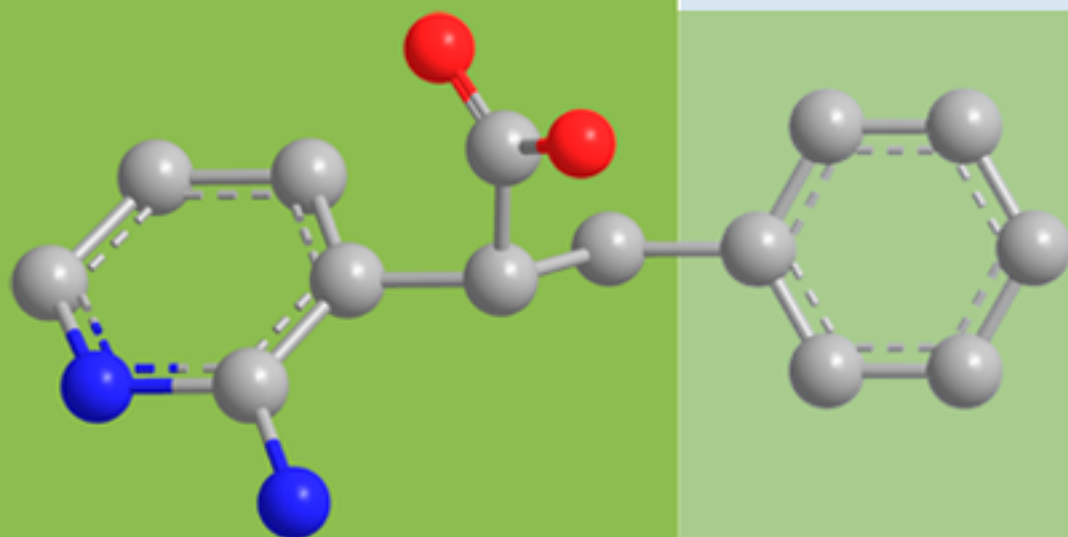
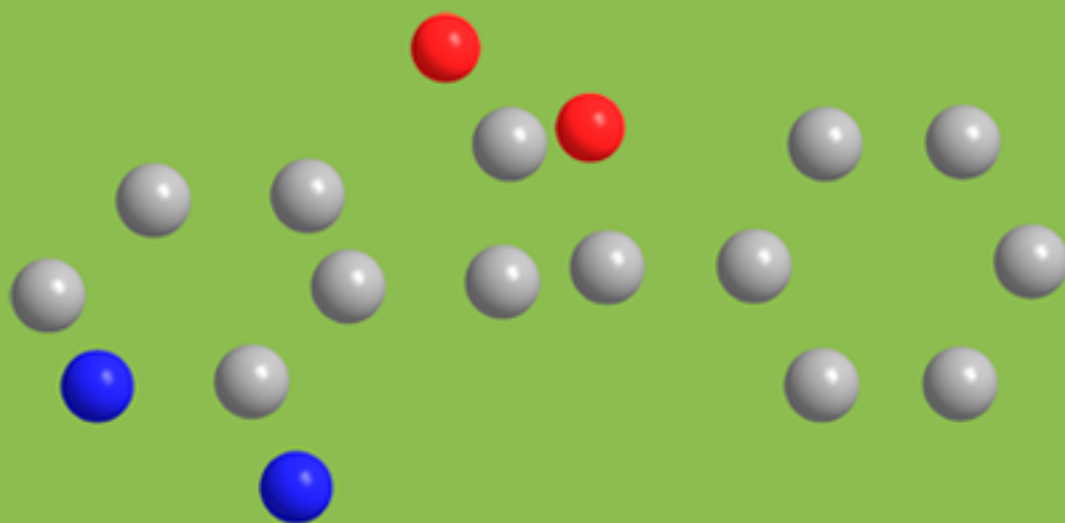




