Supporting Information to

Benchmarking Sets for Molecular Docking

Niu Huang, Brian K. Shoichet, John J. Irwin**

Department of Pharmaceutical Chemistry, University of California San Francisco,

QB3 building, 1700 4th St., Box 2550, San Francisco, CA 94143-2550 U.S.A

RUNNING TITLE: ligand and decoy benchmarking sets for docking

* To whom correspondence should be addressed. B.K.S (Email) shoichet@cgl.ucsf.edu (Phone) 415-514-4126 (Fax) 415-514-4260 J.J.I (Email) jji@cgl.ucsf.edu (Phone) 415-514-4127 (Fax) 415-514-4260

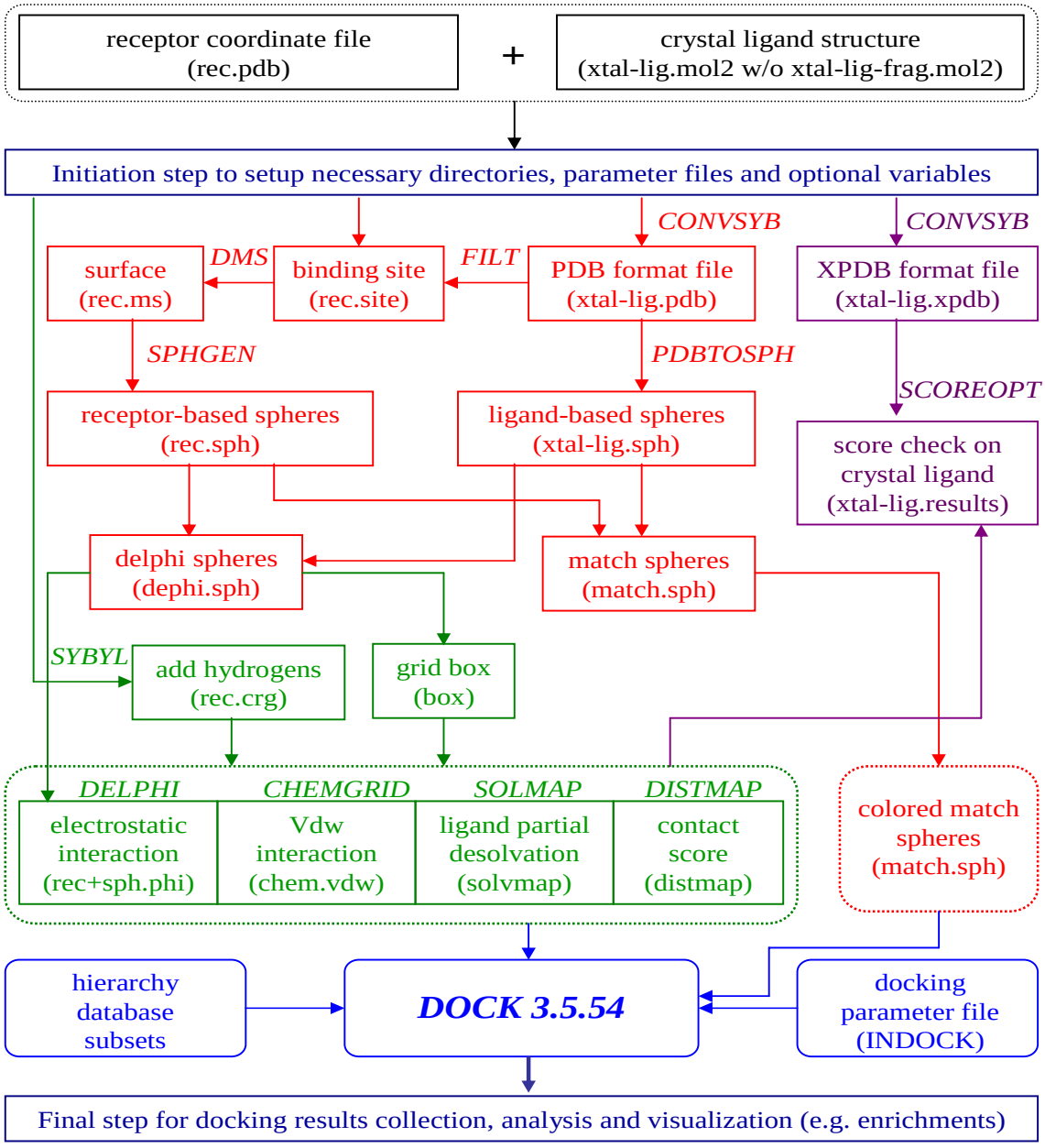

Figure S1. Schematic description of the automated docking pipeline. Steps involved in sphere generation are colored in red, scoring grids computation is in green and scoring the crystallographic ligand is in purple. The programs used in this docking pipeline are in *italic* (*e.g. FILT*, *DMS*, *SPHGEN*), and the associated file names are in parenthesis (e.g. rec.pdb, xtal-lig.mol2).

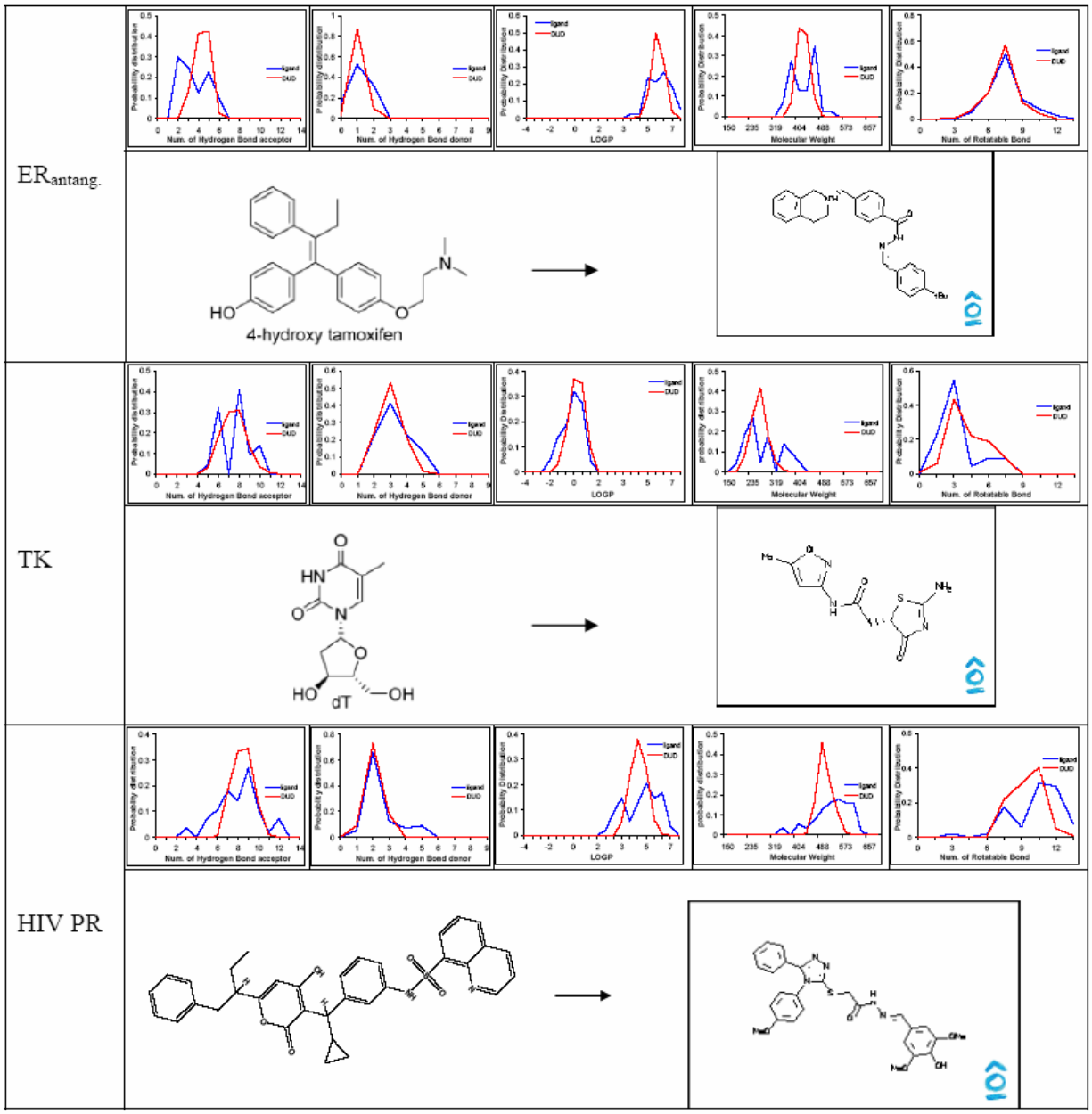

Figure S2. For each system, the blue line represents the property distributions for annotated ligands and the red line is the result from its decoys; the compound shown in left is the native ligand and the compound shown on the right is the selected decoy corresponding to this native ligand.

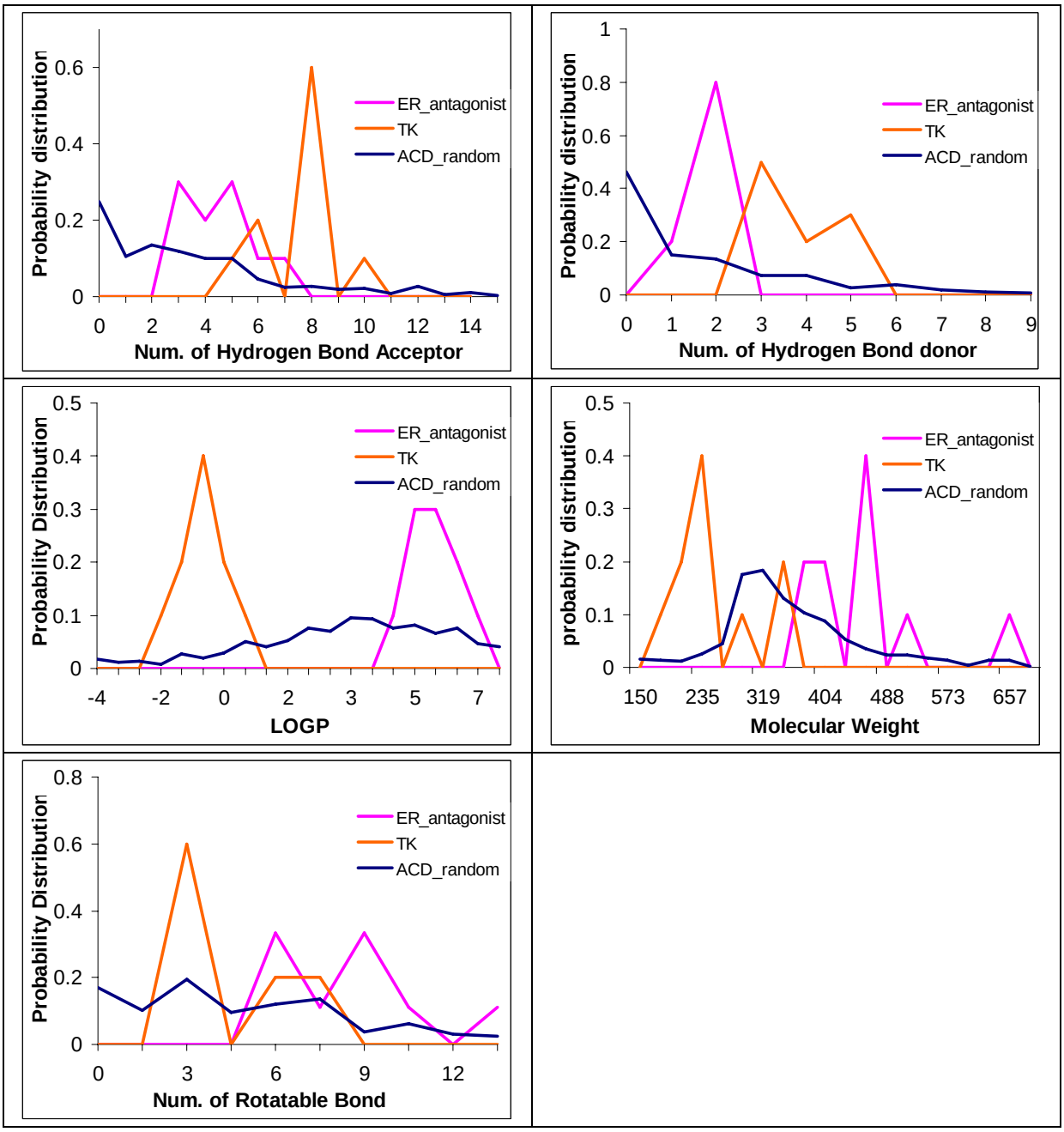

Figure S3. The dark blue line represents the results for 990 ACD random compounds used at Rognan's original dataset, the orange line is the results for 10 TK inhibitors and the purple line is the results for 10 ER antagonists.

