Exploring Drugbank in Virtual Reality Chemical Space

Daniel Probst[1,*] and Jean-Louis Reymond[1,*]

[1]*Department of Chemistry and Biochemistry, National Center for Competence in Research NCCR TransCure, University of Berne, Freiestrasse 3, 3012 Berne, Switzerland*

[*]e-mail: daniel.probst@dcb.unibe.ch, jean-louis.reymond@dcb.unibe.ch

Abstract

The recent general availability of low-cost virtual reality headsets, and accompanying 3D engine support, presents an opportunity to bring the concept of chemical space into virtual environments. While virtual reality applications represent a category of widespread tools in other fields, their use in the visualization and exploration of abstract data such as chemical spaces has been experimental. In our previous work we established the concept of interactive 2D maps of chemical spaces, followed by interactive web-based 3D visualizations, culminating in the interactive web-based 3D visualization of extremely large chemical spaces. Virtual reality chemical spaces are a natural extension of these concepts. As 2D and 3D embeddings, and projections of high-dimensional chemical fingerprint spaces were shown to be valuable tools in chemical space visualization and exploration, existing pipelines of data mining and preparation can be extended to be used in virtual reality applications. Here we present an application based on the Unity engine and the virtual reality toolkit (VRTK), allowing for the interactive exploration of chemical space populated by Drugbank compounds in virtual reality. The source code of the application as well as the most recent build are available on GitHub (https://github.com/reymond-group/virtual-reality-chemical-space).

Introduction

Today, an ever-increasing amount of data on chemical structures is available from databases containing compounds extracted from scientific and patent literature (DrugBank, ChEMBL, SureChEMBL, ZINC, PubChem) as well as from generated databases (GDB).[1–6] This expansion of the chemical space inspired the introduction of new methods and algorithms to analyze and process the data. Among these, interactive visualizations have proven to be a valuable tool for chemists to explore chemical space as they allow for intuitive interaction with the data. Interactive visualizations are embeddings of chemical structures in 2- or 3-dimensional space, created by encoding molecules as high-dimensional feature vectors, or fingerprints, and then applying a suitable dimensionality reduction method. An additional dimension is then added by means of coloring the data points representing chemical descriptors or components of the original, high-dimensional feature vector. The resulting data is then loaded by an application providing an adequate user interface. In previous work, we have shown the viability of this approach by making the generated mappings of a variety of databases available through desktop and web applications. In fact, we iterated through an evolutionary history of the visualization of chemical space by means of desktop and web applications using interactive maps colored by a multitude of molecular descriptors. In 2013 we released a Java applet-based application denoted Mapplet, which enabled chemists to browse a given chemical space in 2D.[7] This made the chemical spaces of GDB-11, GDB-13, DrugBank, ChEMBL, PubChem, ZINC, GDB-11, and GDB-13 intuitively explorable. These visualizations were based on the MQN and SMIfp chemical fingerprints.[7,8] Three years after the initial release of the first Mapplet, we published a web application based on the 3D engine THREE.js, enabling the exploration of the DrugBank chemical space in 3D inside the browser.[9] This functionality was later generalized to allow users to create chemical spaces using custom input data in the form of SMILES strings.[10] The list of chemical fingerprints was extended to also include APfp, Xfp, Sfp and ECFP4.[11–13]

Recently, we published a custom JavaScript 3D engine, capable of visualizing millions of points in a browser, pushing the previous limits by three orders of magnitude.[14] However, navigating densely populated 3-dimensional spaces can be challenging for many users. In fact, 3D user interfaces with a comparatively low information density have been actively investigated for decades, never leaving the experimental phase outside of virtual reality applications.[15–17]

While virtual reality (VR) has been a staple in industrial design, psychology, medicine, and military training for decades, advances in computer hardware have recently pushed the technology into the consumer space.[18–22] The general availability of comparatively low-cost virtual reality headsets has allowed for wide-spread experimentation with the technology in new fields such as education, healthcare, and computational chemistry. After the introduction of VR support in popular 3D engines supporting VR like Unity and Unreal, multiple groups have successfully implemented VR representations of scientific data.[23–28]

Here we present an extension of our previous work on interactive visualizations of chemical spaces to VR. As with our previously released tools, the application allows for the seamless navigation between global and local features of large data sets, enabling an intuitive exploration of chemical spaces.

