MTD-PLS: A PLS-Based Variant of the MTD Method. 2. Mapping Ligand−Receptor Interactions. Enzymatic Acetic Acid Esters Hydrolysis

Ludovic Kurunczi,*,† Marius Olah,‡ Tudor I. Oprea,‡,§ Cristian Bologa,‡ and Zeno Simon‡

Faculty of Pharmacy, University of Medicine and Pharmacy of Timisoara, P-ta Eftimie Murgu 3, 1900-Timisoara, Romania, Department Chemistry, University of West, Str. Pestalozzi 16, 1900-Timisoara, Romania, and EST Lead Informatics, AstraZeneca R and D. Mölndal, S-43183 Mölndal, Sweden

Received December 3, 2001

The PLS variant of the MTD method (T. I. Oprea et al., SAR QSAR *Environ. Res.* **2001**, *12*, 75−92) was applied to a series of 25 acetylcholinesterase hydrolysis substrates. Statistically significant MTD-PLS models (q^2 between 0.7 and 0.8) are in agreement with previous MTD models, with the advantage that local contributions are understood beyond the occupancy/nonoccupancy interpretation in MTD. A "chemically intuitive" approach further forces MTD-PLS coefficients to assume only negative (or zero) values for fragmental volume descriptors and positive (or zero) values for fragmental hydrophobicity descriptors. This further separates the various kinds of local interactions at each vertex of the MTD hypermolecule, making this method suitable for medicinal chemistry synthesis planning.

INTRODUCTION

The first note of this series[1] described a PLS variant of the MTD (minimum topological difference) method[2] aimed at describing steric, hydrophobic, electrostatic, and hydrogen bonding characteristics of the binding site regions in the vicinity of ligand atom positions. The key MTD concept—the hypermolecule, obtained by atom-to-atom superpositions of ligand molecules in a topological manner[2]—is extended in MTD-PLS by attributing additional properties for each atom of molecule ***i*** occupying a vertex ***j*** of the hypermolecule. In the present study these computed properties are the van der Waals fragmental volume (V_{ij}), hydrophobicity constants (H_{ij}), electric charge (S_{ij}), and polarizability constants (P_{ij}). The hypermolecule in itself can be extended to three dimensions, by sampling multiple conformers,[3] but only one conformer per compound was considered in this study. With its simple superimposition method, MTD-PLS is intended as a simple (topologically based) scoring scheme[1] that could—once fully automated and validated—serve as a QSAR-based scoring function in the absence of relevant receptor information.[4] Let us remember that the PLS variant of MTD developed in the first note[1] of this series has several similarities with the Molecular Field Topology Analysis (MFTA) method of the Palyulin group.[5]

In the MTD method,[2] vertices are binned in three categories, beneficial, detrimental, or irrelevant, and are mapped into an occupancy/nonoccupancy type of interaction. The exact nature of this interaction, e.g., steric, hydrophobic, or electrostatic, cannot be directly established from classical MTD. Furthermore, a bias in the vertex attribution—beneficial, detrimental, or irrelevant—may occur in the classical MTD method, due to its rather simplistic optimization method.[2] This bias does no longer occur within the mechanism of extracting the principal components in the MTD-PLS[1] variant, as discussed further in this paper.

If MTD-PLS were an additive approach, then the terms for steric misfit (detrimental) would be expected to yield negative regression coefficients for the V_{ij} terms. By the same token, the hydrophobic effects would generally be expected to have beneficial effects for those regions with "positive" V_{ij} coefficients, i.e., positive H_{ij} coefficients, as the receptor cavities are generally less polar compared to the surrounding water environment. However, this rule of thumb does not always hold true, due to the lack of "chemical awareness" of the software that does not distinguish parameters characterizing nonpolar or polar atoms. It could be also related to the impossibility to obtain, from a QSAR model, information that is more accurate than the data describing the QSAR series, i.e., the information fed into the model via experimental activities and structural parameters describing the ligands.[6]

Here we describe an additive approach based on intermolecular force categories, where the regression coefficients observe the "chemical intuition"—negative for fragmental volumes V_{ij} at the detrimental vertices (steric misfit should always be reflected by coefficients reducing the activity, and V_{ij} values are always positive) and positive (or zero) for H_{ij} fragmental values (i.e., beneficial or neutral). This sign convention for hydrophobicity is enforced because van der Waals attractions and entropic effects related to partial (fragmental) dehydration are usually related to positive contribution to the ligand binding affinity, in particular for hydrophobic fragments. In this convention, the effect of electrostatic interactions for polar groups is given by the corresponding coefficients for S_{ij} and hydrogen bonding

* Corresponding author phone: +40 56 191818; fax: +40 56 191824; e-mail: icht@mail.dnttm.ro or dick@acad-tim.utt.ro. Corresponding address: Romanian Academy, Institute of Chemistry, Bd. Mihai Viteazul nr. 24, RO-1900 Timisoara, Romania.
† University of Medicine and Pharmacy of Timisoara.
‡ University of West.
§ EST Lead Informatics.