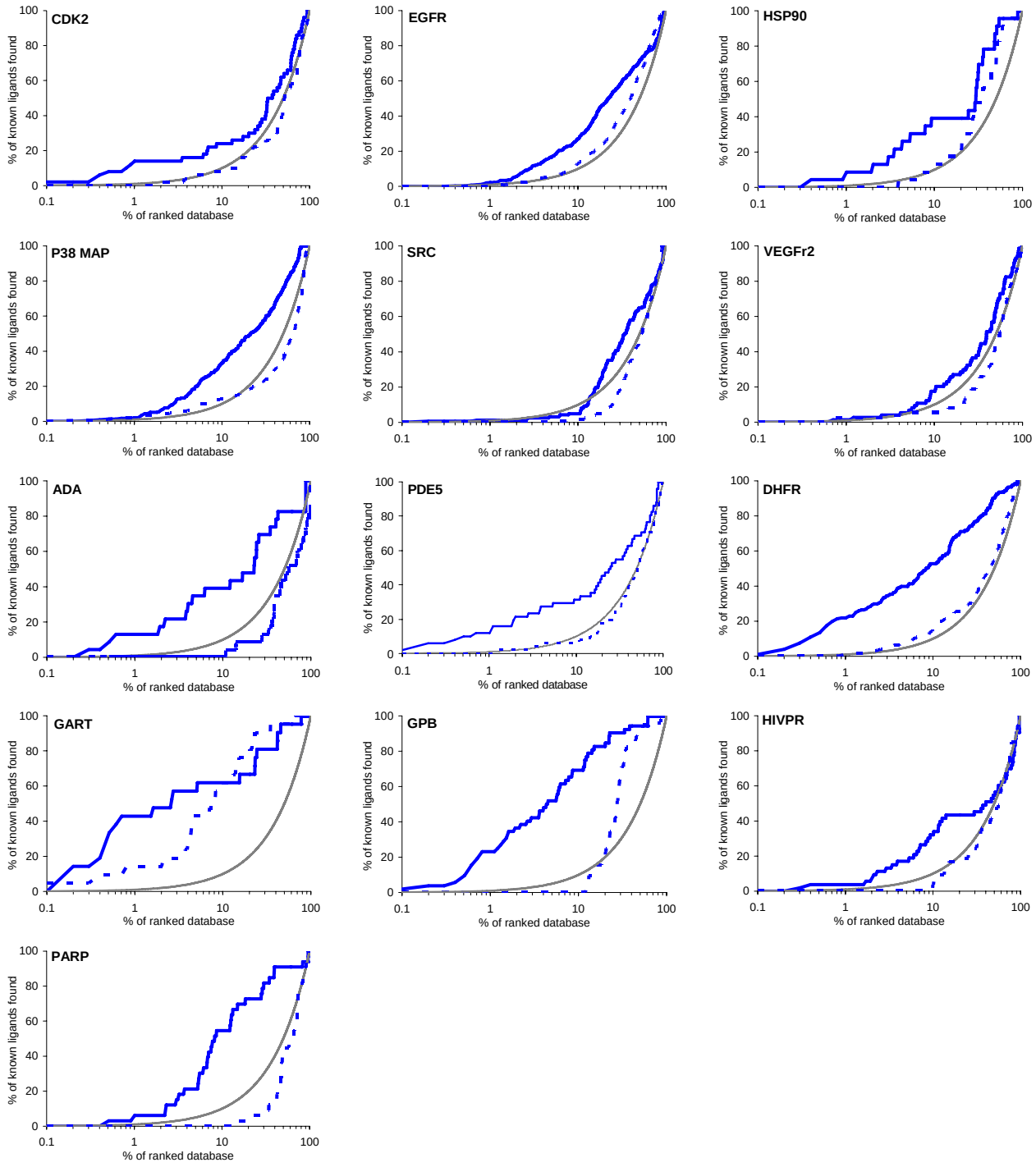

Figure S4. Enrichment plots for 13 targets comparing a semi-automated procedure (blue solid line) with completely automated docking (blue dotted line). The gray line represents the results expected from random selection. The 2,950 DUD ligands and 95,316 DUD decoys were docked in each case. Targets are listed in same order as in Table 1.

Modified parameter files used.