# I-interpret User Manual



**Version 1.0**

---

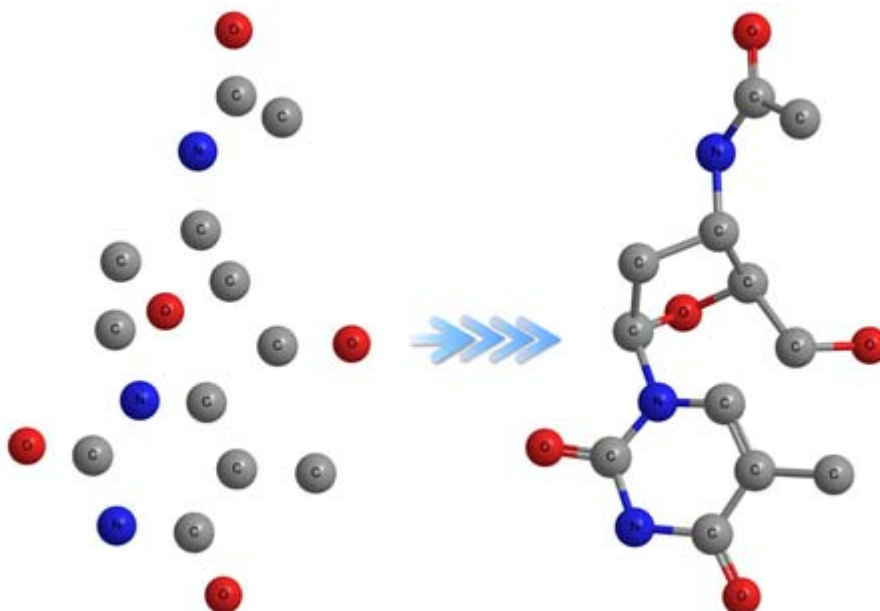
## Table of Contents

---

1. Overview of I-interpret.....	1
2. System Requirements and Installation.....	3
2.1 System requirements.....	3
2.2 Installation.....	3
3. How to Use I-interpret.....	4
3.1 General synopsis for running I-interpret.....	4
3.2 An example of running I-interpret.....	5
4. Parameters of I-interpret.....	6
4.1 Key Parameters.....	7
4.2 Optional Parameters.....	9
4.3 Implicit Parameters.....	10
5. On-line Demo of I-interpret.....	10
Appendix A: Chemical Groups in Neutral and Charged Forms.....	12
Appendix B: Examples of Control Parameters.....	13
Contact Information.....	15

## 1. Overview of I-interpret

I-interpret is an efficient “interpreter” of chemical structures, which can automatically interpret the chemical structure of a given organic molecule merely from its essential structural information, including element identities and three-dimensional coordinates of its component atoms (see Figure 1). I-interpret was tested on thousands of small organic molecules and achieved a high success rate in interpreting their correct chemical structures. It can serve as a valuable tool for processing chemical structures in various molecular modeling studies, especially for high-throughput projects such as processing large databases of organic molecules.



**Figure 1** Interpreting the chemical structure of an organic molecule from a nude model.

The currently released of I-interpret is version 1.0. Besides a high success rate in interpreting chemical structures, I-interpret provides users a number of optional functions, which makes it more flexible and powerful for practical uses.

- **Multiple Supported File Formats.** I-interpret can process a variety of chemical file formats for structure input and output (e.g. PDB, MDL SD/RD/MOL File, SYBYL MOL2). These particular formats are accepted by many molecular modeling programs.

