Eur. J. Med. Chem. 33 (1998) 617–623

The use of molecular similarity indices in the determination of a bioactive conformation

Ana Castro, W. Graham Richards*

New Chemistry Laboratory, South Parks Road, Oxford OX1 3QT, UK

(Received 10 December 1997; accepted 16 February 1998)

Abstract – A QSAR study of a series of thiourea inhibitors of acetylcholinesterase has been carried out with molecular similarity indices used as parameters for correlating the activity and also to suggest the receptor-bound conformation. The availability of the structure of AChE has provided the opportunity to extend the QSAR study to a ligand–receptor docking analysis. © Elsevier, Paris

QSAR / molecular similarity / AChE inhibitors

1. Introduction

The development of cholinomimetic drugs as a therapeutic approach to Alzheimer's disease (AD) has been of great interest since the demonstration of that a cholinergic deficit was involved [1]. Among the different possible approaches, the inhibition of acetylcholinesterase (AChE) has been the most widely studied. The mechanism proposed to explain the memory improvement through AChE inhibition is an increase in acetylcholine (ACh) levels in the central cholinergic synapses involved in the memory circuit [2]. AChE inhibitors produce a sustained and prolonged stimulation of postsynaptic cholinergic receptors by reducing the enzymic degradation of released ACh [3].

The chemical structures of these inhibitors are very diverse so that the successful design of new potent inhibitors of AChE depends on the ability to rationalise the experimental structure – activity relationships. In this area, the application of similarity indices in QSAR analysis has been used with success in recent years [4, 5]. The similarity indices measure properties of molecules that are related to conformation such as the electrostatic potential and the shape. At the same time these properties are considered important in intermolecular interactions. The similarity is calculated from the three dimensional structure of the molecules and is measured in terms of these properties. Following this approach, in using molecular similarity as a variable we are introducing the possibility of incorporating knowledge of conformations into QSAR analysis. In that case, the bioactive conformation should provide more relevant similarity indices, and perhaps a better correlation with the biological activity [6].

The family of [2-(4-piperidinyl)ethyl]thioureas (I) and 1-[2-aminoethoxyalkyl]-3 aroylthioureas (II) (*figure 1*), which have shown potent antiacetylcholinesterase activity [7, 8], were selected for this study. The conformational flexibility of these compounds makes it possible to confirm how the similarity indices can be useful tools in the determination of the receptor-bound conformation.

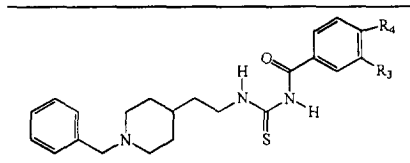
The crystal structure of AChE has been determined by Sussman et al. [9]. Validation of the QSAR analysis can be carried out, using docking and molecular dynamics techniques which permit the establishment of the interaction energy of the enzyme/inhibitor complex.

2. Results and discussion

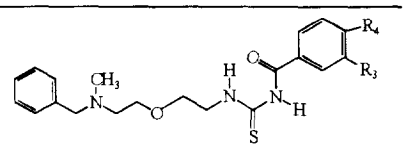
2.1. Conformational analysis

All the molecular modelling techniques described here were performed on Silicon Graphics workstations using the SYBYL 6.3 molecular modelling software [10] and

*Correspondence and reprints



I



II

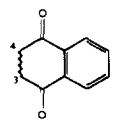
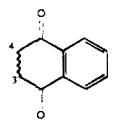
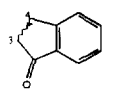
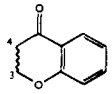
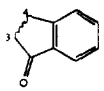
Compound	no.	Compound	no.
	1		13
R_1 = OMe, R_4 = OMe	2		14
R_1 = H, R_4 = SO_2Me	3	R_1 = H, R_4 = OMe	15
R_1 = H, R_4 = COMe	4	R_1 = H, R_4 = Cl	16
R_1 = H, R_4 = OMe	5	R_1 = Cl, R_4 = Cl	17
	6	R_1 = H, R_4 = Me	18
			
	7		19
R_1 = H, R_4 = H	8	R_1 = H, R_4 = H	20
R_1 = H, R_4 = COPh	9	R_1 = H, R_4 = Ph	21
R_1 = H, R_4 = Cl	10		
R_1 = H, R_4 = CF_3	11		
	12		

Figure 1. Structures for the studied thiourea compounds (family I: left-hand column; family II: right-hand column).

the similarity and analysis using the Automated Similarity Program (ASP) [11] and the tools for structure–activity relationships (TSAR) [12].

