ChemAxon

ChemAxon Usermeeting 2019

22-23. May 2019 Budapest

Transforming AnalytiCon‘s Database Infrastructure

Lars Ole Haustedt, Director Projects & Innovation

AnalytiCon discovery

The Natural Product Company

Agenda

1 Introduction AnalytiCon Discovery GmbH

2 Database Infrastructure

3 Plexus Connect

4 Library Enumeration

5 Computational Chemistry

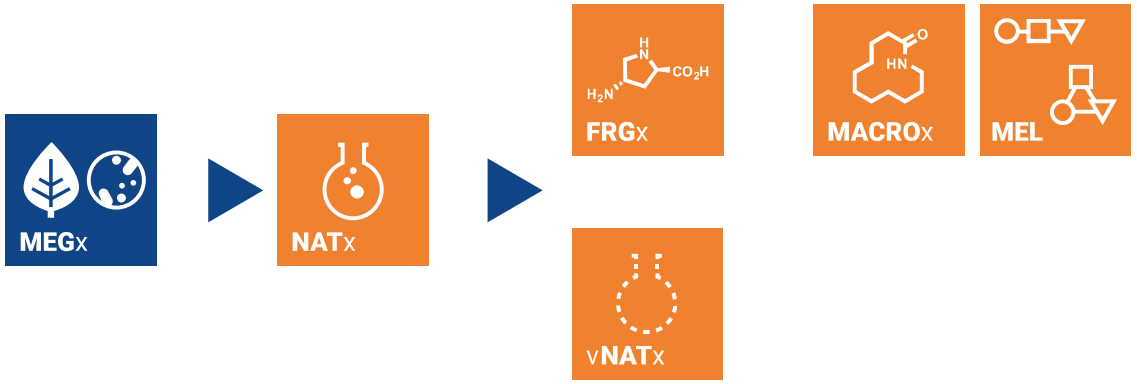
AnalytiCon Discovery GmbH

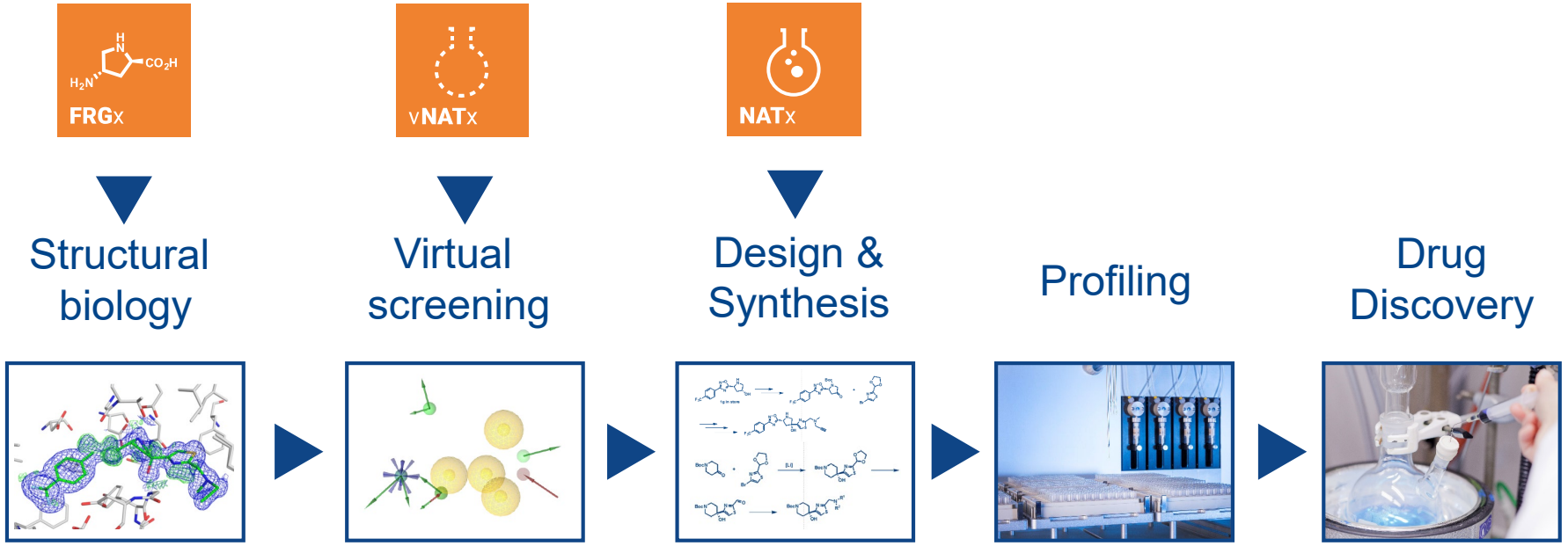
- Headquarters: **Potsdam, Germany**
- Subsidiary: **Rockville** (MD), **USA**
- Privat company founded **2000**
- **65 employees** (17 PhD)

- Microbiology
- Discovery Biology
- Library Design & Generation
- Separation & Analytical Chemistry
- Discovery Chemistry & Medicinal Chemistry
- Structural Biology & Computational Chemistry