1. ionizer.ini (used by ligprep, Schrodinger Inc., New York)
2. tautomer_list (used by ligprep, Schrodinger Inc., New York)
3. torlib.txt (used by Omega, OpenEye, Santa Fe, NM)
4. rules.txt (used by filter.py, an OEChem (OpenEye, Santa Fe, NM) script written by David M. Lorber, SGX, La Jolla, CA)

```
# $Id: ionizer.ini,v 1.40 2004/02/06 19:46:18 reboul Exp $
# Control specs for ionization state expander
# Copyright 2002-2003 Schrodinger, LLC.
# All rights reserved.
#################################################################

#################################################################
#
# Ionization pattern spec syntax
# ==============================
#
# In the Ionizer's processing of this data file, blank lines
# are ignored, and any text to the right of a '#' is treated
# as a comment.
#
# All other text is treated as an ionization spec. Each valid
# ionization spec consists of a command name, followed by the
# command's required parameters. These ion spec commands are
# supported --
#
# Acid:     Specifies an acid group, and how to ionize it
#
# Base:     Specifies a Base group, and how to ionize it
#
# Exclude:  Specifies a group to exclude from consideration
#           as ionizable; used to selectively "forget" group
#           matches Based on prior Acid/Base specs
#
# Ionizable groups are specified using Schrodinger's linear
# substructure notation, which goes by the name "mmsubs". That
# syntax is described in this installed document --
#
```

```
#     $SCHRODINGER/services-v#####/doc/mmsubs_syntax.txt
#
################################################################
#
# Note Well
# =========
#
# Every input CT is tested for any substructures matching
# the specified ionizable groups, in the order in which they
# appear below. The effect of a later pattern match will
# supersede the effect of an earlier pattern match. Thus,
# anyone crafting pattern specs like these must think very
# carefully about the proper order for the desired effects.
#
# For example -- look below -- the sulfonic pattern has to
# come after the sulfinic pattern, because the sulfonic
# pattern is a specialization of the sulfinic pattern. Any
# group which matches sulfonic will also match sulfinic, but
# the opposite assertion is not necessarily true.
#
# If the order of the two patterns were reversed below, then
# a sulfonic match would always be superseded by the sulfinic
# match, thereby losing the intended distinction between the
# two.
#
# Because of the great generality of this syntax, the inter-
# pattern precedences can be much less obvious than for the
# sulfinic vs. sulfonic case. Be careful!
#
################################################################
#
# We now describe the command parameter syntax....
#
#   Acid  <mmsubs_pattern>  <ionized_fragment>  <pKa>
#   Base  <mmsubs_pattern>  <ionized_fragment>  <pKa>
#
# For each input CT, in any substructure found to match the
# specified mmsubs pattern, the atom matching the pattern's
# leading atom will be considered for ionization via fusion
# of the specified ionized fragment. The ion group will be
# treated as acid or Base according to the command used.
#
# The fragment named must be in this program's custom
# fragment library. The library is named "ionized", and it
# currently contains fragments named Ammonium, Hydroxide,
# N-minus, and Thiolate.
```

```
#
# The Acid/Base spec's 3rd parameter, a pKa value for the
# matching group, is used in deciding which ion combinations
# to actually generate in the output, Based on pK and/or
# pH considerations at run time. Note that pKa is used for
# both acids and Bases.
#
# You cannot specify an Acid/Base pattern in this file
# without supplying a pKa value for it. It is assumed that
# a pKa value is at least somewhat accurate, as it is going
# to be used in the program's restriction determinations.
#
# When one Acid/Base pattern is a specialization of another
# Acid/Base pattern, it is probably the case that the pKa
# value for the more specialized pattern can be given more
# precisely. Assuming that the patterns are presented in
# order of increasing specificity, the more precise pKa will
# be assigned ultimately.
#
# Exclude specs don't specify a fragment or pKa....
#
#   Exclude  <mmsubs_pattern>
#
# For each input CT, in any substructure matching the
# specified pattern, the atom matching the pattern's
# leading atom will be _excluded_ from consideration for
# ionization.
#
# An Exclude pattern match causes an effect only if the
# leading atom coincides with a prior Acid/Base pattern
# match's leading atom. If there is no corresponding prior
# Acid/Base match, the Exclude match is simply ignored.
#
# Because of the Ionizer's top-to-bottom processing of these
# pattern specs, each Exclude pattern must be placed after
# all the Acid/Base specs whose matches it might negate.
#
# It is possible for a given input CT atom to be matched as
# an ionization center due to some Acid/Base pattern spec,
# then excluded from consideration due to a later Exclude
# pattern, then re-matched due to an even later Acid/Base
# spec.
#
################################################################
#
# Important note
```

```
# ==============
#
# The following specifications are _not_ an encyclopedic list
# of ionizable groups!
#
# Some users will wish to prepare their own customized data
# file, presumably by adapting a copy from this one, and then
# running with the customized data, specified via command-
# line option.
#
# Such users must understand Schrodinger's "mmsubs" linear
# substructure notation. Correct use of the syntax is not
# trivially easy. Users may need to contact Schrodinger for
# assistance.
#
#####################################################################

#####################################################################
# Specs for acids, to be deprotonated

# carboxylic
#Acid O0(-H0)-C2(=O0)-C0 Hydroxide 4.0
# for now, I don"t want them to be protonated for the CCP pocket
Acid O0(-H0)-C2(=O0)-C0 Hydroxide 2.0

# phosphoric
Acid O0(-H0)-P0(=O0)(-O0)-O0 Hydroxide 2.1

# phosphorylamide
Acid O0(-H0)-P0(=O0)(-N0)-O0 Hydroxide 2.6

# phosphonic
Acid O0(-H0)-P0(=O0)(-O0)-C0 Hydroxide 2.5

# phosphonamide
Acid O0(-H0)-P0(=O0)(-N0)-C0 Hydroxide 2.9

# sulfuric
Acid O0(-H0)-S0(=O0)(=O0)-O0 Hydroxide -2.0

# sulfinic
Acid O0(-H0)-S0(=O0)-C0 Hydroxide 2.0

# sulfonic
# (supersedes match on sulfinic above)
Acid O0(-H0)-S0(=O0)(=O0)-C0 Hydroxide -1.0
```

```
# hydroxamic
Acid O0(-H0)-N0(-H0)-C0=O0 Hydroxide 8.5
Acid O0(-H0)-N0(-C3)-C0=O0 Hydroxide 8.5 # estimated

# sulfonamides
#   aromatic or alkene
Acid N0(-H0)(-C2=C2)-S0(=O0)(=O0)-C0 N-minus 8.3
#   pyridyl
Acid N0(-H0)(-C2=N2)-S0(=O0)(=O0)-C0 N-minus 8.3
#   carbonyl
Acid N0(-H0)(-C2=O2)-S0(=O0)(=O0)-C0 N-minus 4.5
#   alkyl
Acid N0(-H0)(-C3)-S0(=O0)(=O0)-C0 N-minus 11.6
#   Chlorine
Acid N0(-H0)(-Cl)-S0(=O0)(=O0)-C0 N-minus 4.5
# for carbonic anhydrase
#Acid N0(-H0)(-H0)-S0(=O0)(=O0)-C0 N-minus 8.3

# tetrazole
Acid N0(-H0)-N0=N0-N0=C2-1 N-minus 4.5
Acid N0(-H0)-N0=N0-C2=N0-1 N-minus 4.5

# phenol
Acid O3(-H0)-C2*C2(-00)*C2(-00)*C2(-00)*C2(-00)*C2(-00)*3 Hydroxide 10.0

# Regarding the phenol pattern, one co-worker said --
#
# "The phenol pattern above correctly works with most of my tested
# phenols and non-phenols. It is great for single carbon aromatic
# rings, and avoids greedy matches with compounds it should not.
# However, the pattern above does not match some canonical resonance
# structures of polycyclic benzenoid aromatics."
#
# "My opinion is that the pattern is pretty good, and certainly
# better than not having it at all. The polycyclic benzenoid
# compounds are vexing."

# 2-nitrosophenol/2-nitrophenol
Acid O0(-H0)-C2*C2(-N0=O0)*C2(-00)*C2(-00)*C2(-00)*C2(-00)*3 Hydroxide 6.5

# 4-nitrosophenol/4-nitrophenol
Acid O0(-H0)-C2*C2(-00)*C2(-00)*C2(-N0=O0)*C2(-00)*C2(-00)*3 Hydroxide 6.5

# 3,5-dinitrosophenol/3,5-dinitrophenol
Acid O0(-H0)-C2*C2(-00)*C2(-N0=O0)*C2(-00)*C2(-N0=O0)*C2(-00)*3 Hydroxide 6.7
```

```
# alkylthiol
Acid S0(-H0)-C3 Thiolate 9.5

# thiol on thiophene  ring
Acid S0(-H0)-C2*C2*S0*C2*C2*3 Thiolate 6.6

# thiophenol
Acid S0(-H0)-C2*C2(-00)*C2(-00)*C2(-00)*C2(-00)*C2(-00)*3 Thiolate 6.6
Acid S0(-H0)-C2*N2*C2(-00)*C2(-00)*C2(-00)*C2(-00)*3 Thiolate 6.6
Acid S0(-H0)-C2*C2(-00)*N2*C2(-00)*C2(-00)*C2(-00)*3 Thiolate 6.6
Acid S0(-H0)-C2*C2(-00)*C2(-00)*N2*C2(-00)*C2(-00)*3 Thiolate 6.6
Acid S0(-H0)-C2*N2*C2(-00)*N2*C2(-00)*C2(-00)*3 Thiolate 6.6

# Don't match sulfinic acid tautomer
Exclude S0(-H0)(=O0)(=O0)

#################################################################
# In many ring patterns, both above and below here, note the
# use of *C2(-00)*, where -00 (two zeroes) signifies single
# bond to any atom. This restricts the matches to only those
# substructures with aromatic carbons. We have to do this in
# light of our atom types. As one co-worker explained --
#
# "[B]ecause of our broad definition of the aromatic C (*C2*),
# our patterns would [otherwise] match quinone or uracyl type
# compounds. Hence we [...] have to impose a single-bonded
# substituent on every aromatic C to distinguish between them
# and carbonyl type C2."
#################################################################

#################################################################
# Specs for Bases, to be protonated

# We require Base pattern leading Nitrogens to be uncharged,
# to ignore some input molecules' N+ atoms otherwise matching
# these patterns; hence the "{0}" qualifiers, which are the
# mmsubs-extension syntax for zero formal charge.


# dialkylaniline
Base N0{0}(-C3)(-C3)-C2*C2(-00)*C2(-00)*C2(-00)*C2(-00)*C2(-00)*4 Ammonium 4.5

# amine
Base N0{0}(-H0)(-H0)-C3 Ammonium 10.5 # primary
Base N0{0}(-H0)(-C3)-C3 Ammonium 11.0 # secondary
```

```
Base N0{0}(-C3)(-C3)-C3 Ammonium 10.0 # tertiary

# hydrazine
Base N0{0}(-H0)(-H0)-N3 Ammonium 10.0 # primary
Base N0{0}(-H0)(-C3)-N3 Ammonium 10.0 # secondary
Base N0{0}(-C3)(-C3)-N3 Ammonium 10.0 # tertiary

# imine
Base N0{0}(-H0)=C2(-H0)-C3 Ammonium 11.5
Base N0{0}(-H0)=C2(-C3)-C3 Ammonium 11.5
Base N0{0}(-C3)=C2(-H0)-C3 Ammonium 11.5
Base N0{0}(-C3)=C2(-C3)-C3 Ammonium 11.5

# amidine
Base N0{0}(-H0)=C2(-H0)-N0-H0 Ammonium 12.0
Base N0{0}(-H0)=C2(-H0)-N0-C3 Ammonium 12.0
Base N0{0}(-H0)=C2(-C0)-N0-H0 Ammonium 12.0
Base N0{0}(-H0)=C2(-C0)-N0-C3 Ammonium 12.0
Base N0{0}(-C3)=C2(-H0)-N0-H0 Ammonium 12.0
Base N0{0}(-C3)=C2(-H0)-N0-C3 Ammonium 12.0
Base N0{0}(-C3)=C2(-C0)-N0-H0 Ammonium 12.0
Base N0{0}(-C3)=C2(-C0)-N0-C3 Ammonium 12.0

# guanidine
Base N0{0}(-H0)=C2(-N0-H0)-N0-H0 Ammonium 12.5
Base N0{0}(-H0)=C2(-N0-H0)-N0-C3 Ammonium 12.5
Base N0{0}(-H0)=C2(-N0-C3)-N0-H0 Ammonium 12.5
Base N0{0}(-H0)=C2(-N0-C3)-N0-C3 Ammonium 12.5
Base N0{0}(-C3)=C2(-N0-H0)-N0-H0 Ammonium 12.5
Base N0{0}(-C3)=C2(-N0-H0)-N0-C3 Ammonium 12.5
Base N0{0}(-C3)=C2(-N0-C3)-N0-H0 Ammonium 12.5
Base N0{0}(-C3)=C2(-N0-C3)-N0-C3 Ammonium 12.5

# enamine
Base N0{0}(-H0)(-H0)-C2(-H0)=C2 Ammonium 10.5 # primary
Base N0{0}(-H0)(-H0)-C2(-C3)=C2 Ammonium 10.5 # primary
Base N0{0}(-H0)(-C3)-C2(-H0)=C2 Ammonium 10.5 # primary
Base N0{0}(-H0)(-C3)-C2(-C3)=C2 Ammonium 10.5 # primary
Base N0{0}(-C3)(-C3)-C2(-H0)=C2 Ammonium 10.5 # primary
Base N0{0}(-C3)(-C3)-C2(-C3)=C2 Ammonium 10.5 # primary
#
# Screen out some matches on enamines above --
#
Exclude         N0{0}(-H0)(-H0)-C2=C2-C2=O0
Exclude         N0{0}(-H0)(-C3)-C2=C2-C2=O0
Exclude         N0{0}(-C3)(-C3)-C2=C2-C2=O0
```

```
#
Exclude         N0{0}(-H0)(-H0)-C2=C2-C1%N0
Exclude         N0{0}(-H0)(-C3)-C2=C2-C1%N0
Exclude         N0{0}(-C3)(-C3)-C2=C2-C1%N0


# aniline
Base  N0{0}(-H0)(-H0)-C2*C2(-00)*C2(-00)*C2(-00)*C2(-00)*C2(-00)*4  Ammonium  4.5  #
primary
Base  N0{0}(-H0)(-C0)-C2*C2(-00)*C2(-00)*C2(-00)*C2(-00)*C2(-00)*4  Ammonium  4.5  #
secondary
#the problem with the rule above is that compounds also get protonated, if the N is part of a ring
Exclude N0*C2*C2*C2*C2*1          #anniliertes pyrrole
Exclude N0*C2*N0*C2*C2*1          #anniliertes imidazole
Exclude N0*N0*C2*C2*C2*1          #anniliertes pyrazole (?)
#Exclude F or Cl sub.
Exclude N0(-H0)(-00)-C2*C2(-F0)*C2(-00)*C2(-00)*C2(-00)*C2(-00)*4    # F or Cl substiuent
Exclude N0(-H0)(-00)-C2*C2(-Cl)*C2(-00)*C2(-00)*C2(-00)*C2(-00)*4    # F or Cl substiuent
Exclude N0(-H0)(-00)-C2*C2(-00)*C2(-F0)*C2(-00)*C2(-00)*C2(-00)*4    # F or Cl substiuent
Exclude N0(-H0)(-00)-C2*C2(-00)*C2(-Cl)*C2(-00)*C2(-00)*C2(-00)*4    # F or Cl substiuent
Exclude N0(-H0)(-00)-C2*C2(-00)*C2(-00)*C2(-F0)*C2(-00)*C2(-00)*4    # F or Cl substiuent
Exclude N0(-H0)(-00)-C2*C2(-00)*C2(-00)*C2(-Cl)*C2(-00)*C2(-00)*4    # F or Cl substiuent
Exclude N0(-H0)(-00)-C2*C2(-00)*C2(-00)*C2(-00)*C2(-F0)*C2(-00)*4    # F or Cl substiuent
Exclude N0(-H0)(-00)-C2*C2(-00)*C2(-00)*C2(-00)*C2(-Cl)*C2(-00)*4    # F or Cl substiuent
Exclude N0(-H0)(-00)-C2*C2(-00)*C2(-00)*C2(-00)*C2(-00)*C2(-F0)*4    # F or Cl substiuent
Exclude N0(-H0)(-00)-C2*C2(-00)*C2(-00)*C2(-00)*C2(-00)*C2(-Cl)*4    # F or Cl substiuent


Base  N0{0}(-C0)(-C0)-C2*C2(-00)*C2(-00)*C2(-00)*C2(-00)*C2(-00)*4  Ammonium  4.5  #
tertiary

# 1,4-diaminobenzene
Base            N0{0}(-H0)(-H0)-C2*C2(-00)*C2(-00)*C2(-N0(-H0)(-H0))*C2(-00)*C2(-00)*4
Ammonium 6.2
Base            N0{0}(-H0)(-H0)-C2*C2(-00)*C2(-00)*C2(-N0(-H0)(-C3))*C2(-00)*C2(-00)*4
Ammonium 6.2
Base            N0{0}(-H0)(-H0)-C2*C2(-00)*C2(-00)*C2(-N0(-C3)(-C3))*C2(-00)*C2(-00)*4
Ammonium 6.2
Base            N0{0}(-H0)(-C3)-C2*C2(-00)*C2(-00)*C2(-N0(-H0)(-H0))*C2(-00)*C2(-00)*4
Ammonium 6.0
Base            N0{0}(-H0)(-C3)-C2*C2(-00)*C2(-00)*C2(-N0(-H0)(-C3))*C2(-00)*C2(-00)*4
Ammonium 6.0
Base            N0{0}(-H0)(-C3)-C2*C2(-00)*C2(-00)*C2(-N0(-C3)(-C3))*C2(-00)*C2(-00)*4
Ammonium 6.0
Base            N0{0}(-C3)(-C3)-C2*C2(-00)*C2(-00)*C2(-N0(-H0)(-H0))*C2(-00)*C2(-00)*4
Ammonium 6.0
```

