Geoffrey Hutchison[1]

[1]University of Pittsburgh

June 10, 2020

Abstract

The work by García, Alvarez, and López outlines their efforts to automate computational chemistry tasks, particularly adsorption of molecules on surfaces and locating transition states relevant for catalysis. In particular, their workflow automates DFT calculations with VASP relevant to relaxing bound molecules to a slab and location of transition states.

Overall the manuscript describes useful work needed for automation of these tasks relevant for heterogeneous catalysis, but lacks specific details and/or code and should compare this work with other similar efforts.

I would currently recommend Revise before accepting the manuscript.

[You can associate my name if you wish]

1 Referee Report

As noted above, the work by García, Alvarez, and López outlines their efforts to automate computational chemistry tasks, particularly adsorption of molecules on surfaces and locating transition states relevant for catalysis.

My overall impression is that there are useful steps towards automation, but several key issues should be addressed before this is a "publishable" result:

- There are many cases of automated computational frameworks - indeed this uses Fireworks, which has been used by the Materials Project. There does not seem to be much in terms of comparing this specific effort to others. For example, a quick search turns up work by (Montoya & Persson, 2017)
- The authors note automated upload to the ioChem archive and the use of embargo/review processing. These are great, although it's not clear whether the archive itself is used in the algorithm outlined in the work. Does the algorithm "learn" from successful relaxations or transition state searches?
- The authors mention comparisons to graph theory / SMILES - but it's not clear that the work here generates a graph or descriptor. If I use either of these methods, I can generate a particular species or even geometry. Does this work generate a unique descriptor / key like SMILES for accessing the data later? Do you need to know the specific database record number? How do you get that from ioChem?
- There are many cases in the text where specific details are omitted. Several of these are listed below - but as it stands, I cannot learn much from the article because the details required for reproducibility are lacking.

For example, the authors note about the transfer algorithm:

> After a few tests, further improvements were integrated to the transfer algorithm.

What are these improvements? What are the specifics of the bond analysis mentioned? What are the specific corrections to the bonding geometry between the molecule and the slab? In addition to being scientifically relevant, it's hard to review the work because it's not clear whether these corrections make sense.

The authors note efforts to automate preparation, transfer and checking, etc. It would be useful to detail some of the effort involved in these and link to the actual scripts/code, which would likely be useful to a wide number of readers, especially those also using VASP.

> Software written in Python and Bash has been developed in our lab and used to script the preparation, transfer and checking processes. Developed Python libraries and scripts focus on geometry manipulation and input/output parsing, while Bash scripts intend to manage the files related to the calculations, and control the execution of VASP.

The authors also note general success with the automated work, but very little about the specifics. For example, while more optimization steps are noted as required in some cases, the authors don't mention how many were in the initial processing, nor how many additional steps were required later. In all cases, the descriptions in this section are merely qualitative commentary, not a discussion of how many cases fit each pattern (e.g., broken bond, migration to a different adsorption site, etc.)

In general in automation, the associated scripts handle more and more cases with revision - without sharing the failures, it's hard to know how to scientifically improve on this work (or address with code revisions) without more details.

1.0.1

References

A high-throughput framework for determining adsorption energies on solid surfaces. (2017). *Npj Computational Materials*, *3*(1). https://doi.org/10.1038/s41524-017-0017-z