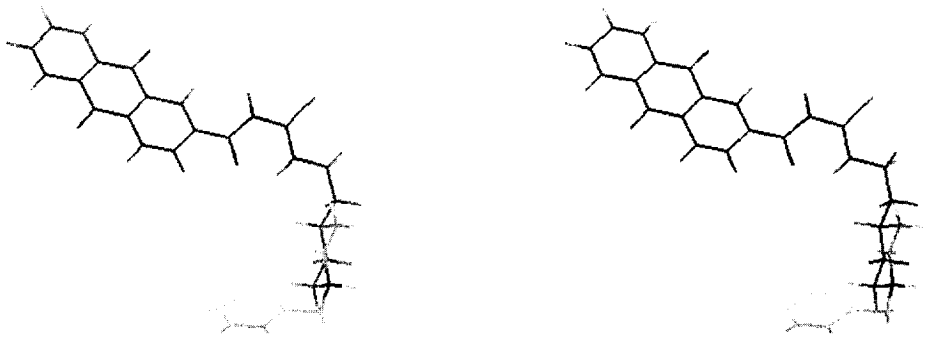
Since the piperidine ring is likely to be protonated at physiological pH, a proton was added to the neutral structures. A conformational analysis was carried out using the molecular dynamics technique at 700 K for 1 ps followed by a run at 200 K for 1 ps with 10 cycles of heating and cooling. From the trajectory only those conformers found between 200-210 K were selected.

Each selected conformation was fully optimized using the standard Tripos molecular mechanics force field. Partial atomic charges required for calculation of the electrostatic potential were calculated using the Gasteiger-Hückel method [13].

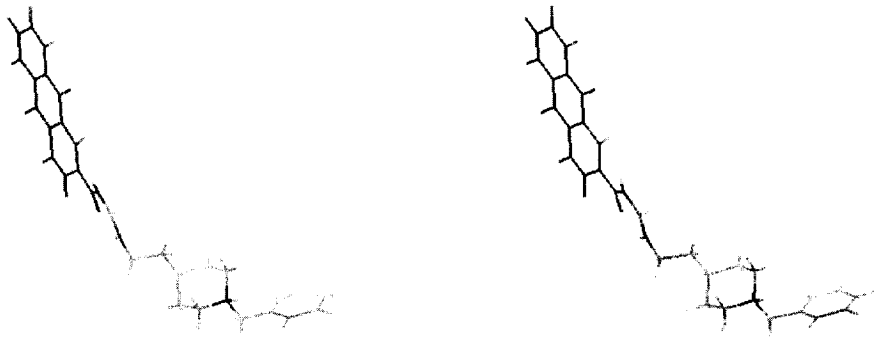
These results were analyzed to obtain three distinct low-energy conformations. Series A are formed by the conformation with the lowest energy found.In order to study some alternative conformations, conformers up to 5 kcal/mol above the series A conformers were selected forming series B and series C. These series of molecules, each representing a distinct conformation, were used as starting points for further study. Intramolecular hydrogen bonding stabilizes the conformations, which are characterized by a six membered pseudocycle. This fact has been confirmed by several studies of similar compounds involving ^1H and ^{13}C NMR and IR experiments [14]. The aromatic system is perpendicular to the piperidine ring as in the active conformation found by Hopfinger et al. [15] for the indanone-benzylpiperidine family (see *figure 2*).

