Dynamic QSAR: A New Search for Active Conformations and Significant Stereoelectronic Indices

Ovanes G. Mekenyan, Higher Institute of Chemical Technology, Bourgas, Bulgaria
Julian M. Ivanov, Higher Institute of Chemical Technology, Bourgas, Bulgaria
Gilman D. Veith, United States Environmental Protection Agency
Steven P. Bradbury, United States Environmental Protection Agency

Available at: https://works.bepress.com/steven_bradbury/64/
Dynamic QSAR: A New Search for Active Conformations and Significant Stereoelectronic Indices

Ovanes G. Mekenyan*, Julian M. Ivanov
Higher Institute of Chemical Technology, 8010 Bourgas, Bulgaria

Gilman D. Veith and Steven P. Bradbury
U.S. Environmental Protection Agency Environmental Research Laboratory – Duluth, 6201 Congdon Boulevard, Duluth, MN 55804 USA

Abstract

A new approach called “dynamic” QSAR is introduced to enhance the exploration of active chemicals and relevant molecular descriptors. In contrast to conventional QSAR methods where chemical structure is described by a single, low energy conformer, “dynamic” QSAR simulates the multiplicity of 3-D molecular shapes that a molecule can assume in complex reaction environments. The core of the new methodology is the coupling of the 3DGEN algorithm which exhaustively generates conformers and a rule-based system to rapidly screen conformers for desired properties. Hypotheses regarding receptor shape and interaction mechanisms are conveniently incorporated into the screening algorithm. A full array of stereoelectronic parameters available to OASIS can be combined with conventional topological and physicochemical indices for all conformations and explored using a variety of mathematical and visualization techniques. The “dynamic” QSAR method is illustrated by modeling the acute toxicity of a series of unsaturated alcohols in fish.

Key words: “dynamic” QSAR; conformer screening; geometry optimization; toxicity

Introduction

QSAR methods are based on the assumption that chemical structures can be defined explicitly and that mathematical descriptors which control chemical activity can be derived from these structures. There is no shortage of topological indices and substituent constants that can be used in the development of QSARs when chemical structure is confined to two-dimensional (2-D) representations. As computational capabilities allow QSAR research to evolve toward more extensive use of the indices derived from 3-D geometry, the proliferation of molecular indices has expanded well beyond our ability to ascribe mechanistic relevance to the indices.

Furthermore, the introduction of 3-D chemical structures in QSAR development has generally focussed on only one conformation of the chemical. Force field, or the more time consuming quantum chemical, calculations usually assist in choosing a low energy conformer of the structure. This selection leaves two questions. The first is that molecules in complex environments such as biological tissues and fluids are likely to exist in conformations other than the lowest, gas phase energy state, particularly when binding to macromolecules. Hence, the chemical behavior (activity, toxicity, reactivity, etc.) may be an integral effect of a set of conformers rather than a property of a single, low energy conformer. Searching for molecules that can assume a specific conformation may be as important as searching low energy conformers for possible activity.

Attempts have been made to account for solvent contributions by force field or quantum-chemical calculations [1-3], however, the results obtained to date have not established the usefulness of these explicit methods. Moreover, the respective time consuming calculations model the effect of single solvent molecules and thus, ultimately transfer the set of “active” conformers again to a single one.

The second question involves the influence of the conformation on the values of the stereoelectronic indices to be computed. Structures which have a wide range of indices among possible conformations could obscure QSAR development if the mechanism involved includes a change in structural conformation at the active site (or during the penetration). If poor correlations were found between chemical activity and a mechanistically plausible stereoelectronic index, it seems equally important to consider models with conformations other than that of lowest energy as it would be to search for different stereoelectronic indices. This paper addresses these two questions.