10.1021/ci010124a CCC: $22.00
Published on Web 06/18/2002

Table 1. Series of Acetic Acid Esters (CH_3COOR)

I	R	Y^a	j^b
1.	C_6H_5	6.72	**2**−**6**
2.	$CH_2CH_2C(CH_3)_3$	6.30	**4**−**8**
3.	$CH_2CH_2SC_2H_5$	5.40	**5**−**7**, **9**
4.	$CH_2SC_2H_5$	5.35	5−7; 4−6; **5**, **6**, **8**
5.	$CH_2CH_2CH(CH_3)_2$	5.32	**4**−**7**; 4−6, 8; 5−8
6.	$CH_2CH_2NO_2$	5.20	**4**−**7**; 4−6, 8; 5−8
7.	CH_2CH_2Cl	5.02	**5**, **6**; 6, 10
8.	$CH_2C{\equiv}CH$	4.81	**6**, **11**
9.	*n*-C_5H_{11}	4.74	**5**−**7**, **9**
10.	*n*-C_7H_{15}	4.75	**5**−**7**, **9**, **12**, **13**
11.	$(CH_2)_4SC_2H_5$	4.73	**5**−**7**, **9**, **12**, **13**
12.	*cyclo*-C_6H_{11}	4.71	**2**−**6**
13.	*n*-C_4H_9	4.72	4−6; 5, 6, 8; **5**−**7**
14.	*n*-C_6H_{13}	4.68	**5**−**7**, **9**, **12**
15.	$(CH_2)_3SC_2H_5$	4.67	**5**−**7**, **9**, **12**
16.	$CH_2C_6H_5$	4.66	**5**−**7**, **10**, **14**, **15**
17.	$CH_2CH(CH_3)_2$	4.32	**5**, **6**, **10**
18.	$CH_2CH{=}CH_2$	4.10	**5**, **6**; 6, 10
19.	*n*-C_3H_7	3.91	**5**, **6**; 6, 10
20.	$CH(CH_3)C_2H_5$	3.69	**1**, **5**, **6**; 2, 5, 6
21.	C_2H_5	3.36	**6**
22.	CH_3	3.00	-
23.	$CH(CH_3)_2$	2.72	1, 2; **2**, **6**
24.	$C(CH_3)_3$	1.30	**1**, **2**, **6**
25.	$C(CH_3)_2C_2H_5$	1.30	1−3, 6; **1**, **2**, **5**, **6**

[a] Y − logarithms of second-order hydrolysis rate (k_2/k_M) according to ref 7. [b] j − vertex occupation for the low energy conformations according to ref 3; bold: structures considered in this work.

parameters (i.e., HD_{ij} for donors and HA_{ij} for acceptors).[1] This method is applied here to hydrolysis data for a series of acetate esters catalyzed by acetylcholinesterase studied by Järv,[7] for which we have previously established MTD QSARs.[3,8] The results are compared with the X-ray crystallography data for acetylcholinesterase (AChE) in complex with acetylcholine.[9]

MATERIALS AND METHODS

Experimental Data. Second-order enzymatic hydrolysis rates (k_2/k_M) for acetic acid esters incubated with AChE from cobra venom (*Naja naja oxiana*) are taken from Järv, Kesvatera, and Aaviksaar.[7] The 25 molecules and their biological activities are listed in Table 1.

The PLS variant of MTD method was described elsewhere.[1] We recall that by atom-per-atom superimposition (neglecting hydrogens) of the molecules of interest within a given series, a hypermolecule with **M** vertices ***j*** results. For each molecule ***i*** in this series, and for each atom occupying a hypermolecular vertex ***j***, we compute the van der Waals volumes V_{ij}, the hydrophobicity constants H_{ij}, the electric charges S_{ij} etc. In MTD-PLS, biological activities are computed using eq 1

$$\hat{Y}_i = a_0 + \sum_{j=1}^{M}\sum_{\mu=1}^{K} a_{j\mu} x_{ij\mu} \qquad (1)$$

where $\hat{Y}_i$ are the calculated biological activities, $a_{j\mu}$ are the regression (pseudo)coefficients, μ refers to the K types of computed properties (steric, hydrophobic, etc.), and $x_{ij\mu}$ are the vertex variables associated with each molecule ***i***, vertex ***j***, and property μ. The $x_{ij\mu}$ values are set equal to the value of the μth structural parameter for the atom (X) or fragment (XH_n) occupying vertex j and molecule i, and zero, if the corresponding vertex is unoccupied in this molecule. So the hydrogen atoms are "contracted" into the X atoms to which they are bound, i.e., no supplementary j vertices are introduced for the H-atoms.

For the acetic acid ester series, N = 25, M = 15, and K = 4 (V, H, S, and P). The $x_{ij\mu}$- structural parameters form a matrix with N = 25 lines and 64 columns (15 × 4 plus the charges S for the first four heavy atoms (groups) of the CH_3-COOCH2- moiety). To follow the "chemical intuition" conditions

$$a_{jV} \le 0 \qquad (2a)$$

$$a_{jH} \ge 0 \qquad (2b)$$

we proceed as follows. The MTD-PLS procedure[1] is reiterated for each case in which conditions (2a) and (2b) are not satisfied in all columns; those columns were eliminated, and the model is recomputed with the new reduced matrix, until conditions (2a) and (2b) are satisfied in all columns.

