KNIME Data Talks

How can KNIME make your Structure-Based Virtual Screening Campaign more successful?

Emilie Pihan – 01/06/2022

discngine

Statement about SBVS setup

Calibration is a key phase

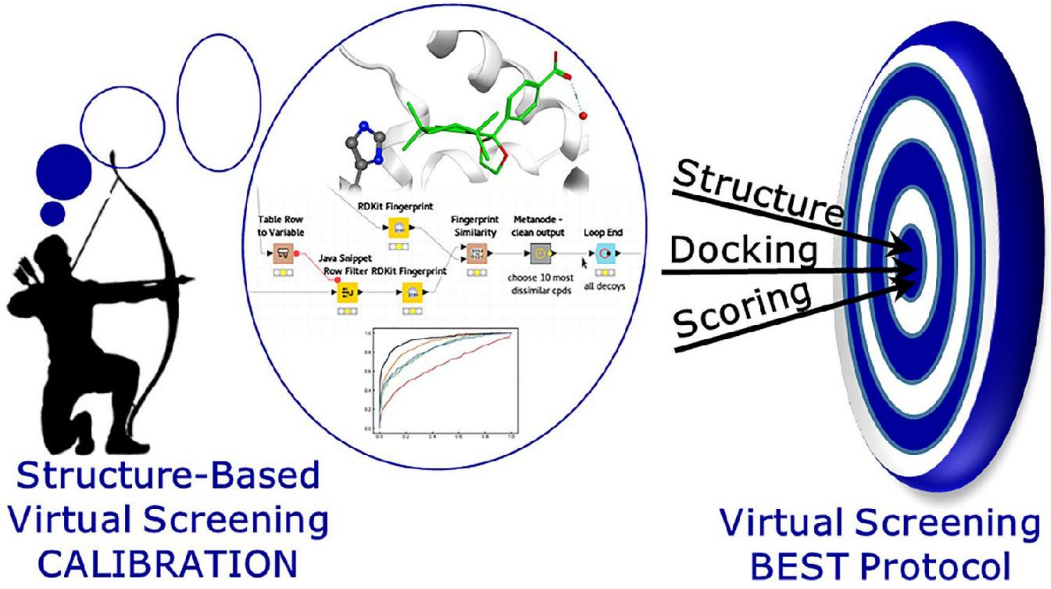
- Impacts the quality
- Increases the hit rate

Process goals

- Automatized
- Standardized
- Shared

Parameters to be evaluated

- Protein structures
- Docking software
- Scoring functions



Pihan E, Kotev M, Rabal O, Beato C, Diaz Gonzalez C. Fine tuning for success in structure-based virtual screening. J Comput Aided Mol Des. 2021 Dec;35(12):1195-1206.

