QSAR Modeling Independent of Input Tautomers

Robert Fraczkiewicz, Marvin Waldman, Robert D. Clark

Simulations Plus, Inc., 42505 10th Street West, Lancaster, CA 93534, USA

Summary

The quality and predictivity of QSAR models used in drug design and development often depend on the tautomeric and valence structures used to represent the molecules of interest. This is because the location of hydrogen bonding groups, bond orders, and formal charges affect the values of atomic and molecular descriptors upon which the models are based. For descriptors such as partial charges, this dependence is typically due to basing atom typing on hybridization. The sp^3-hybridized hydroxyl group in an enol, for example, is treated differently than its sp^2-hybridized counterpart in a carbonyl, even though oxygen appears in both forms in some tautomers of a molecule. In the aqueous environment of interest in drug design applications, representing a compound as any one of its tautomers is likely to distort the QSAR model obtained. Even worse is the danger that choosing a tautomer present at low abundance will bias the model building process or the reliability of predictions or both. To address this problem, we are developing a descriptor generation method which is independent of tautomer and valence structure representation and will illustrate its application to the atomic descriptors used in S+pKa, which is our global model of protic ionization constants. Preliminary results will be shown and compared with the "traditional" approach along with a discussion of advantages and potential pitfalls of the method.

The Problem

- QSAR predictions depend on structural representation:

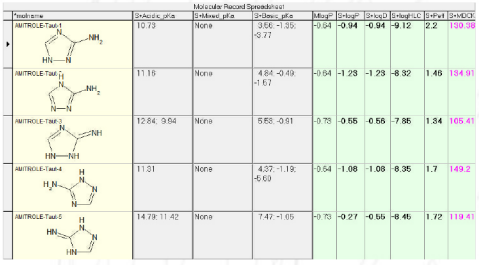

Figure 1. Predictions generated with ADMET Predictor™ v7.1 for different tautomers of amitrole.

- Atomic and molecular descriptors depend on structural representation:

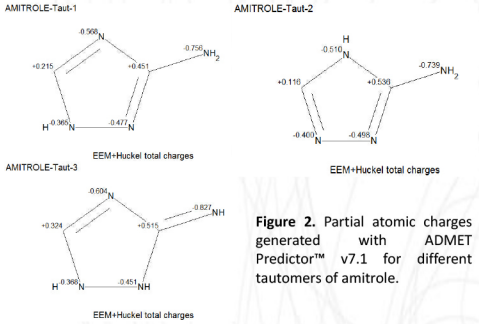

Figure 2. Partial atomic charges generated with ADMET Predictor™ v7.1 for different tautomers of amitrole.

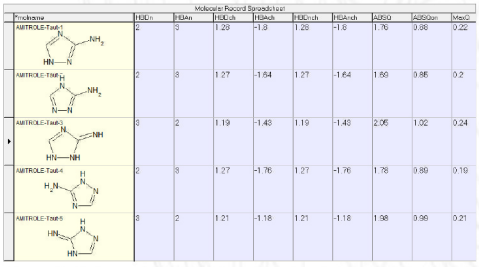

Figure 3. Hydrogen bonding and charge-related molecular descriptors generated with ADMET Predictor™ v7.1 for different tautomers of amitrole.

- There are problems of fundamental nature:

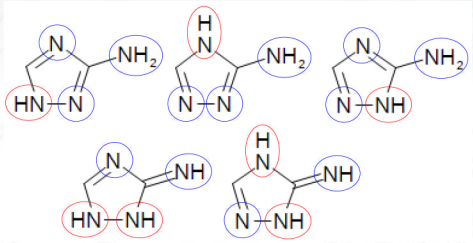

Figure 4. Top three tautomers of amitrole have 1 acidic (red) and 3 basic (blue) groups. Bottom two have 2 acidic and 2 basic groups. So: how many acidic and basic groups does amitrole have?

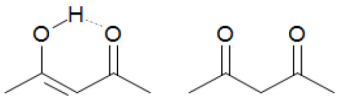

Figure 5. The left tautomer of acetylacetone has 1 internal hydrogen bond. The right tautomer has no internal hydrogen bonds. So: how many internal hydrogen bonds does acetylacetone have?

