Supporting Information

3D-e-Chem-VM: Structural Cheminformatics Research Infrastructure in a Freely Available Virtual Machine

Ross McGuire[1,2#], Stefan Verhoeven[3#]*, Márton Vass[#4], Gerrit Vriend[1], Iwan J. P. de Esch[4], Scott J. Lusher[1,3], Rob Leurs[4], Lars Ridder[3], Albert J. Kooistra[1,4], Tina Ritschel[1], Chris de Graaf[4]**

[1]*Centre for Molecular and Biomolecular Informatics (CMBI), Radboud university medical center, 6525 GA Nijmegen, The Netherlands;* [2]*BioAxis Research, Pivot Park, 5342 AC Oss, The Netherlands;* [3]*Netherlands eScience Center, 1098 XG Amsterdam, The Netherlands;* [4]*Division of Medicinal Chemistry, Faculty of Sciences, Amsterdam Institute for Molecules, Medicines and Systems (AIMMS), Vrije Universiteit Amsterdam, 1081 HZ Amsterdam, The Netherlands*

[#] R.McG, S.V. and M.V. contributed equally

Corresponding Author: * R.McG. e-mail: ross.mcguire@bioaxisresearch.com; S.V. e-mail: S.Verhoeven@esciencecenter.nl; C.d.G. e-mail: c.de.graaf@vu.nl

