FOCUS - Development of a Global Communication and Modeling Platform for Applied and Computational Medicinal Chemists

Nikolaus Stiefl,[*,†] *Peter Gedeck,*[‡] *Donovan Chin,*[§] *Peter Hunt,*[∥] *Mika Lindvall,*[⊥] *Katrin Spiegel,*[∥] *Clayton Springer,*[§] *Scott Biller,*[§] *Christoph Buenemann,*[#] *Takanori Kanazawa,*[▽] *Mitsunori Kato,*[§,▽] *Richard Lewis,*[†] *Eric Martin,*[⊥] *Valery Polyakov,*[⊥] *Ruben Tommasi,*[§] *John van Drie,*[§] *Brian Vash,*[§] *Lewis Whitehead,*[§] *Yongjin Xu,*[⊥] *Ruben Abagyan,*[○] *Eugene Raush,*[○] *and Max Totrov*[○]

[†] Novartis Institute for Biomedical Research, Postfach, CH-4002 Basel, Switzerland

[‡] Novartis Institute for Tropical Diseases Pte. Ltd., 10 Biopolis Road, #05-01 Chromos, Singapore 138670

[§] Novartis Institute for Biomedical Research, 250 Massachusetts Ave., Cambridge, Massachusetts 02139, United States

[∥] Novartis Institute for Biomedical Research, Wimblehurst Road, Horsham, West Sussex RH12 5AB, United Kingdom

[⊥] Novartis Institute for Biomedical Research, 4560 Horton Street, Bldg. 4, Emeryville, California 94608, United States

[#] Novartis Institute for Biomedical Research, 11/2 Muthgasse, 1190 Wien, Austria

[▽] Novartis Institutes for BioMedical Research, Ohkubo 8, Tsukuba, Ibaraki 300-2611, Japan

[○] Molsoft L.L.C., 11199 Sorrento Valley Road, S209, San Diego, California 92121, United States

Corresponding author email: nikolaus.stiefl@novartis.com

S1. FOCUS modules

For the end user FOCUS is organized around interconnected modules with specific tasks to apply in a workflow style. Overall, six modules are enabled by default (Tables S1-S6). In addition to that, there are three modules provided for specific research groups (Table S7) as well as one module for recently introduced functionality or methods under development (Table S8). The tables provide details on the different modules, their functions and information on the execution process (client vs. web service).

Table S1. FOCUS data module

Function	Description	Implementation
Draw molecules	Draw molecule or copy & paste from ISIS/Draw or ChemDraw.	Client
Retrieve Structures from Text	Retrieves structures for identifiers in any sort of text (*eg.* emails).	Web service
Similarity Search	Search vendor or internally available compound databases by similarity using different fingerprints.	Web service
Substructure Search	Search vendor or internally available compound databases by substructure.	Web service
Import from Dart	Import data directly from data hub.	Web service
Copy/Paste from Excel or Avalon	Copy spreadsheet data from Excel or data hub interface into a new table.	Client
Open Structure File or Data Table	Import SD-files, mol/mol2 files, as well as text files.	Client
Extract data from other FOCUS sessions	Direct import from other FOCUS sessions - for example surfaces, molecules or tables.	Client
Add Structures to Table using ID Column	Retrieves structures using IDs in table columns.	Web service
Retrieve Name for structures	Given a set of structures in a table, retrieve internal structure names.	Web service
Check compound availability	Check availability of compounds in in any of the Novartis archives (excluding building block and intermediate stores)	Web service
Ames Mutagenicity Aromatic Amine Building Blocks	Download annotated in-house and public Ames mutagenicity data for aromatic amine building blocks.	Web service

Export table	Export data as comma separated (csv) or SD-file.	Client
Export to Spotfire	Starts spotfire and exports table data directly.	Client

