Seminar 2

Working with databases - continued

Basic operations

- creating a new database – **MOE | File | New | Database**
- adding molecules from MOE to a database - **DBV | Edit | New | Entry**
- sending molecules from a database to MOE – **DBV | Molecule | Send to MOE**
- sequential viewing of individual records **– DBV | File | Browse**

A major advantage of the databases is that some selected molecule operations can be done in bulk - generating tautomers, deprotonating strong acids, protonating strong bases, removing salt-forming partners from the structure, generating 2D or 3D coordinate from SMILES, generating conformers, energy minimization. These bulk operations are primarily used to prepare input data for molecular modelling.

Task 1

Open **cox-2-inhibitors-pdb-prepared.mdb** from the MOE working directory, which you have stored in the latest practice. This database was created by importing molecules in the SMILES format. The database contains ligands from cyclooxygenase-2 (COX-2) crystallographic structures. Currently, the database contains only two columns - 1. The molecule ID (which can be traced to it in the PDB database); 2. SMILES - this field is not only a linear structure entry, but also a 3D coordinate that we generated when importing a database. Rename this column to "Molecule".

- View individual molecules by browsing the database using **DBV | File | Browse**. Verify that the structures actually have 3D coordinates. Verify that strong acids and strong bases are ionized (should have been performed upon applying the Wash function when importing a database).

Task 2

Add a manually created new diclofenac molecule to the opened database.

- Draw a diclofenac molecule in MOE and prepare it according to the practice of the previous practice.
- Add the modified molecule to the open database - **DBV | Edit | New | Entry**. Add Id: DCL to this newly added molecule.

Task 3

Mass calculations

- Open the bulk descriptor calculation window **DBV | Compute | Descriptors | Calculate**. Have a look at possible descriptors and see what a descriptor can be (discussion). 2D descriptors can be calculated from 2D structures, 3D descriptors need 3D coordinates (optimally energy minimized).
- Add the following descriptors to the database as shown below. By appropriately sorting the database by the lip_violation descriptor, you specify the molecules that violate Ro5 and note which parameter violates the rule.

lip_acc	2D	Lipinski Acceptor Count
lip_don	2D	Lipinski Donor Count
lip_druglike	2D	Lipinski Druglike Test
lip_violation	2D	Lipinski Violation Count
logP(o/w)	2D	Log octanol/water partition coefficient
logS	2D	Log Solubility in Water
Weight	2D	Molecular weight (CRC)

- Sort molecules in database based on increasing molecular weight. Remove such molecules that are not likely to be suitable COX-2 ligands (e.g. metal cations, simple oxides, solvents). Such molecules are residues from crystallographic structures of the PDB. In case of doubt, view the controversial molecules in the MOE and consult with the teacher. Save the cleaned database as (**DBV | File | Save As ..**) cox-2-inhibitors-pdb-cleaned.mdb.
- Compute partial charges for compounds of this clean database (**DBV | Compute | Molecule | Partial Charges**) and then do the energy minimization with the default settings (**DBV | Compute | Molecule | Energy Minimize**).
- Save the resulting database as (**DBV | File | Save As..**) cox-2-inhibitors-pdb-cleaned-minimized.mdb

Basic preparation of the protein (receptor)

3D protein structures (or protein-ligand complexes) in the PDB database usually contain a significant amount of errors and incompleteness to be corrected before using for molecular modelling. These drawbacks include:

- Missing coordinates for hydrogen atoms - this is not an error; it is based on the measurement method. In the vast majority of cases, hydrogen atoms are not detected using X-Ray. Therefore, only the coordinates of so-called heavy atoms are usually given in PDB files. However, it is not a problem to add hydrogen atoms to the software, according to the data on bonds created by heavy atoms.
- Missing coordinates of heavy atoms. Atoms that should be present are not visible. This may be due to errors in crystal measurement or errors in the interpretation of obtained electron maps. Most often, however, this problem occurs in parts of the protein that are increasingly mobile (loops, turns, coils). Depending on which atoms are missing, this error may appear as:
    - breaks in the chain
    - the missing amino acid side chain or part of it
- Alternative positions. From electron maps sometimes it is not possible to decide in which positions there are heavy atoms of the side chain of a specific amino acid. For one particular atom, two or more alternative positions can be written in the PDB file. Prior to using for modelling, one specific position must be selected based on an environment analysis and the others deleted.