2.2. *Similarity calculations*

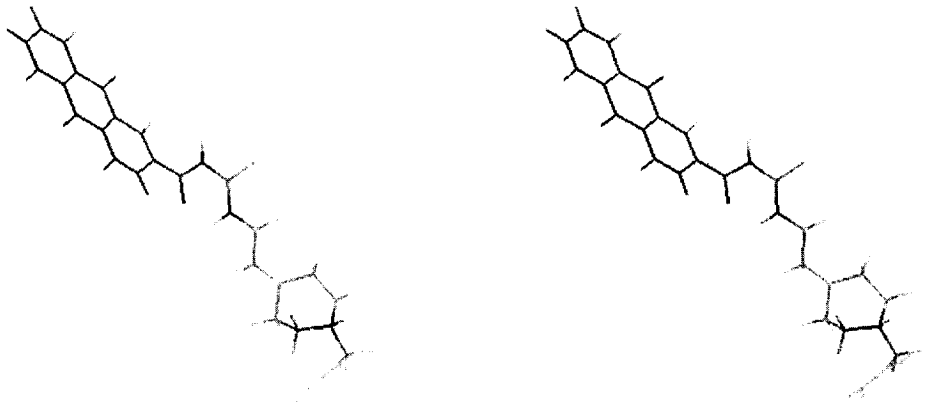
Molecular alignment of the molecules has a major influence on calculated similarity indices[16]. In our case,



SERIES A



SERIES B



SERIES C

Figure 2. Stereoviews of the three distinct low-energy conformations of compound **1** used in the study to generate the different equations.

Table I. Biological and predicted data for Series A derivatives.

n	log 1/IC_{50} obs.	log 1/IC_{50} calc.	residual	*n*	log 1/IC_{50} obs.	log 1/IC_{50} calc.	residual
1	8.699	8.738	−0.039	**13**	8.221	7.838	0.382
2	8.301	7.918	0.382	**14**	7.886	7.564	0.321
3	8.221	8.138	0.082	**15**	7.585	7.664	−0.079
4	8.125	8.241	−0.116	**16**	7.522	7.246	0.275
5	8.000	7.975	0.024	**17**	7.259	7.564	−0.305
6	7.958	7.477	0.480	**18**	7.113	7.044	0.068
7	7.958	8.019	−0.061	**19**	7.000	7.053	−0.053
8	7.886	7.955	−0.069	**20**	6.920	7.277	−0.357
9	7.853	7.762	0.090	**21**	6.638	7.412	−0.774
10	7.698	7.255	0.442				
11	7.481	7.875	−0.394				
12	7.097	7.392	−0.295				

the molecules were fitted using a function based on a mixture of shape, electrostatic charge and lipophilicity. The combined property (cp) for atom i is calculated as follows:

$$\mathrm{cp}(i) = 1.0\, w_s + q_i w_Q + l_i w_L$$

where w_s, w_Q and w_L are user-defined weights (shape, charge and lipophilicity) and q_i and l_i are the partial charge and lipophilicity for atom i respectively. In order to get the best alignment the weights were varied systematically and the optimum values were chosen as judged by the largest cross-validated r^2. The best results were obtained when $w_s = 1$, $w_Q = 1$, $w_L = 0$.

Molecular similarity indices were calculated, for each series of conformers, using the rigid option within the ASP package. This method optimises similarity by changing the relative orientation of the lead and comparison molecules. *Table I* shows the biological data and the calculated parameters for the analysis.

Carbo indices were calculated for the three different series of conformations using the compound **1**, the most active in the series as the reference structure. The Carbo index was calculated for shape alone and for charge alone to investigate the individual contributions of these two descriptors, as well as a combined Carbo index based on an equal contribution of shape and charge.

2.3. *QSAR analysis*

For the QSAR analysis, similarity indices were correlated with the activity through stepwise multiple regression using the TSAR package.

Biological data were taken from Vidaluc et al. [7, 8] for a set of 21 compounds tested as AChE inhibitors and used as log (1/IC_{50}), (where IC_{50} is in nM) as in *table I*.

In order to evaluate the different factors (shape or electrostatic potential) the indices were used separately in the regression analyses. The electrostatic index by itself was found not to be significant. The shape index alone showed results slightly better that the combined index, see *table II*.