Adapt and use the WF

Open for Innovation KNIME Hub

Workflow

Evotec_StructureBasedVirtualScreening_Calibration

Decoy | Sbvs | Calibration | Docking | Virtual screening | +4

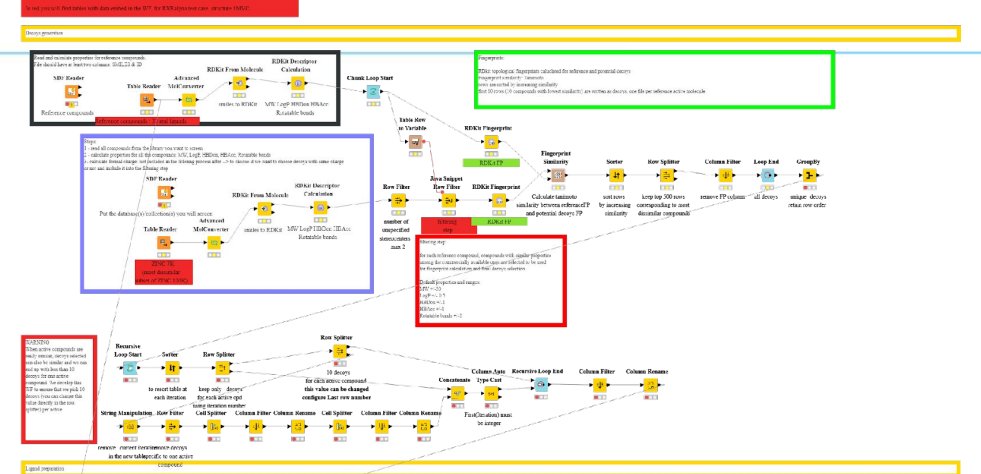
1. **Decoys generation**

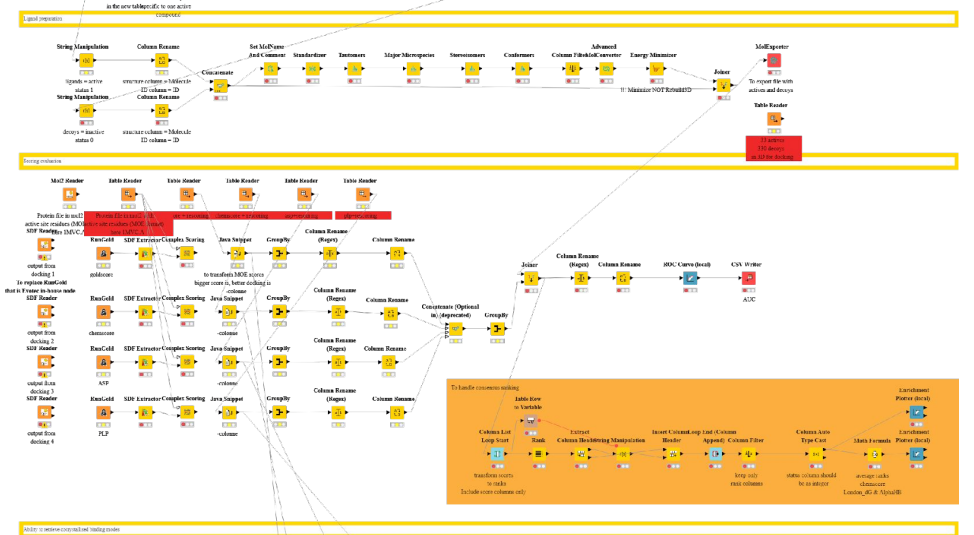
2. **ROC curves generation**
   - > Docking
   - > Rescoring
   - > Ability to retrieve actives before decoys

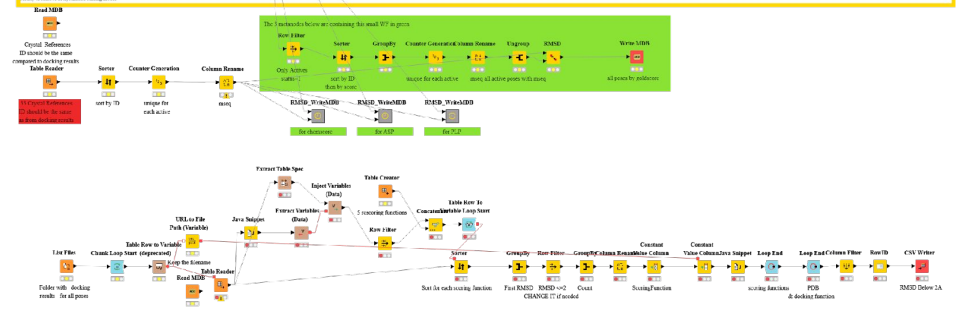
   Adapt with your own tools

3. **RMSD calculation**
   - > Ability to retrieve binding modes

4. **AUC/RMSD plots**
   - > Compromise to select the best protocol







Decoy generation

What is a decoy ?

A molecule that (probably) won't bind to your target (*DUD-E database*)

Why generating decoys?

To have virtual inactive compounds

To build ROC curves and estimate/compare docking model performances

How to generate decoys?

Starting point

- Active molecules

Same physicochemical properties

- MW +/- 50
- logP +/-0,5
- HBDon +/-1
- HBAcc +/-1
- Rotatable bonds +/- 2

BUT dissimilar fingerprints

How many decoys?

10 decoys per active compound is enough (in DUD-E, it is 50)

Huang N, Shoichet BK, Irwin JJ. Benchmarking sets for molecular docking. J Med Chem. 2006 Nov 16;49(23):6789-801.
Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. J Med Chem. 2012 Jul 26;55(14):6582-94.
Chaves AE, Jain AN. Structure- and Ligand-Based Virtual Screening on DUD-E+: Performance Dependence on Approximations to the Binding Pocket. J Chem Inf Model. 2020 Sep 28;60(9):4296-4310.