- **Able to Utilize Additional Structural Information.** In addition to atomic identities and coordinates, sometimes a connection table is also available from input (such as MDL SD/RD/MOL format), providing explicit information of each bond in a given molecule. If user decides to trust such information from the input, I-interpret can use it as well in processing a given molecule.
- **Adjustable Geometrical Parameters.** I-interpret uses a number of cutoffs of bond angles and dihedral angles in its algorithm. All of these cutoffs can be adjusted by users through an external parameter file (see section 4).
- **Able to Process Metal-Containing Organic Molecules.** I-interpret is able to handle metal-containing molecules that few format-converting programs can do. For the sake of convenience, all coordinate bonds formed between metal atoms and organic molecules are labeled as single bonds in its outputs.
- **Able to Fill Up Hydrogen Atoms.** I-interpret is designed to interpret the chemical structure of a given molecule correctly even if hydrogen atoms are not provided implicitly. Upon user's choice, I-interpret can fill up the hydrogen atoms on an organic molecule using standard bond lengths and angles.
- **Adjustable Protonation States.** User is allowed to specify the protonation states of certain chemical groups in the final outputs.
- **Output the largest substructure.** Sometimes an input file may contain some solvent molecules (e.g. water) that users don't care about. These small molecules can be excluded from the interpreting result by outputting the largest substructure.
- **Statistics of Structural Information.** Besides the interpreting results, I-interpret can also output some structural information according to user's request.
- **Able to Process Multiple Molecules in a Single File.** For chemical databases, multiple molecules are written in a single file. I-interpret has ability to process such databases of several hundreds of thousand or even millions of compounds.
- **Error Diagnosis System.** When certain errors are found, I-interpret will records their locations and types to its log file, which will help users to find the reasons and solutions on failed molecules.

## 2. System Requirements and Installation

### 2.1 System requirements

I-interpret is a command line oriented program and has to be executed at a Windows or DOS command prompt or in a shell (e.g., csh, tcsh, or bash) on Linux systems. The following hardware platforms and operating systems are supported:

- x86 platforms running Microsoft Windows® 98/2000/XP (**.NET Framework 2.0** is required. You can download it from <http://www.microsoft.com/downloads/> )
- x86 platforms (32-bit and 64-bit processors) running Linux®, Kernel 2.4 or above.

### 2.2 Installation

The I-interpret program is written in C++ language and has been tested on Windows and Linux platforms. It is provided as executable binary codes for each supported platform. To install the command line version of I-interpret please follow the instructions below.

**1)** Download the program package (<http://www.sioc-ccbq.ac.cn/software/I-interpret/>) for appropriate operating system.

- a) For windows systems, a WinRar archive is provided.
- b) For Linux systems, a Gzip archive is provided.

**2)** Move the package to the directory where you would like the program to be installed and uncompress it. You will get a directory named as [I-interpret/](#), under which there are several subdirectories:

<a href="#">bin/</a>	executable binary file
<a href="#">manual/</a>	user manual in PDF format
<a href="#">example/</a>	examples for practicing I-interpret
<a href="#">parameter/</a>	containing a default parameter file

**3)** Set the environment variables

After step 2, you can run I-interpret in its installation location. However, in order to execute I-interpret from any other directory, you need to configure a few environment variables.

a) In Windows systems, please add the path where I-interpret resides to user variables of your system settings. The user variables (*PATH*) can be accessed from “[Control Panel](#) → [Performance and Maintenance](#) → [System](#) → [Advanced](#) → [Environment Variables](#)”.

b) In Linux systems, if you are using CSH or TCSH, please add the following lines to the ".cshrc" file (or ".login", ".profile", whichever you are more used to) under your login directory:

```
setenv I-INTERPRET_HOME the_installation_directory_of_I-interpret
setenv I-INTERPRET_BIN $I-INTERPRET_HOME/bin
set path = ($path $ I-INTERPRET_BIN)
```

If you are using other types of shell, please add the equivalent contents to your configuration file.

---

## 3. How to Use I-interpret

### 3.1 General synopsis for running I-interpret

The current release of I-interpret can be run at Windows or DOS command prompt or in a shell (e.g., csh, tcsh, or bash on Linux systems). The executable files on MS Windows platforms and Linux systems are **I-interpret.exe** and **I-interpret** respectively.

The general synopsis for running I-interpret is:

**I-interpret [-flag] [input] [output]**

The input and output are the names of input and output files. For convenience, I-interpret

provides several options ([*-flag*]) to help users set a series of parameters at one time. Some of these options can also be specified in the external parameter file (see section 4.3).

- h** Print general synopsis of I-interpret to the screen.
  
- p** Create a parameter file containing parameters with default values. If input file name is given, it will use default parameters to interpret the chemical structure.
  
