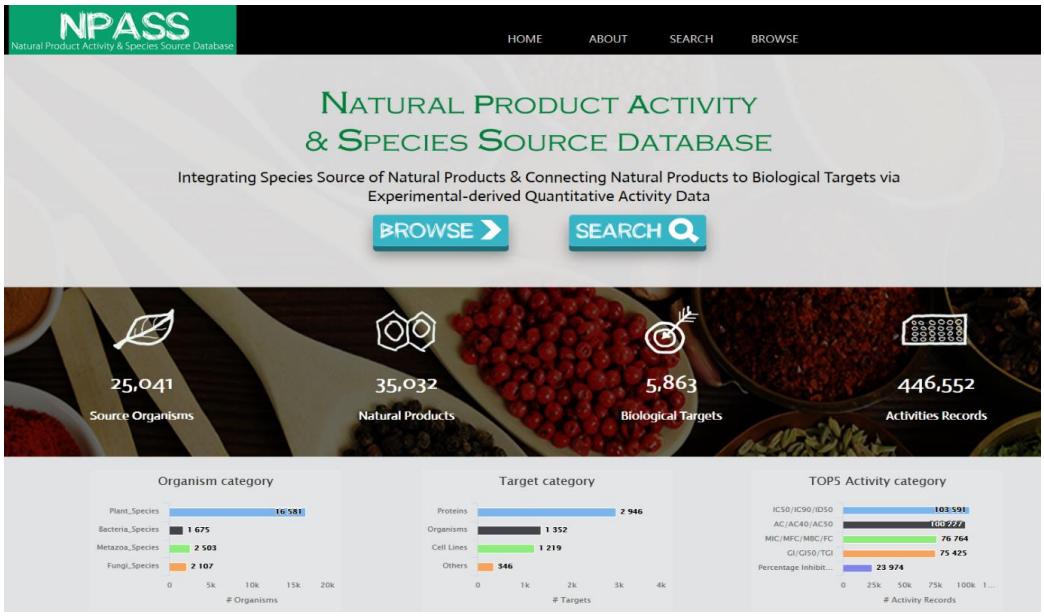
NPASS Database Instruction Manual for Users

Thanks for visiting **NPASS** (**N**atural **P**roduct **A**ctivity & **S**pecies **S**ource database)!

AUG-2017



CONTENTS

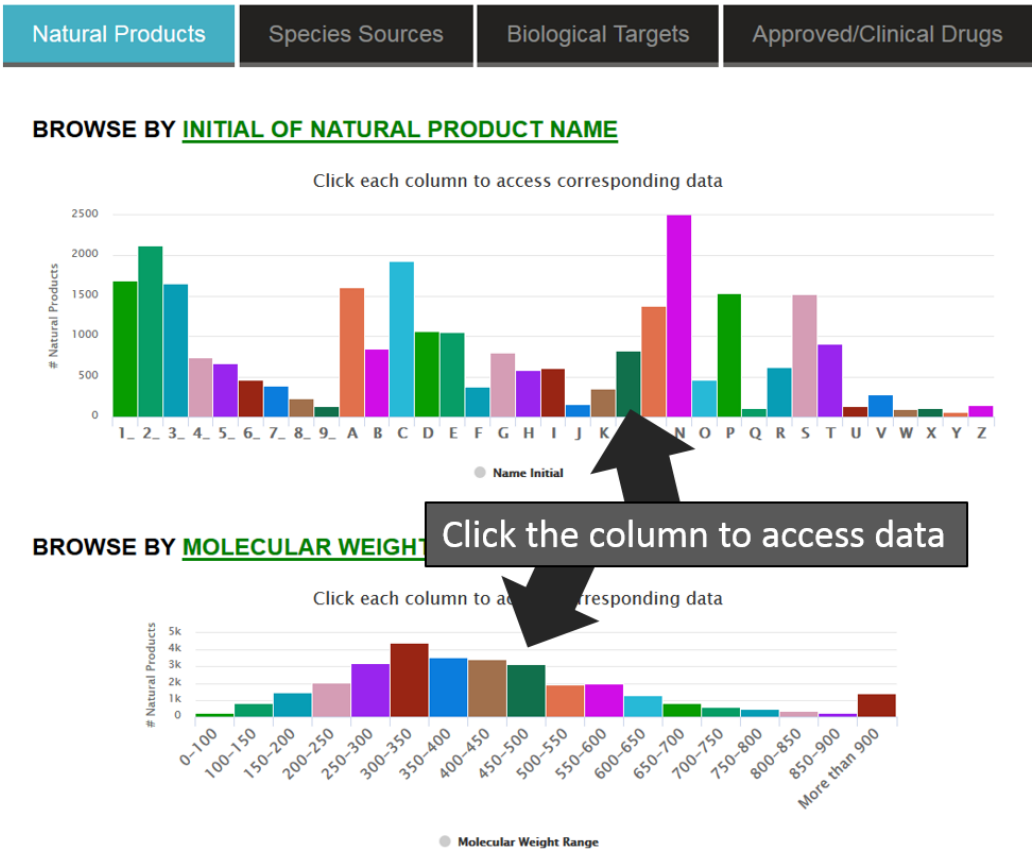
DETAILED INSTRUCTION

BROWSE THE DATABASE

Users can browse the database via clicking the “BROWSE” button on the homepage, and then access specific data via multiple options.