ROC curve and AUC

ROC : Receiver Operator Characteristic

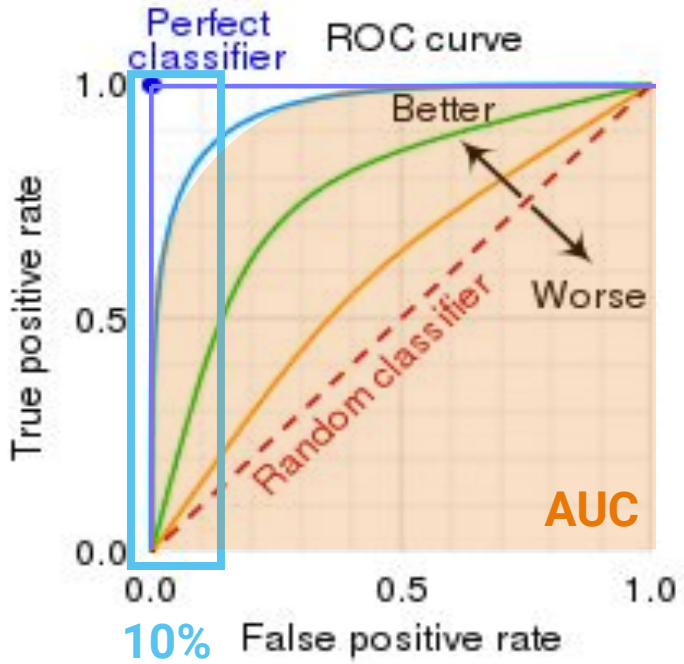
- Performance of a classification model at all classification thresholds
- FPR versus TPR

AUC : Area Under the Curve

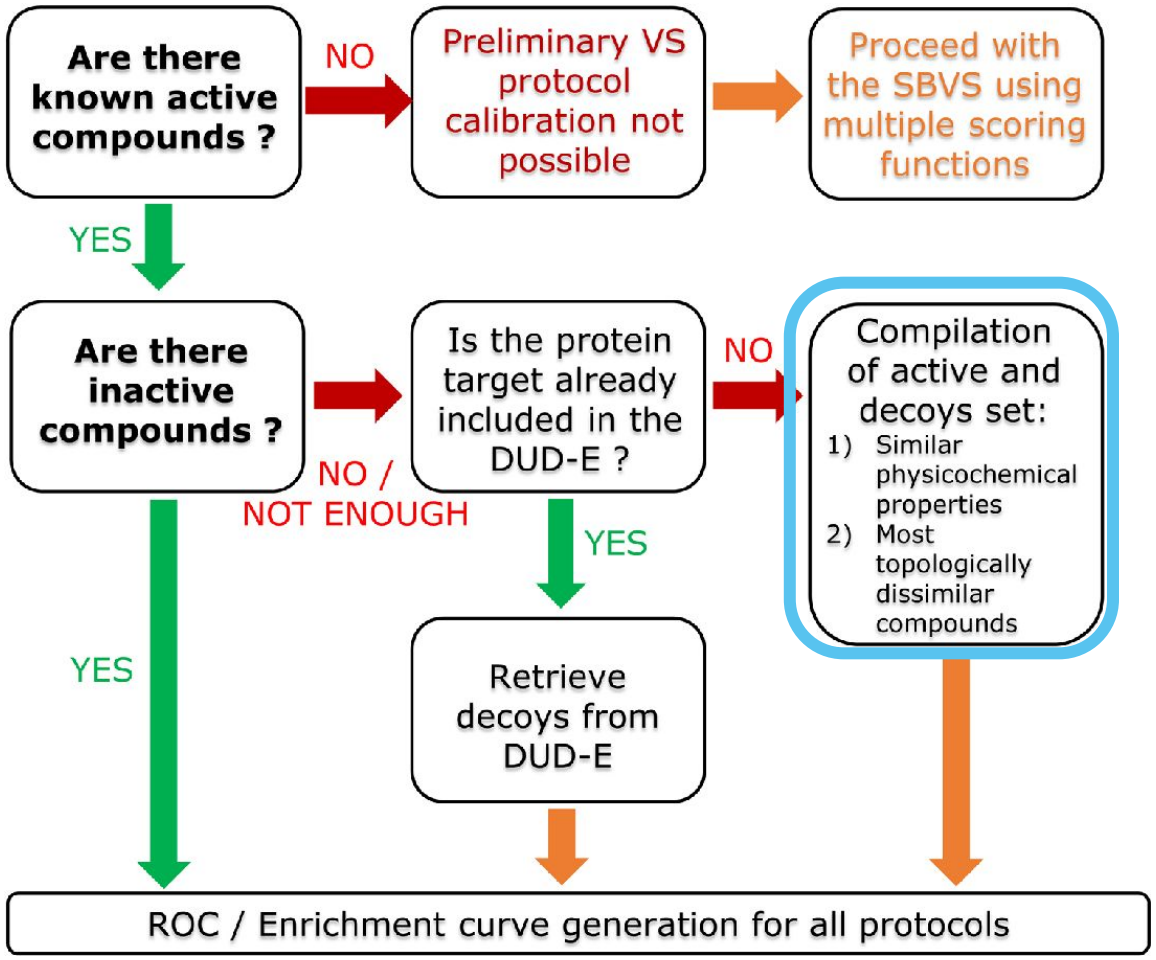
- Global model performance
- Between 0 and 1
- Bigger the better

