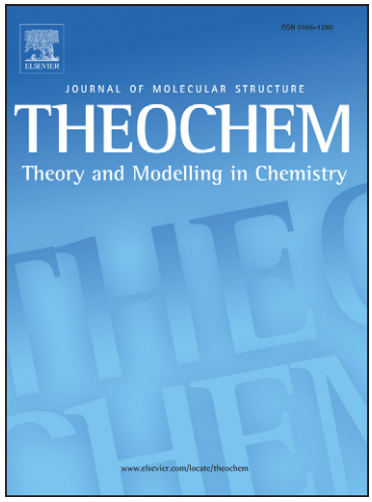
Accepted Manuscript



3D QSAR based on conceptual DFT molecular fields : antituberculotic activity

Sofie Van Damme, Patrick Bultinck

PII: S0166-1280(09)00726-X
DOI: 10.1016/j.theochem.2009.10.031
Reference: THEOCH 11885

To appear in: *Journal of Molecular Structure: THEOCHEM*

Received Date: 14 August 2009
Revised Date: 23 October 2009
Accepted Date: 25 October 2009

Please cite this article as: S. Van Damme, P. Bultinck, 3D QSAR based on conceptual DFT molecular fields : antituberculotic activity, *Journal of Molecular Structure: THEOCHEM* (2009), doi: 10.1016/j.theochem.2009.10.031

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

3D QSAR based on conceptual DFT molecular fields : antituberculotic activity

Sofie Van Damme[a,b], Patrick Bultinck[a,b,*]

[a] *Ghent University, Faculty of Science, Department of Inorganic and Physical Chemistry, Krijgslaan 281, S3, 9000 Ghent, Belgium*
[b] *Members of the QCMM Alliance Ghent-Brussels*

Abstract

We describe the use of conceptual DFT based quantum chemical molecular fields for three-dimensional quantitative structure-activity relations (3D-QSARs) and compare this new approach with the use of the default molecular property fields. The use of the new molecular fields in 3D QSAR is investigated by an application in the field of drug discovery, in which the antituberculotic activity of salycilamide derivatives is investigated. It is shown that conceptual DFT molecular fields have an added value to the default considered CoMFA fields.

Keywords: 3D QSAR, Conceptual DFT

1. Introduction

There can be hardly any doubt that one of the main contributors to our current quality of life has been the development of medicine, including the availability of modern drugs and treatments for many health problems. It can moreover be expected that development of new drugs will continue to be important for at least three reasons. First, there remain diseases that, despite having been known for long time, remain without cure although they affect the quality of life of millions or even billions of people. A typical example is malaria. Second, as life expectation continues to increase, especially in the developed world, some diseases associated with high age become more abundant and require treatment. Third, new diseases show up from time to time.

The development of new drugs is a very time consuming and error prone venture. One of the reasons why is the enormous vast chemical search space where new ligands (i.e. substances that

*Corresponding author
Email address: patrick.bultinck@ugent.be (Patrick Bultinck)

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

eventually can become drugs) have to be sought. From a theoretical/computational chemistry perspective, one of the most promising ways for ligand design would rely on a detailed dynamical description of the ligand/target interaction, including all the essential characteristics of the environment (solvent, electrolytes, ambient temperature, ...) [1]. At present such methods still need to be developed much further to have a sufficient level of detail. Moreover, these methods are so time consuming that they can hardly be used for ligand design where thousands of candidates need to be evaluated.

An alternative to the detailed dynamical description of the interaction are Quantitative Structure Activity Relationships (QSAR), a field of study introduced by Hammet [2], Hansch and Fujita [3] and Free and Wilson [4]. The fundamental assumption for a QSAR is that variations in biological activity among a set of compounds can be related to variations in their chemical structures and properties. In practice one largely abandons the detailed dynamical interaction but tries to establish a mathematical connection between the biological activity of a compound and some key molecular properties. The actual mathematical connection relies on statistics and relates biological activity to so-called molecular descriptors. Such descriptors range in complexity from very simple features like molecular mass, via absence/presence data for e.g. specific functional groups, to full 3D molecular fields like the electron density. The statistical method used then identifies those descriptors that give the best QSAR based on a set of known data for biological activity and descriptors for a so-called training set. Once the QSAR is known, it can then be used to predict the activity of molecules for which only the descriptors are known.