AnalytiCon‘s Chemical EcoSystem

AnalytiCon‘s Chemical Space





AnalytiCon discovery

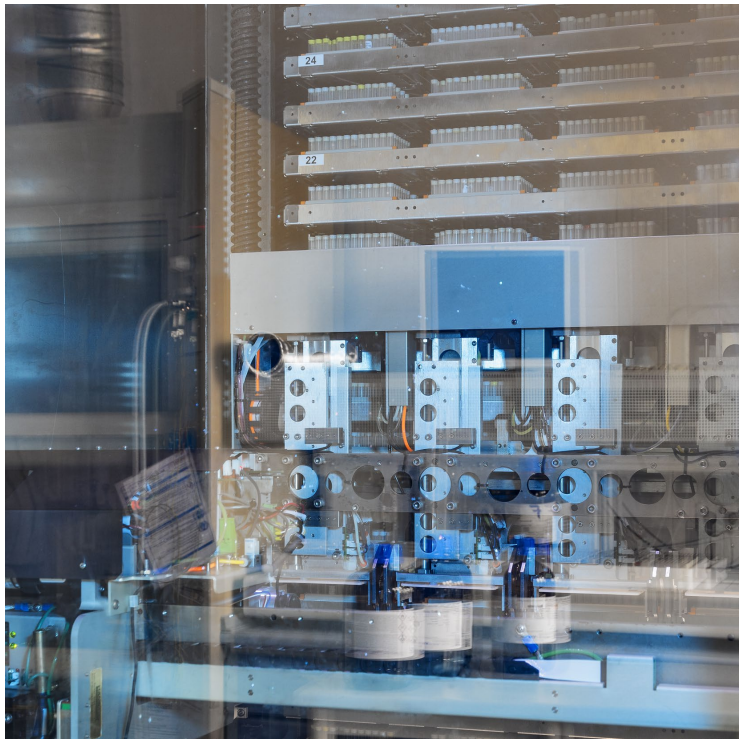
Transformation of Data Managementand Storage

Company Legacy

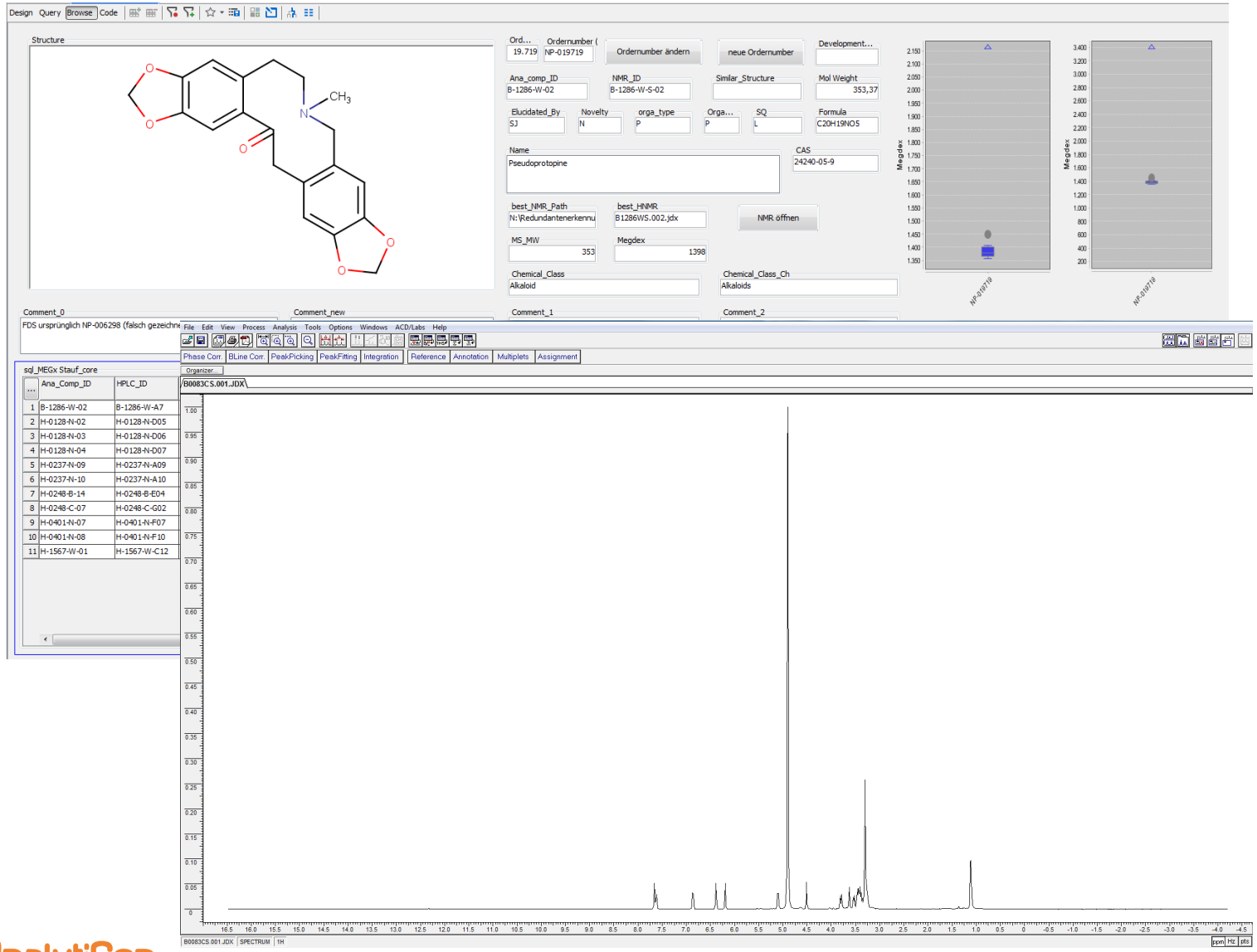
- Complex Interactions of Structural data, Analytical data and Biological data
- Large LCMS and NMR data sets (> 5TB)
- Medium sized structural datasets (< 10GB, 200,000 structure records)
- ChemFinder, Access and Excel based (ISIS Base)
- Tedious maintenance

Perspective

- Relational data base
- MySQLand IJChem based
- Browser based viewing and editing
- Registration, inventory, relation to QC (analytical) data
- Integration of medicinal chemistry & computational chemistry