What is the value for SBVS calibration?

- Comparing different protocols quickly
- Finding the best protocol enriching first part of the ranked database in active molecules

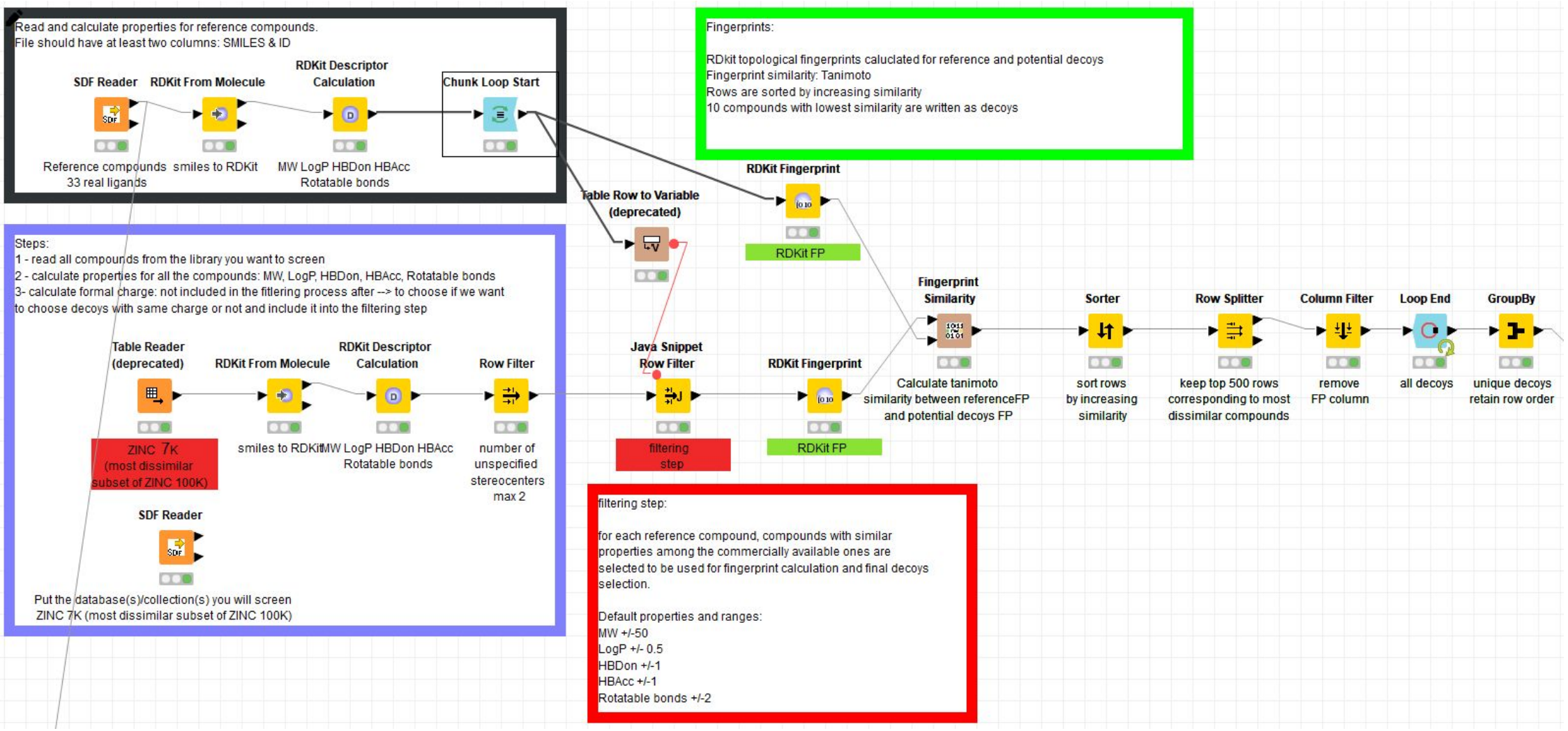


Focus on the decoy generation decision tree



The first section of the workflow will handle this specific case

Focus on the decoy generation workflow



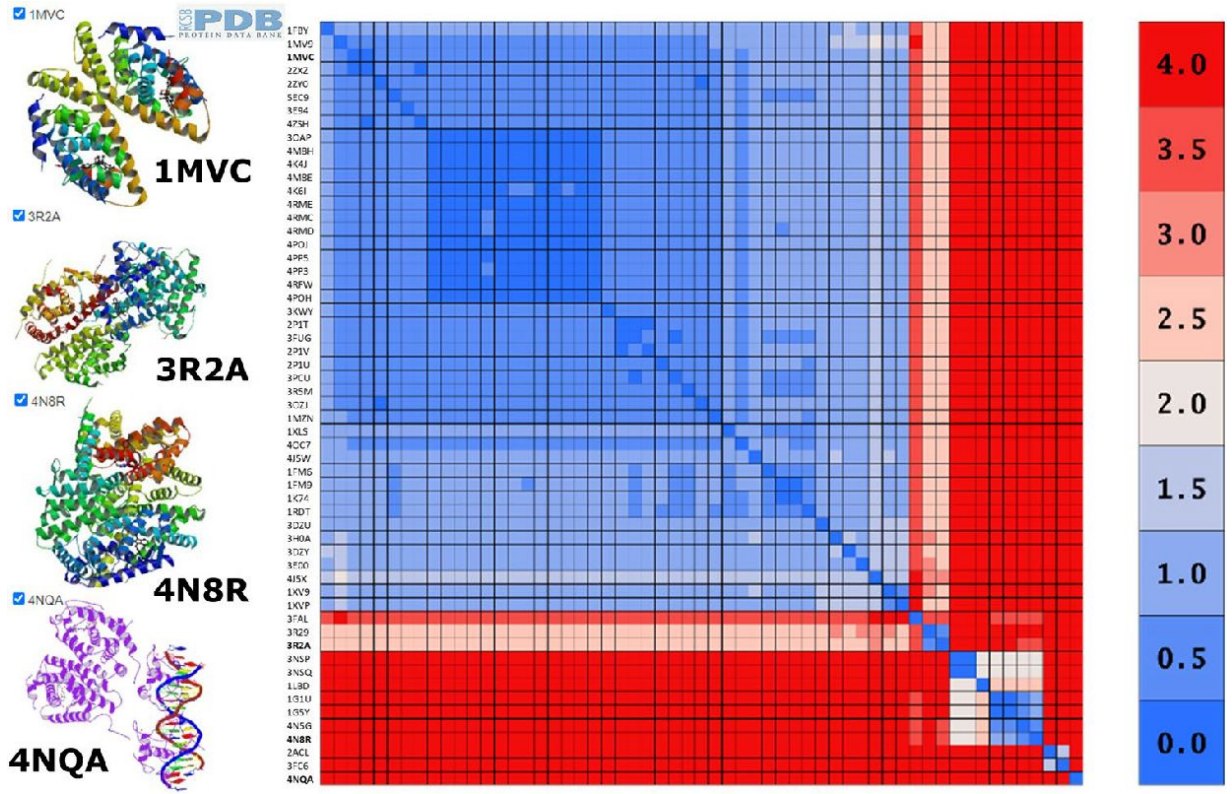
CASE STUDY

Retinoid X Receptor alpha (RXRα)

Input data for RXRα setup

4 representative structures
- 1MVC
- 3R2A
- 4N8R
- 4NQA

- 33 active compounds (cocrystallised ligands)
- 330 decoys generated through the workflow



Target structure selections from the PDB based on clustering for the application case of the RXRα receptor. Selected representative structures are highlighted in bold. The colored rule represents RMSD values, blue to red for more or less similar conformations.

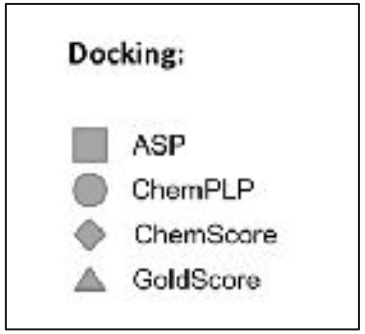
Tools

Chemistry

RDKit
Open-source
cheminformatics
software

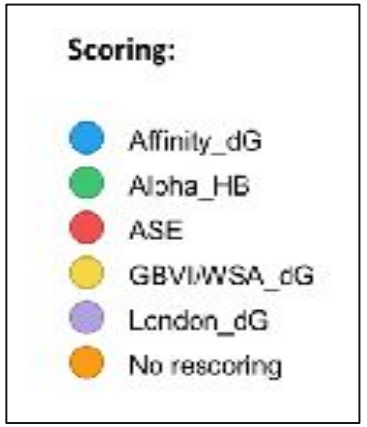
Docking

Gold
CCDC
In-house nodes



Rescoring

MOE
CCG



RMSD

MOE
CCG

Output data

96 Combinations

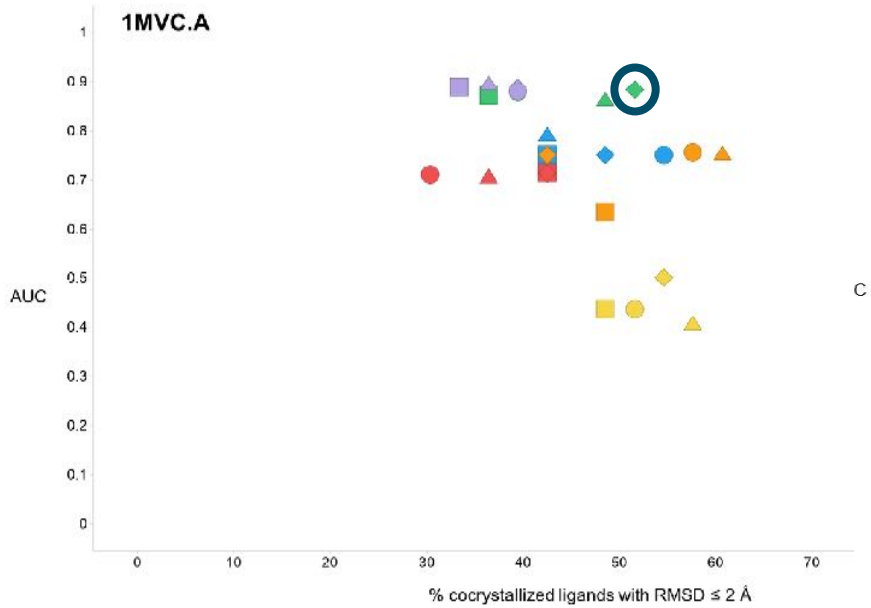
- **4** Structure **1MVC.A**
- **4** Docking **ChemScore**
- **6** Rescoring **Alpha_HB**

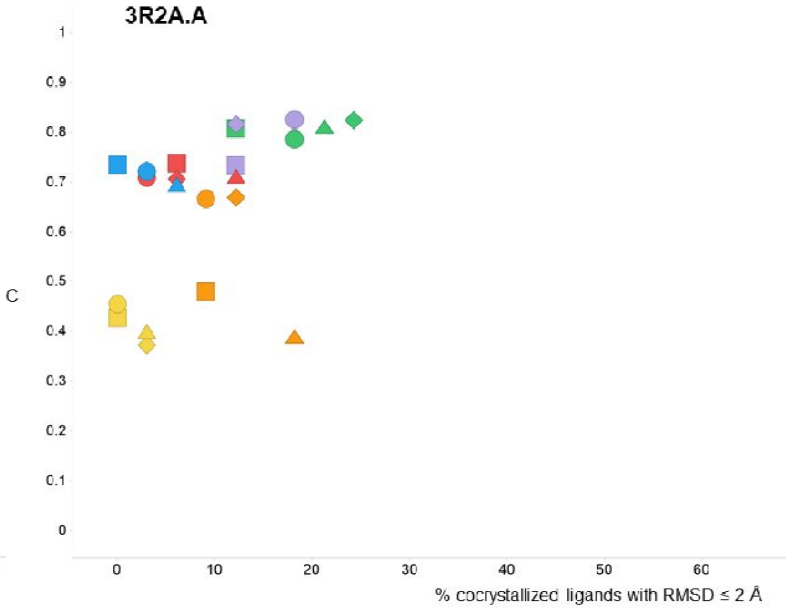
Docking:

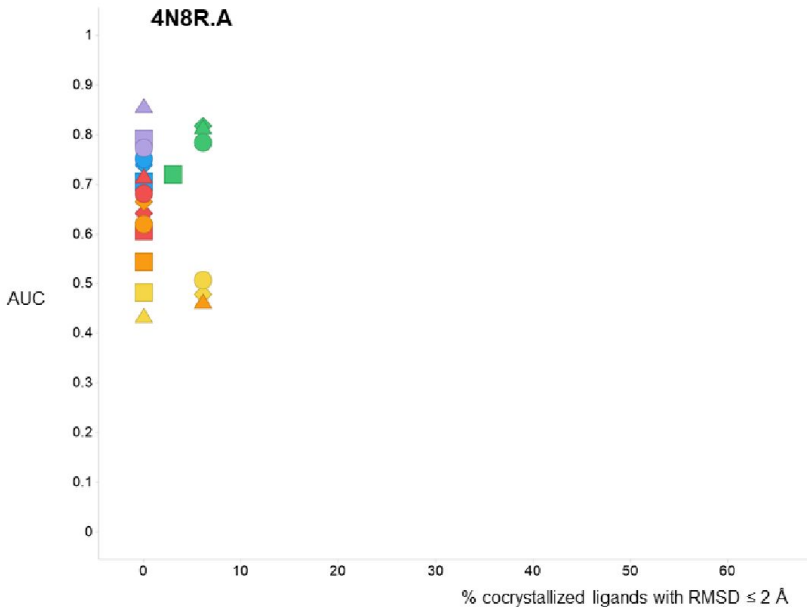
- ASP
- ChemPLP
- ChemScore
- GoldScore

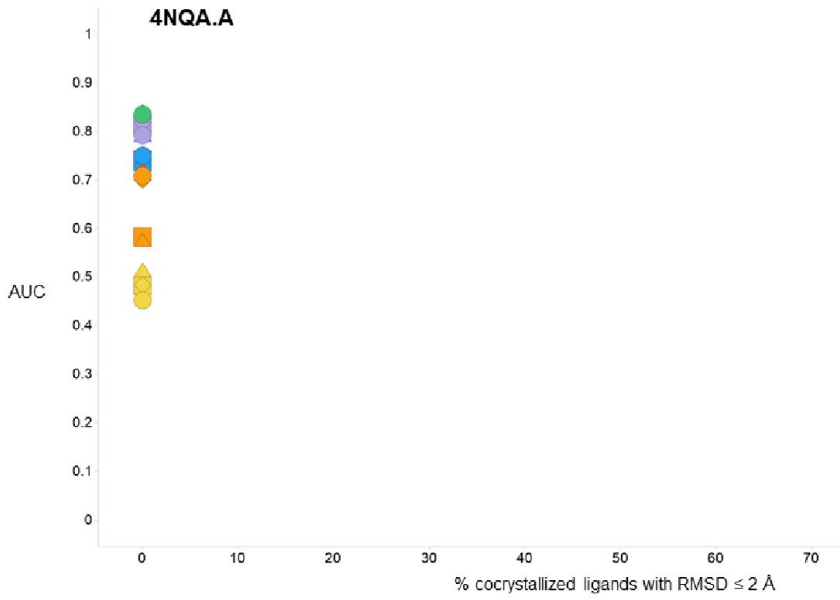
Scoring:

- Affinity_dG
- Alpha_HB
- ASE
- GBVI/WSA_dG
- London_dG
- No rescoring



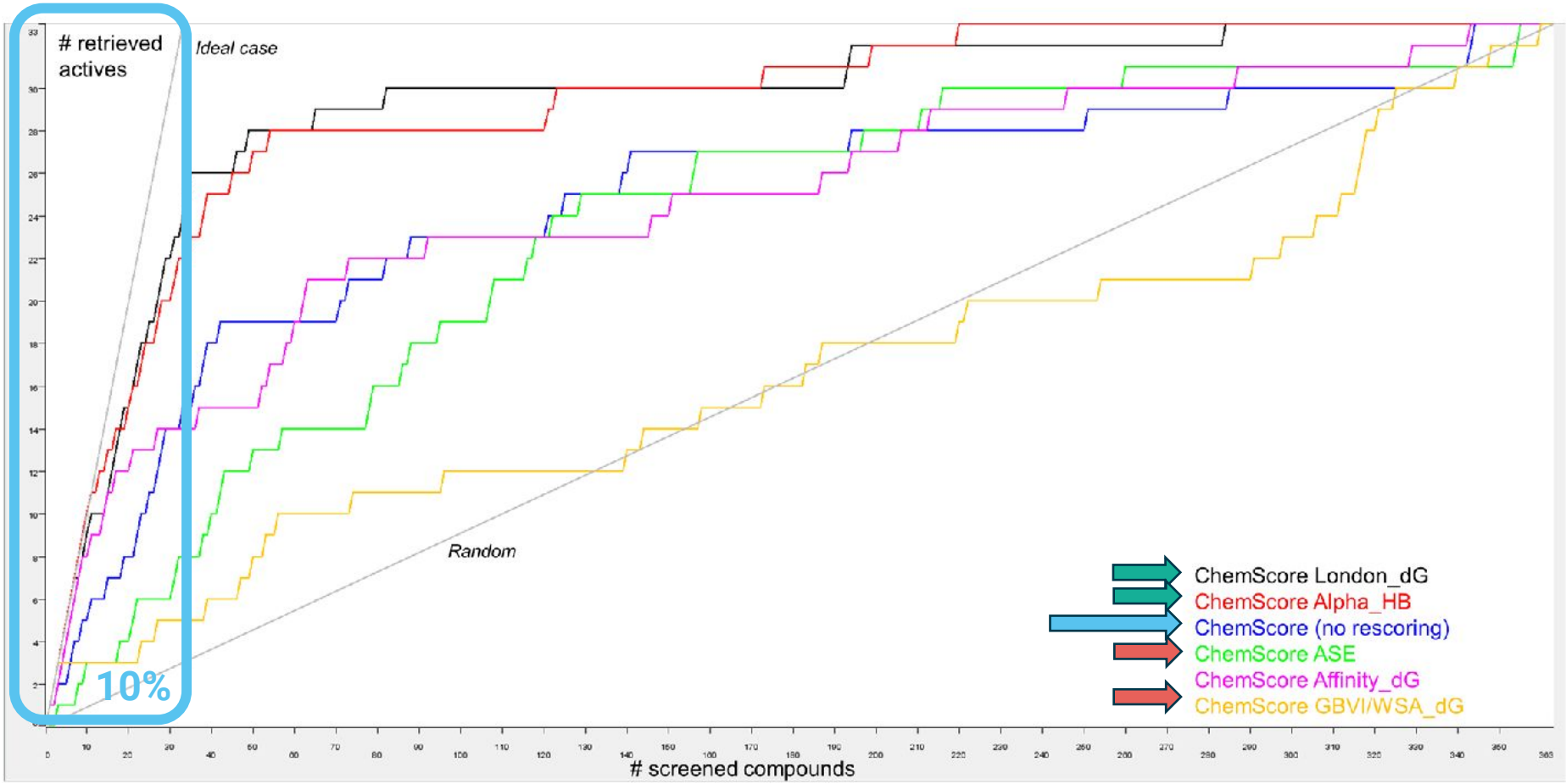






discngine

Enrichment curve 1MVC.A - ChemScore Comparison of rescoring functions



Benefits of calibration

Gain of quality

- Identification of the best parameters for each SBVS
- Clear reporting of SBVS calibration

Gain of time

- Automation through KNIME workflow
- Acceleration of SBVS setup
- Sharing with colleagues

Easy way

- Homogeneity for SBVS setup
- Adapting the workflow to specific docking software

Aknowledgements

Martin Kotev

Obdulia Rabal

Claudia Beato

Constantino Diaz

Danièle de Boyer

discngine

www.discngine.com

contact@discngine.com

discngine

Thanks!

+33 1 80 48 66 70
contact@discngine.com
Discngine, 79 avenue Ledru-Rollin, 75012 Paris – France
www.discngine.com