Structural bioinformatics

VRmol: an Integrative Web-Based Virtual Reality System to Explore Macromolecular Structure

Kui Xu[1,3,4,5], Nan Liu[2,3,4,5], Jingle Xu[1,3], Chunlong Guo[4], Lingyun Zhao[2,3], Hong-Wei Wang[2,3,4], Qiangfeng Cliff Zhang[1,3,4,*]

[1]MOE Key Laboratory of Bioinformatics, School of Life Sciences, Tsinghua University, Beijing, China 100084. [2]Ministry of Education Key Laboratory of Protein Sciences, School of Life Sciences, Tsinghua University, Beijing, China 100084. [3]Beijing Advanced Innovation Center for Structural Biology, Beijing Frotier Research Center for Biological Structures, Tsinghua University, Beijing, China 100084. [4]Tsinghua-Peking Joint Center for Life Sciences, Tsinghua University, Beijing, China 100084. [5]These authors contributed equally to this work.

*To whom correspondence should be addressed. Email: qczhang@tsinghua.edu.cn.

Associate Editor: Yann Ponty

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Summary: Structural visualization and analysis are fundamental to explore macromolecular functions. Here we present a novel integrative web-based virtual reality (VR) system – VRmol, to visualize and study molecular structures in an immersive virtual environment. Importantly, it is integrated with multiple online databases and able to couple structure studies with associated genomic variations and drug information in a visual interface by cloud-based drug docking. VRmol thus can serve as an integrative platform to aid structure-based translational research and drug design.

Availability and implementation: VRmol is freely available (https://VRmol.net), with detailed manual and tutorial (https://VRmol.net/docs). The code of VRmol is available as open source under the MIT license at http://github.com/kuixu/VRmol.

Contact: qczhang@tsinghua.edu.cn

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Visualization and analysis of molecular structures are essential to generating insights into the mechanisms of actions of macromolecules (O'Donoghue, et al., 2010). GRASP (Nicholls, et al., 1991), RasMol (Sayle and Milner-White, 1995), PyMol (Janson, et al., 2017) and UCSF Chimera (Pettersen, et al., 2004) are among widely used software tools for molecular visualization. However, these tools were designed mainly for desktop systems and are often tedious to install and configure. Web-based applications with flexible interfaces to study macromolecular structures have also emerged, including AstexViewer (Hartshorn, 2002), Jmol (Hanson, 2010), 3Dmol.js (Rego and Koes, 2015), NGL Viewer (Rose and Hildebrand, 2015), LiteMol (Sehnal, et al., 2017), Molstar (molstar.org), and Web3DMol (Shi, et al., 2017). But they usually only focus on structure display and lack capabilities for complex structure analysis. Due to the recent advances in hardware, Virtual reality (VR) has been gaining popularity as a powerful technology to create a virtual world and enable users to interact with simulated objects beyond reality (Berg and Vance, 2017; Garcia-Bonete, et al., 2019; Kartiko, et al., 2010), including bioinformatics applications (Sommer, et al., 2018). VR technology has been introduced in structural visualization tools, exemplified by UCSF ChimeraX (Goddard, et al., 2018; Goddard, et al., 2018), Autodesk Molecular Viewer (Balo, et al., 2017), Nanome (Kingsley, et al., 2019), Narupa/iMD-VR (Deeks, et al., 2020; O'Connor, et al., 2018; O'Connor, et al., 2019), UnityMol (Laureanti, et al., 2020), and ProteinVR (Cassidy, et al., 2020). However, they often lack of the integration of several favorable structural analysis options, like automatic-retrieving genomic variations, molecules docking, and structure editing, into one VR platform (Table S1).

Importantly, recent progress in systematic studies of disease and cancer genomes have produced an explosive growth of human genomic variation data, collected by The Cancer Genome Atlas (TCGA) (Cancer Genome Atlas Research, et al., 2013), the Cancer Cell Line Encyclopedia (CCLE) (Barretina, et al., 2012), and the Exome Aggregation Consortium (ExAC) (Lek, et al., 2016). Mapping and studying the variations in a structural context can shed light on the molecular mechanisms of related diseases (Feuk, et al., 2006). Aquaria (O'Donoghue, et al., 2015) and MinOmics (Maes, et al., 2018) are web-based tools integrated with genomic and structural databases, such as PDB (Berman, et al., 2000), InterPro (Hunter, et al., 2012), UniProt (UniProt, 2014), Pfam (Finn, et al., 2014), GO (Ashburner, et al., 2000), and MapMan (Schwacke, et al., 2019), but lack the connection to disease related database like TCGA, CCLE, *etc* mentioned above (Table S1). Moreover, integrative analysis with corresponding potential drug molecules from databases, like the DrugBank (Wishart, et al., 2018) and ChEMBL (Gaulton, et al., 2012), will significantly benefit our understanding on drug action mechanisms and efficacy, and further optimization. However, to our best knowledge, no existing molecular viewing tool has integrated structural analysis with disease-related genomic variation and drug molecule databases in an immersive environment.