However the best predictive ability is obtained by the model using shape and electrostatic potential indices, giving the best values for series A ($r = 0.80$, $r^2 = 0.64$). A list of estimated and predicted activity values for this model can be found in *table I*, while *figure 3* shows the regresion between experimental and predited values. Therefore multiple regression analysis was carried out for the other two series using shape and electrostatic potential as variables. Correlation coefficients for all of them are as follows:

Series A:

$$\mathrm{Log}\ (1/IC_{50}) = 0.44\ \mathrm{Carbo}_{\mathrm{shape}} - 0.22\ \mathrm{Carbo}_{\mathrm{EP}} + 7.68$$
$$n = 21,\ s = 0.33,\ F = 6.88 \times 10^{-5},\ r = 0.80,$$
$$r^2 = 0.64,\ r^2_{cv} = 0.60 \qquad (1)$$

Series B:

$$\mathrm{Log}\ (1/IC_{50}) = 0.30\ \mathrm{Carbo}_{\mathrm{shape}} - 0.05\ \mathrm{Carbo}_{\mathrm{EP}} + 7.68$$
$$n = 21,\ s = 0.47,\ F = 0.01,\ r = 0.53,$$
$$r^2 = 0.28,\ r^2_{cv} = -0.10 \qquad (2)$$

Series C:

$$\mathrm{Log}\ (1/IC_{50}) = 0.22\ \mathrm{Carbo}_{\mathrm{shape}} - 0.07\ \mathrm{Carbo}_{\mathrm{EP}} + 7.68$$
$$n = 21,\ s = 0.51,\ F = 0.19,\ r = 0.40,$$
$$r^2 = 0.16,\ r^2_{cv} = -0.35 \qquad (3)$$

Table II. Summary of statistical results.

Carbo indices	Series A	Series B	Series C
Shape	$r = 0.69$, $r^2 = 0.48$	$r = 0.52$, $r^2 = 0.27$	$r = 0.38$, $r^2 = 0.14$
Shape + E.P.	$r = 0.80$, $r^2 = 0.64$	$r = 0.53$, $r^2 = 0.28$	$r = 0.40$, $r^2 = 0.16$
Combined	$r = 0.67$, $r^2 = 0.44$	$r = 0.51$, $r^2 = 0.26$	$r = 0.36$, $r^2 = 0.13$

From these results it is clear that the best predictive is the *Series A*. Following the idea that the bioactive conformation should provide more relevant similarity indices and perhaps better correlations, series A should be the family of receptor-binding conformations.

2.4. *QSAR validation*

To validate the QSAR analyses we have calculated the total interaction energies between the most potent inhibitor (**1**) and the enzyme, as well as evaluting the binding modes of **1** to provide structural information for the design of more inhibitors, using docking and molecular dynamics techniques.

The coordinates of the protein were obtained from the X-ray structure of AChE isolated from *Torpedo Californica* [9], as deposited in the Brookhaven Protein Data Bank (entry 2ACE). Crystallographic water molecules present in X-ray structure as well as the modeled acetylcholine molecule were deleted. Polar and aromatic hydrogen atoms, in addition to any missing heavy atoms, were added to the protein using the SYBYL program.

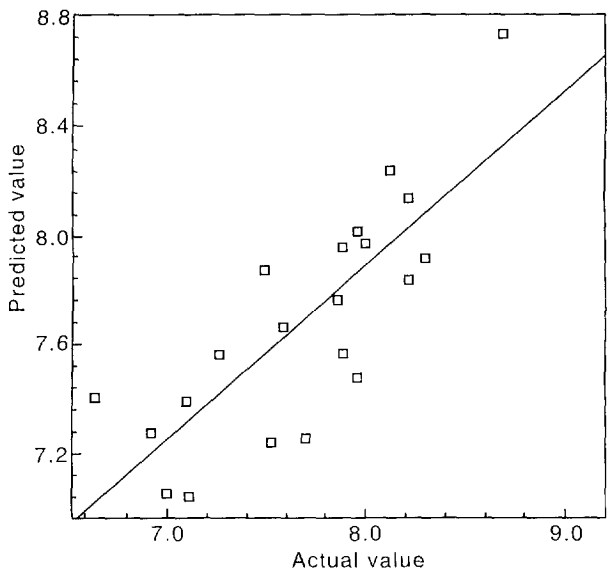


Figure 3. Plot of predicted (Eq. (1)) versus observed log ($1/IC_{50}$) values.

Table III. Results of 50 ps Dynamics simulations of the complex structures of AChE/compound **1** and correlation coefficients.

