Journal of Computer-Aided Molecular Design, **12:** 133–146, 1998.
KLUWER/ESCOM

MTD-ADJ: A multiconformational minimal topologic difference for determining bioactive conformers using adjusted biological activities

Traian Sulea[a,*], Ludovic Kurunczi[b], Tudor I. Oprea[c,**] & Zeno Simon[d]
[a]*FORSLYS (The Group for the Formal Study of Living Systems), P.O. Box 1120, RO-1900 Timisoara-14- România;* [b]*Department of Physical Chemistry, Faculty of Pharmacy and* [d]*Department of Biophysics, Faculty of Medicine, University of Medicine and Pharmacy, P-ta E. Murgu 2, RO-1900 Timisoara, România;* [c]*Theoretical Biology and Biophysics Group (T-10), Los Alamos National Laboratory, Mail Stop K710, Los Alamos, NM 87505, U.S.A.*

Received 26 April 1997; Accepted 23 July 1997

Key words: 3D-QSAR methodology, acetylcholinesterase active site, acetic acid esters, adjusted biological activity, bioactive conformation, MTD

Summary

The active conformation is part of a conformational mixture with experimental activity Y_{exp}, and is used in QSAR studies to extract more information regarding the ligand–receptor interaction. To reflect the relative amount (α) of the active conformation, we adjust Y_{exp} : $Y_{adj} = Y_{exp} - \log \alpha$. We establish a quantitative structure-activity relationship (QSAR) between Y_{adj} and 3D conformational characteristics for the acetylcholinesterase (AChE) hydrolysis rates of 25 acetic esters. The 3D-QSAR model was obtained using the adjusted multiconformational minimal steric/topologic difference (MTD-ADJ) method, optimizing the receptor map based on Y_{adj} for each conformer. Y_{adj} was updated during each step of the optimization process. α and Y_{adj} are based on the Boltzmann distribution calculated using AM1 (MOPAC 6.0) relative energies of the COSMIC 90 derived conformers. The MTD-ADJ results are: (i) the 3D-QSAR models obtained by this procedure have significant statistical parameters and are similar to the unadjusted (MTD-MC, using Y_{exp}) models; (ii) the selected bioactive conformations are extended, occupy cavity vertices and, for the same structures, have the same MTD value; and (iii) the optimized conformational map of the neutral ligands obtained from the MTD-ADJ model fits well in the active site of the crystallographic structure of AChE (from *Torpedo californica*). We propose a neutral ligands binding site model for AChE. Our results show that MTD-ADJ, which can be implemented in any 3D-QSAR method, is capable of providing additional information regarding the active conformations, and can be used to gain further insight into the ligand–receptor models for which no structural data are available.

Introduction

Conformers of biologically active compounds are routinely investigated with computational procedures. Conformational energy can be calculated with acceptable precision [1]. For flexible compounds, only the conformations with the best complementarity fit into the receptor binding site. These bioactive conformers are the major contributors to biological activity, Y. The experimental biological activity, Y_{exp}, of any compound is defined as

$$Y_{exp} = -\log C_T \quad (1)$$

where C_T is any standard concentration of the compound (e.g., IC_{50}). However, a large contribution to Y_{exp} is given by the bioactive conformer(s), which represent(s) only a fraction, α, of the total population of conformers. Y_{exp} could be, therefore, corrected by the relative amount, α, of the active conformation(s),

* Present address: Biotechnology Research Institute, National Research Council of Canada, 6100 Royalmount Avenue, Montréal, PQ, Canada H4P 2R2. ** To whom correspondence should be addressed at: Medicinal Chemistry I, Astra-Hässle AB, S-431 83 Mölndal, Sweden.

in order to extract more information for molecular modeling and QSAR studies. This relative amount α can be calculated by Boltzmann statistics when the active conformation(s) and the global minimum for that particular conformational space can be determined. The difficulty, for QSAR and 3D-QSAR models, stems from the fact that the global minimum conformer is usually not known.

Here we propose an iterative procedure to determine the active conformation, based on the minimum steric/topologic difference (MTD) method [2,3]. In this procedure, the bioactive conformers are considered as having the lowest MTD value with respect to the optimized receptor map (see the section 'The MTD method' in the Appendix). At each step during the optimization, Y_{exp} is adjusted to reflect the relative amount α of the active conformation: $Y_{adj} = Y_{exp} - \log \alpha$ (see the section 'Methods' for details). This adjusted multiconformational MTD method, termed MTD-ADJ, was tested on the acetylcholinesterase (AChE) hydrolysis rates for a series of 25 acetic esters [4], by comparison with MTD results (Ref. 3, and this study) based on unadjusted parameters (Y_{exp}). Conformational analysis was performed within the COSMIC 90 modeling package [5–7], and further refined using the AM1 Hamiltonian, implemented in MOPAC 6.0 [8,9]. The optimized receptor map obtained from the MTD-ADJ procedure shows extended conformations that occupy cavity vertices. For the same structures, the rejected conformers intersect with receptor walls and/or occupy exterior regions. The crystallographically determined structure of AChE from *Torpedo californica* [10] served as the basis for an active site model [11], which we have used in our analysis. We show that MTD-ADJ provides additional information regarding the active conformations of modeled compounds.

Methods

Thermodynamical arguments for adjustment of the experimental biological activity

For pharmacological data (standard biological responses: ED_{50}, IC_{50}, etc.) as well as for direct receptor binding measurements, Y_{exp} (the experimental biological activity) is defined according to Equation 1. Conformationally flexible compounds are mixtures of multiple conformational substates L_i, $i = 1, 2, ..., n$, of which only several fit in the receptor binding site. We refer to these as bioactive conformations, L_F, with their corresponding concentrations C_F (see the section 'Discussion' for details). C_F represents only a fraction α_F of the total concentration, C_T:

$$C_F = \alpha_F C_T \tag{2}$$

As a consequence, we can consider an adjusted biological activity, $Y_{F,adj}$, related to the bioactive conformation L_F:

$$Y_{F,adj} = Y_{exp} - \log \alpha_F \tag{3}$$

When radioligand binding studies are available, simple equilibrium considerations lead to the same Equation 3 for adjusted binding affinities, e.g. by means of the Cheng–Prusoff equation [12]:

$$IC_{50} = K_i(1 + S/K_M) \tag{4}$$

where K_i is the dissociation constant of the enzyme–substrate complex, S is the substrate concentration, and K_M is the Michaelis–Menten constant of the enzyme.

We calculate the relative amount, α_F, of the bioactive conformation, according to Boltzmann statistics, using the following equation:

$$\alpha_F = \frac{g_F \, e^{-U_F/RT}}{\sum_{i=1}^{n} g_i \, e^{-U_i/RT}} \tag{5}$$

where U_i are relative conformational energies, calculated by standard conformational analyses, and g_i are the degeneration degrees of the conformational energy levels due to the appearance of certain symmetric conformers (i.e., gauche forms, as example). U_F and g_F represent the corresponding properties for the bioactive conformation. If there is more than a single bioactive conformation, the corresponding sum is performed in the denominator.

