From systems biology to personalized medicine:
- Tutorial Molecular Drug Design

A.J. Kooistra MSc (a.j.kooistra at vu.nl) and Dr. C. de Graaf (C.de.Graaf at vu.nl)
Division of Medicinal Chemistry, VU University Amsterdam
http://www.vu-compmedchem.nl

Aim of the tutorial

In the following tutorial you will learn how to use computer-aided drug discovery techniques to analyze and predict protein-ligand interactions and solve the following (types of) questions:

1. How molecular interactions determine the selectivity and polypharmacology of GPCR targeting drugs

Why is doxepin (Sineqan/Prudoxin) an antagonist/inverse agonist of multiple aminergic G Protein-Coupled Receptors (including the histamine H_1 receptor and the muscarinic acetylcholine M_1 receptor), while its chemical analogue olopatidine binds H_1R selectively (and has low affinity for M_1R)? How can you use this information to discover/design new selective H_1R antagonists to target inflammation and pruritis?

2. How tyrosine kinase amino acid mutations cause cancer drug resistance

Why is the Bcr-Abl1 tyrosine-kinase inhibitor imatinib (*Gleevec*) not able to inhibit the T315I mutant of Bcr-Abl1 kinase, while ponatinib is able to efficiently inhibit both wild-type and T315I mutant forms of this kinase? How can you use this information to discover/design new inhibitors the Bcr-Abl T315I mutant to target chronic myelogenous leukemia?

3. How cytochrome P450 genetic polymorphisms affect drug metabolism

Why does CYP2D6.49, a polymorphic form of Cytochrome P450 (CYP) 2D6, have a reduced dextromethorphan O-demethylation activity, but an increased dextromethorphan 7-hydroxylation activity than wild-type CYP2D6 and the polymorphic isoform CYP2D6.10? How can you use this information to predict differences in the metabolism of specific drugs by individuals that have the CYP2D6.49 polymorphism?

4. How to solve your own personalized medicine research question

Instead of (or in addition to) addressing questions 1-3 you can also consider these thee questions as examples/templates to address related questions that you are trying to solve in your own research projects (i.e. on other combinations of protein targets and ligands).

In the current tutorial you will learn how to:

- Mine target annotated chemical databases and protein structure databases
- Built a protein homology model
- Analyze protein-ligand complexes and compare ligand binding sites
- Derive important pharmacophore features from ligands and/or protein binding sites
- Use the protein-ligand interactions analysis to explain and predict protein selectivity

Methods and techniques

In the current workshop you will use the general modeling program **MOE** (www.chemcomp.com), and the databases **UniProt** (www.uniprot.org, containing protein information), **Protein databank** (www.rcsb.org, containing protein structures), and **ChEMBLdb** (www.ebi.ac.uk/chembldb/index.php, containing target annotated bioactive molecules).

Timeline

Fig. 1 gives an indication of the timelines for the workshop tutorial.

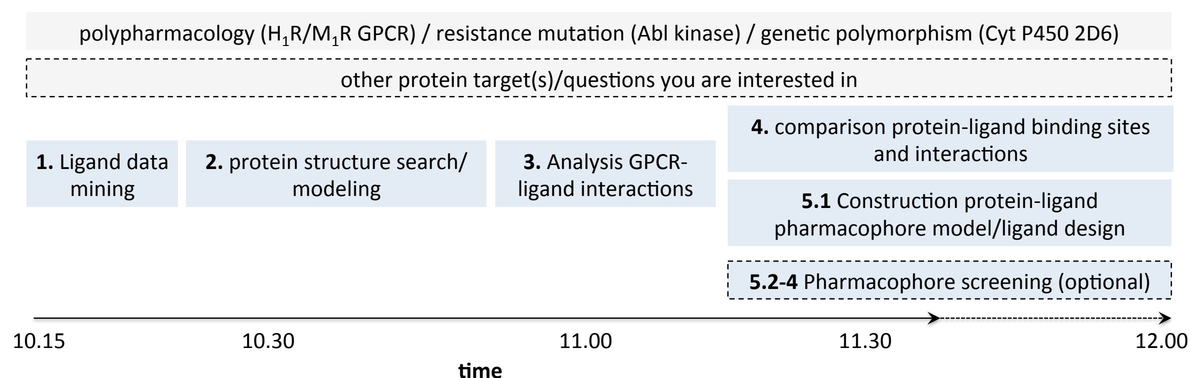

Fig. 1: *Road map following the different steps of the tutorial. The depicted work plan follows a scenario in which only one of the three questions/target classes is investigated, but if one works efficiently, it is in principle possible to address multiple questions and/or try to solve a (similar) question connected to your own research project (please discuss this with one of the instructors).*

