

Siphonify—Managing Chemical Structures with Style

Trung Nguyễn Ajit Jadhav
{nguyenda, ajadhav}@mail.nih.gov

NIH Chemical Genomics Center
<http://www.ncgc.nih.gov>

August 22, 2011



Features Available

- An easy-to-use tool to manage chemical structures similar to how music files and documents are managed
- Provide a convenient repository for searching and browsing of “registered” chemical structures
- Support for basic form of chemical registration (insofar as to recognize when two structures are the “same”)
- Built-in support for structure standardization and normalization
- Batch and interactive compound registration
- Various search forms; e.g., keyword, substructure, similarity, exact, etc.
- Support for “smart” collections and filters
- Advanced set operations on collections and filters



- Importing and searching of well-known public data sources (e.g., ChEMBL, PubChem, etc.)
- Integrate other tools available on <http://tripod.nih.gov>
 - Automated R-group analysis
 - Scaffold-based analyses and visualizations
- Automatic deployment of validated filters and well-known collections (e.g., curated drugs)
- Sharing of filters and collections
- Deployment of QSAR/QSPR models
- Improve on performance

Basic Usage

Batch import

Support for regular expression of individual import fields

Siphonify [DEVEL] — DSSTox (10,453 total)

Import Compounds

Name: hERG

Source: file:/home/trung/herg-fdss.sdf

Scan data source for import fields...

Field	If matches...	Import as...
PUBCHEM_SID		Synonym
SAMPLE_ID		PrimaryID
SAMPLE_NAME		DisplayName
SMILES_ISO		None
SUPPLIER	*Sigma[aA]ldrich*	None
SUPPLIER_ID		Synonym
herg-trans-ntp-p1-AC...		None
hero-trans-ntp-n1-CRC		None

31 unique fields found!

Import Close

1,000 compounds, collection "DSSTox"

Basic Usage

Interactive import

Add a new compound to the “My favorites” collection

The screenshot displays the Siphonify [DEVEL] software interface. The main window title is "Siphonify [DEVEL] – My favorites (2 total)". The interface is divided into several sections:

- Left Panel (Navigation):** Includes "CONTENTS" (Compounds), "FILTERS" (PAINS - m150, PAINS - I150, PAINS - I15), and "COLLECTIONS" (My favorites, DSSTox, DSSTox - m150, NCI, NPC & DSSTox, NPC drugs, WB drugs, curated drugs, hERG, lopac, m150 & I150). A "TASKS" section at the bottom shows "Registering compound ..." as a completed task.
- Top Bar:** Contains navigation icons (home, refresh, save, undo), a search bar labeled "Search Compounds", and window control buttons.
- Central Canvas:** Shows a chemical structure of a complex molecule.
- Right Panel (New compound: phenylpropan):** A form for adding a new compound. It has tabs for "Structure", "Synonyms", and "Therapeutics". The "Name" field contains "phenylpropan". The "URL" field is empty. The "Structure" field displays the chemical structure of phenylpropan (CC(O)c1ccccc1). Below the structure, it says "Double-click to edit structure 0 stereocenters defined (out of 1 possible)". The "Comments" field contains the text "a intersting compound found while reading a paper". At the bottom of the panel are "Add", "Clear", and "Close" buttons.

At the bottom of the window, a status bar indicates "2 compounds, collection 'My favorites'" and a scrollbar is visible.

Basic Usage

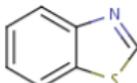
Initiating (sub-) structure search

The screenshot displays the Siphonify [DEVEL] software interface. The main window title is "Siphonify [DEVEL] — Compounds (23,231 total)". The search bar at the top right contains the query "sub:S1C=NC2=CC=CC=C12".

The left sidebar contains a "CONTENTS" panel with sections for "Compounds", "FILTERS", and "COLLECTIONS". The "COLLECTIONS" section includes "My favorites", "DSSTox", "DSSTox - m150", "NCI", "NPC & DSSTox", "NPC drugs", "WB drugs", "curated drugs", "hERG", "lopac", and "m150 & I150". A "TASKS" section at the bottom shows "Registering compound ...".

The main workspace shows a grid of chemical structures. The top row includes a structure labeled "2-(4-Aminophenyl)-...", a structure labeled "Rifamexil", and a structure labeled "Structure". The bottom row shows a structure labeled "Structure Search".

The "Structure Search" dialog box is open, showing the following options:

- Search Type: Substructure, Similarity, Exact, Topology
- Tanimoto: Top N:
- Structure: 
- Double-click to edit structure
- Buttons: Search, Close

The bottom status bar indicates "91 compounds".