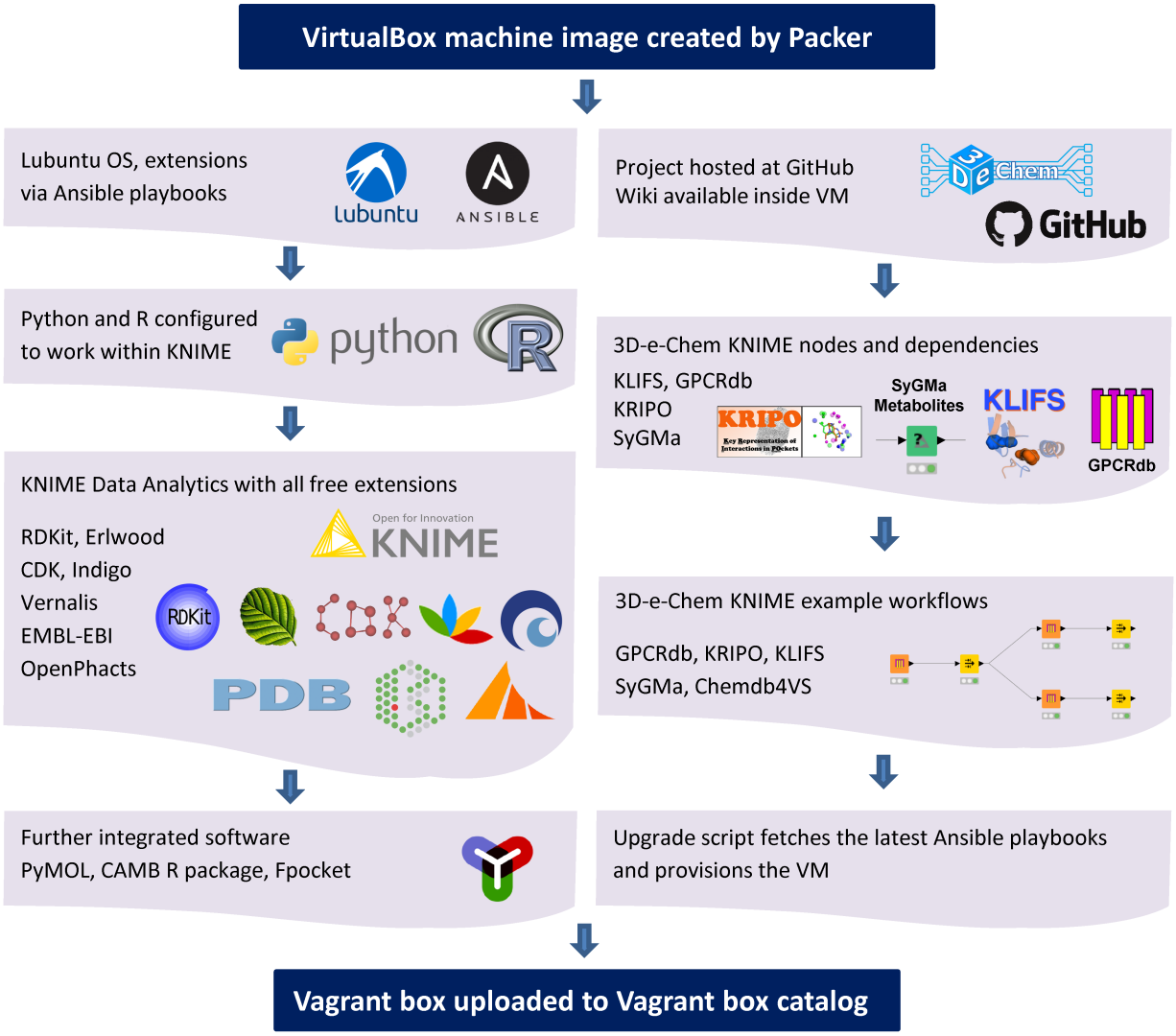


Figure S1 Contents of the 3D-e-Chem-VM virtual machine containing publicly available open source software and tools (left) complemented by novel KNIME nodes and workflows (right).