1. Biological activity data mining/ligand analysis

*This exercise will help you to extract biological activity data of specific compound(s) and compare their selectivity for different protein targets. Vice versa, you can search biological activity data for all ligands of a specific protein target (see optional exercise **5.2**).*

ChEMBLdb (www.ebi.ac.uk/chembldb/index.php) contains a target-annotated database of bioactive molecules.

> Fill in the name of the compound you are interested in and search for "*Compounds*".

> Alternatively, you can draw the chemical structure of the compound (or a chemical substructure) and search on similar molecules and/or molecules that share the same chemical substructure. You can retrieve the "smile" string that represents the molecular structure of a specific compound from, for example, *PubChem* (www.ncbi.nlm.nih.gov/pccompound), and copy and paste this smile string in the "Ligand Search" Tab under "*Smiles Search*".

> Check the boxes of the compound(s) and indicate the type of data you want to download (e.g. *IC50, Inhibition, Ki, Kd* data if you are interested in inhibition capacity and/or binding affinity): *Please Select … - Filter Bioactivities*: add the data you want to download)

> In the next step, indicate that you want to download the data in Excel format (*Please Select … - Download All Data (XLS)*).

> Open the Excel file and search for the biological activity data for the protein targets you are interested in (by text search and/or ordering according to the *PREF_NAME* column). Note that you can also search for the Uniprot *PROTEIN_ACCESSION* code (www.uniprot.org) of the target.

P.S.: Please note that ChEMBLdb does not cover scientific journals such as *Molecular Pharmacology* and *Drug Metabolism and Disposition*, therefore, please don't hesitate to perform a parallel "manual" literature search to obtain the relevant biological activity data.

Q1*: What are the (relevant differences in the) biological activity data of the compound(s) for the protein(s) you are investigating?*

2. Protein structure search/homology modeling

This exercise will help you to analyze what structural information is available for your protein target(s), and build a protein homology model if necessary. This can for example help you to guide rational ligand design, compare different ligand binding modes for your protein target, and compare ligand binding sites of your target and homologous proteins.

Determine whether there is/are experimentally determined structure(s) (e.g. crystal structure) available for the protein you are investigating using one or more of the following databases (choose the approach that is most appropriate/efficient for your search):

- **2.1 Protein Data Bank** (if there is a crystal structure available for your target protein(s))
- **2.2 Build a protein homology model**

N.B.: In case you have to construct a protein homology model for your target protein, please be aware that the model will be rough. In particular protein modeling based on protein crystal structure templates that have low homology to the protein target can be very challenging (as for example illustrated in the GPCR DOCK 2010 competition: Kufareva (2011) *Structure* 19: 1108-1126, doi: 10.1016/j.str.2011.05.012).

2.1 Protein Data Bank

Search in the **Protein Data Bank, or PDB** (http://www.pdb.org), for your target protein.

> You can for search on keyword: (e.g. “histamine H_1 receptor”), a Uniprot ID (e.g. P35367), or extract the relevant PDB accession codes from the scientific article describing the protein crystal structure (e.g. Shimamura (2011) *Nature* 475, 65-70. doi: 10.1038/nature10236; pdb accession code: 3RZE).

> Save the pdb file: click on “Download files” next to the bold PDB code → PDB file (text)

Q2: *What ligand is co-crystallized in the protein-ligand complex (and what is its pdb code)?*

If you only want to change a small number of specific amino acid residues in your protein (i.e. create specific point mutants), you can do open the pdb file in MOE and change the specific amino acid residues via the Sequence Editor:

> Find the residue that you want to mutate by selecting at least one atom in the structure and then go to the Sequence Editor (**Window** | **Sequence Editor)**, alternative if you already know the residue number you can directly go the sequence editor. Find the selected residue in the **Sequence Editor** or find the correct residue number by showing the sequence numbers above the sequence (**Sequence Editor** | lower right corner, **button “13A/ALA”**). Subsequently go to **SE | Edit | Mutate** and select the target residue-type that you want to introduce, then click on the selected residue and you have created your mutant (hit escape or click on the red X in the upper-left corner to exit the mutation process).

