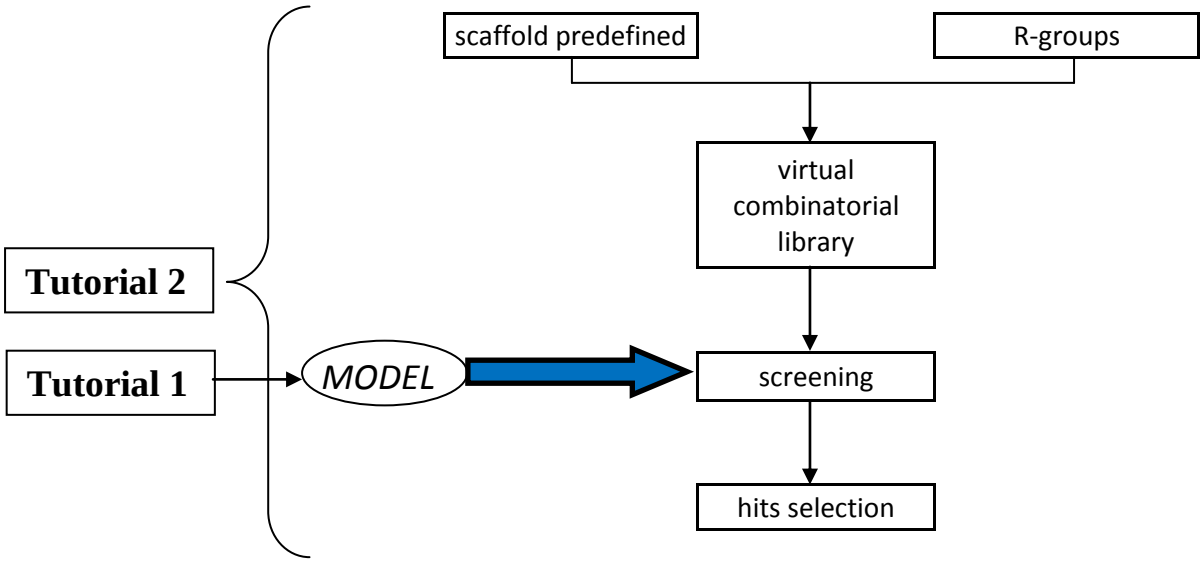
Tutorials on Library Design

E. Lounkine and J. Bajorath (University of Bonn)
C. Muller and A. Varnek (University of Strasbourg)

The purpose of this tutorial is to generate a library of potential inhibitors for cGMP.



Tutorial 1. QSAR modeling - Binary model.

Formatting data.

Open gmp_inhibitors.mdb.
File > open > Open in Database Viewer.

Construction of binary models needs two distinct values for actives and inactives. Here the activity *pIC50* take many possible values. Then we decided to consider as active inhibitors only molecule with **pIC50 > 5**.

DBV | *Edit > New > Field*.
New Field Type: **int** *Name:* **ACTIVITY** *Value:* **0**, and press OK.

DBV | *Edit > Clear > Entry Selection.*

DBV | *Edit > Select.*
Add To Entry selection: **pIC50** *matches* **> 5**.

DBV | *Compute > Calculator.*
Check toggle ON *Use Selected Entries Only*.

For Available Fields and Destination Field choose *ACTIVITY*.
Click on the number 1 in the calculator panel.
Press Evaluate.

Calculating Molecular Descriptors.

DBV | Compute > Descriptors > Calculate:
Choose "2D" as *descriptor Class.*
Enter "sl" as *Filter* and select all SlogP_VSA descriptors.
Enter "sm" as *Filter* and select all SMR_VSA descriptors.
Press OK.

QSAR Model.

DBV | *Compute > Model > QSAR.*
Activity Field: ACTIVITY.
Method: Binary.
Select *SlogP* and *SMR* descriptors.
Click on Fit, then Report.
(The button Validate permits to perform a cross-validation (Leave-one-out method), then predictive performance on cross-validation is available in the Report section).
Press Save for saving the model.

<u>Tutorial 2.</u> Quinazoline Library – CombiGen

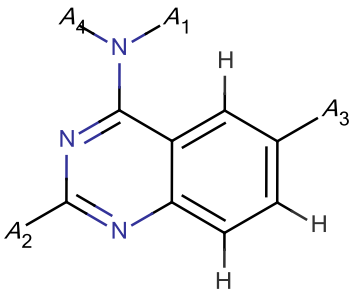
Scaffold library

Open the database file: $MOE/sample/mol/gmp_scaffold.mdb
File > open > Open in Database Viewer.