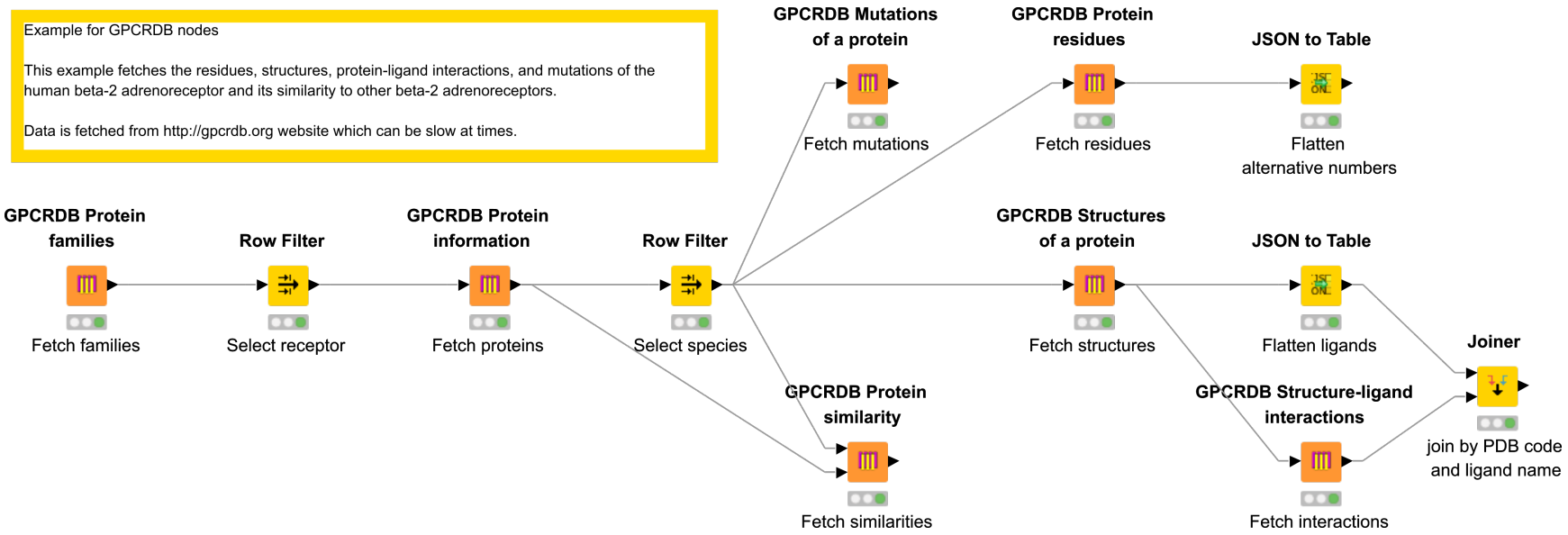


Figure S2 Full version of GPCRdb nodes example workflow

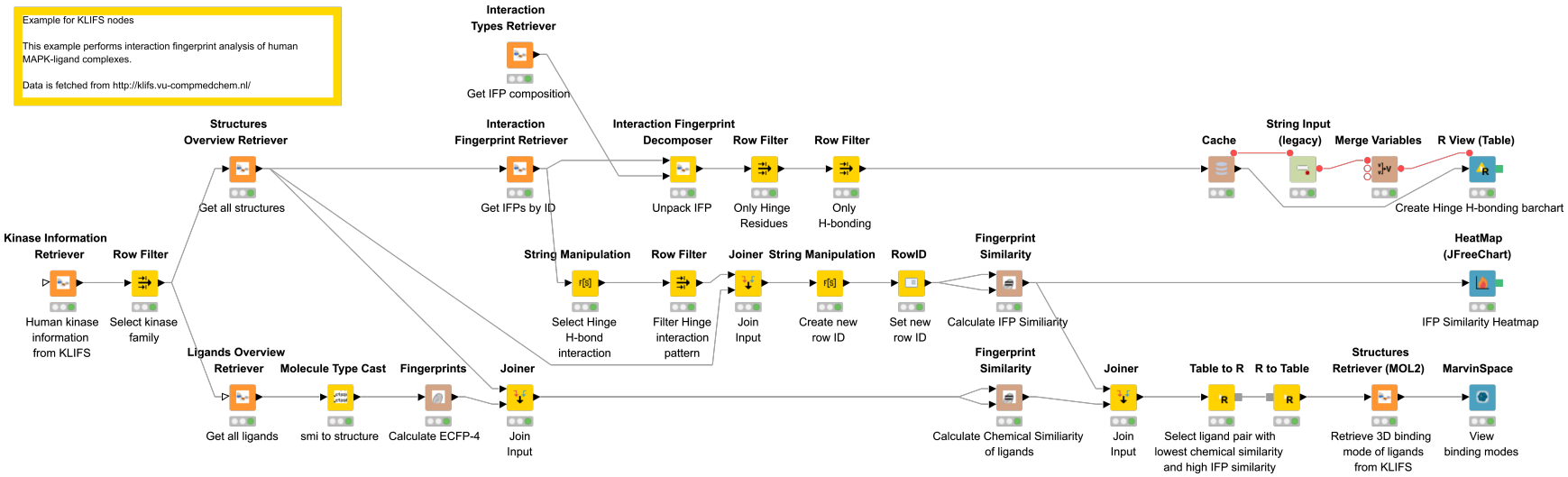


Figure S3 Full version of KLIFS nodes example workflow.

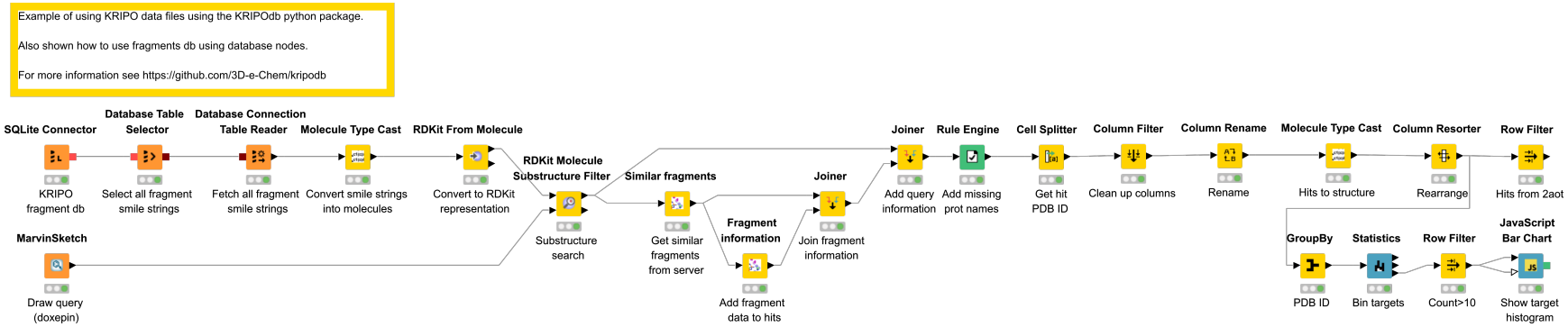


Figure S4 Full version of KRIPO nodes example workflow.