Structural Atomic Parameters. Atomic van der Waals volumes, V_{ij}, are calculated by a MLR procedure described in previous notes.[1,10] Hydrophobic fragments, H_{ij}, are taken mainly from the data of Hansch et al.[11] Partial atomic charges, S_{ij}, are calculated using the semiempirical AM1-method,[12] as implemented in Hyperchem program package.[13]

Atomic (fragment) polarizabilities, P_{ij}, are computed from molecular refractivity,[14] as follows. The first step is the calculation of molecular refractivity from bonding refractivities for some simple molecules. The second step is to calculate, from these molecular refractivities, the atomic (fragmental) refractivities. For example:

$$-CH_3: R_{CH_3-CH_3} = 6\cdot R_{C-H} + R_{C-C} = 11.352 \Rightarrow R_{CH_3-} = R_{CH_3-CH_3}/2 = 5.676;$$

$$>C<: R_{C(CH_3)_4} = 12\cdot R_{C-H} + 4\cdot R_{C-C} = 25.296 \Rightarrow R_{>C<} = R_{C(CH_3)_4} - 4\cdot R_{CH_3} \Rightarrow R_{>C<} = 2.59;$$

$$CH_{ar}: R_{C_6H_6} = 6\cdot R_{C-H} + 6\cdot R_{C_{ar}-C_{ar}} = 26.184 \Rightarrow R_{CH_{ar}} = R_{C_6H_6}/6 = 4.36.$$

This method was applied for all fragmental refractivities. The structural parameters used here are listed in Table 2. Atomic (fragment) polarizabilities, P_{ij}, were considered since they intervene in charge-induced dipole interactions and in augmented van der Waals interactions, e.g., between stacked, aromatic moieties. There is generally a good correlation between V and P parameters, but for atoms within aromatic systems, *P* values are sensibly higher than usual. Since no atoms within the alcohol moiety (with one exception, the NO_2 group for compound **6**) are hydrogen bonding, the corresponding parameters were not found useful for this series.

Atomic volumes, V, are given in Å^3, hydrophobicities, H, in the usual Hansch units, S, in fractional electronic charges, P, in refractivity increments (cm^3/mol). The values used for V, H, S, and P, for different types of atoms and XH_n fragments are listed in Table 2.

RESULTS AND DISCUSSIONS

MTD-PLS, in Its Previous Version.[1] The hypermolecule (M = 15 vertices) derived from the series of 25 esters (Table

Table 2. Fragmental vdW Volumes, V; Hydrophobicity, H; and Polarizability Constants, P[a]

fragment	V	H	P
-CH_3	22.73	0.86	5.68
-CH_2-	17.02	0.53	4.62
>CH-	11.41	0.16	3.62
>C<	5.59	−0.16	2.59
H_2C=	19.82	0.67	5.44
=CH-	14.15	0.30	4.40
HC≡	18.84	0.19	4.61
≡C−	12.43	−0.13	3.77
CH (arom)	13.57	0.36	4.36
C (arom)	8.36	0.22	3.34
-S− (aliph)	19.14	−0.82	7.92
N (NO_2) (aliph)	8.85	0.00	1.34
O (NO_2) (aliph)	9.68	−0.56	2.68
-Cl (aliph)	20.24	0.10	5.86

[a] V − in Å^3; H − Hansch type hydrophobicities; P − molar polarizabilities, cm^3.

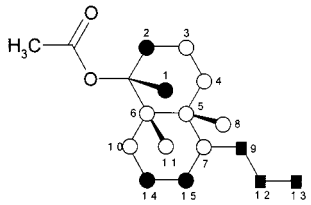

Figure 1. Hypermolecule. Series of acetic acid esters. Vertex numbering. Open circles − beneficial, black circles − detrimental, squares − irrelevant vertices according to the results of the classical MTD method.[3]

1) is shown in Figure 1. As some of these compounds have a large conformational flexibility, only the conformers with the lowest MTD (lowest misfit to the receptor site) were selected, as previously described.[3] The hypermolecular vertex occupancies corresponding to some low energy conformations are also listed in Table 1 (bold: the structures considered in this work). We have also investigated the effect of adjusting biological activity in accordance to conformational energy for this series.[8]