Table S2. FOCUS Design module

Function	Description	Implementation
Load processed reagents	Load internal or external reagents (preferred vendor, compound/intermediate archive, ACD).	Web service
Upload Reagents from File	Upload your own reagents from Excel or SD-file.	Client
Convert Reagents to R-groups	Convert reagents to R-groups suitable for Markush-type library enumeration.	Client
Submit to preferred vendor	Submit all compounds with valid catalogue number in current table to preferred vendor ordering system.	Web service
Draw a Reaction	Use the molecular editor to add a reaction to an existing or a new table.	Client
Enumerate Selected Reaction	Select a reaction in a table and enumerate.	Client
Draw a Scaffold	Use the molecular editor to draw a scaffold. Use R1, R2, ... to define the attachment points.	Client
Enumerate Scaffold	Select a scaffold in a table and enumerate.	Client
Load CraigPlot Substituents	Load R-group table of substituents with experimental Hammett (electronic property) values and calculated Hansch values (lipophilic property).	Web service

Table S3. FOCUS Insilico module

Function	Description	Implementation
In-silico properties and filters	Calculate common ADME like properties (e.g. cLogP, PSA, hERG risk, solubility).	Client and web service
NIBR pKa values	Calculate pKa values for compounds in current table using internally trained Moka model.	Web service
NIBR logD values	Calculate logD values for compounds in current table using internally trained Moka model.	Web service
LogD curve	Add logD curve graph to current table.	Web service
Species distribution	Add species distribution graph to current table	Web service
SmartCyp predictions	Predict site of metabolism using SmartCyp.	Web service
Phototox predictions	Assess the phototoxicity potential of compounds.	Web service

Other QSAR models	Access additional QSAR models provided by the CADD group.	Web service
Calculate basic 2D descriptors	Choose among hundreds of other 2D descriptors	Web service
Clear annotation and selection	Remove annotation and atom selection from structures in current table.	Client
Medicinal chemistry rules (Lilly)	Apply rules published by Bruns and Watson [J Med Chem 55 (2012) 9763-9772].	Web service
PAINS annotation	Apply substructure queries published by Baell and Holloway [J Med Chem 53 (2010) 2719-2740].	Web service
QED prediction	Calculate quantitative estimate of drug-likeness (QED) published by Bockerton *et al.* [Nature Chemistry 4 (2012) 90-98.	Web service
CNS MPO prediction	The CNS multiparameter optimization approach published by Wagner *et al.* [ACS Chem Neuroscience 1 (2010) 435-449] captures the property profile of marketed CNS drugs..	Web service
CNS MPO prediction (verbose output)	Displays indidividual contributions to CNS:MPO.	Web service
Lipinski filter	Filter table on Lipinski-type properties.	Client
'Golden Triangle' plot	Create a golden triangle plot of MW over logD(pH=7.4) with the 'Golden Triangle' area highlighted.	Client and web service
Passive Gut Absorption (Egan Egg plot)	Calculate required properties and create plot in current table.	Client and web service
Remove Compounds outside Egan Egg	Remove compounds in current table that are predicted to have poor passive gut absorption.	Client
Shape analysis NPR	The normalized ratios (NPR) of the principal moments of inertia (PMI) can be used to analyze the shape diversity of compound sets as published by Sauer and Schwarz (J Chem Inf Comput Sci 43 (2003) 987).	Web service
Shape analysis Fraction Csp3	The ratio of sp3 carbon atoms in a molecule is an indicator for three-dimensionality of a compound.	Web service

Table S4. FOCUS Analysis module

Function	**Description**	**Implementation**
Substructure and similarity search	Substructure and similarity search in tables. Remove compounds with a given substructure from table	Client
Unwanted Group Filter	Removal of predefined groups or multiple substructure retrieval	Client
Ligand efficiency index	Calculate ligand efficiency metrics from activity data.	Client
Scatterplot and Histogram	Create simple scatterplots and histograms. For full control use the Plot/Histogram button next to the tables.	Client

Principal component analysis (PCA)	Creates a principal component analysis of the data for the current table and adds a scatter plot.	Client
Transformation	Apply transformations like smiles to molecule or IC50 to pIC50 to table columns.	Client
Categorization	Add column or colors to table based on categorization of data.	Client
Extract selected rows	Extract selected rows from current table and move to a new table.	Client
Cluster structures	Cluster molecules in current table	Client
Extract cluster centers	Extract cluster centers from current table (clusters if required)	Client
R group decomposition	Decomposes molecules in table based on common core and defined R-group positions	Client

