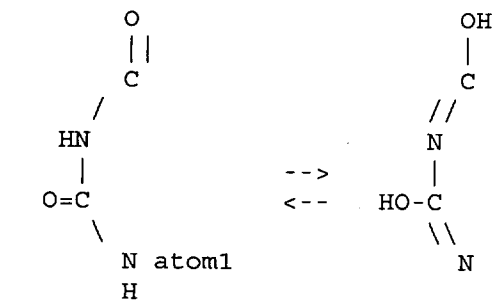
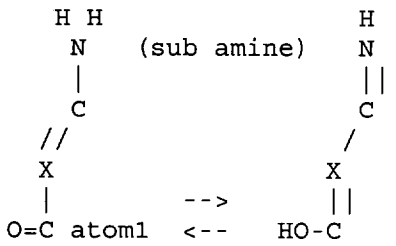
```
*****************************************************************************
*                            ~~~~~~~~~~
*                            PERCEPTION
*                            ~~~~~~~~~~
*Knowledge Base Name     :tautomerism.kbase
*Description             :A knowledge base to identify molecules that undergo
*                         tautomerism and generate new tautomers.
*Created by              :A.K.T. Ting
*Date of creation        :16-7-1997
*Last modified           :9-2-99
*Version                 :2.3
*****************************************************************************


Information :
~~~~~~~~~~~~~

New CAESA Rule Base substitution types :
      add-taut-hydrogen

Tautomerism rules are labelled by number
Explanation gives the type of tautomerism the rule identifies
Diagrams above rules show an example of the application of that tautomerism
rule.

This knowledge base is not comprehensive but serves to illustrate the rules.
```

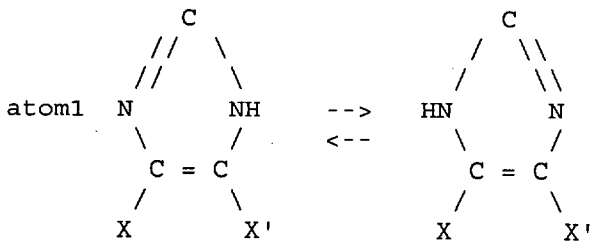



```
CHEMICAL-LABEL <tautomerism rule 1>
...STARTP
...N[HS>0];[ARYL=YES]-C[ARYL=YES](=O)-N[HS>0];[ARYL=YES]-C[ARYL=YES](=O)
...ENDP

RULE
EXPLANATION Extended Keto/Enol system.
IF  tautomerism rule 1
THEN substitute-bond 1 with =
        substitute-bond 2 with -
        substitute-bond 4 with =
        substitute-bond 5 with -
        add-taut-hydrogen 3
        add-taut-hydrogen 6
END-THEN
```



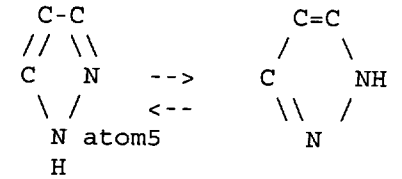
Enaminoketone/enolimine tautomerism

```
CHEMICAL-LABEL <tautomerism rule 2>
...STARTP
...C[ARYL=YES](=O)-X[ARYL=YES]=C[ARYL=YES](-N[HS>0];[RINGS=NO])
...ENDP

RULE
EXPLANATION Enaminoketone/enolimine tautomerism.
IF  tautomerism rule 2
THEN substitute-bond 1 with -
        substitute-bond 2 with =
        substitute-bond 3 with -
        substitute-bond 4 with =
        add-taut-hydrogen 2
END-THEN
```



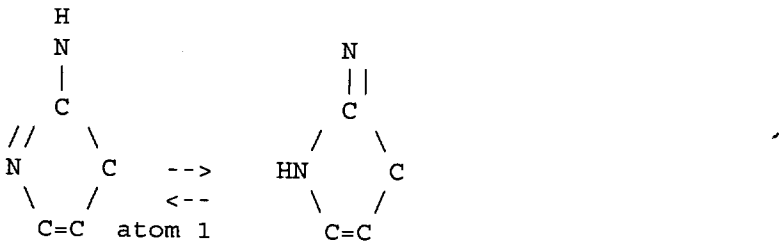
Imidazole

```
CHEMICAL-LABEL <tautomerism rule 3>
...STARTP
...N[HS=0];[ARYL=YES]=C[ARYL=YES]-N[HS>0];[ARYL=YES]-C[ARYL=YES]
...=C[ARYL=YES]-@1
...ENDP


RULE
EXPLANATION Imidazole ring.
IF  tautomerism rule 3
THEN substitute-bond 1 with -
        substitute-bond 2 with =
        add-taut-hydrogen 1
END-THEN
```



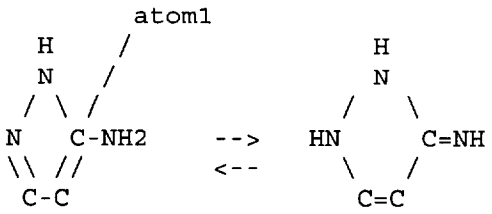
Pyrazole

```
CHEMICAL-LABEL <tautomerism rule 4>
...STARTP
...N[HS=0];[ARYL=YES]=C[ARYL=YES]-C[ARYL=YES]=C[ARYL=YES]-N[HS>0];
...[ARYL=YES]-@1
...ENDP

RULE
EXPLANATION Pyrazole ring.
IF  tautomerism rule 4
THEN substitute-bond 1 with -
        substitute-bond 2 with =
        substitute-bond 3 with -
        substitute-bond 4 with =
        add-taut-hydrogen 1
END-THEN
```



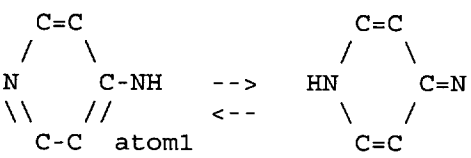
2-amino-pyrrole

