

## Structure-Based Alignment and Comparative Molecular Field Analysis of Acetylcholinesterase Inhibitors

Sung Jin Cho, Maria Luisa Serrano Garsia,<sup>†</sup> Jim Bier,<sup>‡</sup> and Alexander Tropsha\*

The Laboratory for Molecular Modeling, Division of Medicinal Chemistry and Natural Products, School of Pharmacy, University of North Carolina, Chapel Hill, North Carolina 27599

Received May 22, 1996<sup>⊗</sup>

The method of comparative molecular field analysis (CoMFA) was used to develop quantitative structure–activity relationships for physostigmine, 9-amino-1,2,3,4-tetrahydroacridine (THA), edrophonium (EDR), and other structurally diverse inhibitors of acetylcholinesterase (AChE). The availability of the crystal structures of enzyme/inhibitor complexes (EDR/AChE, THA/AChE, and decamethonium (DCM)/AChE) (Harel, M.; et al. Quaternary ligand binding to aromatic residues in the active-site gorge of acetylcholinesterase. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 9031–9035) provided information regarding not only the active conformation of the inhibitors but also the relative mutual orientation of the inhibitors in the active site of the enzyme. Crystallographic conformations of EDR and THA were used as templates onto which additional inhibitors were superimposed. The application of cross-validated  $R^2$  guided region selection method, recently developed in this laboratory (Cho, S. J.; Tropsha, A. Cross-Validated  $R^2$  Guided Region Selection for Comparative Molecular Field Analysis (CoMFA): A Simple Method to Achieve Consistent Results. *J. Med. Chem.* **1995**, *38*, 1060–1066), to 60 AChE inhibitors led to a highly predictive CoMFA model with the  $q^2$  of 0.734.

### Introduction

Modern methods for computer-assisted drug design fall into two major categories generally known as ligand-based and receptor-based methods. The former, which include conventional quantitative structure–activity relationships (QSAR),<sup>1</sup> active analog approach,<sup>2,3</sup> and recently comparative molecular field analysis (CoMFA),<sup>4</sup> are based entirely on experimental structure–activity relationships for receptor ligands or enzyme inhibitors, and their application in the last 30 years led to several drugs currently on the market (reviewed in ref 5). The latter methods which include docking and advanced molecular simulations require that the structural information about the receptor or enzyme should be available from X-ray crystallography, NMR, or protein homology model building. This strategy has become available only recently, with rapid advances in structure elucidation methods, and it already leads to several promising drug candidates (reviewed in ref 6).

The ligand-based methods of analysis are used widely since they are not very computationally intensive and afford rapid generation of QSAR's from which biological activity of newly designed compounds can be predicted. Most of the existing methods require generation of a 3D pharmacophore hypothesis (i.e., unique 3D arrangements of important functional groups common to all or the majority of the receptor ligands). In many cases, when the receptor ligands are not very diverse structurally and include conformationally rigid compounds, a pharmacophore can be generated in a reasonably unbiased and unique way, using either automated (e.g., DISCO<sup>7</sup>) or semiautomated<sup>2</sup> pharmacophore prediction methods. However, this task of unique pharmacophore

generation becomes less feasible for more structurally diverse and/or conformationally flexible compounds. In general, in the absence of detailed structural information about the receptor binding site any pharmacophore inferred from only the ligand structure remains hypothetical.

Structure-based docking algorithms can generate fairly accurate orientation of known or designed receptor ligands in the active site (e.g., GRID,<sup>8</sup> DOCK<sup>9</sup>). However, predictions of binding affinity of receptor ligands either are fast but very inaccurate due to insufficient accuracy of scoring functions used in docking algorithms or are fairly accurate but very computationally intensive, when free energy simulation methods are used.<sup>10,11</sup> Therefore, it is still impractical to use structure-based drug design methods for practical purposes of QSAR.

It is quite appealing to combine the accuracy of the structure-based alignment and the computational efficiency of ligand-based methods. According to this approach, the ligand alignment is generated on the basis of the experimental or predicted orientation of molecules in the actual active site, and 3D QSAR methods are applied to this alignment afterward. For example, this strategy was successfully employed earlier to 59 inhibitors of HIV protease.<sup>12</sup> In this paper, we employ the combination of structure-based alignment and CoMFA to obtain three-dimensional QSAR for 60 chemically diverse inhibitors of acetylcholinesterase (AChE).

Inhibition of AChE is considered as one of the most promising strategies for the treatment of Alzheimer disease,<sup>13,14</sup> and the possible therapeutic applications in the treatment of Parkinson's disease,<sup>15</sup> aging,<sup>15</sup> and myasthenia gravis<sup>16</sup> have been implicated. Over the years, hundreds, if not thousands, of compounds have been synthesized and tested for anticholinesterase activity, and several of them have found clinical applications.<sup>13</sup> The chemical structures of these inhibitors are very diverse, ranging from bis-quaternary compounds

\* To whom correspondence should be addressed.

<sup>†</sup> On sabbatical leave from Universidad de Centrale de Venezuela, Caracas, Venezuela.

<sup>‡</sup> On sabbatical leave from Ferrum College, Ferrum, VA.

<sup>⊗</sup> Abstract published in *Advance ACS Abstracts*, November 15, 1996.

such as decamethonium (DME) to simple monocationic compounds such as edrophonium (EDR) to formally neutral tricyclic compounds such as THA (Table 1).

The successful design of new potent inhibitors of AChE depends on our ability to rationalize the experimental structure–activity relationships among the enzyme inhibitors. Such rationalization usually requires the construction of a three-dimensional (3D) pharmacophore that unifies common structural features among enzyme inhibitors or receptor ligands.<sup>2</sup> However, the great structural diversity of the AChE inhibitors makes it practically impossible to structurally align all the inhibitors in any unbiased way and generate a unique three-dimensional pharmacophore. As a result, earlier SAR studies were limited to series of structurally congeneric ligands.<sup>17–21</sup>

Recent X-ray crystallographic analysis of AChE from *Torpedo californica* (EC 3.1.1.7)<sup>22</sup> followed by X-ray determination of the complexes of the enzyme with three structurally diverse inhibitors, THA, EDR, and DME,<sup>23</sup> provided crucial information with respect to the orientation of these inhibitors in the active site of the enzyme. The crystallographic data indicated that each of the three inhibitors has a unique binding orientation in the active site of the enzyme. Their natural structural alignment would probably never have been predicted by any of the existing automated algorithms for ligand alignment or even by the researcher's imagination based on the ligand chemical structure alone.