Complex enz/inh	E_{inter} (kcal/mol)	r	r^2	r^2_{cv}
Series A	− 202.36	0.80	0.64	0.60
Series B	− 188.45	0.53	0.28	− 0.10
Series C	− 177.04	0.40	0.16	− 0.35

The distinct low-energy conformations of compound **1** were manually docked into the active site of AChE with the aid of computer graphics. These initial ligand/AChE complexes were minimized using MAXIMIN2, and the minimized complex structures were used as starting points for molecular dynamics simulations, which were carried out by heating the system over 10 ps from 0 to 300 K, with a time step 0.001 ps. The system was then equilibrated for 30 ps, and a constant-temperature dynamics simulation was then performed for 50 ps. Snapshots were taken at 0.1 ps intervals during the simulation period and were then subsequently analyzed to obtain the total interaction energies between the inhibitor and the enzyme.

The total interaction energy was defined as:

$$E_{int} = E_{complex} - (E_{enz} + E_{inh})$$

If we assume: (1) that E_{int} is proportional to the enthalpy of binding (ΔH_{bind}) and (2) that the entropy of binding (ΔS_{bind}) would be constant for this series, then E_{int} should be proportional to the binding free energy (ΔG_{bind}), i.e.:

(1) $E_{int} \approx \Delta H_{bind}$

$$\Rightarrow \Delta G_{bind} = \Delta H_{bind} - T\Delta S_{bind} \Rightarrow \Delta G_{bind} \approx E_{int}$$

(2) $\Delta S_{bind} \approx$ constant

The results for each of the three complexes are sumarized in *table III* which lists the interaction energies (E_{int}) and correlation coeficients. These data suggest that conformation A, with the lowest value of E_{int}, forms the most energetically favoured complex with AChE. This is in agreement with the series of conformers that showed the best correlation with biological activity (Eq. (1)).

From the molecular dynamics simulation it is posible to explain the high activity of compound **1** through hydrogen bonds and hydrophobic interaction, shown in *figure 4*. Complex AChE/**1** was established by: (a) the hydrogen bond between the positively charged NH pip-

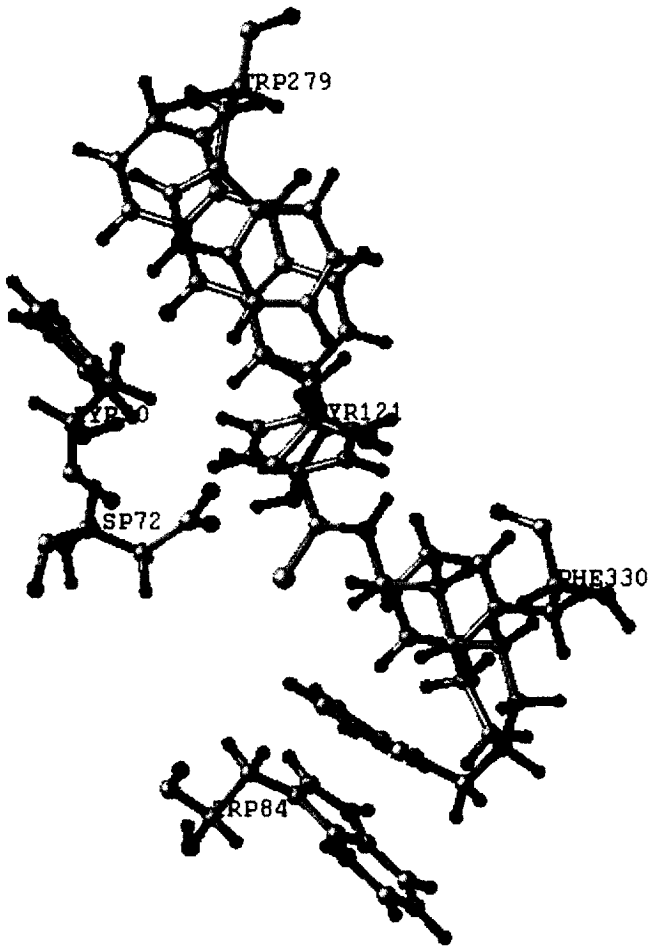


Figure 4. Binding modes of **1** to the active site gorge of acetylcholinesterase.

eridine and the negatively charged carboxylate side chain of Asp 72 and two hydrophobic interactions: (b) the aromatic ring is parallel to the indole side chain of Trp 279, (c) the N-benzyl substituent forms an off-center π-stacking interaction with the indole chain of Trp 84.