Experimental data

For this series of 25 acetic acid esters, hydrolysis rates were determined using AChE from cobra venom (*Naja naja oxiana*) [4]. These rates, used as Y_{exp} (experimental activity) values, are given in Table 1. This series, also studied with a previous multiconformational minimal steric/topologic difference (MTD-MC) version [3], meets several requirements: (i) the ligands present high, yet simple, conformational flexibility, which allows an exhaustive conformational search; (ii) the experimental data are homogeneously distributed and cover a large range; (iii) the common pharmacophore exhibits a good degree of structural variation

Table 1. The acetic acid ester series: Structure, experimental AChE hydrolysis rates, QSAR parameters, total number of selected conformers, MTD vertex attributions, MOPAC-AM1 conformational energies and degeneration degrees and adjusted activities used in the MTD-ADJ method

i	R	Y_{exp}	σ*	CLogP	k	j($x_{ijk} = 1$)	E_{AM1}	g_i	Y_{adj}
1	C_6H_5	6.72	0.60	1.49	2	11,53,54,55,56	–55.159	1	7.01
						1,**2**,**7**,**11**,**17***	–55.139	1	7.03
2	$CH_2CH_2C(CH_3)_3$	6.30	0.00	2.57	3	**1**,**9**,**15**,**16**,**26***	–121.390	2	7.14
						1,2,3,11,12	–122.338	1	6.45
3	$CH_2CH_2SC_2H_5$	5.40	0.22	1.44	4	**1**,**2**,**11**,**21***	–106.161	2	6.39
						7,17,19,20	–106.960	2	5.80
4	$CH_2SC_2H_5$	5.35	0.56	1.05	8	**1**,**9**,**26**	–97.741	2	6.26
						7,**14**,**22**	–97.343	2	6.55
						1,9,51	–97.825	2	6.19
						1,9,52	–97.927	2	6.12
5	$CH_2CH_2CH(CH_3)_2$	5.32	0.00	2.17	6	**1**,**9**,**16**,**26***	–119.321	2	6.39
						1,**9**,**15**,**26***	–119.582	2	6.20
						1,**2**,**3**,**11**[a]	–120.048	1	5.86
						1,2,3,12[a]	–120.038	1	5.87
6	$CH_2CH_2NO_2$	5.20	0.62	–0.12	4	**1**,**9**,**32**,**49***	–100.153	2	6.26
						7,14,30,31	–101.073	2	5.58
7	CH_2CH_2Cl	5.02	0.39	0.77	5	**1**,**2***	–109.383	1	5.89
						1,**9***[b]	–109.132	1	6.07
						7,17	–109.933	2	5.48
						1,**10**[b]	–109.119	1	6.08
8	$CH_2C{\equiv}CH$	4.81	0.55	0.28	1	**1**,**13***	–36.766	1	4.81
9	*n*-C_5H_{11}	4.74	0.00	2.30	3	1,2,3,4	–122.631	1	5.01
						1,**9**,**26**,**27***	–122.141	2	5.37
10	*n*-C_7H_{15}	4.75	0.00	3.36	5	**1**,**2**,**3**,**4**,**5**,**6**	–136.321	1	5.25
						1,**9**,**26**,**27**,**28**,**29***	–135.893	2	5.56
						7,14,22,23,24,25	–135.996	2	5.48
11	$(CH_2)_4SC_2H_5$	4.73	0.03	1.92	19	**1**,**2**,**3**,**4**,**5**,**6**	–119.814	1	5.93
						1,**9**,**26**,**27**,**28**,**29***	–119.412	2	6.22
						1,**2**,**3**,**4**,**33**,**34**	–120.124	2	5.70
						1,2,3,36,37,38	–119.156	2	6.41
						1,**9**,**26**,**27**,**41**,**42***	–119.662	2	6.04
						1,**9**,**26**,**27**,**39**,**40***	–119.741	2	5.98
						1,**2**,**3**,**4**,**5**,**35**	–119.904	2	5.86
						1,**9**,**26**,**27**,**28**,**44***	–119.509	2	6.15
						1,**9**,**26**,**27**,**28**,**43***	–119.552	2	6.12
						7,14,22,23,24,25	–119.670	2	6.03
12	cyclo-C_6H_{11}	4.71	0.00	2.21	2	**1**,**2**,**7**,**11**,**17***	–120.484	2	5.01
13	*n*-C_4H_9	4.72	0.00	1.77	7	**1**,**2**,**3**	–115.790	1	5.47
						1,**9**,**26***	–115.305	2	5.82
						7,14,22	–115.435	2	5.73
						7,17,18	–115.946	2	5.35
14	*n*-C_6H_{13}	4.68	0.00	2.83	3	**1**,**2**,**3**,**4**,**5**	–129.480	1	4.95
						1,**9**,**26**,**27**,**28***	–128.988	2	5.31
15	$(CH_2)_3SC_2H_5$	4.67	0.08	1.39	17	1,2,3,4,5	–112.798	1	5.90
						1,9,26,27,28 *	–112.524	2	6.10
						1,**2**,**11**,**21**,**50***	–112.157	2	6.37
						1,9,26,45,47 *	–112.978	2	5.77
						1,2,3,36,37	–113.111	2	5.67
						1,9,26,46,48 *	–112.801	2	5.90

Table 1. (Continued)

i	R	Y_{exp}	σ^*	CLogP	k	j($x_{ijk} = 1$)	E_{AM1}	g_i	Y_{adj}
						1,2,3,4,33	–112.946	2	5.79
						1,9,26,27,39 *	–112.684	2	5.98
						1,9,26,27,41 *	–112.683	2	5.98
16	$CH_2C_6H_5$	4.66	0.25	1.96	6	**1**,**9**,**26**,**57**,**58**,**59***	–67.288	2	5.81
						1,2,3,59,63,64 *	–67.63	2	5.56
						7,14,22,60,61,62	–68.16	2	5.18
17	$CH_2CH(CH_3)_2$	4.32	0.00	1.64	4	**1**,**2**,**9***[c]	–113.425	1	4.98
						1,**2**,**10**[c]	–113.331	1	5.05
						7,14,17	–113.617	2	4.84
18	$CH_2CH{=}CH_2$	4.10	0.23	0.76	1	**1**,**2***	–76.724	1	4.10
19	*n*-C_3H_7	3.91	0.00	1.24	5	**1**,**2***	–108.990	1	4.56
						1,**9***	–108.440	2	4.96
						7,17	–109.157	2	4.44
20	$CH(CH_3)C_2H_5$	3.69	0.00	1.55	2	**1**,**7**,**10**[d]	–112.364	1	3.99
						1,**8**,**9***[d]	–112.364	1	3.99
21	C_2H_5	3.36	0.00	0.71	1	**1***	–102.231	1	3.36
22	CH_3	3.00	0.20	0.18	1	–	–96.433	1	3.00
23	$CH(CH_3)_2$	2.72	0.00	1.02	2	**1**,**7***[e]	–106.224	1	3.02
						1,**8***[e]	–106.219	1	3.02
24	$C(CH_3)_3$	1.30	0.00	1.42	1	**1**,**7**,**8***	–107.027	1	1.30
25	$C(CH_3)_2C_2H_5$	1.30	0.00	1.95	2	**1**,**7**,**8**,**14***	–111.957	2	1.60

Compounds are arranged in decreasing order, i, of the biological activity Y_{exp}. R is the variable substituent in the acetic acid esters CH_3COOR. $Y_{exp} = \log(k_{cat}/K_M)$, in $s^{-1}\ M^{-1}$ [4], is the experimental activity for the conformational mixture. σ^* are Taft substituent electronic constants [4, 15–18]. CLogP are calculated logarithmic *n*-octanol/water partition coefficients, using MacLogP v. 1.0.3 [19]. k is the total number of ligand conformations selected for the MTD-ADJ QSAR. j ($x_{ijk} = 1$) represents the vertex attribution in the hypermolecule shown in Figure 1. Marked with * are the conformers that lead to the best correlation in the MTD-MC model; in bold are marked the conformations selected using MTD-ADJ. E_{AM1} are the AM1 conformational (internal) energies in kcal/mol, calculated with MOPAC 6.0 [8,9]. The degeneration degree g_i corresponds to the conformational energetic level i. Y_{adj} is the adjusted Y_{exp}, used in MTD-ADJ.

[a] These conformers are mirror images and correspond to a double degenerate energetic level. Due to the fact that both vertices 11 and 12 are occupied by other conformers of the series, both these conformers were considered for the MTD-ADJ. Similar situations occur for conformers marked with b,c,d, with respect to the vertices 9 and 10, and for conformers marked with e, with respect to the vertices 7 and 8.

at the substituents; and (iv) the catalytic hydrolysis mechanism is common to all compounds, well understood, and occurs at the same binding site.