Basic Usage

Implicit neighbor search

Right-click on any structure to perform implicit neighbor/similarity search

The screenshot displays the Siphonify [DEVEL] software interface. The title bar reads "Siphonify [DEVEL] - Compounds (23,231 total)". The search bar contains the query "sub:S1C=NC2=CC=CC=C12". The main window shows a list of 15/91 compounds. A chemical structure of ETHOXZOLAMIDE is highlighted with a blue border. A right-click context menu is open over this structure, with the following options: "Show similar compounds", "Edit", "Save As..." (with a keyboard shortcut of Ctrl+Shift-S), "Save Selection...", and "Save Image...". Below the main structure, three smaller chemical structures are shown, labeled "DSSTox", "WB drugs", and "hERG". The left sidebar contains a "CONTENTS" panel with "Compounds" selected, and a "FILTERS" section with "PAINS - m150", "PAINS - I150", and "PAINS - I15". Below filters are "COLLECTIONS" including "My favorites", "DSSTox", "DSSTox - m150", "NCI", "NPC & DSSTox", "NPC drugs", "WB drugs", "curated drugs", "hERG", "lopac", and "m150 & I150". At the bottom of the sidebar, there is a "TASKS" section with "Registering compound...". The bottom status bar shows "91 compounds" and a scroll bar.

Collections and Filters

Creating collections

Collections and filters are “tags” on chemical structures

The screenshot shows the Siphonify [DEVEL] interface with the collection "NPC & DSSTox" selected, containing 667 total compounds. A context menu is open over the first row, showing options: Compound, Filter, Collection, and Dynamic Collection. The table lists the following compounds:

Name	Collections	Synonyms
L-Ascorbic acid	NPC drugs DSSTox NPC & DSSTox HERG PAINS - I15	sodium (2R)-2,3:4,5-diolactone-5-phosphate 20105 105 L-ascorbic acid 20106 5785
atropine	NPC drugs DSSTox NPC & DSSTox WB drugs curated drugs	(3-endo)-8-azabicyclo[3.2.1]octan-3-yl 3-hydroxybenzoate 20113 12310716 17184 5318386 154417
Azacitidine	NPC drugs DSSTox NPC & DSSTox WB drugs curated drugs HERG lopac	4-amino-1,2,3,4-tetrahydro-2H-1,2,4-triazin-6(1H)-one 20116 9444 1265902 460485 1805

Collections and Filters

Editing collections

Right-click on a selected set of structures to add to a collection

The screenshot displays the Siphonify [DEVEL] software interface. The title bar reads "Siphonify [DEVEL] — PAINS - m150 (852 total)". The interface includes a toolbar with icons for home, refresh, save, undo, add, edit, and delete. A search bar is labeled "Search Compounds".

The left sidebar contains a "CONTENTS" section with "Compounds". Below it is a "FILTERS" section with "PAINS - m150" selected, along with "PAINS - I150" and "PAINS - I15". A "COLLECTIONS" section lists "My favorites", "DSSTox", "DSSTox - m150", "NCI", "NPC & DSSTox" (667), "NPC drugs", "WB drugs", "curated drugs", "hERG" (1000), "lopac", and "m150 & I150". A "TASKS" section shows "Registering compound ...".

The main area shows a grid of chemical structures. A context menu is open over a selected set of structures, with options: "Select All", "Add to Collections", "Add selections to collections", "My favorites", "DSSTox", "DSSTox - m150", "NCI", "NPC & DSSTox", "NPC drugs", "WB drugs", "curated drugs", "hERG", "lopac", and "m150 & I150".

The status bar at the bottom indicates "852 compounds, collection 'PAINS - m150', 8 selected".

Collections and Filters

Creating filters

Below are the PAINS filters available at

<http://blog.rguha.net/?p=850>

Siphonify [DEVEL] – PAINS - I15 (465 total)

Update Filter: PAINS - I15

Name: PAINS - I15

Description: less than 15 block

Operation: Inclusion Exclusion

Filter ID	Operation	Expression
1	Substructure	<chem>[*]1[*]([*6;X4])N(-[*]([*1]</chem>
2	Substructure	<chem>[*1A]NN=Cc1c([*1A])c([*1A]</chem>
3	Substructure	<chem>[*1A]Oc1c(-[*]([*1]),[*]([*6]-[</chem>
4	Substructure	<chem>[*]([*1]),[*]([*6]:[*6])~c1c(~[*]</chem>

409 expressions loaded!