Base N0{0}(-C3)(-C3)-C2*C2(-00)*C2(-00)*C2(-N0(-H0)(-C3))*C2(-00)*C2(-00)*4 Ammonium 6.0
Base N0{0}(-C3)(-C3)-C2*C2(-00)*C2(-00)*C2(-N0(-C3)(-C3))*C2(-00)*C2(-00)*4 Ammonium 6.0

N-heterocycles....

imidazole
Base N0{0}=C2-N0-C2=C2-1 Ammonium 7.0
#RB no Cl or F substiuents
Exclude N0=C2(-F0)-N0-C2=C2-1
Exclude N0=C2(-Cl)-N0-C2=C2-1
Exclude N0=C2-N0-C2=C2(-F0)-1
Exclude N0=C2-N0-C2=C2(-Cl)-1
Exclude N0=C2-N0-C2(-F0)=C2-1
Exclude N0=C2-N0-C2(-Cl)=C2-1

pyridine
Base N2{0}*C2(-00)*C2(-00)*C2(-00)*C2(-00)*C2(-00)*1 Ammonium 5.5

#3-aminopyridine has got an pKa of 6.2 in ACD labs
#RB
Base N2{0}*C2(-00)*C2(-N0(-H0)(-H0))*C2(-00)*C2(-00)*C2(-00)*1 Ammonium 6.2

pyridine with C=O, C=C or halide
Exclude N2*C2(-F0)*C2*C2*C2*C2*1
Exclude N2*C2(-Cl)*C2*C2*C2*C2*1
Exclude N2*C2*C2(-F0)*C2*C2*C2*1
Exclude N2*C2*C2(-Cl)*C2*C2*C2*1
Exclude N2*C2*C2*C2(-F0)*C2*C2*1
Exclude N2*C2*C2*C2(-Cl)*C2*C2*1
Exclude N2*C2(*C2=O2)*C2*C2*C2*C2*1
Exclude N2*C2*C2(*C2=O2)*C2*C2*C2*1
Exclude N2*C2*C2*C2(*C2=O2)*C2*C2*1
Exclude N2*C2(*C2=N2)*C2*C2*C2*C2*1
Exclude N2*C2*C2(*C2=N2)*C2*C2*C2*1
Exclude N2*C2*C2*C2(*C2=N2)*C2*C2*1

2-aminopyridine (ACD labs says 6.7) (they should still be protonated at pH 4.5, even with F or Cl subs.)
Base N2{0}*C2(-N0(-H0)(-H0))*C2(-00)*C2(-00)*C2(-00)*C2(-00)*1 Ammonium 7.5
Base N2{0}*C2(-N0(-H0)(-C3))*C2(-00)*C2(-00)*C2(-00)*C2(-00)*1 Ammonium 7.5
Base N2{0}*C2(-N0(-C3)(-C3))*C2(-00)*C2(-00)*C2(-00)*C2(-00)*1 Ammonium 7.5

4-aminopyridine (they should still be protonated at pH 4.5, even with F or Cl subs. the better way to go would probably to define diffrent pkas
#for the subtituet ones)
Base N2{0}*C2(-00)*C2(-00)*C2(-N0(-H0)(-H0))*C2(-00)*C2(-00)*1 Ammonium 9.0
Base N2{0}*C2(-00)*C2(-00)*C2(-N0(-H0)(-C3))*C2(-00)*C2(-00)*1 Ammonium 9.0
Base N2{0}*C2(-00)*C2(-00)*C2(-N0(-C3)(-C3))*C2(-00)*C2(-00)*1 Ammonium 9.0

4-methoxypyridine
Base N2{0}*C2(-00)*C2(-00)*C2(-O0-C3)*C2(-00)*C2(-00)*1 Ammonium 6.5

new rules for 3-aminopyridazine (pyridazine itself has got a pKa of about 2.5)
Base N2{0}*N2*C2(-N0(-H0)(-H0))*C2*C2*C2*1 Ammonium 5.0
#No substituentns on the N-N pattern
Exclude N2(*00)*N2*C2(-N0(-H0)(-H0))*C2*C2*C2*1
Exclude N2*N2(*00)*C2(-N0(-H0)(-H0))*C2*C2*C2*1
#Exclude 3,5, Diaminopyridazines, since they would be again covered by the following rule (RB)
Exclude N2*N2*C2(-N0(-H0)(-H0))*C2*C2(-N0(-H0)(-H0))*C2*1

new rules for 4-aminopyridazine (pyridazine itself has got a pKa of about 2.5)
Base N2{0}*N2*C2*C2(-N0(-H0)(-H0))*C2*C2*1 Ammonium 6.5
Exclude N2(*00)*N2*C2*C2(-N0(-H0)(-H0))*C2*C2*1
Exclude N2*N2(*00)*C2*C2(-N0(-H0)(-H0))*C2*C2*1

new rules for 2-aminopyrazine (pyrazine itself has got a pKa of about -0.5)
Base N2{0}*C2(-N0(-H0)(-H0))*C2*N2*C2*C2*1 Ammonium 3.0

#new rule for 4-aminopyrimidne
Base N2{0}*C2(-00)*N0*C2(-N0(-H0)(-H0))*C2*C2*1 Ammonium 5.5

new rule for 2,4-aminopyrimidine (2-aminopyrimidine has got a pKa of about 2.6)
Base N2{0}*C2(-N0(-H0)(-H0))*N2*C2(-N0(-H0)(-H0))*C2*C2*1 Ammonium 6.5

#exclude pyrimidines with a Cl or F substituent (RB)
Exclude N2*C2(-F0)*N0*C2(-N0(-H0)(-H0))*C2*C2*1
Exclude N2*C2(-Cl)*N0*C2(-N0(-H0)(-H0))*C2*C2*1
Exclude N2*C2*N0*C2(-N0(-H0)(-H0))*C2(-F0)*C2*1
Exclude N2*C2*N0*C2(-N0(-H0)(-H0))*C2(-Cl)*C2*1
Exclude N2*C2*N0*C2(-N0(-H0)(-H0))*C2*C2(-F0)*1
Exclude N2*C2*N0*C2(-N0(-H0)(-H0))*C2*C2(-Cl)*1

#exclude all 6-membered N-containing rings with more than 1 F or Cl (RB)
must update. currently just F
Exclude N0*00(-F0)*00(-F0)*00*00*00*1
Exclude N0*00(-F0)*00(-F0)*00*00*00*1

```
Exclude N0*00(-F0)*00*00(-F0)*00*00*1
Exclude N0*00(-F0)*00*00*00(-F0)*00*1
Exclude N0*00(-F0)*00*00*00*00(-F0)*1
Exclude N0*00*00(-F0)*00(-F0)*00*00*1
Exclude N0*00*00(-F0)*00*00(-F0)*00*1



#new rule for 2amino-thiazole
Base N2{0}*C2(-N0(-H0)(-H0))*S0*C2(-00)*C2(-00)*1 Ammonium 5.5


# purine
# (supersedes match on imidazole above) and other heterocycles (new placement, RB)
#Exclude N0=C2-N0-C2*N0*C2(-00)*N0*C2*C2(-1)(*5)
#Exclude N2*C2*N2*C2*C2*1
Exclude N0*C0*N0*C0*N0*C0*N0*C0*C0(*1)(*4)
#Why does this rule not get applied?????


# benzimidazole family
Exclude N2{0}*C2*N2*C2*C2(*02*02*02*02*4)*1

#####################################################################
# Alternative method for not matching N+ atoms
#
# Don't specify "{0}" in the Base N* patterns above, but
# then, after all the Base N* patterns, Exclude matches on
# Nitrogens with total bond order 4, which must be N+ atoms.
#
# With "{0}" qualifiers still in place above, the following
# is redundant, that is, it has no effect.
#
# Because the following patterns are so simple, they may hit
# some Nitrogens not matched by the Base patterns above. Any
# such matches are simply ignored, since there are no prior
# matches to undo.

Exclude N0(-00)(-00)(-00)(-00)    # 4 single bonds
Exclude N0(-00)(-00)(=00)         # 2 single + 1 double
Exclude N0(=00)(=00)              # 2 double
Exclude N0(-00)(%00)              # 1 single + 1 triple

# exclude amides
Exclude  N0*C0-O0-H0          # one aromatic, one OH
Exclude  N0-C0=O0           # one aromatic, one OH
```

aliphatic ethylenediamines
the real pKas for piperazine (most frequent) are 9.83 and 5.56
the real pKas for ethylenediamine are 10.81 and 6.99
but we can return the monoprotonated forms by giving an average value
Base N3{0}(-H0)(-H0)-C3-C3-N3(-H0)(-H0) Ammonium 7.7
Base N3{0}(-H0)(-C3)-C3-C3-N3(-H0)(-H0) Ammonium 7.7
Base N3{0}(-C3)(-C3)-C3-C3-N3(-H0)(-H0) Ammonium 7.7
Base N3{0}(-H0)(-H0)-C3-C3-N3(-H0)(-C3) Ammonium 7.7
Base N3{0}(-H0)(-C3)-C3-C3-N3(-H0)(-C3) Ammonium 7.7
Base N3{0}(-C3)(-C3)-C3-C3-N3(-H0)(-C3) Ammonium 7.7
Base N3{0}(-H0)(-H0)-C3-C3-N3(-C3)(-C3) Ammonium 7.7
Base N3{0}(-H0)(-C3)-C3-C3-N3(-C3)(-C3) Ammonium 7.7
Base N3{0}(-C3)(-C3)-C3-C3-N3(-C3)(-C3) Ammonium 7.7
for explicit quaternary ammonium we can drop the second amine further
Base N3{0}(-C3)(-C3)-C3-C3-N5{1}(-00)(-00)(-00) Ammonium 6.0
Base N3{0}(-H0)(-C3)-C3-C3-N5{1}(-00)(-00)(-00) Ammonium 6.0
Base N3{0}(-H0)(-H0)-C3-C3-N5{1}(-00)(-00)(-00) Ammonium 6.0

##
Please do not alter or remove the comments below capturing
the RCS revision info on this data file! The Ionizer's
handling of the -v|-ver|-version option expects to find
these here, in exactly this form --
#
VERSION $Revision: 1.40 $
VERSION $Date: 2004/02/06 19:46:18 $

instructions:
Within a set of tautomers
Covalent bonds involving Hydrogen that are moved in any tautomer must be represented
explicitly and proceeded by a -.
Hydrogens that move must be respresented using a [#1]
Bonds that change in any tautomer must be represented explicitly.
Heavy atoms must appear in the same order in all tautomers.
All instances of asymmetric tautomers must be represented
(see for example double-sided_ketol-enol).