- c** Assign atom and bond types in default charged forms for some functional groups. It makes the carboxylic, sulfonic, and phosphoric acid groups to be deprotonated; while aliphatic amine groups, guanidine and amidine groups are supposed to be protonated. (see *Appendix A*).
  
- n** Assign atom and bond types in neutral forms for some functional groups. It makes the carboxylic, sulfonic, and phosphoric acid groups to be protonated; while aliphatic amine groups, guanidine and amidine groups are supposed to be deprotonated. (see *Appendix A*).

Note: The flag at the command line is preferred over the external parameter file. Only if the flag is not given, I-interpret will search the parameter file in your working directory. You can also specify “-c” or “-n” flag in the parameter file.

The current release of I-interpret can accept PDB, MDL SD/RD/MOL and SYBYL MOL2 format files as valid inputs. The output formats supported by I-interpret include PDB, SD and MOL2. There is no need to specify the input and output format as I-interpret can automatically identify them from their extensions. If the output file name was not specified, e.g. **I-interpret example.pdb**, I-interpret will use the base name of the input file (**example**) and default output format (**MOL2**) instead, so the output file name will be **example.mol2**.

### 3.2 An example of running I-interpret

The example file *example.pdb* provided with the distribution contains the structural information of three molecules in PDB format. We'll take it as an example to run I-interpret. Please copy this file into your working directory and type the following command at the command line prompt:

### I-interpret example.pdb result.mol2

I-interpret will generate a default parameter file and output *result.mol2*. Furthermore, a log file named *I-interpret.log* will be created in your working directory which records information on the I-interpret run, such as computation time, total number of processed molecules, total number of successful and problematic molecules and warning messages (figure 2).

```

=====
| II      ii      tt                                tt |
| II      tt                                tt |
| II      ii nnnnnn ttttt eeeeeee rr rr pppppp rr rr eeeeeee ttttt |
| II **** ii nn  nn  tt  ee   ee rrr  pp  pp rrr  ee   ee tt |
| II      ii nn  nn  tt  eeeeeeee rr  pp  pp rr  eeeeeeee tt |
| II      ii nn  nn  tt  ee    rr  pp  pp rr  ee    tt |
| II      ii nn  nn  tttt eeeeeee rr  pppppp rr  eeeeeee tttt |
|                                     pp |
|                                     pp |
|                                     version 1.0 |
=====

```

```

Interpreting finished at: Fri Jul 20 16:40:02 2007
Total number of molecules: 3
3 molecules have been interpreted.
Total elapsed time: 15 ms

```

**Figure 2** An example of the log file created by I-interpret

If you want to assign atom and bond types in charged forms for the chemical groups in *Appendix A*, you can do:

### I-interpret -c example.pdb result.mol2

The parameters in the "*parameter.txt*" will be changed for "-c" flag. It is to note that in the interpreting result the terminal amine group in the first molecule and the carboxyl group in the third molecule are assigned in their charged forms.

## 4. Parameters of I-interpret

The flag in the command line is helpful in setting a group of parameters for special usage. However, you can also adjust each of the parameters to meet your own purpose. An example of default parameter file is given in the "*parameter/*" subdirectory (see figure 3).



```
#####  
#           Parameters for I-interpret           #  
#####  
#### Geometry Parameters for Determine  
#### Initial Hybridization State  
      SP1_CUTOFF_ANGLE = 155.0  
      SP2_CUTOFF_ANGLE = 115.0  
      TORSION_RING_FIVE = 7.5  
      TORSION_RING_SIX = 15.0  
      FLAT_TORSION_ANGLE = 30.0  
  
#### Parameters for Controlling Output  
#### of Certain Functional Groups  
      CHARGED_CARBOXYL = NO  
      ALIPHATIC_NITROGEN = NO  
      CHARGED_GUANIDINIUM = NO  
      ODD_RING_AROMATICITY = YES  
      CONVERT_DATIVE_BONDS = NO  
  
#### Parameters for Optional Functions  
      LARGEST_SUBSTRUCTURE = NO  
      ADD_HYDROGEN_ATOMS = NO  
      HEAVY_ATOM_STATISTICS = NO
```

**Figure 3** Default parameter file of I-interpret

Note that all of the lines started with a "#" sign in the parameter file are considered as comments and therefore are neglected by the program. You can disable a certain parameter by putting a "#" sign at the beginning of that line. You can also re-enable it by removing the "#" sign.