If there is no experimental structure model available of your (wild-type or mutant) target protein, you can construct a homology model based on homologous protein(s), as described in section **2.2**.

2.2. Build a protein homology model (optional)

First assess whether there are experimental structures available of homologous proteins of your target protein using a sequence alignment search.

> Find the Uniprot identifier (e.g. P35367) and *Sequence* of your target protein in Uniprot (www.uniprot.org) and save the amino acid sequence in FASTA format (*target.fasta*).

> Go to **BLAST** (http://blast.ncbi.nlm.nih.gov), select "protein blast" and paste the Uniprot identifier (e.g. P35367) into the *Enter accession number(s)* box and select "Protein Data Bank proteins (pdb)" in the Database list to search only the structure (homologs) that are deposited in the Protein Databank. And click on the "BLAST" button.

> Note the top PDB entries (Accession column) with the BLAST search and note their similarity to your target protein. Carefully check:
- whether the *PDB* entries of the identified homologue(s) (see *3D structure databases*) are indeed experimentally determined (e.g. *X-ray*) or computational *models*
- what part (amino acid residue *Positions*) of the protein the 3D structure is determined.

Q2: ***Which crystal structure is a suitable modeling template for your target protein?***

> Search for the PDB entry (or entries) in the **PDB** (see **2.1**)

> Start the general modeling program **MOE** and open a pdb file of the homologous protein that you want to use as a template to built a homology model of the target protein. Check whether the co-crystallized *ligand* in the crystal structure template (see **Q2**) also binds to the target protein you want to model (via ChEMBLdb or manual literature search, see **1**).

> If the pdb file contains multiple chains you can choose to use only one chain for building the model, and delete the other chain: In the **Sequence Editor**: select the chains you want to delete with the left mouse button and *Delete* them.

> You can furthermore delete residues that are not present in the original protein (e.g. that are part of fusion proteins that are used to stabilize the protein construct for crystallization): In the **Sequence Editor**: select the residues you want to delete with the left mouse button (while holding the shift button), click on the right mouse button, and *Delete* them. Tip: check the **PDB** for the residue numbers actually matching your target.

> Open the sequence of the target protein (that you saved earlier in fasta format in **2.2**) in the **Sequence Editor**.

> Perform an alignment between the amino acid sequences of the homologous protein (template) and target protein: **Sequence Editor | Alignment | Align**, select the template and target and *Apply* (the sequences of the template and target should now be aligned in the *Sequence Editor*)

> In order to include the co-crystallized ligand in the homology-modeling template in the modeling procedure, select (only) the ligand (Ctrl + left mouse button and select the ligand).

Alternatively, you can include (a substructure of) a ligand that is similar to the co-crystallized ligand or try to dock a known active compound into your homology model. In case you want to do this, please ask the instructors for advice.

> Construct a homology model of the protein target: **Sequence Editor | Protein | Homology Model** (you can *Include Selected Atoms as Environment for Induced Fit* if you want to include the co-crystallized ligand (or similar/docked ligand) in the homology modeling template in the modeling procedure). Select your protein target in the first *sequence* box and the template in the *template* box, change models from 10 to 1 (to save time) and click on OK.

3. Analysis protein-ligand interactions

In the current exercise you will analyze the molecular interactions between your target protein and its ligands in order to determine the molecular determinants of ligand binding.

> Open the downloaded pdb file of one protein-ligand complexes (or homology model) of your target protein in MOE.

> Analyze protein-ligand interactions (**Compute | Ligand Interactions**). N.B.: Check if the "**Protein:**" and "**Ligand:**" are properly defined in the **Ligand Interactions** interface (you can use the scroll buttons on the right to check different combinations). Then click **Apply** and **Isolate** to zoom into the protein-ligand binding site.

> Select and highlight the ligand (**Select | Ligand,** or select the ligand with your left mouse button while holding the *Ctrl* key) and change the color of the carbon atoms (under **Render | Atoms**).