Table S5. FOCUS 3D module

Function	Description	Implementation
Minimize / Generate Conformer	Generate a single minimized structure.	Client
Generate multiple conformations	Generation of multiple conformations per molecule.	Client
Flexible Overlay of Molecules	Find best overlay of two or three flexible molecules.	Web service
Overlay Molecules to Rigid Template	Flexible alignment of molecules to a rigid template.	Web service
Retrieve Results from Overlay	Check status of (asynchronous) overlay jobs.	Web service

Table S6. FOCUS Protein module

Function	Description	Implementation
Retrieve Protein	Search internal and external protein structure databases by keyword or PDB ID.	Web service and client
Ligand Search	Substructure and similarity search of ligands in internal and external protein structure databases.	Web service and client
Overlay Proteins	Align one protein structure to another.	Client
Convert Protein	Necessary to highlight H-bonds, display surface properties etc.	Client
Add Various Surfaces	Create surfaces around selected ligand	Client

Binding Site Analysis	Create visualization of important regions in binding site around ligand	Client
Binding Site View	Create view of binding site around ligand	Client
Prepare table for docking	Modify the current table so that structure in table can be used in docking.	Client
Dock compounds in table	Dock all compounds in table. Needs pre-aligned structures for minimization.	Client
Kinase ABC	Open the Kinase ABC for Focus	Web service and Client

Table S7. FOCUS specialist modules (Metabolism, multi objective optimization and protein structures)

Function	**Description**	**Implementation**
Metabolism		
SmartCyp (full version)	Predict site of metabolism using SmartCyp.	Web service
Permeability Guide	Create guide for permeability diagnosis.	Web service and Client
Reformat pKa values	Create a new table with the pKa values reformatted in a format suitable for %fraction ionized calculations	Client
Multi objective optimization		
Pareto ranking	Please select the numerical columns in the table to use for the Pareto ranking.	Web service
Create ideas	Create ideas for the molecules in current table.	Web service
Protein structure		
Import protein or sequence data	Import multiple pdb or sequence files. Sequence files should be .fas or .seq format.	Client
Extract protein sequences	Extract sequences from protein chains in workspace.	Client
Align multiple proteins by structure	Align multiple proteins to single template by structure.	Client
Align multiple sequences	Align multiple sequences from the workspace (either extracted or read in).	Client
Link structures to alignment	Relink structures in workspace to currently active sequence alignment (important for selection etc.).	Client
Assign residue labels	Assign standard residue labels around selected ligand.	Client
Create alignment analysis page	For the currently displayed alignment create a new html for standard tasks like highlighting, labelling etc.	Client

Table S8. FOCUS experimental module

Function	Description	Implementation
Predict PAMPA	Predict PAMPA using internal model.	Web service
Predict Caco-2 permeability	Predict Caco-2 passive permeability using a risk-bin model.	Web service
Predict Caco-2 efflux	Predict Caco-2 efflux using a risk-bin model.	Web service
Predict MDR1-MDCK permeability	Predict MDR1-MDCK passive permeability using a risk-bin model.	Web service
Predict MDR1-MDCK efflux	Predict MDR1-MDCK efflux using a risk-bin model	Web service
Predict pK_a values	Uses retrained models with improved prediction on internal data.	Web service
Calculate fraction ionized	Calculates fraction ionized based on Moka results.	Web service
CADD QSAR models	Access QSAR models from CADD.	Web service
Amber	Generates Amber input files to run minimizations or MD simulations.	Client
MOE ligand interaction plot	Generates MOE ligand interaction plots for a protein and a table of (docked) molecules.	Web service
Molecular Matched Pair analysis (CIx)	Creates pair analysis table which can be used for transformation vector work etc.	Web service
Import assay data	Retrieve assay data from internal data hub for compounds in the current table.	Web service
Deprotect molecules	Remove common protecting groups from molecules in table.	Web service
Calculate tautomers or protomers	Calculate tautomers or protomers using Moka.	Web service
Free Wilson analysis	Given an R-group decomposition, run a Free-Wilson analysis.	Client
Shape-based overlay to rigid Template	Alignment of molecules to a rigid template based on molecular shape only.	Web service
Isostere replacement tool	Replace parts of a molecule with fragments based on PDB ligands and CSD crystal structures.	Web service