Throughout the PCA and PLS calculations the data were preprocessed by means of centering and autoscaling to unit variance. In a first attempt we used the S, H, and V parameters that were found relevant for the dioxin and biphenyl series,[1] to which charges for the acetate moiety atoms were added. In addition to these 49 variables, those used in our previous work,[3] namely the σ^* Taft constants, the π Hansch hydrophobicities, and the E_S steric constants (for the R group in CH_3COOR, see Table 1) were included in initial model. However, the σ^*, π, and E_S values were eliminated during the preliminary PCA models. The first PLS model (**M4**) with the nominalized parameters and the carboxy moiety charges, with N = 25 molecules, 49 structural variables and one PLS component, had the following statistical characteristics

$$r_x^2 = 0.121;\ r_y^2 = 0.835;\ q^2 = 0.597 \qquad (3)$$

where r_x^2 and r_y^2 are the fractions of the explained variance in the X and the Y block, respectively, and q^2 is the cross-validated coefficient for the PLS model, as provided in SIMCA program package.[15]

After three successive PLS models, compounds **1**, **16**, and **22** were eliminated based on the DMod(Y) criterion.[16] A rationale for this outlier behavior is that compounds **1** and **16** are the only substrates that contain phenyl groups, and it is likely that this structure does not overlap well with others in the hypermolecule of Figure 1. Compound **22** is the smallest (R = CH_3) and does not occupy any of the numbered vertices in the hypermolecule.

Table 3. VIP[15,16] Values and $a_{j\mu}$ Coefficients (Cumulative Coeff [2][15] Free Term a_0 = 20.43) for Eq 1 for the PLS Model **M8**[a]

variable	VIP[2] (cum)	coeff [2]	variable	VIP[2] (cum)	coeff [2]
1S	2.14	−12.22	3V	0.53	0.02
1V	2.06	−0.02	3H	0.53	0.52
1H	2.06	−0.43	7S	0.48	−0.67
2H	1.83	−0.44	V12	0.44	0.00(1)
2S	1.77	−9.99	9H	0.43	0.12
2V	1.75	−0.01	5V	0.42	0.01
SdC	1.71	−28.48	12H	0.42	0.03
7V	1.41	0.01	13V	0.37	−0.00(0)
SdO	1.40	22.79	13H	0.37	−0.01
4V	1.26	0.02	6S	0.32	0.39
SsC	1.23	1.79	11H	0.30	−0.11
6V	1.11	−0.03	11V	0.30	−0.00(0)
8V	1.04	0.01	11S	0.30	−0.03
8H	1.04	0.39	13S	0.30	−0.36
6H	0.97	−0.19	9S	0.26	0.71
4H	0.82	0.29	5S	0.24	0.24
7H	0.81	0.14	10S	0.20	−3.56
8S	0.78	5.29	10V	0.20	−0.00(4)
9V	0.69	0.00(4)	10H	0.20	−0.10
5H	0.66	−0.16	12S	0.17	0.54
SsO	0.62	−3.24	4S	0.15	0.05
3S	0.53	25.07			

[a] See text for details. **M8** − PLS model (N = 22 molecules). SdC, SsC, SdO, SsO: charges at atoms of the CH_3COOCH_2 moiety (>C=, C in CH_3, carboxylic =O, and respectively −O-).

In this condition a stable, final PLS model (**M8**) was obtained, with two PLS components:

$$r_x^2 = 0.271\begin{pmatrix}0.168\\0.103\end{pmatrix};\ r_y^2 = 0.951\begin{pmatrix}0.888\\0.063\end{pmatrix};\quad q^2 = 0.792\begin{pmatrix}0.754\\0.154\end{pmatrix} \qquad (4)$$

The elimination of the above three compounds also eliminated vertices {$\boldsymbol{j}$ = 14, 15}; therefore, the model contains 22 molecules and 43 variables (V_{ij}, H_{ij}, S_{ij}). Compared to our previous results, r_y^2 = 0.867 and q^2 = 0.795, **M8** (see eq 4) is quite similar, despite the fact that **M8** does not include σ^* Taft constants, as our previous study.[3] The $a_{j\mu}$ and VIP[16] (variable importance in projection) values for each descriptor are listed in Table 3.

The polarizability parameter, P_{ij}, for all 25 molecules, was added in the **M5** model, with a total of 64 variables, including the charge parameters SsC, SsO, SdC, SdO of the acetate moiety—see Table 3. Statistical parameters of **M5** can be summarized as

$$r_x^2 = 0.121,\ r_y^2 = 0.847,\ q^2 = 0.614 \qquad (5)$$

The improvement of **M5**, based on DMod(Y) and t-score outliers, yields the stable PLS model **M13**—that is similar to **M8** (22 molecules, 56 structural parameters, compounds **1**, **16**, and **22** were eliminated). The correlation results are

$$r_x^2 = 0.288\begin{pmatrix}0.168\\0.121\end{pmatrix},\ r_y^2 = 0.952\begin{pmatrix}0.895\\0.057\end{pmatrix},\ q^2 = 0.815\begin{pmatrix}0.787\\0.233\end{pmatrix} \quad (6)$$

Concerning the above results, and also the $a_{j\mu}$ and the corresponding VIP values, the addition of structural parameter P yields only small modifications.