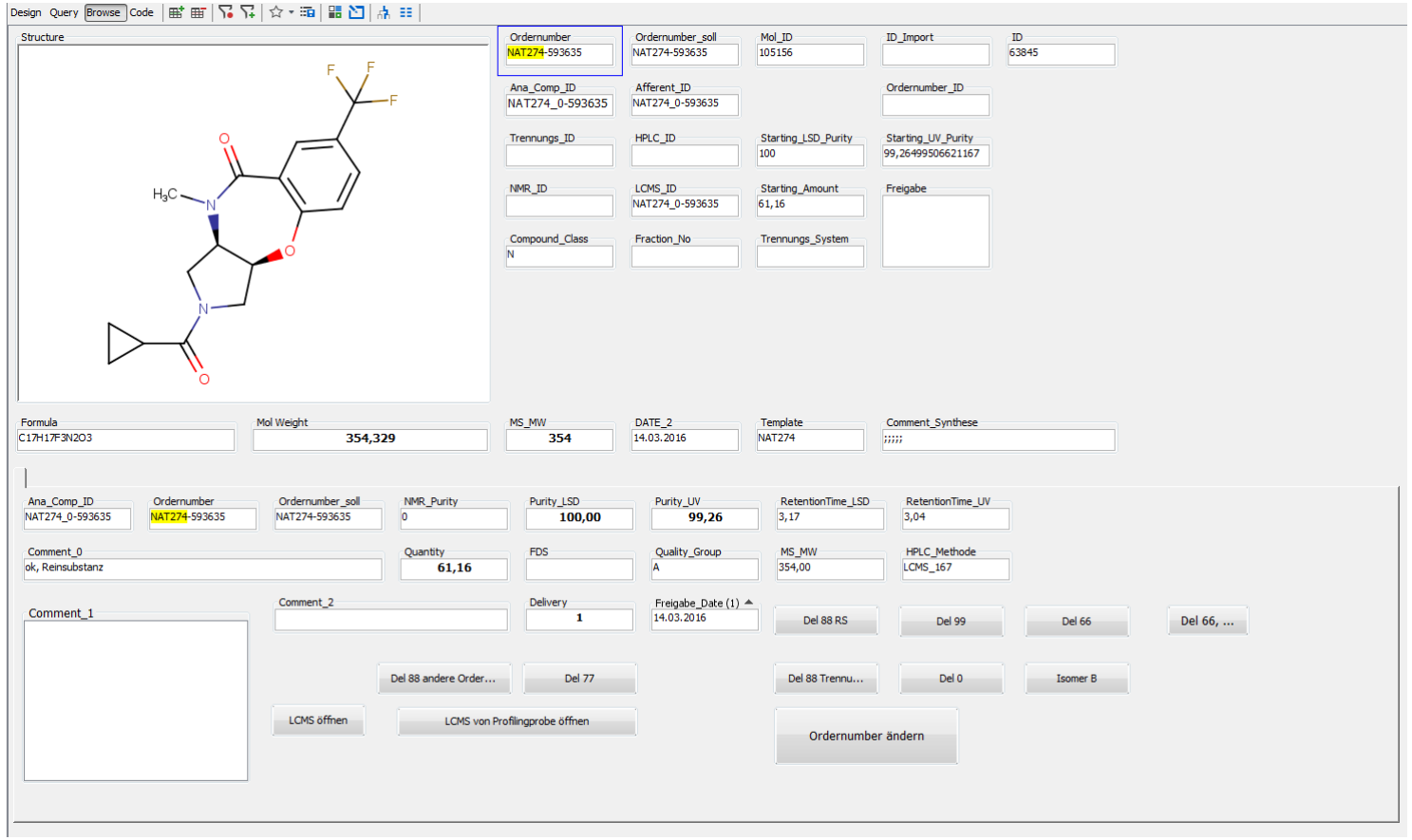
IJChem MySQL



Natural Product Database

- Storage of Chemical Structure
- Relational table with all batches
- Direct link to NMR data (JCAMP-DX,)
- Direct link to LCMS pdf
- Polarity Indicator

IJChem MySQL



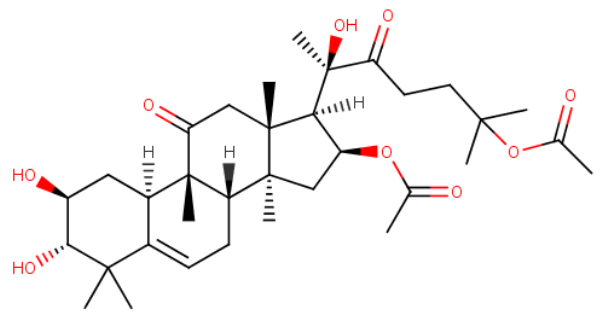
Synthetic Cpd Database

- Storage of Chemical Structure
- Direct link to LCMS pdf
- Synthetic Synthon
- Compound Management / Storage

Plexus Connect

MEGx_Stauf _Plexus

Structure



Ordernumber	Development_ID	Mol Weight	Formula
NP-019961		604,8	C34H52O9

Ana_comp_ID	NMR_ID	Similar_Struct...	Chemical_Clas...
C-1878-E-15	C-1878-E-S-17		Terpenoids

