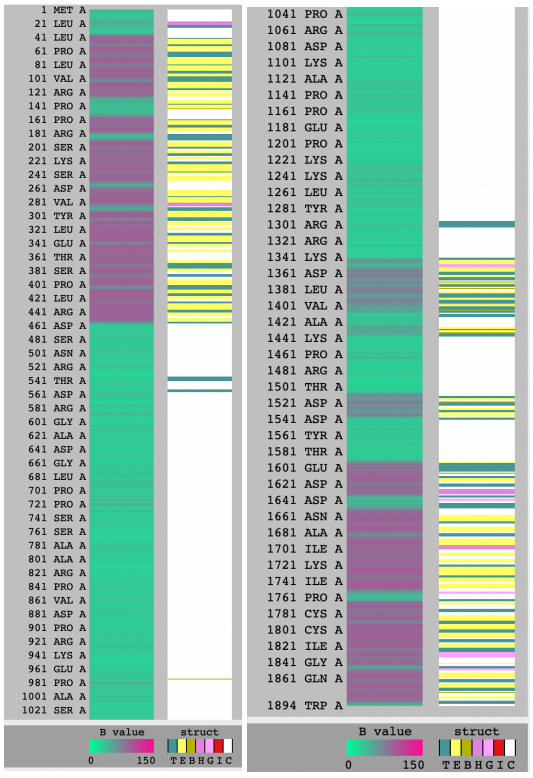

Figure 2: Sequence Viewer of Q4ZHG4

Data from the terminal of VMD for the Q4ZHG4 accession states the structure contains 14501 atoms, 14893 bonds, and 1899 residues. Using Gromos96, more data was established about the protein. Before "cleaning", there were 31302 pairs and 35786 dihedrals. After the torsions, there are 10580 dihedrals, 27305 angles, 8864 impropers, 31302 pairs, and 18802 bonds. The total charge is 32.000 e and there are 1894 residues. The starting terminus was MET-1, which is NH3+, and the ending terminus was TRP-1894 residue, which is COO-. For histidine protonation, there are 3080 donors and 2822 acceptors.

For the J3KNQ2, the VMD terminal found 2638 atoms, 2701 bonds, and 330 residues. There are no ligands or crystallizing agents. Before "cleaning", there were 5625 pairs and 6597 dihedrals. After cmap torsions, the terminal stated that there were 1861 dihedrals, 1684 impropers, 5094 angles, and 5625 pairs, and 3459 bonds. The total charge is 8.000 e and there are 330 residues. Residue LEU-14, which

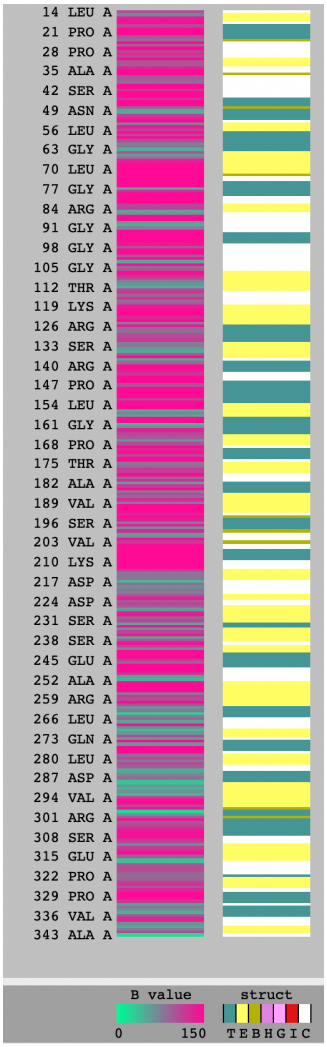

Figure 3: Sequence Viewer of J3KNQ2

is NH3+, was found to be the starting terminus and the ending terminus is ALA-343, which is COO-. For the histidine protonation, the terminal states there are 531 donors and 497 acceptors. Residue LEU-14, which is NH3+, was found to be the starting terminus and the ending.