We have used this natural structural alignment of the three inhibitors, THA, EDR, and DME, as a template onto which other structurally analogous AChE inhibitors have been superimposed. In order to obtain quantitative relationship between the structure and biological activities of the inhibitors, we have employed CoMFA<sup>4</sup> and cross-validated  $R^2$  guided region selection ( $q^2$ -GRS) routine, recently developed in this laboratory.<sup>24,25</sup> We report herein that the combination of structure-based alignment of 60 AChE inhibitors and CoMFA/ $q^2$ -GRS generates a highly predictive QSAR model with the  $q^2$  of 0.734.

### Computational Details

SYBYL Molecular Modeling Software<sup>26</sup> was used for structure generation and CoMFA. Molecular mechanics calculations including field-fit optimization were performed with the standard Tripos force field<sup>27</sup> and the minimum energy change of 0.05 kcal/mol as a convergence criteria. Charges were calculated using the Gasteiger–Huckel method as implemented in SYBYL. SYBYL grid (10° increment) and systematic (10° increment; energy threshold of 0.1) search methods were used to obtain the lowest-energy conformers. All calculations were performed on an IBM RS6000 Model 340.

**Biological Activity Data.** For this work we have selected 60 chemically diverse inhibitors of AChE (Table 1) whose activity was measured by four different research groups.<sup>28–31</sup> This data set included two of the three inhibitors (THA and EDR) whose complexes with AChE were characterized by X-ray crystallography.<sup>22</sup> The inhibitory activity of the compounds was expressed as IC<sub>50</sub> values (Table 1); the data were either presented in this form in the original publication<sup>31</sup> or calculated from the originally reported  $K_i$  values<sup>28–30</sup> using the Cheng–Prusoff equation<sup>32</sup>

$$IC_{50} = K_i(1 + S/K_m) \quad (1)$$

where  $K_i$  is the dissociation constant of the enzyme–inhibitor complex,  $S$  is the substrate concentration, and  $K_m$  is the Michaelis constant of the substrate.

In general, tabulating biological activity data from different sources should be done very carefully. All authors used similar methods to determine anticholinesterase activity of the inhibitors, and the enzyme used in these assays was obtained from the same source, an electric eel. Furthermore, the  $K_i$  values for the same compounds measured by different authors agreed with each other very closely. For example, the  $K_i$  values of choline reported in refs 28 and 14 were  $9.3 \times 10^{-4}$  and  $9.6 \times 10^{-4}$  M, respectively, and  $K_i$  values of (3-hydroxyphenyl)-trimethylammonium in refs 28 and 30 were  $2.10 \times 10^{-7}$  and  $3.10 \times 10^{-7}$  M, respectively. Thus, we felt confident that all the data are compatible with each other.

**Conventional CoMFA.** Conventional CoMFA was performed with the QSAR option of SYBYL. For all steps of conventional CoMFA, the default Sybyl settings were used except otherwise noted. For each CoMFA analysis, the minimum  $\sigma$  was set to 2.0 to expedite the calculation. The steric and electrostatic field energies were calculated using sp<sup>3</sup> carbon probe atoms with +1 charge. The CoMFA grid spacing was 2.0 Å in all three dimensions within the defined region, which extended beyond the van der Waals envelopes of all molecules by at least 4.0 Å. The CoMFA QSAR equations were calculated with the partial least square (PLS) algorithm. The optimal number of components (ONC) in the final PLS model was determined by the  $q^2$  and standard error of prediction (SDEP) values, obtained from the leave-one-out cross-validation technique; the number of components with the lowest SDEP value was selected as ONC.

**$q^2$ -GRS Routine.** The  $q^2$ -GRS process and its application to various data sets were described in detail elsewhere.<sup>24,25</sup> This method was developed due to our observation that the  $q^2$  values obtained as a result of conventional CoMFA are sensitive to the orientation of aligned molecules on the user's terminal and can vary by as much as  $0.5q^2$  unit if this orientation is systematically changed. Thus, if the researcher applies conventional CoMFA to one orientation, both the orientation and results should be viewed in general as a random sample. On the contrary, the  $q^2$ -GRS method leads to reproducible  $q^2$  values that do not depend on the orientation of molecular aggregate of aligned molecules on the user terminal.<sup>24</sup> This routine includes several steps as follows: (1) a conventional CoMFA is performed initially using an automatically generated region file (rectangular grid); (2) the rectangular grid, encompassing aligned molecules, is then divided into 125 small boxes of equal size; (3) for each of this newly generated subregion files, a separate CoMFA is performed with the step size of 1.0 Å; (4) the regions with the  $q^2$  value greater than the specified threshold value are selected for further analysis; (5) the selected regions are combined to generate a master region file; and (6) the final CoMFA is performed using the master region file.

**Structure Alignment.** In order to obtain a "native" alignment of THA, EDR, and DME, the polypeptide backbones of the AChE from three AChE/inhibitor complexes<sup>23</sup> were superimposed using rigid rms fit routine (Figure 1). The rms deviations between the backbone atoms of the AChE chains of AChE/THA and AChE/EDR complexes and AChE/THA and AChE/DME complexes were 0.2868 and 0.3461 Å, respectively. The coordinates of all three compounds were then extracted from their respective crystallographic complexes, and their geometry was optimized individually with the Tripos force field with no constraints on the internal geometry of the molecules.

The structure of each of the three inhibitors (EDR, THA, and DME) was then used as a template onto which their close analogs (Table 1) were superimposed as follows. Compound **6** was initially superimposed with EDR using the quaternary nitrogen (N.4), the centroid of the heterocyclic six-membered ring, and the oxygen (O.3) adjacent to the ring as the respective template and target atoms to generate the lowest rms fit. Compound **6** was then field-fitted to EDR and reoptimized with the field-fit option turned off. Using compound **6** as the template, compounds **1–5** and **7–9** were rms-fitted employing the same sets of atoms used to fit compound **6** to EDR.

