

Computation of Ligand States in Preparation for Design Against Biological Targets

Amit Kulkarni, Ph.D., Accelrys Inc.

Swati Puri, Ph.D., Accelrys Inc.

Key Products: Structure Checker, Accord for Excel

Industry Sectors: Pharmaceuticals

Companies: Accelrys Inc.

Introduction

Increasing volumes of chemical information present a significant data management challenge to drug discovery research organizations. Such chemical information is typically chemically correct in the context of medicinal chemistry, but not always enumerated or represented appropriately for the physiological conditions under which their potential biological targets exist. Correct chemistry, bond orders, tautomeric states, and ionization states are necessary before chemicals are registered in corporate databases or used for computational experiments, such as docking, QSAR, or molecular dynamics simulation. The necessity of providing a correct representation arises when the molecules are to be correlated with biological activities or responses ascribed to the chemical entities being represented.

Accelrys' Structure Checker helps to meet this challenge by ensuring consistency and quality of chemical data across the organization. Structure Checker can automatically correct structures based on a set of customizable chemistry rules. Accelrys Structure Checker 1.0 enhances the ability to understand and unify chemical information based on a customizable set of Cheminformatics rules - created to validate and interpret the chemical structure. It can be used to ensure that structures conform to specific quality standards.

Shown in this case study is one such application of Structure Checker to process ligands before docking experiments. Structure-based drug design is the method used to identify and optimize pharmaceutical leads when an X-ray crystallography, NMR structure, or homology model of a specific target protein is known. Virtual screening of corporate libraries, external compound collections, and virtual compounds using various docking methods is routine in the drug discovery process. We now have in large part the ability to represent - reasonably accurately - single - or dozens - of ligands in a charged states appropriate to milieu pH, with correct configurations (stereochemistry, cis-trans isomers, E-Z isomers, etc), and potential tautomeric forms. However it is difficult to perform such customized treatments of ligands on a large scale.

Tools that are commonly used for ligand preparation often have hard coded rules. So a transformation of chemistry outside the defined rules is generally not possible. For example, most of the tools available will not ionize hydroxamic acids, as they are weak acids and not usually considered as ionized. Structure checker provides an open box solution. It allows users to define sets of rules specific to the local requirements of their targets.

Methodology and Work-Flow

The Structure Checker component is deployed as a web service with a SOAP interface. The component runs on the Java 2 runtime environment and is hosted by the JBoss application server. Tailor-made graphical user interface (GUI) for the Structure Checker can be easily coded based upon the SOAP interface

Structure checker accepts the following input format for ligands:

- Accord Hex Format (Hex encoded native Accord data)
- MolFile

- SMILES.

The heart and soul of structure checker is the Rules file (scRules.xml). The xml format of scRules.xml file is shown in Figure 1.

```

- <chem:Rule id="SCP009" description="Structure outside molecular weight range" optional="yes" applyCorrection="no" lastModified="2003-07-31 01:00:00" dormant="no">
  <chem:class name="com.accelrys.chem.gaffe.checker.rulechecker.MolecularWeightViolation" param1="50" param2="1000" />
</chem:Rule>
- <chem:Rule id="SCP010" description="Illegal stereobonds" optional="no" applyCorrection="yes" lastModified="2003-06-20 00:04:01" dormant="no">
  <chem:class name="com.accelrys.chem.gaffe.checker.rulechecker.IllegalStereoBonds" />
</chem:Rule>
- <chem:Rule id="SCC001" description="charge separated nitro groups" optional="no" applyCorrection="yes" lastModified="2003-07-31 01:00:00" dormant="no">
  - <chem:chemical>
    <chem:transform>C(['R1'])(O)=O>>C([O-])(=O)['R1']</chem:transform>
  </chem:chemical>
</chem:Rule>
</chem:scRules>

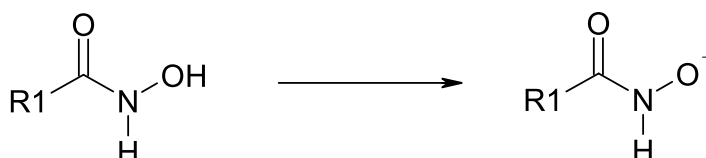
```

Figure 1: XML format of scRules.xml file

Structure checker allows import of user defined rules. Given below are the steps to create chemistry rules.