CAS	Novelty	orga_t...	Orga_type_V
586960-44-3	N	P	P

Name

Hemslecin D

best_NMR_Path	best_HNMR	Megd...	MS_MW
N:\Jcamp\	C1878ES.017.dx	2330	604

Comment_0	Comment_new	Comment_1	Comment_2

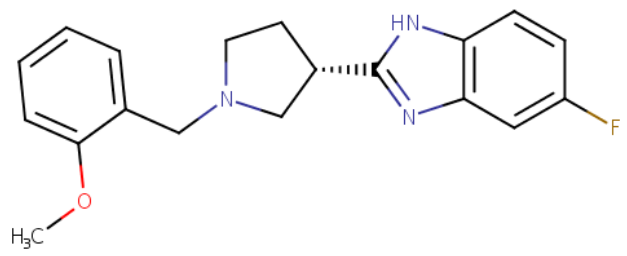
sql_MEGx Stauf_core

Ana_Co...	HPLC_ID	NMR_ID	Fraction_No	MS...	Meg...	NMR_P...	Purity	Quant...	Formiat	Orga_...	Ana Orga ID	Analyticon ID	Genus	Species	Comment Material	Comment Profiling
C-1878-E-14	C-1878-E-E12	C-1878-E-S-16	69+70	604	2330	2	99,8	146,5		P	15687-1	ACD-V-21978-W-0	Hemsleya	carnosiflora		
C-1878-E-15	C-1878-E-F01	C-1878-E-S-17	71+72+73	604	2332	1	99,3	212,7		P	15687-1	ACD-V-21978-W-0	Hemsleya	carnosiflora		
C-1878-E-16	C-1878-E-F02	C-1878-E-S-18	74+75+76	604	2331	2	94,7	73,3		P	15687-1	ACD-V-21978-W-0	Hemsleya	carnosiflora		
C-1878-D-26	C-1878-D-D08	C-1878-D-S-29	94+95+96+97	604	2329	3	90,3	9,1		P	15687-1	ACD-V-21978-W-0	Hemsleya	carnosiflora		
C-1878-D-26		C-1878-D-26_L	94+95+96+97	604	2329	3	90,3	8,1	1	P	15687-1	ACD-V-21978-W-0	Hemsleya	carnosiflora		
C-1878-E-14		C-1878-E-14_L	69+70	604	2330	2	99,8	131,7	1	P	15687-1	ACD-V-21978-W-0	Hemsleya	carnosiflora		
C-1878-E-15		C-1878-E-15_L	71+72+73	604	2332	1	99,3	198,7	1	P	15687-1	ACD-V-21978-W-0	Hemsleya	carnosiflora		
C-1878-E-16		C-1878-E-16_L	74+75+76	604	2331	2	94,7	67,8	1	P	15687-1	ACD-V-21978-W-0	Hemsleya	carnosiflora		
C-2049-H-11	C-2049-H-A11	C-2049-H-S-12	57+58+59+60	604	2345	1	97,8	6,0	0	P	15685-9	ACD-V-21976-W-0	Phlomis	betonicoides		
C-2049-H-12	C-2049-H-A12	C-2049-H-S-13	62+63+64+65	604	2349	4	50,4	4,7	0	P	15685-9	ACD-V-21976-W-0	Phlomis	betonicoides		

Plexus Connect

NATx_Production_Plexxus

Structure



Ordernumber	Ordernumber_...	Mol_ID	ID_Import	ID
NAT31-457216	NAT31-457216	30846		45014

Ana_Comp_ID	Afferent_ID	Ordernumber_ID
NAT31 1-457216	NAT31 1-457216	

Trennungs_ID	HPLC_ID	Starting_LSD_...	Starting_UV_P...
	NAT31 1-457216	98,482	98,454

NMR_ID	LCMS_ID	Starting_Amount	Freigabe
		106,50	

Compound_Class	Fraction_No	Trennungs_Sys...
B		

Formula	Mol Weight	MS_MW	DATE_2	Template	Comment_Synthese
C19H20FN3O	325,387	325	22.09.2009	123	;;;;

Ana_Comp_ID	Ordernumber	Ordernumber_...	NMR_Purity	Purity_LSD	Purity_UV	RetentionTime...	RetentionTime...
NAT31_1-457216	NAT31-457216	NAT31-457216	0	98,48	98,45	7,22	7,01

Comment_0	Quantity	FDS	Quality_Group	MS_MW	HPLC_Methode
ok, Reinsubstanz	106,50		A	325,00	B31

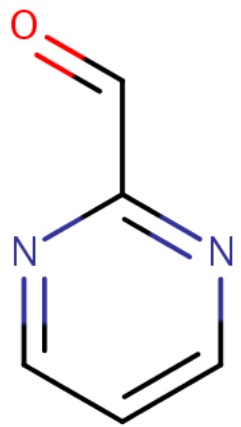
Comment_1	Comment_2	Delivery	Freigabe_Date
		1	22.09.2009