In order to align compounds containing ester, thioester, amide, and alkyl side chains that are absent in EDR, neostigmine was used. After the grid search was performed to obtain

**Table 1.** Acetylcholinesterase Inhibitors Included in CoMFA/ $q^2$ -GRS Study

No.	Structure	IC <sub>50</sub>	Ref	No.	Structure	IC <sub>50</sub>	Ref
1	R =  , X = Br <sup>-</sup>	2.07 × 10 <sup>-3</sup>	28	35		5.34 × 10 <sup>-6</sup>	30
2	R =  , X = Br <sup>-</sup>	6.91 × 10 <sup>-4</sup>	28	36		6.59 × 10 <sup>-7</sup>	30
3	R =  , X = Br <sup>-</sup>	8.13 × 10 <sup>-3</sup>	28	37		5.97 × 10 <sup>-6</sup>	30
4	R =  , X = Br <sup>-</sup>	1.16 × 10 <sup>-2</sup>	28	38		7.54 × 10 <sup>-4</sup>	30
5	R =  , X = Br <sup>-</sup>	5.12 × 10 <sup>-3</sup>	28	39		6.91 × 10 <sup>-5</sup>	30
6	R = H, X = Cl <sup>-</sup>	1.76 × 10 <sup>-3</sup>	28				
7	R = CH <sub>3</sub> , X = Br <sup>-</sup>	1.73 × 10 <sup>-3</sup>	28	40		3.77 × 10 <sup>-6</sup>	30
8	R = C(CH <sub>3</sub> ) <sub>3</sub> , X = Br <sup>-</sup>	3.71 × 10 <sup>-3</sup>	28	41		5.97 × 10 <sup>-4</sup>	30
9	R = CF <sub>3</sub> , X = Br <sup>-</sup>	3.83 × 10 <sup>-3</sup>	28	42		2.39 × 10 <sup>-4</sup>	30
10		3.01 × 10 <sup>-3</sup>	29	43		3.45 × 10 <sup>-4</sup>	30
11		2.39 × 10 <sup>-3</sup>	29	44		1.22 × 10 <sup>-4</sup>	30
12		4.40 × 10 <sup>-4</sup>	29				
13		1.16 × 10 <sup>-3</sup>	29				
14		1.51 × 10 <sup>-3</sup>	29				
15		2.29 × 10 <sup>-4</sup>	29				
16		1.26 × 10 <sup>-4</sup>	29				
17		8.48 × 10 <sup>-5</sup>	29				
18		2.92 × 10 <sup>-3</sup>	29				
19		8.48 × 10 <sup>-4</sup>	29				
20		1.13 × 10 <sup>-3</sup>	29				
21		8.79 × 10 <sup>-5</sup>	29				
22		5.97 × 10 <sup>-4</sup>	29				
23		4.71 × 10 <sup>-4</sup>	29				
24		5.34 × 10 <sup>-4</sup>	29				
25		8.16 × 10 <sup>-4</sup>	29				
26		6.28 × 10 <sup>-4</sup>	29				
27		1.00 × 10 <sup>-3</sup>	29	45	R <sub>1</sub> = H; R <sub>2</sub> =	5.7 × 10 <sup>-8</sup>	31
28		1.92 × 10 <sup>-4</sup>	29	46	R <sub>1</sub> = H; R <sub>2</sub> =	1.10 × 10 <sup>-7</sup>	31
29		9.73 × 10 <sup>-7</sup>	30	47	R <sub>1</sub> = H; R <sub>2</sub> =	1.52 × 10 <sup>-7</sup>	31
30		1.41 × 10 <sup>-4</sup>	30	48	R <sub>1</sub> = H; R <sub>2</sub> =	4.60 × 10 <sup>-7</sup>	31
31		2.83 × 10 <sup>-6</sup>	30	49	R <sub>1</sub> = H; R <sub>2</sub> =	3.50 × 10 <sup>-7</sup>	31
32		3.11 × 10 <sup>-6</sup>	30	50	R <sub>1</sub> = CH <sub>3</sub> ; R <sub>2</sub> =	3.4 × 10 <sup>-8</sup>	31
33		3.01 × 10 <sup>-6</sup>	30	51	R <sub>1</sub> = H; R <sub>2</sub> =	1.7 × 10 <sup>-6</sup>	31
34		4.08 × 10 <sup>-5</sup>	30	52	R <sub>1</sub> = H; R <sub>2</sub> =	9.70 × 10 <sup>-7</sup>	31
				53	R <sub>1</sub> = H; R <sub>2</sub> =	1.8 × 10 <sup>-6</sup>	31
				54	R <sub>1</sub> = H; R <sub>2</sub> =	6.3 × 10 <sup>-6</sup>	31
				55	R <sub>1</sub> = H; R <sub>2</sub> =	4.0 × 10 <sup>-5</sup>	31

Table 1 (Continued)

No.	Structure	IC <sub>50</sub>	Ref	No.	Structure	IC <sub>50</sub>	Ref
56	R <sub>1</sub> = H; R <sub>2</sub> =	3.5 × 10 <sup>-5</sup>	31	THA		7.6 × 10 <sup>-8</sup>	31
EDR		7.8 × 10 <sup>-7</sup>	31	BW284C51		8.0 × 10 <sup>-9</sup>	31
Neostigmine		9.1 × 10 <sup>-8</sup>	31				

Table 2.  $q^2$  Values Obtained after Performing CoMFA/ $q^2$ -GRS with Different  $q^2$  Threshold Values for 60 AChE Inhibitors<sup>a</sup>

$q^2$ threshold	number of components								
	1	2	3	4	5	6	7	8	9
none <sup>b</sup>	0.475	0.551	<b>0.621</b>	0.624	0.624	0.636	0.630	0.638	0.653
0.1	0.472	0.566	0.650	0.658	0.666	0.688	<b>0.701</b>	0.698	0.677
0.2	0.472	0.565	0.648	0.657	0.663	0.686	<b>0.699</b>	0.697	0.682
0.3	0.497	0.572	0.644	0.662	0.658	0.680	<b>0.694</b>	0.686	0.680
0.4	0.498	0.583	0.644	0.683	0.670	0.714	<b>0.735</b>	0.718	0.715
0.5	0.518	0.598	0.683	0.698	0.701	0.718	<b>0.734</b>	0.737	0.720
0.6	0.430	0.588	0.676	0.706	0.707	0.703	0.707	0.724	<b>0.743</b>

<sup>a</sup> The numbers in bold represent the  $q^2$  values at the optimal number of components. <sup>b</sup> The results of conventional CoMFA for arbitrary orientation of aligned molecules.