Addressing these gaps, we developed a novel web-based framework, VRmol, for molecular structure visualization and integrative analysis (Table S1). VRmol implements both an immersive VR and a traditional non-VR environment with the ability to smoothly switch between the two modes. Notably, VRmol connects to multi-source diseases and drugs related database and performs drug docking from the cloud and visualizes results in a fully immersive VR environment, providing an integrated platform for translational study. Users can access VRmol at https://VRmol.net through various WebXR enabled web browsers, such as Google Chrome (v81+), Microsoft Edge (v81+) and Firefox Reality.

2 Materials and implementation

The VRmol system consists of three modules (Figure 1A and Table S2): (1) a resource module to retrieve and parse different types of structural and translational data from the web by connecting to online databases including the protein data bank (PDB) (Berman, et al., 2000), Consurf-DB (Ashkenazy, et al., 2010), disease-related genomic mutation and variations data (like TCGA), and drug databases (like DrugBank); (2) a structure modeling module to generate, process, and render 3D structural models in VR environment implemented by native JavaScript with WebGL, WebXR, and Three.js; (3) an interaction module to analyze, manipulate, and interact with structural models in an immersive VR environment using different VR devices such as HTC Vive and Microsoft Mix Reality in compatible web browsers (Table S3). Implementation details are described in detail in the Supplementary material.

3 Results

3.1 Structural visualization and Operation

VRmol supports various structural representation styles, including the ribbon, tube, stick, and ball & stick styles (Figure 1B, Figure S4), as well as different structure surfaces, like the solvent-accessible surface and Van de Waals surface (Figure 1C). Users can color a surface by hydrophobicity or secondary structure, and render it with mesh or different level of transparency. To improve the computational efficiency and visualization experience of large molecules, VRmol introduces a mechanism which defines a vision sphere centered at the camera position. Only the atoms in the sphere are rendered in high resolution, while the others are in low resolution.

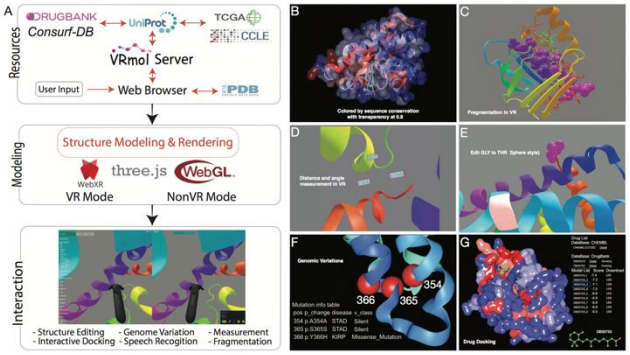

Figure 1. A: System design of VRmol. B: Surface and secondary structural style colored by sequence conservation (PDB id: 1MBS) (Scouloudi and Baker, 1978). C: Selected fragments in Sphere and Sticks styles (PDB id: 1MBS). D: Distance and angle measurement (PDB id: 1MBS). E: Structural editing: replacing the first residue (glycine) of Seal myoglobin (PDB id: 1MBS) by threonine. F: Genomic variations: three variation sites (from TCGA) were mapped onto the structure of human glucose transporter GLUT1 (PDB id: 4PYP) (Deng, et al., 2014) and highlighted with red balls. G: Drug docking: docking of the drug DB00755 from DrugBank onto the cellular retinoic acid binding protein (PDB id: 1CBS) (Kleywegt, et al., 1994).