The "Chemically Intuitive" MTD-PLS Variant. Starting from the **M5** model, with all 25 molecules and the constraints imposed by conditions (2a) and (2b), the V columns for {$\boldsymbol{j}$ = 3−5, 7−15} and the H columns for {$\boldsymbol{j}$ = 1, 2, 5, 6} were eliminated. Model **M14**, based on 25 molecules and 48 structural parameters, resulted:

$$r_x^2 = 0.117,\ r_y^2 = 0.853,\ q^2 = 0.576 \quad (7)$$

Starting from the PLS stable **M13** model, the (2a) and (2b) constraints removed the V columns for {$\boldsymbol{j}$ = 3−5, 7−9, 11, 12} and the H j columns for {$\boldsymbol{j}$ = 1, 2, 5, 6, 10, 12, 13}. We recall that vertices {$\boldsymbol{j}$ = 14, 15}, and their respective columns, were also eliminated together with compounds **1**, **16**, and **22**. Thus the model **M16** is based on 22 molecules and 41 structural parameters:

$$r_x^2 = 0.280\begin{pmatrix}0.171\\0.109\end{pmatrix},\ r_y^2 = 0.949\begin{pmatrix}0.883\\0.066\end{pmatrix},\ q^2 = 0.796\begin{pmatrix}0.749\\0.190\end{pmatrix} \quad (8)$$

The predictive q^2 values for the models respecting eqs 2a and 2b could be questionable. Unfortunately our series is too small to partition it in learning and test sets. Nevertheless we try to estimate the general validity of the resulted models by a comparison with X-ray crystallography data (see below the corresponding section).

The $a_{j\mu}$ coefficients are listed in Table 4 for model **M16**. The results are listed as an M×K matrix; the first figure is the $a_{j\mu}$ coefficient; the figure in parentheses is the VIP-value. Comparing the VIP values from Tables 3 (**M8**) and 4 (**M16**), it can be seen that the significant contributions of the original variables to the models are quite similar. The role of the "eliminated" coefficients (a_{jV} and a_{jH}, for $\boldsymbol{j}$ values see above) is successfully superseded by the corresponding a_{jP} values. A single exception exists: a_{6H} is three times more important in Table 3 than the corresponding a_{jP} in Table 4. Model **M16**, derived from **M13**, should be considered more reliable from the statistical viewpoint.

Comments Concerning the MTD-PLS Method and Interpretation of the Results. In this work eq 1 approximates ligand−receptor affinities by the sum of interaction potentials of different types between ligand atoms (fragments) and binding sites. The electrostatic interaction, for example, is proportional to the charges bore by the ligand atoms (fragments). Thus the use of $x_{ij\mu} = 0$ (μ = charge) value for the absence of a real atom or fragment occupying a vertex in a molecule has the same effect as the zero charge in the same position. This is true also for other structural parameters.

The values for volumes of atoms (or XH_n fragments) occupying vertices do not differ very much (the same holds for polarizabilities). One could simply replace them by an indicator variable (0 unoccupied, 1 occupied) without a dramatic change of the statistics. We tried this using instead of volumes the indicator variables, and the statistics do not differ much (final model very similar to **M16**, with r_y^2 = 0.924, q^2 = 0.749 and the significant contributions of the original variables almost the same). We nevertheless do not regard as correct, from the physical, molecular point of view, to consider the absence of an atom (fragment) in a vertex belonging to a molecule as missing value. Namely, an atom in a vertex, if this falls in a region occupied by the receptor site atoms (in the nondeformed receptor), will "distort" the site (and also the molecule), reducing molecule−receptor affinity. The absence of an atom in that vertex will have zero effect on affinity.

Table 4. Regression $a_{j\mu}$ Coefficients and VIP[15,16] Values (in Parentheses) for the PLS Model **M16** − Free Term a_0 = 21.97[a]

	μ			
vertex J	S	V	P	H
sC	1.84 (1.22)			
sO	−1.05 (0.55)			
dC	−30.65 (1.69)			
dO	23.72 (1.40)			
1	−12.68 (2.14)	−0.017 (2.05)	−0.067 (2.05)	
2	−10.40 (1.77)	−0.015 (1.75)	−0.058 (1.71)	
3	25.00 (0.51)		0.059 (0.51)	0.519 (0.51)
4	0.05 (0.15)		0.075 (1.25)	0.293 (0.81)
5	0.24 (0.23)		0.029 (0.69)	
6	0.40 (0.32)	−0.031 (1.10)	−0.023 (0.34)	
7	−0.70 (0.48)		0.042 (1.32)	0.149 (0.80)
8	6.56 (0.81)		0.067 (1.06)	0.443 (1.06)
9	0.79 (0.26)		0.012 (0.67)	0.113 (0.43)
10	−4.27 (0.23)	−0.004 (0.23)	−0.018 (0.23)	
11	0.53 (0.26)		0.006 (0.26)	0.139 (0.26)
12	0.58 (0.17)		0.002 (0.46)	
13	−0.60 (0.32)	−0.001 (0.39)	−0.002 (0.39)	

[a] See text for details. Note: $a_{j\mu}$ are the regression coefficients from eq 1; these are cumulative values, corresponding to those for two PLS components. Values in brackets indicate cumulative VIP-values.