We recently reported a PC-based algorithm which exhaustively derives all isomers, enantiomers, and other 3-D conformations for a given molecular topology [4]. The “dynamic” QSAR method described herein was formed by coupling that algorithm for generating 3-D conformations for a set of chemicals with a hierarchical rule-based program which selects subsets of structures for subsequent QSAR analysis. The rule-based program screens the conformers for noteworthy features. For example, conformers which can assume a near planar structure or with specific hydrogen bonding alignments can be identified. Often such conformers are not in a low energy form of the chemicals. The combined program enables one to incorporate specific hypotheses for the mechanism of interaction and test the hypotheses with a more realistic subset of 3-D conformers in the congeneric series.

The “dynamic” QSAR method also allows optimization of molecular geometry as an optional step in the analysis. We note this because geometry optimization is often the first task in QSAR analyses involving stereoelectronic indices. However, we have often
found greater variation in indices among different conformers of a given molecule than with different geometry optimization procedures. In many instances, at least when radicals or excited states are not involved, the indices of electronic structures, without geometry optimization (1SCF or one-point calculations) may be sufficient for estimating of chemical activity. If so, delaying geometry optimization until necessary would save a significant amount of computing time and facilitate the modeling process.

To illustrate the use of the dynamic QSAR approach, we have reanalyzed QSARs for the aquatic toxicity of a series of unsaturated alcohols [5]. The unsaturated alcohols become reactive toxicants when metabolized to aldehydes or ketones if the parent alcohols are α, β-unsaturated. We previously showed that several reactivity descriptors of soft electrophilicity as described by local acceptor superdelocalizability of the unsaturated carbon atom (S1N) and the frontier charge on LUMO of carbonyl carbon (f3LUMO), unambiguously discriminated reactive and narcotic alcohols [5]. The reactivity descriptors were calculated by MNDO hamiltonian after the geometry optimization by the same quantum-chemical method. Using logP01 as a measure of hydrophobicity and either reactivity parameter, QSARs for which r2 = 0.85 and 0.84 were obtained when S1N and f3LUMO, respectively, were used.

**Method**

Algorithms like CONCORD are already available [6] for converting 2-D structures to single 3-D geometries which approximate low energy forms of the molecules. Each of these methods, however, is specialized in a particular area of the manifold procedures of 3-D structure elucidation. We developed a combinatorial and template-free algorithm (3DGEN) for converting the 2-D structures to 3-D [4]. Proceeding from atom-atom connectivity augmented by the type of atoms and multiplicity of chemical bonds, it provides an exhaustive generation of all stereoisomers, enantiomers and conformers corresponding to the given molecular topochemistry. The number of generated conformers can be controlled by managing the values of specific geometric constraints such as strain and interatomic distances.

The 3DGEN system and screening algorithms are subroutines of the 3DMOL module of a new version of the OASIS program package [7–9]. This module controls input of chemical structures and associated activity data, database management, generation of 3-D structures for each chemical, and interactive screening of the conformations for sought-after molecular features. The molecular topochemistry is the only information needed to begin the procedure. In a single session, all possible stereo and optical isomers can be generated, statistical summaries of distributions in the conformational space displayed, and specific structures or parameters can be imposed to focus the QSAR analysis.

The structural constraints inherent to the 3DGEN system include distances between nonbonded atoms, ring-closure parameters and torsional resolution which, when specified by the user, prevent the generation of unlikely high energy conformers. There are several rule-based systems which provide interactive screening of the generated 3-D isomers. One screening capability uses structural rules that have been developed to select conformers only when specific features are found in the structure. For example, conformers which have internal hydrogen bonding can be distinguished from those conformers which sterically do not allow formation of such bonds. Also, the high rotational barriers of conjugated fragments provide a relatively secure reduction of associated torsional angles to 180° and 0°. Maximizing distances in saturated fragments also often results in low energy conformers. Energy estimates obtained from strain minimization are also used [4]. Finally, configurations with respect to stereocenters (i.e. cis-trans orientations around a double bond or /– parity of chiral centers) have also been established.

The rule-base and other user requirements can be established either before or after stating the generation of conformers. However, formulating hypotheses in the form of these constraints prior to QSAR analysis reduces the complexity of the combinatorial task of 3DGEN and significantly speeds the exploration for most relevant conformers.