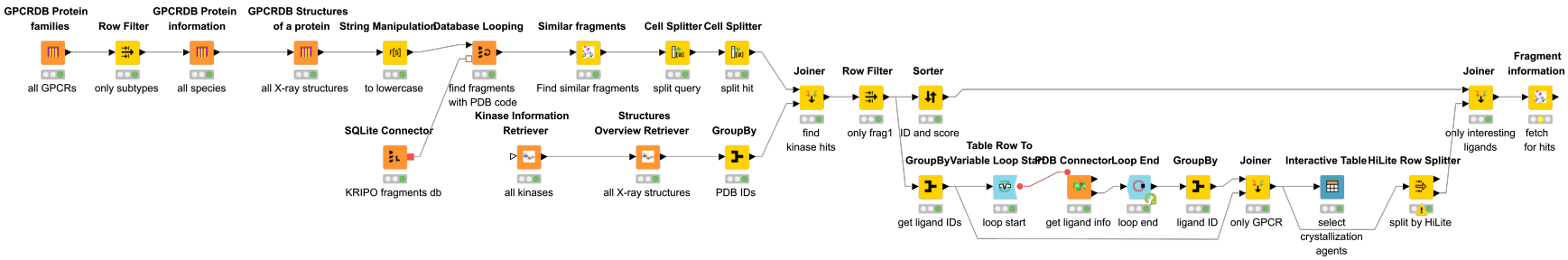
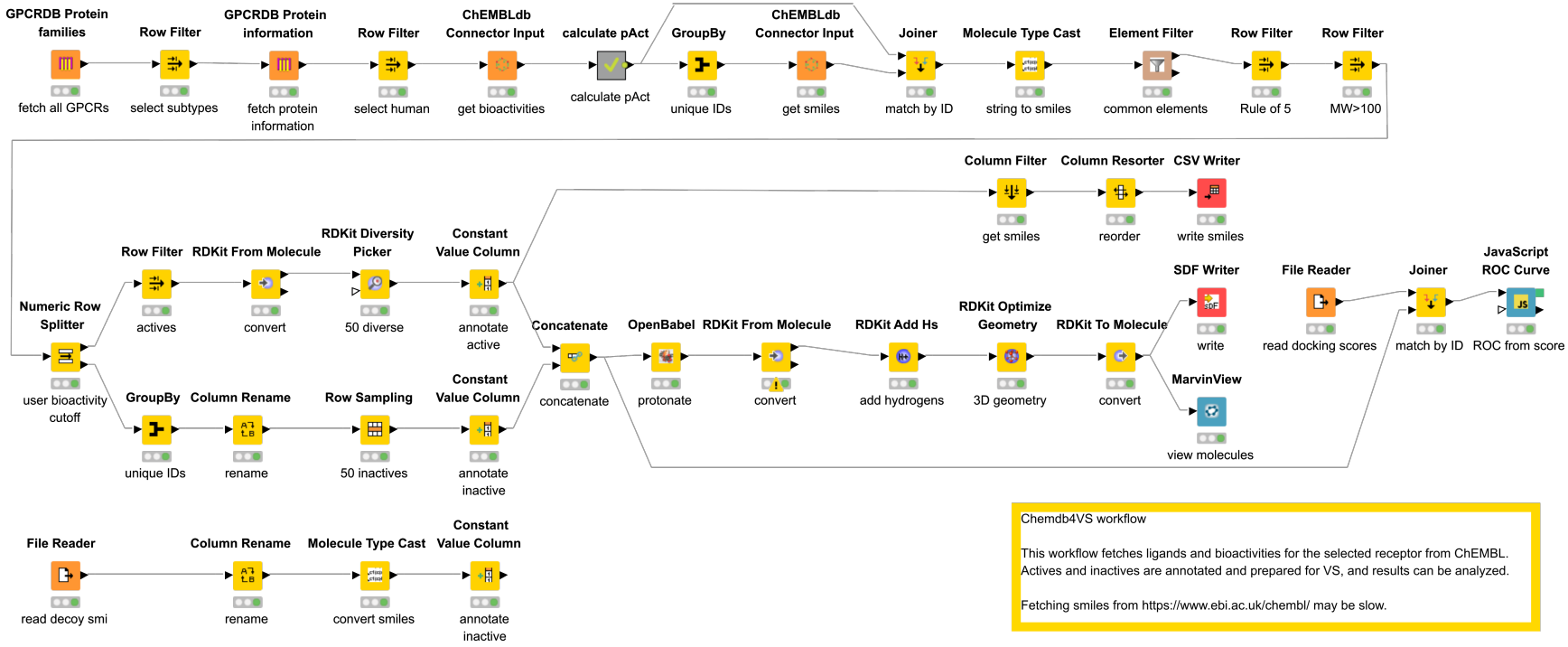