Methods

Data Acquisition and Preparation

The data used is a subset of DrugBank ($n = 5,818$), containing only compounds labelled as *Approved, Experimental*, or *Investigational* with a heavy atom count of ≤ 50. To obtain the 3D coordinates for the visual representation of the molecules in VR, their respective 42-dimensional MQN fingerprints are calculated. The high-dimensional fingerprint vectors are then embedded in a 3-dimensional space using principal component analysis (PCA). We have shown that this representation can retain a substantial amount of information with an intuitively navigable distribution. As with our previously published tools,

the three spatial dimensions can be extended by a fourth through the colorization of data points. This

fourth dimension, termed a map, contains the values of a molecular descriptor scaled to the interval

[0,360] which represents the hue value in a cylindrical HSV (hue, saturation, value) color model. The

hue value is encoded as a triplet in RGB space.

In addition, the structural information files (SDF) for all molecules in the Drugbank subset are

also acquired from the Drugbank website. These files are required for rendering the 3D models of the

compounds in the VR application.

Implementation

As previously published implementations have shown that the Unity game engine is capable of

rendering point clouds or large sets of scientific data in a VR context, we selected the freely available

development environment as our starting point. In addition, the virtual reality tool kit (VRTK) available

for Unity supports headset-agnostic development.
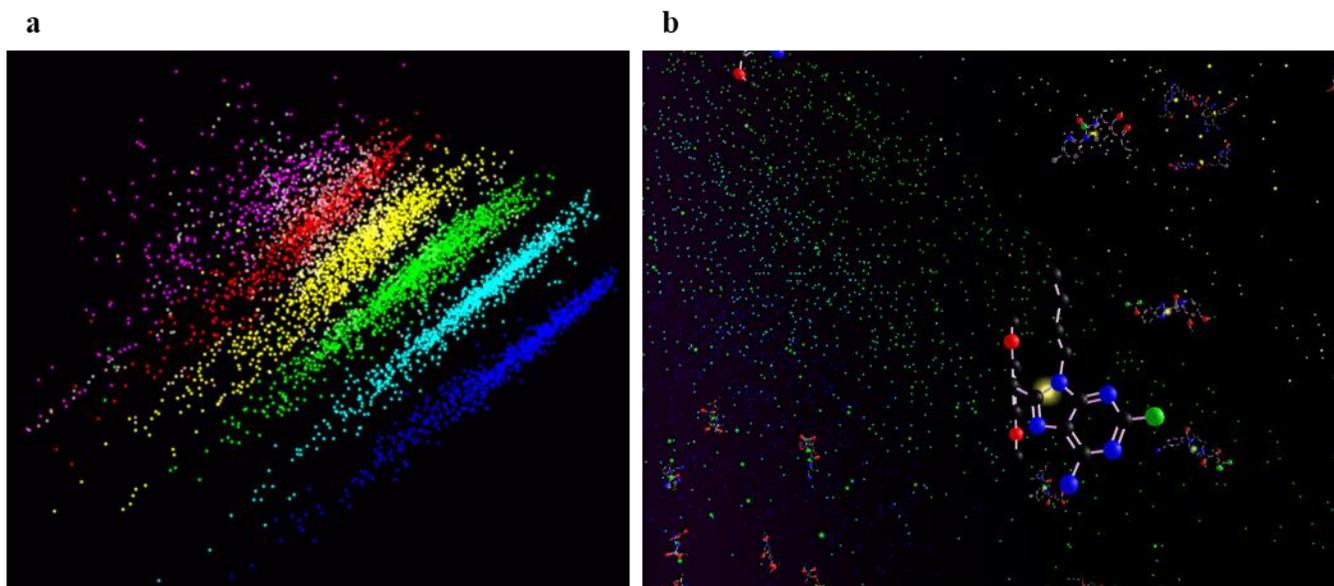
Fig. 1 Rendering Drugbank chemical space in VR. (**a**) Real-time rendering of the Drugbank chemical space using Unity's particle system. The colors represent the ring number descriptor (ascending in number blue to magenta). (**b**) Shows a closeup of the chemical space where the player was moved inside the point cloud and molecular models of nearby compounds are loaded as a secondary level of detail.