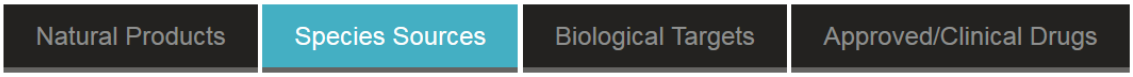
1. Browse NPASS by initial of NPs name or by molecular weight range of NPs

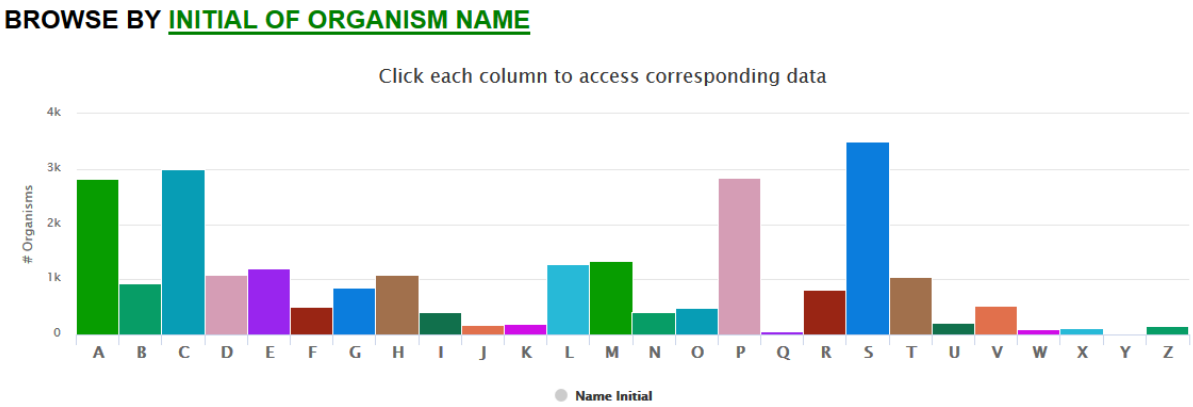
Users can click specific column on the bar charts to access corresponding data. A new webpage will be shown to present a table that contains matched NPs on you click. Detailed information of individual NPs can be accessed by clicking the ‘Natural Product ID’.



2. Browse NPASS by initial of species source name

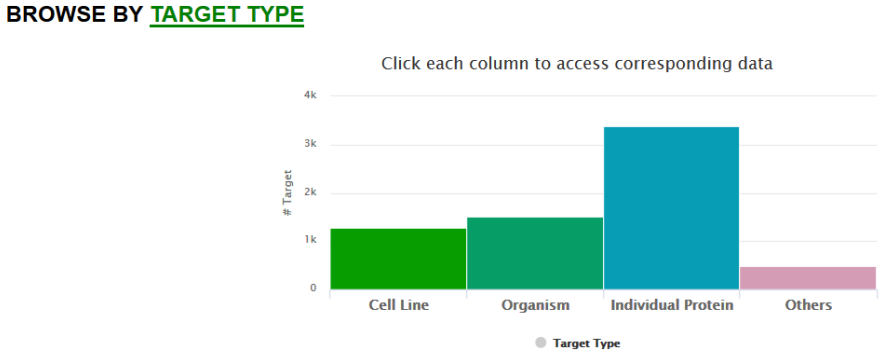
Users can click specific column on the bar charts to access corresponding data. A new webpage will be shown to present a table that contains matched species on you click. This table includes organism name, brief taxonomic information, and number of NPs isolated from this species. Detailed information of individual species can be accessed by clicking the 'Organism ID'.