Unbound	Unbound

AnalytiCon discovery

Plexus Connect

Chemikalien-DB_Plexus

Structure	Name
	Pyrimidine-2-carboxaldehyde

CAS	Größe	Einheit	Reinheit
27427-92-5	1	g	

Anbieter	Best.-Nr.	Preis	Währung
Manchester Organics	D45360	184,00	Euro

Raum	Ablage	Entleiher	Kuehlung
EG8	KS1		

Bestellt	Datum erstellt	Erhalten	Geöffnet am
Am	2019-03-26	Ras	

CdId	Formel	Mol Weight	Dichte
7600	C5H4N2O	108,100	

Kommentar

Brennbar	Lagerklasse	H-Sätze
		H315-H317-H319-H335

P-Sätze
P233-P260-P261-P264-P271-P272-P280-P302+P352-P304+P340-P312-P321-P332+P313

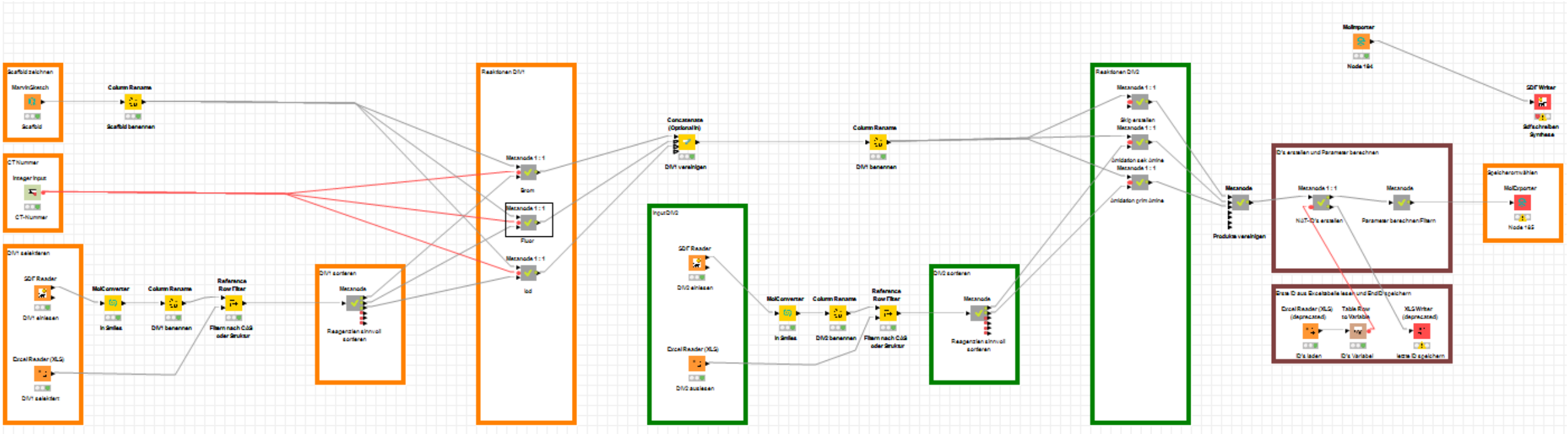
ReagentDatabase

- Storage of chemical structure
- Provider / price Info
- Inventory information
- Browser access for lab staff