In the present paper we examine how well 3D molecular fields from conceptual or chemical DFT perform as descriptors for antituberculotic activity of salicylamides. Do they offer added value compared to simpler descriptors? In order to answer this question, we first describe in some more detail the QSAR algorithm used, then the fields considered and eventually the results of the QSAR in detail.

2. Quantitative structure activity relationships

For the construction of QSAR models, most often a linear model is sought, which can be expressed as equation (1)

$$A = k_1 * D_1 + k_2 * D_2 + ... + \text{const} \tag{1}$$

ACCEPTED MANUSCRIPT

Here A stands for the biological activity, D_i stands for the value of the i^{th} molecular descriptor retained in the QSAR, k_i is the regression constant belonging to descriptor D_i and C is a constant. The core machinery of deriving a QSAR are statistical algorithms that pick the most discriminating descriptors for each problem from a large pool of possible descriptors and yield the coefficients k_i. Obviously, in order to develop a QSAR, one first needs a set of molecules for which the activities are known as well as a set of descriptors from which usually only a small fraction will be kept in the final QSAR. Obviously, the better one can describe the molecule in terms of a set of molecular descriptors, the bigger the chance of a better QSAR. As a consequence, there has been a lot of research on the development of new descriptors that may be better at representing different aspects of molecular behavior, e.g. reactivity [5]. Note that extending the pool of descriptors from which to pick the significant descriptors for the problem considered is fruitful, provided the descriptors are not correlated. Increasing the number of descriptors retained in the QSAR model has to be done with great care in order to avoid overfitting and so is fundamentally different from increasing the pool of descriptors.

2.1. Three-dimensional (3D) QSAR

In 3D-QSAR [6], one uses 3D molecular fields as molecular descriptors. Assuming that the compounds considered in some application all bind in more or less the same way to the target, one could use the value of a 3D molecular field at every point in space as a descriptor. If then one assumes that variation in biological activity can be related to the change in these 3D molecular field values between the molecules, a 3D-QSAR model can be obtained.

The first commercially available 3D QSAR algorithm is known as CoMFA [7], where the default interaction fields are the steric and electrostatic fields, evaluated on a common grid of points. The same grid is used for all molecules, so molecular alignment plays an important role. The Partial Least Squares (PLS) technique [8] is used to handle the high number of descriptors.

3D QSAR models are relatively easily interpretable through visualization of where the important regions that correspond to descriptors retained in the QSAR are located. Once it is known in what region these reside and what effect they have on the QSAR, drug designers can manipulate the ligand candidates to maximally exploit these regions and hence increase the (predicted) activity of the compound.

CoMFA is still used very often as it does often give good working QSAR. Over the years, several

ACCEPTED MANUSCRIPT

workers have introduced several new types of fields, adding hydrophobic [9], hydrogen donor and hydrogen acceptor [10] or lipophilic [11] information to the model. Our interest lies mainly in the use of fields originating from so-called conceptual or chemical DFT [12]. The use of these fields is very attractive in 3D QSAR, as they can be interpreted as reactivity indices representing the response of a system to a perturbation[13]. In as far as the system modeled can be assumed to represent the actual drug or ligand in its working environment, the biological activity could be expressed completely in terms of the electron density, or the shape function, or the response functions such as the Fukui function or the local softness.

In the next section the 3D fields are introduced. The final section presents the application for a real test case.

3. 3D QSAR molecular property fields

3.1. Default considered CoMFA fields

In standard CoMFA, two fields are generated on a regular grid : the electrostatic and the steric field. The electrostatic field at a point j is expressed as its approximate Coulomb point-charge based potential function. The steric field is based on the Lennard-Jones potential function. Both fields are extensively described in reference work about 3D QSAR [7].

3.2. Fields based on conceptual DFT

It is known that the insufficient representation of the investigated structures in CoMFA is still an inherent deficiency in the algorithm, despite the widespread use and success of 3D QSAR. Representing a molecule by a set of atom-centered partial charges to calculate its Coulomb interaction is a very crude approximation, corresponding to the assumption that the charge distribution is locally isotropic close to the atoms. As a consequence, the models do not always represent the real situation. Therefore, together with the invention of the 3D QSAR technique, a search for new descriptor fields originated as well.

The fields introduced here are obtained from conceptual density functional theory and are extensively reviewed in reference [12]. The solid theoretical foundation of these fields does not stand any intuitive interpretation in the way, which is important for use as QSAR descriptors.