```
CHEMICAL-LABEL <tautomerism rule 5>
...STARTP
...C=C-N[HS=0];[ARYL=YES]=C[ARYL=YES](-N[HS>0])-C,S-@1
...ENDP

RULE
EXPLANATION 2-amino-pyrrole or 2-amino-thiazole.
IF  tautomerism rule 5
THEN substitute-bond 3 with -
        substitute-bond 4 with =
        add-taut-hydrogen 3
END-THEN
```



5-amino-pyrazole

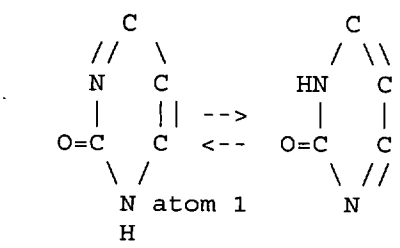
```
CHEMICAL-LABEL <tautomerism rule 6>
...STARTP
...C[ARYL=YES](-N[HS>0])=C[ARYL=YES]-C[ARYL=YES]=N[HS=0];[SAMERING=6];
...[ARYL=YES]
...-N[SAMERING=5]-@1
...ENDP


RULE
EXPLANATION 5-amino-pyrazole.
IF  tautomerism rule 6
THEN substitute-bond 1 with =
        substitute-bond 2 with -
        substitute-bond 3 with =
        substitute-bond 4 with -
        add-taut-hydrogen 5
END-THEN


N.B. SAMERING=atomnum : means in same ring as atomnum
```



4-amino-pyridine

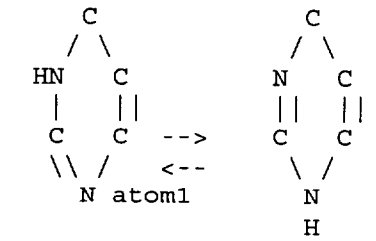
```
CHEMICAL-LABEL <tautomerism rule 7>
...STARTP
...C[ARYL=YES]-C[ARYL=YES]=N[HS=0];[ARYL=YES]-C=C-C(-N[HS>0])=@1
...ENDP
```

```
RULE
EXPLANATION 4-amino-pyridine.
IF  tautomerism rule 7
THEN substitute-bond 1 with =
        substitute-bond 2 with -
        substitute-bond 6 with =
        substitute-bond 7 with -
        add-taut-hydrogen 3
END-THEN
```



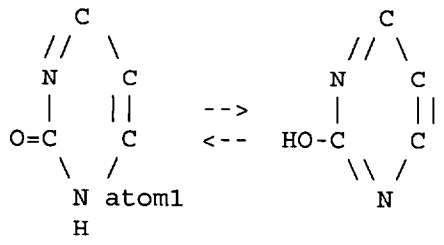
2-pyrimidone tautomerism type 1

```
CHEMICAL-LABEL <tautomerism rule 8>
...STARTP
...N[HS>0];[ARYL=YES]-C[ARYL=YES](=O)-N[HS=0];[ARYL=YES]=C[ARYL=YES]
...-C[ARYL=YES]=C[ARYL=YES]-@1
...ENDP


RULE
EXPLANATION 2-pyrimidone tautomerism type 1.
IF  tautomerism rule 8
THEN substitute-bond 4 with -
        substitute-bond 5 with =
        substitute-bond 6 with -
        substitute-bond 7 with =
        add-taut-hydrogen 4
END-THEN
```



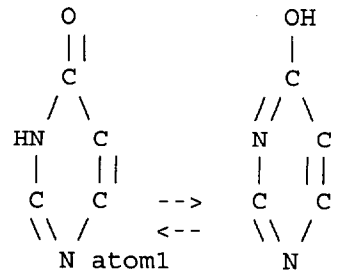
pyrimidine tautomerism type 1

```
CHEMICAL-LABEL <tautomerism rule 9>
...STARTP
...N[HS=0];[ARYL=YES]=C,N[ARYL=YES]-N[HS>0];[SAMERING=1];[ARYL=YES]
...ENDP


RULE
EXPLANATION pyrimidine tautomerism type 1, also cover tetrazole, triazole,
oxy-triazine, oxy-imidazole.
IF  tautomerism rule 9
THEN substitute-bond 1 with -
        substitute-bond 2 with =
        add-taut-hydrogen 1
END-THEN
```



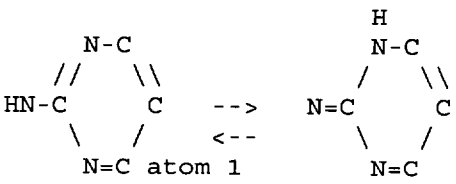
2(1H)-pyrimidone

```
CHEMICAL-LABEL <tautomerism rule 10>
...STARTP
...N[HS>0];[ARYL=YES]-C[ARYL=YES](=O)-N[HS=0];[ARYL=YES]=C[ARYL=YES]
...-C[ARYL=YES]=C[ARYL=YES]-@1
...ENDP


RULE
EXPLANATION Amide/hydroxyimine tautomerism in 2(1H)-pyrimidone.