Conformational analysis

We used a combined [13] molecular mechanics and semiempirical quantum mechanical approach for the conformational analysis of the 25 ligand molecules of Table 1. First, the conformational ligand space was investigated by the PCMIN01 program of the COSMIC molecular modeling package [5–7]. Computational conditions were chosen as follows: rotation increment of 10° for each rotatable bond, 30 randomizations, 10 cyclic randomizations, dielectric constant of 4.0, and computational accuracy of 0.1 cal/mol. Conformers within 1 kcal/mol above the global minimum were stored. Finally, a filter procedure was applied for the selected conformations, in order to retain only geometrically different conformers [7]. The 1 kcal/mol limit provides a reasonable approximation, covering about 85% of the conformational mixture within the Boltzmann distribution Equation (5). The COSMIC results provided, on the average, 10 different conformations stored within 1 kcal/mol under the conditions described above. These conformers were used as input structures for a MOPAC optimization [8,9] (keywords: AM1, PRECISE, NOINTER). COSMIC and AM1-optimized conformers, stored within 1 kcal/mol, are listed in Table 1. Their relative AM1 internal energies

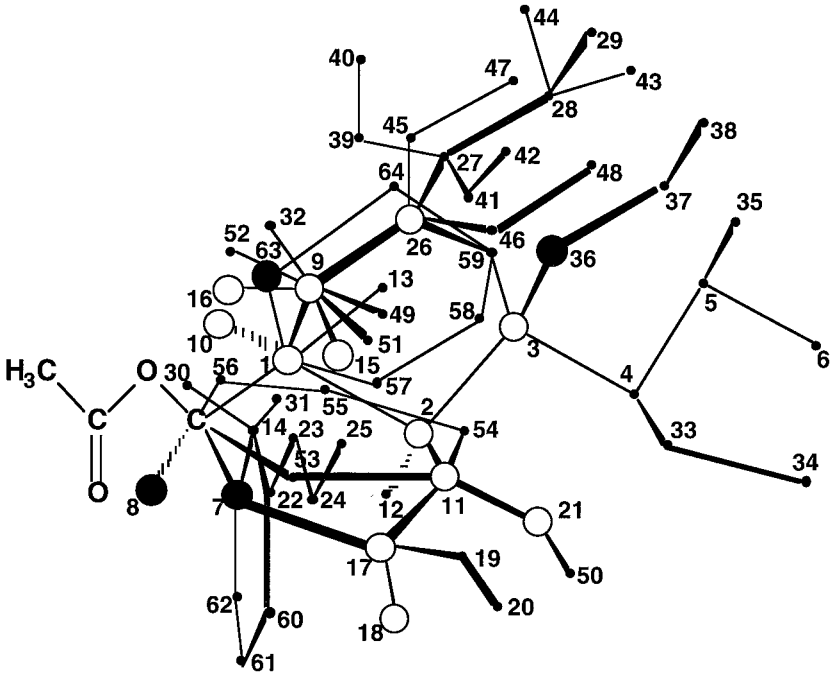

Figure 1. The MTD hypermolecule, showing the assignment of non-common vertices and the optimized receptor map for the biparametric MTD-σ^* model using the connex MTD-ADJ method (equation 7). The vertex attributions are: open circles – cavity (beneficial) vertices; full circles – wall (detrimental) vertices; dots – exterior (irrelevant) vertices.

were used to calculate the Boltzmann population of conformational levels and to adjust Y_{exp}.

Some degenerated conformers can be constructed for several symmetric gauche torsional angles (degeneration degrees are listed in Table 1). The energy minimizations performed with the COSMIC force field do not routinely indicate such degenerated states. However, these are important when calculating α using Equation 5 and when constructing the MTD hypermolecule, because the degenerate gauche conformers occupy distinct vertices. Statistically, the MTD parameter can be considered as an aggregate of ϵ_j attributions at each j vertex of the H hypermolecule, when occupied by the ith molecule in the kth conformation (see the section 'The MTD method' in the Appendix). For the construction of the hypermolecule, we considered only one of the two degenerate gauche conformers, thus reducing the number of adjustable parameters. This simplification is necessary because the AChE binding site is not symmetric, and there is no possibility for a formal split of the binding site in two parts as mirror images. Only one of the two symmetric gauche conformers interacts and binds, depending on its relative spatial orientation and the disposition of the corresponding key atoms in the receptor binding site. The conformational degeneracy was considered for a correct calculation of Y_{adj} and α, prior to applying the MTD-ADJ method.

Construction of the MTD hypermolecule

The AM1 conformations listed in Table 1 were analyzed with the COSMIC molecular modeling package [5–7]. Systematic measurements of relevant torsion angles using COSMIC allowed the rms fit of different structures, resulting in the superimposition of selected conformers for all compounds of the series.

The superimposition was made using the common CH_3COOC frame. Similar atoms of different structures occupy the same vertex if the interatomic distance between them is less than 0.7 Å, and different vertices if this distance is greater than 0.7 Å. This interatomic distance limit can be justified on the basis of van der Waals interactions [14]. The 0.5 Å limit previously used in Ref. 3 was increased to 0.7 Å, because of the exact superimposition of the common CH_3COOC frame. A relaxation of this frame superimposition allows an improved fit of the other atoms. The result of this superimposition is the hypermolecule with $M = 64$ vertices depicted in Figure 1.

3D-QSAR models and structural parameters: The MTD-ADJ procedure

The adjusted multiconformational MTD method (MTD-ADJ) is a modified version of the MTD method, which accounts for the relative amount of conformers (using Y_{adj}). This method, based on the multiconformational MTD (MTD-MC) version [3], is detailed in the Appendix (sections 'The MTD-MC method' and 'Implementation of the adjusted version in the MTD-MC method'). In deriving 3D-QSAR models with the MTD-ADJ method on the same series of acetic esters as MTD-MC [3], we addressed the following aspects:

(1) During conformational evaluation, multiple conformers were considered to find if a unique (convergent) solution is obtained, i.e., if the binding site is unique.

(2) The adjusted model (MTD-ADJ, 'refined') and the unadjusted model (MTD-MC, 'crude') were compared using the exact set of conformers.

(3) Additional QSAR descriptors were included in the MTD-like optimization procedure to develop further 3D-QSAR. These were compared with the monoparametric MTD models (i.e., MTD as the only parameter). Biparametric correlations including the σ^*-Taft substituent electronic constants [4, 15–18] and the calculated n-octanol/water partition coefficients (CLogP [19] values) were performed (see Tables 1 and 2).

(4) The QSAR results were cross-validated using a leave-half-out (LHO) cross-validation-like procedure [20], as well as the leave-one-out (LOO) procedure (see the section 'Linear regression and cross-validation procedures' in the Appendix for details).

In all the above models, we applied the connectivity condition as an additional constraint of the receptor map with respect to the cavity (beneficial) vertices: all cavity vertices have to form, topologically, a single *connected* network, i.e., the receptor cavity is not fragmented in several subcavities [2].

The adjustment step is a distinctive feature of MTD-ADJ, but its basic principle remains the minimum topologic difference between the binding site and ligand geometry. Central to this approach is the pharmacophore concept, that all structures that bind should exhibit similar 3D features, including geometry. Based on this principle, the first step in each optimization cycle is conformational selection (step (iii) in the section 'Implementation of the adjusted version in the MTD-MC method' in the Appendix). In MTD-ADJ, one or more kth conformers of compound ith are characterized by an MTD_{ik} value, with respect to the current vertex attribution (instantaneous receptor map). Experimental activity is then adjusted for each conformer, during each MTD-ADJ cycle. We call 'bioactive conformers' those that best fit their Y_{adj} at the end of the iterative procedure. Cross-validation is performed only for the final MTD-ADJ model, by predicting Y_{adj} using MTD values, with the LOO and LHO procedures [3]. When more than one bioactive conformer is selected in MTD-ADJ for the same compound, all these conformers have the same MTD value, while Y_{adj} is calculated considering conformer contributions to the total population (i.e., $\Sigma\alpha$).