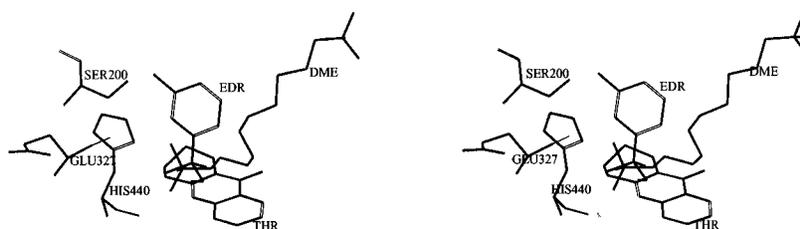


Figure 1. Superposition of THA, EDR, and DME in the active site of AChE.

the lowest energy conformer, neostigmine was rms-fitted to EDR using the quaternary nitrogen (N.4), the centroid of the phenyl ring, and the oxygen (O.3) adjacent to the phenyl ring. Since neostigmine may not bind to the enzyme in its lowest energy conformation, we have examined the orientation of the carbamoyl moiety of neostigmine in the active site of AChE. The carbonyl oxygen (O.2) was found to be facing toward the side chain of Ser 200 residue of the enzyme. Since the hydrolysis by acetylcholinesterase requires that the hydroxyl oxygen atom of the Ser 200 side chain attacks the carbonyl carbon of acetylcholine, we felt that it is not likely to have the conformation favoring the carbonyl oxygen (O.2) facing toward the Ser 200. To adjust the conformation, we have manually modified the torsional angle of the bond adjacent to the phenyl ring of neostigmine, and the structure was reoptimized, resulting in the final torsion angle of 57.9°. This torsional angle was used to manually modify the torsional angles of compounds with the carbonyl oxygen (O.2) facing Ser 200.

In order to align compounds **10–25**, we selected compound **15** as the template for this group of compounds. Compound **15** (fully extended conformation) was rms-fitted to neostigmine using the quaternary nitrogen (N.4), oxygen (O.3), and carbonyl oxygen (O.2). Compounds **10–14** and **16–25** (fully extended conformations) were rms-fitted to compound **15** using the quaternary amine (N.4), oxygen (O.3), and carbonyl oxygen (O.2) (if either oxygen (O.3) or carbonyl oxygen (O.2) is missing from the compound, the carbon before or after oxygen (O.3) was selected instead).

Compounds **26–28** and **29–44** were rms-fitted to EDR and neostigmine, respectively, the quaternary nitrogen (N.4), the centroid of the phenyl ring, and the oxygen (O.3) (for compounds **29–44**) or a nitrogen atom (for compounds **26–28**) adjacent to the phenyl ring.

To align compounds **46–56**, compound **45** was selected as the template for this group of compounds. After compound

**45** was rms-fitted to neostigmine using the quaternary nitrogen (N.4), oxygen (O.3), and carbonyl oxygen (O.2), it was subjected to field-fit minimization using neostigmine as the reference compound. The subsequent optimization with the field-fit option turned off yielded the final structure. Compounds **46–56** were then rms-fitted to compound **45** employing the same sets of atoms used to fit compound **45** to neostigmine.

BW284C51 was built by modifying DME and then subjected to field-fit geometry optimization. The subsequent structure optimization with the field-fit option turned off yielded the final structure of BW284C51.

## Results

**CoMFA/ $q^2$ -GRS of AChE Inhibitors.** Sixty inhibitors of AChE, structurally aligned on the basis of the knowledge of crystallographic coordinates of several enzyme/inhibitor complexes, were subjected to CoMFA/ $q^2$ -GRS, and the results are presented in Tables 3 and 4. The predictability of the CoMFA model was initially assessed by conventional CoMFA (cf. Tables 2 and 3). The  $q^2$ -GRS routine was then applied to optimize the initial CoMFA model. Various  $q^2$  thresholds (0.1–0.6) were used to isolate the regions of the lattice surrounding the aligned molecules where the change in the field values correlated strongly with biological activity. This procedure can be interpreted as elimination of the irrelevant variables in the PLS analysis. As the  $q^2$  threshold increases from 0.1 to 0.6, the  $q^2$  values for the optimal number of components increase, reaching a maximum at 0.4- and 0.5 $q^2$  threshold, and then decrease again (Table 2). Correspondingly, the SDEP values

**Table 3.** Standard Errors of Prediction Obtained after Performing CoMFA/ $q^2$ -GRS with Different  $q^2$  Threshold Values for 60 AChE Inhibitors<sup>a</sup>

$q^2$ threshold	number of components								
	1	2	3	4	5	6	7	8	9
none <sup>b</sup>	1.190	1.110	<b>1.029</b>	1.033	1.043	1.037	1.054	1.054	1.041
0.1	1.193	1.090	0.988	0.986	0.983	0.959	<b>0.949</b>	0.963	1.005
0.2	1.194	1.092	0.991	0.987	0.987	0.963	<b>0.951</b>	0.963	0.996
0.3	1.165	1.083	0.997	0.980	0.996	0.972	<b>0.960</b>	0.981	1.000
0.4	1.163	1.069	0.998	0.949	0.977	0.918	<b>0.893</b>	0.930	0.944
0.5	1.140	1.050	0.941	0.926	0.931	0.912	<b>0.894</b>	0.898	0.936
0.6	1.240	1.063	0.951	0.914	0.922	0.935	0.938	0.919	<b>0.897</b>

<sup>a</sup> The numbers in bold represent the standard errors of prediction at the optimal number of components. <sup>b</sup> The results of conventional CoMFA for arbitrary orientation of aligned molecules.