IF  tautomerism rule 10
THEN substitute-bond 1 with =
        substitute-bond 2 with -
        add-taut-hydrogen 3
END-THEN
```



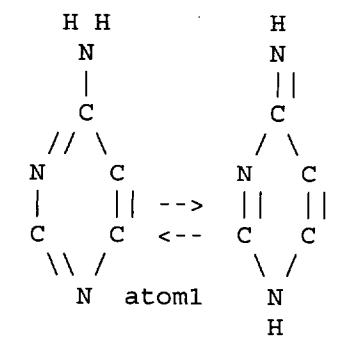
4(3H)-pyrimidone

```
CHEMICAL-LABEL <tautomerism rule 11>
...STARTP
...N[HS=0];[ARYL=YES]=C[ARYL=YES]-N[HS>0];[SAMERING=1];[ARYL=YES]-C[ARYL=YES]
...(=X[HS=0])-C[ARYL=YES]=C[ARYL=YES]-@1
...ENDP


RULE
EXPLANATION Amide/hydroxyimine on 4(3H)-pyrimidone ring.
IF  tautomerism rule 11
THEN substitute-bond 3 with =
        substitute-bond 4 with -
        add-taut-hydrogen 5
END-THEN
```



2(1H)-amino-pyrimidine

```
CHEMICAL-LABEL <tautomerism rule 12>
...STARTP
...C[HS=1]=N[HS=0]-C[ARYL=YES](-N[HS>0])=N[HS=0];[ARYL=YES]-C[HS=1]=C[HS=1]-@1
...ENDP


RULE
EXPLANATION 2(1H)-amino-pyrimidine tautomerism.
IF  tautomerism rule 12
THEN substitute-bond 3 with =
        substitute-bond 4 with -
        add-taut-hydrogen 5
END-THEN
```



4(1H)-amino-pyrimidine

```
CHEMICAL-LABEL <tautomerism rule 13>
...STARTP
...N[HS=0];[ARYL=YES]=C[HS=1];[ARYL=YES]-N[HS=0];[ARYL=YES]=C[ARYL=YES]
...(-N[HS>0])
...ENDP


RULE
EXPLANATION 4(1H)-amino-pyrimidine tautomerism.
IF  tautomerism rule 13
THEN substitute-bond 1 with -
        substitute-bond 2 with =
        substitute-bond 3 with -
        substitute-bond 4 with =
        add-taut-hydrogen 1
END-THEN
```

```
*************************************************************************
                          ~~~~~~~~~~
                          PERCEPTION
                          ~~~~~~~~~~
*Knowledge Base Name    :charge.kbase
*Description            :A knowledge base to assign protonation states
*                        according to estimated pKa values.
*Created by             :A.K.T. Ting
*Date of creation       :28-8-97
*Last modified          :8-2-99
*Version                :3.1
*************************************************************************


Information :
~~~~~~~~~~~~~

New CAESA Rule Base substitution types :
      add-prot-hydrogen
   del-prot-hydrogen

Protonation behaviour is assigned by considering pKa value of compounds at
pH7

1. pKa > 7 molecule will exist mainly in the protonated form at pH7
2. pKa < 7 molecule will exist mainly in the deprotonated form at pH7


Charges set in this kbase only represent the sign of the charge and not
the  magnitude.

Later rules override the effects of earlier ones.

This knowledge base is not comprehensive but serves to illustrate the rules.


pKa Reference :
~~~~~~~~~~~~~~~

(1) Joule,J. A.; Mills, K.; Smith, G.F.  Heterocyclic Chemistry.
London Chapman and Hall, 1994.

(2)  Physical Methods in Heterocyclic Chemistry}, Katritzky, A.R., Ed.,
Academic Press, 1963; Vol. I.

(3)  Physical Methods in Heterocyclic Chemistry}, Katritzky, A.R., Ed.,
Academic Press. 1971; Vol. III.

(4) Davies, D. T.   Aromatic Heterocyclic Chemistry}.
Oxford Science Publications, 1992.

(5)  Comprehensive Heterocyclic Chemistry -- Part 2B: Six--membered Rings
with Oxygen, Sulfur or Two or More Nitrogen Atoms}. Katritzky, A.R., Ed.,
Pergamin, 1984.
```

(6) Comprehensive Heterocyclic Chemistry -- Part 4B: Five--membered Rings with Two or More Oxygen, Sulfur or Nitrogen Atoms}. Katritzky, A.R., Ed., Pergamin, 1984.

(7) Handbook of Heterocyclic Chemistry}. Katritzky, A.R., Ed.; Pergamin, 1985.

(8) Merck Index}. Budavari, S., Ed.; MERCK; Co., Inc, 1996.

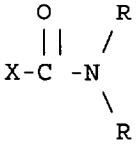
Sections:
~~~~~~~~~
1. Simple functional groups
2. 6 membered aromatic heterocyclic compounds {1 or more heteroatom(s)}
3. 5 membered aromatic heterocyclic compounds {2 or more heteroatoms}
4. Others

Simple functional groups
~~~~~~~~~~~~~~~~~~~~~~~~

```
CHEMICAL-LABEL <Amine>
...STARTP
...C[ARYL=NO]-N[EPS=1];[SPCENTRE=3];[ARYL=NO]
...ENDP

RULE
EXPLANATION Amine protonated at pH 7.
IF Amine
THEN set-charge 2 to +1
     add-prot-hydrogen 2
END-THEN
```

ArNH2
Aromatic Amine

```
CHEMICAL-LABEL <Aromatic Amine>
...STARTP
...C[ARYL=YES]-N[HS>0];[ARYL=NO]
...ENDP

RULE
EXPLANATION Aromatic amine is not protonated at pH 7.
IF Aromatic Amine
THEN set-charge 2 to 0
END-THEN
```



```
Amide

CHEMICAL-LABEL <Amide>
...STARTP
...X-C(=O)-N[HS=0,1,2]
...ENDP

RULE
EXPLANATION Amide, the N has no charge.
IF Amide
THEN set-charge 4 to 0
END-THEN



CHEMICAL-LABEL <Carboxylic Acid>
...STARTP
...O[HS=1];[HETS=0]-C[HS=0,1];[RINGS=NO]=O
...ENDP