- User friendly selection of reagent sets

Feature Requests Plexus Connect

- Implementation of scripts and buttons in form view
- Pull down menus in form view
- Import functionality (SD- file)
- Adding records in form view
- Definition of mandatory fields in form view

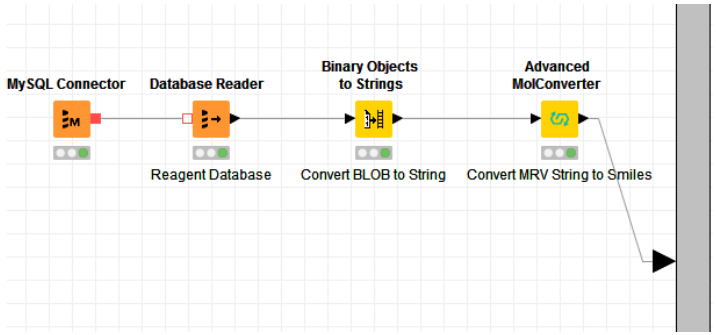
ChemAxon Knime Nodes: Library Enumeration



Knime Workflow

- Library Enumeration
- Generation of unique compound identifier
- Tractable synthon information
- Property calculation (TPSA, clogP)
- Generation of conformers (Inteligand)
- Export: MySQL, SDF, Table of SMILES

Library Enumeration



File Hilite Navigation View

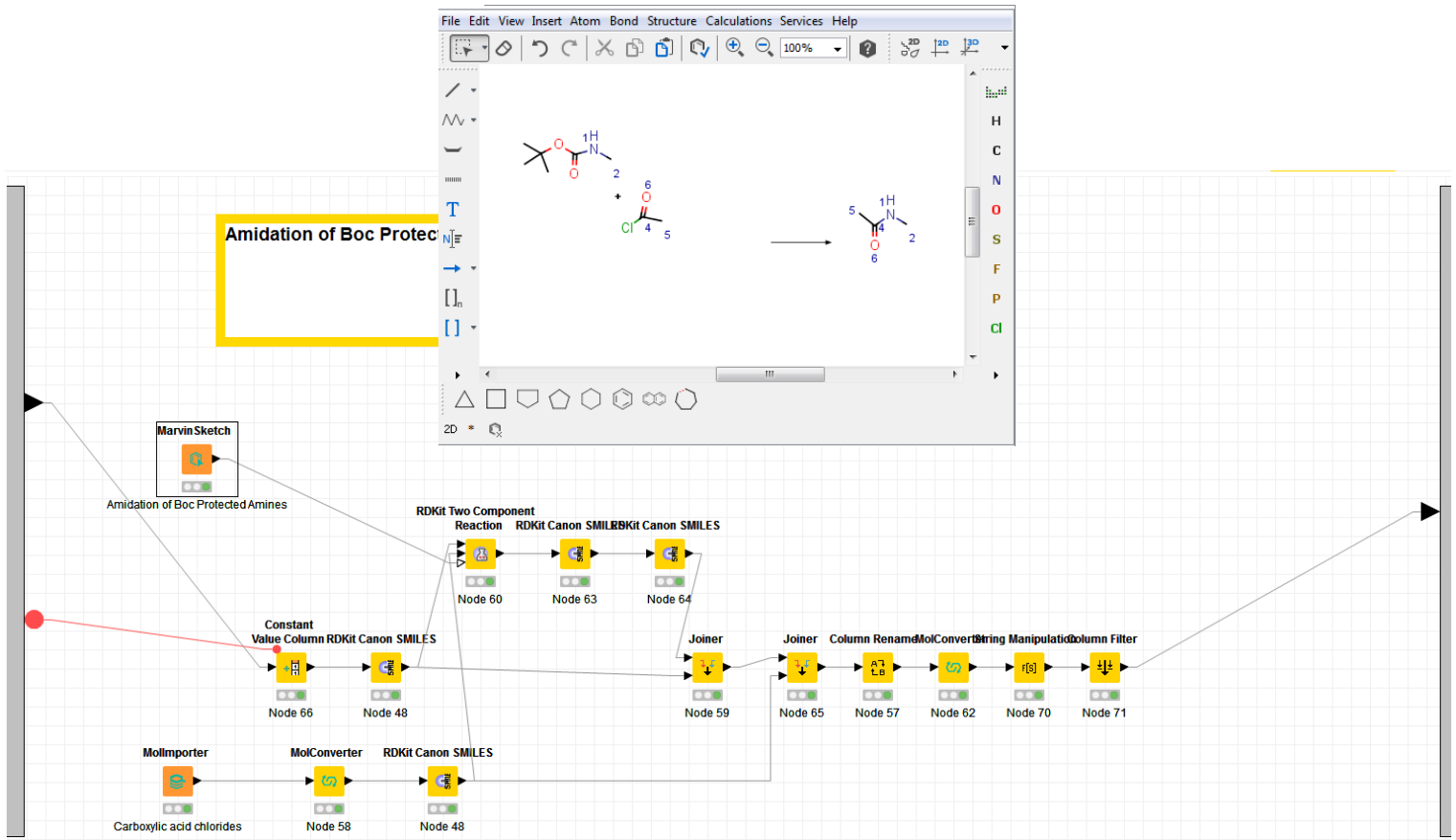
Table "default" - Rows: 7229 | Spec - Columns: 2 | Properties | Flow Variables