3. Browse NPASS by target category

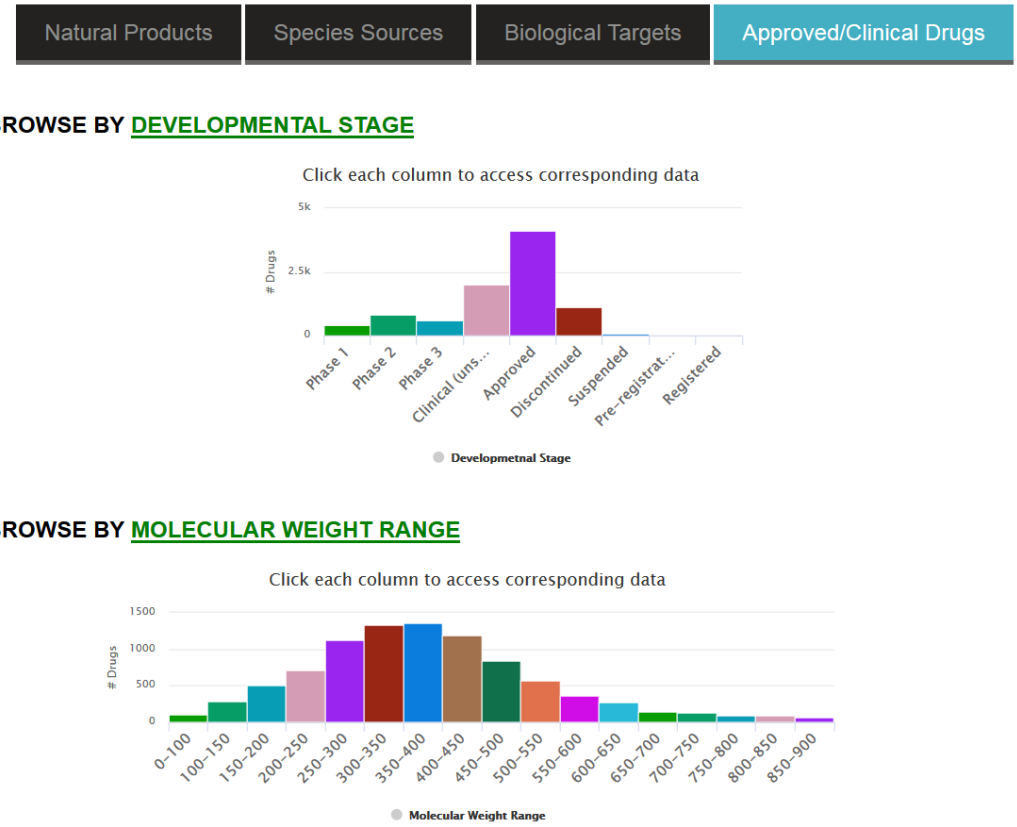
Users can click specific column on the bar charts to access corresponding data. A new webpage will be shown to present a table that contains matched targets on you click. The table includes number of NPs that have activity measurement on against this target and number of total activity records. The number of activity records may be larger than number of NPs because there are multiple activity records for the same NP-Target pair. Detailed information of each target can be accessed by clicking the "Target ID".



4. Browse NPASS by Approved/Clinical drugs

Users can click specific column on the bar charts to access corresponding data. A new webpage will be shown to present a summary of matched drugs. This table includes information of drugs (drug name, max developmental stage) and numbers of similar NPs at each similarity levels (High similarity level, Intermediate similarity level, and Remote similarity level).

The similarity levels are defined based on Tanimoto coefficient (Tc), a widely used index for descripting structural similarity between two molecules, at different cutoffs (Tc>=0.85, 0.7=<Tc<0.85, 0.58<=Tc<0.7, respectively) as descripted previously (Ref: Zhang C. *et al*. *J Mol Graph Model.* 2017, 76:136-142). By clicking the "Drug ID", users can access the detailed information of the drug, which includes detailed information of the drug and similar NPs.