The fundamental property in conceptual DFT is the molecular electron density ρ, which is related to the shape and the size of a molecule. A molecular 3-dimensional descriptor closely related to

ACCEPTED MANUSCRIPT

the electron density is the shape function. The molecular shape function, or shape factor $\sigma(r)$ as introduced by Parr and Bartolotti in reference [14] characterizes the shape of the electron distribution and carries relative information about its electron distribution as described in references [15, 16, 17].
The next field used in this study is the Fukui function, a 3-dimensional descriptor that allows understanding and predicting relative reactivities of different sites in a molecule. Due to the discontinuity in the ρ versus N curve, two fukui functions need to be distinguished [18, 19], where at the point r, $f^+(r)$ and $f^-(r)$ are direct measures of reactivity toward nucleophilic and electrophilic attack. That is, regions where $f^+(r)$ is large capably stabilize additional electron density and hence are especially reactive towards electron-rich reactants. Regions where $f^-(r)$ is large readily give up their electrons, and are thus reactive towards electron-poor reactants.
The final conceptual DFT field introduced in the 3D QSAR context is the local softness, which describes the tendency of a particular site to be involved in "frontier-controlled" interactions [20]. As for the fukui function, two softness functions need to be distinguished. These has been shown to play a key-role in the application of the HSAB principle at local level [21, 22], and has been used in a variety of studies on regioselectivity [12].

3.3. Pretreatment of fields

It is well known in the QSAR methodology that a proper pretreatment of variables is crucial for the outcome of the analysis.
Intrinsic to 3D QSAR, there are many grid points with only minor variation in the field values throughout the compound set, e.g. the field values far outside the molecules. To eliminate such grid points, a 'minimum sigma' condition is defined. The grid points with a variation in the field value among the molecules lower than the minimum sigma value are not considered in the subsequent PLS analysis.
Some fields, as the default CoMFA fields and the electron density, have in close proximity to the surface of the atoms very steep slopes. It is common to truncate the field values at some arbitrary level to eliminate these points. It is necessary to use only well investigated cutoff values as the model quality depends critically on this value. Both the Fukui function and the softness field do not require any cutoff, as they are computed as finite differences. Hence, the large extremes in the unperturbed neutral system are compensated by nearly exactly the same extremes in the perturbed, anion or cation, density function.

ACCEPTED MANUSCRIPT

4. Antituberculotic activity of Salicylamides : a 3D QSAR case study

4.1. Dataset

44 derivatives of salicylamides [23], with possible antituberculotic activity, are used for 3D QSAR model development. All molecules have one of the five basic structures shown in figure 1, presented together with the possible substituents **R**. The compounds cover a promising group of potential drugs with a new, yet unrevealed, mechanism of action. They are assumed to serve as structural templates of inhibitors of the two-component regulatory system in bacteria[24]. The derivatives of salicylamides are tested for in vitro activity against *Mycobacterium tuberculosis*. As tuberculosis is still a considerable illness causing the death of more than a million people a year, it remains a main target for drug discovery. Analysis of structure-activity relationships of salicylamide derivates can produce fruitful suggestions for further research of antituberculotics. Using computed properties of the salicylamide derivatives we wish to identify the essential conditions depending on which the compounds elicit a stronger/weaker biological response.

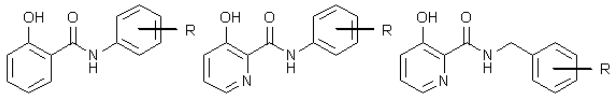
In vitro antimycobacterial activities of compounds 1 - 5 against Mycobacterium tuberculosis CNCTC My 331/88 are taken from the literature[23]. The antimycobacterial activities of the compounds are determined as minimal inhibitory concentrations (MIC) after incubation at 37 °C for 14 days. The concentrations of the compounds applied in the assay are 1000, 500, 250, 125, 62.5, 32, 31, 16, 8, 4 and 2 μmol/l, giving rise to a discrete character of MICs. The activities for the tested set of molecules are presented in table 1, with the compound name composed corresponding to the numbering of the basic structures and the name of the substituents in figure 1.

4.2. Algorithmic details

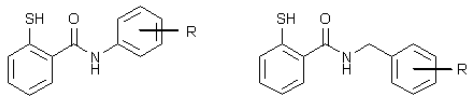
Any 3D QSAR analysis requires a list of algorithms and parameters which all need to be established beforehand. These are all investigated in depth in reference [25] and the final selected algorithms and parameters will be summarized here.