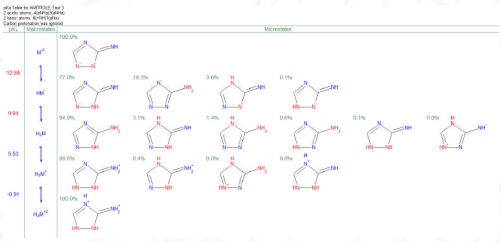

Figure 6. Tautomers are ionization microstates. They always exist in aqueous environment, simultaneously, albeit in varying proportions.

Proposed Solution

In order to make the calculation of atomic descriptors independent of tautomerism for pK_a prediction purposes, we have investigated the following novel structural representation:

- Protonate all basic nitrogen atoms
- Deprotonate all acidic –OH, -SH, and (where applicable) –CH< groups
- Find all resonance structures in which a minimal number of atoms bear a formal charge
- Average atomic descriptors over the resonance structures so obtained

This procedure is applied to each isolated p-electronic system in the molecule separately, which produces a collection of representative resonance structures with fixed numbers and positions of mobile hydrogen atoms.

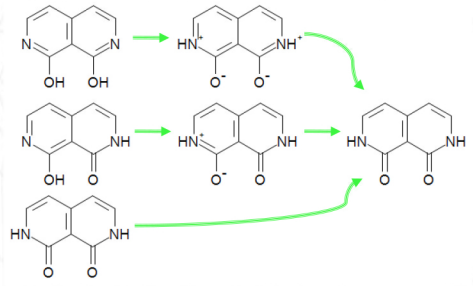

Figure 7. Applying our procedure to three different tautomers of 1,8-dihydroxy-2,7-diazanaphthalene leads to the same base resonance structure.

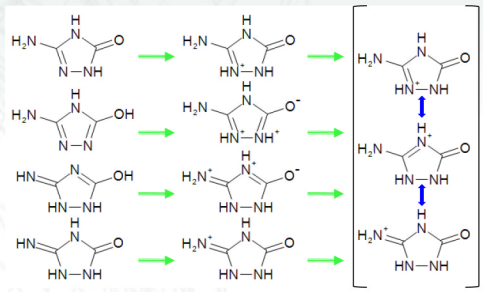

Figure 8. In this case, applying our procedure to any of four tautomers of hydroxyamitrole results in three equivalent resonance structures.

Preliminary Results

- Consistent atomic descriptors:

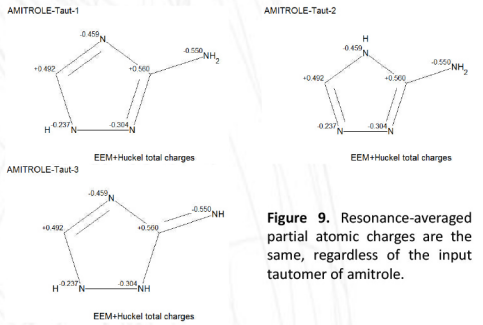

Figure 9. Resonance-averaged partial atomic charges are the same, regardless of the input tautomer of amitrole.

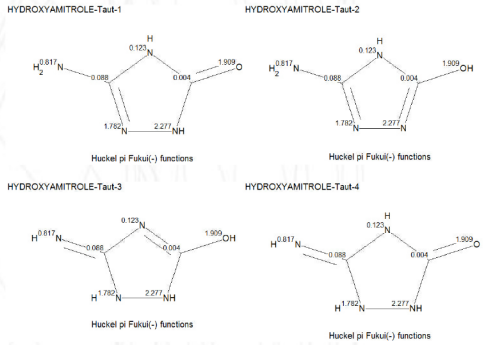

Figure 10. Resonance-averaged π-system Fukui(-) functions (measures of electrophilic reactivity) are the same, regardless of which hydroxyamitrole tautomer is input.

Conclusions

Results of our preliminary work are very encouraging: our resonance averaging procedure produces atomic descriptors that are independent of input tautomers. In fact, each of the 140-plus atomic descriptors implemented in the next version of ADMET Predictor has this invariant property.

The new atomic descriptors are being used to build a tautomer-independent version of our S+pKa model for predicting ionization constants.

For additional information, please leave your e-mail address or write directly to robert@simulations-plus.com.

simulations*plus, inc.*
integrating science and software