Compared to traditional applications, VR software is extremely demanding on GPUs, as each frame must be rendered from two view-points, one for each eye, to enable a stereoscopic view of the scene. However, a relatively high number of visible data is required to provide a global perspective while exploring local features, which is an important aspect of the interactive exploration of chemical spaces. These two contrasting requirements can be satisfied by drawing from a wide range of optimization techniques such as frustum culling, a level-of-detail (LOD) system, or per-layer cull distances. The aim of these techniques is to reduce the number of polygons, or vertices, rendered by the GPU, while not negatively influencing visual fidelity. The system we have chosen to implement draws heavily from the level of details system which has been a mainstay in the game development industry for decades. These systems have the main goal of simplifying geometry and thus the number of rendered polygons depending on the distance from the rendered object to the user's position in the virtual world. For instance, a tree positioned at 1,000 meters from the user can be rendered using a lower level of detail compared to a tree positioned 1 meter from the user, without losing visual fidelity. In our implementation of such a system, we introduced two levels of detail: (1) Compounds close to the user's current position are depicted including a 3D representation of the molecule. (2) Outside a specified perimeter, compounds are only represented by a colored sphere. Although this optimization leads to a massive increase in performance due to only rendering a fraction of polygons contained within the set of all structure models, the rendering of thousands or even millions of simple spheres on midlevel hardware in VR is still intractable.

Rather than optimizing the rendering of thousands of spheres in VR, the rendering of colored spheres representing the low level of detail is omitted entirely. Instead, we make use of the particle system that ships with Unity. In real-time graphics, particle systems are usually used to mimic non-solid objects which cannot be emulated using meshes such as liquids, smoke or clouds.[29] Using a static variant

of the particle system, our implementation can take advantage of its shaders that are optimized for rendering large numbers of small sprites (Figure 1a) while avoiding any additional overhead that would be generated by the particle systems physics component. This approach effectively reduces the number of vertices per data point from 42 to 1, assuming low-polygon sphere primitives.

The molecular models representing the structure of compounds found in Drugbank as shown in Figure 1b are loaded from SDF files containing 3D structural information. To render the models in Unity, we used the Unity scripting environment to implement an SDF loader which dynamically translates SDF information to Unity game objects. In order to optimize performance of rendering molecular models, our application does not make use the internal sphere and cylinder primitives but requires custom low polygon models created in Blender.[30] This optimization reduces the number of vertices per sphere (atom) from 515 to 42 and the number of vertices per cylinder (bond) from 88 to 20. Furthermore, meshes of spheres and cylinders within a molecular model that share the same material (color) are combined into a single mesh, thus reducing the number of draw calls. However, this optimization comes at the cost of memory and while the increase in memory usage can be substantial for high-polygon meshes, our use of low-polygon meshes leverages this drawback.

User Interaction Design

An important factor of VR applications is how the user interacts with objects and the virtual environment. To enable natural interaction inside virtual worlds, VR headsets such as the HTC Vive or the Oculus Rift ship with controllers whose position is, just like the position of the headset, asserted by low latency outside-in or inside-out tracking respectively.