RULE
EXPLANATION Carboxylic Acid is not protonated at pH7 and the double bond
normalied between the O's hence both have -ve charge.
IF Carboxylic Acid
THEN set-charge 1 to -1
        set-charge 3 to -1
END-THEN



CHEMICAL-LABEL <Carboxylate>
...STARTP
...O[HS=0];[HETS=0];[CHARGE=ANION]-C[HS=0,1];[RINGS=NO]=O
...ENDP

RULE
EXPLANATION In Carboxylate the double bond normalised between the Os
hence both have -ve charge.
IF Carboxylate
THEN set-charge 1 to -1
        set-charge 3 to -1
END-THEN



CHEMICAL-LABEL <Nitro>
...STARTP
...C-N[CHARGE=CATION](=O)-O[CHARGE=ANION]
...ENDP

RULE
EXPLANATION Nitro group has +ve charge on N and the double bond normalised
between the Os hence both Os have -ve charge.
IF Nitro
THEN set-charge 2 to +1
        set-charge 3 to -1
        set-charge 4 to -1
END-THEN
```

```
    O⁻
    |+
-C-S-C-
```

sulfoxide

```
CHEMICAL-LABEL <Sulphoxide>
...STARTP
...C(-S[CHARGE=NEUTRAL];[CONNECTIONS=3]=O[CHARGE=NEUTRAL])-C
...ENDP

RULE
EXPLANATION Sulphoxide has +ve charge on S and -ve charge on O.
IF  Sulphoxide
THEN set-charge 2 to +1
        set-charge 3 to -1
END-THEN
```

```
    H
    N
    ||
H2N-C-NH2
```

Guanidine

```
CHEMICAL-LABEL <Guanidine>
...STARTP
...N[HS=0,1];[ARYL=NO]=C[ARYL=NO](-N[HS=0,1,2];[ARYL=NO])
...-N[HS=0,1,2];[ARYL=NO]
...ENDP

RULE
EXPLANATION Guanidine protonated at =N and cation is stabliized through
reasonance.
IF Guanidine
THEN  set-charge 1 to +1
            set-charge 3 to +1
            set-charge 4 to +1
            add-prot-hydrogen 1
END-THEN
```

```
    H
    N
    ||
H2N-C-C
```

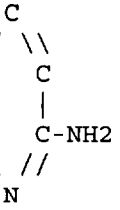
Amidine

```
CHEMICAL-LABEL <Amidine>
...STARTP
...N[HS=0,1];[ARYL=NO]=C[ARYL=NO](-N[HS=0,1,2];[ARYL=NO])-C[ARYL=NO]
...ENDP

RULE
EXPLANATION Amidine protonated at =N and cation is stabliized through
reasonance.
IF Amidine
THEN  set-charge 1 to +1
            set-charge 3 to +1
            set-charge 4 to 0
            add-prot-hydrogen 1
END-THEN
```

6 membered aromatic heterocyclic compounds

Pyridine

Parent pyridine pKa = 5.12, therefore does not protonate at pH7.
Molecules substituted with amino gp and without chloro or nitro gp present at the same time give pKa > 7



2-amino pyridine

```
CHEMICAL-LABEL <2-Amino Pyridine>
...STARTP
...N-,%C(-N[HS>0])=,%C-,%C=,%C-,%C=,%@1
...ENDP


RULE
EXPLANATION 2-Amino Pyridine protonated at ring N.
IF 2-Amino Pyridine
THEN  set-charge 1 to +1
            add-prot-hydrogen 1
END-THEN

CHEMICAL-LABEL <4-Amino Pyridine>
...STARTP
...N-,%C=,%C-,%C(-N[EPS=1])=,%C-,%C=,%@1
...ENDP
```

```
RULE
EXPLANATION 4-Amino Pyridine protonated at ring N.
IF 4-Amino Pyridine
THEN  set-charge 1 to +1
            add-prot-hydrogen 1
END-THEN



CHEMICAL-LABEL <3-Nitro-Pyridine>
...STARTP
...N-,%C=,%C(-N(-O)=O)-,%C=,%C-,%C=,%@1
...ENDP

RULE
EXPLANATION 3-Nitro Pyridine is not protonated at pH7.
IF 3-Nitro-Pyridine
THEN  set-charge 1 to 0
            set-charge 4 to 0
            del-prot-hydrogen 1
END-THEN



CHEMICAL-LABEL <2-Chloro-Pyridine>
...STARTP
...N-,%C(-Cl)=,%C-,%C=,%C-,%C=,%@1
...ENDP


RULE
EXPLANATION 2-Chloro Pyridine is not protonated at pH7.
IF 2-Chloro-Pyridine
THEN  set-charge 1 to 0
      del-prot-hydrogen 1
END-THEN



CHEMICAL-LABEL <4-Chloro-Pyridine>
...STARTP
...N-,%C=,%C-,%C(-Cl)=,%C-,%C=,%@1
...ENDP


RULE
EXPLANATION 4-Chloro Pyridine is not protonated at pH7.
IF 4-Chloro-Pyridine
THEN  set-charge 1 to 0
      del-prot-hydrogen 1
END-THEN
```

```
5 membered aromatic heterocyclic compounds
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~


Imidazole
=========

 C-N3
// \\
C   C
 \ /
  N1
  H

imidazole

Parent imidazole pKa = 7.0
Molecules with amino gp substitutents also have pKa > 7,
in absence of 2- or 4- chloro or nitro gps.


