ACToR Chemical Structure processing using Open Source ChemInformatics Libraries

Jayaram Kancherla[1], Kamel Mansouri[1], Hao Truong[1], Ann M. Richard[1], Richard S. Judson[1]

(1) U.S. Environmental Protection Agency, Office of Research and Development, National Center for Computational Toxicology, Research Triangle Park NC, USA,

Jayaram Kancherla | Kancherla.Jayaram@epa.gov | 919-541-3089

www.epa.gov

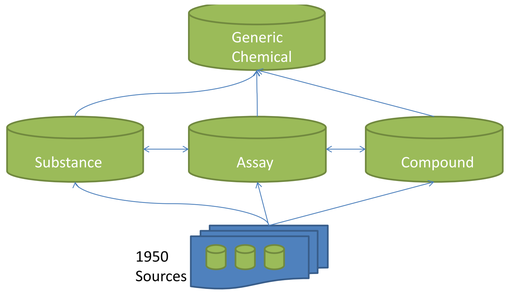
Introduction

ACToR (Aggregated Computational Toxicology Resource) is a centralized database repository developed by the National Center for Computational Toxicology (NCCT) at US EPA. Open source tools were used to compile toxicity data from over 1950 public sources.

ACToR contains chemical structure information and toxicological data for over 558,000 unique chemicals. The database primarily includes data from NCCT research programs e.g., **ToxRef** (in vivo toxicity data), **ExpoCast** (human exposure data), **ToxCast** (high-throughput screening data) and **DSSTox** (chemical inventory for 16000 unique structures). Included are also data from PubChem, ChemSpider, USDA, FDA, NIH and other public international data sources.

Organizing and managing this huge collection of data and improving the chemical structure quality has posed major challenges.

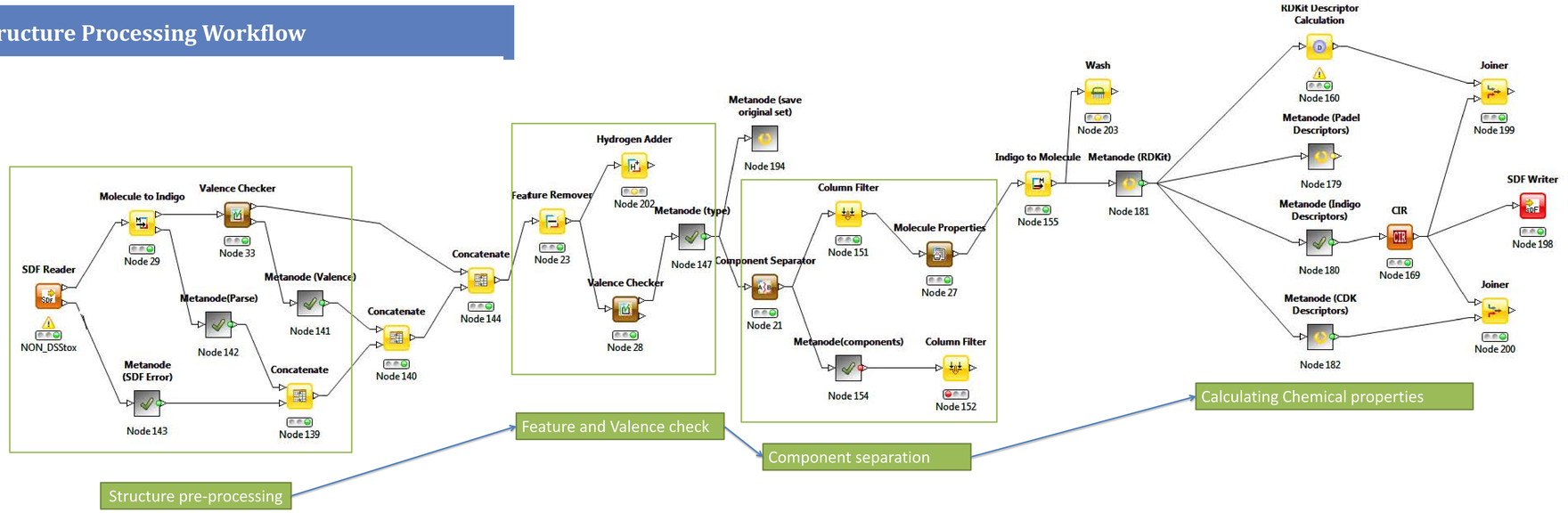
ACToR Database



Issues and Problems

- Valence errors , incomplete and parsing issues with structure files
- No **One to One Relationship** between CAS, Structure and Name
- Missing/issues with chemical properties from different sources.

Structure Processing Workflow



Structure <-> Name <-> CAS relationship

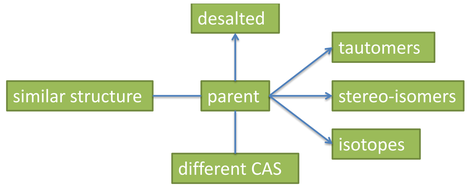
Source	Structure	Name	CAS
EPA DSSTox	YES	YES	YES
EPA SRS	NO	YES	YES
EPA HPV TSCA	YES	YES	YES
PubChem	YES	NO*	NO*
ChemSpider	YES	NO*	NO*
	... for all 1950 sources		

* One to One relationship between CAS, Structure and Name does not exist

ACToR Statistics

Category	Count
Data Collections	1940
Substances	2,850,553
Compounds	1,259,087
Generic Chemicals	558,063
Generic Chemicals (with structure)	460,297
Assays	3683
Assay Results	41,769,608

Structure Relationships



Future Plans

- Build a QSAR ready version of all the compounds available in ACToR
- Identify different structural relationships between compounds