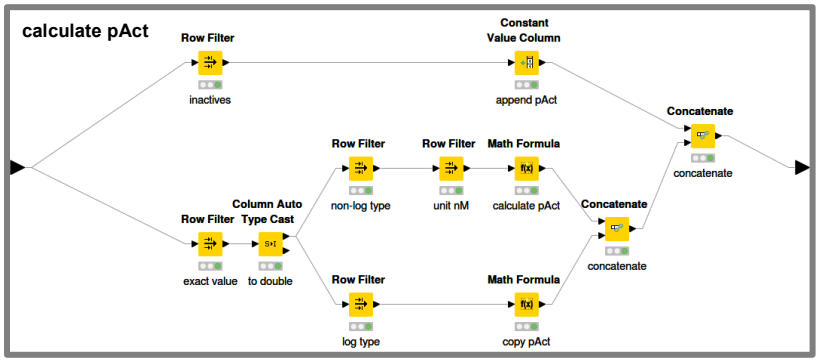
Figure S5 Full version of an integrated 3D-e-Chem workflow for the identification of similar GPCR and kinase binding pockets. First the GPCRdb and KLIFS nodes are used to fetch all experimentally determined structures of ligand-protein complexes in the two drug target families. The KRIPO nodes are subsequently used to assess the structure-based pharmacophore similarity between all GPCR and kinase binding sites, yielding 1428 similar GPCR-kinase pairs (modified Tanimoto coefficient > 0.5, excluding crystallization agents). Although the scope of this work was not to examine all kinase-GPCR associations in depth, this analysis provides information on potential cross-reactivities between kinase and GPCR ligands that can be used for off-target identification, ligand repurposing, or the discovery of ligands with a desired GPCR-kinase polypharmacological profile.[1] The KRIPO pharmacophore analysis identified the similar ergotamine bound serotonin 5-HT_{2B} receptor (PDB: 4ib4) and Sorafenib bound p38α MAP kinase (PDB: 3heg) binding site pair (modified Tc = 0.55), which is consistent with the recent experimental identification of sorafenib as a high affinity 5-HT_{2B} ligand (K_i = 56 nM).[2] The pharmacophore-based similarity search furthermore identified doxepin bound histamine H_1 receptor binding site (PDB: 3rze) and the maternal embryonic leucine zipper kinase (MELK, PDB: 4umt) as targets that share acidic and aromatic/hydrophobic binding site features[3, 4] and can potentially be targeted by dual ligands.



Chemdb4VS workflow

This workflow fetches ligands and bioactivities for the selected receptor from ChEMBL. Actives and inactives are annotated and prepared for VS, and results can be analyzed.

Fetching smiles from https://www.ebi.ac.uk/chembl/ may be slow.