In the parameter file, there are three types of parameters ---- a) key parameters; b) optional parameters and c) implicit parameters. All of them will be explained below.

## 4.1 Key Parameters

There are ten key parameters, five of which are related with geometry and the others are utilized to control the output for certain functional groups. These parameters should be explicitly written in the parameter file. If one of them is missing, I-interpret will create a default parameter file to overwrite the current one.

Geometrical Parameters	Type	Default	Description
<a href="#">SP1_CUTOFF_ANGLE</a>	Float	155	Atoms as bond angle centers greater than this value are typed as <i>sp</i> hybridized.
<a href="#">SP2_CUTOFF_ANGLE</a>	Float	115	Atoms except for <i>sp</i> -hybridized that have an average bond angle greater than this value are typed as <i>sp</i> <sup>2</sup> hybridized.
<a href="#">TORSION_RING_FIVE</a>	Float	7.5	A five-member or smaller ring will be identified as planar if its average in-ring torsion angle is lower than this value.
<a href="#">TORSION_RING_SIX</a>	Float	15	A six-member or larger ring will be identified as planar if its average in-ring torsion angle is lower than this value.
<a href="#">FLAT_TORSION_ANGLE</a>	Float	30	If a bond has a dihedral angle larger than this value, it will be considered as single bond. It constrain the atoms involved in a double bond to be in a plane
<p>Note: These parameters only affect the determination of initial hybridization states. In the other interpreting steps, the hybridization state may be rectified. If there are no significant geometrical errors in the input structure, the default values are recommended.</p>			

Controlling Parameters	Values	Default	Description
<a href="#">CHARGED_CARBOXYL</a>	YES or NO	NO	Whether to output carboxylic, sulfonic, and phosphoric acid groups as deprotonated.
<a href="#">ALIPHATIC_NITROGEN</a>	YES or NO	NO	Whether to output aliphatic amine groups as protonated.
<a href="#">CHARGED_GUANIDINIUM</a>	YES or NO	NO	Whether to output guanidine and amidine groups as protonated.

	NO		
<a href="#">ODD_RING_AROMATICITY</a>	YES or NO	YES	Whether to assign aromaticity on odd-number rings if applicable.
<a href="#">CONVERT_DATIVE_BONDS</a>	YES or NO	NO	Whether to convert dative bonds (e.g., N-O bond in nitro group, N-N bond in azide group).
Note: Some examples related to these parameters are shown in <i>Appendix B: Examples of Controlling Parameters</i> .			

## 4.2 Optional Parameters

In the parameter file, there are three optional parameters that are designed to control some optional functions. Under default condition, these functions are disabled. If you want to use these functions, just change the corresponding parameter to “YES”.

Optional Parameters	Values	Default	Description
<a href="#">LARGEST_SUBSTRUCTURE</a>	YES or NO	NO	If a molecule contains more than one substructures, enable this parameter will force I-interpret to output the substructure with the most heavy atoms.
<a href="#">ADD_HYDROGEN_ATOMS</a>	YES or NO	NO	Enable this parameter will force I-interpret to fill up the hydrogen atoms on the organic molecule using standard bond lengths and angles.
<a href="#">HEAVY_ATOM_STATISTICS</a>	YES or NO	NO	A file named as “statistics.txt” will be created together with the interpreting result as long as you enable this parameter. The number of heavy atoms in each chemical structure as well as its name is

			written in this file.
--	--	--	-----------------------

### 4.3 Implicit Parameters


The following parameters are referred to as implicit parameters because they are not written in the default parameter file. You need to manually add them to the parameter file if you want to use them. These parameters are mutually exclusive, so you can only enable one of them at a time.