STEP 1:

Draw the chemical reaction of interest using either ISIS Draw or ChemDraw. Shown below is a simple example of converting hydroxamic acid to its ionized form.



STEP 2:

DS Accord for Excel is a chemistry aware excel spreadsheet that allows scientists to display chemistry and reactions, perform chemical calculations, analyze R-groups, calculate QSAR descriptors and query by similarity or substructure, all within the worlds' most popular spreadsheet environment, Figure 2.

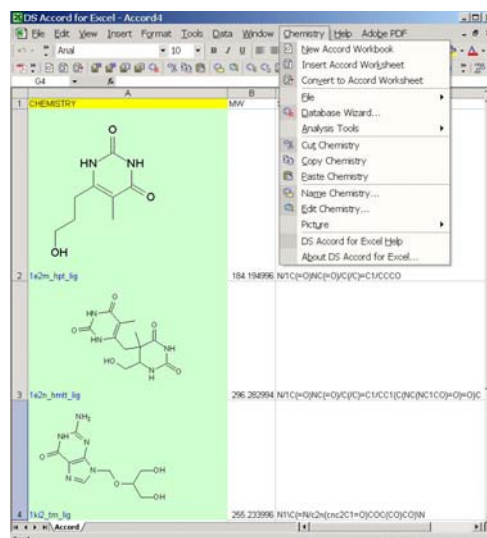


Figure 2: Accord for Excel Spreadsheet

Copy the above reaction and paste it in Accord for Excel (using the paste chemistry option), Figure 3. Then, use Accord for Excel to generate SMILES notation for the reaction, Figure 4

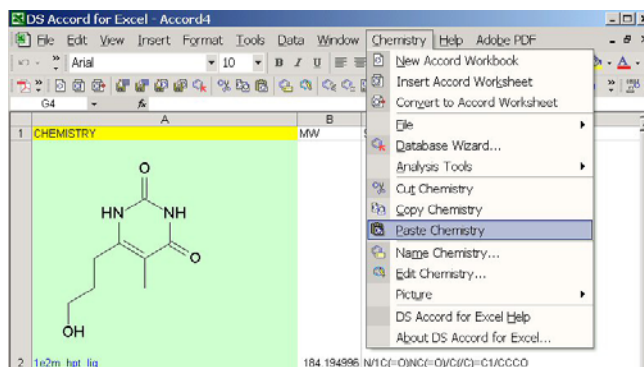


Figure 3: Chemistry paste option in Accord for Excel.

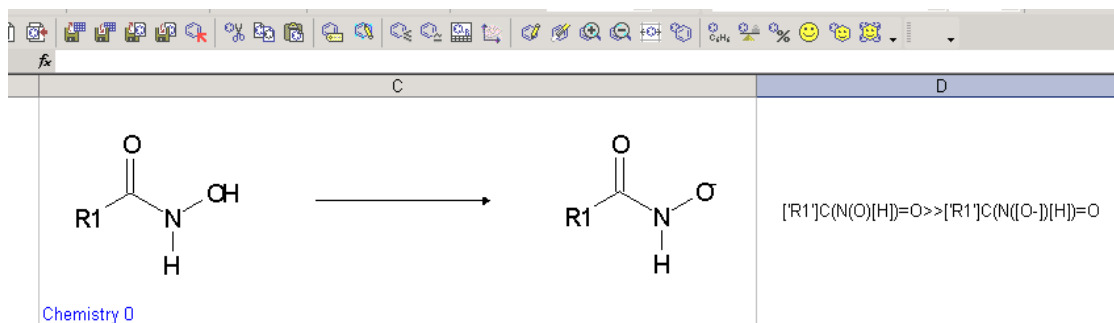


Figure 4: Generate SMILES string for chemical reaction.

STEP 3:

Open the following file in text editor.

C:\ProgramFiles\Accelrys\Accord50\StructureChecker\server\default\deploy\structurechecker.war\WEB-INF\classes\com\accelrys\chem\gaffe\rulestore\scRules.xml

Copy and paste the last part of the file from <Chem:Rule id> to </chem.:Rule>.