If one obtains hereby a realistic description of the receptor atom−ligand atom interactions, they can be confirmed only by comparison with other results concerning the same series, or even better, with X-ray crystallography data for receptor−ligand complexes.

Comparison with Other QSAR Results. Let us recall the results of Järv et al.,[7] generalized for all 25 esters:[3]

$$\hat{Y} = 3.111 + 3.011\cdot\sigma^* + 0.968\cdot\pi + 1.935\cdot E_S;\ r^2 = 0.779 \quad (9)$$

and the results of Ciubotariu et al.[3] (see Figure 1 for vertex attributions):

$$\hat{Y} = 8.434 + 1.683\cdot\sigma^* - 0.753\cdot MTD;\ r^2 = 0.867 \quad (10)$$

where the σ^* are the Taft constants, π are the Hansch hydrophobicities, and the E_S are the Taft inductive constants for the alcoholic (R) moiety of CH_3COOR.

The favorable electron withdrawing effect of the R group is evident. In our results, this can be connected to the a_{jS} coefficients for the atoms in the CH_3COOCH_2 moiety. We found that a_{SsO} is negative, with all SsO values being also negative in the model **M16**, i.e., higher negative partial charges lead to higher activity. The negative a_{SdC} for the double bonded carbons prevails as absolute value (all SdO

> 0), i.e., smaller positive charges lead to higher activity. These interpretations are consistent with the effect of σ^*.

The beneficial (cavity) vertices for the classical MTD-method[3] (see Figure 1) are {$j = 3-8, 10, 11$}. Current results show positive a_{jH} regions (values between 0.1 and 0.5) for vertices {$j = 3, 4, 7-9, 11$}, both for **M14** and **M16** (Table 4). For all these vertices, $a_{jV} = 0$ (i.e., no steric repulsion was observed), while the a_{jP} coefficients were positive. These can be interpreted as "beneficial" interactions in these regions. Vertices {$j = 6, 10, 13$} show negative a_{jV} values (i.e., steric repulsion) and small ({$j = 10, 13$} in **M14**, not shown here) or zero a_{jH} coefficients. We interpret these vertices as being detrimental; we have previously found that vertex 13 (the same as vertex 6 in our previous work[8]) "extends in the receptor wall formed by Trp^{84} and Tyr^{442}" of AChE.[8]

Concerning the a_{jS} coefficients for electrostatic interactions, their beneficial or detrimental effects depend on the partial charges S_{ij} of those atoms that occupy the hypermolecular vertices. We have discussed above the influence of the carboxyl moiety. The electrostatic interaction between two 1*e* charges at the usual van der Waals contact interatomic distance (3.5 Å) corresponds to about 380 kJ/mol in a vacuum or about 65 logarithmic units of biological activity (since 2.3RT ≅ 5.9kJ/mol at 300 K). Taking a maximal value $a_{SdC} \cong -30$ (Table 4), the corresponding double bonded carbon atom of the CH_3COOCH_2 moiety should be in the electric field of an about −0.45e charges from the receptor site (taking an average 3.5 Å interatomic distance). We note that one of the two imidazole nitrogens from a histidine side chain, which in the crystal structure of the complexed acetylcholine ligand[9] (*vide infra*) is located at approximately 3 Å from the double bonded carbon, has a partial charge of −0.191e, as computed by Mullikan population analysis after an AM1 calculation.

Detrimental (wall) vertices from the classical MTD method[3] (Figure 1) are {$j = 1, 2, 14, 15$}. Vertices {$j = 14, 15$} are present only in the previously eliminated outliers **1**, **16**, and **22** (compare **M5** to **M13**). Vertices {$j = 1, 2$} have $a_{jV} \cong -0.016$ and $a_{jP} \cong -0.06$ (see Table 4), and with mean values of $V_{ij} \cong 15$ and $P_{ij} \cong 4$ the detrimental effect for such a vertex is approximately 0.5 logarithmic units of Y activity. This is comparable with the −0.75 coefficient for MTD in eq 10. Our current results indicate that vertex 6 is also detrimental (as discussed above).

Irrelevant classical MTD vertices[3] (Figure 1) are {$j = 9, 12, 13$}. Indeed, vertices {$j = 9, 12$} have small or zero a_{jV}, a_{jH}, and a_{jP} coefficients (see Table 4). However, vertex 13 (although with relatively small VIP value) is found to be detrimental in model **M16**.

Comparison with X-ray Crystallography Data. We have analyzed data for the native AChE from *Torpedo Californica* (electric organ, electroplaque tissue), with acetylcholine manually docked in the active site and covalently attached to Ser_{200} in the tetrahedral intermediate state.[9] Although the analyzed experimental data are determined with AChE from *Naja naja oxiana*, this protein bears 76% primary sequence identity to the protein used in the X-ray structures, in particular for the first 173 amino acid residues discussed in the above-mentioned enzyme.[8] Distances between acetylcholine atoms and nearby atoms of amino acid residues of the enzyme site are available in the PDB file[9]—see also Figure 2 for the numbering scheme of acetylcholine atoms. A total of 18 amino acid residues from several portions of the enzyme's polypeptide chain are implied in these vicinal contacts.