3.2 Visualization of Genomic Variations

Analyzing the distribution of genomic variations on protein structures can shed light on the genetic basis of many complex diseases. VRmol retrieves data from online genomic variation databases, which integrated TCGA, CCLE, ExAC, and dbSNP, and maps variations onto structures (Figure S3) (Supplementary material). This approach can reveal whether and how frequently mutated residues spatially cluster on a structure. As an example, Figure 1F shows the human glucose transporter GLUT1 (PDB ID: 4PYP.) (Deng, et al., 2014) with highlighted silence and missense variations.

3.3 Drug Docking

VRmol can automatically search and load relevant drug molecules or small ligands of the target structure from multiple databases, such as DrugBank, ChEMBL, BindingDB (Chen, et al., 2001), SwissLipids (Aimo, et al., 2015), and GuidetoPHARMACOLOGY (Harding, et al., 2018) (Table S4), and performs drug docking (Video S2) through a cloud computing configure on the VRmol server. The best docking modes are returned with docking scores displayed in the screen (Figure 1G). It is important to note that, the VR environment of VRmol offers an opportunity for intuition-guided docking by manually defining a region on the structure surface as the preferred docking location. The docking pipeline and scoring function are described in Supplementary information.

Conclusion

VRmol provides users a convenient platform to explore molecular structures in a fully immersive 3D virtual environment. It compares favorably with state-of-the-art structural visualization and analysis tools, with various structural visualization and analysis functions (Table S1). Importantly, VRmol performs drug docking with a

server and connects to many disease-related mutation databases and drug databases, and thus provides an integrative platform for translational researches.

Acknowledgements

We thank Xun Ran and Dawei Zhang for the help on developing VRmol, and we thank Haitao Li and Maoxiang Shi (Tsinghua University) for helpful comments on the development of VRmol.

Funding

This project is supported by the State Key Research Development Program of China (Grant No. 2018YFA0107603 to Q.C.Z.) and the National Natural Science Foundation of China (Grants No. 91740204, 31761163007 and 31621063 to Q.C.Z.), the Beijing Advanced Innovation Center for Structural Biology, and the Tsinghua-Peking Joint Center for Life Sciences to Q.C.Z. and H.W..

Conflict of Interest: none declared.

References

Aimo, L., *et al.* (2015) The SwissLipids knowledgebase for lipid biology. *Bioinformatics*, 31(17), 2860-2866.

Ashburner, M., *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics*, 25(1), 25-29.

Ashkenazy, H., *et al.* (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic acids research*, 38(Web Server issue), W529-533.

Balo, A.R., Wang, M. and Ernst, O.P. (2017) Accessible virtual reality of biomolecular structural models using the Autodesk Molecule Viewer. *Nat Methods*, 14(12), 1122-1123.

Barretina, J., *et al.* (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483(7391), 603-607.

Berg, L.P. and Vance, J.M. (2017) Industry use of virtual reality in product design and manufacturing: a survey. *Virtual Reality*, 21(1), 1-17.

Berman, H.M., *et al.* (2000) The Protein Data Bank. *Nucleic acids research*, 28(1), 235-242.

Cancer Genome Atlas Research, N., *et al.* (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nature genetics*, 45(10), 1113-1120.

Cassidy, K.C., *et al.* (2020) ProteinVR: Web-based molecular visualization in virtual reality. *Plos Comput Biol*, 16(3), e1007747.

Chen, X., Liu, M. and Gilson, M.K. (2001) BindingDB: a web-accessible molecular recognition database. *Comb Chem High Throughput Screen*, 4(8), 719-725.

Deeks, H.M., *et al.* (2020) Interactive molecular dynamics in virtual reality for accurate flexible protein-ligand docking. *PloS one*, 15(3), e0228461.

Deng, D., *et al.* (2014) Crystal structure of the human glucose transporter GLUT1. *Nature*, 510(7503), 121-125.

Feuk, L., Carson, A.R. and Scherer, S.W. (2006) Structural variation in the human genome. *Nat Rev Genet*, 7, 85.

Finn, R.D., *et al.* (2014) Pfam: the protein families database. *Nucleic acids research*, 42(Database issue), D222-230.

Garcia-Bonete, M.J., Jensen, M. and Katona, G. (2019) A practical guide to developing virtual and augmented reality exercises for teaching structural biology. *Biochem Mol Biol Edu*, 47(1), 16-24.

Gaulton, A., *et al.* (2012) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic acids research*, 40(Database issue), D1100-1107.

Goddard, T.D., *et al.* (2018) Molecular Visualization on the Holodeck. *Journal of molecular biology*.