In addition, conformers can be screened by electronic properties. The stereoelectronic indices for each allowed configuration can be computed using semi-empirical quantum chemical methods, either before or after optimizing geometry. 1SCF calculations of indices using the OASIS modification for reactivity parameters [7–9] on the unoptimized conformers give reasonable approximations for most structures because bond lengths and angles used in 3DGEN are parameterized in advance [4]. Comparing the computed geometries from the 3DGEN with those derived from quantum chemical or force field calculations, we have often found that such optimizations do not improve the resulting QSAR. Alternatively, having a rapid, good approximation enables one to screen and select conformers based on ranges of specific local and/or global parameter indices.

Conformer distributions based on energy or any other selected geometric, physico-chemical, and/or stereoelectronic parameter can be examined and then used in the search for a reasonable, mechanistically relevant QSAR. There is a variety of schemes to examine the distributions of values among the conformers. In summary form, they are:

(a) Whole Range of Parameter Variation

The number of conformers examined in detail can be reduced by dividing the parameter distribution into equal portions with user-specified quintals. The quintals simply segment the distribution frequency into ranges encompassing an equal number of conformers. Selecting a conformer from each quintal reduces the number of conformers in subsequent analysis while maintaining some assurance that the data set is representative of the entire range of parameter values. If the number of conformers is smaller than the number of quintals, all conformers are included in subsequent analysis.

For example, the conformer distribution of 2-decyn-1-ol based on the maximum geometric distance in the molecules (Lmax[Å]) is illustrated in Figure 1. The number of quintals chosen in this example was seven, and the conformer selected from each quintal to represent the parameter range is marked with an asterisk. Together, the seven conformers provide a uniform representation of the whole range of Lmax [Å] variation.
(b) Prevailing Range of the Parameter

The distribution of each parameter can also be displayed among a selected number of parameter windows, each having an equal percentage of the whole parameter range. The conformers belonging to the most populated window are considered the conformers with the prevailing range of parameter values as illustrated in Figure 2. These conformers are assumed to provide a subset of the most prevalent values among all conformers.

(c) Extreme Values of the Parameter

Conformers which have extreme (maximum or minimum) values of the parameter populate the outermost windows of the parameter distribution. Figure 3 illustrates the clustering of conformers with extreme values of $L_{\text{max}}$ (Å). Other common uses of this technique includes the selection of the lower energy structures after displaying the distribution of calculated heat of formation. Conformers with the highest electron acceptor or donor properties can be selected by clustering those having minimum values of $E_{\text{LUMO}}$ or maximum values of $E_{\text{HOMO}}$ respectively.

(d) Specified Population of Energy Level

Conformer energies can also be displayed by Boltzmann’s distribution using the calculated heat of formation, $\Delta H_i^f$, for each conformer (at $T = 298$ K). Selecting a cluster of conformers which are characterized by having energies corresponding to a specified percentage of the Boltzmann’s distribution can be accomplished by specifying the percentage of energy level population:

$$ P = 100 \times p_i = 100 \times \exp\left[-(\Delta H_i^f - \Delta H_i^{\text{th}})/RT\right]\Sigma \exp\left[-(\Delta H_i^f - \Delta H_i^{\text{th}})/RT\right] $$

(1)

(e) Parameter Weighing

In the first three screening methods, conformers are weighted equally when used in QSAR analysis. However, each parameter value can be normalized with respect to mean values, where the mean is computed as the arithmetic (used in Table 1), geometric, quadratic or harmonic mean. Thus, if one denotes by $E_{\text{LUMO}}^1, E_{\text{LUMO}}^2, E_{\text{LUMO}}^3, \ldots, E_{\text{LUMO}}^n$ the $E_{\text{LUMO}}$ values for the conformers with the respective energy level populations $p_1, p_2, p_3, \ldots, p_n$, then:

$$ E_{\text{LUMO}}^{\text{q}} = \Sigma p_i E_{\text{LUMO}}^{\text{q}} $$

(2)

Using the 3DGEN algorithm with the set of unsaturated alcohols [5], 968 conformers were obtained for the 20 alcohols or their aldehyde and ketone metabolites. High energy conformations were precluded using steric constraints such as a non-bonded cutoff at 1.8 Å and a ring-closure distance of 1.2–1.8 Å (see ref. [4]). Reactivity descriptors were calculated by PM3, AM1 and MND0 quantum-chemical methods by using MOPAC6 package [10].