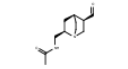
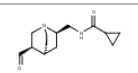
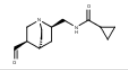
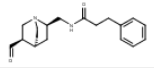
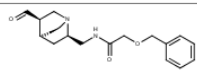
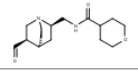
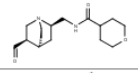
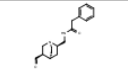
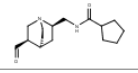
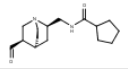
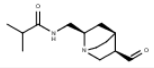
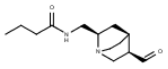
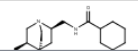
Row ID	cd_structure	CAS
Row0		25016-11-9
Row1		3034-50-2
Row2		123-08-0
Row3		1074-86-8
Row4		10200-59-6
Row5		1196-69-6
Row6		35344-95-7
Row7		65873-72-5
Row8		10070-92-5
Row9		27258-33-9
Row10		4649-09-6
Row11		487-89-8
Row12		112758-40-4
Row13		123-11-5

ReagentDatabase

- Direct import from MySQL reagent database
- Import of Sub Selections
- No redundant storage of information

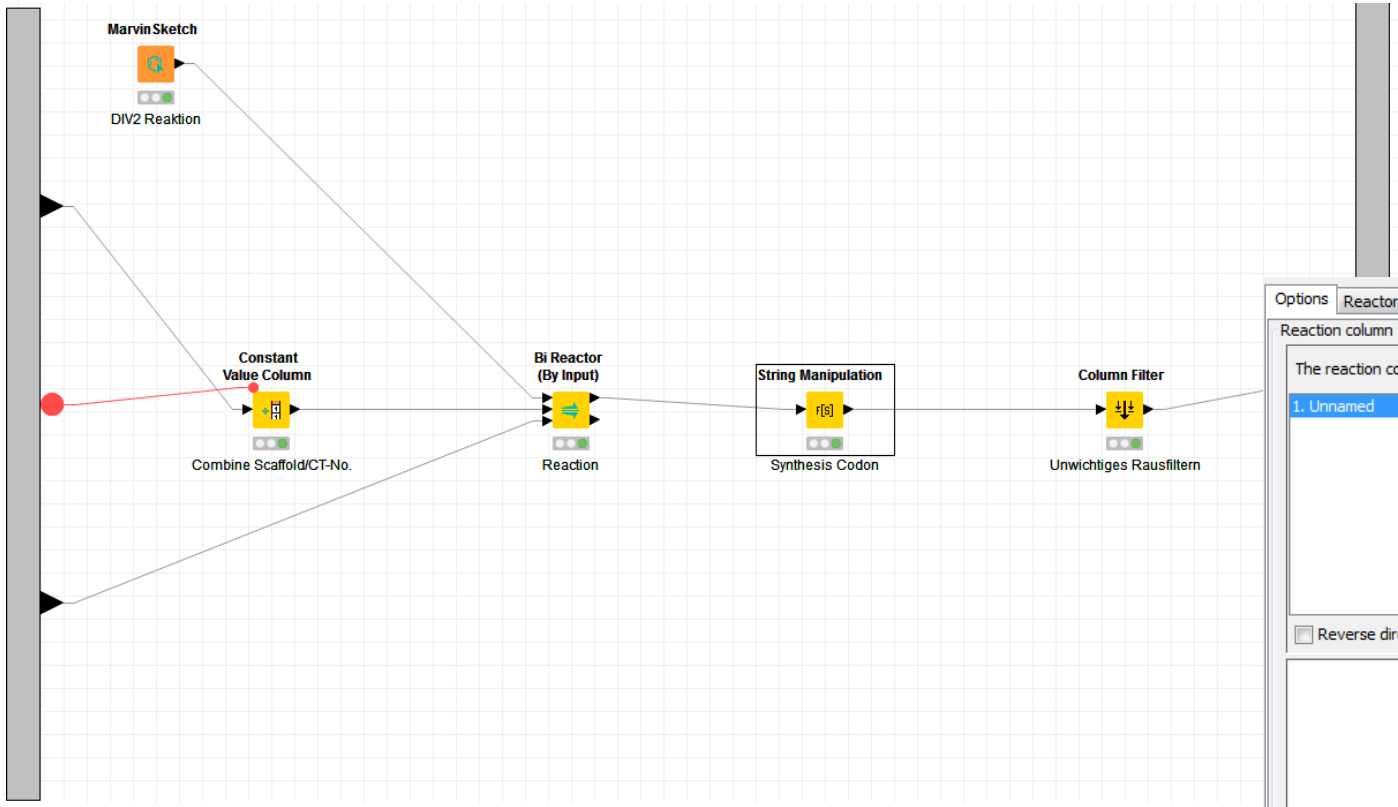
Workflow without JChem Knime Nodes

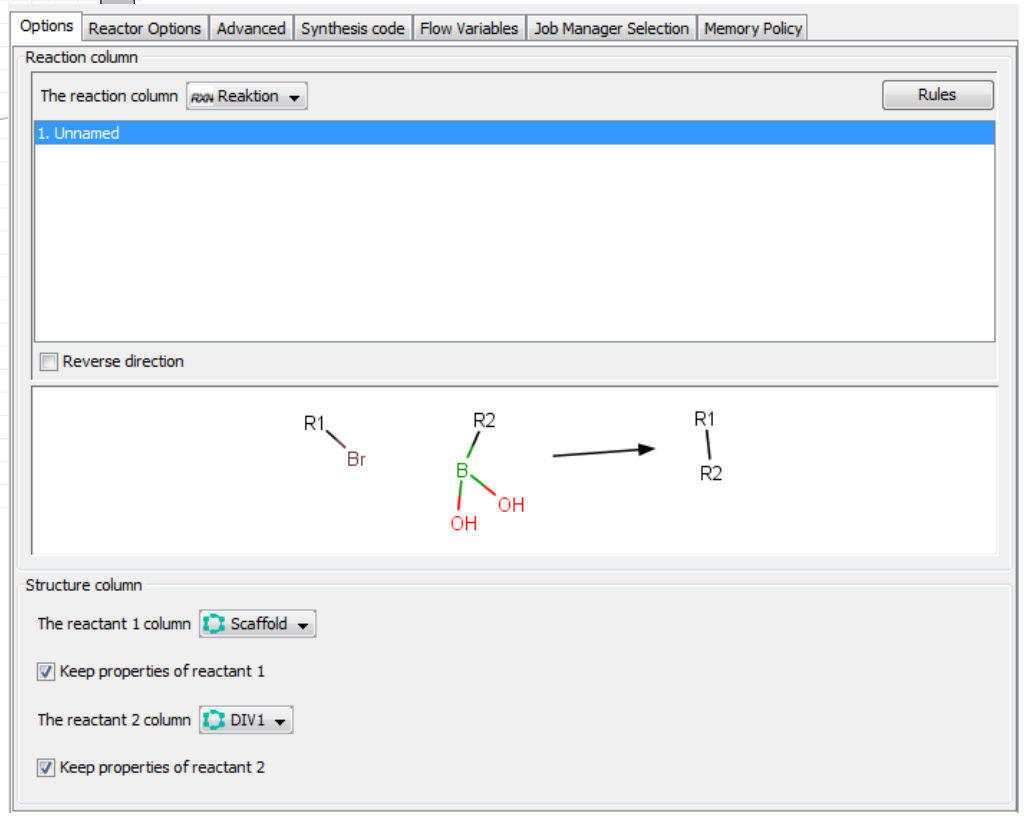


Row ID	Alkyne	Library_Intermediates-1	DIV1-ID
0_0_0_0_Ro...			31--CAS-75-36-5
0_1_0_0_Ro...			31--CAS-4023-34-1
0_1_0_0_Ro...			31--CAS-4023-34-1
0_2_0_0_Ro...			31--CAS-645-45-4
0_3_0_0_Ro...			31--CAS-19810-31-2
0_4_0_0_Ro...			31--CAS-40191-32-0
0_4_0_0_Ro...			31--CAS-40191-32-0
0_5_0_0_Ro...			31--CAS-103-80-0
0_6_0_0_Ro...			31--CAS-4524-93-0
0_6_0_0_Ro...			31--CAS-4524-93-0
0_7_0_0_Ro...			31--CAS-79-30-1
0_8_0_0_Ro...			31--CAS-141-75-3
0_9_0_0_Ro...			31--CAS-2719-27-9

- Traction of synthon information more tedious
- Careful inspection (filtering) of results necessary

