NWChem Computational Modeling of Metabolites in KBase

Neeraj Kumar[1]* (Neeraj.Kumar@pnnl.gov), Andrew McNaughton[1], Rajendra Joshi[1], Eric Bylaska[1], Edoardo Apra[1], Lee Ann McCue[1], Gu Tianhao[2], Shane Canon[3], **Christopher Henry,[2] Adam P. Arkin[3], and Bob Cottingham[4]**

[1]Pacific Northwest National Laboratory, Richland, WA; [2]Argonne National Laboratory, Argonne, IL; [3]Lawrence Berkeley National Laboratory, Berkeley, CA; [4]Oak Ridge National Laboratory, Oak Ridge, TN

https://github.com/nkkchem/CompMolNWChem and https://github.com/nwchemgit/nwchem-dockerfiles

Project Goals: The DOE Systems Biology Knowledgebase (KBase) is a free, open-source software and data platform that enables researchers to collaboratively generate, test, compare, and share hypotheses about biological functions; analyze their own data along with public and collaborator data; and combine experimental evidence and conclusions to model plant and microbial physiology and community dynamics. KBase's analytical capabilities currently include (meta)genome assembly, annotation, comparative genomics, transcriptomics, and metabolic modeling. Its web-based user interface supports building, sharing, and publishing reproducible, annotated analysis workflows with integrated data. Additionally, KBase has a software development kit that enables the community to add functionality to the system.

Accurate prediction of quantum chemical properties of metabolites and their structure is needed to understand metabolism and to design new metabolic pathways for engineered systems. Such values can be precomputed for biochemistry databases like the ModelSEED resource that underlies KBase, but support is also needed for metabolites that do not yet appear in these databases (e.g. compounds proposed by cheminformatics applications or compounds involved in newly discovered synthetic pathways). Thus, it is not sufficient just to run property prediction apps in existing databases, but also to make these apps available for users to run themselves on new compounds of interest. To achieve this task, we recently deployed NWChem[1] quantum mechanical code in KBase to automate computational chemistry calculations and obtain optimized chemical structure with partial charges. This app accepts an arbitrary compound-set as input and enables users to select individual compounds from the input set to compute a structure for. The app adds mol2 structures to the input compound set for subsequent use in downstream applications like AutoDock Vina. Of course, it is also useful to precompute structures for all current compounds in the ModelSEED database. However, physics-based computational chemistry calculations are quite challenging to apply to a large database as they grow exponentially with the system size and require high-performance computing resources.[2] Thus, we are in the process of applying NWChem to compute high-quality predicted structures for as many compounds in the ModelSEED as possible, prioritizing the compounds that appear most prevalently in models in KBase. Thus far, structures have been precomputed for all 800

compounds in a *Yarrowia lipolytica* Yeast model. We are also currently exploring how to use HPC resources to power the NWChem app in KBase to greatly speed the pace of structure prediction in this app. In our talk, we will demonstrate this app on some example compounds and discuss the results of our analysis of the *Yarrowia* metabolites. Overall, this new integration in KBase empowers the synthetic biology community to test new compounds with known enzyme targets via docking calculations and paves the way for computation of a broad set of thermodynamic properties (e.g. free energy of formation of metabolites and free energy of metabolic reaction) that are useful for numerous systems biology applications.

References

1. M. Valiev, E.J. Bylaska, et al. "NWChem: a comprehensive and scalable open-source solution for large scale molecular simulations" *Comput. Phys. Commun*. 181, 1477 (2010)
2. N. Kumar, et al. “Mechanistic Implications of Reductive Co–C Bond Cleavage in B_{12}-Dependent Methylmalonyl CoA Mutase” *J. Phys. Chem. B* 123, 10, 2210 (2019).

Funding statement. KBase is funded by the Genomic Science program within the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under award numbers DE-AC02-05CH11231, DE-AC02-06CH11357, DE-AC05- 00OR22725, and DE-AC02-98CH10886. N.K. was supported by the Laboratory Directed Research and Development Program at Pacific Northwest National Laboratory (PNNL). PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830. A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by OBER and located at PNNL.