Import Clear Update Close

PHOSPHAMIDON PIROXICAM C.I. Pigment red 23

465 compounds, collection "PAINS - I15"

Collections and Filters

Examining individual filter rules

Siphonify [DEVEL] — PAINS - m150 (852 total)

659/852

Identifiers Properties Therapeutics

Name
vofopitant

PAINS - m150
8: [#1A]Oc1ccccc1-[#6;X4]-[#7]
PAINS - I150
22: [#1A]N([#1A])c1c([#1A])c(-[(\$(#8),\$(#7)),(\$
PAINS - I15
243: [#1A]C([#1A])c1nnnn1-c2c([#1A])c([#1A])c(OC([#1

Collections Filters

Substances Comments

WB drugs NPC drugs

852 compounds, collection "PAINS - m150"

The screenshot displays the Siphonify software interface. The main window title is "Siphonify [DEVEL] — PAINS - m150 (852 total)". The interface is divided into several sections. On the left is a sidebar with a "CONTENTS" menu containing "Compounds", "FILTERS", and "COLLECTIONS". Under "FILTERS", "PAINS - m150" is selected. Under "COLLECTIONS", "My favorites" is expanded, showing "DSSTox", "DSSTox - m150", "NCI", "NPC & DSSTox" (with a badge of 667), "NPC drugs", "WB drugs", "curated drugs", "hERG" (with a badge of 1000), "lopac", and "m150 & I150". The main area shows a chemical structure of vofopitant, a benzamide derivative with a piperidine ring and a fluorinated benzimidazole group. Below the structure is the name "vofopitant". To the right of the structure is a panel with tabs for "Identifiers", "Properties", and "Therapeutics". The "Identifiers" tab is active, showing the name "vofopitant" and a list of filter rules. The first rule is "PAINS - m150" with a count of 8 and the SMILES pattern "[#1A]Oc1ccccc1-[#6;X4]-[#7]". The second rule is "PAINS - I150" with a count of 22 and a complex SMILES pattern. The third rule is "PAINS - I15" with a count of 243 and a complex SMILES pattern. Below the filter rules are tabs for "Collections" and "Filters". At the bottom of the main area are two icons for "WB drugs" and "NPC drugs". The bottom status bar indicates "852 compounds, collection 'PAINS - m150'".

Collections and Filters

Operations on collections and filters

Basic set operations (e.g., AND, OR, XOR) are supported as well as advanced logic expressions

The screenshot shows the Siphonify [DEVEL] interface with a dialog box titled "Update Dynamic Collection: m150 & I150". The dialog contains the following fields and options:

- Name: m150 & I150
- Description: (empty text area)
- Operation: Custom (selected), with other options: Overlap, Combine, Exclusive, Not In.
- Logic Expression: $ABP \ \&\& \ ((ABP \ != \ ABZ) \ \&\& \ (ABP \ != \ ACA))$
- Filters table:

Filter	Collection
ABP	PAINS - m150
ABZ	PAINS - I150
ACA	PAINS - I15

Buttons: Clear, Update, Close

Background interface details:

- Window title: Siphonify [DEVEL] -- m150 & I150 (432 total)
- Search bar: Search Compounds
- Left sidebar: CONTENTS (Compounds), FILTERS (PAINS - m150, PAINS - I150, PAINS - I15), COLLECTIONS (My favorites, DSSTox, DSSTox - m150, NCI, NPC & DSSTox, NPC drugs, WB drugs, curated drugs, hERG, lopac, m150 & I150), TASKS (Registering compound ...)
- Status bar: 432 compounds, collection "m150 & I150"

Collections and Filters

More operations on collections and filters

The screenshot shows the Siphonify [DEVEL] interface with a dialog box titled "Update Dynamic Collection: curated drugs". The dialog contains the following fields and options:

- Name:** curated drugs
- Description:** confident drug structures from three curated data sources
- Operation:** Overlap, Combine, Exclusive, Not In, Custom
- fx:** AAC && AAM && AAV
- Collection List:**

<input type="checkbox"/>	<input checked="" type="checkbox"/>	AAC	Collection	NPC drugs
<input type="checkbox"/>	<input checked="" type="checkbox"/>	AAM	Collection	DSSTox
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	AAV	Collection	WB drugs
- Buttons:** Clear, Update, Close

The background interface shows a sidebar with "CONTENTS" (Compounds, FILTERS, COLLECTIONS) and "TASKS". The main area displays chemical structures, including atropine. The status bar at the bottom indicates "568 compounds, collection 'curated drugs'".

- Siphonify is available for download from here
<http://tripod.nih.gov/siphonify>
- Please give us feedback either via email or the feedback button within the software 

Our informatics colleagues have been instrumental in the development of Siphonify:

- Rajarshi Guha's web-based filter tool—*siphon*—provides the initial inspiration.
- Noel Southall's original work on the `CanonicalForm` code serves as guidance for the structure standardization code within Siphonify. Any shortcomings are certainly our fault.