Workflow Making Use of JChem Knime Nodes

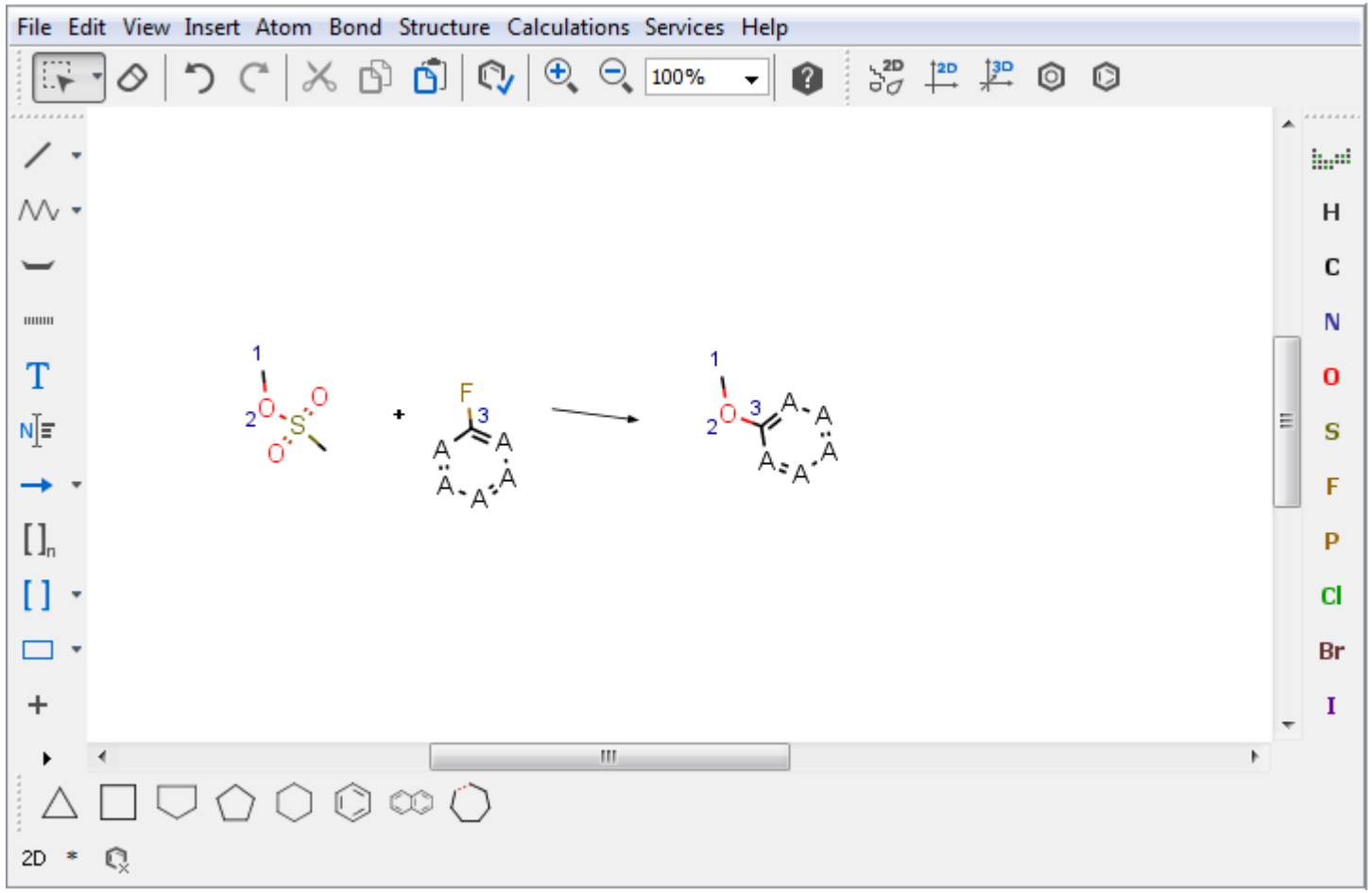




- User friendly traction of reagent properties
- Better control on reaction products

AnalytiCon discovery

Definition of Reactions for Enumeration



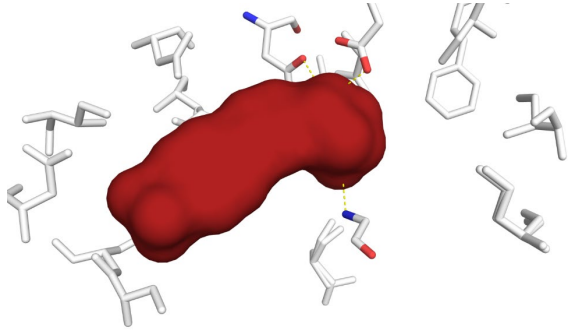
Definition of Reactions with Marvin

- Very flexible
- Chemical support by Marvin (SMARTS)
- Good atom mapping
- Intuitive no need for non chemical hacks

Computational Chemistry

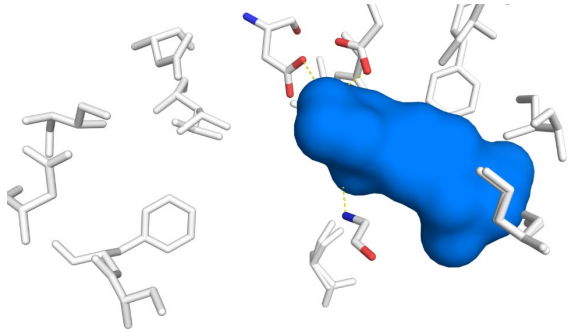
Virtual Libraries

- Design of virtual libraries based on pharmacophores or established scaffolds
- Virtual Library Enumeration
- Calculation of properties; calculation of conformations , calculation of fingerprints



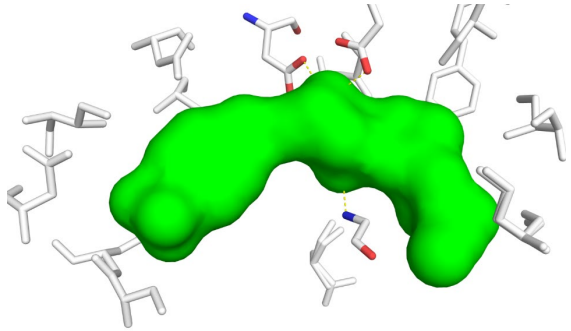
Virtual Screening

- Pharmacophore Search(LigandScout)
- Docking (SeeSAR, LeadIT)



Compound Selection

- Results from virtual screening
- Property based selection (similarity or diversity)
- Fingerprint based selection (similarity or diversity)



Summary

Achievements

- Sustainable, stable high performance relational structural database (MySQL,IJChem)
- Less complex more reliable infrastructure
- Company wide access via Plexus Connect
- Library enumeration integrated (reagent database, synthesis selection, JChem Knime nodes)
- Better integration with compound storage system
- Better integration into Microsoft Office (JChem for Office, Marvin)

Requests

- More functionality and customizable dynamic content in Plexus Connect
- Even more intuitive format menue (color, superscript, subscript) in Marvin

Thank you.

AnalytiCon Discovery GmbH

Oliver Kunz
Karsten Siems
Martina Jaensch
Karsten Vandieken

ChemAxon

Luca Buzasi
Anna Tomin

Lars Ole Haustedt
AnalytiCon Discovery GmbH
Hermannswerder Haus
14473 Potsdam | Germany

Email: l.haustedt@ac-discovery.com
www.ac-discovery.com

AnalytiCon discovery

The Natural Product Company