Starting from the set of molecules, a subset of compounds is selected as a training set by K-means clustering [26]. The test set contains five molecules, indicated with an asterisk in table 1.

ACCEPTED MANUSCRIPT



1 **2** **3**



4 **5**

R					
a	H	**f**	4-Br	**k**	3-NO_2
b	4-CH_3	**g**	4-F	**l**	4-OCH_3
c	4-Cl	**h**	3-F	**m**	N(CH_3)$_2$
d	3-Cl	**i**	4-CF_3	**n**	COOEt
e	3,4-Cl_2	**j**	4-NO_2	**o**	CN

Figure 1: Molecular set investigated for antituberculotic activity.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 1: Biological activities for the molecular set tested for antituberculotic activity on *Mycobacterium tuberculosis*. Molecules constituting the test set are indicated by an asterisk.

	MIC	-log(MIC)		MIC	-log(MIC)
1b	62.5	4.20	3e	62.5	4.20
1c	31	4.51	3g	250	3.60
1d	16	4.80	3i	125	3.90
1e	8	5.09	3j*	125	3.90
1f	16	4.80	4a	32	4.50
1g	62.5	4.20	4b	16	4.80
1h*	31	4.51	4c	32	4.50
1i	8	5.09	4d	32	4.50
1j	8	5.09	4e	62.5	4.20
1k	16	4.80	4f	62.5	4.20
1l	62.5	4.20	4g*	16	4.80
2a	250	3.60	4h	32	4.50
2b	125	3.90	4l	16	4.80
2c	62.5	4.20	5a	8	5.09
2d	62.5	4.20	5b	8	5.09
2f	62.5	4.20	5c	8	5.09
2g	125	3.90	5d	8	5.09
2i*	32	4.50	5e	32	4.50
2l	125	3.90	5f	8	5.09
3a	500	3.30	5g	8	5.09
3c	125	3.90	5j*	8	5.09
3d	125	3.90	5k	8	5.09

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

The molecular structures have been geometry optimized at the DFT level using the B3LYP functional and 6-31G* basis set as implemented in Gaussian03 [27].

As the electrostatic field is calculated as an atomic point charge approximation, an optimal partial point charge calculation scheme has to be selected. Several charge calculation methods has been investigated in [25], as the partial charges derived from Gasteiger-Marsili[28, 29], the MMFF94 force field [30], Mulliken population analysis [31], Natural population analysis [32], Hirshfeld analysis [33] and the iterative Hirshfeld analysis [34]. It was found that simpler methods, such as the Gasteiger-Marsili method, yield results of good quality, and thus can confidently be used to construct the electrostatic field.

There is no structural information available on the target or receptor environment. As a consequence, the alignment is restricted to these algorithms which do not make use of any information about the target. The molecular set is subjected to different so called ligand-based alignment techniques, in order to select the most promising alignment [25]. Out of several alignment techniques (TGSA[35], ROCS[36] and QSSA[37]), the ROCS algorithm implemented in the Openeye Software Suite [38] performed the best. ROCS stands for 'rapid overlay of chemical structures', a rigid body optimization process that maximizes the overlap volume between the compounds.

For the default CoMFA studies, with the electrostatic and steric field only, a spacing of 1 Å in a box extending to 4 Å outside the molecular van der Waals surface is preferred in order to have an accurate description of the field. These values are selected throughout history as the set of values that invokes a good and consistent CoMFA outcome. As there has been a limited number of preceding studies on the optimal parameters for quantum chemical fields [39, 40], it was necessary to set-up a study for the optimal spacing for these conceptual DFT quantum chemical fields. For this set of molecules, a grid spacing of 0.6 Å in a box extending to 1 Å outside the molecular van der Waals surface can be advocated. The current parameters are chosen as a balance between computational expense and consistent outcomes.

Each of the fields - except the Fukui function and the softness field - need to be allocated a cut-off value, as specified before. An optimal cut-off value can be found by scanning the performance of the models, slightly varying the cut-off value from a minimum to a maximum value. The optimal cut-off values are mentioned in table 2.

ACCEPTED MANUSCRIPT

Table 2: 3D QSAR model cut-off values.