Expand the molecule cell in the Database Viewer (DBV) to view the scaffold.

Right click on the structure and *Send To MOE*.

Label the scaffold atoms by name in the MOE window to view the connection points.
MOE | Render > Atoms > Label: Name



R-Groups Library

Open the databases in $moe/sample/mol:

gmp_r1.mdb
gmp_r2.mdb
gmp_r3.mdb
gmp_r4.mdb

Expand a molecule cell to view the A0 port.

Library Generation

Open the panel *MOE | Compute > QuaSAR > QuaSAR-CombiGen*

Choose *gmp_library.mdb* as Output Database name.
Check ON the Open Validation Report option.
Click on Define and Edit Connections.

Click *MOE* to use the scaffold present in MOE Window.
Select sequentially R-Group databases (gmp_r1.mdb, gmp_r2.mdb, gmp_r3.mdb, gmp_r4.mdb) and click on Add Connection.
Press OK when finished.

The defined connections should appear in the QuaSAR-CombiGen panel.
Press OK to generate the library.

Check in the report file if there is no error.
6480 products have been generated.
Each product molecule is encoded as:

C1./A1.1./.../A3.2./.../AX.N.

C1 is the scaffold, AX refers to the x-th connection point AX; the number N is the R-Group database entry number of the R-Group substituted at the x-th position.

Library Enumeration – Evaluate the completed library.

Open the panel *DBV | Compute > Model > Evaluate.*

Browse to select *gmp_binary.fit* for the Model File.
Press OK.

Predicted results are written to a new database field: $PRED.
These results are probabilities that the entry will be active.

Open the panel *DBV | Compute > Sort*.

Sort by Field $PRED.
Select Descending.
Press OK.

In this case only results for which value in $PRED field is higher than 0.5 are considered to be active.

Select *DBV | Edit > Select*.

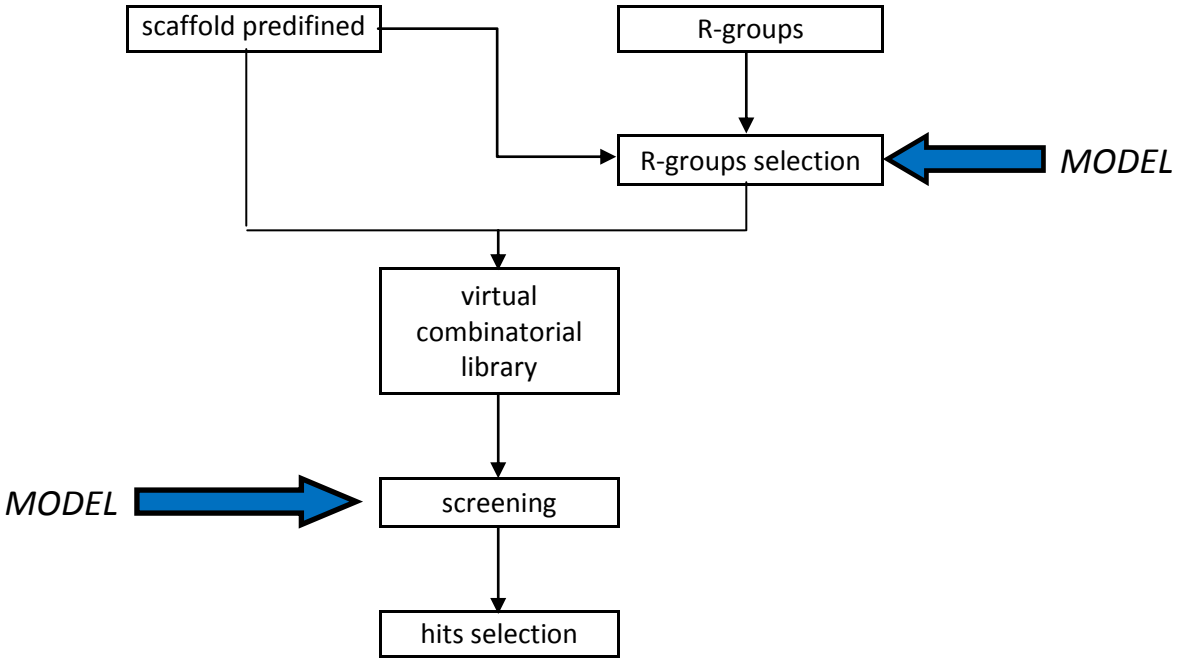
Select Entries where $PRED >= 0.5:
Add To Entry selection: **$PRED** *matches* **>= 0.5**.
Press OK.

Tutorial 3. Quinazoline Library – Reagent

R-Groups selection with QSAR Models.

QSAR Reagent for cGMP.

The R-Group (and scaffold, if applicable) databases are randomly sampled and the binary model applied. The more times an R-Group appears in a compound deemed active by the model, that R-Group's "score" goes up.



Open the QSAR Reagent panel:
MOE | Compute > QuaSAR > QuaSAR-Reagent.

For Scaffold Output Database browse into your personal directory and give a name for the database. (default: *qcrsel_c.mdb*).
Open the Define Connections panel and do as for the Tutorial 1.
Press OK.

The defined connections should appear in the QuaSAR-Reagent panel.
Press Next.

Use the default Lipinski Rule-of-five filter default values.
(if the generated compound is out of the panel ranges, then it is skipped.)
Press Next.

Browse to find Model File “gmp_binary.fit”.
Set Iterations to 20.
Press Next.
Press Start.

QSAR Reagent Results.

Open the four A*.qcrsel_gmp_r*.mdb created.

The $ACT field contains an estimate of the proportion of active compounds in the entire virtual library that contain the corresponding component.
Sort each database in descending order on the $ACT field: right click on the field name and choose sort descending.

QSAR Reagent Enumeration.

For A1.qcrsel_gmp_r1.mdb select top 5 reagents.

Delete the other reagents in the database:
Select DBV | Edit > Invert > Entry Selection.
Edit > Delete > Selected Entries.

For A2.~r2.mdb and A3.~r3.mdb select top 3 reagents.
For A4.~r4.mdb select only the hydrogen.
(These four databases will be the new reagent databases for use in generating a virtual library.)

Open *MOE | Compute > QuaSAR > QuaSAR CombiGen*.
For output Database browse into your directory and give a name to the database. (ex: gmp_library_enriched.mdb)
Select Define and Edit Connections:
Browse to select the scaffold database.
Select sequentially R-Group databases:
A1.qcrsel_gmp_r1.mdb
A2.qcrsel_gmp_r2.mdb
A3.qcrsel_gmp_r3.mdb
A4.qcrsel_gmp_r4.mdb
Press OK when finished.
Press OK to generate the library.

Open *DBV | Compute > Model > Evaluate*.
Browse to select *gmp_binary.fit* for the Model File.
Press OK.

Sort down the predicted probability of activity ($PRED).

In this case only results for which value in $PRED field is higher than 0.5 are considered to be active.

Select *DBV | Edit > Select*.

Select Entries where $PRED >= 0.5:
Add To Entry selection: **$PRED** *matches* **>= 0.5**.
Press Apply.

Tutorial 4. RECAP – Simple Retrosynthetic Analysis

Examination of structural features common to the most active compounds.

RECAP Analysis for GMP Inhibitors.

Open ***gmp_inhibitors.mdb***: *MOE | File > Open.*

Select *DBV | Compute > RECAP > Analysis*.

For Activity Field choose *pIC50 > 5*.
(The Activity Field, operator and threshold value determine whether the input compounds are treated as active or inactive.)
Press OK.

The results will be written in *gmp_inhibitors_recap.mdb*.

Sort Down the *freqA* field. This field corresponds to an activity adjusted frequency value.

apo: the number of attachment points in each RECAP fragment.
freq0, freq1: frequency of each RECAP fragment in inactives and actives, respectively.

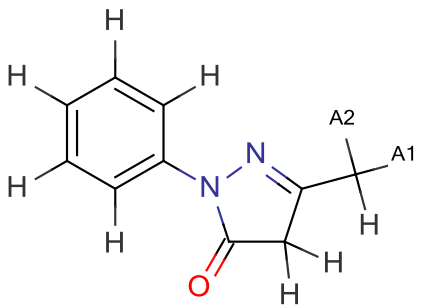
Tutorial 5. Preparing R-Groups and Scaffold.

Pyrazolone Library Scaffold.

Construct the pyrazolone scaffold with the Molecule Builder:
MOE | Edit > Build > Molecule.

Sequentially select two of the H atoms and use the Molecule Builder to rename as A1 and A2.

Save the molecule as a MOE file pyrazolone_scaffold.moe:
File > Save.



R-Groups from SDF Files.

Open the SD File alkyl_halides.sdf:
MOE | File > Open > Import SD File.
Toggle ON the *New Database* and *Open Database Viewer* options.
Press OK.

DBV | Compute > CombiChem > Clip R-Groups.
Clip Modes: Leaving Groups (the group is clipped and the atom at the clip point is marked as A0).
Select the *alkyl halide* entry from the list of clipping groups.
Destination Field: Other, *Specify*: clip.
Press Apply.