**Table 4.** Summary of CoMFA/ $q^2$ -GRS Results

	conventional CoMFA	CoMFA/ $q^2$ -GRS	
$q^2$ threshold	none	0.4	0.5
minimum $\sigma$	0	0	0
number of small boxes	none	18	11
number of lattice points	2652	3150	1925
optimal number of components	3	7	7
$q^2$	0.621	0.735	0.734
standard error of prediction	1.029	0.893	0.894
standard error of estimate	0.739	0.478	0.451
$r^2$	0.804	0.924	0.932
$F$ values	76.805 <sup>a</sup>	90.259 <sup>b</sup>	102.337 <sup>b</sup>
probability of $R^2 = 0$	0 <sup>a</sup>	0 <sup>b</sup>	0 <sup>b</sup>
relative contributions			
steric	0.772	0.683	0.670
electrostatic	0.228	0.317	0.330

<sup>a</sup>  $n_1 = 3$ ,  $n_2 = 56$ . <sup>b</sup>  $n_1 = 7$ ,  $n_2 = 52$ .

reached the minimum at 0.4- and 0.5 $q^2$  threshold for ONC (Table 3).

Since the values of both  $q^2$  and SDEP for both 0.4- and 0.5 $q^2$  thresholds were very close to each other (cf. Tables 2 and 3), we have examined both models. The results obtained from CoMFA/ $q^2$ -GRS at 0.4- and 0.5 $q^2$  thresholds are summarized in Table 4. Non-cross-validated CoMFA calculations showed that the 0.5 $q^2$  threshold exhibits slightly better overall statistics compared to that with the 0.4 $q^2$  threshold. Table 4 also presents the number of lattice points for the two different CoMFA runs; obviously, a significant number of lattice points are excluded from the analysis as the  $q^2$  threshold value increases (3150 vs 1925 lattice points at 0.4- and 0.5 $q^2$  thresholds, respectively). This suggests that 1225 additional lattice points (i.e., 2450 variables) present in the 0.4 $q^2$  threshold model most likely do not contribute to the predictability of the CoMFA model. On the basis of the above considerations, we have selected a 0.5 $q^2$  threshold at seven principal components as the final CoMFA model. The actual, calculated, and residual activities of the 0.5 $q^2$  threshold model are shown in Table 5. The plot of actual vs calculated activities of this model is shown in Figure 2.

**CoMFA Fields.** The CoMFA steric and electrostatic fields obtained using an  $sp^3$  carbon with +1 charge and the amino acids comprising the active site are shown in Figures 3–6. The field values were calculated by multiplying the  $\beta$ -coefficient and standard deviation of columns in the QSAR table (stdev\*coeff). The green (sterically favorable) and yellow (sterically unfavorable) contours shown in Figure 3 represent 80% and 20% level contributions, respectively. A part of the structure of

the active compound, neostigmine, protrudes into the green (sterically favorable) contour region whereas the structure of less active compound **19** does not occupy this region at all (Figure 3). Other less active compounds either partially occupy sterically unfavorable areas or do not occupy sterically favorable regions (not shown). Thus, in general CoMFA steric fields explain differences in the activity of inhibitors. Active site residues (defined as all residues that have at least one heavy atom within a union of 7 Å radius spheres centered on all heavy atoms of EDR) and the CoMFA steric fields are shown in Figure 4. The contours are compatible with the environment of the active site as indicated by the absence of amino acids near the green contours (sterically favorable) and the presence of His 440, Ser 226, Glu 199, Ser 200, and Phe 288 residues near yellow contours (sterically unfavorable; cf. Figure 4).

Electrostatic contours are shown in Figure 5. The blue (positive charge favored region) and red (negative charge favored region) contours represent 80% and 20% level contribution, respectively. The contours reflect the physicochemical environment around the structure of active compound and are able to differentiate the active compound, neostigmine, from the less active compound **19** (Figure 5); the carbonyl oxygen and carbon of neostigmine are surrounded by the red and blue contours, respectively. The compound **19**, however, lacks this functional group and does not effectively contribute to the electrostatic portion of the CoMFA equation. Active site residues and the CoMFA electrostatic contours are shown in Figure 6. The positive charge favorable regions (shown as blue contours in Figure 6) are surrounded by Glu 199, Ser 200, Ser 226, and Glu 327, which can accommodate the increase in the positive charge around this region. The negative charge favorable regions (shown as red contours in Figure 6) are surrounded by Phe 288, Phe 290, Phe 330, and Phe 331 residues.

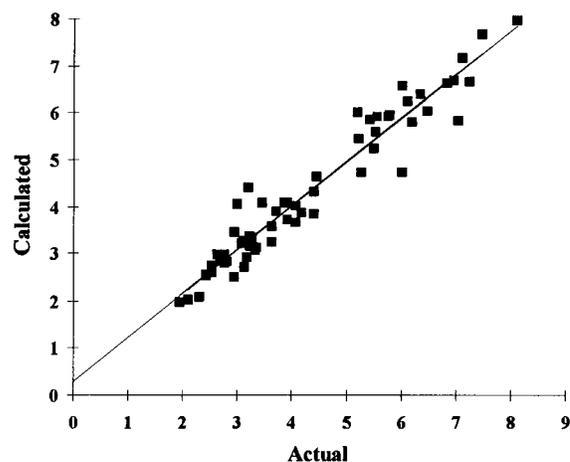
## Discussion

The most crucial aspect of this work is the use of X-ray crystallographic information about several AChE-inhibitor complexes in order to generate the structure-based alignment of 60 inhibitors of the enzyme. We should emphasize that without this knowledge any unbiased structural alignment of so diverse inhibitors would be simply impossible. Furthermore, in some cases this information helped to adjust the (low-energy) internal geometry of the inhibitors (e.g., neostigmine) to make sure that this geometry was compatible with the active site geometry.

