What Makes Chemgine Different

Andrei-Dan Filip

October 2023

1 What exists so far

1.1 Molecular Dynamics Programs

Molecular dynamics systems are used to simulate the physical interaction between molecules and are useful in biophysics and physics. These either use established physical formulas or numeric methods and the simulation is generally as accurate as the input properties are.

Some examples are: HyperChem, VMD (Visual Molecular Dynamics)Charm, OpenMM.

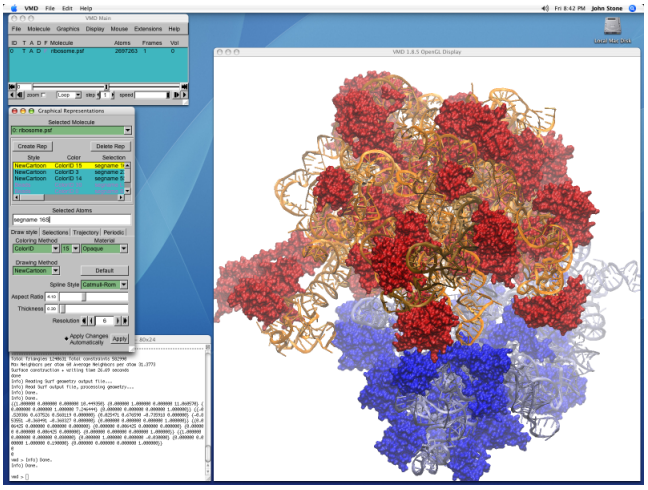

Figure 1: Screenshot from VMD

Molecular dynamics systems were designed to operate at the atomic scale making them impractical for simulating the macroscopic world, where instead of working with thousands of atoms they would have to simulate roughly 10^{23} atoms for just 1ml of water.

1.2 Virtual Laboratory Programs

As the name suggests these are tools that simulate the practical laboratory experience and are primarily used educationally or recreationally.

Some examples are: ChemCollective, Virtual ChemLab, Labster, Beaker.

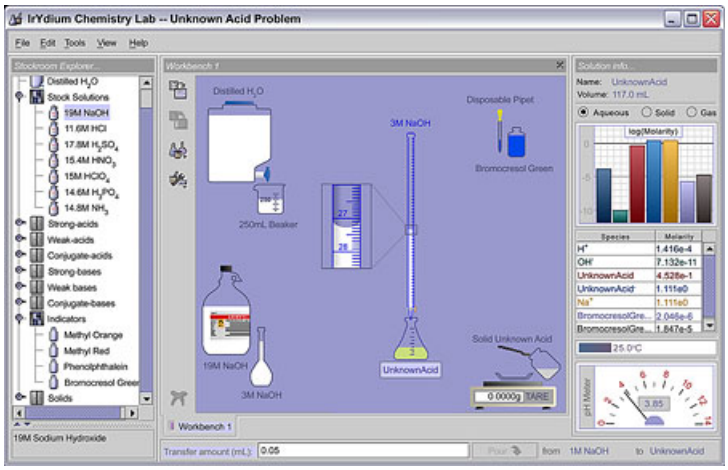

Figure 2: Screenshot from ChemCollective

These programs usually showcase scripted laboratory activities, or only simulate basic inorganic chemistry sometimes with a few predefined common organic substances. This is enough for demonstrating introductory textbook principles but they can hardly be called simulators.

1.3 Chemistry libraries and APIs

These libraries implement common utilities in cheminformatics, like naming conventions, structure rendering and property approximation and are generally implemented in Python.

Some examples are: ChemPy, RDKit.

2 What will Chemgine do

Chemgine will also be virtual laboratory engine but with a much larger set of possibilities, achieved be defining universal rules rather than pre-scripted behaviours.

The term "rule" is used as a synonym to "generic reaction", reactions being considered similar to rules in logic programming languages. An example of such a rule can be:

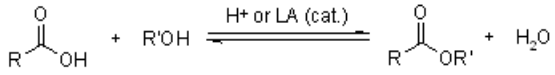

Figure 3: Fischer Esterification Reaction

The reason why this is called a generic reaction is because R and R' can be substituted with any chain of atoms and as a result this reaction is theoretically applicable to an infinite set of reactants producing an infinite set of products.

To compare Chemgine with existing virtual laboratories we could define a metric S_{max} representing the total numbers of different molecules that can be obtained within the systems. In the case of Chemgine we can compute S_{max} as follows.

Let M_0 be the set of initially known molecules.
Let R be a set of rules.
Let F(M, R) be a function the returns all the molecules that can be obtained by applying every rule in R on the set of molecules M.

Then:

$$S_{max} = ||M_i|| \text{ , when: } M_i - F(M_i, R) = \emptyset$$

where:

$$M_{i+1} = M_i \cup F(M_i, R)$$

The number of iterations depends primarily on the generality of the rule set R but it is clear that S_{max} grows quickly with the size of R. In a typical system where organic reactions and molecules have to be hardcoded it is simply infeasible to reach a. high S_{max} value. Whereas with Chemgine the addition of one rule could bring s drastic S_{max} increase.

Additionally the rule set of Chemgine can be enriched or modified by the user. Thus S_{max} is theoretically infinite and allowing for specialized systems to be created using only a subset of the total rules. The practical limit is imposed by the hardware but with certain optimizations this limit should not be a problem for the common user.

When it comes to laboratory setups, Chemgine treats every piece of equipment or glassware as a component with a purpose. The user can connect components in any way and observe the consequences or perhaps benefits of non-standard setups and improvisations. Some such components can work be reactors themselves, after all chemistry doesn't just work inside flasks. This can apply to the atmosphere as well, letting the user know in what dangers he or she is.

Finally, Chemgine will take the form of an open-source library that can be integrated in any new or existing application.