CHEMICAL-LABEL <Imidazole>
...STARTP
...N[HS=1]-C=N-C=C-@1
...ENDP


RULE
EXPLANATION Imidazole protonated at pH7.
IF Imidazole
THEN  set-charge 1 to 0
            set-charge 3 to +1
            add-prot-hydrogen 3
END-THEN




CHEMICAL-LABEL <2-Chloro Imidazole>
...STARTP
...N[HS=1]-C(-Cl)=N-C=C-@1
...ENDP


RULE
EXPLANATION 2-Chloro Imidazole is not protonated at pH7.
IF  2-Chloro Imidazole
THEN  set-charge 1 to 0
            set-charge 4 to 0
            del-prot-hydrogen 4
END-THEN



CHEMICAL-LABEL <4-Chloro Imidazole>
...STARTP
...N[HS=1]-C=N-C(-Cl)=C-@1
...ENDP


RULE
```

```
EXPLANATION 4-Chloro Imidazole is not protonated at pH7.
IF  4-Chloro Imidazole
THEN  set-charge 1 to 0
            set-charge 3 to 0
            del-prot-hydrogen 3
END-THEN


CHEMICAL-LABEL <5-Chloro Imidazole>
...STARTP
...N[HS=1]-C=N-C=C(-Cl)-@1
...ENDP


RULE
EXPLANATION 5-Chloro Imidazole is not protonated at pH7.
IF  5-Chloro Imidazole
THEN  set-charge 1 to 0
            set-charge 3 to 0
            del-prot-hydrogen 3
END-THEN


CHEMICAL-LABEL <2-Nitro Imidazole>
...STARTP
...N[HS=1]-C(-N[CHARGE=NEUTRAL,CATION](-O[CHARGE=NEUTRAL,ANION])=O)
...=N-C=C-@1
...ENDP


RULE
EXPLANATION 2-Nitro Imidazole is not protonated at pH7.
IF 2-Nitro Imidazole
THEN  set-charge 1 to 0
      set-charge 6 to 0
            del-prot-hydrogen 6
END-THEN



CHEMICAL-LABEL <4-Nitro Imidazole>
...STARTP
...N[HS=1]-C=N-C(-N[CHARGE=NEUTRAL,CATION](-O[CHARGE=NEUTRAL,ANION])=O)
...=C-@1
...ENDP


RULE
EXPLANATION 4-Nitro Imidazole is not protonated at pH7.
IF 4-Nitro Imidazole
THEN  set-charge 1 to 0
      set-charge 3 to 0
            del-prot-hydrogen 3
END-THEN
```

```
CHEMICAL-LABEL <5-Nitro Imidazole>
...STARTP
...N[HS=1]-C=N-C=C(-N[CHARGE=NEUTRAL,CATION](-O[CHARGE=NEUTRAL,ANION])=O)-@1
...ENDP


RULE
EXPLANATION 5-Nitro Imidazole is not protonated at pH7.
IF 5-Nitro Imidazole
THEN  set-charge 1 to 0
      set-charge 3 to 0
            del-prot-hydrogen 3
END-THEN
```

Others
~~~~~~

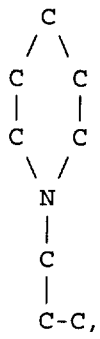
Piperidine
==========

Parental piperidine has pKa > 7 and is cover by <Amine> rule

```
  C
 / \
C   C
|   |
C   C-C,N,S=O
 \ /
  N
  H
```

2-substituted-piperidine with electron withdrawing group

```
CHEMICAL-LABEL <2-Substituted Piperidine>
...STARTP
...N[HS=1]-C(-C,N,S=O)-C-C-C-C-@1
...ENDP


RULE
EXPLANATION 2-Substituted Piperidine not protonated at pH7.
IF 2-Substituted Piperidine
THEN  set-charge 1 to 0
      del-prot-hydrogen 1
END-THEN
```
~~~~~~



N-substituted Piperidine with electron withdrawing group

```
CHEMICAL-LABEL <N-Substituted Piperidine>
...STARTP
...N[HS=0](-C-C,N,S#,=N,O)-C-C-C-C-C-@1
...ENDP


RULE
EXPLANATION N-Substituted Piperidine is not protonated at pH7.
IF N-Substituted Piperidine
THEN  set-charge 1 to 0
      del-prot-hydrogen 1
END-THEN
```

```
*******************************************************************************
                          ~~~~~~~~~~
                          PERCEPTION
                          ~~~~~~~~~~
*Knowledge Base Name     :Hproperty.kbase
*Description             :A simple knowledge base to identify HD and HA bonding
*                         properties
*Created by              :A.K.T. Ting
*Date of creation        :13-5-1997
*Last modified           :23-4-99
*Version                 :1.2
*******************************************************************************


Information :
~~~~~~~~~~~~~~

New CAESA Rule Base substitution types :
   assign-atom

New atom assignment (assign-atom) allowed :
   HACCEPTOR
   HDONOR
   HBOTH
   HNEITHER
   HDXORHA

New atom features added to the  PATRAN language :
HDONOR HACCEPTOR HBOTH HEITHER HNEITHER NOT_HDONOR NOT_HACCEPTOR
p.s. for more information about PATRAN see Reference Guide version 2.0.

Note :
1. The order of the rules in this kbase are ordered according to specifity.
i.e. the more general rules come first and the more specific ones follow.
e.g. Nitrogen with H rule will label all Nitrogens with H as both HA and HD
and primary & secondary amide rule will change that to just HD.

2.    CONNECTIONS - is defined as the number of non-hydrogen atom(s)
connected to the target atom.

This knowledge base is not comprehensive but serves to illustrate the rules.

Sections:
~~~~~~~~~

                        1  Nitrogen
                        2  Oxygen
                        3  Halogens
                        4  Sulphur
                        5  Nitrogen groups
                        6  Oxygen groups
                        7  Sulphur groups
                        8  Phosphorus groups
                        9  Others
```