Figure 1. Conformer distribution of 2-decyn-1-ol based on the maximum geometric distances in the molecules ($L_{\text{max}}$ (Å)). The number of quintals is seven.

Figure 2. Selected conformers for 2-decyn-1-ol (72 out of 222) with prevailing values of $L_{\text{max}}$ in the range 10.290 Å to 11.223 Å. The introduced number of quintals is again seven.

Figure 3. Selected conformers for 2-decyn-1-ol (5 out of 222) with maximum values of $L_{\text{max}}$ in the range 13.090 Å to 14.023 Å. The introduced number of quintals is seven.
### Results and Discussion

The statistics of derived QSARs obtained after conformer screening according to $E_{\text{LUMO}}$ used as a reactivity parameter, are presented in Table 1. The size of the correlation sample ($n$), explained variance ($r^2$), standard deviation ($s^2$) and Fisher criterion ($F$) are given here. All models in Table 1 were found to be significant at 95% confidence level.

2-Decyn-1-ol was the most flexible structure and 3DGEN produced 617 conformers for this alcohol. The fact that over 50% of all conformers come from one compound will cause a bias in the regression analysis. In order to minimize such possible biases, an initial conformation selection was made using the screening approach which provided representation of the whole range of a steric parameter variation but reduced the set of conformers to 100. The geometric analogue of Wiener topological index [11] was chosen as a geometric screening parameter. Additional conformer screening rules were based on the distribution of stereo-electronic parameter related to molecular soft electrophilicity, as the energy of the LUMO orbitals, $E_{\text{LUMO}}$, acceptor superdelocalizability indices, $S_{1}^N$, $S_{3}^N$ (for the unsaturated carbon atom and carbonyl carbon atom, respectively), frontier charge on LUMO, $f_{1}^{\text{LUMO}}$ and volume polarizability (Vol. Pol) [12]. Reactivity indices were calculated using ISCF methods before and after geometry optimization. Each conformer was treated as a discrete chemical in the correlation sample. The possible influence of this on the regression analysis is discussed in relation to Table 2.

The acceptor superdelocalizability indices, $S_{1}^N$, $S_{3}^N$ and frontier charge on LUMO, $f_{1}^{\text{LUMO}}$ are best combined with hydrophobicity ($\log P$; determined by CLOGP [13] as in ref. 5). This is an indication for the existence of two reaction sites, one at the unsaturated carbon atom (position #1) and the other at the carbonyl carbon (#3) of the presumed metabolites. As opposed to our previous report [5] the carbonyl carbon appears to be the more likely reaction site. As one can see from Table 1, the $S_{3}^N$ index combined with $\log P$ is found in most of the derived QSARs (the models with $S_{3}^N$ are denoted by asterisks). A similar conclusion can be drawn from the data presented in Table 2, where the modeling results are presented by combining $\log P$ with both the frontier charges and acceptor super-delocalizabilities for the two possible reaction sites. One can see that either frontier charge or acceptor super-delocalizability of carbonyl carbon are included in all best derived QSARs.

The incorporation of all conformers from the initial set into the correlation sample yield models which appear to be better than those in reference 5. The geometry optimization improved the QSARs, but almost insignificantly when considering the dramatic increase in computing time to do so. With the reduction of initial correlation sample by representation of the whole range of $E_{\text{LUMO}}$ by a smaller number of conformers (up to 10, 30 or 50 for each compound), $r^2$ values decreased to $0.82 < r^2 < 0.86$. Again, when the optimized conformers were included in the correlation samples, the statistics of the derived models were not significantly changed. The latter decreased to $r^2 < 0.79$ when conformers with the greatest electron acceptor properties (minimum of $E_{\text{LUMO}}$) were selected.