larger matches have precedence
earlier matches of the same size have precedence

for changing in bonding for aromatic N's (n)

```
# all bond orders which should change need to be represented by ':'
# see imidazole for an example

tautomer_set{
   name: imidazole

   tautomer{
     name: form1
     pattern: c1n(-[#1])-c=[nX2]c1
     probability: 0.50
   }

   tautomer{
     name: form2
     pattern: c1[nX2]=c-n(-[#1])c1
     probability: 0.50
   }
}

tautomer_set{
   name: double-ket-enol
# may want to make special allowances for outside enol (C=C outside
# central region)
# From: Handbook of organic chemistry

   tautomer{
     name: enol_ket
     pattern: [CX3](-[*;!#8;!#7])(=[CX3]([*;!#8;!#7])-O-[#1])-[CX3]([*;!#8;!#7])=O
     probability: 0.381
   }

   tautomer{
     name: diket
     pattern: C(-[#1])(-[*;!#8;!#7])(-[CX3]([*;!#8;!#7])=O)-[CX3]([*;!#8;!#7])=O
     probability: 0.238
   }

   tautomer{
     name: ket_enol
     pattern: [CX3](-[*;!#8;!#7])(-[CX3]([*;!#8;!#7])=O)=[CX3]([*;!#8;!#7])-O-[#1]
     probability: 0.381
   }
}

tautomer_set{
   name: single-sided_ket-enol
```

```
# From: Handbook of organic chemistry

   tautomer{
     name: enol
     pattern: [CX3]=[CX3]([*;!#8;!#7])-O-[#1]
     probability: 0.00005
   }

   tautomer{
     name: ket
     pattern: [CX4](-[#1])-[CX3]([*;!#8;!#7])=O
     probability: 0.99995
   }
}

tautomer_set{
   name: double-sided_ket-enol
# From: Handbook of organic chemistry

   tautomer{
     name: 1enol
     pattern: [CX3]=[CX3](-[CX4]-[#1])-O-[#1]
     probability: 0.00000001
   }

   tautomer{
     name: ket
     pattern: [CX4](-[#1])-[CX3](-[CX4]-[#1])=O
     probability: .99999998
   }

   tautomer{
     name: 2enol
     pattern: [CX4](-[#1])-[CX3](=[CX3])-O-[#1]
     probability: 0.00000001
   }
}

tautomer_set{
   name: single-sided_thio_ket-enol

   tautomer{
     name: thiol
     pattern: [CX3]=[CX3]([*;!#8;!#7])-S-[#1]
     probability: 0.00005
   }
```

```
    tautomer{
      name: thio-acetal
      pattern: [CX4](-[#1])-[CX3]([*;!#8;!#7])=S
      probability: 0.99995
    }
}

tautomer_set{
    name: double-sided_thio_ket-enol

    tautomer{
      name: 1thiol
      pattern: [CX3]=[CX3](-[CX4]-[#1])-S-[#1]
      probability: 0.00005
    }

    tautomer{
      name: thio-ketone
      pattern: [CX4](-[#1])-[CX3](-[CX4]-[#1])=S
      probability: 0.9999
    }

    tautomer{
      name: 2thiol
      pattern: [CX4](-[#1])-[CX3](=[CX3])-S-[#1]
      probability: 0.00005
    }
}

tautomer_set{
    name: single-sided_N

    tautomer{
      name: enol
      pattern: [CX3]=[CX3]([*;!#8;!#7])-[NX3](-[CX4,#1])-[#1]
      probability: 0.000014
    }

    tautomer{
      name: ket
      pattern: [CX4](-[#1])-[CX3]([*;!#8;!#7])=[NX2](-[CX4,#1])
      probability: 0.999986
    }
}
```

```
tautomer_set{
   name: double-sided_N

   tautomer{
     name: 1enol
     pattern: [CX3]=[CX3](-[CX4]-[#1])-[NX3](-[CX4,#1])-[#1]
     probability: 0.000008
   }

   tautomer{
     name: ket
     pattern: [CX4](-[#1])-[CX3](-[CX4]-[#1])=[NX2]-[CX4,#1]
     probability: 0.999984
   }

   tautomer{
     name: 2enol
     pattern: [CX4](-[#1])-[CX3](=[CX3])-[NX3](-[CX4,#1])-[#1]
     probability: 0.000008
   }
}

tautomer_set{
   name: diethenylamine

   tautomer{
     name: ket1
     pattern: [CX3]=[CX3]([*;!#8;!#7])-[NX2]=[CX3]-[CX4]-[#1]
     probability: 0.499993
   }

   tautomer{
     name: enol
     pattern: [CX3]=[CX3]([*;!#8;!#7])-[NX3](-[CX3]=[CX3])-[#1]
     probability: 0.000014
   }

   tautomer{
     name: ket2
     pattern: [CX4](-[#1])-[CX3]([*;!#8;!#7])=[NX2]-[CX3]=[CX3]
     probability: 0.499993
   }
}


tautomer_set{
```

```
   name: 1,2-diazole

   tautomer{
     name: 1H
     pattern: c-1=c-c=[nX2]-[nX3](-[#1])-1
     probability: 0.50
   }
   tautomer{
     name: 2H
     pattern: c=1-c=c-[nX3](-[#1])-[nX2]=1
     probability: 0.50
   }
}


tautomer_set{
   name: triazole

   tautomer{
     name: 1H
     pattern: c-1=c-[nX2]=[nX2]-[nX3](-[#1])-1
     probability: 0.25
   }
   tautomer{
     name: 2H
     pattern: c=1-c=[nX2]-[nX3](-[#1])-[nX2]=1
     probability: 0.5
   }
   tautomer{
     name: 3H
     pattern: c-1=c-[nX3](-[#1])-[nX2]=[nX2]-1
     probability: 0.25
   }
}

tautomer_set{
   name: tetraazole

   tautomer{
     name: 1H
     pattern: c=1-[nX3](-[#1])-[nX2]=[nX2]-[nX2]=1
     probability: 0.80
   }

   tautomer{
     name: 2H
```

```
    pattern: c-1=[nX2]-[nX3](-[#1])-[nX2]=[nX2]-1
    probability: 0.20
  }

# Disabled since these are equivalent due to symmetry
#   tautomer{
#     name: 3H
#     pattern: c=1-[nX2]=[nX2]-[nX3](-[#1])-[nX2]=1
#     probability: 0.10
#   }

#   tautomer{
#     name: 4H
#     pattern: c-1=[nX2]-[nX2]=[nX2]-[nX3](-[#1])-1
#     probability: 0.40
#   }
}

#tautomer_set{
#   name: 2,4-oxy-pyridine
# see if non-one or hydroxy c can be changed to a
#   tautomer{
#     name: one
#     pattern: [nX3]-1(-[#1])-c(=O)-[a]=[a]-n=[a]-1
#     pattern: [nX3]-1(-[#1])-c(=O)-c=c-c=c-1
#     probability: 0.50
#   }
#   tautomer{
#     name: two
    #pattern: [nX2]=1-c(=O)-[a]=[a]-[nX3](-[#1])-[a]=1
#     pattern: [nX3]-1(-[#1])-c(=O)-c=c-c=c-1
    #probability: 0.30
  #}
  #tautomer{
  #  name: hydroxy
  #  pattern: [nX2]=1-c(-O-[#1])=[a]-[a]=n-[a]=1
#     pattern: [nX2]=1-c(-O-[#1])=c-c=c-c=1
  #  probability: 0.20
  #}
#}

tautomer_set{
  name: 4-oxy-pyridine
  tautomer{
    name: one
    pattern: [nX3]-1(-[#1])-[a]=[a]-c(=O)-[a]=[a]-1
```

```
    probability: 0.80
  }
  tautomer{
    name: hydroxy
    pattern: [nX2]=1-[a]=[a]-c(-O-[#1])=[a]-[a]=1
    probability: 0.20
  }
}
tautomer_set{
  name: 2-oxy-pyridine
# see if non-one or hydroxy c can be changed to a
  tautomer{
    name: one
    pattern: [nX3]-1(-[#1])-c(=O)-[a]=[a]-[a]=[a]-1
#     pattern: [nX3]-1(-[#1])-c(=O)-c=c-c=c-1
    probability: 0.80
  }
  tautomer{
    name: hydroxy
    pattern: [nX2]=1-c(-O-[#1])=[a]-[a]=[a]-[a]=1
#     pattern: [nX2]=1-c(-O-[#1])=c-c=c-c=1
    probability: 0.20
  }
}

# disabled because not all forms interconvert readily
#tautomer_set{
#   name: buten_aldehyde
#
#   tautomer{
#     name: but-3-en-aldehyde
#     pattern: [CX3](=O)-[CX4](-[#1])-[CX3]=[CX3](!#8)!#8
#     probability: 0.005
#   }
#   tautomer{
#     name: but-2-en-aldehyde
#     pattern: [CX3](=O)-[CX3]=[CX3]-[CX4](-[#1])(!#8)!#8
#     probability: 0.99
#   }
#   tautomer{
#     name: but-1,3-dien-1-ol
#     pattern: [CX3](-O-[#1])=[CX3]-[CX3]=[CX3](!#8)!#8
#     probability: 0.005
#   }
#}
```

```
# disabled because not all forms interconvert readily
#tautomer_set{
#    name: penten_aldehyde
#
#    tautomer{
#      name: pent-3-en-aldehyde
#      pattern: [CX3](=O)-[CX4](-[#1])-[CX3]=[CX3]-[CX3]=O
#      probability: 0.45
#    }
#    tautomer{
#      name: pent-3-en-1-aldehyde-5-ol
#      pattern: [CX3](-O-[#1])=[CX3]-[CX3]=[CX3]-[CX3]=O
#      probability: 0.05
#    }
#    tautomer{
#      name: pent-3-en-aldehyde
#      pattern: [CX3](=O)-[CX3]=[CX3]-[CX4](-[#1])-[CX3]=O
#      probability: 0.45
#    }
#    tautomer{
#      name: pent-3-en-aldehyde
#      pattern: [CX3](=O)-[CX3]=[CX3]-[CX3]=[CX3]-O-[#1]
#      probability: 0.05
#    }
#}

tautomer_set{
   name: guanine
# Tuatomer naming from Hanus et a. JACS 125, 7678 (2003)
# Their frame work atom numbering:
# (bond orders left out)
#
#             O
#             |
#       N     C
#     /   \ /   \
#    C     C     N
#    |     |     |
#    N-----C     C
#           \   / \
#             N    N
#
#
#             10
#             |
#       7     6
```

```
#     /  \/  \
#    8    5    1
#    |    |    |
#    9-----4    2
#          \  / \
#           3   N
# So that tautomer names equivalences are:
# Used Here      Hanus et al.
#  1-5          1-9
#  1-7          1-7
#  3-7          3-7
#  3-5          3-9
#  5-7          7-9
#  5-O          9-O
#  7-O          7-O
   tautomer{
    name: 1H,9H,2-amino,6-one
    pattern: [nX3]-1(-[#1])-c(-[NX3]-[#1])=[nX2]-c=2[nX3](-[#1])-c=[nX2]c=2-c-1=O
    probability: 0.7781
   }
   tautomer{
    name: 1H,7H,2-amino,6-one
    pattern: [nX3]-1(-[#1])-c(-[NX3]-[#1])=[nX2]-c=2[nX2]=c-[nX3](-[#1])c=2-c-1=O
    probability: 0.2194
   }
   tautomer{
    name: 3H,7H,2-amino,6-one
    pattern: [nX2]-1=c(-[NX3]-[#1])-[nX3](-[#1])-c=2[nX2]=c-[nX3](-[#1])c=2-c-1=O
    probability: 0.0025
   }
   tautomer{
    name: 3H,9H,2-amino,6-one
    pattern: [nX2]-1=c(-[NX3]-[#1])-[nX3](-[#1])-c=2[nX3](-[#1])-c=[nX2]c=2-c-1=O
    probability: 1.67e-05
   }
   tautomer{
    name: 9H,2-amino,6-hydroxy
    pattern: [nX2]-1=c(-[NX3]-[#1])-[nX2]=c-2[nX3](-[#1])-c=[nX2]c-2=c-1-O-[#1]
    probability: 2.39e-06
   }
   tautomer{
    name: 7H,2-amino,6-hydroxy
    pattern: [nX2]-1=c(-[NX3]-[#1])-[nX2]=c-2[nX2]=c-[nX3](-[#1])c-2=c-1-O-[#1]
    probability: 4.11e-7
   }
   tautomer{
```

```
    name: 1H,3H,7H,2-imino,6-one
    pattern: [nX3]-1(-[#1])-c(=[NX2])-[nX3](-[#1])-c=2[nX2]=c-[nX3](-[#1])c=2-c-1=O
    probability: 2.34e-7
  }
  tautomer{
    name: 1H,3H,9H,2-imino,6-one
    pattern: [nX3]-1(-[#1])-c(=[NX2])-[nX3](-[#1])-c=2[nX3](-[#1])-c=[nX2]c=2-c-1=O
    probability: 2.18e-8
  }
#   tautomer{
#     name: 1H,7H,2-imino,6-hydroxy
#     pattern: [nX3]-1(-[#1])-c(=[NX2])-[nX2]=c-2[nX2]=c-[nX3](-[#1])c-2=c-1-O-[#1]
#     probability:  2.63e-19
#   }

}

tautomer_set{
  name: guanosine
# Actually a methyl is used in place of the sugar which is OK for
# tautomers.
# This is similar to guanine except that there is one less H atom
# to move around and position 9 is blocked by the methyl
#
#
#           10
#           |
#       7    6
#     /  \/  \
#    8    5    1
#    |    |    |
#    9-----4    2
#   /      \  / \
#  Me        3   N

# based upon guanine

  tautomer{
    name: 1H,2-amino,6-one
    pattern: [nX3]-1(-[#1])-c(-[NX3]-[#1])=[nX2]-c=2[nX3](-[!#1])-c=[nX2]c=2-c-1=O
    probability: 0.9999785
  }
  tautomer{
    name: 3H,2-amino,6-one
    pattern: [nX2]-1=c(-[NX3]-[#1])-[nX3](-[#1])-c=2[nX3](-[!#1])-c=[nX2]c=2-c-1=O
    probability: 2.15e-05
```

```
   }
   tautomer{
