



## NSisToolkit

---

Noesis flagship product is the **NSisToolkit** consisting of two software suites, the **NSisUtilities** Suite and the **NSisApps** Suite. **NSisUtilities** provides robust, efficient, flexible, fundamental chemoinformatics utilities to the expert user. The utilities include chemical structure and property databases, fragmentation and substructure mining tools, a compound selection module and a chemical dataset clustering utility.

## NSisUtilities Suite

---

### Utilities/Modules

NSisChemDB – A chemical structure database module

NSisBioDB – A molecular property database module

NSisFragment – A fragmentation and substructure mining module

NSisProfile – A fragment characterization and profiling module

NSisOrganize – A chemical data clustering module

NSisSelect – A compound selection tool

## Application Domains

---

Noesis' software can be of use to research organizations from the pharmaceutical, agrochemical and biotechnology fields that are actively engaged in the quest for discovery of new small molecules with desirable biological properties. Noesis decision support system can also be of use to enterprises from the cosmetics and food supplements fields which regularly perform investigations to detect molecules with e.g. specific odor, skin effect, favorable taste or minimal environmental toxicity effects.



# NSisFragment

---

A molecular fragmentation and substructure mining module

## **Description**

A powerful molecular fragmentation module implementing a variety of methods designed to extract substructures from molecules. The module accepts chemical datasets in sd and smiles format, as well as molecules already stored in a NSisChemDB. The fragments produced can be stored using a specially designed NSisChemDB utility or placed in ascii files for subsequent manipulation by the user. The fragments produced are annotated with information about their attachment points and the type of substituents connected at each attachment point (ongoing). The output of the NSisFragment module can be supplied as input to the NSisProfile module for fragment profiling according to biological and chemical properties.

## **Synergies**

NSisChemDB, NSisProfile, NSisOrganize (internal)

## **Fragmentation Methods\***

RECAP, Molecular Frameworks, Ring Systems, Linear Substructures, Maximum Common Substructures

## **Interface**

NSisFragment has a highly flexible command line driven interface. Users can indicate preferences via a plain ascii text parameter file conforming to the NSisFragment parameter specifications.

## **Performance**

The performance of NSisFragment depends on the number and type of fragmentation methods selected and the size and complexity of the chemical dataset. Indicatively, a set of 5,000 compounds can be fragmented with the RECAP and molecular frameworks methods in 2:13 minutes on a Windows XP PC with an Intel Core 2 Duo 3 GHz processor. The much larger dataset of 2 million compounds can be processed with the same fragmentation methods on the same PC in less than 18 hours.

\* References: [For a detailed list of references check Noesis website](#)



## **NSisOrganize**

---

A chemical data clustering module

### **Description**

A graph-based chemical dataset clustering module that ensures that each cluster produced will contain a common core substructure, the scaffold, characteristic of the compounds in the cluster. The module accepts chemical datasets in sd and smiles format, as well as molecules already stored in a NSisChemDB. The clusters produced are placed in ascii files for subsequent manipulation by the user. The information about each cluster includes an ID, the scaffold, the compound IDs containing the scaffold, the MCS of the compounds in the cluster (optional), and a cluster tightness measure.

### **Synergies**

NSisChemDB, NSisFragment (internal)

### **Clustering Method\***

Approximate Scaffold-based Technique

### **Interface**

NSisOrganize has a highly flexible command line driven interface. Users can indicate their preferences through a parameter file written in ascii text format and conforming to the NSisOrganize parameter naming specifications.

### **Performance**

The performance of NSisOrganize depends mainly on the specific settings of the approximate scaffold generation method used and the size and complexity of the chemical dataset. Indicatively, a set of 5,000 can be clustered with the ring-system method in 3:35 minutes on a Windows XP PC with an Intel Core 2 Duo 3 GHz processor. A larger dataset of 220,000 compounds can be processed in about 6:45 hours with the same clustering settings and on the same PC.

\* [References: For a detailed list of references check Noesis website](#)



## **NSisSelect**

---

A compound selection module.

### **Description**

A fast molecule selection module enabling users to implement molecular filters to pick compound subsets from much larger datasets. The module allows the implementation of three types of filters: substructure-based filters, calculable molecular property filters and filters utilizing compound properties as stored in the NSisBioDB. Any number of filters can be implemented. The module accepts chemical datasets in sd and smiles format, as well as molecules already stored in a NSisChemDB. The compounds selected can be stored in a NSisChemDB table or simply placed in ascii files. A detailed collection of log files records information related to the rejection of compounds by the module according to the filters set by the user.

### **Synergies**

NSisChemDB, NSisBioDB.

### **Selection Methods / Filter Types**

- Substructure filters: Collection of valid SMARTS queries
- Calculable property filters: Hydrogen-bond donors/acceptors, molecular weight, flexibility, complexity, number of rings.
- Stored property filters: Properties available in NSisBioDB.

### **Interface**

NSisSelect has a simple, easy to set, command line driven interface. Users can indicate their preferences through a parameter file written in ascii text format and conforming to the NSisSelect parameter naming specifications.

### **Performance**

The performance of NSisSelect depends on the number and type of filters implemented and the size and complexity of the chemical dataset searched. In internal tests a set of 1 million compounds is processed with the Rule-of-Five and a substructure filter in about 3 hours on a Windows XP PC with an Intel Core 2 Duo E8400 @ 3 GHz