> Investigate the residues surrounding the ligand and try to find the residues that are making key interactions with the ligands.

Q3: Use the MOE protein-ligand interaction diagram (and visual analysis) to assign complementary pharmacophore features in the ligand and in the sidechains of the residues of the protein binding site.

Which pharmacophore features (H-bond acceptor, H-bond donor, anionic, cationic, aromatic) did you define?

Repeat the analysis for the other protein-ligand complex you want to investigate to address you research question (e.g. to compare H_1R-doxepin vs. M_1R-doxepin).

4. Comparison protein-ligand binding sites

This exercise will help you to compare different ligand binding sites and ligand binding modes of different proteins. This can for example help you to guide ligand design by e.g. targeting specific residues in the binding site (that are similar ("polypharmacology") or different ("protein selectivity") between protein targets.

Please note that for:

- *GPCRs* (e.g. Kooistra et al. (2013) *Br J Pharmacol* 170: 101-126. doi: 10.1111/bph.12248)

- *Kinases* (e.g. Van Linden et al. (2014) *J Med Chem* 57: 249-277. doi: 10.1021/jm400378w)
- *CYPs* (e.g. de Graaf et al. (2005) *J Med Chem* 48: 2725-2755. doi: 10.1021/jm040180d)

general amino acid residue numbering schemes have been defined to facilitate the comparison of the structures/binding sites of different receptor subtypes/isoenzymes.

4.1 Structural alignment of protein structures

Investigate similarities and differences in the ligand binding sites of different proteins (e.g. H_1R vs. M_1R, wild-type vs. T315I mutant Bcr-Abl kinase, wild-type CYP2D6.1 vs. polymorphic CYP2D6.49) by superposing their crystal structures and/or homology models (**2.2**).

> Superpose the structures (**Window** | **Sequence Editor**, then in the sequence editor **Alignment | Superpose**) but not before cleaning up the structures (removing fusion proteins, antibodies, multiple chains etc.) to avoid problems in the structural alignment.

> In the **Sequence Editor**: **Display**, check *Actual Secondary Structure* (icon in the bottom right corner with an arrow over “ALA”) and color the different structures (**Render** | **Ribbon** | **Color**: *Chain*).

4.2 Comparison of binding site residues

> You can identify the name and number of a certain residue by selecting it with your left mouse button while holding the *Ctrl* key and go with your right button to **Atoms** > *Label: Residue*. Try to select a specific residue which forms interactions with the ligand and identify its residue name and number.

> Vice versa, you can also select a specific residue in the **Sequence Viewer** (**Selection | Atoms | Of Selected Residues**) and see the selection in the **MOE Viewer**. Try to select a specific residue which is known to be important for protein-ligand binding in the **Sequence Viewer** and identify its location in MOE.

Q4: Identify the similarities and differences in protein-ligand interactions and amino acid residues in the ligand binding sites of the different proteins.

Identify/predict which residues and protein-ligand interactions determine ligand selectivity between the different proteins?

5. Pharmacophore modeling

A pharmacophore model describes the spatial arrangement of molecular features a ligand must contain to bind a target protein. In the current exercise you will construct customized pharmacophore models and investigate how you can use a pharmacophore model to predict protein-ligand binding modes and steer ligand discovery and design.

5.1 Construction pharmacophore model

> Based on your analyses in exercises **1-4**, create and refine 1-2 pharmacophore models for your protein target (and/or homologues). First create a new (empty) pharmacophore model (**File | New | Pharmacophore Query**).

PQ *-> Pharmacophore query (not the main MOE screen, but the pharmacophore query screen)*

Based on your answers to **Q3** and **Q4**:

> Add/delete pharmacophore features in the **Pharmacophore Query Interface** by selecting the pharmacophore feature points on the ligand that have now appeared in the main screen of MOE.

> You can furthermore add/delete ligand-based inclusion volumes (**V**: *Included*) and/or protein-based exclusion volumes (**V**: *Excluded*). Keep in mind that certain regions in the proteins are more/less flexible than others.