Model	Field	|Cut-off value|
1	$\rho(r)$	1.07 $bohr^{-3}$
2	$\sigma(r)$	0.03 $bohr^{-3}$
3	f^{+}(r)	No Cut-off
4	f^{-}(r)	No Cut-off
5	s^{+}(r)	No Cut-off
6	s^{-}(r)	No Cut-off
7	Electrostatic potential	36.9 $\frac{kcal}{mol}$
8	Steric potential	30.1 $\frac{kcal}{mol}$

4.3. Results

The default molecular interaction and the quantum chemical molecular fields are subjected to a cross-validated 3D QSAR analysis. The models are obtained in a PLS model building with a leave-10-out cross-validation. The results are summarized in table 3. The quality of a QSAR model is characterized by its complexity, i.e., the number of latent variables (LV) necessary to describe the dataset and by the quality of prediction, the predictive correlation coefficient q^2 obtained by the leave-10-out procedure (q^2_{L10O}). A value higher than 0.50 indicates a model with predictive capacities.

Table 3: 3D QSAR model characteristics.

Model	Field	LV	q^2_{L10O}
1	$\rho(r)$	4	0.68
2	$\sigma(r)$	2	0.69
3	f^{+}(r)	4	0.68
4	f^{-}(r)	2	0.75
5	s^{+}(r)	4	0.67
6	s^{-}(r)	2	0.76
7	Electrostatic potential	4	0.66
8	Steric potential	3	0.65

Model 6 using the softness s^{-} field as independent variable is, based on the internal predictive power, statistically the most significant within the set of quantum chemical fields. The model can be set up successfully using only two latent variables. In order to make a fair comparison between the classical and the quantum chemical approach for the problem, a full statistical analysis has been performed for model 6 as well as for the two default CoMFA fields (model 7 & model 8). The results are summarized in table 4. The number of latent variables and the predictive power are repeated from the previous table, the goodness-of-fit is given by R^2. The internal predictivity test indicates

ACCEPTED MANUSCRIPT

the predictivity of the model, but the only true test on the predictive power of a model can be obtained through the application of the final model on the external test set and is quantified by the value of q^2_{ext} and the root mean square error (rmse) on the predicted values.

Table 4: Statistical parameters for selected models

Model	LV	R^2	q^2_{L10O}	q^2_{ext}	rmse
6	2	0.81	0.76	0.77	0.22
7	4	0.80	0.66	0.88	0.16
8	3	0.84	0.65	0.89	0.14

4.4. Discussion

4.4.1. Behavior of the Quantum chemical fields

As 3D QSAR results are primarily judged based on the predictivity, the quantum chemical fields will primarily be compared with each other based on the q^2_{L10O} value. The fields presented in the first part of table 3 can all be classified as predictive. Especially the f^- and s^- fields yield promising results. The electron density gives a model with a lower predictivity. The shape function gives a slightly more predictive model than the electron density at a lower number of latent variables, but the improvement in quality is rather small. This is not surprising as the molecules in the training set are of similar size and thus of comparable electron number. Therefore, the extra information included in the shape function disappears.

The fact that the Fukui f^- and the softness s^- fields give predictive models at only 2 latent variables can be a hint towards the mechanisms of action. Both fields indicate an electrophilic attack from the receptor towards the ligands.

4.4.2. Comparison of 2D QSAR and 3D QSAR

Another well investigated kind of QSAR analysis is 2D QSAR. In contrast to 3D QSAR, 2D QSAR does not make use of local field descriptors. As a consequence, far less number of descriptors are fed into the analysis, which simplifies the statistics to multiple linear regression.

It is clear that 2D QSAR - by its very nature - is much more straightforward than 3D QSAR. More user intervention is necessary in each step in 3D QSAR. As the current set of molecules has been investigated in a 2D QSAR analysis[41], a comparison can be made between 2D QSAR and 3D QSAR. The results for the 2D QSAR analysis are summarized in table 5, together with the 3D QSAR analysis results. Despite the fact that the 2D QSAR analysis was performed without an

ACCEPTED MANUSCRIPT

Table 5: Comparison between 2D QSAR and 3D QSAR models. The 2D QSAR model is based on 45 molecules, no test set is selected. The 3D QSAR models are based on 39 molecules, with a test set of 5 molecules.

	2D QSAR	3D QSAR		
Model		6	7	8
Number of molecules	45	39	39	39
Goodness-of-fit	0.91	0.81	0.80	0.84
Number of descriptors	5	2	4	3
Internal predictivity	0.89	0.76	0.66	0.65
External predicitivity	/	0.77	0.88	0.89

external test set, the internal predictivity and the goodness-of-fit indicate the superiority of the 2D QSAR model. Based on the results presented, one would favor 2D QSAR model for the current problem.