```
Nitrogen
~~~~~~~~


CHEMICAL-LABEL <Nitrogen without H>
...STARTP
...X&N[HS=0];[CHARGE=NEUTRAL]
...ENDP

RULE
EXPLANATION Nitrogen without H is HA.
IF Nitrogen without H
THEN assign-atom 2 as HACCEPTOR
END-THEN

N.B. e.g. <Nitrogen without H> rule covers nitriles.


CHEMICAL-LABEL <Nitrogen with H>
...STARTP
...X&N[HS>0];[CHARGE=NEUTRAL]
...ENDP

RULE
EXPLANATION Nitrogen with H is both HA and HD.
IF Nitrogen with H
THEN assign-atom 2 as HBOTH
END-THEN



Oxygen
~~~~~~


CHEMICAL-LABEL <Oxygen without H>
...STARTP
...X&O[HS=0]
...ENDP

RULE
EXPLANATION Oxygen without H is HA.
IF Oxygen without H
THEN assign-atom 2 as HACCEPTOR
END-THEN


N.B.
1. e.g. <Oxygen without H> rule covers carbonyls.
2. Oxygen with H case will be cover as <Alcohol> rule.


Halogens
~~~~~~~~


CHEMICAL-LABEL <Fluoride>
...STARTP
...X-F
...ENDP
```

```
RULE
EXPLANATION Fluoride is  a weak HA.
IF Fluoride
THEN assign-atom 2 as HACCEPTOR
END-THEN

N.B. Fluoride is a weak Hydrogen acceptor.



CHEMICAL-LABEL <Chloride>
...STARTP
...X-Cl
...ENDP

RULE
EXPLANATION Chloride is HA.
IF  Chloride
THEN  assign-atom 2 as HACCEPTOR
END-THEN



Sulphur
~~~~~~~


Note : The order of the rule : S without H & EPS
                               S with H
                               S with H & EPS


CHEMICAL-LABEL <Sulphur without H and EPS>
...STARTP
...C&S[HS=0];[EPS=0]
...ENDP

RULE
EXPLANATION Sulphur without H and EPS is neither HA nor HD.
IF Sulphur without H and EPS
THEN assign-atom 2 as HNEITHER
END-THEN



CHEMICAL-LABEL <Sulphur with H>
...STARTP
...C&S-H
...ENDP

RULE
EXPLANATION Sulphur with H is HD.
IF Sulphur with H
THEN assign-atom 2 as HDONOR
END-THEN

N.B. e.g. <Sulphur with H> covers thiols.
```

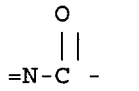
```
CHEMICAL-LABEL <Sulphur with H and EPS>
...STARTP
...C&S[HS>0];[EPS>0]
...ENDP

RULE
EXPLANATION Sulphur with H and EPS is both HA and HD.
IF Sulphur with H and EPS
THEN assign-atom 2 as HBOTH
END-THEN



CHEMICAL-LABEL <Sulphur with EPS and no H>
...STARTP
...C&S[HS=0];[EPS>0]
...ENDP

RULE
EXPLANATION Sulphur with EPS and no H is HA.
IF Sulphur with EPS and no H
THEN assign-atom 2 as HACCEPTOR
END-THEN



CHEMICAL-LABEL <Sulphur with H and no EPS>
...STARTP
...C&S[HS>0];[EPS=0]
...ENDP

RULE
EXPLANATION Sulphur with H and no EPS is HD.
IF Sulphur with H and no EPS
THEN assign-atom 2 as HDONOR
END-THEN


Nitrogen groups
~~~~~~~~~~~~~~~

 |
-NH  and -NH2 are donors only
Ref: Hydrogen bonding in biological structures by G A Jeffrey and W. Saenger.


CHEMICAL-LABEL <sp3 NH>
...STARTP
...X-N[HS=1,2];[SPCENTRE=3]
...ENDP

RULE
EXPLANATION sp3 NH is HD.
IF sp3 NH
THEN assign-atom 2 as HDONOR
END-THEN
```

```
CHEMICAL-LABEL <aromatic NHs>
...STARTP
...X%N[HS>0]
...ENDP

RULE
EXPLANATION aromatic NHs is HD.
IF aromatic NHs
THEN assign-atom 2 as HDONOR
END-THEN
```

e.g. the aromatic N in Pyrrole.

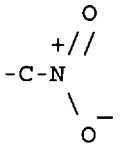
```
cover cases : N+ -H
              NH4+
              RNH3+
              R2NH2+
              R3NH+

CHEMICAL-LABEL <N cation>
...STARTP
...N[HS>0];[CHARGE=CATION]
...ENDP

RULE
EXPLANATION N cation is HD.
IF N cation
THEN assign-atom 1 as HDONOR
END-THEN
```

```
  O    H
  ||  /
X-C -N
     \
      H/R

CHEMICAL-LABEL <Primary and Secondary Amide>
...STARTP
...X-C(=O)-N[HS=1,2]
...ENDP

RULE
EXPLANATION Primary and Secondary Amide, O is HA and N is HD.
IF Primary and Secondary Amide
THEN assign-atom 3 as HACCEPTOR
     assign-atom 4 as HDONOR
END-THEN
```