Implicit Parameters	Values	Default	Description
NEUTRAL_FORM	YES or NO	----	Output chemical structures in neutral form. It has the same effect as type “-n” option in the command line.
CHARGED_FORM	YES or NO	----	Output chemical structures in charged form. It has the same effect as type “-c” option in the command line.

## 5. On-line Demo of I-interpret

We provide a web-based demo of I-interpret (Figure 4) for testing at: <http://www.sioc-ccbg.ac.cn/software/I-interpret/>. You can upload your chemical structures and set the parameters on the web. After submitting, a page with the information of input and output files will pop up, from which you can download the interpreting result.

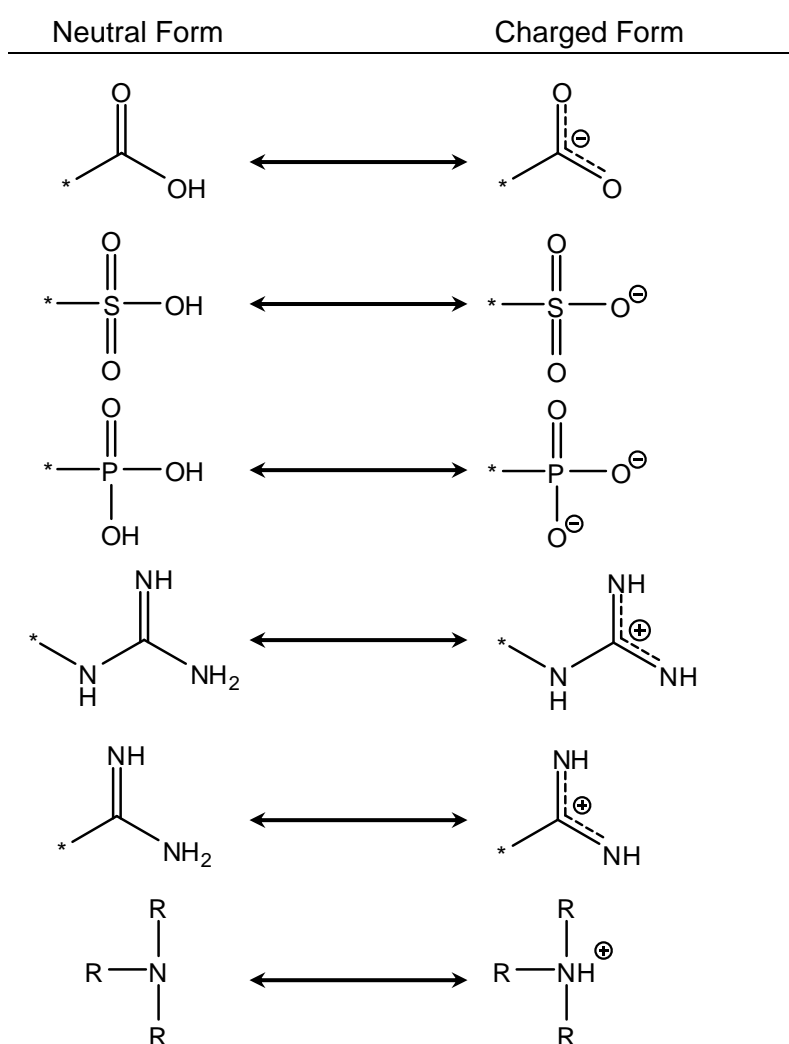
## I-interpret Online

input file name:	<input type="text"/>	<input type="button" value="浏览..."/>
input file format:	PDB <input type="button" value="v"/>	
output file name:	<input type="text"/>	
output file format:	MOL2 <input type="button" value="v"/>	
<b>Parameters</b>		
Protonation state <input checked="" type="radio"/> Neutral form <input type="radio"/> Charged form		
add hydrogens <input type="radio"/> Yes <input checked="" type="radio"/> No		
output largest substructure <input type="radio"/> Yes <input checked="" type="radio"/> No		
<input type="button" value="Submit"/>		
		