Results

The MTD-ADJ results using selected conformers, as described above and listed in Table 1, are given in this section only for the biparametric 3D-QSAR based on MTD and σ^*-Taft. All other QSARs are available as supplementary material. For all QSAR models, the start receptor map, S°, used to initiate the MTD-like optimization process, was the same: S° was selected by visual inspection of the vertex occupancy distribution for all ligand conformations, after ranking the series in decreasing order of Y_{exp}. Cross-validation was applied, in both the LHO and LOO variants, starting the optimization process on each subseries with the same initial S° receptor map, previously used to derive the original MTD model, as suggested for cross-validation techniques [21]. This start receptor map was

$$S^\circ \begin{cases} j(\epsilon_j = -1): 1, 2, 7, 9, 11, 15\text{–}17, 21, 26 \\ j(\epsilon_j = 0): 3\text{–}6, 10, 12\text{–}14, 18\text{–}20, 22\text{–}25, \\ \qquad\qquad 27\text{–}64 \\ j(\epsilon_j = +1): \text{none} \end{cases}$$

Below are given the results for the optimized receptor map (S^*), the correlation equation and its associated statistical parameters (n, r^2, s, F), along with cross-validation results (q^2 and its corresponding PRESS, in both LHO and LOO). Statistical parameters and techniques are briefly described in the section 'Linear regression and cross-validation procedures' in the Appendix. The hypermolecule based on the optimized receptor map obtained in the biparametric (MTD-σ^*) QSAR model by a connex MTD-ADJ is represented in Figure 1.

MTD and σ as parameters in the connex MTD optimization process*

Unadjusted model (MTD-MC, crude)

$$S^* \begin{cases} j(\epsilon_j = -1): 1, 2, 7, 9, 11, 15\text{–}17, 21, 26 \\ j(\epsilon_j = 0): 3\text{–}6, 10, 12\text{–}14, 18\text{–}25, 27\text{–}56, 58\text{–}64 \\ j(\epsilon_j = +1): 7, 8, 57 \end{cases}$$

$$Y = -0.757(\pm 0.042)\,\mathrm{MTD} + 2.452(\pm 0.283)\sigma^* + 8.466(\pm 0.259) \quad (6)$$

$n = 25;\ r^2 = 0.948;\ s = 0.306;\ F = 200.78$

$q^2_{(LHO)} = 0.806;\ \mathrm{PRESS}_{(LHO)} = 7.672$

$q^2_{(LOO)} = 0.834;\ \mathrm{PRESS}_{(LOO)} = 6.571$

Adjusted model (MTD-ADJ, refined)

$$S^* \begin{cases} j(\epsilon_j = -1): 1\text{–}3, 9\text{–}11, 15\text{–}18, 21, 26 \\ j(\epsilon_j = 0): 4\text{–}6, 12\text{–}14, 19, 20, 22\text{–}25, 27\text{–}35, 37\text{–}62, 64 \\ j(\epsilon_j = +1): 7, 8, 36, 63 \end{cases}$$

$$Y = -0.878(\pm 0.040)\mathrm{MTD} + 2.956(\pm 0.280)\sigma^* + 12.913(\pm 0.394) \quad (7)$$

$n = 25;\ r^2 = 0.966;\ s = 0.302;\ F = 308.28$

$q^2_{(LHO)} = 0.810;\ \mathrm{PRESS}_{(LHO)} = 7.532$

$q^2_{(LOO)} = 0.777;\ \mathrm{PRESS}_{(LOO)} = 8.844$

Actual versus predicted LHO cross-validation plots are given, for both methods, in Figure 2. For the adjusted model, Y_{adj} was re-transformed into Y_{exp} and used as the 'predicted' activity. The final selection of bioactive conformations is presented in Figure 3, together with their incidence on the beneficial vertices of the optimized receptor map of Figure 1. We propose this final selection as the active conformation preferred by the considered compounds. Table 2 summarizes the statistical results for all QSAR models.

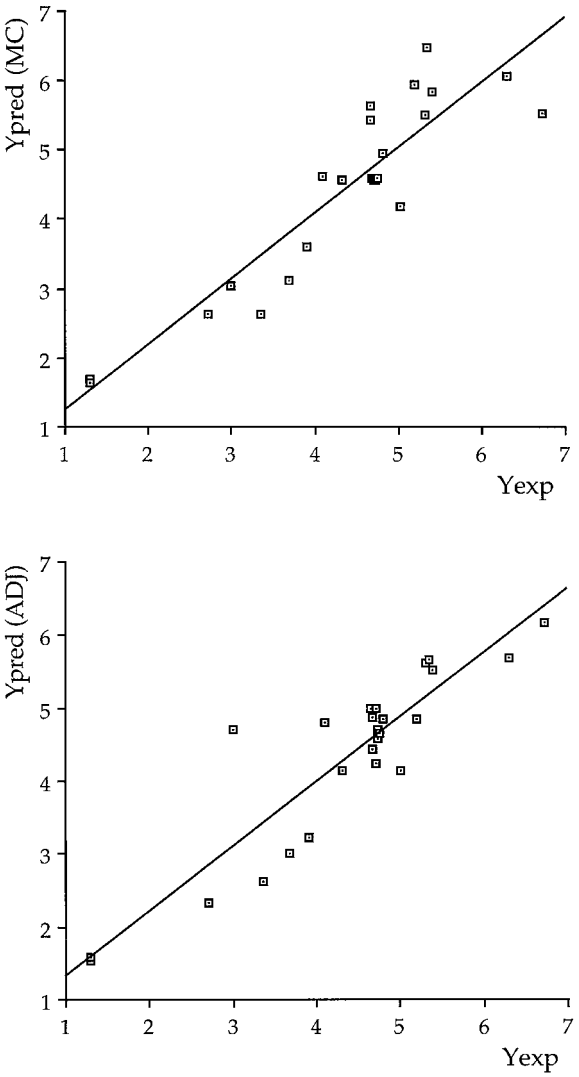

Figure 2. Actual versus LHO-predicted activities for the MTD-MC (top) and MTD-ADJ (bottom) connex biparametric MTD-σ* models; the corresponding correlations (Equations 6 and 7) and statistical parameters are given in the text.

Discussion

Statistical and topological aspects: Comparison between the crude and refined models

For a comparison of the QSAR results obtained using the different variants of the MTD method for the 25 acetic acid esters, the reader is referred to Table 2. The MTD monoparametric model, MTD, explains 90.3% of the variance, but is not as robust in the LHO cross-validation procedure, when compared with the other models. The LHO results for the biparametric MTD-CLogP, while improved, are worse than the MTD-σ* model. This latter model gives the highest correlations and predictive abilities. A triparametric correlation on the same series [3], involving σ*-Taft, π-Hansch hydrophobicity and E_s-steric constant, yielded an r^2 of 0.779, with $q^2_{(LHO)} = 0.658$ and $q^2_{(LOO)} = 0.577$. The positive slope of σ*-Taft (electron-withdrawing ester R group) in the MTD-σ* model is in agreement with the nucleophilic substitution character of this enzymatic hydrolysis. This slope (2.45 for MTD-MC and 2.96 for MTD-ADJ) is in qualitative agreement with five QSARs derived for alkaline hydrolysis which contain a σ-Hammett parameter with a slope of 2.25 ± 0.25

Table 2. Summary of QSAR results

QSAR model	MTD		MTD, σ*		MTD, CLogP	
	ADJ	MC	ADJ	MC	ADJ	MC
r^2	0.903	0.836	0.966	0.948	0.910	0.874
$q^2_{(LHO)}$	0.412	0.414	0.810	0.806	0.566	0.642
$q^2_{(LOO)}$	0.745	0.729	0.777	0.834	0.742	0.798