The mechanistic interpretation of the 2D QSAR model indicates electrostatic interactions at some points [41]. And at this point, the 3D QSAR models are favored on the 2D QSAR models as with these models one can pinpoint the exact locations where the electrostatic interactions with the environment are favored.

Thus based on the simplicity and the statistical parameters, one would favor the 2D QSAR model building, but the 3D QSAR models still have their own benefit as they give a 3-dimensional interpretation of the results.

4.4.3. Comparison of the Default CoMFA and the Quantum chemical models

As new fields are introduced in 3D QSAR, their performance has to be compared with the performance of the classical fields in CoMFA, in order to investigate whether these fields are worth the effort to calculate for future applications. Therefore the most promising quantum chemical 3D QSAR model, the model based on the s^- (model 6), will be highlighted and confronted with the models based on the default CoMFA fields (model 7 & model 8) and tested on some essential QSAR conditions, published as preliminary guidelines from the European Commission in reference [42]. Table 4 summarizes the statistical results for the three models.

Robustness : A clear correlation between the chosen descriptors and the target property, as can be seen by the goodness-of-fit, R^2, is apparent. The three models have a similar behavior concerning the goodness-of-fit, so no distinction can be made between the fields based on this one statistical criterion.

ACCEPTED MANUSCRIPT

Predictivity : The main goal of any QSAR is the prediction of the activity of new compounds, so the internal predictivity is an important criterion. In table 4 it is clear that the internal predictive capacity of the model with the quantum chemical field, softness s^-, is higher than for both default considered CoMFA fields. The application of the models on the external test set of compounds establishes the real predictive power of the model. However, the external test set may not be involved in the model building at any point. Thus the results on the external test set may not be used in any decision-making about the superiority of one field over the others, but only as an external confirmation of the predictive power of a model. Based on these external results, the three fields all give rise to predictive models.

Explanatory power : Knowledge about the explanatory power enables the user to suggest a possible mode of action for the active compounds and to propose ameliorations on possible future ligands. A graphical analysis of the coefficient fields is given in figures 2-4. These give a rough location from which structure activity relationship statements can be inferred, discriminating areas where fields are important from those that have no significance. Positive values for the steric field (orange) in figure 2 indicate areas where an increase in steric hindrance would produce higher activity, and areas showing negative values (green mesh) are those where steric bulk would decrease the activity of the ligands. Positive values for the electrostatic field (red mesh) in figure 3 indicate those areas where a highly positive electrostatic potential would produce higher activity, promoting electropositive ligands at these sites. Areas showing negative values (blue) are those where a highly negative electrostatic potential would increase the activity, thus promoting electronegative ligands at these sites. The coefficients for the softness field are given in figure 4. While the electrostatic potential indicates those regions that an electrophile or an nucleophile is likely to approach, it is also important to know how readily available the electrons in those regions are. The map of the softness function indicates the susceptibility to electrophilic or nucleophilic attack. Positive values (green) for the softness s^- map indicate those areas where an electron-rich center would increase the activity. Negative values (yellow mesh) for the softness map indicate those areas where an electron-rich center would decrease the activity. It is thus clear that the three fields considered all can be interpreted readily for the problem at hand.

ACCEPTED MANUSCRIPT

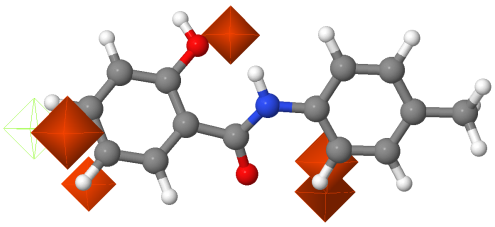


Figure 2: Stdev*coeff map showing areas of favorable and unfavorable steric field contributions to the activity.

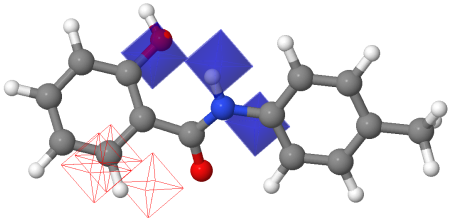


Figure 3: Stdev*coeff map showing areas of electrostatic field contributions to the activity.