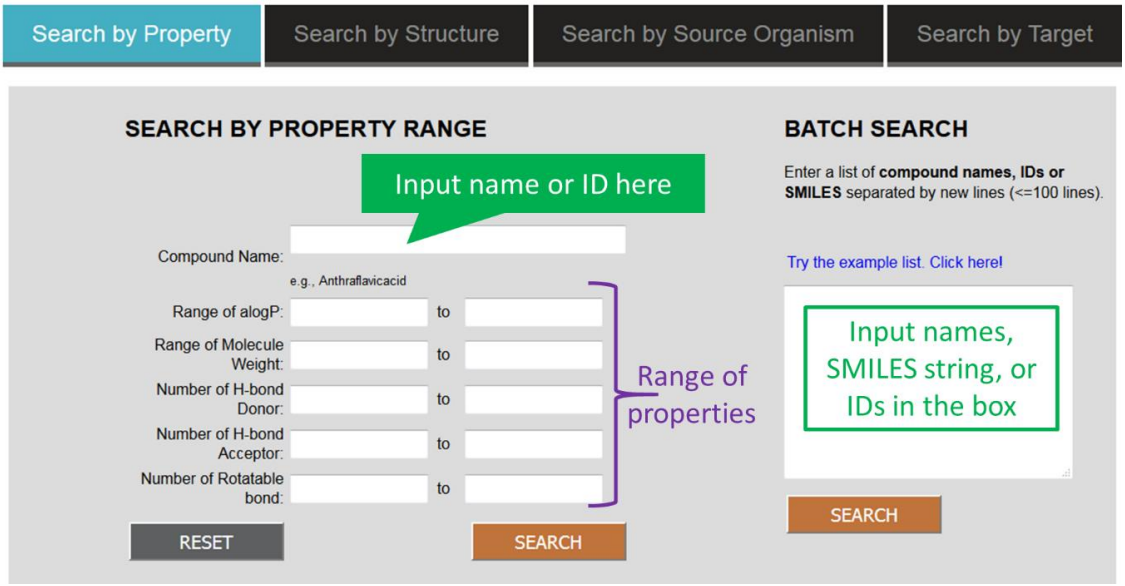


SEARCH THE DATABASE

1. Search by NPs name/IDs, properties ranges, and batch search

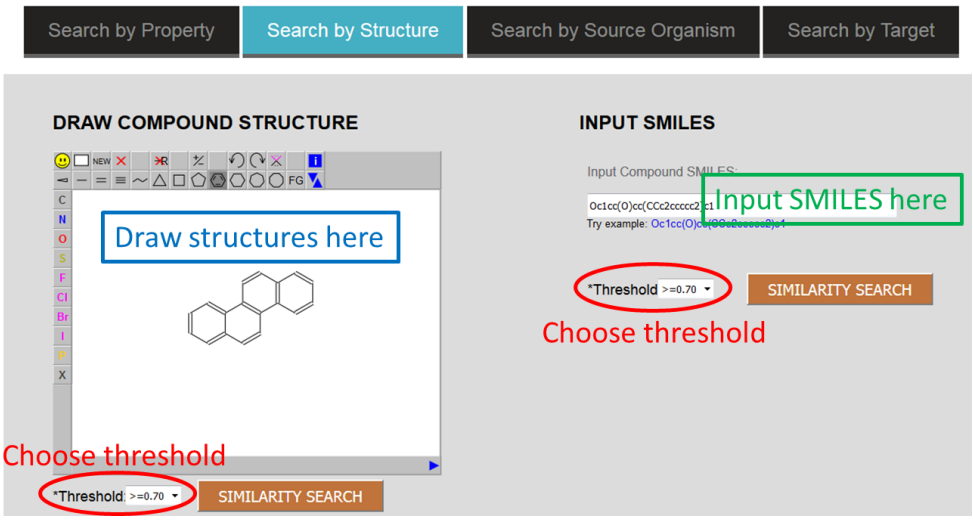
Users can input compound name (can be common name, synonyms, or IUPAC name), range of ALogP, range of molecular weight, range of number of hydrogen bond donor (HBD), range of number of hydrogen bond acceptor (HBA), range of number of rotatable bond. Users can use one or combination of multiple of these options to customize their search.

Besides, users can also input a list of names, or smiles strings, or IDs (can be NPASS Natural Product ID, ChEMBL ID, or PubChem CID) in the input box to do a batch search. In the batch search box, names/SMILES/IDs should be separated by new line (one line one name/SMILES/ID). After clicking the "SEARCH" button, a new webpage will display matched NPs.



2. Search by NP's structure (Similarity search)

To do a similarity search, users can either draw a structure or input a SMILES string on this page. Users can choose a similarity threshold to indicate the similarity level. The threshold is Tanimoto coefficient Tc (Tc lies in [0,1], "1" indicates the most similar level, "0" indicates the most dissimilar level). The similarity search will return a new webpage to display: a histogram to present the similarity distribution between query molecule and all NPs in NPASS, a histogram to present the similarity distribution between query molecule and all Approved/Clinical drugs deposited in NPASS, and a table included NPs of intermediate and high similarity level (Tc>=0.7). Clicking "Natural Product ID" in the table can access detailed information of individual NPs.



3. Search by species source

Users can search specific NP-derived species by inputting species taxonomic name or NCBI taxonomy IDs. A new webpage will be presented to display matched species and specific species can be accessed by clicking the “Organism ID” in the table.

4. Search by target

Users can search a specific target by inputting target name (can be protein name, microorganism name, etc.) or target IDs (Uniprot IDs and NPASS Target IDs are acceptable at the moment).

Besides, we are developing and testing the protein sequence similarity search function to enable users input a protein sequence and search similar protein targets by sequence similarity. This function will be done by Sep-15, 2017.

DOWNLOAD DATA

Uses can freely and easily download data from the NPASS. In the webpage of natural product, there is a download summary table that includes several data sections users can choose depends on their interest.

