# IMProv Video Tutorial

by

https://www.msstudio.ca/

dschriem@ucalgary.ca
https://www.msstudio.ca/mss-improv/

## GETTING STARTED

In order to prepare the IMProv deployment bundle and execute a modelling run. We are going to look at a sample project available from github to get things started.

## PREPARE IMP TOPOLOGY AND CONFIG FILES

Pull up the MassSpecStudio IMProv project online tutorial

https://github.com/pellst/imp_msstudio_init/blob/master/IMProv_msstudio_tut.md

This presents us with the steps that need to be taken in order to pull together the various data files ( fasta, pdb, em, xl etc. ). The wizard enables us to define the Topology file representation of, for example, the protein structures involved. We also need a configuration file to define the various parameters, such as the number of cpu's for the MPI job run, the number of frames as well as names and characteristics of the structures being included.

The Activity Diagram for Integrative Modeling using IMP presents an diagramatic overview of the Integrative Modeling lifecycle and the Stages involved.

Stage 1 is what we are covering here. This involves defining the various data files that need to be included. This includes preparing the YAML config file and Topology file.

Stage 2 and 3 cover the inner workings of a IMP modeling run and the generation of the rmf files for viewing with Chimera.

Stage 4 is the Analysis phase.

## NEW INTEGRATIVE MODELLING PLATFORM PROJECT

Start MassSpecStudio. Open a new project and select the Integrative Modelling Platform Project template. Give your project a name ( eg: IMP_PRC2 ) and optionally choose an alternative location.

## ADD PROTEINS

Using the Add Proteins wizard screen. Select the FASTA or PDB files to add reference sequences. This will then show the Name and give the the opportunity to customize the Topology by clicking the Manage button under the Topology column for the row with a protein name.

## ADD PROTEIN TOPOLOGY

The Topology record can be edited to set the start and end of the sequence together with the PDB Offset etc. Once you click Ok you will be returned to the Add Proteins wizard screen so that you can do the same for each of the Proteins involved. Once you have completed all the Proteins that you wish to amend. You can click the Next button which will take you to the Add Link Data wizard screen

ADD LINK DATA

Add Link Data wizard screen is where you can add additional data files including Cross-Linking, Hydrogen Exchange, Covalent Labeling and Electron Microscopy. These files will be included in their respective folders for the final output that is generated. Once you have completed your file selections you can click the Next button at the bottom right hand corner of the screen. This will take you to the Configure IMP wizard screen.

CONFIGURE IMP

The Configure IMP wizard screen is where we define the Directory path to export the data files and modeling scripts to. We also set the Sampling Frames and States here. The Ridgid Body and Super Ridgid Body assignments are available through the pick lists provided. The final step is to click the Export button at the bottom right hand corner of the screen. This will produce the folder structure containing the Topology and YAML Config file together with the raw data files that you selected in the wizard steps ( data folder ). It also adds a folder with the modeling scripts needed (imp_model) to perform the job run using the python driver script provided.

DEPLOYMENT STEPS

Pull up the IMProv script deployment, online tutorial. We will start with deploying to Cedar on the Compute Canada cluster. AWS deployment will be covered later. In both cases the pre-requisite steps to setup an account and login to those services is mentioned in the online tutorial and includes links to their getting started guides.

https://github.com/pellst/imp_msstudio_init/blob/master/IMProv_on_Cedar_tut.md

https://github.com/pellst/imp_msstudio_init/blob/master/IMProv_on_AWS_tut.md