     name: 2-amino,6-hydroxy
     pattern: [nX2]-1=c(-[NX3]-[#1])-[nX2]=c-2[nX3](-[!#1])-c=[nX2]c-2=c-1-O-[#1]
     probability: 3.07e-06
   }
   tautomer{
     name: 1H,3H,2-imino,6-one
     pattern: [nX3]-1(-[#1])-c(=[NX2])-[nX3](-[#1])-c=2[nX3](-[!#1])-c=[nX2]c=2-c-1=O
     probability: 2.80e-8
   }
}

tautomer_set{
   name: uracil-thymine
# Here we use (not counting H's and the N side-chain):
#             O
#             ||
#             C
#           /   \
#          N     C
#          |     |
#          C     C
#         // \   /
#        O     N
#
#
#
#             O2
#             ||
#             4
#           /   \
#          3     5
#          |     |
#          2     6
#         // \   /
#        O1    1


   tautomer{
     name: 1H,3H,2,4-dione
     pattern: n-1(-[#1])-c(=O)-n(-[#1])-c(=O)-c=c-1
     probability: 0.999999707
   }

   tautomer{
```

```
    name: 1H,2-hydroxy,4-one
    pattern: n-1(-[#1])-c(-O-[#1])=[nX2]-c(=O)-c=c-1
    probability: 1.87e-10
  }

  tautomer{
    name: 1H,2-one,4-hydroxy
    pattern: n-1(-[#1])-c(=O)-[nX2]=c(-O-[#1])-c=c-1
    probability: 2.93e-7
  }

  tautomer{
    name: 3H,2-hydroxy,4-one
    pattern: [nX2]-1=c(-O-[#1])-n(-[#1])-c(=O)-c=c-1
    probability: 4.59e-9
  }

  tautomer{
    name: 2-hydroxy,4-hydroxy
    pattern: [nX2]-1=c(-O-[#1])-[nX2]=c(-O-[#1])-c=c-1
    probability: 4.72e-12
  }

  tautomer{
    name: 3H,2-one,4-hydroxy
    pattern: [nX2]=1-c(=O)-n(-[#1])-c(-O-[#1])=c-c=1
    probability: 1.09e-09
  }

}

tautomer_set{
  name: uridine-thymidine
# Actually a methyl is used in place of the sugar which is OK for
# tautomers.
# This is similar to uracil except that there is one less H atom to move
# around and position 1 is blocked by the methyl

  tautomer{
    name: 3H,2,4-dione
    pattern: n1(-[!#1])-c(=O)-n(-[#1])-c(=O)cc1
    probability: 0.99999724
  }

  tautomer{
    name: 2-hydroxy,4-one
```

```
    pattern: n1(-[!#1])-c(=O)-[nX2]=c(-O-[#1])cc1
    probability: 2.76e-6
  }

  tautomer{
    name: 2-one,4-hydroxy
    pattern: n1(-[!#1])-c(-O-[#1])=[nX2]-c(=O)cc1
    probability: 1.87e-10
  }
}

tautomer_set{
  name: cytosine
#Is this complete what about 1-3?
# Here we use (not counting H's and the N side-chain):
#            NH2
#            |
#            C
#          /   \
#         N     C
#         |     |
#         C     C
#        // \   /
#       O     N
#
#
#
#            N
#            |
#            4
#          /   \
#         3     5
#         |     |
#         2     6
#        // \   /
#       O     1

  tautomer{
    name: 1H,2-one,4-amino
    pattern: n-1(-[#1])-c(=O)-[nX2]=c(-[NX3]-[#1])-c=c-1
    probability: 0.996
  }

  tautomer{
    name: 3H,2-one,4-amino
    pattern: [nX2]=1-c(=O)-n(-[#1])-c(-[NX3]-[#1])=c-c=1
```

```
    probability: 0.00398
  }

  tautomer{
    name: 2-hydroxy,4-amino
    pattern: [nX2]-1=c(-O-[#1])-[nX2]=c(-[NX3]-[#1])-c=c-1
    probability: 5.63e-7
  }

  tautomer{
    name: 1H,3H,2-one,4-imino
    pattern: n-1(-[#1])-c(=O)-n(-[#1])-c(=[NX2])-c=c-1
    probability: 3.056e-6
  }

}


tautomer_set{
  name: cytidine

#probabilities based upon corresponding structures for
#cytosine
  tautomer{
    name: 2-one,4-amino
    pattern: n1(-[!#1])c(=O)-[nX2]=c(-[NX3]-[#1])cc1
    probability: 0.997
  }

  tautomer{
    name: 3H,2-one,4-imino
    pattern: n1(-[!#1])c(=O)-n(-[#1])-c(=[NX2])cc1
    probability: 3.056e-6
  }

}

tautomer_set{
  name: adenine

  tautomer{
    name: 9H,6-amino
    pattern: [nX2]=1-c=[nX2]c=2-[nX3](-[#1])-c=[nX2]c=2-c=1-[NX3]-[#1]
    probability: 0.911
  }
```

```
  tautomer{
    name: 7H,6-amino
    pattern: [nX2]=1-c=[nX2]c=2-[nX2]=c-[nX3](-[#1])c=2-c=1-[NX3]-[#1]
    probability: 0.084
  }

  tautomer{
    name: 3H,6-amino
    pattern: [nX2]-1=c-[nX3](-[#1])c-2=[nX2]-c=[nX2]c-2=c-1-[NX3]-[#1]
    probability: 0.0046
  }

  tautomer{
    name: 1H,6-amino
    pattern: [nX3]-1(-[#1])-c=[nX2]c-2=[nX2]-c=[nX2]c-2=c-1-[NX3]-[#1]
    probability: 0.00043
  }

  tautomer{
    name: 1H,7H,6-imino
    pattern: [nX3]-1(-[#1])-c=[nX2]c=2-[nX2]=c-[nX3](-[#1])c=2-c-1=[NX2]
    probability: 2.88e-7
  }

  tautomer{
    name: 1H,9H,6-imino
    pattern: [nX3]-1(-[#1])-c=[nX2]c=2-[nX3](-[#1])-c=[nX2]c=2-c-1=[NX2]
    probability: 1.01e-6
  }

  tautomer{
    name: 3H,7H,6-imino
    pattern: [nX2]-1=c-[nX3](-[#1])c=2-[nX3](-[#1])-c=[nX2]c=2-c-1=[NX2]
    probability: 8.75e-10
  }

  tautomer{
    name: 3H,9H,6-imino
    pattern: [nX2]-1=c-[nX3](-[#1])c=2-[nX2]=c-[nX3](-[#1])c=2-c-1=[NX2]
    probability: 7.16e-12
  }
}

tautomer_set{
  name: adenosine
```

```
# based upon adenine probabilities
   tautomer{
     name: 6-amino
     pattern: [nX2]=1-c=[nX2]c=2-[nX3](-[!#1])-c=[nX2]c=2-c=1-[NX3]-[#1]
     probability: 0.9999989
   }
   tautomer{
     name: 1H,6-imino
     pattern: [nX3]-1(-[#1])-c=[nX2]c=2-[nX3](-[!#1])-c=[nX2]c=2-c-1=[NX2]
     probability: 1.1e-6
   }
   tautomer{
     name: 3H,6-imino
     pattern: [nX2]-1=c-[nX3](-[#1])c=2-[nX3](-[!#1])-c=[nX2]c=2-c-1=[NX2]
     probability: 7.8e-12
   }

}


#tautomer_set{
#    name: isoxazol
#
##The frame work atom and numbering are below
## (Bond orders are left out)
##
##                O
##               / \
##              C   N
##                  |   |
##                  C-----C
##                   \
##                    O
##
##                3
##               /  \
##              2    4
##              |    |
##                  1-----5
##                   \
##                    O
##
##
##The names of tautomers are derived from ChemDraw
##
#    tautomer{
```

```
#     name: isoxazol-3-ol
#     pattern: c1=c-c(-O-[#1])=n-o1
#     probability: 0.5
#   }
#   tautomer{
#     name: isoxazol-3(2H)-one
#     pattern: c1=c-c(=O)-n(-[#1])-o1
#     probability: 0.5
#   }
#
#}
#
#tautomer_set{
#   name: Hydroxypyridazin
#
##The frame work atom and numbering are below
## (Bond orders are left out)
##
##              O
##                  |
##                  C
##             / \
##            C   N
##            |   |
##            C   N
##            \ /
##             C
##             |
##             O
##
##
##             O
##                 |
##                 3
##             / \
##            2   4
##            |   |
##            1   5
##             \ /
##              6
##              |
##              O
##
##The names of tautomers are derived from ChemDraw
##
#   tautomer{
```

```
#     name: 6-hydroxypyridazin-3(2H)-one
#     pattern: c1=c-c(-O-[#1])=[nX2]-[nX3](-[#1])-c(=O)1
#     probability: 0.33
#   }
#
#   tautomer{
#     name: 1,2-dihydropyridazine-3,6-dione
#     pattern: c1=c-c(=O)-[nX3](-[#1])-[nX3](-[#1])-c(=O)1
#     probability: 0.33
#   }
#
#   tautomer{
#     name: 1,2-dihydropyridazine-3,6-dione
#     pattern: c1=c-c(=O)-[nX3](-[#1])-[nX2]=c(-O-[#1])1
#     probability: 0.33
#   }
#
#
#}
#
#
#tautomer_set{
#    name: Isothiazol
#
##The frame work atom and numbering are below
## (Bond orders are left out)
##
##              N     O
##             /  \  /
##            S     C
##            |     |
##            C-----C
##
##
##
##               4   O
##             /  \ /
##            3     5
##                |     |
##                2-----1
###
##
##
###The names of tautomers are derived from ChemDraw
##
#
```

```
#    tautomer{
#      name:  Isothiazol-3(2H)-one
#      pattern: [cX3]1[cX3]s[nX3](-[#1])-[cX3]1(=O)
#      probability: 0.5
#    }
#
#    tautomer{
#      name: Isothiazol-3-ol
#      pattern:  [cX3]1[cX3]s[nX2]=[cX3]1(-O-[#1])
#      probability: 0.5
#    }
#
#}

#tautomer_set{
#    name: isoxazol
#
##The frame work atom and numbering are below
## (Bond orders are left out)
##
##              O    O
##             / \  /
##            N    C
##            |    |
##                 C----C
##
##
##
##               4       O
##             /    \  /
##            3      5
##            |      |
##                 2------1
##
##
##
##The names of tautomers are derived from ChemDraw
##
#
#    tautomer{
#      name: isoxazol-5(2H)-one
#      pattern: c-1=c-[nX3](-[#1])-o-c-1(=O)
#      probability: 0.33
#    }
#
#    tautomer{
```

```
#     name: Isoxazol-5-ol
#     pattern: c=1-c=[nX2]-o-[cX3]=1(-O-[#1])
#     probability: 0.33
#   }
#
#   tautomer{
#     name: Isoxazol-5(4H)-one
#     pattern: c-1(-[#1])-c=[nX2]-o-c-1(=O)
#     probability: 0.33
#   }
#
#}


tautomer_set{
   name:  3-O-pyrazole

#The frame work atom and numbering are below
# (Bond orders are left out)
#
#               C
#              /  \
#             N    C
#                  |    |
#                  N-----C
#                  \
#                   O
#
#
#The names of tautomers are derived from ChemDraw
#
   tautomer{
     name: 1H-pyrazol-3-ol
     pattern: c=1-c(-O-[#1])=[nX2]-[nX3](-[#1])-c=1
     probability: 0.690
   }

   tautomer{
     name: 1H,2H-pyrazol-3-one
     pattern: c=1-c(=O)-[nX3](-[#1])-[nX3](-[#1])-c=1
     probability: 0.237
   }

   tautomer{
     name: 4H-pyrazol-3-one
     pattern:  C-1(-[#1])-C(=O)-[NX3](-[#1])-[NX2]=C-1
```

```
    probability: 0.0716
  }

  tautomer{
    name: 2H-pyrazol-3-ol
    pattern: c-1=c(-O-[#1])-[nX3](-[#1])-[nX2]=c-1
    probability: 0.00124
  }

}
tautomer_set{
  name: pyrimidine
# (based on cytosine model)
#Is this complete what about 1-3?
# Here we use (not counting H's and the N side-chain):
#
#
#           C
#         /   \
#        N     C
#        |     |
#        C     C
#         \   /
#           N
#
#
#
  tautomer{
    name: 1H,2-one,4-amino
    pattern: [nX2]-1=c-[nX3+](-[#1])=c-c=c-1
    probability: 0.5
  }

  tautomer{
    name: 3H,2-one,4-amino
    pattern: [nX3+]-1(-[#1])=c-[nX2]=c-c=c-1
    probability: 0.5
  }
}
```

```
SP3-SP3   60.0 -60.0 180.0
SP2-SP2   0.0 180.0 -30.0 30.0 150.0 -150.0
SP3-SP2   0.0 30.0 -30.0 60.0 -60.0 120.0 -120.0 -150.0 150.0 180.0 -90.0 90.0
#the follow res works better for strained structures
#SP2-SP2   0.0 180.0 -30.0 30.0 150.0 -150.0 -60.0 60.0 -120.0 120.0

#polysaccaride bridges
```

```