Table 1: Structural Information of Accessions

Data	J3KNQ2	Q4ZHG4
Waters	0	0
Total Mass (a.m.u)	37333.531	205582.946
Fragments	1	6

From the user study, there were 29 total responses from a range of demographics. The background of the participants ranged from computer science, to business, to chemical sciences, to biological sciences and more. The age range was also wide with most of

the users in 18-55. For the primary accession of J3KNQ2, there were two questions about the difficulty in finding structures in the molecule in a ranking of 1 to 5, 1 being extremely easy and 5 being extremely difficult. For finding alpha helices and beta sheets, the mean response was 3.22.The minimum ranking was 1 and the maximum ranking was 5. The standard deviation for this question was 1.05 and the variance was 1.102. The standard error of the mean (SEM) was 0.202. For finding sulfur atoms, which were marked in yellow in CPK form, the mean response was 3.29. The minimum was 1 and the maximum was 5. Standard deviation for this group was 1.203 and the variance was 1.45. The SEM was 0.231. The more complex molecule whose primary accession was Q4ZHG4 also had the same questions with the same ranking type. For the alpha helices and beta sheets, the mean response was 2.00. The minimum was 1 and the maximum was 5. The standard deviation was 1.07 and the variance was 1.15. The SEM was 0.207. The sulfur atoms garnered a mean ranking of 1.85. The minimum was 1 and this time the maximum was 4. Lastly, the variance was 0.977 and the standard deviation is 0.988. The SEM was 0.19.

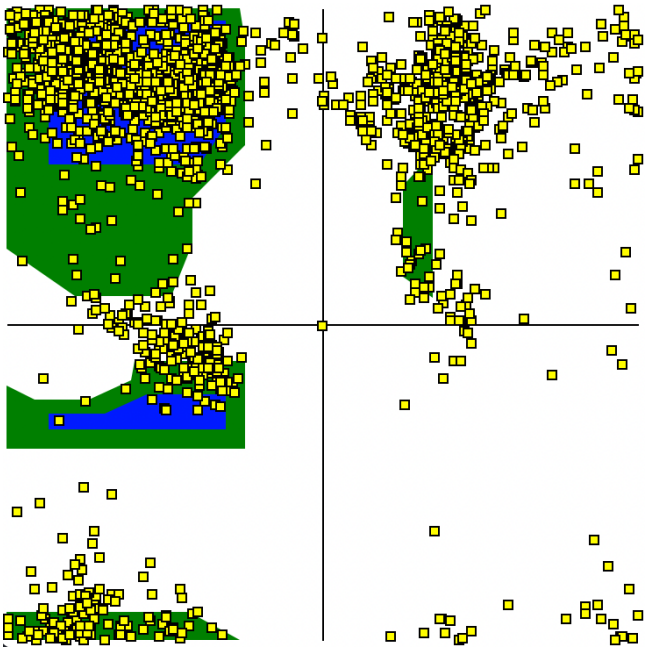

Figure 4: Ramachandran Plot of Q4ZHG4

5 DISCUSSIONS

From knowing all of the information conducted through the study and analysis through VMD, there are many conclusions that could be made. Knowing each residue, starting, and terminus could help us understand which segments are the most important in cancer development or pathways that upregulate. By knowing the precise secondary structure, also confirmed by the Ramachandran plot, it is also easy to see the folding and possibly misfolding of this protein. The Sequence Viewer provided information on the structure of each residue as well, which also allows for more advanced research. It

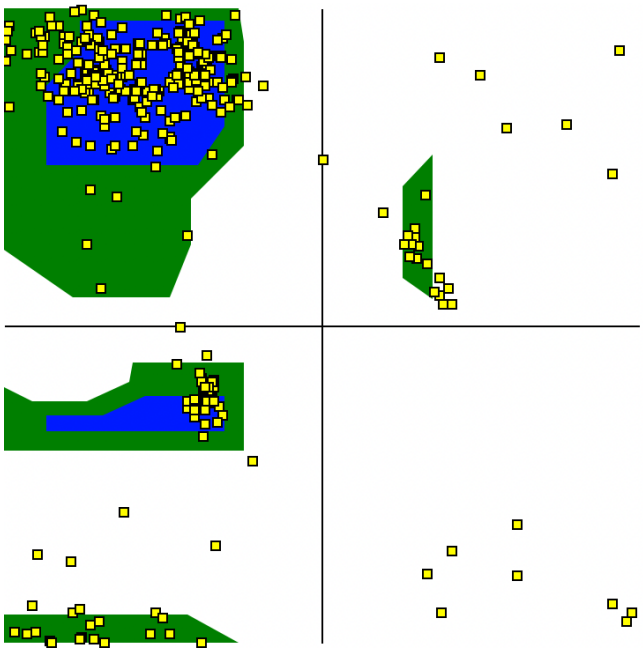

Figure 5: Ramachandran Plot of J3KNQ2

has the B-factor, which is the uncertainty of each atom's position, the residue number, the amino acid, and each residue secondary structure.