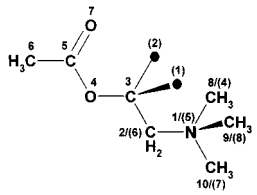

Figure 2. Numbering of acetylcholine atoms in the complex with AChE, in the X-ray crystallography derived PDB file:[9] (*j*) vertex numbering in parentheses.

The general view of the complex[9] is with $-CH_2-$ groups 2 and 3 in a narrow gorge of the enzyme and the $-NMe_3^+-$ group in a hydrophobic pocket defined by the side chains of electron-rich aromatic residues. Acetylcholine occupies vertices {$j = 4-8$}—see Figure 2. The $-CH_2-$ group 2, corresponding to vertex 6, is at a rather short, 2.8 Å, interatomic distance from the CD2 atom of His^{440}, in agreement with the "detrimental" attribution for this vertex in the current study. In Table 4, $a_{6V} = -0.03$, $a_{6P} = -0.02$. The "detrimental" vertices {$j = 1, 2$} are connected in the hypermolecule to the $-CH_2-$ group 3 (not numbered as j-vertex, since it is common to all ligands), which is within the narrow gorge of the enzyme. These vertices should "penetrate" into the walls of the enzyme site (see the negative a_{1V}, a_{2V}, a_{1P}, and a_{2P} coefficients in Table 4). Indeed, inspecting the X-ray structure it appears that Tyr^{130} (previously studied with a hypermolecule[8]) and Glu^{199} are good candidates for this steric interference.

The $-NMe_3^+-$ group occupies vertices {$j = 4, 5, 7, 8$}. The three CH_3- groups correspond to groups 8, 9, and 10 (as numbered in the PDB file[9]—see Figure 2) and make numerous contacts to the aromatic residues Phe^{330}, His^{440}, Trp^{84} of the enzyme site. Concerning the regression coefficients of Table 4, the electrostatic a_{jS}- coefficients are rather small, except $a_{8S} \cong 6$ (VIP = 0.8). The a_{jV} are all zero, the polarizability a_{jP} and hydrophobicity a_{jH} coefficients are rather large and positive (except $a_{5H} = 0$, corresponding to the sterically buried quaternary nitrogen atom). The large and positive a_{jP} and a_{jH} coefficients are, from the "chemical intuition" point of view, expected to favor interactions with aromatic side chains from the enzyme site. As for $a_{8S} \cong 6$, which corresponds to the CH_3- group 9, it matches to the single CH_3- group of acetylcholine interacting with the negatively charged OE1 atom of the acidic side chain of Glu^{199} of the active site in what appears to be a strong attractive interaction.

Vertices {$j = 3, 9-15$} are more difficult to discuss without docking other esters, as none are occupied by acetylcholine. Vertices {$j = 9, 12$} could be placed in the binding pocket in the region occupied by water molecules cocrystallized around CH_3 10 in acetylcholine (vertex {$j = 7$}). This may explain the irrelevant character of these vertices in MTD. However, we found that vertex 13 penetrates the receptor wall formed by Trp^{84}, as discussed above.

A successful structure-based QSAR analysis of AChE inhibitors was recently published by Sippl and co-workers.[17] Although we discuss here substrates, not inhibitors, one

similarity was found: one of the favorable regions of hydrophobic interaction of the active site discussed[17] by Sippl et al. coincides with our beneficial vertices {j = 4, 7, 8, 11}. However, these inhibitor molecules are much larger than the substrates, so detailed comparisons are rather difficult.

CONCLUSIONS

The statistical model stability of the $a_{j\mu}$ regression coefficients, when analyzing the different models, does not change dramatically. In this series, the various $a_{j\mu}$ values obtained are in good agreement with those observed in acetylcholine-AChE contacts, as indicated by X-ray crystallography.[9] The high a_{8S} coefficient derived only for the CH_3-group in contact with the negative Glu_{199}, or the negative a_{6V} and a_{6P}, a_{13V}, and a_{13P} values for vertices {j = 6, 13}, assigned in this study to be detrimental (but not in the previous MTD-studies[3,8]), are some of the atomic-level details that were correctly captured in this study. This "chemically intuitive" approach further forces MTD-PLS coefficients to assume only negative (or zero) values for fragmental volume descriptors and positive (or zero) values for fragmental hydrophobicity descriptors. This further separates the various kinds of local interactions at each vertex of the MTD hypermolecule, making this method suitable for medicinal chemistry synthesis planning. Other studies of this kind, on QSAR-series that can include more data at the structural atomic detail (e.g., by X-ray crystallography), are required to further validate this MTD-PLS approach.