O@[CD3]O[CD3]([#1])@C          2 3 4 5 0.0 30.0 -30.0 180.0
O@[CD3]([#1])O[CD3]([#1])@C    3 2 4 5 0.0 30.0 -30.0 180.0

#acids
[OD1]~C(~[OD1])[CX4](*)*    1 2 4 5  30.0 -30.0 -60.0 60.0 90.0 -90.0 0.0 180.0
[a]cC([OD1])=O              1 2 3 4  0.0 20.0 -20.0
[OD1]C(=O)[CD2]C      1 2 4 5  0.0 45.0 90.0

#sulfonamides
NS(=O)(=O)c1[cD2][cD2]a[cD2][cD2]1  1 2 5 6 90.0
c([aD2])S(=O)(=O)[ND2][CD2] 1 3 6 7   60.0 -60.0
O=S(=O)N[CX4D3]*           2 4 5 6  -90.0 90.0 120.0 -120.0
O=S(=O)N[CX4D2]*           2 4 5 6  -90.0 90.0 120.0 -120.0
[c]S(=O)(=O)NC             1 2 5 6   -70.0 70.0 90.0 -90.0 50.0 -50.0
*=*-S(=O)(=O)C             1 2 3 6  90.0 -90.0 60.0 -60.0
O=S(=O)N[CH2]              1 2 4 5  -60.0 60.0 180.0 0.0 30.0 -30.0
[aD2]c([aD2])S(=O)(=O)[ND2^3]   1 2 4 7 90.0 -90.0 120.0 -120.0 60.0 -60.0
[aD2]c([aD3])S(=O)(=O)[ND2^3]   1 2 4 7 80.0 -80.0 110.0 -110.0
[aD3]c([aD3])S(=O)(=O)[ND2^3]   1 2 4 7 70.0 -70.0 110.0 -110.0
[aD2]c([aD2])S(=O)(=O)[CD2^3]   1 2 4 7 90.0 -90.0 110.0 -110.0 70.0 -70.0
[a]cS(=O)(=O)[C,N]         1 2 3 6  60.0 -60.0 90.0 -90.0 120.0 -120.0 0.0 180.0
aS(=O)(=O)[ND2]a           1 2 5 6  90.0 -90.0
aaNS(=O)(=O)               1 2 3 4 90.0 -90.0

#sulfone
O=S(=O)[CD2][CD3][#1]        2 4 5 6 30.0 -30.0

#hydrazides
[O,S]=C[ND2][ND2]          1 2 3 4  0.0 180.0
[O,S]=C[ND2][ND2]-,=*       2 3 4 5  180.0 90.0 -90.0

#cyclopropyl-ketones
O=CC1([#1])[CD2][CD2]1      1 2 3 4  180.0
O=CC1([#1])CC1             1 2 3 4  180.0 160.0 -160.0 0.0 20.0 -20.0
O=CC1([*])CC1              1 2 3 4  180.0 160.0 -160.0 0.0 120.0 -120.0 90.0 -90.0 30.0 -30.0
#epoxy-ketone
O=C([*D2])C1([#1])O[CD2,CD3]1   1 2 4 5  0.0 180.0

#opposite end of tert amide
O=C([ND3])[CD2]*           1 2 4 5  0.0 30.0 -30.0 100.0 -100.0 80.0 -80.0
O=C([CD3^3])[CD2]*          1 2 4 5  0.0 30.0 -30.0
O=C([ND3])[CD3][#1]         1 2 4 5  180.0 150.0 -150.0 120.0 -120.0

#misc
[CD2]C(=O)[ND2]-!@[CD3][#1]    2 4 5 6  0.0 30.0 -30.0 60.0 -60.0 180.0
[cD2]c([cD2])-!@[CD2^3][CD3^3] 1 2 4 5  90.0 -90.0 70.0 -70.0 110.0 -110.0
```

```
c[CD2][ND3](C)c              1 2 3 4  90.0 -90.0 60.0 -60.0 120.0 -120.0

#carbonyls
O=CC=O                   1 2 3 4  180.0 0.0 120.0 -120.0 90.0 -90.0
C=CC=O                   1 2 3 4  0.0 180.0 20.0 -20.0 160.0 -160.0
O=C[CD2][ND2]              1 2 3 4  0.0 -30.0 30.0 150.0 -150.0 180.0
O=C[CD2]C=O              1 2 3 4  0.0 -30.0 30.0 60.0 -60.0 130.0 -130.0
O=C(c)[ND2][CD3][#1]       2 4 5 6  0.0 -30.0 30.0
O=C[ND2][CD3]*             2 3 4 5  20.0 -20.0 120 -120.0 60.0 -60.0 0.0
O=CN[CD2]*               2 3 4 5  180.0 150.0 -150.0 -120.0 120.0 0.0 30.0 -30.0
O=Ccc[OD1]               1 2 3 4  0.0 180.0 90.0 -90.0 30.0 -30.0
O=C[CD4][CD1]              1 2 3 4  0.0 30.0 -30.0 60.0 -60.0 120.0 -120.0
O=C[CD3][OD1]              1 2 3 4  0.0 30.0 -30.0 60.0 -60.0 120.0 -120.0
O=C[CD2][CD1]              1 2 3 4  0.0 30.0 -30.0 60.0 -60.0 90.0 -90.0 120.0 -120.0
O=C[CD3][#1]             1 2 3 4  0.0 30.0 -30.0 180.0
#O=C[CD3]*               1 2 3 4  0.0 90.0 -90.0 30.0 -30.0 -120.0 120.0 60.0 -60.0

#amidene and guanidine
[aD3]cC(~[ND1])~[ND1]          1 2 3 4  0.0 30.0
[a]cC(~[ND1])~[ND1]            1 2 3 4  0.0 30.0
*[ND2]~C(~[ND1])~[ND1]         1 2 3 4  0.0 30.0
[CD2][CD2][ND2]~C(~[ND1])~[ND1] 1 2 3 4  -70.0 70.0 90.0 -90.0 110.0 -110.0

#ether
aCO[CD2]                 1 2 3 4 180.0 100.0 -100.0

#isoprene
C=C[CX4D2]*              1 2 3 4  0.0 180.0 90.0 -90.0 60.0 -60.0 30.0 -30.0
C=Cc[a]                1 2 3 4  0.0 90.0 -90.0 180.0 30.0 -30.0 150.0 -150.0

#aryl secondary amines
[aD2]c([aD2])[ND2][CD2]    1 2 4 5  0.0 180.0
[aD2]c([aD3])[ND2][CD2]    1 2 4 5  0.0
[aD2]c([aD2])[ND2][CD1]    1 2 4 5  0.0 90.0 -90.0 180.0
ac[ND2][CD2]             1 2 3 4  90.0 -90.0 160.0 -160.0 20.0 -20.0

#aromatic subtituents
[aD3]c([aD3])[CD2]C          1 2 4 5  90.0 -90.0 60.0 -60.0 120.0 -120.0
[aD2]c([aD2])[ND3]([CD1])[CD2]  1 2 4 5  0.0 180.0
[aD3][c,n]([aD2])[C^3D3][#1]    1 2 4 5  0.0 -30.0 30.0 60.0 -60.0 160.0 -160.0
a[CD2X4][ND3^3]*         1 2 3 4  60.0 -60.0 180.0 160.0 -160.0 90.0 -90.0 120.0 -120.0
an[CD2X4][CD1]           1 2 3 4  90.0 -90.0
[aD3]c([aD2])C(=O)[C^3]    1 2 4 5  0.0 20.0 -20.0 150.0 -150.0 180.0
[aD3]c([aD2])O[CD2]        1 2 3 4  180.0
a[ND2][CD2X4][CD2X4]       1 2 3 4  180.0 160.0 -160.0 80.0 -80.0 60.0 -60.0
[ND1]C(=O)c([aD3])         1 2 4 5  0.0 180.0 30.0 -30.0 150.0 -150.0
```

```
[aD2]c([aD2])c([aD2])[aD2]  1 2 4 5  -150.0 -30.0 30.0 150.0
[a]c[CD2][*D2]              1 2 3 4  -90.0 90.0 180.0 0.0 30.0 -30.0 150.0 -150.0
[a]cC(=O)c[a]               1 2 3 4  -150.0 -30.0 0.0 30.0 150.0 180.0
[a]cC(=O)[*D2]              1 2 3 4  0.0 180.0 30.0 -30.0 150.0 -150.0
[a]cOC                      1 2 3 4  0.0 180.0 30.0 -30.0 150.0 -150.0

#borderline low-res
[CD2]C(=O)[ND2][CD3][#1]    2 4 5 6  90.0 -90.0 60.0 -60.0 120.0 -120.0

#conjugated substituents
a[CD2]C=*                   1 2 3 4  150.0 -150.0 180.0 30.0 -30.0 0.0
C=CC=C                      1 2 3 4  0.0 180.0 30.0 -30.0 150.0 -150.0 60.0 -60.0 120.0 -120.0
cO[CD2]*                    1 2 3 4  0.0 30.0 -30.0 60.0 -60.0 90.0 -90.0 180.0
C=N[ND2]*=,:*               2 3 4 5  0.0 30.0 -30.0 150.0 -150.0 180.0
c[CD2][ND2]c                1 2 3 4  60.0 -60.0 80.0 -80.0 180.0
C=[CD3][ND3]*               1 2 3 4  30.0 -30.0 60.0 -60.0 90.0 -90.0 0.0 180.0

#ureas
[ND2]C(=O)Nc[nD2]           2 4 5 6  0.0 180.0
[ND2]C(=O)[ND2]*            1 2 4 5  0.0 180.0

#carbamates
C[ND2]C(=O)O                1 2 3 4  0.0 180.0
[ND2]C(=O)OC                3 2 4 5  0.0
OC(=O)N*                    3 2 4 5  0.0 20.0 -20.0 120.0 -120.0 160.0 -160.0 180.0

#piperidine amide
O=CN1[CD2][CD2][CD2][CD2][CD2]1    1 2 3 4  0.0

#amides and esters
[*D2]C(=O)O[CD3][#1]        2 4 5 6  0.0 30.0 -30.0
[OD2]C(=O)[CD2][CD2^3]      3 2 4 5  0.0 30.0 -30.0 120.0 -120.0 180.0
[O,SD1]=C(C)[ND2]C=[O,S]    1 2 4 5  0.0 180.0
[O,SD1]=C(C)[ND2][#7,#8]=*  1 2 4 5  0.0 180.0
[O,SD1]=C(C)[ND2]N          1 2 4 5  0.0 180.0
[O,SD1]=C(C)cn              1 2 4 5  0.0 180.0
[O,SD1]=C([#6])[ND2]*       1 2 4 5  0.0 20.0 -20.0
[O,SD1]=C[ND2]*             1 2 3 4  0.0 20.0 -20.0 180
O=C[ND3][CD3X4][#1]         2 3 4 5  0.0 180.0 20.0 -20.0
O=CNc([aD2,aD3])[aD3]       2 3 4 5  20.0 -20.0 -90.0 90.0 60.0 -60.0 120.0 -120.0 0.0
O=CNc[a]                    2 3 4 5  -20.0 20.0 90.0 -90.0 -160.0 160.0
O=C([CD2,CD3])O[CD2]        1 2 4 5  0.0
O=C([CD1])O[CD1]            1 2 4 5  0.0
[O,S]=CO[CD1]               1 2 3 4  0.0 20.0 -20.0 180.0
O=CO[CD2][CD1]              2 3 4 5  180.0
O=CO[CD2]*                  2 3 4 5  180.0 60.0 -60.0 90.0 -90.0
```

```
O=CO[CD3]*                  2 3 4 5  120.0 -120.0 180.0 0.0 60.0 -60.0
O=CO[CD4]*                  2 3 4 5  -60.0 60.0 120.0 80.0 -80.0
O=CO[CD3,CD4]               1 2 3 4  0.0 30.0 -30.0 60.0 -60.0
O=CO*                       1 2 3 4  0.0 30.0 -30.0 60.0 -60.0
O=C[ND3]([*D3])[*D3]        1 2 3 4  20.0 -20.0 0.0 180 150.0 -150.0
O=C[ND3]*                   1 2 3 4  0.0 180.0
CC[ND3](CC)[CD2,CD3]*       2 3 6 7  180.0 60.0 -60.0 120.0 -120.0 0.0 180.0 30.0 -30.0
[a][CD2][CD2][ND3]          1 2 3 4  90.0 -90.0 180.0 60.0 -60.0
[ND3]C(=O)[nD3]*            1 2 4 5  90.0 -90.0 60.0 -60.0 120.0 -120.0
[CD2]OC(=O)[CD2][CD3]       4 3 5 6  0.0 150.0 -150.0

#t-butyl
C([CD1])([CD1])([CD1])c[a]    2 1 5 6  90.0 30.0
**C([CD1])([CD1])[CD1]        1 2 3 4  180.0 150.0

#propyl
[CD1]C([CD1])([#1])[CD2]*     4 2 5 6 60.0 -60.0

#highly substituted alkane
#[CD2]C(=O)[ND2]-!@[CD3][#1] 2 4 5 6

*[CD2X4][CD3X4]([#1])[CD3] 1 2 3 4 180.0 60.0 -60.0 40.0 -40.0
c[CD2^3][CD3^3][#1]         1 2 3 4  180.0 60.0 -60.0
[CD2^3][CD2^3][CD3^3][#1]   1 2 3 4  60.0 -60.0 180.0 30.0 -30.0 0.0 160.0 -160.0 120.0 -
120.0
[*D2][CD2][CRH]([*R])[*R]  1 2 3 4  30.0 -30.0 120.0 -120.0 150.0 -150.0 60.0 -60.0 180.0 0.0
[*D2][CD2][CX4D3][*D2]     1 2 3 4  30.0 -30.0 120.0 -120.0 150.0 -150.0 60.0 -60.0 180.0
*[CHD3][CH2D2]*            1 2 3 4  150.0 -150.0 60.0 -60.0 180.0 -90.0 90.0 0.0 30.0 -30.0
[CD1]C([CD1])[CD2]*        1 2 4 5  60.0 -60.0 180.0 80.0 -80.0 30.0 -30.0

#long unsubstituted alkanes
#temporary removals of the following:
#[CD3][OD2][CD2][OD2]       1 2 3 4  60.0 -60.0 180.0 100.0 -100.0
#[CD1][CD2][CD2][*D2][*D2][*D2]  2 3 4 5  180.0
#[CD1][CD2][CD2][*D2][*D2][*D2]  1 2 3 4  180.0
#[*D2^3][*D2^3][*D2^3][*D2^3][*D2^3][*D2^3]  2 3 4 5  180.0

#nitro
[aD3]cN(~[OD1])~[OD1]       1 2 3 4  0.0 60.0 -60.0
[a]cN(~[OD1])~[OD1]         1 2 3 4  0.0

#trifluoromethyl
**C(F)(F)F                  1 2 3 4  0.0

#trichloromethyl
[a]cC(Cl)(Cl)Cl             1 2 3 4 0.0
```

```
#CSD SPECIFIC RULES
a[PD3](a)-[PD3](a)a        1 2 4 5 180.0 60.0 -60.0
PPcc                       1 2 3 4 60.0 -60.0

#phosphorus containing groups
#[OD1]~PO*                 1 2 3 4  0.0 -30.0 30.0 -60.0 60.0 120.0 -120.0
#[OD1]~P(~[OD1])(~[OD1])[OD2][CD2]* 2 5 6 7 0.0 60.0 120.0 180.0 -120.0 -60.0
#S=POc                     1 2 3 4  0.0 -60.0 60.0 90.0 -90.0
#[a]cCP(c)(c)c             1 2 3 4  90.0 -90.0
```

```
# rules.txt – for filter.py – by David M. Lorber. Uses OEChem (OpenEye, Santa Fe, NM)
# special flags
50.0 600.0 MOLWT
STRIPSALTS yes
0 10 CHIRALITY enumerate
ALLOWED_ATOMS C N O S P Cl F Br I H

# normal format is (min, max, name, SMARTS)

#rules
5 50 Non-Hydrogen_atoms [a,A]
2 49 carbons [#6]
1 20 N,O,S [#7,#8,#16]
0 1 Sulfonyl_halides S(=O)(=O)[Cl,Br]
0 1 Acid_halides [S,C](=[O,S])[F,Br,Cl,I]
0 3 Alkyl_halides [Br,Cl,I][CX4;CH,CH2]
0 0 Phosphenes cPc
0 0 Heptanes [CD1][CD2][CD2][CD2][CD2][CD2][CD2]
0 0 Perchlorates OCl(O)(O)(O)
0 7 Fluorines F
0 6 Cl,Br,I [Cl,Br,I]
0 0 Carbazides O=CN=[N+]=[N-]
0 0 Acid_anhydrides C(=O)OC(=O)
0 0 Peroxides OO
0 1 Iso(thio)cyanates N=C=[S,O]
0 1 Thiocyanates SC#N
0 0 Phosphoranes C=P
0 0 P/S_halides [P,S][Cl,Br,F,I]
0 0 Carbodiimides N=C=N
0 0 Cyanohydrines N#CC[OH]
0 0 Carbazides O=CN=[N+]=[N-]
0 1 Sulfate_esters COS(=O)O[C,c]
0 1 Sulfonates COS(=O)(=O)[C,c]
0 0 Pentafluorophenyl_esters C(=O)Oc1c(F)c(F)c(F)c(F)c1(F)
0 0 Paranitrophenyl_esters C(=O)Oc1ccc(N(=O)=O)cc1
```

