Seaweed metabolite database (SWMD): A database of natural compounds from marine algae

G Dicky John Davis[1] & A Hannah Rachel Vasanthi[2]

[1]Department of Bioinformatics, Sri Ramachandra University, Porur, Chennai, 600116 India; [2]Herbal & Indian Medicine Research Laboratory (HIMRL), Department of Biochemistry, Sri Ramachandra University, Porur, Chennai, 600116 India. G. Dicky John Davis - Email: Dicky.John@gmail.com; Phone: +914424768042; Fax: +914424765995; *Corresponding Author.

Abstract:
The cataloguing of marine chemicals is a fundamental aspect for bioprospecting. This has applications in the development of drugs from marine sources. A publicly accessible database that provides comprehensive information about these compounds is therefore helpful. The Seaweed Metabolite Database (SWMD) is designed to provide information about the known compounds and their biological activity described in the literature. Geographical origin of the seaweed, extraction method and the chemical descriptors of each the compounds are recorded to enable effective chemo-informatics analysis. Cross-links to other databases are also introduced to facilitate the access of information about 3D Structure by X-ray and NMR activity, drug properties and related literature for each compound. This database currently contains entries for 517 compounds encompassing 25 descriptive fields mostly from the Red algae of the genus Laurencia (Ceramiales, Rhodomelaceae). The customized search engine of this database will enable wildcard querying, which includes Accession Number, Compound type, Seaweed Binomial name, IUPAC name, SMILES notation or InChI.

Availability: http://www.swmd.co.in

Background:
Marine chemicals have novel structures with pronounced biological activity and pharmacology. The study of such chemicals therefore is promising. High throughput screening of marine metabolites for a given drug target can be achieved only if natural compounds are available as a database. Creating a database of natural products and sharing it with huge scientific community facilitates the understanding of basic mechanism of compounds and can reduce the timeline in drug discovery [1]. A publicly accessible database that provides comprehensive information about these compounds is therefore helpful to the relevant communities.

Seaweeds are among the first marine organisms chemically analyzed, with more than 3,600 articles published describing 3,300 secondary metabolites from marine plants and algae, and they still remain an almost endless source of new bioactive compounds. This database is focused on bioactive compounds that target the pharmaceutical market, along with the spectrum of biological activities (**Table 1**). Among macroalgae, significantly more rich in secondary metabolites appear the brown and red algae, with the latter being the top producers of halogenated metabolites. Red algae of the genus Laurencia (Ceramiales, Rhodomelaceae) are some of the most prolific producers of secondary metabolites in the marine environment. Secondary metabolites from these algae are predominantly sesquiterpenes, diterpenes, triterpenes and C15-acetogenins, characterized by the presence of halogen atoms in their chemical structures. Most Laurencia species accumulate a characteristic major metabolite or a class of compounds not widely distributed within the genus [2].

Database Structure:
The entries of this database are generated from a text mining of published articles. Our database currently contains 356 entries of compounds found from literature. SWMD is designed in MySQL 5.1.36 and PHP 5.3.0. These compounds cover 37 different species of Laurencia and other genera, which is shown is Table 2A, 2B respectively. Geographical origin and extraction method directed for each of these compounds were searched and included in the database along with the biological activity exhibited.

Compounds in SWMD are annotated by molecular property. These include molecular weight, Monoisotopic Mass, Molar Refractivity, number of rotatable bonds, calculated LogP, number of hydrogen-bond donors, number of hydrogen-bond acceptors, Polar Surface Area and Van der Waals surface area. The chemical descriptors and 3D structure for each compound were calculated using Marvinsketch [3] and Chemsketch [4], respectively. Lipophilicity or calculated LogP is predicted using ALOGPS 2.1 program [5]. For molecular visualization, the user needs the free Chime-Plugin from MDL (available for Windows, SGI, Mac) or the Java2 Runtime Environment.