An unsolved problem in VR is virtual reality sickness, which causes symptoms similar to these of motion sickness and is currently not well understood.[31] As moving in a virtual environment that represents chemical space through a point cloud is disorienting, even to people not prone to virtual

reality sickness, we chose to implement two well-known optimizations that allow users to explore the virtual environment with fewer negative impacts. (1) To keep with the notion of chemical space, we introduced an independent visual background in the form of an outer space scene.[32] This background provides visual cues as to the orientation of the player in the virtual environment while being subtle enough to not distract from the data points being presented. (2) Research has shown that if the user remains stationary both in the real and virtual world, induction of virtual reality sickness is less likely.[33] Rather than letting the user move through the virtual environment, they use a controller to move the point cloud. In order to reinforce the notion of being stationary while moving the data points, we introduced a floor in the form of a grid acting as a static frame of reference.[34]

We used the engine's collision detection system to allow for the presentation of additional information on molecular structures. Arbitrary information, such as molecular descriptors or names can be shown when the user touches a molecular model with the controller's avatar. This requires each model to be positioned with a collider—an invisible boundary triggering the appropriate text to be shown upon intersection with the controller's collider.

Results and Discussion

As our previous applications for chemical space visualization, their extension into VR allows an intuitive exploration of large sets of compounds. The generation of coordinates and color maps is based on a well-established pipeline applying chemical fingerprints such as MQN and PCA dimensionality reduction. By using the Unity engine and VRTK,[35] the application can be used on both the Oculus Rift and the HTC Vive (through SteamVR) headsets. Additionally, Ximmerse View and Google Daydream are currently supported experimentally. Unity also supports WebVR as a build target. However, the browser as well as device support for WebVR is currently limited, and if available, experimental.
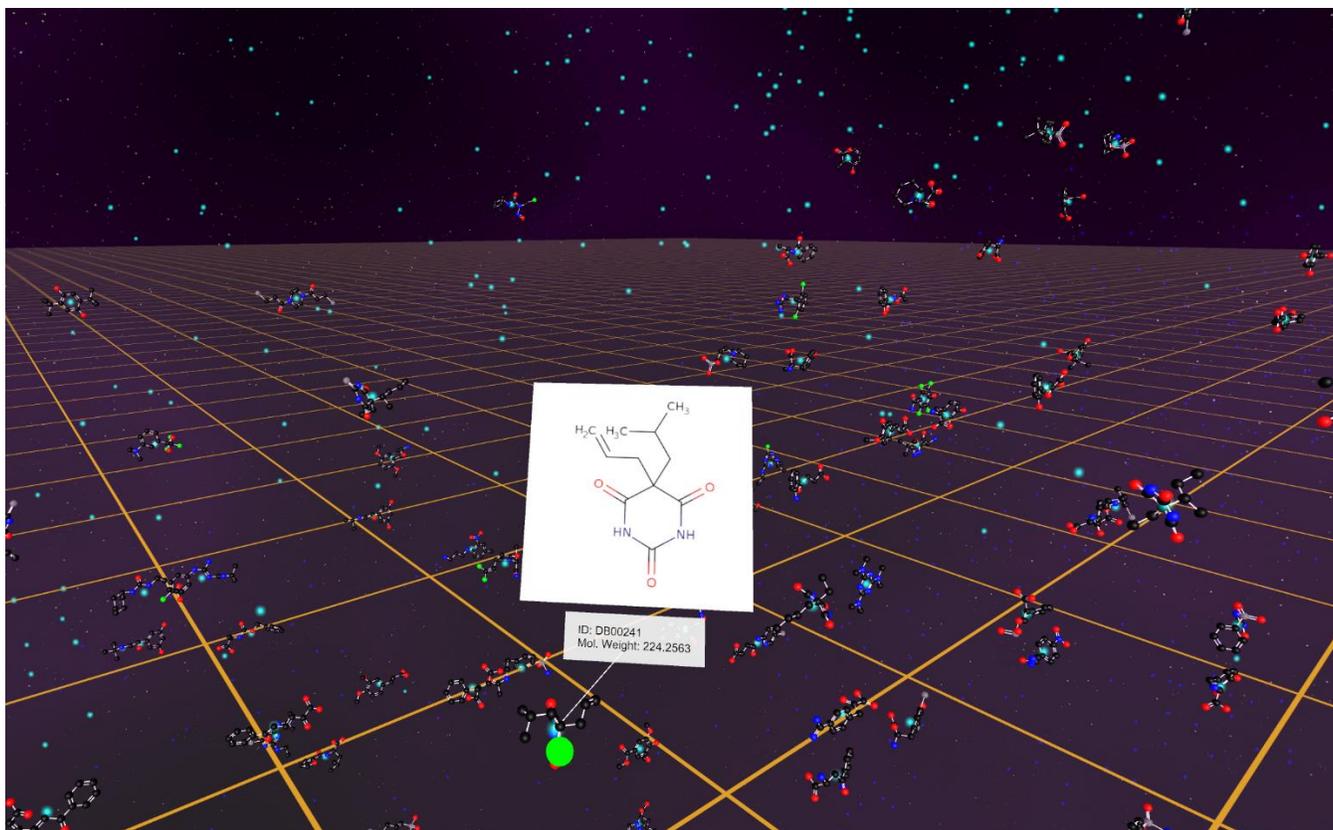
Fig. 2 Optimized virtual reality chemical space. Virtual reality chemical space with a subtle rendering of an outer space scene as an independent visual background. A virtual floor represented by a translucent grid acts as a static frame of reference for the player. The molecular models of close-by data points are dynamically rendered from SDF files. The compounds are depicted using a ball-and-stick model and are colored based on the CPK coloring convention. More distant compounds are represented by a single colored sprite, with the hue ranging from blue to magenta representing the number of rings. When a user touches a molecular model using the controller, the 2D structure drawing as well as additional information, such as the Drugbank ID and the molecular weight, are displayed.