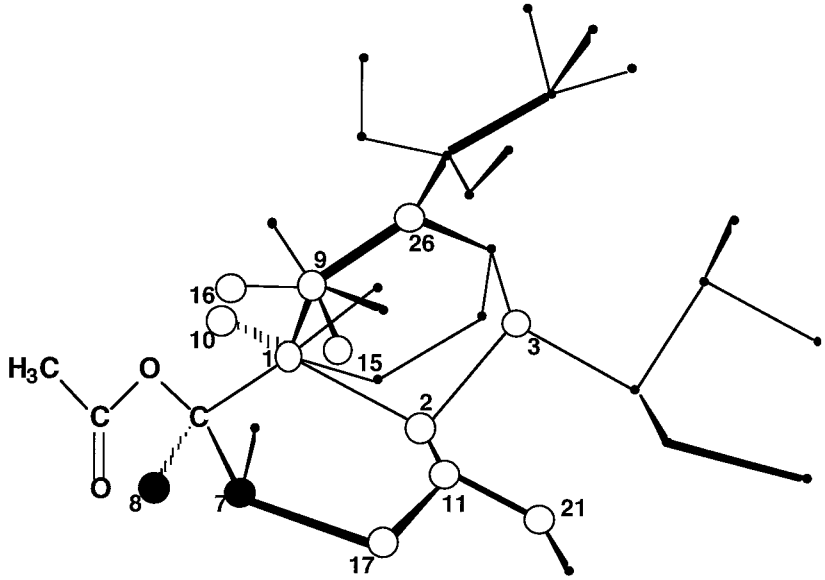

Figure 3. Selected (bioactive) conformations at the end of the connex MTD-ADJ optimization process involving a biparametric QSAR model with MTD and σ*. The incidence of these conformations upon the cavity vertices of the optimized receptor map of Figure 1 is also shown.

[22]. The statistical results, consistent with the lateral validation [22] trend, indicate that the biparametric (MTD-σ*) QSAR is more reliable in describing the binding of these esters to the active site of AChE. This model is used for further discussions below.

No information regarding electrostatic interactions could be obtained from our model. The 25 esters are neutral molecules, albeit the natural substrate, acetylcholine (ACh), is cationic. Hydrophobicity parameters (CLogP, π-Hansch) did not produce a significant improvement of this model. In the event that transmembrane transport is involved in this process, it does not constitute a rate-limiting step.

After superimposing all the compounds on the final hypermolecule, some interesting observations emerged. The conformational map of Figure 3 contains 11 cavity vertices (1, 2, 3, 9, 10, 11, 15, 16, 17, 21 and 26) and two wall vertices (7 and 8). All these vertices are mostly occupied by C atoms. For cavity vertices, the exceptions are: sulfur occupies vertices 2 and 17 in two conformations of compound **3**, vertex 1 in three conformations of compound **4**, and vertices 3, 11 and 26 in nine conformations of compound **15**; nitrogen occupies vertex 9 in one conformation of compound **6**; and chlorine occupies vertices 2, 9, 10 and 17 in four conformations of compound **7**. For wall vertices, only one exception is present: sulfur occupies vertex 7 in one conformation of compound **4**. For comparative purposes, conformers selected by both the MTD-MC and MTD-ADJ models are given in Table 1, along with Y_{adj} values for all the conformers used in QSAR. Because MTD-ADJ is an iterative procedure, only final conformer selections were presented in Table 1. Similar conformational selections were obtained for structurally different ligands.

Biological activities of heteroatom-containing structures differ, and in most cases are improved, in comparison with their isosteric (carbon-atom) counterparts. For example, Y_{exp} for compound **3** ($R = CH_2CH_2SC_2H_5$) is 5.40, while Y_{exp} for

the isosteric compound **9** ($R = n\text{-}C_5H_{11}$) is just 4.74; similarly, replacing a sulfur (compound **4**, $R = CH_2SC_2H_5$) by a carbon (compound **13**, $R = n\text{-}C_4H_9$) leads to a decrease in Y_{exp} from 5.35 to 4.74. However, this difference is absent when comparing the isosteric structures **15** ($R = (CH_2)_3SC_2H_5$) and **14** ($R = n\text{-}C_6H_{13}$): Y_{exp} values are 4.68 and 4.67, respectively. Replacing the Cl atom of compound **7** ($R = CH_2CH_2Cl$) by a carbon in the isosteric structure **19** ($R = n\text{-}C_3H_7$) results in a significant drop in activity, from 5.29 to 3.91. The topological evaluation of the nitro-derivative **6** ($R = CH_2CH_2NO_2$) is more difficult: it can be compared either with the similar compound **2** ($R = CH_2CH_2C(CH_3)_3$, isosteric with ACh) or with the chloro-derivative **7**. In the first case, Y_{exp} increases from 5.20 to 6.30, while in the second case it drops slightly, from 5.20 to 5.02.

Heteroatom occupancy of vertices can be summarized as follows. A sulfur atom (characterized by increased electronegativity and reduced hydrophobicity) generally leads to an improved biological activity when placed in the positions corresponding to β-C and quaternary ammonium atoms of ACh; when this sulfur is located further from the ester group, no changes in the biological activity occur. Chlorine and nitrogen atoms located in the β-C position of ACh contribute significantly to the increase of Y_{exp}, due to their more pronounced electronegativity, a dramatic decrease in lipophilicity, and an increase in the attracting inductive effect. The oxygen atoms of compound **6** are not oriented towards the cavity region, and therefore do not contribute to binding. The uncharged ACh isostere **2** has a high activity, and its conformers occupy almost all cavity vertices of the map shown in Figure 3. Because these ligands are neutral, one could expect a drop in activity, when compared to cationic molecules. However, the anion–cation interaction is replaced, in several cases, by the presence of heteroatoms in the R group. Their electronegativity increases the electrophilic character of the carbonyl carbon of the ester group and facilitates the attack of the Ser^{200} oxygen atom (as found in the AChE crystal structure [21]). In the next section we use the results of this qualitative analysis to discuss our hypothetical model for the binding mode of uncharged ligands in the active site of AChE.

Similar results were obtained using both the adjusted (MTD-ADJ) and unadjusted (MTD-MC) models. The r^2 values are somewhat better for the adjusted model, but it is difficult to interpret cross-validation results. However, that the 'refined' and the 'crude' models are comparable supports the predictive value of the unadjusted models.

The MTD-ADJ optimization process is performed on Y_{adj} data over all conformations, including the degenerated ones, aiming at the best correlation of the type

$$Y_{adj} = aMTD + b, \quad a < 0 \tag{8}$$

which includes the energetic penalty for strained conformers. This can be rearranged using original Y_{exp} values as follows:

$$Y_{exp} = aMTD + \log\alpha + b, \quad a < 0 \tag{9}$$

By comparison with the unadjusted model equation,

$$Y_{exp} = a'MTD' + b', \quad a' < 0 \tag{10}$$

it appears that MTD-ADJ is introducing just an energetic penalty (we thank Jeff Blaney and Eric Martin (Chiron Corporation) for useful comments regarding this aspect), i.e., $\log\alpha$, to optimize the correlation. The MTD values and the resulting optimized receptor map, regression coefficients, and statistical tests (including cross-validation) are not obtained using a variable selection procedure, because only MTD is optimized, and no parameters are eliminated. The purpose of MTD-ADJ, which maintains Y_{exp} and $\log\alpha$ as *invariant* during one iteration, is to select those conformers that, from all conformers considered, lead to the best correlation. Y_{exp} is adjusted separately for each conformer during the optimization; therefore, in MTD-ADJ, Y_{adj} is not just a linear combination of Y_{exp} and energetic penalty ($\log\alpha$). Besides the best r^2 value, MTD-ADJ seeks an internally consistent set of conformers that are likely to correspond to the active conformation. The quality of QSARs obtained with or without biological activity adjustment is not significantly different, but, in addition, MTD-ADJ provides the hypothetical map of the bioactive conformers. This result can be utilized for lead compound optimization and for the general understanding of ligand–receptor interactions, especially for receptors of unknown structure.