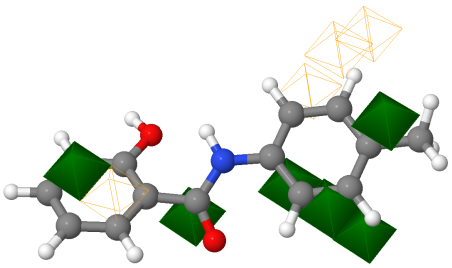


Figure 4: Stdev*coeff map showing areas of softness field contributions to the activity.

ACCEPTED MANUSCRIPT

Table 6: Summary of conditions for QSAR model building, applied on model 6, 7 & 8

	Model 6	Model 7	Model 8
Robustness	+	+	+
Predictive power			
Internal	++	+	+
External	+	++	++
Explanatory power	+	+	+
Simplicity			
Computer time	-	+	+
Uniqueness	+	-	-

Simplicity and Uniqueness : The fact that both CoMFA fields are much easier to calculate than the quantum chemical fields in 3D QSAR, favors them clearly. The screening of the molecular set of 39 molecules takes 1 minute for both classical fields on one computer and 5 minutes for the softness field on a cluster of 20 computers. However, the main drawback of the default considered CoMFA field is their lack in uniqueness, due to the variable parameters that constitute their expression, as e.g. the plethora of possibilities in calculating the condensed charges.

Gathering these four criteria together, it is clear that none of the 3D fields can be favored above the other fields. The three models behave well for the conditions mentioned, as is summarized in table 6. Thus, although the default 3D QSAR fields are useful in obtaining predictive 3D QSAR models for the current problem, the quantum chemical approach is worth the effort, due to the extra information obtained and the uniqueness in obtaining and using the quantum chemical field.

The 3D QSAR results can be ameliorated by combining several 3D molecular fields into one model. As the main purpose of the current article was to indicate the usefulness of individual quantum chemical fields in 3D QSAR, we do not pursue this further but rather refer the reader to Van Damme et al. [43] for an example where different fields are combined.

5. Conclusion

In this study the behavior of the default CoMFA fields is confronted with the conceptual DFT fields in 3D QSAR analysis. Properties like electron densities or Fukui functions for electrophilic and nucleophilic attack can be used as physically meaningful descriptors. The example indicates that the newly introduced fields are worth the effort to be used in 3D QSAR problems as they can

ACCEPTED MANUSCRIPT

have a similar behavior compared to the default 3D QSAR field and they deliver new interpretable information, not included in the default 3D QSAR fields. Especially in those cases where the default considered CoMFA fields do not carry the necessary information for a predictive 3D QSAR model, it is our advise that conceptual DFT quantum chemical fields should be considered.

6. References

[1] R. Car, M. Parrinello, Phys. Rev. Lett. 55 (1985) 2471.

[2] L. P. Hammett, Physical Organic Chemistry. Reaction rates, Equilibria and Mechanisms, Mc Graw-Hill, New York, 1970.

[3] C. Hansch, T. Fujita, J. Am. Chem. Soc. 86 (1964) 1616.

[4] S. M. Free, J. W. Wilson, J. Med. Chem. 7 (1964) 395.

[5] G. M. Downs, in : P. Bultinck, H. De Winter, W. Langenaecker, J. P. Tollenaere (Eds.), Similarity and Quantitative Structure-Activity Relationships in Computational medicinal Chemistry for Drug Discovery, Decker Inc., New York, 2004, pp. 295-322.

[6] R. D. Cramer, M. Milne, Abstr. Paper. Am. Chem. Soc. Natl. Meet. (1979) COMP44.

[7] R. D. Cramer, D. E. Patterson, J. D. Bunce, J. Am. Chem. Soc. 110 (1988) 5959.

[8] S. Wold, M. Sjöström, L. Eriksson, in : J. Gasteiger (Ed.), Encyclopedia of Computational Chemistry, John Wiley & Sons Inc., New York, 1994.

[9] D. J. Abraham, G. E. Kellogg, J. Comput.-Aided Mol. Des. 8 (1994) 41.

[10] K. H. Kim, G. Greco, E. Novellino, C. Silipo, A. Vittoria, J. Comput.-Aided Mol. Des. 7 (1993) 263.

[11] P. Gaillard, P. A. Carrupt, B. Testa, A. Boudon, J. Comput.-Aided Mol. Des. 8 (1994) 83.