The best QSARs were obtained by the conformer screening with the prevailing value of the reactivity parameter ($E_{\text{LUMO}}$). Here, the QSARs obtained after geometry optimization of conformers are slightly better than those derived after ISCF calculations.

### Table 1. Statistics of the derived QSARs with $\log P$ and SN(3) (SN(1) in the cases denoted by asterisks) as regressors at the different conformer screening algorithms according to $E_{\text{LUMO}}$. The reactivity parameters are determined with and without performing molecular geometry optimization. PM3 hamiltonian is used.

<table>
<thead>
<tr>
<th>Screening method</th>
<th>Number of Quintals</th>
<th>Number of Conform</th>
<th>%</th>
<th>n</th>
<th>$r^2$</th>
<th>$s^2$</th>
<th>$F$</th>
<th>n</th>
<th>$r^2$</th>
<th>$s^2$</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Conform. Set</td>
<td>436</td>
<td>0.882</td>
<td>0.133</td>
<td>1603.5</td>
<td>436</td>
<td>0.887</td>
<td>0.126</td>
<td>1709.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole Range</td>
<td>10</td>
<td>0.821</td>
<td>0.181</td>
<td>266.3</td>
<td>119</td>
<td>0.826</td>
<td>0.177</td>
<td>274.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.846</td>
<td>0.175</td>
<td>640.3</td>
<td>235</td>
<td>0.854</td>
<td>0.166</td>
<td>678.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.857</td>
<td>0.154</td>
<td>970.7</td>
<td>328</td>
<td>0.863</td>
<td>0.146</td>
<td>1028.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prevailing</td>
<td>3</td>
<td>0.900</td>
<td>0.117</td>
<td>1107.5</td>
<td>254</td>
<td>0.927</td>
<td>0.078</td>
<td>1298.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.901</td>
<td>0.122</td>
<td>964.9</td>
<td>217</td>
<td>0.889</td>
<td>0.125</td>
<td>528.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>10</td>
<td>0.885</td>
<td>0.126</td>
<td>611.0</td>
<td>162</td>
<td>0.876</td>
<td>0.124</td>
<td>417.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.865</td>
<td>0.156</td>
<td>270.0</td>
<td>87</td>
<td>0.865</td>
<td>0.160</td>
<td>214.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum Range</td>
<td>3</td>
<td>0.759</td>
<td>0.149</td>
<td>136.8</td>
<td>90</td>
<td>0.854</td>
<td>0.098</td>
<td>589.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.787</td>
<td>0.162</td>
<td>106.5</td>
<td>61</td>
<td>0.856</td>
<td>0.078</td>
<td>357.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.787</td>
<td>0.210</td>
<td>75.4</td>
<td>44</td>
<td>0.808</td>
<td>0.151</td>
<td>82.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-Population</td>
<td>10</td>
<td>0.810</td>
<td>0.234</td>
<td>34.2</td>
<td>20*</td>
<td>0.796</td>
<td>0.251</td>
<td>31.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.810</td>
<td>0.220</td>
<td>36.2</td>
<td>21*</td>
<td>0.834</td>
<td>0.198</td>
<td>62.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.746</td>
<td>0.217</td>
<td>33.9</td>
<td>28*</td>
<td>0.876</td>
<td>0.143</td>
<td>148.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parameter</td>
<td>10</td>
<td>0.810</td>
<td>0.234</td>
<td>34.2</td>
<td>20*</td>
<td>0.796</td>
<td>0.251</td>
<td>31.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.810</td>
<td>0.234</td>
<td>34.2</td>
<td>20</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.810</td>
<td>0.234</td>
<td>34.2</td>
<td>20</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dynamic Search for 3-D QSAR 305

As already mentioned in the Method section, the energy factor was neglected in the above screening algorithms and the conformers were assumed to be equally stable. To examine the influence of energy, two additional screening methods were used. Conformers were selected to provide 10%, 30% and 50% of energy level population according to the Boltzmann's distribution at 298 K. The best models without geometry optimization of 3-D isomers in the correlation sample have 0.75 < \( r^2 \) < 0.81. After the geometry optimization, we found an improvement in the explained variance in that 0.80 < \( r^2 \) < 0.88 for the lower energy conformations. In other words, the screening approach providing the most stable conformers for each chemical, is more appropriate if the molecules are energetically optimized in advance. Still, in general, the modeling results are worse than in the case of the screening providing prevailing values of \( E_{\text{LUMO}} \).

