

InterMol: User's Guide

Shirts Group
University of Virginia
Department of Chemical Engineering

May 21, 2014

Contents

1	Introduction	5
1.1	Interoperability of Simulation Formats	5
1.2	Prerequisites	5
1.3	Installation	5
1.4	Abstract Representation	6

Chapter 1

Introduction

1.1 Interoperability of Simulation Formats

Currently, simulation preparation requires a complex sequence of steps to produce a complete, simulation-ready set of molecular coordinates and parameters. These steps are often carried out by hand, or by writing custom scripts for each job which depend on a particular simulation package and set of tools. While numerous tools for specific steps in this process exist, virtually all are tied to specific formats and/or simulation packages, greatly limiting their utility, and needlessly limiting research productivity. These interoperability limitations also make it difficult to compare simulations of the same system across different researcher groups and simulation packages.

This work serves to facilitate the simple conversion of atomistic representations between molecular simulation packages. There have been many prior examples of scripts to convert between different formats such as `acpype`[1] which converts between the Amber simulation format to the GROMACS. Due to the one way nature of such scripts one would require N^2 total scripts to convert between N different simulation packages. Through the use of a universal abstract representation of classical molecular systems, it is possible to reduce the number of necessary scripts to N in order to convert between N packages. Essentially to convert to or from a particular format one must code a translator to and from the universal abstract. Thus a typical workflow to convert to and from one format to another, for example the Maestro¹ format to the GROMACS would require reading in a Maestro simulation into the abstract representation. Then from the abstract representation convert to a GROMACS compatible simulation.

Using InterMol to bridge the file format gap, it will be possible for previously setup simulations to be ported across simulation packages to take advantage of specific tools offered by one package but not another. This ability to reuse previous simulations will be invaluable as it prevents the unnecessary duplication of previous work. Furthermore, the introduction of InterMol will allow the easy interconversion between file formats and force fields which presents researchers with a novel method to benchmark and compare across simulation packages and force fields. Finally, the ability to use simulation setup pipelines such as Maestro across all simulation packages will vastly reduce the burden for researchers to learn how to build systems of interest. Automation will also facilitate the spread of best practices in molecular simulations to prevent human errors.

1.2 Prerequisites

GMXLIB, gromacs binary, python 2.7

1.3 Installation

Fill in this section after we create the egg.

¹Maestro, version 9.2, Schrödinger, LLC, New York, NY, 2011.

1.4 Abstract Representation

System Objects The InterMol abstract representation serves as an intermediate conversion layer between any two simulation packages. The representation is a custom build database to generally store coordinate, and force field parameter for a simulation. Any MD simulation consists of sets of molecules in a simulation box. A list of all **Molecules** in the system as well as box parameters, and other system configuration settings are stored in the **System** object.

Each **Molecule** is a list of **Atoms** which store the name, position, velocity information. as well as the basic atom specific parameters such as σ and ϵ from the LJ-potential or the net charge, and mass. The bonded interactions are stored in a **MoleculeType** each type of molecule in the system is binned into its corresponding **MoleculeType**. All of the bonds, pairs, angles, dihedrals etc. are stored as hashed maps for quick lookup.

In between each of these custom objects, the basic parameters of any general MD simulation can be represented.

Bibliography

- (1) Sousa da Silva, A. W.; Vranken, W. F. *BMC research notes* **Jan. 2012**, 5, 367.