> Finally, notice that you can customize the pharmacophore models by making specific features essential (**F:** *Essential*), defining the number of features to be satisfied (*Enable Partial Match*), and adapting the feature radius (**R:**). Base these settings as much as possible on your knowledge of protein-ligand interactions (e.g. if you want to emphasize specific interactions that are selective for a specific protein/mutant/polymorphic isoform, see **4.2**).

> Save your pharmacophore model (**PQ | Save**) and give the file the name *target_ph4_1/2.ph4* and click OK.

Pharmacophore screening (e.g. Sirici et al. (2012) *J Chem Inf Model* 52: 3308-3324. doi: 10.1021/ci3004094) or molecular docking (e.g. de Graaf and Kooistra et al. *J Med Chem.* 2011 54: 8195-8206. doi: 10.1021/jm2011589) studies can be used to discover new biologically active molecules. If you are interested in the validation and/or application of your pharmacophore model for virtual screening, you can go through optional steps **5.2-5.4**).

5.2 Retrieve ligands of your target protein (OPTIONAL)

This exercise will help you to extract ligands and biological activity data of the protein target you are interested in. This can, for example, help you to select known reference ligands for benchmarking of experimental assays and/or identify useful structure-activity relationship (SAR) ligand data sets for your protein target.

> In **ChEMBLdb** (www.ebi.ac.uk/chembldb/index.php) search for “Targets” and check the box of the target (make sure you also select the correct species)

> You can display or filter the data based on specific endpoints (within specific values for the type of molecules that you want to obtain, see the first question) via the button on the top right (*Please Select …*).

Q5*: What type of ligands do you want to investigate for the protein you are interested in (e.g. agonist, antagonist/inverse agonist, inhibitor, inducer)?*

> You can save the data after filtering/displaying as an Excel file (top right button *Please Select – Download All Data (TAB)*). We suggest that you filter the ligand data to make sure you have molecules in your selection with the biological activity data you are interested in (see **Q5**, e.g. filter on Ki/IC50/EC50 ≤ 1000 (nM), pKi/pIC50/pEC50 ≥ 6).

If the ChEMBLdb does *not* contain any ligands for your target, but you have found ligands from another source or if you want to add a specific set of ligands (e.g. your own), you can construct the molecules in **MOE** by using the builder, then add them to the database (**DBV | Edit | New | Entry**)**.** If necessary create a new database (**MOE** | **File | New | Database**).

> **File | Open** select the tab file and assign a name (*ligands.mdb*) to the MOE db.
DBV *-> Database viewer (not the main MOE screen, but the database screen)*

> Remove the column ASSAY_TYPE (**right-click on the column name** – **delete**)

> Protonate the compounds (**DBV | Compute | Molecule | Wash**, set *Source Field* to *CANONICAL_SMILES* and *Destination Field* to *CANONICAL_SMILES*, check the *Add Explicit Hydrogens, Protonate Strong Bases* and Deprotonate Strong Acids boxes).
Perform a conformational search (**Compute| Pharmacophore | Conformation Import**, explicitly define the Output Database and (**Add**) the Input File)).

5.3. Pharmacophore search (OPTIONAL)

> Perform pharmacophore searches of the conformational databases against the pharmacophore model(s) you constructed. Open one the pharmacophore queries (if you closed it), push *Search*, define the conformational database from the previous step as *Search Database*, select 1 hit per molecule (**Hits**: *Best Per molecule*: **N**: 1), and give the output database a name (**Output**: *ph4_screen.mdb*).

> Determine how many molecules fit the pharmacophore. Evaluate your pharmacophore models and settings and select the models/settings with which you retrieve many known ligands of the protein target and:

- few decoys
- few known ligands of the undesired "off" target
- many known ligands of the desired "second" target to identify dual target ligands (ADVANCED)

Q6*:* What is the *enrichment* of known ligands of the desired protein target over decoys/ligands of the undesired protein target?

5.4. Analysis pharmacophore fits (OPTIONAL)

> Browse the fits of actives/decoys (**DBV | Browse**) in the different (optimized) pharmacophore models (display the fits in the corresponding protein-ligand complex). Not only the compounds that fit the model are interesting, also try to assess why some of the actives did not fit into the pharmacophore model and why some decoys could be fitted into the pharmacophore. Could you further optimize the pharmacophore model to make it more sensitive or selective?