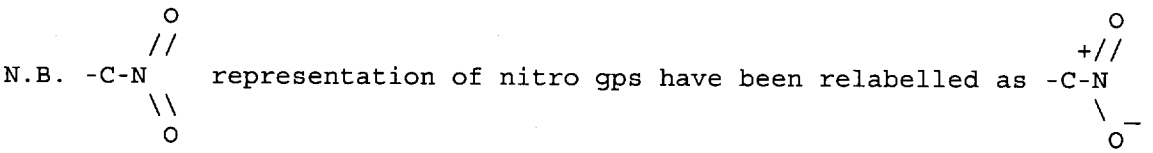


```
CHEMICAL-LABEL <Tertiary Amide>
...STARTP
...X-C(=O)-N[HS=0]
...ENDP

RULE
EXPLANATION Tertiary Amide, O is HA and N is neither HD nor HA.
IF Tertiary Amide
THEN assign-atom 3 as HACCEPTOR
     assign-atom 4 as HNEITHER
END-THEN
```

```
CHEMICAL-LABEL <Tertiary Amine with sp2 character>
...STARTP
...C[SPCENTRE=2]-N[HS=0];[SPCENTRE=3]
...ENDP

RULE
EXPLANATION Sp3 N next to sp2 centre,  N assumed to be sp2
Tertiary Amine with sp2 character is neither HA nor HD because
electron delocalised into the sp2 system.
IF Tertiary Amine with sp2 character
THEN assign-atom 2 as HNEITHER
END-THEN
```



```
CHEMICAL-LABEL <Nitro>
...STARTP
...C-N[CHARGE=CATION](=O)-O[CHARGE=ANION]
...ENDP

RULE
EXPLANATION Nitro, N is neither HD nor HA, the two O are HAs due to
normalisation.
IF Nitro
THEN assign-atom 2 as HNEITHER
        assign-atom 3 as HACCEPTOR
        assign-atom 4 as HACCEPTOR
END-THEN
```



```
     H
     N
     ||
H2N-C-NH2

CHEMICAL-LABEL <Unsubstituted guanidine>
...STARTP
...N[HS=1];[ARYL=NO]=C[ARYL=NO](-N[HS=1,2];[ARYL=NO])-N[HS=1,2];[ARYL=NO]
...ENDP

RULE
EXPLANATION Unsubstituted guanidine, all Ns are HDs.
IF Unsubstituted guanidine
THEN assign-atom 1 as HDONOR
        assign-atom 3 as HDONOR
        assign-atom 4 as HDONOR
END-THEN
```

```
    H
    N
    ||
R2N-C-NH2


CHEMICAL-LABEL <1,1-Disubstituted guanidine>
...STARTP
...N[HS=1,2]=C[ARYL=NO](-N[HS=0];[ARYL=NO])-N[HS=1,2]
...ENDP

RULE
EXPLANATION 1,1-Disubstituted guanidine, N without H is neither HA nor HD and
other Ns are HDs.
IF 1,1-Disubstituted guanidine
THEN assign-atom 1 as HDONOR
        assign-atom 3 as HNEITHER
        assign-atom 4 as HDONOR
END-THEN
```

```
    H
    N
    ||
R2N-C-NR2

CHEMICAL-LABEL <1,1,3,3,-tetrasubstituted guanidine>
...STARTP
...N[HS=1];[ARYL=NO]=C[ARYL=NO](-N[HS=0];[ARYL=NO])-N[HS=0];[ARYL=NO]
...ENDP

RULE
EXPLANATION 1,1,3,3,-tetrasubstituted guanidine, only the NHs group is HD.

IF 1,1,3,3,-tetrasubstituted guanidine
THEN assign-atom 1 as HDONOR
        assign-atom 3 as HNEITHER
        assign-atom 4 as HNEITHER
END-THEN
```

```
CHEMICAL-LABEL <Amidine>
...STARTP
...C[ARYL=NO](=N[HS=1];[ARYL=NO])-N[HS=1,2];[ARYL=NO]
...ENDP

RULE
EXPLANATION Amidine, both Ns are HDs.
IF  Amidine
THEN assign-atom 2 as HDONOR
        assign-atom 3 as HDONOR
END-THEN
```

Oxygen groups

```
CHEMICAL-LABEL <Alcohol>
...STARTP
...X-O[HS=1]
...ENDP

RULE
EXPLANATION Alcohol is both HA and HD.
IF Alcohol
THEN assign-atom 2 as HBOTH
END-THEN
```

```
-C=O
 |
 OH

CHEMICAL-LABEL <Non-normalised Carboxylic Acid>
...STARTP
...O[HS=1];[HETS=0]-C[HS=0,1];[RINGS=NO]=O[HETS=0];[HS=0]
...ENDP

RULE
EXPLANATION Carboxylic Acid deprotonated at pH7 and the double bond is
normalised between the Os and hence both HAs.
IF Non-normalised Carboxylic Acid
THEN assign-atom 1 as  HACCEPTOR
        assign-atom 3 as  HACCEPTOR
END-THEN


CHEMICAL-LABEL <Deprotonated Carboxylic Acid>
...STARTP
...O[CHARGE=ANION];[HETS=0]-C[HS=0,1];[RINGS=NO]=O[CHARGE=ANION];[HETS=0]
...ENDP
```

```
RULE
EXPLANATION Carboxylic Acid deprotonated at pH7 and the double bond is
normalised between the Os and hence both HAs.
IF Deprotonated Carboxylic Acid
THEN assign-atom 1 as  HACCEPTOR
        assign-atom 3 as  HACCEPTOR
END-THEN
```