From the QSAR analysis we can conclude that shape index has presented a better description than electrostatic potential index, but all the correlations have been improved when the electrostatic potential index has also been considered. This shape dependence could be related to the fact that one of the features of AChE structure is its deep and narrow channel [17]. This gorge penetrates halfway into the enzyme widening out close to its end. The active site of AChE lies close to the wide end of this deep and narrow aromatic gorge. Thus inhibitors with a cylindrical shape can fit into and slide along the inner surface of the tube. In this case, the conformational flexibility of the compounds allows the adoption of the corresponding shape. This means that steric factors play an important role in the biological activity of these compounds, and this fact is shown through shape similarity indices.

On the other hand, AChE has a remarkably large dipole moment [18]. The active-site gorge posseses a positive charge at one end and a negative charge at the other, hence the preferred orientation and position of the inhibitors in the tube is dictated, in part, by electrostatic interactions. This could explain why both shape and electrostatic potential indices combined have been shown to be the best description.

3. Conclusions

With this study we have tried to illustrate the potential application of the molecular similarity techniques to the determination of bioactive conformation. The results indicate that, in this instance, steric factors play a more important role than electrostatic factors, but both are needed to explain biological activity.

Molecular mechanics calculations and molecular dynamics simulations were used to validate QSAR study, as well as to predict the binding modes of the most potent inhibitor.

Acknowledgement

A.C. acknowledges the financial support provided by the Ramsay Memorial Fellowships Trust and C.S.I.C.; and by the Spanish Ministerio de Educación y Ciencia through a F.P.U. scholarship.

References

[1] Coyle J.T., Price D.L., Delong M.R., Science 219 (1983) 1184-1190.

[2] Inove A., Kawai T., Wakita M., Iimura Y., Sugimoto H., Kawakami Y., J. Med. Chem. 39 (1996) 4460-4470.

[3] Appleyard M.E., Biochem. Soc. Trans. 22 (1994) 749-755.

[4] Good A.C., So S.S., Richards W.G., J. Med. Chem. 36 (1993) 433-438.

[5] Good A.C., Peterson S.J., Richards W.G., J. Med. Chem. 36 (1993) 2929-2937.

[6] Montanari C.A., Tute M.S., Beezer A.E., Mitchell J.C., J. Comput-Aided Mol. Design 10 (1996) 67-73.

[7] Vidaluc J.L., Calmel F., Bigg D., Carilla E., Stenger A., Chopin Ph., Brilley M., J. Med. Chem. 37 (1994) 689-695.

[8] Vidaluc J.L., Calmel F., Bigg D., Carilla E., Brilley M., J. Med. Chem. 38 (1995) 2969-2973.

[9] Sussman J.L., Harel M., Frolow F., Oefner C., Goldman A., Toker L., Silman I., Science 253 (1991) 872-879.

[10] SYBYL Molecular Modelling System, Version 6.3, Tripos, Inc., St. Louis, MO 63144.

[11] ASP v. 3.11, Oxford Molecular Ltd., Oxford, 1994.

[12] TSAR v. 3.1, Oxford Molecular Ltd., Oxford, 1997.

[13] Purcel W.P., Singer J.A., J. Chem. Eng. Data 12 (1967) 235-246.

[14] Suda L.V., Stahyanarayana D.N., J. Mol. Struct. 13(3-4) (1985) 253-259.

[15] Cardozo M.G., Kawai T., Iimura Y., Sugimoto H., Yamanishi Y., Hopfinger A.J., J. Med. Chem. 35 (1992) 590-601.

[16] Parretti M.F., Kroemer R.T., Rothman J.H., Richards W.G., J. Comput. Chem. 18 (1997) 1344-1353.

[17] Iounoe A., Takatoshi K., Wakita M., Iimura Y., Sugimoto H., Kawakami Y., J. Med. Chem. 39 (1996) 4460-4470.

[18] Ripoll D.R., Faerman C.H., Axelsen P.H., Silman I., Sussman J.L., Proc. Natl. Acad. Sci. USA 90 (1993) 5128-5132.