Task 4

Open c: /MOE/sample/mol/1err.pdb, which contains the estrogen receptor complex with raloxifene (SERM). The enzyme crystallized as a dimer with two identical subunits (homodimer). Use **System Manager** or **SEQ** to delete all "B" strings (receptor, water, and ligand).

Then follow the instructions here - **Example 2: Manual Preparation of 1ERR**. Pay attention to the types of mistakes you will correct. Save the resulting protein as **1err_prep_basic.moe**.

The above-mentioned protein preparation is necessary for the use of a given structure to molecular modelling. Not all bugs must always be fixed. Obviously, a bug at a ligand binding site or near it will have a major impact on the docking study. By contrast, the same error in the marginal part of the protein may not affect the result at all. In any case, the user should be aware of the imperfections of the protein (receptor).

Other possible modifications of the protein

Protonate3D

Automated application in MOE, which mainly determines protonic states, rotamers and tautomers of individual amino acid residues to achieve the state of the system with the lowest energy. Furthermore, the 3D network of hydrogen bonds (both within the protein and the ligand - solvent - protein complex) is optimized. Below are some examples where experimental X-ray diffraction measurement of a crystal cannot usually determine which condition is more likely. You can use Protonate3D for all these purposes.
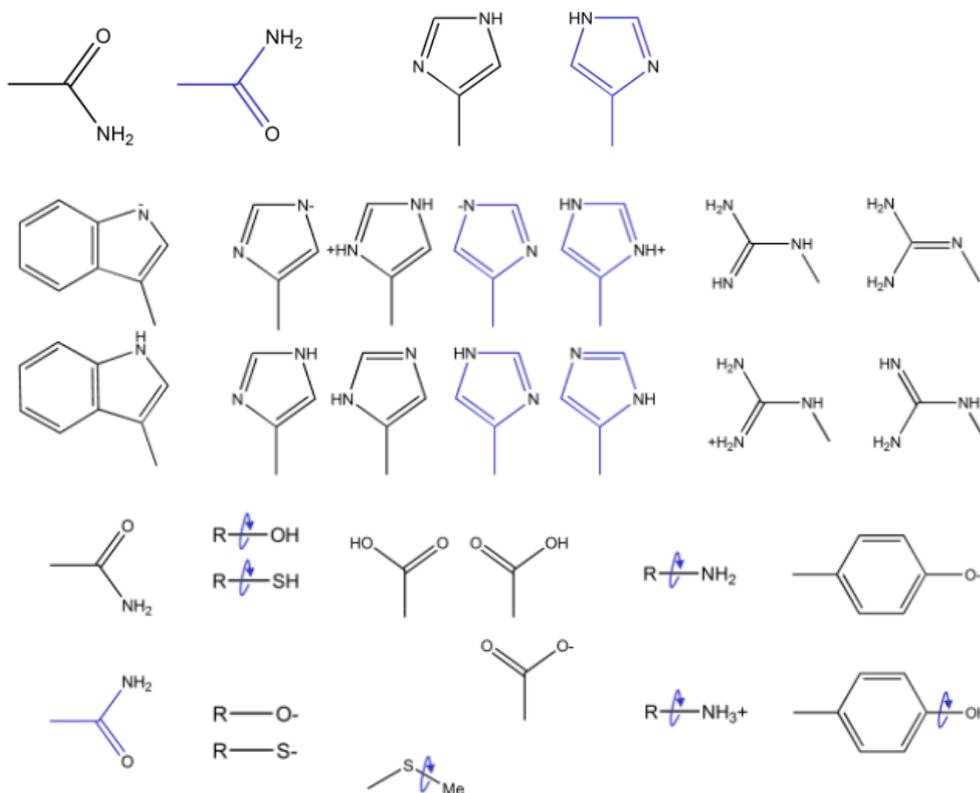
Modify protein using Protonate3D (**MOE | Compute | Prepare | Protonate3D**) with the default settings. Save the result as **1err_prep_Protonate3D.moe**.

Constrained energy minimization of protein

Optional step. For docking, it is usually unnecessary, as docking can minimize the energy of ligand-receptor complex. On the contrary, the minimization is necessary for molecularly dynamic studies where a well-minimized system is required as input. Minimization takes place on the basis of molecular mechanics (force field) and we usually apply some constraints to it so that the position of important atoms does not change too much from experimentally determined positions.

Use the **MOE | Compute | Energy Minimize** command to minimize system. Set as instructed by the assistant. Save the resulting system as **1err_prep_minim.moe**

Follow ligand-receptor interactions as instructed by the assistant. **RHS | Ligand | Ligand Interactions**. Save the interactive diagram as an image.