
ABSTRACT

DESIGNING AND USING A DATABASE
SYSTEM TO IMPROVE THE PREDICTION OF
CHEMICAL PROPERTIES: WADING
THROUGH THE BUZZWORDS

PETER MEULBROEK, RICHARD P. MULLER, AND WILLIAM GODDARD III

The advance in abilities of simulation software when paired with the advance of low-cost “commodity” supercomputers has led to rapid expansions in computational chemistry groups, with a resultant explosion in available simulation results. With the riches in results comes the difficulty of managing large amounts of data. The MSC, like any simulation group, faces issues of data control. These include difficulties in relating input to output, finding and accessing the results of previous simulations, and viewing the results of simulations in a meaningful way. A good solution to managing this data is built on a relational database system accessed using SQL. This solution gives the advantages of a powerful search interface, fast access, and large capacity. We have begun beta-testing a new system, the “MPDB” that has the potential to alleviate many of these data control issues.

The database is designed to address the general problem of using atomistic simulations to predict macroscopic properties. The paper will focus on demonstrating some of the capabilities of the MPDB using an example project. The particular implementation described here is designed to archive, retrieve, and predict chemical properties of hydrocarbons using both experimental and simulation results.

The DOE Basins project requires an extensive chemical properties database, and hence forms a good testing ground for the MPDB. The basins project is concerned with predicting the chemical evolution of sedimentary basins. The focus of the project here at the MSC is to model the chemical behavior of hydrocarbon fluids as they mature and move in the subsurface. Since petroleum is such a complex mixture, a project that hopes to predict the properties of the fluid under a vast range of pressure and temperature conditions needs to access the chemical properties of a large number of compounds. Though these properties are available from the literature for many species, there exist species of interest (such as biomarkers) for which chemical properties of interest are not available, and need to be predicted using molecular simulations. The issues faced by the project are therefore:

- Archiving large number of chemical properties.
- Using simulations to generate unknown chemical properties,
- Visualizing the results of simulations.
- Relating the microscopic simulations to macroscopic properties.

The MPDB is intended to address three important simulation facets: controlling simulations, visualizing the results, and archiving results in a database (see figure 1). For the basins project, the focus is on archiving known properties and calculating new properties.