The application loads data in two steps: (1) Coordinate and color data is loaded by parsing the file containing the data line by line. (2) The molecular models are loaded asynchronously from the SDF files. Rather than let the user wait until all the molecular models have been loaded, the user is free to start exploring the chemical space while the models load, and the percentage of models loaded is shown to the user in a head-up display. Once the coordinate and color data are loaded, the user can move the loaded set of data points using the grip-button on the Oculus Touch or HTC Vive controller respectively. Touching the loaded data points in virtual space using the controller displays information on the associated compound, such as its ID and molecular weight. Additionally, a 2D structure drawing of the

molecule is displayed. The 2D structure drawings are loaded as PNG images from the Drugbank website.

The 3D models of the different molecules are depicted as ball-and-stick models and colored according to the CPK coloring convention. Even though the meshes of the models have been heavily optimized by lowering their polygon count, Unity's standard shader implementing the Phong reflection and shading models provide a pleasant level of visual fidelity.[36] The implementation of a two level of detail system, while chosen based on performance constraints has also proven to be advantageous for navigating the chemical space. The global perspective is preserved by clearly indicating positions and properties for compounds further away, while preserving the ability of the user to compare highly localized features such as structural isomers sharing substructures.

Conclusion

Our implementation of an application to explore chemical space in VR provides both a tool to explore Drugbank in VR as well as a framework to create further visualizations of different databases. Virtual reality has shown itself to be a useful tool in fields such as architecture, psychology and aviation training with the goal of simulating real world environments. The application we developed differs in that it does not visualize an environment with an analogue in the real world, but a mathematical construct embedded in 3-dimensional space. Compared to traditional computer applications, the bulkiness of VR headsets as well as the often-occurring eye strain and virtual reality sickness are significant drawbacks. Based on our experience and the current state of technology, we see the application of virtual reality in chemistry mainly in education and training. While interactive visualizations have proven to be an undeniably valuable tool for chemists, it will have to be evaluated whether there is an additional benefit in extending them into VR.

Authors' contributions. D. P. designed and developed the Virtual Reality Chemical Space application and wrote the paper. J.-L. R. co-designed, supervised the project and wrote the paper.

References

1. Wishart, D. S.; Feunang, Y. D.; Guo, A. C.; Lo, E. J.; Marcu, A.; Grant, J. R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; Assempour, N.; Iynkkaran, I.; Liu, Y.; Maciejewski, A.; Gale, N.; Wilson, A.; Chin, L.; Cummings, R.; Le, D.; Pon, A.; Knox, C.; Wilson, M. DrugBank 5.0: A Major Update to the DrugBank Database for 2018. *Nucleic Acids Res* **2018**, *46*, D1074–D1082.