These sections includes:

1. General Info & Identifiers & Properties: *This section includes all data displayed in the table of "Natural Product" part of this webpage and data in the "External Identifiers" part.*

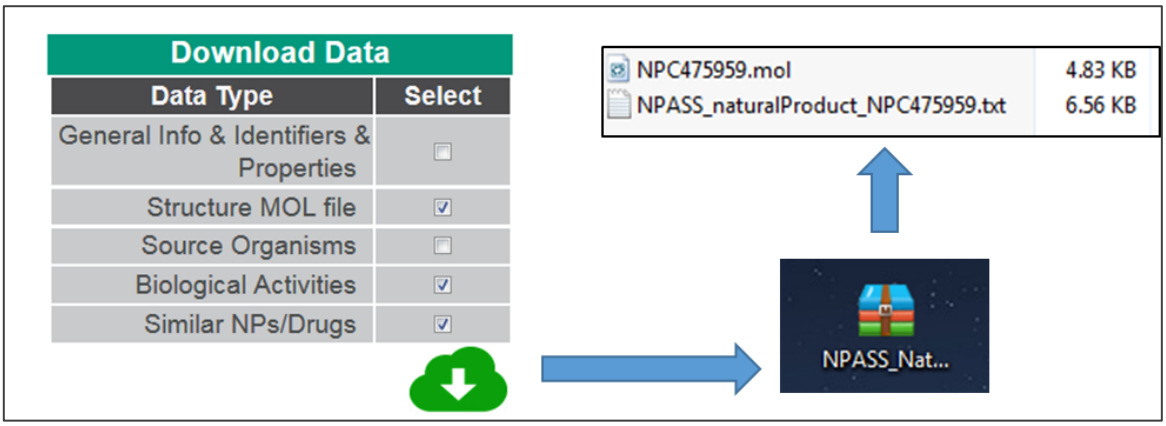
2. Structure MOL file: *This section includes a MOL file of the NP.*

3. Species Sources: *This section includes all data displayed in the table of the "Species Source" of the webpage.*

4. Biological Activities: *This section includes all data displayed in the table of the "Biological Activity" part of the webpage.*

5. Similar NPs/Drugs: *This section includes all data displayed in the tables of the "Similar Natural Products in NPASS" and "Similar Approved/Clinical Drugs" parts of the webpage.*

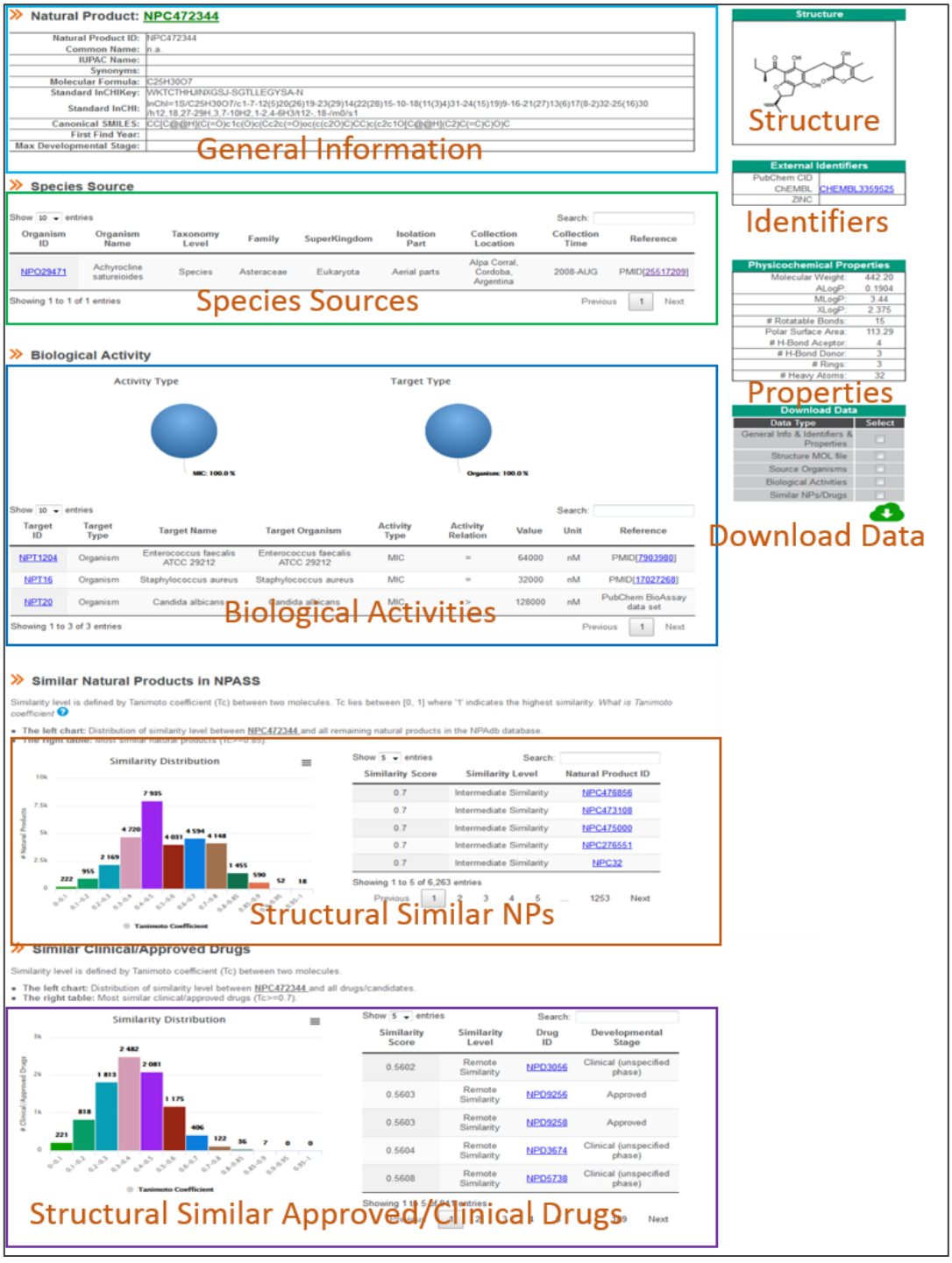
Users can choose part or all of these sections to download. The downloaded file is a compressed zip file that includes a txt file and/or a MOL file. A detailed instruction can also be found in the txt file.



DESCRIPTION OF INDIVIDUAL WEBPAGE

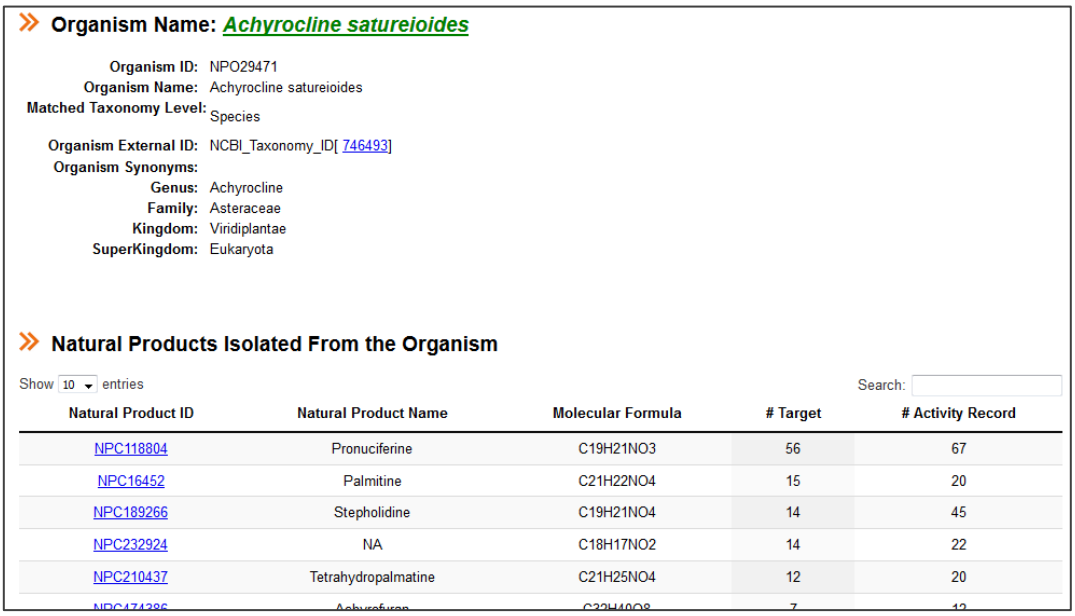
1. Webpage of a specific Natural Product

The NP page includes six sections: 1. Natural Product, 2. Species Source, 3. Biological Activity, 4. Similar Natural Products in NPASS, 5. Similar Approved/Clinical Drugs, 6. The right side part of the webpage includes image of structure, external identifiers, physicochemical properties, and download data tables.



2. Webpage of a specific species source

The species source page includes two sections: 1. Organism name, 2. Natural Products Isolated From the Organism. Users can find taxonomic information and external links (if applicable) of the species in the section 1 and all NPs from this organism. Users can access specific NPs by clicking “Natural Product ID” in the table of section 2.



Organism Name: *Achyrocline satureioides*

Organism ID: NPO29471
Organism Name: Achyrocline satureioides
Matched Taxonomy Level: Species
Organism External ID: NCBI_Taxonomy_ID[746493]
Organism Synonyms:
Genus: Achyrocline
Family: Asteraceae
Kingdom: Viridiplantae
SuperKingdom: Eukaryota

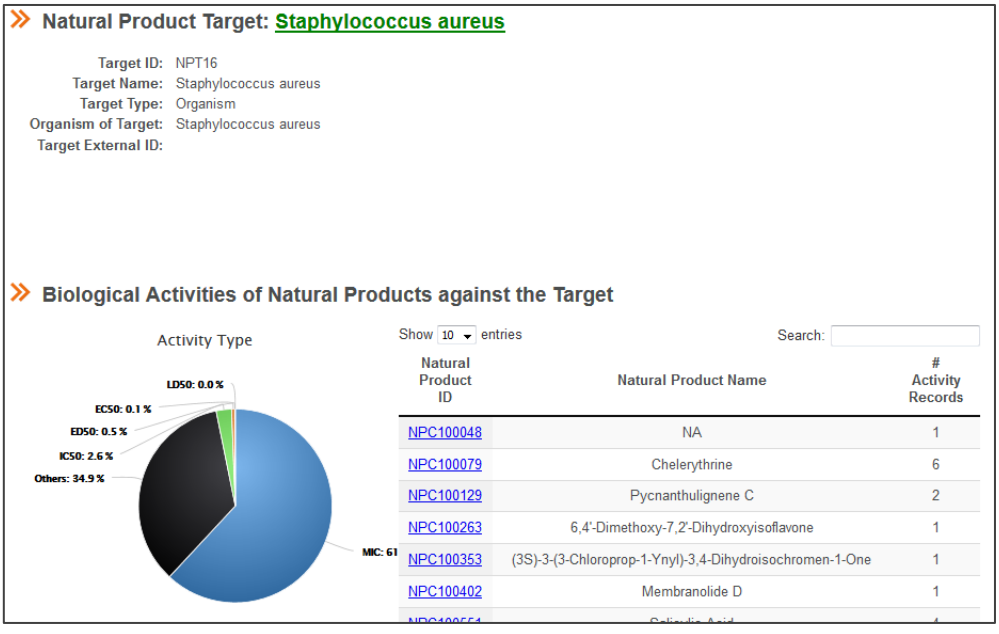
Natural Products Isolated From the Organism

Show 10 entries Search:

Natural Product ID	Natural Product Name	Molecular Formula	# Target	# Activity Record
NPC118804	Pronuciferine	C19H21NO3	56	67
NPC16452	Palmitine	C21H22NO4	15	20
NPC189266	Stepholidine	C19H21NO4	14	45
NPC232924	NA	C18H17NO2	14	22
NPC210437	Tetrahydropalmatine	C21H25NO4	12	20

3. Webpage of a specific target

The target page includes two sections: 1. Natural Product Target, 2. Biological Activities of Natural Products against the Target. Users can find detailed information of the target in section 1 and access each NPs that have activity records against this target by clicking the “Natural Product ID” in the table of section 2.



Natural Product Target: Staphylococcus aureus

Target ID: NPT16
Target Name: Staphylococcus aureus
Target Type: Organism
Organism of Target: Staphylococcus aureus
Target External ID:

Biological Activities of Natural Products against the Target

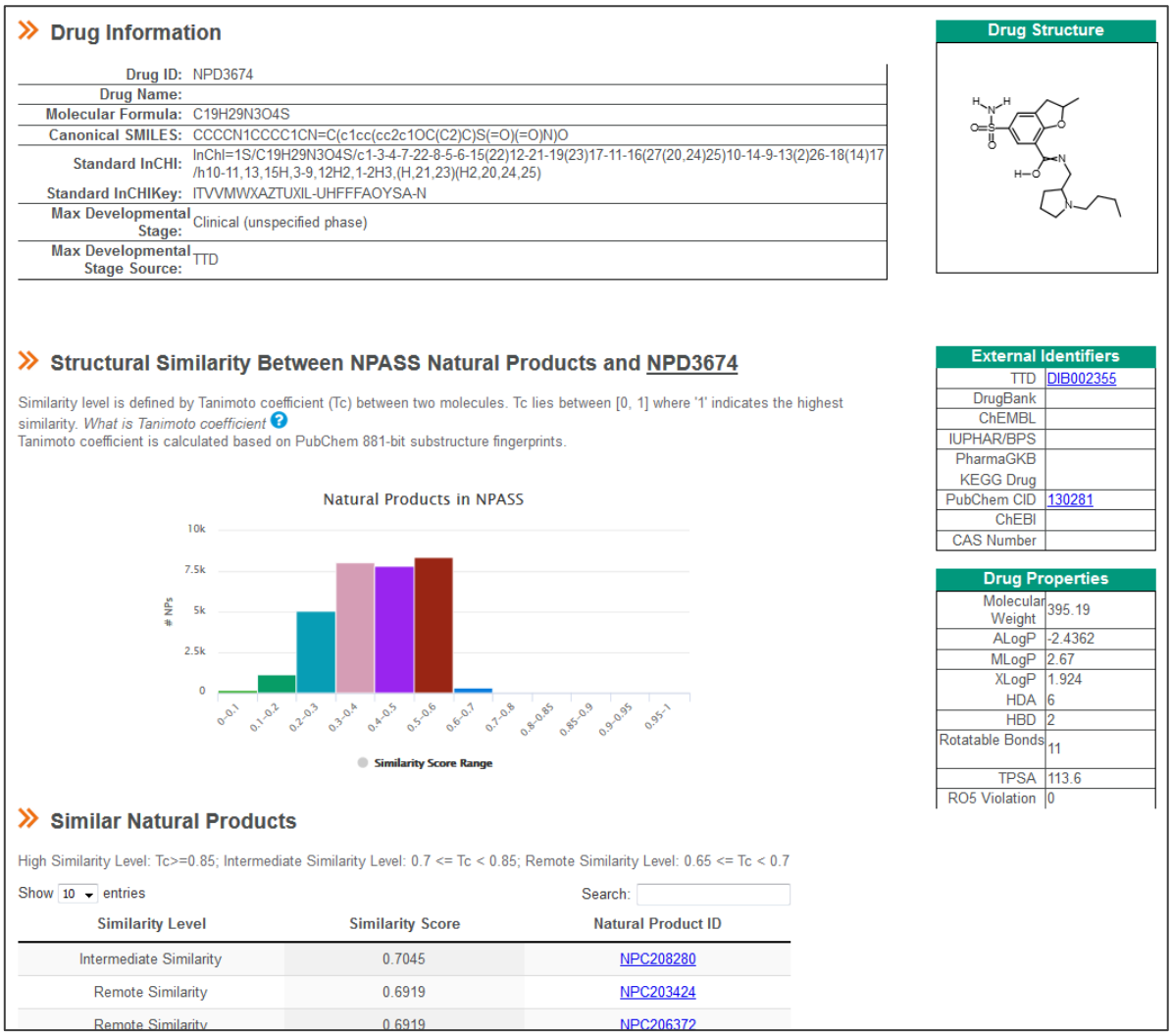
Activity Type

Show 10 entries Search:

Natural Product ID	Natural Product Name	# Activity Records
NPC100048	NA	1
NPC100079	Chelerythrine	6
NPC100129	Pycnanthulignene C	2
NPC100263	6,4'-Dimethoxy-7,2'-Dihydroxyisoflavone	1
NPC100353	(3S)-3-(3-Chloroprop-1-Ynyl)-3,4-Dihydroisochromen-1-One	1
NPC100402	Membranolide D	1

4. Webpage of a specific Approved/Clinical drug

The drug page includes three sections: 1. Drug Information, 2. Similarity Distribution between the Drug and all NPs in NPASS, 3. Similar NPs. Section 1 includes brief information of the drug including max developmental stage, several external identifiers which links to each database, and properties of the drug. Users can have an overview of similarity distribution of all NPs in NPASS to this drug from the histogram of section 2. From section 3, users can access each similar NPs.



Drug Information

Drug ID:	NPD3674
Drug Name:	
Molecular Formula:	C19H29N3O4S
Canonical SMILES:	CCCCN1CCCC1CN=C(c1cc(cc2c1OC(C2)C)S(=O)(=O)N)O
Standard InCHI:	InChI=1S/C19H29N3O4S/c1-3-4-7-22-8-5-6-15(22)12-21-19(23)17-11-16(27(20,24)25)10-14-9-13(2)26-18(14)17/h10-11,13,15H,3-9,12H2,1-2H3,(H,21,23)(H2,20,24,25)
Standard InCHIKey:	ITVVMWXAZTUXIL-UHFFFAOYSA-N
Max Developmental Stage:	Clinical (unspecified phase)
Max Developmental Stage Source:	TTD

Drug Structure

External Identifiers

TTD	DIB002355
DrugBank	
ChEMBL	
IUPHAR/BPS	
PharmaGKB	
KEGG Drug	
PubChem CID	130281
ChEBI	
CAS Number	

Drug Properties

Molecular Weight	395.19
ALogP	-2.4362
MLogP	2.67
XLogP	1.924
HDA	6
HBD	2
Rotatable Bonds	11
TPSA	113.6
RO5 Violation	0

Structural Similarity Between NPASS Natural Products and NPD3674

Similarity level is defined by Tanimoto coefficient (Tc) between two molecules. Tc lies between [0, 1] where '1' indicates the highest similarity. *What is Tanimoto coefficient*
Tanimoto coefficient is calculated based on PubChem 881-bit substructure fingerprints.

Similar Natural Products

High Similarity Level: Tc>=0.85; Intermediate Similarity Level: 0.7 <= Tc < 0.85; Remote Similarity Level: 0.65 <= Tc < 0.7

Show 10 entries Search:

Similarity Level	Similarity Score	Natural Product ID
Intermediate Similarity	0.7045	NPC208280
Remote Similarity	0.6919	NPC203424
Remote Similarity	0.6919	NPC206372