Conformational aspects: The AChE binding site for neutral ligands

The active conformers (see Figure 3) are preferentially extended, of antiperiplanar type, and occupy most cavity vertices in the binding site. For molecules with several possible active conformers, different conformations are characterized by the same MTD value. Most of the cavity vertices in Figure 3 match two

conformations of ACh (or its isostere, compound **2**): one of antiperiplanar type and the other with a *gauche* dihedral at the β-C atom. There is a general agreement between the MTD-ADJ conformational map and the X-ray crystallography analysis of substrate binding at the active site of AChE [10]. The extended type of the active conformations is also in agreement with the molecular dynamics study of AChE [23], which suggested that substrate molecules protrude to the active site via transient short channels (i.e., there is not much exceeding space available). Our MTD-ADJ results are also in agreement with the previous [3] MTD-MC results.

The AChE active site, at which the hydrolysis of the ester linkage of the neurotransmitter acetylcholine takes place, comprises four major domains [11, 24]: (i) an esteratic site, comprising Ser^{200}, His^{440}, and Glu^{327} as the catalytic triad of serine proteases; in addition, Glu^{199} and Ala^{201} are involved in the ester linkage breaking; (ii) an anionic site, which binds the quaternary ammonium moiety; anionic residues like Glu^{199}, and aromatic residues, e.g., Trp^{84}, are proposed to play an important role in binding the cationic function; (iii) a hydrophobic region to accommodate the uncharged and lipophilic regions of the ligands; the aromatic gorge of AChE consists of 14 aromatic residues lining its inner surface; and (iv) a peripheral anionic site for binding cationic ligands such as decametonium. Residue numbers are those of the *Torpedo californica* AChE structure [11]. This protein was shown to bear 76% primary sequence identity for the first 173 residues, also sequenced for the AChE from *Naja naja oxiana* [25], for which Y_{exp} values were determined [4] and used in this study. However, it was suggested that the latter does not have a peripheral anionic binding site [26].

Binding of neutral substrates to AChE appears to be somewhat different: a model for active site dynamics during the acylation stage of AChE-catalyzed hydrolyses of neutral esters suggests an *induced fit* as a rate-controlling step, consistent with the demonstrated conformational flexibility of the active site [24]. We combined the information provided by MTD-ADJ (particularly the map of selected bioactive conformations shown in Figure 3) with the qualitative analysis of vertex occupancy by heteroatoms and the available data about the AChE active site, derived from kinetic experimental data [24], X-ray crystallography data [10], and theoretical molecular modeling studies [11]. This resulted in a hypothetical model for the neutral ligands binding site of AChE; see Figure 4. In the following discussion, the spatial attribution of residues lining the active site at the bottom of the aromatic gorge (e.g. binding site positions, sites of favorable interactions) as well as the relative strengths of the interactions involved are based on the minimized structures obtained by Pullman [11] by simulation of the ACh–AChE complex, adapted to our MTD-ADJ optimized receptor map for neutral ligands.

With respect to the esteratic site, several details are provided by our model. This region consists of the catalytic triad Ser^{200}, His^{440}, and Glu^{199}, the latter showing a minor participation in the ester group surrounding, probably being involved in the anionic site as well. The position of the His^{440} side chain allows, besides the construction of a proper single-proton bridge with Ser^{200}, some attractive interactions with ligand atoms that occupy cavity vertices 9, 15, and 16. Ala^{201} is involved in the ester linkage breaking by a distorted hydrogen bond between its amidic hydrogen and the carbonyl oxygen of the ester. Glu^{327}, initially thought to contribute in this region, appears to be involved in the anionic site. The ester group is strained during the interaction of the ligand with AChE, as suggested in Figure 4. Obviously, this high-energy conformer was not analyzed in our conformational search, but this does not affect our results due to the fact that our study was aimed at the variable R group.

Larger differences were observed in the anionic site, which binds cationic ligands (e.g., ACh, decametonium). Even though the same residues, i.e., Glu^{199}, Glu^{327} and Trp^{84}, are involved, the binding contribution is smaller, since the ligands are neutral. We consider that Glu^{199}, which shows the dominant attractive contribution for ACh binding due to its proper orientation and negative charge, has a reduced role in the binding of neutral molecules to AChE. That electrostatic (attractive) interactions are lost and replaced by dispersive (van der Waals London) attractions is consistent with the vertex occupancy by heteroatoms, as discussed above. It is possible that the Glu^{199} side chain transitions to a different rotamer, to allow contact between its aliphatic part of the side chain and the corresponding atoms in the ligand. In our optimized receptor map, Glu^{199} exhibits beneficial interactions with cavity vertices 1, 9 and 15, and with C^{α} (not numbered in the hypermolecule). Glu^{327}, located further from the aromatic gorge, plays a structural role and is probably not involved in direct interactions with the ligand.

The hydrophobic site contains two tyrosines (Tyr^{130} and Tyr^{442}), two tryptophans (Trp^{84} and

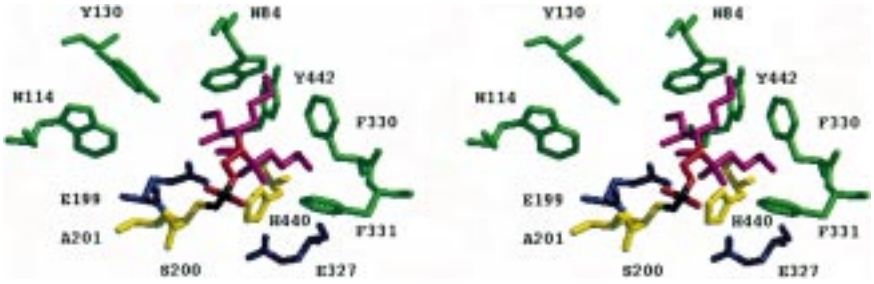

Figure 4. Stereo drawing of the AChE active site model showing the active conformers of the neutral ligands. The numbering of the residues utilized is that of the *Torpedo californica* enzyme (1ace), with the one-letter code for the amino acids. The following color code has been used: magenta shows the cavity vertices, red the initial substrate, green the hydrophobic site, yellow the esteratic site, and blue the anionic site. The Ser^{200}-ester bond is shown in black.

Trp^{114}), and two phenylalanines (Phe^{330} and Phe^{331}). In comparison with the binding site model for cationic ligands [11], we propose some modifications for binding neutral ligands: hydrophobic residues (Trp^{84}, Trp^{114}, Phe^{330}, and Phe^{331}) are closer to the active region, whereas polar side chains (Tyr^{130} and Tyr^{442}) move further from the ligands.

As can be seen in Figure 4, one of the irrelevant vertices (6) extends in the receptor wall formed by Trp^{84} and Tyr^{442}. The major contribution of this aromatic and lipophilic region is due to its spatial position and hydrophobicity rather than cation–π interactions (which occur with cationic ligands [11]). The modification of the Trp^{84} position, along with the relocation of Tyr^{442} further from the active site, is likely to eliminate this steric clash (Figure 4). Thermal fluctuations could relocate Trp^{84} in a more central position, favoring hydrophobic interactions with the aliphatic chains that occupy cavity vertices 2, 11, 17, and 21. The absence of interactions between Tyr^{442} and the ligands is in agreement with the Pullman model [11].

Another hydrophobic contact is observed between cavity vertices and two phenylalanines, 330 (cavity vertices 3 and 26) and 331 (cavity vertices 1, 2, 9, 10 and 16, and with C^{α}); see Figure 4. This suggests a more pronounced contribution of the two phenylalanines, in contrast with their calculated contributions for ACh binding [11]. Conformational transitions in the active site, during uncharged ligand binding, could give a more important role to Phe^{330}. Along with this side chain, Trp^{114} and Tyr^{130} (cavity vertices 17 and 21) may flank the all-trans ligand chains, which end in the exterior regions. Trp^{114} is not involved in the ACh binding [11], but wall vertices 7 and 8 are buried in its side chain. This explains the repulsive interactions proposed by MTD-ADJ during the optimization, and increases the importance of this residue in the neutral

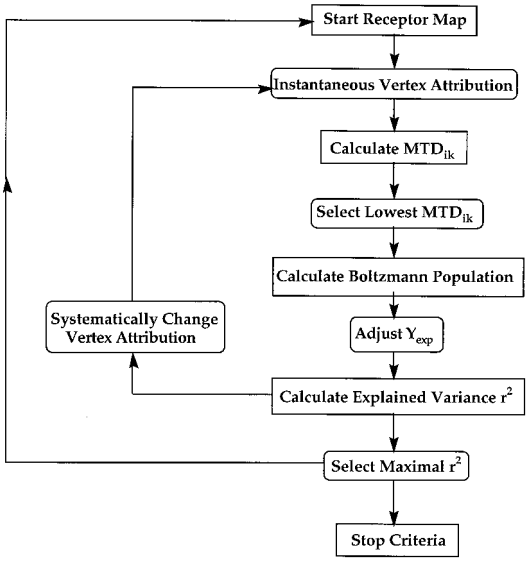

Figure 5. Overview of the MTD-ADJ program flow chart.

binding site model (also in favorable interaction with cavity vertex 17 and with C^{α}) .

The presence of heteroatoms affects the hydrolysis rate by modifying the charge density at the ester carbonyl carbon. In addition, we postulate a contribution of charge-transfer interactions involving π-electrons of the aromatic residues, based on the fact that the presence of third period atoms increases the biological activity when located upon vertices 1, 2, 9, 10, or 17.

Conclusions

The binding of a ligand to a receptor is a multistep process. Any accessible conformation can, in principle, transition into the active one during this process. Entropy is lost during temporary/irreversible binding

and is gained in the desolvation process. The total free energy of binding, ΔG_{bind}, can be attributed to this process as a whole, from independent ligands and receptors in the surrounding physiological environment, up to the ligand–receptor complex, but only if these intermediate steps are rapid enough. For each case, one should ascertain that there is no rate-limiting step during intermediate stages.

Any QSAR model is a rationalization between chemical structure and the experimentally measured property. Any mathematical (empirical) SAR model provides a framework to test new, potentially more active, molecular structures, and a rationale for the initial hypothesis. However, all the parameters that cannot be directly measured (e.g., geometric variations of the binding site, entropic changes, etc.) remain *hidden*. For this reason, one should exercise caution in 3D-QSAR studies. In the MTD-MC and MTD-ADJ methods, we use a time-sliced (frozen) model. In principle, one has to account for kinetic bottlenecks in the intermediate steps, but the measure of the binding process (ΔG_{bind}) remains thermodynamic (not kinetic) in nature. In the AChE case, with flexible substrates, our model proposes a certain binding geometry without direct structural evidence from e.g. X-ray crystallography. Thus, a certain bias exists, regarding both the receptor site geometry and the adjustment of biological activities.

The 3D-QSAR models obtained using the MTD-ADJ procedure were characterized by significant statistical parameters and were similar to the unadjusted models. Of interest was the comparison between the adjusted (refined) and the unadjusted (crude) model. Since the differences in results were small, we have more confidence in both the unadjusted and the adjusted results. The refined model yields more information in the optimized receptor map. The selected bioactive conformers are extended, occupy cavity vertices and, for the same structures, differ in irrelevant regions. The optimized conformational map of the neutral ligands obtained from the MTD-ADJ model was graphically overlapped in the active site of the crystallized AChE from *Torpedo californica*. We have discussed the AChE binding site for neutral ligands. Our QSAR method is limited by the available information, and does not reflect exactly the physicochemical reality. This limit is apparent in our model, where two irrelevant vertices interact with the Trp^{84}-Tyr^{442} side chains. However, no active site minimization was undertaken, and, as pointed out earlier, the Y_{exp} values used for the optimization procedure were not determined using AChE from *Torpedo californica* [4]. Thus, differences in the active site may exist, which could account for some of the inadequacies observed in our model.

The activity adjustment is required by the fact that in the 3D-QSAR methodology, a single (active) conformer is used to explain Y_{exp}. Our adjusted model attempts to extract maximum of information, thereby improving the signal-to-noise ratio. Its utility is apparent in predictivity applications. When a new molecule is designed, its active conformation can be predicted by the refined QSAR model based on the final MTD-ADJ map. Its predicted Y_{exp} can then be obtained from Y_{adj}, because $\log \alpha$ is pre-computed. One could then select compounds for which the predicted $\Delta Y = Y_{adj} - Y_{exp}$ is minimal, in which case the entropic cost of transitioning to the active conformation is reduced.

The adjustment of biological activities, for series with conformationally flexible ligands, can be implemented in any 3D-QSAR method. This method can provide information regarding the active conformations, and can be used to gain further insight into the ligand–receptor models for which no structural data are available.

Acknowledgements

We thank Dr. Andy Vinter (Cambridge University, U.K.) for COSMIC, and Prof. Al Leo (Pomona College, Claremont, CA, U.S.A.) for MacLogP. The Department of Physiology, University of Medicine and Pharmacy, and the Institute of Chemistry, Timisoara Branch of the Romanian Academy (both from Timisoara, România), are acknowledged for support. Part of this work (T.I.O) was supported by the U.S. Department of Energy through a Los Alamos National Laboratory LDRD-CD grant. Intermediate and final statistical results, as well as the optimized receptor maps and selected conformations obtained for all MTD-QSAR variants tabulated in Table 2, are available from the authors upon request.

References

1. Gundertofte, K., Palm, J., Petterson, I. and Stamvik, A., J. Comput. Chem., 12 (1991) 200.
2. Simon, Z., Chiriac, A., Holban, S., Ciubotariu, D. and Mihalas, G.I., Minimum Steric Difference. The MTD Method for QSAR Studies, Research Studies Press, Letchworth, and Wiley, NY, 1984.