Change the Rule id number and add description of what the rule does. Then cut and paste the SMILES string notation for the reaction from Accord for Excel after <chem.:transform>. Figure 5 shows the addition of a new rule that will ionize all hydroxamic acids.

```

- <chem:Rule id="SCP009" description="Structure outside molecular weight range" optional="yes" applyCorrection="no"
  lastModified="2003-07-31 01:00:00" dormant="no">
  <chem:class name="com.accelrys.chem.gaffe.checker.rulechecker.MolecularWeightViolation" param1="50"
    param2="1000" />
</chem:Rule>
- <chem:Rule id="SCP010" description="Illegal stereobonds" optional="no" applyCorrection="yes" lastModified="2003-06-20 00:04:01"
  dormant="no">
  <chem:class name="com.accelrys.chem.gaffe.checker.rulechecker.IllegalStereoBonds" />
</chem:Rule>
- <chem:Rule id="SCC001" description="charge separated nitro groups" optional="no" applyCorrection="yes" lastModified="2003-07-
  31 01:00:00" dormant="no">
- <chem:chemical>
  <chem:transform>C([R1])(O)=O>>C([O-])(=O)[R1]</chem:transform>
</chem:chemical>
</chem:Rule>
- <chem:Rule id="SCC002" description="Charge Hydroxamic Acids" optional="no" applyCorrection="yes" lastModified="2003-07-31
  01:00:00" dormant="no">
- <chem:chemical>
  <chem:transform>[R1]C(N(O)[H])=O>>[R1]C(N([O-])[H])=O</chem:transform>
</chem:chemical>
</chem:Rule>
</chem:scRules>

```

Figure 5: Editing of scRules.xml file.

Table 1 shows some other chemistry rules that can be easily defined using Accord for Excel and Structure Checker.


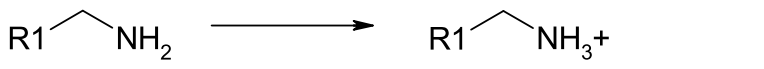
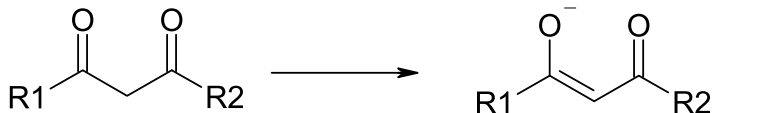
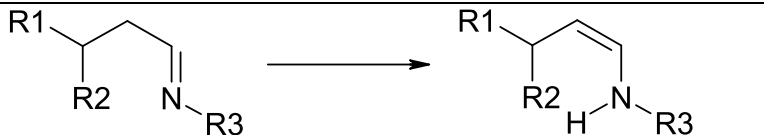
Chemistry Type	Chemical Reaction
Ionization	
Ionization	
Keto-Enol Tautomers (1,3 Diketones)	
Imine-Enamine (Tautomers)	

Table 1: User defined chemistry rules.

Results

When a structure is submitted for checking against the Rules database, the following outcomes are possible:

The structure is correct. No errors are returned and no structural corrections are necessary.

The structure is changed to become correct. If possible the Structure Checker will correct the structure and annotate the structure to show the changes that have been made. A structure failing any rule will be corrected (if possible) so that it passes the rule and then rechecked against all the rules.

The structure cannot be changed to become correct. If the Structure Checker attempts to make suitable corrections based on one rule but then fails based on another rule, you can configure the program to simply fail the molecule or to generate warnings and report on the failure.

The structure fails without any corrections being made. If the Structure Checker cannot make any corrections to enable the structure to pass a specific rule, the structure will be failed.

The Structure Checker will output a structure in the same format. Where no corrections were made to the structure during the course of the checks, this output structure will be an exact copy of the input data. This avoids any potential loss of data through conversion from one format to another, and back again.

In addition to the output structure, a second structure will be output in native Accord format, containing mark-up information indicating which atoms or bonds were involved in any corrections or rules that failed.

The structure Checker is supported on Windows XP and 2000 and on Linux (Advanced Server 2.1)

Conclusion

Structure Checker is a streamlined, customizable, and high-throughput tool for checking chemistries, ionization states, tautomeric states etc. Although the primary use of Structure Checker is to check chemistry during pre-registration process, as described here, it can also be used to pre-process ligands prior to docking experiments in structure based design.

System Requirements

Structure Checker is supported on Windows XP and 2000 and on Linux (Advanced Server 2.1).

Acknowledgements

The authors of this case study are grateful to Dr. Rita Wilby and Dr. Mariana Vaschetto from Accelrys Inc for many helpful discussions and advice. We also acknowledge the considerable intellectual input from Amol Kulkarni at University of Buffalo, NY