Goddard, T.D., *et al.* (2018) UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci*, 27(1), 14-25.

Hanson, R.M. (2010) Jmol– a paradigm shift in crystallographic visualization. *Journal of Applied Crystallography*, 43(5-2), 1250-1260.

Harding, S.D., *et al.* (2018) The IUPHAR/BPS Guide to PHARMACOLOGY in 2018: updates and expansion to encompass the new guide to IMMUNOPHARMACOLOGY. *Nucleic acids research*, 46(D1), D1091-D1106.

Hartshorn, M.J. (2002) AstexViewer: a visualisation aid for structure-based drug design. *J Comput Aided Mol Des*, 16(12), 871-881.

Hunter, S., *et al.* (2012) InterPro in 2011: new developments in the family and domain prediction database. *Nucleic acids research*, 40(Database issue), D306-312.

Janson, G., *et al.* (2017) PyMod 2.0: improvements in protein sequence-structure analysis and homology modeling within PyMOL. *Bioinformatics*, 33(3), 444-446.

Kartiko, I., Kavakli, M. and Cheng, K. (2010) Learning science in a virtual reality application: The impacts of animated-virtual actors' visual complexity. *Comput. Educ.*, 55(2), 881-891.

Kingsley, L.J., *et al.* (2019) Development of a virtual reality platform for effective communication of structural data in drug discovery. *J Mol Graph Model*, 89, 234-241.

Kleywegt, G.J., *et al.* (1994) Crystal-Structures of Cellular Retinoic Acid-Binding Protein-I and Protein-Ii in Complex with All-Trans-Retinoic Acid and a Synthetic Retinoid. *Structure*, 2(12), 1241-1258.

Laureanti, J., *et al.* (2020) Visualizing biomolecular electrostatics in virtual reality with UnityMol-APBS. *Protein Sci*, 29(1), 237-246.

Lek, M., *et al.* (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536(7616), 285-291.

Nicholls, A., Sharp, K.A. and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, 11(4), 281-296.

O'Connor, M., *et al.* (2018) Sampling molecular conformations and dynamics in a multiuser virtual reality framework. *Sci Adv*, 4(6), eaat2731.

O'Connor, M.B., *et al.* (2019) Interactive molecular dynamics in virtual reality from quantum chemistry to drug binding: An open-source multi-person framework. *J Chem Phys*, 150(22), 220901.

O'Donoghue, S.I., *et al.* (2010) Visualization of macromolecular structures. *Nat Methods*, 7(3 Suppl), S42-55.

O'Donoghue, S.I., *et al.* (2015) Aquaria: simplifying discovery and insight from protein structures. *Nat Methods*, 12(2), 98-99.

Pettersen, E.F., *et al.* (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*, 25(13), 1605-1612.

Rego, N. and Koes, D. (2015) 3Dmol.js: molecular visualization with WebGL. *Bioinformatics*, 31(8), 1322-1324.

Rose, A.S. and Hildebrand, P.W. (2015) NGL Viewer: a web application for molecular visualization. *Nucleic acids research*, 43(W1), W576-579.

Sayle, R.A. and Milner-White, E.J. (1995) RASMOL: biomolecular graphics for all. *Trends Biochem Sci*, 20(9), 374.

Schwacke, R., *et al.* (2019) MapMan4: A Refined Protein Classification and Annotation Framework Applicable to Multi-Omics Data Analysis. *Mol Plant*, 12(6), 879-892.

Scouloudi, H. and Baker, E.N. (1978) X-ray crystallographic studies of seal myoglobin. The molecule at 2.5 A resolution. *Journal of molecular biology*, 126(4), 637-660.

Sehnal, D., *et al.* (2017) LiteMol suite: interactive web-based visualization of large-scale macromolecular structure data. *Nat Methods*, 14(12), 1121-1122.

Shi, M., Gao, J. and Zhang, M.Q. (2017) Web3DMol: interactive protein structure visualization based on WebGL. *Nucleic acids research*, 45(W1), W523-W527.

Sommer, B., *et al.* (2018) From Virtual Reality to Immersive Analytics in Bioinformatics. *J Integr Bioinform*, 15(2).

UniProt, C. (2014) Activities at the Universal Protein Resource (UniProt). *Nucleic acids research*, 42(Database issue), D191-198.

Wishart, D.S., *et al.* (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic acids research*, 46(D1), D1074-D1082.