

Review of the thesis entitled “Conformational Analysis of an Organometallic Compound with Data Science Inspired Workflow”

The thesis under review by Urmas Pitsi is in the area of computational chemistry and uses conformational analysis to find conformer ensembles for particular organometallic molecules.

To achieve the research goals, the author developed an open-source tool, 'Molli'. Observing achieved results, providing code and literature references, one may conclude that complexity of the problem and proposed solution satisfies the MSc. level thesis requirements of TalTech.

The strengths of the research are the references of the used techniques, which indicates that the author possesses good overview of the field, and open-source tool which may be applied to solve similar problems. Also, conformational analysis is the novel topic among those defending in the Computer Science (Informatics) curricula. Finally, the author explained in detail computational resources need to complete his research and provides time duration references which can be used by the other researchers to plan their work accordingly.

The main criticism of the reviewer concerns the style of presentation, which may prevent one from understanding the results of the thesis and their value.

1. The title of the thesis suggests a stronger connection to data science. The reviewer was unable to find the problem statement in terms used in data science. Also, the workflow itself was not properly described.
2. Background information is given from the chemistry or computational chemistry perspective, whereas the computational part is basically missing. When presenting the concepts of computational chemistry, the author did not give any examples, allowing the reader to understand, try, and follow the concept. It is recommended to present such an example during the defence, preferably illustrated by a proper diagram. Also, immediately in the beginning, the author refers to “level of PBE0/cc-pVTZ” without explicitly defining it. There are a number of similar references/statements throughout the text.
3. Despite the large references section, the motivation to study this particular problem is not explained with references and is not positioned with respect to the literature. The absence of positioning makes it difficult to evaluate the novelty.
4. In the first part of the thesis, the author did not present any validation framework and numeric thresholds to compare the results with. This makes it difficult to follow Subsections 5.6 and 5.8.
5. The methods used by the author rely on the density functional theory; again, no examples of the functionals have been provided. Overall, the presentation is very poor with respect to technical details.
6. Throughout the thesis, the author cites many existing software solutions. Without proper flow chart style diagrams, it is difficult to follow their organization and proportion in the software developed by the author. The reviewer suggests including

such a diagram in the defence presentation slides. In addition, the diagram explaining the general workflow of the experiments would be a welcomed addition.

The reviewer suggests grading the thesis 4-, nevertheless depending on the defence presentation, it may be adjusted by the committee.

Dr. Sven Nõmm
Senior researcher
Department of Software Science
TalTech