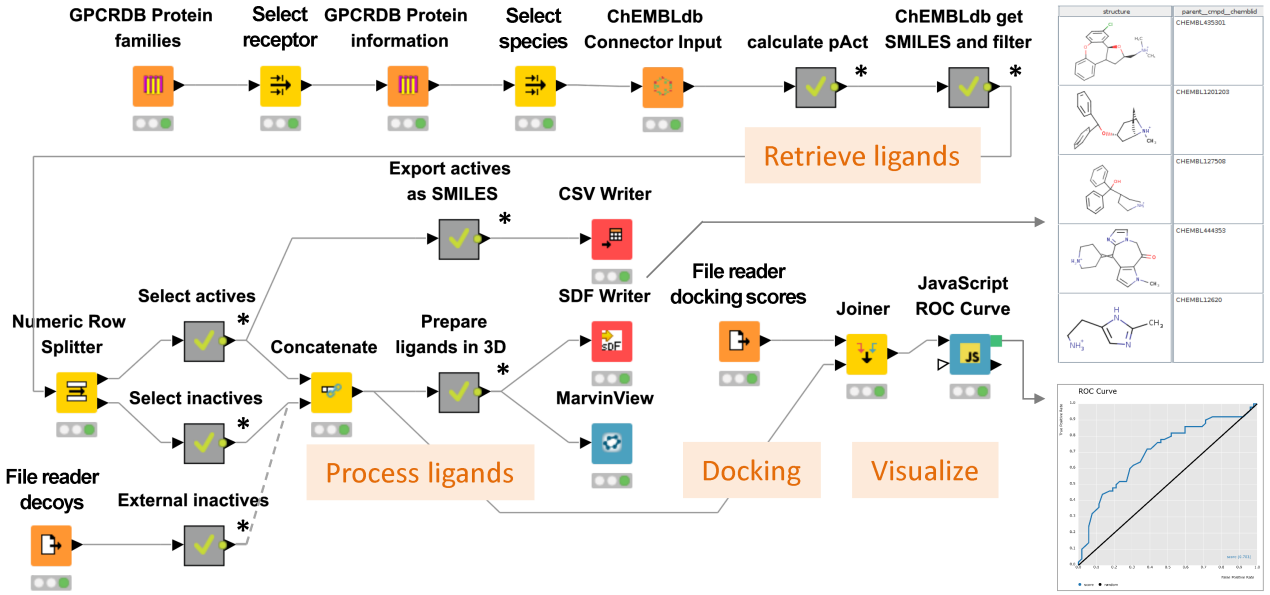


Figure S6 Full and small versions of Chemdb4VS workflow and the "calculate pAct" metanode expanded. This workflow is designed to collect experimentally supported ligand data sets for the retrospective validation of virtual screening methods. In contrast to traditional decoy collection methods where decoys are only assumed to be inactives (as e.g. in the DUD-E[5]) the worflow collects both actives and known inactives from the ChEMBL database in order to use validated datasets for the evaluation of virtual screens. In the example workflow, the novel GPCRdb nodes were utilized in combination with a ChEMBLdb Connector Input node for the download of all bioactivity data available for a specified receptor. Bioactivity data are filtered and processed separately for non-logarithmic (EC_{50}, IC_{50}, K_i, K_D, K_B) and logarithmic (pK_i, log K_i etc.) activity types and for DRUGMATRIX assays, and the pActivity value is calculated for the data points passing the filters. The user can set the active/inactive bioactivity cutoff (in nM) in a Numeric Row Splitter node. SMILES strings for the filtered ChEMBL IDs are downloaded and are filtered for druglike properties and their number is reduced to a user defined limit. Representative actives are selected by the RDKit Diversity Picker node and inactives are sampled randomly. Alternatively, external ligands can be imported and decoy sets can be created by for example using the DUD-E server.[5] Ligands are protonated using the OpenBabel node and their three-dimensional geometries are optimized by the RDKit Optimize Geometry node. The results of an external docking simulation of the generated sets are imported, and based on the active/inactive annotation a ROC curve is plotted using the JavaScript ROC Curve node.

References

1. Elkins, J. M.; Fedele, V.; Szklarz, M.; Abdul Azeez, K. R.; Salah, E.; Mikolajczyk, J.; Romanov, S.; Sepetov, N.; Huang, X. P.; Roth, B. L.; Al Haj Zen, A.; Fourches, D.; Muratov, E.; Tropsha, A.; Morris, J.; Teicher, B. A.; Kunkel, M.; Polley, E.; Lackey, K. E.; Atkinson, F. L.; Overington, J. P.; Bamborough, P.; Muller, S.; Price, D. J.; Willson, T. M.; Drewry, D. H.; Knapp, S.; Zuercher, W. J., Comprehensive characterization of the Published Kinase Inhibitor Set. *Nat Biotechnol* **2016**, 34, 95-103.
2. Lin, X.; Huang, X. P.; Chen, G.; Whaley, R.; Peng, S.; Wang, Y.; Zhang, G.; Wang, S. X.; Wang, S.; Roth, B. L.; Huang, N., Life beyond kinases: structure-based discovery of sorafenib as nanomolar antagonist of 5-HT receptors. *J Med Chem* **2012**, 55, 5749-59.
3. Kooistra, A. J.; Kuhne, S.; de Esch, I. J.; Leurs, R.; de Graaf, C., A structural chemogenomics analysis of aminergic GPCRs: lessons for histamine receptor ligand design. *Br J Pharmacol* **2013**, 170, 101-26.
4. van Linden, O. P.; Kooistra, A. J.; Leurs, R.; de Esch, I. J.; de Graaf, C., KLIFS: a knowledge-based structural database to navigate kinase–ligand interaction space. *J. Med. Chem.* **2013**, 57, 249-277.
5. Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K., Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, 55, 6582–6594.