2. Bento, A. P.; Gaulton, A.; Hersey, A.; Bellis, L. J.; Chambers, J.; Davies, M.; Krüger, F. A.; Light, Y.; Mak, L.; McGlinchey, S.; Nowotka, M.; Papadatos, G.; Santos, R.; Overington, J. P. The ChEMBL Bioactivity Database: An Update. *Nucleic Acids Res* **2014**, *42*, D1083–D1090.

3. Papadatos, G.; Davies, M.; Dedman, N.; Chambers, J.; Gaulton, A.; Siddle, J.; Koks, R.; Irvine, S. A.; Pettersson, J.; Goncharoff, N.; Hersey, A.; Overington, J. P. SureChEMBL: A Large-Scale, Chemically Annotated Patent Document Database. *Nucleic Acids Res* **2016**, *44*, D1220–D1228.

4. Ruddigkeit, L.; van Deursen, R.; Blum, L. C.; Reymond, J.-L. Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17. *J. Chem. Inf. Model* **2012**, *52*, 2864–2875.

5. Wang, Y.; Xiao, J.; Suzek, T. O.; Zhang, J.; Wang, J.; Bryant, S. H. PubChem: A Public Information System for Analyzing Bioactivities of Small Molecules. *Nucleic Acids Res* **2009**, *37*, W623–W633.

6. Irwin, J. J.; Shoichet, B. K. ZINC − A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model* **2005**, *45*, 177–182.

7. Awale, M.; van Deursen, R.; Reymond, J.-L. MQN-Mapplet: Visualization of Chemical Space with Interactive Maps of DrugBank, ChEMBL, PubChem, GDB-11, and GDB-13. *J. Chem. Inf. Model* **2013**, *53*, 509–518.

8. Schwartz, J.; Awale, M.; Reymond, J.-L. SMIfp (SMILES Fingerprint) Chemical Space for Virtual Screening and Visualization of Large Databases of Organic Molecules. *J. Chem. Inf. Model* **2013**, *53*, 1979–1989.

9.   Awale, M.; Reymond, J.-L. Web-Based 3D-Visualization of the DrugBank Chemical Space. *J. Cheminformatics* **2016**, *8*.

10.   Awale, M.; Probst, D.; Reymond, J.-L. WebMolCS: A Web-Based Interface for Visualizing Molecules in Three-Dimensional Chemical Spaces. *J. Chem. Inf. Model* **2017**, *57*, 643–649.

11.   Awale, M.; Reymond, J.-L. Atom Pair 2D-Fingerprints Perceive 3D-Molecular Shape and Pharmacophores for Very Fast Virtual Screening of ZINC and GDB-17. *J. Chem. Inf. Model* **2014**, *54*, 1892–1907.

12.   Willett, P.; Barnard, J. M.; Downs, G. M. Chemical Similarity Searching. *J. Chem. Inf. Model* **1998**, *38*, 983–996.

13.   Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model* **2010**, *50*, 742–754.

14.   Probst, D.; Reymond, J.-L. FUn: A Framework for Interactive Visualizations of Large, High-Dimensional Datasets on the Web. *Bioinformatics* **2018**, *34*, 1433–1435.

15.   Ark, W.; Dryer, D. C.; Selker, T.; Zhai, S. Representation Matters: The Effect of 3D Objects and a Spatial Metaphor in a Graphical User Interface. In *People and Computers XIII*; Johnson, H., Nigay, L., Roast, C., Eds.; Springer London: London, 1998; pp 209–219.

16.   Bowman, D. A.; Kruijff, E.; LaViola, J. J.; Poupyrev, I. An Introduction to 3-D User Interface Design. *Presence (Camb)* **2001**, *10*, 96–108.