The SWMD database web interface is shown in **Figure 1**. This database is searchable by Accession Number, Compound type, Seaweed Binomial name, IUPAC name, SMILES notation or InChI. The search is case insensitive. In a query, a user can specify full name or any part of the name in a text field. Wild characters of '%' and '_' are supported in text field.
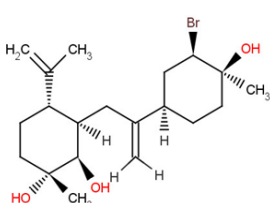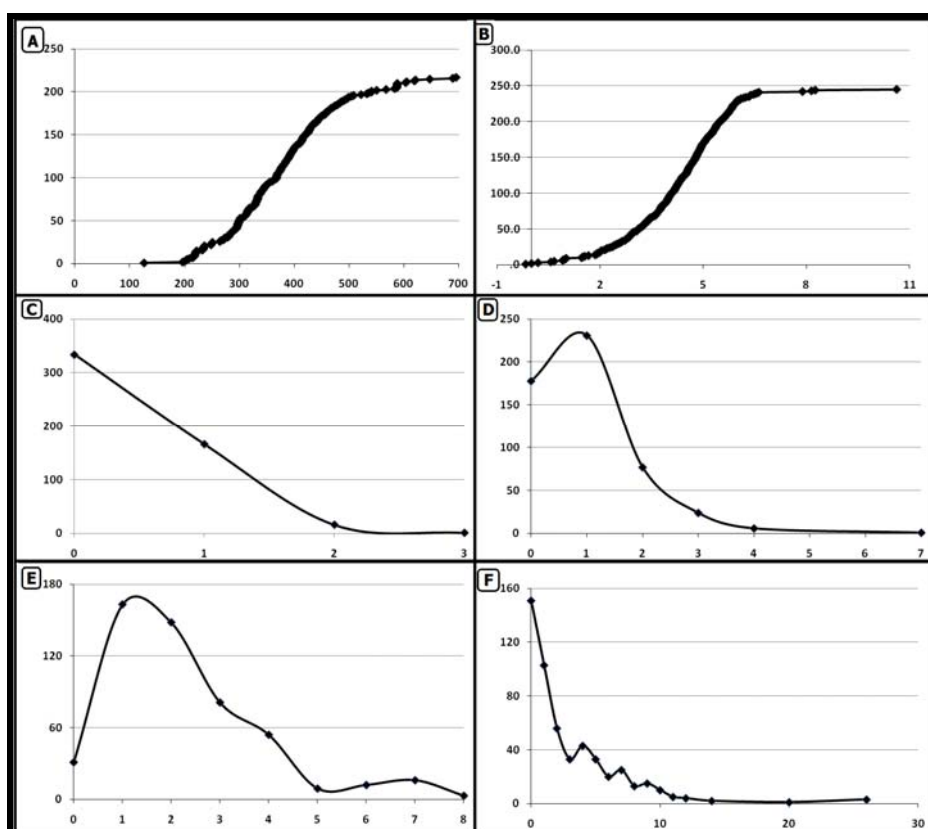
Figure 1: The search result page of SWMD.

Figure 2: Molecular properties of compounds in SWMD. (A) molecular weight; (B) calculated LogP; (C) violations of Lipinski's rule-of-fives; (D) hydrogen-bond donors; (E) hydrogen-bond acceptors and (F) rotatable bonds.

Database features:
SWMD has a web interface at http://www.swmd.co.in. The database is unique in providing comprehensive information of compounds from seaweeds via 25 descriptive fields. Each entry in the database is categorized into sections such as General information, Structure information, Predicted properties and Bibliographic references. The general information part of the database entry displays the compound's unique SWMD accession number viz. XY123 where X represents the Macroalgae - Brown, Green and Red by B, G and R respectively and Y represent the first letter of the genus. It also encompasses compound type, and an external links to the compound's PUBCHEM ID and Chemspider ID (if available) are provided. Binomial name is followed by geographical origin and biological activity which was curated from literature sources. The Structure of the compound, its name in IUPAC, SMILES notation and InChI are displayed in structure information along with atomic coordinates in MOL and PDB format which can be downloaded for 3D molecular visualization [6]. The predicted properties display the pre-computed chemical descriptors of the compounds and reference section lists the citations relevant to the respective compounds with external links to PubMed if available.

SWMD currently contains entries for 517 compounds encompassing 25 descriptive fields mostly from the Red algae of the genus Laurencia (Ceramiales, Rhodomelaceae) **(Table 2 in supplementary material)**. The number of compounds in SWMD is growing, and the numbers reported here should be considered a representative snapshot; see the Web-page for up-to-date statistics. Of these 517 compounds, 331 are Lipinski compliant [7], with the caveat that we have used ALOGPS 2.1 program [5] as a surrogate for c Log P between -3.5 and 5, MW $\leq$ 500 g•mol-1, a maximum of 10 H bond acceptors and 5 H bond donors. Of these, 107 are "lead-like" molecules [8, 9], which have MW = 150-350 g•mol-1, c Log P < 4, H bond donors $\leq$ 3, and H bond acceptors $\leq$ 6. A total of 27 molecules are "fragment-like" [10] with c Log P between − 2 and 3, MW < 250 g•mol-1, H bond donors < 3, H bond acceptors < 6 and rotatable bonds < 3 **(Figure 2).**

Conclusion and Future Perspectives:
The data presented in SWMD can be effectively used for chemo-informatics studies like QSAR analysis, pharmacophore search, molecular docking etc. pertaining to drug discovery. It also portrays the span of secondary metabolites available in seaweeds and the need to preserve the perishing marine ecosystem. The database will be extended with more data on molecular interactions, embedded interactive visualization tools and additional chemical descriptors. The users are also welcome to contribute relevant data to the database via email to authors. The dataset and web interface shall be upgraded periodically.

References:
[1] JS Lazo *et al. Mol Pharmacol*. **72:** 1 (2007) [PMID: 17405872]
[2] W Fenical, *Phytochem*. **15**: 511 (1976)
[3] F Csizmadia, *J Chem Inf Comput Sci*. **40**: 323 (2000) [PMID: 10761134] http://www.chemaxon.com
[4] http://www.acdlabs.com
[5] IV Tetko & VY Tanchuk, *J Chem Inf Comput Sci*. **42**: 1136 (2002) [PMID: 12377001] http://www.vcclab.org/lab/alogps
[6] U Vetrivel *et al. Bioinformation* **4**(2): 71 (2009) [PMID: 20198172]
[7] CA Lipinski, *J Pharmacol Toxicol Methods*. **44**: 235 (2000) [PMID: 11274893]
[8] G Schneider, *Curr Med Chem*. **9**: 2095 (2002) [PMID: 12470249]
[9] TI Oprea, *J Comput-Aided Mol Des,* **16**: 325 (2002) [PMID: 12489682]
[10] ML Verdonk *et al. Proteins* **52**: 609 (2003) [PMID: 12910460]

Supplementary material:

Table 1: Biological Activity of Some Seaweed Metabolites

Compound	Biological activity
Laurinterol	Cytotoxic - K562(IC_{50}=128.3µM); MCF7(IC_{50}=67.2µM); PC3(IC_{50}=76.6µM); HeLa(IC_{50}=83.9µM); A431(IC_{50}=74.6µM); CHO(IC_{50}=165.8µM); NSCLC-N6(IC_{50}=26.5µM)
(+)-α-Isobromo-cuparene	Cytotoxic - HT29(IC_{50}=130.4µM); MCF7(IC_{50}=177.6µM); PC3(IC_{50}=191.2µM); HeLa(IC_{50}=204.3µM); A431(IC_{50}=198.4µM)
Isolaurenisol	Cytotoxic - K562(IC_{50}=127.4µM); MCF7(IC_{50}=95.5µM); PC3(IC_{50}=103.2µM); HeLa(IC_{50}=88.6µM); A431(IC_{50}=122.0µM); CHO(IC_{50}=165.5µM)
Caespitenone	Cytotoxic - HT29(IC_{50}=18.9µM); MCF7(IC_{50}=19.7µM); A431(IC_{50}=21.6µM)
(8R*)-8-bromo-10-epi-β-snyderol	Antimalarial - *Plasmodium falciparum* D6 clones(IC_{50}=2700ng/mL); W2 clones(IC_{50}=4000ng/mL)
Majapolene B	Antibacterial - *Chromobacterium violaceum*(MIC=20µg/disc); *Proteus mirabilis*(MIC=20µg/disc); *Proteus vulgaris*(MIC=20µg/disc); *Erwinia sp*(MIC=10µg/disc); *Vibrio parahaemolyticus*(MIC=20µg/disc); *Vibrio alginolyticus*(MIC=20µg/disc);
Laurenditerpenol	Inhibits hypoxia-activated (hypoxia-inducible factor-1) HIF-1 (IC_{50}=0.4µM) and hypoxia-induced VEGF (a potent angiogenic factor) in T47D cells

Table 2: Marine algae and *Laurencia species* listed in SWMD with number of entries

Marine algae	number	Laurencia species	number
Boergeseniella fruticulosa	1	*Laurencia aldingensis*	4
Corallina granifera	1	*Laurencia caduciramulosa*	5
Cutleria multifida	1	*Laurencia calliclada*	1
Cystoseira mediterranea	1	*Laurencia cartilaginea*	4
Dictyota dichotoma	4	*Laurencia catarinensis*	14
Enteromorpha compressa	1	*Laurencia claviformis*	1
Galaxaura marginata	14	*Laurencia composita*	12
Gelidium crinale	1	*Laurencia decumbens*	14
Halymenia floresii	1	*Laurencia glandulifera*	12
Hypnea musciformis	1	*Laurencia intermedia*	3
Jania Rubens	7	*Laurencia intricata*	5
Laurencia	465	*Laurencia karlae*	6
Padina pavonica	1	*Laurencia luzonensis*	21
Phyllophora crispa	1	*Laurencia majuscula*	38
Polysiphonia morrowii	1	*Laurencia mariannensis*	21
Sphaerococcus coronopifolius	9	*Laurencia microcladia*	25
Sporochnus pedunculatus	1	*Laurencia nidifica*	2
Taonia atomaria	1	*Laurencia nipponica*	12
Undaria pinnatifida	5	*Laurencia obtusa*	95
TOTAL	517	*Laurencia okamurai*	7
		Laurencia omaezakiana	4
		Laurencia paniculata	1
		Laurencia pannosa	3
		Laurencia papillosa	1
		Laurencia perforata	3
		Laurencia saitoi	15
		Laurencia scoparia	19
		Laurencia similis	28
		Laurencia snyderiae	2
		Laurencia sp.	42
		Laurencia subopposita	12
		Laurencia thyrsifera	1
		Laurencia tristicha	14
		Laurencia undulata	1
		Laurencia venusta	3
		Laurencia viridis	12
		Laurencia yonaguniensis	2
		TOTAL	465