Arithmetic averaging of the molecular descriptors for each compound after weighting the contribution of conformers with respect to the energy level population did not improve results. Moreover, even for the series with optimized conformers, we found \( r^2 < 0.80 \). As expected, the incorporation of energetically unfavorable conformers in the averaging process prevented any reasonable QSAR from being developed.

The comparison of the results by PM3, AM1, and MNDO methods, after conformer screenings according to \( E_{\text{LUMO}} \) showed that the results are essentially the same. Once again, the QSAR results for these largely aliphatic chemicals are not dependent on the Hamiltonian used in the chemical modeling.

The results in Table 2 were obtained after conformer screening with respect to \( E_{\text{LUMO}} \), using both frontier charges on LUMO and acceptor superdelocalizabilities as reactivity parameter. As is evident, the frontier charge on the carbonyl carbon combines statistically slightly better with \( \log P \) than the acceptor superdelocalizability. Thus, the biparametric models based on ISCF calculations have 0.89 < \( r^2 \) < 0.91. Statistically significant three-parameter models were also derived from ISCF calculations before (0.93 < \( r^2 \) < 0.95) and after (0.92 < \( r^2 \) < 0.95) geometry optimizations. It can also be seen that all of the three-parameter models are linear combinations between hydrophobicity and electron acceptor properties of the two reaction sites.

The screening of conformers was also performed using reactivity indices such as \( S_1^N, S_3^N \) and \( f_1 \text{LUMO} \). The statistics of the models again pointed out \( S_1^N, S_3^N \) as best combined reactivity parameters with \( \log P \). The screening results do not differ qualitatively from those in Table 1. Similarly to the screening according to \( E_{\text{LUMO}} \), the best QSARs are obtained by conformers providing prevailing values of the parameters assessing local acceptor properties of molecules (0.88 < \( r^2 \) < 0.93).

Recognizing that chemicals may have many conformations, and that those conformations can significantly alter computed properties, creates a somewhat intractable problem for the use of regression analysis in QSAR development. We have created a system which permits one to explore different strategies to detect the most important factors in chemical activity. As long as the screening method results in approximately the same number of conformers for each chemical, regressions on all conformers or the mean values of each descriptor should give similar results.

However, as it is often the case, the series of chemicals vary widely in flexibility and the number of possible conformers will vary. When the conformation significantly alters the stereoelectronic indices, averaging methods to keep the number of conformers constant for regression analysis will obscure what may possibly be the true structure-activity relationship. At the same time, allowing the variance of the dependent variables to change violates the principles of regression analysis [14].

Table 2: QSARs with \( \log P \) and reactivity parameters used as regressors. The conformer screening was performed according to the prevailing range of \( E_{\text{LUMO}} \). PM3 Hamiltonian is used.

<table>
<thead>
<tr>
<th>Number</th>
<th>n</th>
<th>( r^2 )</th>
<th>( s^2 )</th>
<th>F</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>254</td>
<td>0.945</td>
<td>0.064</td>
<td>1423.2</td>
<td>( f(3), S(1), \log P )</td>
</tr>
<tr>
<td>3</td>
<td>254</td>
<td>0.910</td>
<td>0.103</td>
<td>1269.5</td>
<td>( f(3), \log P )</td>
</tr>
<tr>
<td>5</td>
<td>217</td>
<td>0.943</td>
<td>0.070</td>
<td>1165.4</td>
<td>( f(3), S(1), \log P )</td>
</tr>
<tr>
<td>5</td>
<td>217</td>
<td>0.906</td>
<td>0.114</td>
<td>1037.0</td>
<td>( f(3), \log P )</td>
</tr>
<tr>
<td>10</td>
<td>162</td>
<td>0.933</td>
<td>0.075</td>
<td>724.3</td>
<td>( f(3), S(1), \log P )</td>
</tr>
<tr>
<td>10</td>
<td>162</td>
<td>0.885</td>
<td>0.126</td>
<td>611.0</td>
<td>( f(3), \log P )</td>
</tr>
</tbody>
</table>

As expected, the incorporation of energetically unfavorable conformers in the averaging process prevented any reasonable QSAR from being developed.