0 0 HOBt_esters C(=O)Onnn
0 0 Triflates OS(=O)(=O)C(F)(F)F
0 0 Lawesson's_reagents P(=S)(S)S
0 0 Phosphoramides NP(=O)(N)N
0 0 Aromatic_azides cN=[N+]=[N-]
0 2 Quaternary_C,Cl,I,P,S [C+,Cl+,I+,P+,S+]
0 2 Beta_carbonyl_quaternary_N C(=O)C[N+,n+]
0 2 Acylhydrazides [N;R0][N;R0]C(=O)
0 0 Chloramidines [Cl]C([C&R0])=N
0 0 Isonitriles [N+]#[C-]
0 0 Triacyloximes C(=O)N(C(=O))OC(=O)
0 0 Acyl_cyanides N#CC(=O)
0 0 Sulfonyl_cyanides S(=O)(=O)C#N
0 0 Cyanophosphonates P(OCC)(OCC)(=O)C#N
0 0 Azocyanamides [N;R0]=[N;R0]C#N
0 0 Azoalkanals [N;R0]=[N;R0]CC=O
0 2 (Thio)epoxides,aziridines C1[O,S,N]C1
0 2 Benzylic_quaternary_N cC[N+]
0 2 Thioesters C[O,S;R0][C;R0](=S)
0 0 Diand_Triphosphates P(=O)([OH])OP(=O)[OH]
0 2 Aminooxy(oxo) [#7]O[#6,#16]=O
0 2 nitros N(~[OD1])~[OD1]
0 2 Imines C=[N;R0]*
0 2 Acrylonitriles N#CC=C
0 2 Propenals C=CC(=O)[!#7;!#8]
0 4 Quaternary_N [n+,N+]