17.   Teather, R. J.; Stuerzlinger, W.; Pavlovych, A. Fishtank Fitts: A Desktop VR Testbed for Evaluating 3D Pointing Techniques; ACM Press, 2014; pp 519–522.

18.   Mathur, A. S. Low Cost Virtual Reality for Medical Training. In *IEEE Virtual Reality (VR)*; IEEE, 2015; pp 345–346.

19.   Riva, G. Virtual Reality: An Experiential Tool for Clinical Psychology. *Br J Guid Counc* **2009**, *37*, 337–345.

20.   Schultheis, M. T.; Rizzo, A. A. The Application of Virtual Reality Technology in Rehabilitation. *Rehabil Psychol* **2001**, *46*, 296–311.

21.   Pausch, R.; Crea, T.; Conway, M. A Literature Survey for Virtual Environments: Military Flight Simulator Visual Systems and Simulator Sickness. *Presence (Camb)* **1992**, *1*, 344–363.

22.   Merchant, Z.; Goetz, E. T.; Cifuentes, L.; Keeney-Kennicutt, W.; Davis, T. J. Effectiveness of Virtual Reality-Based Instruction on Students' Learning Outcomes in K-12 and Higher Education: A Meta-Analysis. *Comput Educ* **2014**, *70*, 29–40.

23.   Donalek, C.; Djorgovski, S. G.; Davidoff, S.; Cioc, A.; Wang, A.; Longo, G.; Norris, J. S.; Zhang, J.; Lawler, E.; Yeh, S.; Mahabal, A.; Graham, M.; Drake, A. Immersive and Collaborative Data Visualization Using Virtual Reality Platforms. *arXiv:1410.7670 [astro-ph]* **2014**, 609–614.

24. Marks, S.; Estevez, J. E.; Connor, A. M. Towards the Holodeck: Fully Immersive Virtual Reality Visualisation of Scientific and Engineering Data. In *Proceedings of the 29th International Conference on Image and Vision Computing New Zealand*; ACM Press: New Zealand, 2014; pp 42–47.

25. Millais, P.; Jones, S. L.; Kelly, R. Exploring Data in Virtual Reality: Comparisons with 2D Data Visualizations; ACM Press, 2018; pp 1–6.

26. Moran, A.; Gadepally, V.; Hubbell, M.; Kepner, J. Improving Big Data Visual Analytics with Interactive Virtual Reality. In *High Performance Extreme Computing Conference (HPEC)*; IEEE, 2015; pp 1–6.

27. Unity3d www.unity3d.com (accessed Jun 20, 2018).

28. Unreal Engine www.unityengine.com (accessed Jun 20, 2018).

29. Reeves, W. T. Particle Systems---a Technique for Modeling a Class of Fuzzy Objects. *ACM Trans. Graph.* **1983**, *2*, 91–108.

30. Blender www.blender.org (accessed Jun 20, 2018).

31. LaViola, J. J. A Discussion of Cybersickness in Virtual Environments. *ACM SIGCHI Bulletin* **2000**, *32*, 47–56.

32. Lin, J. J.-W.; Abi-Rached, H.; Kim, D.-H.; Parker, D. E.; Furness, T. A. A "Natural" Independent Visual Background Reduced Simulator Sickness. *Proc Hum Factors Ergon Soc Annu Meet* **2002**, *46*, 2124–2128.

33. Ruddle, R. A. The Effect of Environment Characteristics and User Interaction on Levels of Virtual Environment Sickness. In *IEEE Virtual Reality 2004*; IEEE: Chicago, IL, USA, 2004; pp 141–285.

34. Nguyen-Vo, T.; Riecke, B. E.; Stuerzlinger, W. Simulated Reference Frame: A Cost-Effective Solution to Improve Spatial Orientation in VR; IEEE: Reutlingen, Germany; p 8.

35. VRTK (Virtual Reality Toolkit) www.vrtk.io (accessed Jun 20, 2018).

36. Phong, B. T. Illumination for Computer Generated Pictures. *Commun. ACM* **1975**, *18*, 311–317.

Graphics for the Table of Contents: