

# Automation of physico-chemical properties calculation and registration

Prioritizing the most promising drug candidates for further evaluation.

## Getting information to medicinal chemists in a timely manner

The development of drug candidates that combine an acceptable biological activity and an appropriate physico-chemical profile is a key challenge. Therefore, in the drug discovery process, physico-chemical properties are important parameters for the characterization of compounds. In the search for new drug candidates, medicinal chemists routinely evaluate data such as biological activities and physico-chemical properties associated to numerous compounds. This is to prioritize the most promising ones for further optimization or study and discard the others. The present workflow helps chemists evaluate whether the compounds possess desirable physico-chemical properties such as solubility, pKa, and Lipinski criteria.

The goal of this project was to provide essential information to all medicinal chemists in a timely manner. All scientists, about one hundred people – and not only MedChem or computational chemists - all benefit from these calculated properties at various stages of the discovery process.

The specific requirements of this project included:

- Integrate physico-chemical properties (calculated by ACD/Labs Percepta) to the corporate database.
- Implement a user-friendly data visualization format for pKa values to simplify association of calculated property to compound functional group.
- Automate the entire process without any human intervention (e-mail if workflow ends with success).
- Run the workflow once a day, during night.
- Extract newly registered (i.e. without calculated physico-chemical properties) molecules from dedicated ORACLE™ view.
- Interact with ACD/Labs Percepta in order to calculate physico-chemical properties.
- Parse the calculated results and upload to the corporate database.
- Render the annotated structures in PNG image and upload to the corporate database.

## Fully automating the evaluation of drug compounds

To aid chemists in evaluating whether compounds have desirable physico-chemical properties, a KNIME workflow was developed that routinely updates compounds registered in the proprietary database, with the corresponding predicted physico-chemical properties (LogP, LogD, LogS and pKa). A commercial program for property calculation (ACD/Labs Percepta) has been coupled with KNIME Analytics Platform and KNIME Server to fully automate this procedure for all new chemical entities registered in the company database. The KNIME workflow, deployed on KNIME Server, is executed automatically at a given time, and results are stored in the Chiesi proprietary corporate DB.

### Company

Based in Parma, Italy, Chiesi Farmaceutici is an international research-focused healthcare group (Chiesi Group) with 85 years of experience in the pharmaceutical industry and a global presence in 29 countries. Chiesi Group researches, develops, and markets innovative drugs in the respiratory therapeutics, specialist medicine, neonatology and rare disease areas and employs nearly 6,000 people. Its R&D organization is headquartered in Parma (Italy), and is integrated with R&D centers in France, the USA, the UK, and Sweden to advance Chiesi's pre-clinical, clinical, and registration programs. Chiesi Group is a certified Benefit corporation.

### Solution highlights

# 50,000

compounds with calculated properties over a time-frame of more than 4 years - without any problem or intervention.

Pharma

Drug Discovery

Data Analysis

*"The free and open source KNIME Analytics Platform played an important role – largely due to the significant cost advantages that an open source software has, as well as the number of internal KNIME advocates who had already been using KNIME."*

*Andrea Ciacci, Application Scientist, S-IN Soluzioni Informatiche Srl (KNIME Partner).*

The project started with verifying that ACD/Labs Percepta (batch module) could calculate all the needed properties via command-line and that the results were compatible with standard KNIME nodes – specifically SDF Reader and CSV Reader. A set of molecule structures was then received from a public structure database to use in setting up a properties calculation different enough to cover most calculation problems. Then the structure's format of the input table coming from the ORACLE™ view was defined (i.e. SDF or SMILES format of molecules, identifier, primary keys, other fields), as well as the output format to write to ORACLE™ tables (table names, fields name and type, accessory columns).

The construction of the workflow looked like this:

- Build the ORACLE™ table similar to production environment, for the read and write/update.
- Export SMILES structures, transform and write an SDF file, and instruct external applications to read the last one using the External Tool node.
- Gather the output files (CSV and SDF).
- Split data between a single-type property (LogP, LogD and LogS at pH 7.4) to be registered in the molecule table, multiple-type property (pKa values) to be registered in its proper table and warning message to be archived as logfile.txt – all using the CSV Reader Output.
- Annotate the pKa values in the SDF structure, render as PNG image and load as binary object in a third table – all using SDF Reader Output.
- Send email to administrator (assuming all is fine) automatically using KNIME Server.

## Project Results

This project has resulted in approximately 50,000 compounds with calculated properties over a time-frame of more than 4 years - without any problem or intervention. The biggest impact that this solution has had, is the time saved by scientists who no longer need to calculate properties on demand – as a result, customer satisfaction has also increased considerably. The biggest lesson learned: solving a real-life business case using integration and automation increase productivity and user experience.

## Why KNIME?

Before the project began, two key features were required. The first being the ability to interact (read, write and update permissions) with ORACLE™ Database. The second: the ability to interact via command line with third party software (ACD/Labs Percepta Batch). KNIME Analytics Platform enabled us to do these two things.

To start with, the free and open source KNIME Analytics Platform played

an important role – largely due to the significant cost advantages that an open source software has, as well as the number of internal KNIME advocates who had already been using KNIME. Once acceptance for KNIME Analytics Platform grew, getting a license for KNIME Server was much simpler. Furthermore, adoption of KNIME Server was driven by the possibility to solve other use cases across different departments.