**Table 5.** Results of CoMFA-Based QSAR for Compounds in the Training Set

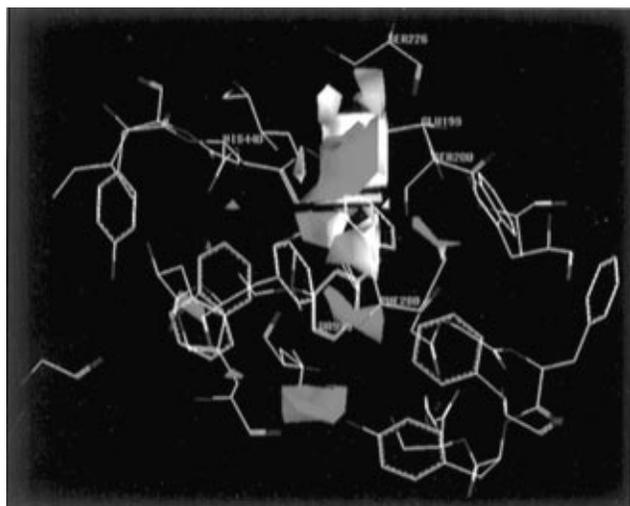
compound	actual	calculated	residual
29	6.012	4.735	1.277
neostigmine	7.041	5.842	1.199
26	3.202	4.398	-1.196
27	3.000	4.042	-1.042
54	5.201	6.007	-0.806
43	3.462	4.063	-0.601
45	7.244	6.671	0.573
52	6.013	6.565	-0.552
34	4.389	3.847	0.542
35	5.273	4.732	0.541
20	2.947	3.447	-0.500
40	5.424	5.869	-0.445
13	2.936	2.504	0.432
49	6.456	6.039	0.417
38	3.123	2.716	0.407
21	4.056	3.670	0.386
42	3.622	3.243	0.379
36	6.181	5.813	0.368
31	5.549	5.905	-0.356
11	2.622	2.961	-0.339
39	4.161	3.861	0.300
32	5.507	5.227	0.280
46	6.959	6.704	0.255
23	3.327	3.078	0.249
2	3.161	2.916	0.245
12	3.357	3.115	0.242
30	3.850	4.078	-0.228
50	7.469	7.685	-0.216
37	5.224	5.437	-0.213
6	2.754	2.963	-0.209
18	2.535	2.739	-0.204
5	2.291	2.088	0.203
28	3.717	3.907	-0.190
56	4.456	4.644	-0.188
25	3.088	3.274	-0.186
44	3.912	3.727	0.185
53	5.745	5.928	-0.183
51	5.770	5.953	-0.183
47	6.818	6.638	0.180
16	3.900	4.071	-0.171
1	2.684	2.843	-0.159
EDR	6.108	6.253	-0.145
41	3.224	3.356	-0.132
9	2.417	2.549	-0.132
19	3.072	3.199	-0.127
BW284C51	8.097	7.975	0.122
8	2.431	2.523	-0.092
15	3.640	3.558	0.082
22	3.224	3.144	0.080
55	4.398	4.321	0.077
17	4.072	4.002	0.070
24	3.272	3.336	-0.064
10	2.521	2.582	-0.061
48	6.337	6.397	-0.060
33	5.521	5.577	-0.056
3	2.090	2.036	0.054
THA	7.119	7.158	-0.039
4	1.936	1.973	-0.037
7	2.762	2.790	-0.028
14	2.821	2.827	-0.006

The quantitative results of our CoMFA analysis, i.e., high  $q^2$  value, benefited from our recent improvements to standard CoMFA methodology.<sup>24,25</sup> The most important practical aspect of this approach as compared to conventional CoMFA is due to the fact that the  $q^2$  values, obtained as a result of conventional CoMFA, are sensitive to the overall orientation of rigidly aligned molecules on a computer terminal.<sup>24</sup> As we discussed earlier, conventional CoMFA may often generate a reasonably high  $q^2$  value (e.g., in this study, cf. Table 3). However, due to the problem discussed above, the results of conventional CoMFA should not be presented

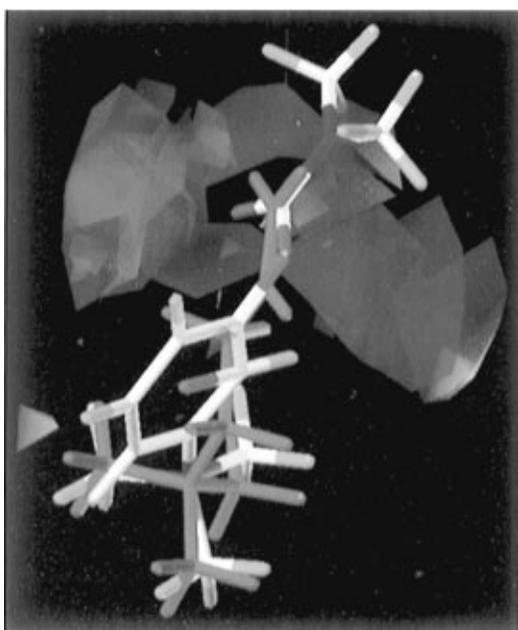
**Figure 2.** Comparison of actual vs calculated  $-\log IC_{50}$  (based on the data of Table 5).**Figure 3.** The CoMFA steric stdev\*coeff contour plot. Green regions represent a contribution level of 80%, i.e., sterically favored areas. Yellow regions represent a contribution level of 20%, i.e., sterically disfavored areas.

as a single value but rather as a range of values for different, systematically varied orientations of rigidly aligned molecules on the computer terminal.<sup>24</sup> Our  $q^2$ -GRS routine<sup>24,25</sup> deals with this problem effectively and generates a reproducible, high  $q^2$  value (Tables 3 and 4). This value is ca.  $0.1q^2$  unit higher than the one obtained with conventional CoMFA (Table 3), and the statistics is also better (cf. Table 4). Therefore, we suggest that the  $q^2$ -GRS method should be used routinely in CoMFA studies. This viewpoint is confirmed by recent successful application of our method by others.<sup>33</sup>

Since the structure of AChE was known, it was quite interesting to compare the qualitative results of CoMFA, i.e. location of steric and electrostatic contours, with the chemical and geometrical properties of the active site of the enzyme. It is important to note that, in general, such comparisons should be attempted very carefully. In fact, the authors of the original CoMFA paper cautioned strongly against "the temptation to overin-

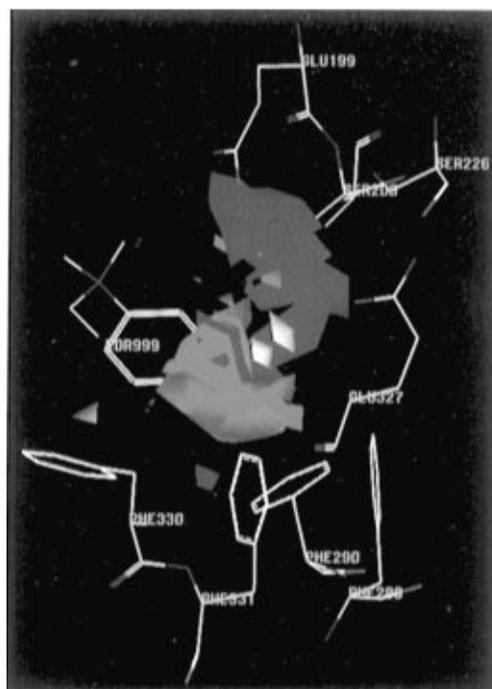


**Figure 4.** Superposition of the CoMFA steric stdev\*coeff contour plot and active site residues.



**Figure 5.** The CoMFA electrostatic stdev\*coeff contour plot. Blue regions represent a contribution level of 80%, i.e., positive charge favored areas. Red regions represent a contribution level of 20%, i.e., negative charge favored areas.

interpret the 'contour coefficient' maps, for example, as 'receptor maps'.<sup>4</sup> Nevertheless, when the alignment is based on the receptor structure, one might expect certain correlation. Thus, we have analyzed the contours shown in Figures 3 and 4 in terms of their correspondence to the steric and electrostatic environment in the AChE active site. In general, we find that the location of the contour coefficient maps is consistent with what is known about the active site of AChE. More specifically, we find that sterically favorable regions around training set molecules (shown in green in Figure 4) occupy cavities in the AChE active site whereas sterically unfavorable regions (shown in yellow in Figure 4) overlap with enzyme atoms. This correlation is less obvious in the case of electrostatic fields, although we do find that positive charge favorable fields (blue contours in Figure 4) occupy areas in the vicinity of residues that can accommodate positive charge (Glu 199, Ser 200, Ser 226, and Glu 327). In the case of negative charge



**Figure 6.** Superposition of the CoMFA electrostatic stdev\*coeff contour plot and active site residues.

favorable fields (shown as red contours in Figure 6), interpretation is much more difficult, and we could not explain the presence of Phe 288, Phe 290, Phe 330, and Phe 331 near this region. One possibility, of course, is that the orientation of the carbonyl oxygen of neostigmine might be wrong since we did not have the experimental knowledge of the orientation of this group when we started this project. These considerations suggest that when the alignment is based on the geometry of the active site like in this and other<sup>12</sup> publications, the CoMFA steric and electrostatic fields may correspond to the steric and electrostatic environments of the active site. However, the degree of this correspondence depends on the chemical structure of training set.

In conclusion, the synergistic application of structure-based alignment and CoMFA provides exciting avenues for future drug design studies. The 3D QSAR model proposed in this study for AChE inhibitors can be used directly to predict the activity of newly designed compounds. Thus, a new molecule can be first aligned with the most structurally similar template compounds (i.e., one of the 60 inhibitors, considered in this paper), its orientation and conformation can be then corrected on the basis of the geometry of the enzyme active site, and its activity can be predicted using our QSAR equation. More generally, the approach considered in this paper can also be applied in combination with docking studies. As mentioned above, docking algorithms may generate a reasonable orientation of receptor ligands in the active site but generally fail to accurately rank the activity or affinity of ligands due to the inaccuracy of the docking energy function.<sup>9</sup> If known ligands of the given receptor system are first subjected to CoMFA study, and a CoMFA model is generated, then the latter can be used to predict the activity of compounds proposed by docking algorithm. This idea is currently under investigation in our laboratory.

**Note.** The atomic coordinates and partial atomic charges of all structurally aligned AChE inhibitors as

well as the script to run  $q^2$ -GRS routine are available from the authors upon request (jin@gibbs.oit.unc.edu).

**Acknowledgment.** This work was supported in part by the Hoechst Celanese Corporation/UNC Research Partnership, PHS Grant MH 40537 and Center Grants HD03310 and MH33127. The authors thank Dr. Joel L. Sussman for supplying crystal coordinates for the enzyme/inhibitor complexes and Tripos, Inc. for the software grant. M.L.S. acknowledges the support from "Programa de Nuevas Tecnologías" CONICIT-Venezuela and CDCH of the Universidad Central de Venezuela Training Grant, and J. Bier appreciates the support from Wellcome Foundation.