I-interpret version: v1.0. <a href="#">Latest update on 26/6/2007.</a>		

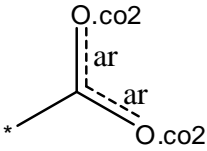
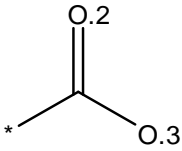
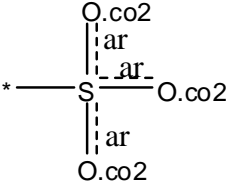
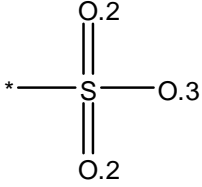
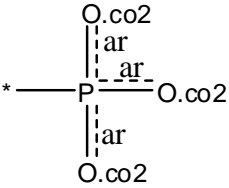
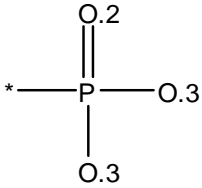
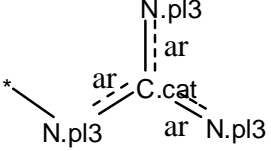
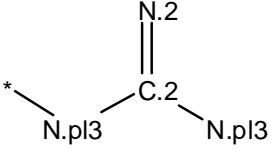
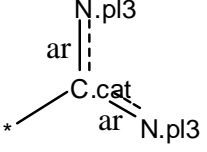
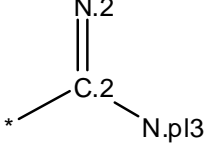
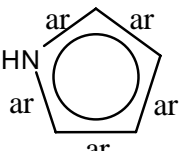
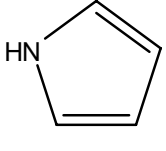
**Figure 4** Web-based demo of I-interpret

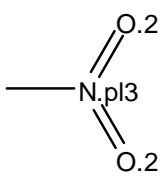
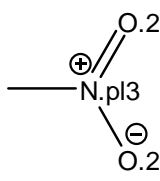
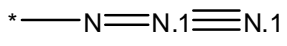
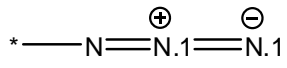
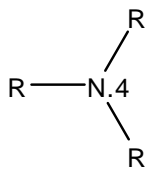
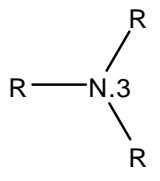
## Appendix A: Chemical Groups in Neutral and Charged Forms

The following groups are controlled by the flags in the command line or the implicit parameter in the external parameter file. The flag “-n” will result in the neutral forms of these groups (equal to “NEUTRAL\_FORM = YES” in the parameter file) in the output, while the flag “-c” will convert them into charged forms (equal to “CHARGED\_FORM = YES” in the parameter file).



## Appendix B: Examples of Control Parameters

Controlling Parameters	Examples	
	Value = YES	Value = NO
CHARGED_CARBOXYL		
		
		
CHARGED_GUANIDINIUM		
		
ODD_RING_AROMATICITY		

CONVERT_DATIVE_BONDS		
		
ALIPHATIC_NITROGEN		
A detailed description of SYBYL MOL2 file formats is available as a PDF document at <a href="http://www.tripos.com">http://www.tripos.com</a> .		



## Contact Information

If you have any technical concerns on this program, the primary way for getting responses is to register on our forum at <http://www.sioc-ccbq.ac.cn/forum/> and log into the special section created for I-interpret. Alternatively, you may contact us as following:

*Prof. Renxiao Wang*

*State Key Lab of Bio-organic and Natural Products Chemistry*

*Shanghai Institute of Organic Chemistry (SIOC), Chinese Academy of Sciences.*

*354 Fenglin Road, Shanghai, China.*

*Email: [wangrx@mail.sioc.ac.cn](mailto:wangrx@mail.sioc.ac.cn)*

Any suggestion or comment on this program will be highly appreciated.

## Reference

1. Zhao, Y.; Cheng, T. J.; Wang, R. X. Automatic Perception of Organic Molecules Based on Essential Structural Information. *J. Chem. Inf. Model.* **2007**, *47*, 1379-1385