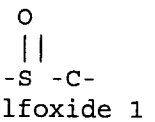
N.B. Since after protonation perception, acid groups are deprotonated and both Os have negative charges, therefore Deprotonated Carboxylic Acid rule needed to correctly identify Hproperties.

```
-C=O
  \ _
   O

CHEMICAL-LABEL <Carboxylate>
...STARTP
...O[HS=0];[HETS=0];[CHARGE=ANION]-C[HS=0,1];[RINGS=NO]=O[HETS=0];[HS=0]
...ENDP

RULE
EXPLANATION Carboxylate, the double bond is normalised between the Os
and hence both Os are HAs.
IF Carboxylate
THEN assign-atom 1 as  HACCEPTOR
        assign-atom 3 as  HACCEPTOR
END-THEN
```

```
  O
  ||
C-C-O-C

CHEMICAL-LABEL <Ester>
...STARTP
...C-O[HS=0];[HETS=0];[CHARGE=NEUTRAL]-C(=O)-C
...ENDP

RULE
EXPLANATION Ester, both Os are HAs, -O- a weak HA.
IF Ester
THEN assign-atom 2 as  HACCEPTOR
        assign-atom 4 as  HACCEPTOR
END-THEN
```

Sulphur groups
~~~~~~~~~~~~~~

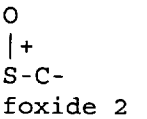
Sulfoxide :



sulfoxide 1

Ref : (O=S)gps is an acceptor gp according to Jeffrey Hydrogen bonding bk p.22

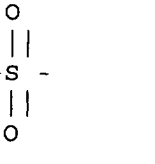
```
CHEMICAL-LABEL <Sulphoxide without charge>
...STARTP
...C(-S[CHARGE=NEUTRAL]=O[CHARGE=NEUTRAL])-C
...ENDP

RULE
EXPLANATION Sulphoxide without charge, both O and S are HAs.
IF  Sulphoxide without charge
THEN assign-atom 2 as  HACCEPTOR
     assign-atom 3 as  HACCEPTOR
END-THEN
```



sulfoxide 2

```
CHEMICAL-LABEL <Sulphoxide with charge>
...STARTP
...C(-S[CHARGE=CATION]-O[CHARGE=ANION])-C
...ENDP

RULE
EXPLANATION Sulphoxide with charge, both O and S are HAs.
IF  Sulphoxide with charge
THEN assign-atom 2 as  HACCEPTOR
     assign-atom 3 as  HACCEPTOR
END-THEN
```

Sulfonamide (R2N-SO2R') :




```
CHEMICAL-LABEL <Sulphonamide with acidic H>
...STARTP
...X&N[HS=1,2](-S[CONNECTIONS=4];[HS=0](=O)=O)
...ENDP
```
~~~~~~~~~~~~~~

```
RULE
EXPLANATION Sulphonamide with acidic H, the N with acidic H is HD and both Os
are HA and S neither HA nor HD.
IF  Sulphonamide with acidic H
THEN assign-atom 2 as HDONOR
     assign-atom 3 as HNEITHER
     assign-atom 4 as HACCEPTOR
     assign-atom 5 as HACCEPTOR
END-THEN
```

```
     O
     ||
R2-N-S -    (where R cannot be H)
     ||
     O
```

```
CHEMICAL-LABEL <Sulphonamide without acidic H>
...STARTP
...X&N[HS=0](-S[CONNECTIONS=4];[HS=0](=O)=O)
...ENDP

RULE
EXPLANATION Sulphonamide without acidic H, both Os are HA, and N without acidic
H and S are neither HA nor HD.
IF  Sulphonamide without acidic H
THEN assign-atom 2 as HNEITHER
     assign-atom 3 as HNEITHER
     assign-atom 4 as HACCEPTOR
     assign-atom 5 as HACCEPTOR
END-THEN
```

```
  O
  ||
R-S=O
  |
  OH
```

```
CHEMICAL-LABEL <Sulfonic acid>
...STARTP
...O[HS=1]-S[CONNECTIONS=4];[RINGS=NO](=O)=O
...ENDP

RULE
EXPLANATION Sulfonic Acid deprotonated in pH7 and the double bond is normalised
between the Os and hence both HAs.
IF Sulfonic acid
THEN assign-atom 1 as  HACCEPTOR
     assign-atom 2 as  HNEITHER
        assign-atom 3 as  HACCEPTOR
        assign-atom 4 as  HACCEPTOR
END-THEN
```

Phosphorus group
~~~~~~~~~~~~~~~~~

```
   O
   ||
-N-P -
   |
```

CHEMICAL-LABEL <Phosphoramide with fully substituted N>
...STARTP
...N[HS=0]-P=O
...ENDP

RULE
EXPLANATION Phosphoramide with N only, N is neither HD nor HA, and O is a HA.

IF Phosphoramide with fully substituted N
THEN assign-atom 1 as HNEITHER
     assign-atom 3 as HACCEPTOR
END-THEN

N.B. similar to sulphonamide case where N is never HA and O is a HA.

CHEMICAL-LABEL <Phosphoramide with NH>
...STARTP
...N[HS>0]-P=O
...ENDP

RULE
EXPLANATION Phosphoramide with NH, N is HD and O is a HA.
IF Phosphoramide with NH
THEN assign-atom 1 as HDONOR
     assign-atom 3 as HACCEPTOR
END-THEN

N.B. Other cases where O=P[HETS=0] will be cover by rule <O without H>

```
  OH
  |
R-P=O
  |
  OH
```

CHEMICAL-LABEL <Phosphonic acid>
...STARTP
...O[HS=1]-P[CONNECTIONS=4];[RINGS=NO](=O)-O[HS=1]
...ENDP

RULE
EXPLANATION Phosphonic Acids are strong acids and therefore deprotonated at
both OH gps at pH7 .
IF Phosphonic acid
THEN assign-atom 1 as HACCEPTOR
        assign-atom 3 as HACCEPTOR
        assign-atom 4 as HACCEPTOR
END-THEN
~~~~~~~~~~~~~~~~~

```
Others
~~~~~~



CHEMICAL-LABEL <Water>
...STARTP
...H-O-H
...ENDP

RULE
EXPLANATION water is both HA and HD.
IF Water
THEN assign-atom 2 as HBOTH
END-THEN
```