3. Ciubotariu, D., Deretey, E., Oprea, T.I., Sulea, T., Simon, Z., Kurunczi, L. and Chiriac, A., Quant. Struct.–Act. Relatsh., 12 (1993) 367.
4. Järv, J., Kesvatera, T. and Aakvisaar, A., Eur. J. Biochem., 67 (1976) 315.
5. Vinter, J.G., Davis, A. and Saunders, M.R., J. Comput.-Aided Mol. Design, 1 (1987) 31.
6. Morley, S.D., Abraham, R.J., Haworth, I.S., Jackson, D.E., Saunders, M.R. and Vinter, J.G., J. Comput.-Aided Mol. Design, 5 (1991) 475.
7. Vinter, J.G., COSMIC 2.0 PC, using DBOS and INTERACTOR, Polyhedron Software, 1992.
8. Gilbert, K.E., MOPAC 6.0, 1991, program recompiled for 40 heavy and 40 light atoms by Mateoc, F. and Kurunczi, L., 1992. The accuracy of this version was tested using the MOPAC 4.0 manual examples (QCPE No. 445).
9. Stewart, J.J.P., J. Comput.-Aided Mol. Design, 4 (1991) 1.
10. Sussman, J.L., Harel, M., Frolow, F., Oefner, C., Goldman, A., Toker, L. and Silman, I., Science, 253 (1991) 872.
11. Pullman, A., In Pullman, A., Jortner, J. and Pullman, B. (Eds.) Modelling of Biomolecular Structures and Mechanisms, Kluwer Academic Publishers, Dordrecht, 1995, pp. 11–23.
12. Cheng, Y.-C. and Prusoff, W., Biochem. Pharmacol., 22 (1973) 3099.
13. Kurunczi, L., Sulea, T. and Oprea, T.I., J. Mol. Struct. (THEOCHEM), 306 (1994) 93.
14. Gergen, I., Bohl, M., Simon, H. and Simon, Z., Rev. Roum. Chim., 34 (1989) 995.
15. Taft, R.V., In Newman, M.S. (Ed.) Steric Effects in Organic Chemistry, Wiley, New York, NY, 1956, pp. ?.
16. Koppel, I.A., Karelson, M.M. and Palm, V.A., Org. React., 10 (1973) 479.
17. Palm, V.A., Püssa, T.O., Nummert, V.M. and Talvik, I.V., Org. React., 10 (1973) 243.
18. Talvik, I.V. and Palm, V.A., Org. React., 8 (1971) 445.
19. MacLogP v. 1.0.3, Biobyte Corporation, Claremont, CA.
20. Oprea, T.I., Ciubotariu, D., Sulea, T. and Simon, Z., Quant. Struct.-Act. Relatsh., 12 (1993) 21.
21. Lanteri, S., Chemometr. Intell. Lab. Syst., 15 (1992) 159.
22. Hansch, C., Acc. Chem. Res., 26 (1993) 147.
23. Gilson, M.K., Straatsma, T.P., McCammon, J.A., Ripoll, D.R., Faerman, C.H., Axelsen, P.H., Silman, I. and Sussman, J.L., Science, 263 (1994) 1276.
24. Quinn, D.M., Chem. Rev., 87 (1987) 955.
25. Weise, C., Kreienkamp, H.J., Raba, R., Aaviksaar, A. and Hucho, F., J. Protein Chem., 9 (1990) 53.
26. Kreienkamp, H.J., Weise, C., Raba, R., Aaviksaar, A. and Hucho, F., Proc. Natl. Acad. Sci. USA, 88 (1991) 6117.

APPENDIX

The MTD method [2]

The hypermolecule is the result of the approximate (non-hydrogen) atom per atom superimposition of the molecules i= 1, 2, …, N of the investigated series. The vertices j= 1, 2, …, M of the hypermolecule correspond to the positions of these atoms. If molecule i occupies vertex j, $x_{ij} = 1$, while $x_{ij} = 0$ if the vertex is not occupied. The minimal steric difference MTD_i of the molecule i with respect to the receptor is calculated according to the formula:

$$MTD_i = s + \sum \epsilon_j x_{ij} \quad \text{(A1)}$$

with $\epsilon_j = -1$, 0 or $+1$ for vertices attributed to the receptor cavity (beneficial), to the exterior (irrelevant) region and to the receptor walls (detrimental), respectively; s is the total number of cavity vertices. Thus, MTD_i is a measure of the steric misfit of the molecule i with respect to the receptor cavity and it is equal to the number of occupied wall vertices plus the number of unoccupied cavity vertices. The attributions of $\epsilon_j = -1$, 0 or $+1$ to the vertices j are performed according to an optimization procedure. One starts from an initial attribution of ϵ_j's ($\epsilon_j = +1$ for vertices occupied only in inactive molecules, $\epsilon_j = -1$ for vertices occupied only in active molecules) and changes these attributions aiming at an increase of the correlation coefficient, r, for a regression equation of the type

$$Y = a\,MTD + b, \quad a < 0 \quad \text{(A2)}$$

The MTD-MC method [3]

In this MTD version, the molecules M_i can be described by a vectorial, topographical descriptor $X\{x_{ijk}\}$, which contains, for each conformation k of the molecule M_i, a number of M structural variables x_{ijk} (i = 1, 2, …, N; j = 1, 2, …, M; k = 1, 2, …, C_N) corresponding to the M vertices of the hypermolecule H with $x_{ijk} = 1$ if vertex j of H is occupied by the molecule M_i in conformation k and $x_{ijk} = 0$ if this vertex is not occupied. If a molecule has several low-energy conformations, it will adopt the one that best fits to the receptor, i.e. the conformation with the lowest MTD value. For the molecule M_i in the conformation k and a given receptor, the distribution of vertices to the receptor is: $\epsilon_j = -1$ for cavity, $\epsilon_j = 1$ for receptor walls, and $\epsilon_j = 0$ for exterior vertices:

$$MTD_{ik} = s + \sum \epsilon_j\, x_{ijk} \quad \text{(A3)}$$

We consider that MTD_i is the minimum over k of the MTD_{ik} values.

Linear regression and cross-validation procedures

The correlation equations are obtained by the least-squares method. Corresponding statistical parameters

are given for each linear regression: the number of independent variables (n), the squared regression coefficient (r^2), the standard error of estimate (s), and the Fisher test (F). The ability of derived QSARs to predict the biological activity is tested by means of two cross-validation techniques: leave-half-out (LHO) and leave-one-out (LOO). In LHO [18], the compounds, in decreasing order of biological activities, are separated in two subseries: ODD, i = 1, 3, 5, ..., and EVEN, i = 2, 4, 6, ... MTD and biological activity values for the compounds of the ODD subseries are calculated with the optimized receptor map and correlation equation of the EVEN subseries, and *vice versa*. In LOO, all compounds of the series are left out once and only once from the model. The biological activity of the removed compound is calculated using the optimized receptor map and correlation equation obtained for the remaining N–1 compounds. In both techniques, the predictive q^2 is calculated according to

$$q^2 = 1 - \frac{\text{PRESS}}{\text{SD}} = 1 - \frac{\sum_{i}^{N}(Y_{i,exp} - Y_{i,pred})^2}{\sum_{i}^{N}(Y_{i,exp} - Y_{i,mean})^2} \quad \text{(A4)}$$

where $Y_{i,exp}$ is the actual value for the biological activity, $Y_{i,pred}$ is the predicted value for the biological activity, and $Y_{i,mean}$ is the average biological activity for the given series. PRESS (predictive residual sum of squares) represents the standard errors of the cross-validated predictions, and it is the sum, over all compounds, of the squared differences between the actual and predicted biological activity values. SD is the sum of squared deviations of biological activities from their mean, Y_{mean}. For predictivity analyses during cross-validation, $Y_{adj}$was reconverted to Y_{exp}: $Y_{exp,pred} = Y_{adj,pred} + \log \alpha$.

Implementation of the adjusted version in the MTD-MC method

The MTD-ADJ program structure, written in FORTRAN 77, includes two loops and follows several steps, as presented below (see the program flow chart in Figure 5):

(i) Selection of the start vertex attribution (i.e., the initial receptor map, S°).

(ii) Calculation of the MTD_{ik} values for each conformation k for all compounds i.

(iii) Selection of the conformer(s) with the minimal MTD_{ik} value for each compound i.

(iv) Calculation of the Boltzmann population (i.e., relative amount α or $\Sigma\alpha$) for the conformer(s) selected at step (iii).

(v) Calculation of the adjusted activity data for conformer(s) selected in step (iii), according to Equation (5).

(vi) Calculation of the conventional r^2 value (using Y_{adj}) and the corresponding correlation equation for the vertex attribution of step (i) (statistical aspects are given in the section 'Discussion').

(vii) Systematic variation of vertex attribution (which consists of 2M possibilities, where M is the number of vertices in the MTD hypermolecule).

(viii) Recycling steps (i) – (vii) 2M times.

(ix) Selection of the maximal conventional r^2 value along with the corresponding correlation equation and receptor map.

(x) Starting the computations with the map selected in step (ix) beginning from step (i).

(xi) Recycling steps (i) – (x) until one reaches the stop criterion (xii).

(xii) Stop criterion: three consecutive equal maximal conventional r values, or a new maximal conventional r^2 value lower than the previous one.