## References

- Martin, Y. C.; Danaher, E. B.; May, C. S.; Weininger, D.; Van Drie, J. H. Strategies in Drug Design Based on 3D-Structures of Ligands. In *QSAR: Quantitative Structure-Activity Relationships in Drug Design*; Fauchere, J. L., Ed.; Alan R. Liss: New York, 1989; pp 177-181.
- Marshall, G. R.; Barry, C. D.; Bosshard, H. E.; Dammkoehler, R. A.; Dunn, D. A. The Conformational Parameter in Drug Design: The Active Analog Approach. In *Computer-Assisted Drug Design*; Olsen, E. C., Christoffersen, R. E., Eds.; ACS Symposium Series No. 112; American Chemical Society: Washington, DC, 1979; pp 205-226.
- Martin, Y. C. Overview of Concepts and Methods in Computer-Assisted Rational Drug Design. *Methods Enzymol.* **1991**, *203*, 587-613.
- Cramer, R. D. III; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959-5967.
- Boyd, D. B. Successes of Computer-Assisted Molecular Design. In *Reviews in Computational Chemistry I*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1990; pp 355-372.
- Bugg, C. E.; Carson, W. M.; Montgomery, J. A. Drugs by Design. *Sci. Am.* **1993**, *269*, 92-98.
- Martin, Y. C.; Bures, M. G.; Danaher, E. A.; DeLazzer, J.; Lico, I.; Pavlik, P. A. A Fast New Approach to Pharmacophore Mapping and Its Application to Dopaminergic and Benzodiazepine Agonists. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 83-102.
- Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849-957.
- Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automated Docking with Grid-Based Energy Evaluation. *J. Comput. Chem.* **1992**, *13*, 505-524.
- Ferguson, D. M.; Radmer, R. J.; Kollman, P. A. Determination of the Relative Binding Free Energies of Peptide Inhibitors to the HIV-1 Protease. *J. Med. Chem.* **1991**, *34*, 2654-2659.
- Tropsha, A.; Hermans, J. Application of Free Energy Simulations to the Binding of a Transition-State-Analog Inhibitors to HIV Protease. *Protein Eng* **1992**, *5*, 29-33.
- Waller, C. L.; Oprea, T. I.; Giolitti, A.; Marshall, G. R. Three-Dimensional QSAR of Human Immunodeficiency Virus (I) Protease Inhibitors. 1. A CoMFA Study Employing Experimentally-Determined Alignment Rules. *J. Med. Chem.* **1993**, *36*, 4152-4160.
- Taylor, P. Anticholinesterase Agents. In *The Pharmacological Basis of Therapeutics*; Gilman, A. G., Rall, T. W., Nies, A. S., Taylor, P., Eds.; Pergamon: New York, 1990; pp 131-149.
- Hakansson, L. Mechanism of Action of Cholinesterase Inhibitors in Alzheimer's Disease. *Acta Neurol. Scand. Suppl.* **1993**, *149*, 7-9.
- Sirvio, J.; Riekkinen, P. J. Brain and Cerebrospinal Fluid Cholinesterases in Alzheimer's Disease, Parkinson's Disease and Aging. A Critical Review of Clinical and Experimental Studies. *J. Neural Transm.: Parkinson's Dis. Dementia Sect.* **1992**, *4*, 337-358.
- Linton, D. M.; Philcox, D. Myasthenia Gravis. *DM, Dis. Mon.* **1990**, *36*, 593-637.
- Villalobos, A.; Blake, J. F.; Biggers, C. K.; Butler, T. W.; Chapin, D. S.; Chen, Y. L.; Ives, J. L.; Jones, S. B.; Liston, D. R.; Nagel, A. A. Novel Benzoisoxazole Derivatives as Potent and Selective Inhibitors of Acetylcholinesterase. *J. Med. Chem.* **1994**, *37*, 2721-2734.
- Ishihara, Y.; Hirai, K.; Miyamoto, M.; Goto, G. Central Cholinergic Agents. 6. Synthesis and Evaluation of 3-[1-(Phenylmethyl)-4-piperidinyl]-1-(2,3,4,5-tetrahydro-1H-1-benzazepin-8-yl)-1-propanones and Their Analogs as Central Selective Acetylcholinesterase Inhibitors. *J. Med. Chem.* **1994**, *37*, 2292-2299.
- Chen, Y. L.; Liston, D.; Nielsen, J.; Chapin, D.; Dunaiskis, A.; Hedberg, K.; Ives, J.; Johnson, J., Jr.; Jones, S. Syntheses and Anticholinesterase Activity of Tetrahydrobenzazepine Carbamates. *J. Med. Chem.* **1994**, *37*, 1996-2000.
- Vidaluc, J. L.; Calmel, F.; Bigg, D.; Carilla, E.; Stenger, A.; Chopin, P.; Briley, M. Novel [2-(4-Piperidinyl) ethyl](thio)ureas: Synthesis and Antiacetylcholinesterase Activity. *J. Med. Chem.* **1994**, *37*, 689-695.
- Sasho, S.; Obase, H.; Ichikawa, S.; Kitazawa, T.; Nonaka, H.; Yoshizaki, R.; Ishii, A.; Shuto, K. Synthesis of 2-Imidazolidinylidenepranedinitrile Derivatives as Stimulators of Gastrointestinal Motility. *J. Med. Chem.* **1993**, *36*, 572-579.
- Sussman, J. L.; Harel, M.; Frolow, F.; Oefner, C.; Goldman, A.; Toker, L.; Silman, I. Atomic Structure of Acetylcholinesterase from *Torpedo californica*: A Prototypic Acetylcholine-Binding Protein. *Science* **1991**, *253*, 8872-8879.
- Harel, M.; Schalk, I.; Ehret-Sabatier, L.; Bouet, F.; Goeldner, M.; Hirth, C.; Axelsen, P. H.; Silman, I.; Sussman, J. L. Quaternary ligand binding to aromatic residues in the active-site gorge of acetylcholinesterase. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 9031-9035.
- Cho, S. J.; Tropsha, A. Cross-Validated  $R^2$  Guided Region Selection for Comparative Molecular Field Analysis (CoMFA): A Simple Method to Achieve Consistent Results. *J. Med. Chem.* **1995**, *38*, 1060-1066.
- Cho, J.; Tropsha, A.; Suffness, M.; Lee, K.-H. Antitumor Agents. 163. Three-Dimensional QSAR Study of 4'-O-Demethylpiperodophylloxin Analogs Using the CoMFA/ $q^2$ -GRS Approach. *J. Med. Chem.* **1996**, *39*, 1383-1395.
- The program Sybyl 6.0 is available from Tripos Associates, 1699 South Hanley Road, St Louis, MO 63144.
- Clark, M.; Cramer, R. D., III; Van Opdenbosch, N. Validation of the General Purpose Tripos 5.2 Force Field. *J. Comput. Chem.* **1989**, *10*, 982-1012.
- Lee, B. H.; Stelly, T. C.; Colucci, W. J.; Garcia, G.; Gandour, R. D.; Quinn, D. M. Inhibition of Acetylcholinesterase by Hemicholiniums, Conformationally Constrained Choline Analogues. Evaluation of Aryl and Alkyl Substituents. Comparisons with Choline and (3-Hydroxyphenyl)trimethylammonium. *Chem. Res. Toxicol.* **1992**, *5*, 411-418.
- Thanei-Wyss, P.; Waser, P. G. Interaction of Quaternary Ammonium Compounds with Acetylcholinesterase: Characterization of the Active Site. *Eur. J. Pharmacol.* **1989**, *172*, 165-173.
- Coccolas, G. H.; Cranford, J. G.; Choi, H. S. Y. Studies on the Conformational Requirements of Substrate and Inhibitor on Acetylcholinesterase. *J. Med. Chem.* **1974**, *17*, 938-943.
- Atack, J. R.; Yu, Q. S.; Soncrant, T. T.; Brossi, A.; Rapoport, S. I. Comparative Inhibitory Effects of Various Physostigmine Analogs Against Acetyl- and Butyrylcholinesterases. *J. Pharmacol. Exp. Ther.* **1989**, *249*, 194-202.
- Cheng, Y.-C.; Prusoff, W. H. Relationship between the inhibition constant ( $K_i$ ) and the concentration of the inhibitor which causes 50 percent inhibition ( $IC_{50}$ ). *Biochem. Pharmacol.* **1973**, *22*, 3099-3108.
- Jones, J. P.; He, M.; Trager, W. F.; Rettie, A. E. Three-Dimensional Quantitative Structure-Activity Relationship for Inhibitors of Cytochrome P4502C9. *Drug Metab. Dispos.* **1996**, *24*, 1-6.

JM950771R