The screening algorithm results do not differ qualitatively from those in Table 1. Similarly to the screening according to \( E_{\text{LUMO}} \), the best QSARs are obtained by conformers providing prevailing values of the parameters assessing local acceptor properties of molecules (0.88 < \( r^2 \) < 0.93).

Recognizing that chemicals may have many conformations, and that those conformations can significantly alter computed properties, creates a somewhat intractable problem for the use of regression analysis in QSAR development. We have created a system which permits one to explore different strategies to detect the most important factors in chemical activity. As long as the screening method results in approximately the same number of conformers for each chemical, regressions on all conformers or the mean values of each descriptor should give similar results.

However, as it is often the case, the series of chemicals vary widely in flexibility and the number of possible conformers will vary. When the conformation significantly alters the stereoelectronic indices, averaging methods to keep the number of conformers constant for regression analysis will obscure what may possibly be the true structure-activity relationship. At the same time, allowing the variance of the dependent variables to change violates the principles of regression analysis [14].

Perhaps the only way to address the issue of varying conformations is to use visualization techniques, and to allow the researcher to propose different screening strategies interactively based on visualization of the data. This dynamic approach to QSAR is illustrated in Figure 4 in only two dimensions. Figure 4 shows not only the variation in the number of conformers among the 20 chemicals, but also the variation in the computed toxicity for the various conformers. The regression line is biased on those structures which have widely varying indices. One can remove this bias by...
selecting one conformation; however, the true structure-activity relationship could be missed if that conformer is not mechanistically relevant. In fact, Figure 4 suggests that there could be several best-fit lines in addition to the regression line. Whether a line represents a structure-activity relationship or a statistical aberration is best explored by proposing a mechanistic hypothesis and objectively ordering conformations to adequately test the proposal. This iterative approach becomes a dynamic exploration when the investigator can manage large data arrays and pose powerful questions with ease.

Conclusions

The new QSAR method proposed in this paper is an attempt to take into account the multiplicity of conformers taking part in interactions carried out in complex reaction environments. This was achieved by generating all conformers fulfilling some requirement for their stereoelectronic structure, according to user’s hypothesis on the reaction mechanism. The new method is an alternative to the conventional QSAR, where chemicals are represented by a single conformation.

The results of application of the “dynamic” QSAR were illustrated by modeling acute toxicity of a series of unsaturated alcohols and their respective metabolites. It was found that the inclusion of all conformers of the studied molecules improves the statistics of the models obtained by ISCF assessments of molecular reactivity. Though further validation is necessary, this result can facilitate QSAR studies which are usually hampered by the tremendous calculation for molecular geometry optimization.

Acknowledgement

The authors thank Dr. Stoyan Karabunarliev and Dr. Daniel Bonchev of the Higher School of Chemical Technology, Dr. Jose Kanetti of the Bulgarian Academy of Sciences and Dr. Daniel Call of the University of Wisconsin-Superior for the fruitful discussions and assistance in preparing this work for publication. Mention of specific modeling systems or approaches does not constitute endorsement by EPA.

References


Received on April 28th, 1994; accepted on June 26th, 1994

Figure 4. Graphical comparison of the estimated and measured aquatic toxicity of alcohols using the dynamic approach.

formed by rotations around single bonds only and they do not take into account the conformational flexibility of (saturated) cyclic molecular fragments.

The geometry optimization of the conformers under investigation does not improve significantly the statistics of the models obtained by ISCF assessments of molecular reactivity. Though further validation is necessary, this result can facilitate QSAR studies which are usually hampered by the tremendous calculation for molecular geometry optimization.