[12] P. Geerlings, F. De Proft, W. Langenaeker, Chem. Rev. 103 (2003) 1793.

[13] P. W. Ayers, J. S. M. Anderson, L. J. Bartolotti, Int. J. Quantum Chem. 101 (2005) 520.

[14] R. G. Parr, L. J. Bartolotti, J. Phys. Chem. 87 (1983) 2810.

ACCEPTED MANUSCRIPT

[15] P. W. Ayers, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 1959.

[16] P. Bultinck, R. Carbo-Dorca, J. Math. Chem. 36 (2004) 191.

[17] P. W. Ayers, Phys. Rev. A 71 (2005) 62506.

[18] P. W. Ayers, M. Levy, Theor. Chem. Acc. 103 (2000) 353.

[19] P. W. Ayers, R. G. Parr, J. Am. Chem. Soc. 122 (2000) 2010.

[20] W. T. Yang, R. G. Parr, Proc. Natl. Acad. Sci. U. S. A. 82 (1985) 6723.

[21] P. K. Chattaraj, H. Lee, R. G. Parr, J. Am. Chem. Soc. 113 (1991) 1855.

[22] F. Mendez, J. L. Gazquez, J. Am. Chem. Soc. 116 (1994) 9298.

[23] K. Waisser, M. Pesina, P. Holy, M. Pour, O. Bures, J. Kunes, V. Klimesova, V. Buchta, P. Kubanova, J. Kaustova, Arch. Pharm. 336 (2003) 322.

[24] M. J. Macielag, J. P. Demers, S. A. Fraga-Spano, D. J. Hlasta, S. G. Johnson, R. M. Kanojia, R. K. Russel, Z. H. Sui, M. A. Weidner-Wells, H. Werblood et al., J. Med. Chem. 41 (1998) 2939.

[25] S. Van Damme, Quantum Chemistry in QSAR. Quantum chemical descriptors : Use, Benefits and drawbacks., Ph.D. thesis, Ghent University, 2009.

[26] A. Golbraikh, M. Shen, Z. Y. Xiao, Y. D. Xiao, K. H. Lee, A. Tropsha, J. Comput.-Aided Mol. Des. 17 (2003) 241.

[27] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant et al., Gaussian 03, Revision E.01 (2007).

[28] J. Gasteiger, M. Marsili, Tetrahedron Lett. 34 (1978) 3181.

[29] J. Gasteiger, M. Marsili, Tetrahedron 36 (1980) 3219.

[30] T. A. Halgren, J. Comput. Chem. 17 (1996) 490.

[31] R. S. Mulliken, J. Chem. Phys. 23 (1955) 1833.

[32] A. E. Reed, L. A. Curtiss, F. Weinhold, Chem. Rev. 88 (1988) 899.

ACCEPTED MANUSCRIPT

[33] F. L. Hirshfeld, Theor. Chim. Acta 44 (1977) 129.

[34] P. Bultinck, C. Van Alsenoy, P. W. Ayers, R. Carbo-Dorca, J. Chem. Phys. 126 (2007) 144111.

[35] X. Girones, D. Robert, R. Carbo-Dorca, J. Comput. Chem. 22 (2001) 255.

[36] J. A. Grant, M. A. Gallardo, B. T. Pickup, J. Comput. Chem. 17 (1996) 1653.

[37] P. Bultinck, T. Kuppens, X. Girones, R. Carbo-Dorca, J. Chem. Inf. Comp. Sci. 43 (2003) 1143.

[38] Oechem, Openeye Scientific Software, Inc., Santa Fe (2008).

[39] A. Poso, R. Juvonen, J. Gynther, QSAR Comb. Sci. 14 (1995) 507.

[40] M. Rahnasto, H. Raunio, A. Poso, C. Wittekindt, R. O. Juvonen, J. Med. Chem. 48 (2005) 440.

[41] R. Dolezal, P. Bultinck, S. Van Damme, K. Waisser, Eur. J. Med. Chem. 44 (2009) 869.

[42] A. P. Worth, A. Bassan, A. Gallegos, T. I. Netzeva, G. Patlewicz, M. Pavan, I. Tsakovska, M. Vracko, Technical report, European Commission, Joint Research Centre, 2007.

[43] S. Van Damme, P. Bultinck, J. Comput. Chem. 30 (2009) 1749.

ACCEPTED MANUSCRIPT