On the other hand, the user study was very useful in analyzing VMD as a tool for visualization. The study did have its own limitations with a sample size of 29, it does not speak for an entire population. Also, because part of the study depended on the audience's understanding of the textual description and video, there could be human error. There was a large age range of data from 18 to 55. The background field of the audience ranged from biological, chemical, and computer sciences to humanities. Also included was public health, statistics, and environmental sciences. The data from the study showed that on average the questions about the simple structure, J3KNQ2, had an ease ranking of around 3.20. The more complex molecule was bumped up in difficulty with an average ranking of 1.9. The variance and standard error of mean for the questions were mostly considered low and significant. For the identification of which molecule made which Ramachandran plot, 61.5% were able to correctly identify the Q4ZHG4 molecule and 63.0% were able to identify the J3KNQ2 molecule. This could point to users being able to infer from the previous question about alpha helices and beta sheets, as the Ramachandran plot depends on those secondary structures and regions of allowance and disallowance. Overall, the study points to issues with VMD as a tool when complexity increases. This could be combatted with more explanatory descriptions and visual aids in the official user guide. Another solution could also be more labeling in the visualization space in VMD to aid users.

6 CONCLUSIONS AND FUTURE WORK

Future research is necessary as this protein is an integral figure in learning more about cancer and other diseases. This research serves as an overview of VMD as a software to analyze protein structures but also as an examination of two different modeled structures of one protein. From the results of the user study, a lot of the open-end responses to the question of if people believed VMD was an appropriate tool for molecular analysis were that yes, it is an appropriate tool, however there are many visual issues with complex molecules and that they would have wanted to know how to use it properly before trying to analyze. In the future, VMD may want to update their user guide and tutorial to have more visual aid and even possibly including videos. The results and structural information found from this research could serve as a building block for future research in cancer and other diseases that FNDC1 has an effect on. With structural biology, cancer research can become easier, especially in this situation with FNDC1.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation, Research Experience for Undergraduates Award No. 2050532.

REFERENCES

[1] Priyanka Ghosh and Olorunseun O Ogunwobi. 2023. Intracellular interaction of FNDC1/FN1/AR/cMYC in prostate cancer cells. *bioRxiv* (2023), 2023–02.

[2] Mohammed I Jameel, Arshed H Yaseen, Marwan Q Al-Samarraie, and Bahez A Abdulla. 2018. Visualization and analysis of Tumor protein P53 by VMD. *International Journal for Sciences and Technology* 143, 6042 (2018), 1–9.

[3] Hui Jiang, Bo Ling Chu, Jiao He, Zhi Liu, and Ling Yang. 2022. Expression and prognosis analyses of the fibronectin type-III domain-containing (FNDC) protein family in human cancers: A Review. *Medicine* 101, 49 (2022), e31854.

[4] Yao Lu, Panpan Huang, Xueliang Zeng, Wenyu Liu, Rui Zhao, Jing Li, Gaolu Cao, Yaqiong Hu, Qiuxiang Xiao, Meng Wu, et al. 2023. Inhibition of FNDC1 suppresses gastric cancer progression by interfering with G$\beta\gamma$-VEGFR2 complex formation. *Iscience* 26, 9 (2023).

[5] Charles Marion, Joachim Pouderoux, Julien Jomier, Sebastien Jourdain, Marcus Hanwell, and Utkarsh Ayachit. 2013. A hybrid visualization system for molecular models. In *Proceedings of the 18th International Conference on 3D Web Technology*. 117–120.

[6] Yang Song, Jun-Feng Guo, Pei-Shu Lan, Miao Wang, and Quan-Yu Du. 2024. Investigation of the pan-cancer property of FNDC1 and its molecular mechanism to promote lung adenocarcinoma metastasis. *Translational Oncology* 44 (2024), 101953.

[7] Mariano Spivak, John E Stone, João Ribeiro, Jan Saam, Peter L Freddolino, Rafael C Bernardi, and Emad Tajkhorshid. 2023. VMD as a platform for interactive small molecule preparation and visualization in quantum and classical simulations. *Journal of chemical information and modeling* 63, 15 (2023), 4664–4678.

[8] Mythreye Venkatesan, Harini Mohan, Justin R Ryan, Christian M Schürch, Garry P Nolan, David H Frakes, and Ahmet F Coskun. 2021. Virtual and augmented reality for biomedical applications. *Cell reports medicine* 2, 7 (2021).

[9] Samantha Vos, Oyewole Oyekoya, and Olorunseun Ogunwobi. 2023. Visualization of Point Mutations in Fibronectin Type-III Domain-Containing Protein 3 in Prostate Cancer. In *Companion Proceedings of the 2023 Conference on Interactive Surfaces and Spaces*. 10–13.