

# CHEMISTRY COMPONENT COLLECTION

The Chemistry Collection offers a comprehensive suite of molecular property calculators, filters, and manipulators. This collection of modular components extends the standard capabilities of Pipeline Pilot to include compound processing and cheminformatics analyses. With these enhancements, you can create protocols for a broad array of applications including:

## WITH THE CHEMISTRY COMPONENT COLLECTION YOU CAN:

### Compound library acquisition:

- Library cleanup and standardization
- Multiple library comparisons
- Substructure and similarity searching
- Extensive property profiling and subset selection

### Combinatorial library design:

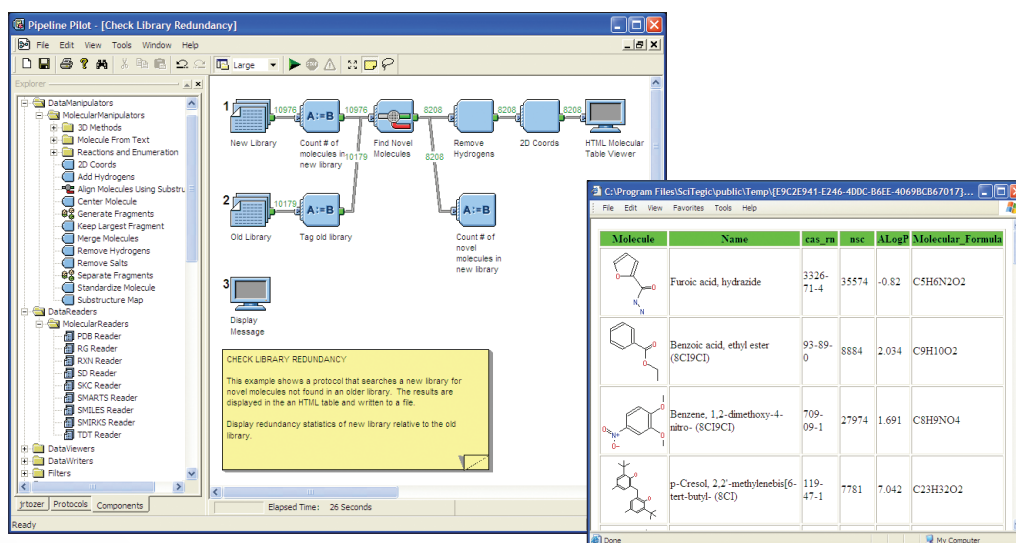
- Scaffold and reaction driven enumeration

## Combined with the separately available Modeling Collection:

- Structure activity modeling
- Compound clustering
- Maximal common substructure searching

## READERS AND WRITERS





The Chemistry Collection includes file readers and writers for most industry-standard chemistry formats. It includes SD, RG, and RXN (from MDL), SMILES, SMIRKS, SMARTS, and TDT (from Daylight),



**CHECK LIBRARY REDUNDANCY**

This example shows a protocol that searches a new library for novel molecules not found in an older library. The results are displayed in the an HTML table and written to a file.

Display redundancy statistics of new library relative to the old library.

Molecule	Name	cas_rn	usc	ALogP	Molecular_Formula
	Furoic acid, hydrazide	3326-71-4	35574	-0.82	CSH6N2O2
	Benzoic acid, ethyl ester (SC19CT)	93-89-0	8884	2.034	C9H10O2
	Benzene, 1,2-dimethoxy-4-nitro- (SC19CT)	709-09-1	27974	1.691	CSH9NO4
	p-Cresol, 2,2-methylenebis[6-tert-butyl- (SC1)	119-47-1	7781	7.042	C23H32O2

MOL2 (from Tripos) and public formats such as PDB. You can also read molecular data from databases such as ISIS™ and MDL Direct (from MDL), ActivityBase (from IDBS), and Accord (from Accelrys).

## VIEWERS

In addition to Pipeline Pilot's standard viewers, the Chemistry Collection seamlessly integrates with several popular structural viewers including DS ViewerPro and MDL's ISIS for Excel. Furthermore, a variety of Web viewers are available providing a way to graphically display your molecular data.

## MANIPULATORS

The Chemistry Collection includes manipulators that modify structures according to their salts, tautomers, stereochemistry, charges. You can normalize sets of molecules before comparing them, apply chemical reactions, merge collections of molecules and even enumerate combinatorial libraries. For optimal display characteristics, you can employ manipulators for structural alignment and for fast, high quality 2D and 3D layout.

## FILTERS

Numerous filter components are provided to separate molecules into groups based on chemical properties or calculations. These filters range from simple property value thresholds, like those in the Lipinski Filter, to more sophisticated criteria, such as occurrences of substructures or duplicates. Filters are available to select representative subsets based on diversity or similarity. An interactive filter allows you to select a subset of molecules for further processing. What's more, you can create custom chemistry filters for any property values that you require.

## PROPERTY CALCULATORS

The Chemistry Collection includes a variety of fast molecular property calculators. You can characterize hundred or thousands of molecules per second. Whole molecule properties include AlogP, logD, pKa and solubility, and molecular weight. Other properties include electrotopological indices, solvent accessible surface area and volume descriptors and a large variety of molecular property counts.

## STRUCTURAL FINGERPRINTS

A proprietary method for calculating structural fingerprints is included in the Chemistry Collection. Known as Extended Connectivity Fingerprints (ECFP), this method offers excellent characterization of molecules, indexing the environments of every atom in a molecule using up to four billion different structural features. It is an extremely rapid method useful in searching, clustering, and modeling applications. When used with Bayesian learning technologies in the separately available Modeling Collection, ECFP provides interpretable predictive models for large data sets. You can also use Daylight, MDL or user defined keys within the system.

## MOLECULAR TOOLKIT

When combined with the Integration Collection, the Molecular Toolkit's Java and Perl API's provide programmatic access to the molecular data model's searching methods.

To learn more about Pipeline Pilot, go to [accelrys.com/pipeline-pilot](http://accelrys.com/pipeline-pilot)