Visualization of Molecular Data

Alexander Rabinowitsch & Gustav Keppler

Abstract

.

1. Introduction

-leave Vis of Molec. Dynamics out

challenges: -efficient vis. techniques for large datasets –> not enough technology to implment that. Big data is the future! -multiscale vis –> could be an option, implementin this with CellView for example -vis of quantum effects in molecular systems –> note: define quantum effect, probably also possible with cellvie -interactive ray-tracing –> probably high gpu cost, alforithmically and technically complex

In the chapter of molecular data we will describe the data we are using to implement our methods. This are espcially RNA and DNA, as well as Lipids and Proteins.They are the three major biological macromolecules that are essential for all known forms of life. Quelle:https://www.rnasociety.org/what-is-rna

In the methods chapter we will explain the most used techniques in the visualization of molecular data. –> one sentence to every method

In the discussion chapter we discuss the results of our System of the Art Report and give outlooks on the future.

2. Molecular Data

2.1. Biomolecules

The types of molecular data, which are especially interesting for visualization, are biomolecules. Those are any substances produced by cells and living organisms. https://www.britannica.com/science/RNA The major subcategories which are analyezd are nucleic acids, RNA and DNA, as well as proteins.

2.1.1. Nucleic Acids

Deoxyribonucleic acid, short DNA, is the hereditary material in all organisms. It is mostly found in the cell nucleus but also in the mitochondria. DNA information is stored as a code of ade up of four chemical bases: adenine (A), guanine (G), cytosine (C), and thymine (T). Human DNA consists of about 3 billion bases, and more than 99 percent of those bases are the same in all people.The base sequences determines the information available for building and maintaining an organism. Quelle:https://ghr.nlm.nih.gov/primer/basics/dna

RNA, abbreviation of ribonucleic acid, is a complex compound of high molecular weight that functions in cellular protein synthesis and replaces DNA (deoxyribonucleic acid) as a carrier of genetic codes in some viruses. RNA consists of ribose nucleotides (nitrogenous bases appended to a ribose sugar) attached by phosphodiester bonds, forming strands of varying lengths. The nitrogenous bases in RNA are adenine, guanine, cytosine, and uracil, which replaces thymine in DNA. Quelle:https://www.britannica.com/science/RNA

Quelle: https://www.nature.com/scitable/topicpage/protein-structure-14122136/

2.1.2. Proteins

Proteins are generally defined as large molecules which consist of one or more amino acid chains in specific orders, determined by the base sequence of nucleotides in the DNA coding for the protein. Quelle: https://www.medicinenet.com/script/main/art.asp?articlekey=15380 Proteins consist of amino acids which are linked by peptide bonds, The linear sequence of amino acids within a protein is considered the primary structure of the protein.

Quelle: https://www.nature.com/scitable/topicpage/protein-structure-14122136/

Proteines have critical functions in structure, transport and defense of living organisms. Furthermore tremendously imprtant types of proteines are hormons and enzymes. https://courses.lumenlearning.com/wm-biology1/chapter/reading-function-of-proteins/

von wikipedia irgendwo

3. Representation Methods

There are different representation methods for the visualization of the molecular model. The choice depends on analysis task that is intended for the visualization. The models can be categorized into atomistic (3.1) and abstract (3.4) types.

3.1. simple shape

graphical or schematic 2D representation

3.2. arbitrary shape

missleading as structural data

3.3. data-driven shape

Protine databank why so important? ->strucure determines function

3.3.1. atomic

-> conveys scale, complexity, shown atoms, cant deline chains, visualy complex

3.3.2. ribbon style

-> highlights intermediate structure, functional components, domains and linkers

3.3.3. coarse surface style

-> distinguish chains, shows general general shape

3.3.4. tight surface style

-> shows complexity of the structure

3.4. Illustrative and Abstract Models

The models destribed above show the actuals molecules and atoms. Abstract models dilate other features of the molecular data. Those features are not or not unambiguously evident in the atomistic model. This can benefit the understanding due to a a less dense representation or due to reduced coverage in very large molecules. In those complex molecules the general shape is much more important than the single atoms. [**?**]

3.4.1. molecular architecture

Visual abstraction of the molecular architecture can depict structural features more distinctly than an atomistic representation [**?**]. Those abstractions can be seen as different levels of detail that correspond to the underlying structural hierarchy of the molecules [**?**].

In 1981, Richardson [**?**] presented the *cartoon representation* for proteins, which delineates the secondary structure as ribbons and arrows. From that point forward, an assortment of carton renderings have been built up that differ the graphical appearance.

4. Rendering

4.1. Multiscale rendering

4.2. Futher methods

5. Discussion

5.1. Conclusions