Should MTD-PLS prove effective, it could serve as a basis for receptor site mapping within a modified (2D-based) CoMFA-like method. The $a_{j\mu}$ regression coefficients could thus offer information related to the characteristics of nearby receptor atoms, in close contact to hypermolecular vertices (conceptually similar to the CoMFA lattice), e.g., electric charge, steric rigidity of the site-walls, polar-nonpolar contacts, etc. The method also suggests that in CoMFA somewhat different potential terms should be used, for example the steric field based on intersection volumes of ligand-probe atoms van der Waals envelopes[18,19] or, for partial dehydration effects, the free energy field of Sulea and Purisima.[20] Different cutoffs, based on van der Waals contacts at minimal distances, could then be used for modeling electrostatic and van der Waals interactions.

ACKNOWLEDGMENT

We thank Dr. Erik Johansson from Umetrics (Umeå, Sweden) for the SIMCA package and Prof. Dr. M. Mracec for hardware and software support. This work was partially funded by the Ministry of Education and Research, contract no. 7012/1997/theme 13/GRANT CNCS 1248.

REFERENCES AND NOTES

(1) Oprea, T. I.; Kurunczi, L.; Olah, M.; Simon, Z. MTD-PLS: A PLS-based variant of the MTD method. A 3D-QSAR analysis of receptor affinities for a series of halogenated dibenzodioxin and biphenyl derivatives. *SAR QSAR Environ. Res.* **2001**, *12*, 75−92.

(2) Simon, Z.; Chiriac, A.; Holban, S.; Ciubotariu, D.; Mihalas, G. I. *Minimum Steric Difference. The MTD Method for QSAR Studies*; Research Studies Press: Letchworth, 1984.

(3) Ciubotariu, D.; Deretey, E.; Oprea, T. I.; Sulea, T.; Simon, Z.; Kurunczi, L.; Chiriac, A. Multiconformational minimal steric difference. Structure − acetylcholinesterase hydrolysis rates relations for acetic acid esters. *Quant. Struct.-Act. Relat.* **1993**, *12*, 367−372.

(4) Oprea, T. I.; Zamora, I.; Svensson, P. Qvo Vadis, Scoring Functions? Toward an Integrated Pharmacokinetic and Binding Affinity Prediction Framework. In *Combinatorial Library Design and Evaluation for Drug Design*; Ghose, A. K., Viswanadhan, V. N., Eds.; Marcel Dekker Inc.: New York, 2001; pp 233−266.

(5) Palyulin, V. A.; Radchenko, E. V.; Zefirov, N. S. Molecular field topology analysis method in QSAR studies of organic compounds. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 659−667.

(6) Simon, Z. Stereochemical and informational aspects of QSAR and MTD. *Rev. Roum. Chim.* **1987**, *32*, 1103−1107.

(7) Järv, J.; Kesvatera, T.; Aaviksaar, A. Structure activity relations in acetylcholin-esterase reactions. Hydrolysis of nonionic acetic acid esters. *Eur. J. Biochem.* **1967**, *67*, 315−322.

(8) Sulea, T.; Kurunczi, L.; Oprea, T. I.; Simon, Z. MTD-ADJ: A multiconformational minimal topologic difference for determining bioactive conformers using adjusted biological activities. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 133−146.

(9) Analysis of Ligand − Protein Contact in PDB entry 2ACE: http://pdb-browsers.ebi.ac.uk/pdb-docs/LIGIM/AC/1ACL_lig01/1ACL_lig01.html.

(10) Olah, M. Molecular fragment volume calculation for QSAR studies. *Rev. Roum. Chim.* **2000**, *45*, 1123−1125.

(11) Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR. Hydrophobic, Electronic and Steric Constants*; American Chemical Society: Washington, DC, 1995.

(12) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. The development and use of quantum-mechanical molecular models. 76. AM1: A new general-purpose quantum-mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3910.

(13) Hyperchem 5.11 and ChemPlus 1.6 are available from HyperCube, Inc. http://www.hyper.com.

(14) Volkenstein, M. V. *Stroenie i fizicheskie svoistva molecul*; Izd. Akad. Nauk: Moscow, 1955; p 279.

(15) The SIMCA program package is available from Umetrics, Umeå, Sweden, http://www.umetrics.com.

(16) Höskuldson, A. PLS regression methods. *J. Chemom.* **1998**, *2*, 211−228.

(17) Sippl, W.; Contreras, J. M.; Parrot, I.; Rival, Y. M.; Wermuth, C. G. Structure-based 3D QSAR and design of novel acetylcholinesterase inhibitors. *J. Comput.-Aided Mol. Design* **2000**, *15*, 395−410.

(18) Mureşan, S.; Sulea, T.; Ciubotariu, D.; Kurunczi, L.; Simon, Z. van der Waals intersection envelopes as a possible basis for steric interactions in CoMFA. *Quant. Struct. − Act. Relat.* **1996**, *15*, 31−32.

(19) Sulea, T.; Oprea, T. I.; Mureşan, S.; Chan, S. L. A different method for steric field evaluation in CoMFA improves model robustness. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1162−1168.

(20) Sulea, T.; Purisima, E. O. Desolvation free energy field derived from boundary element continuum dielectric calculations. *Quant. Struct. − Act. Relat.* **1999**, *18*, 154−158.

CI010124A