Journal of Computer-Aided Molecular Design, **15:** 247–258, 2001.
KLUWER/ESCOM

An improved nicotinic pharmacophore and a stereoselective CoMFA-model for nicotinic agonists acting at the central nicotinic acetylcholine receptors labelled by [^{3}H]-*N*-methylcarbamylcholine

Janne E. Tønder[a], Preben H. Olesen[a], John Bondo Hansen[a], Mikael Begtrup[b] & Ingrid Pettersson[a,*]

[a]*Health Care Discovery, Novo Nordisk A/S, Novo Nordisk Park, DK-2760 Måløv, Denmark;* [b]*Department of Medicinal Chemistry, Royal Danish School of Pharmacy, Universitetsparken 2, DK-2100 Copenhagen, Denmark*

Received 15 March 2000; accepted 6 November 2000

Key words: CoMFA model, 3D-QSAR, nicotinic agonists, pharmacophore model

Summary

A study of a series of compounds with agonistic effect at the α4β2 nicotinic acetylcholine receptors resulted in an improved pharmacophore model as well as a CoMFA model. The pharmacophore was composed of three pharmacophoric elements: (1) a site point (**a**) corresponding to a protonated nitrogen atom, (2) a site point (**b**) corresponding to an electronegative atom capable of forming a hydrogen bond, and (3) the centre of a heteroaromatic ring or a C=O bond (**c**). The pharmacophoric elements were related by the following parameters: (**a–b**) 7.3–8.0 Å, (**a–c**) 6.5–7.4 Å, and the angle between the two distance vectors (Δ**bac**) 30.4–35.8°. In addition to this, a stereoselective CoMFA model was developed, which showed good predictability even for compound classes not present in the training set.

Introduction

The nicotinic pharmacophore has been developing since 1970, where Beers and Reich were the first to propose that the specific binding of nicotinic agonists to the nicotinic acetylcholine receptors (nAChRs) is mediated by (1) a coulombic interaction between a positively charged nitrogen atom and (2) a hydrogen bond forming from a carbonyl oxygen atom (in acetylcholine) or a nitrogen atom (in (*S*)-nicotine) (Figure 1) [1]. The pharmacophores published since then all have a cationic centre, an electronegative atom capable of forming a hydrogen bond, and the centre of the pyridine ring in (*S*)-nicotine as pharmacophoric elements [2, 3]. The latest model also has a lipophilic region close to the 3′,4′ carbon atoms of the pyrrolidine ring in (*S*)-nicotine [3].

The stereoselectivity of the central nAChRs has been well known due to the fact that (*S*)-nicotine is 10–100 times more potent than (*R*)-nicotine [4, 5]. By combining the van der Waals volumes of (+)-anatoxin-a, (−)-cytisine, and isoarecolone methiodide (Figures 1 and 2), Hacksell and Mellin described a receptor excluded volume for the central nAChRs, and by incorporation of (−)-anatoxin-a (Figure 2) which has a much lower affinity for central nAChRs, they described a part of the receptor essential volume [6].

The stereoselectivity seen for (+)- and (−)-anatoxin-a is also present within the enantiomeric pair of ligands UB-165 and *ent*-5 [7]. In this case there is a difference in affinity of a factor 26 between the two enantiomers [7]. UB-165 being the most potent enantiomer, and the one that possesses the same configuration as (+)-anatoxin-a (Figure 2).

In the later years much attention has been devoted to the internitrogen distance (N^+–N distance), where suggestions have ranged from 4.5 Å to more than 6 Å [2, 9–12]. Very recently however, it has been argued, that the internitrogen distance has no influence on nicotinic affinity, whereas a distance of 7.0–8.0 Å

*To whom correspondence should be addressed.
E-mail: inpe@novo.dk

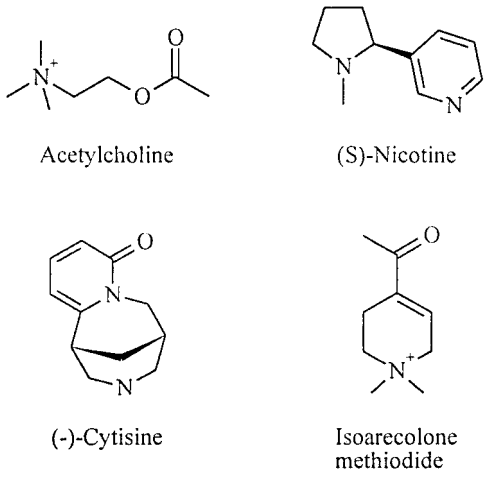


Figure 1. Compounds acting at the nAChRs.

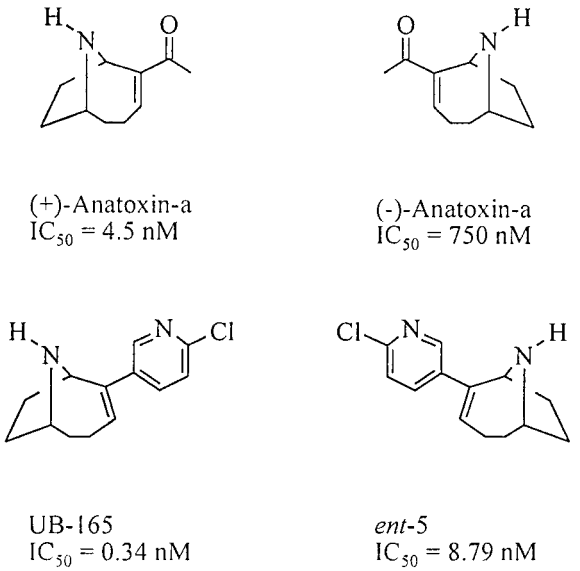


Figure 2. The IC_{50} values for (+)- and (−)-anatoxin-a are determined in rat cortex using [^{3}H]acetylcholine as radioligand [8], whereas the IC_{50} values for UB-165 and *ent*-5 are determined in rat P2 brain membranes using [^{3}H]nicotine as radioligand [7].

between two site points (**a**,**b**) complementary to the protonated nitrogen atom and the hydrogen bond acceptor is crucial [13]. Additionally, the biologically active conformations of (*R*)- and (*S*)-epibatidine were suggested, making it possible to superimpose new nicotinic ligands on (*R*)- and (*S*)-epibatidine [13].

In 1994 a CoMFA study of various protonated nicotinic ligands all containing a pyridine ring was published [14]. In this study the compounds were aligned using the protonated nitrogen atom, the centroid of the pyridine ring, and the pyridine nitrogen atom. The resulting CoMFA maps revealed, that increased electron density was favoured near the pyridine ring and decreased electron density was favoured near the protonated nitrogen atom. Additionally, favourable steric interactions were found around the azabicyclic ring of epibatidine.

Apart from this, only two other 3D-QSAR studies regarding compounds acting at nicotinic receptors are published. These studies are of insecticides structurally related to imidacloprid, which act at the α7 nAChRs (measured by α-bungarotoxin). As this receptor subtype has a low (micromolar) affinity for (*S*)-nicotine [15], the results are not likely to be comparable [16, 17].

From our previous work [13, 18] a large set of compounds with affinity for the nicotinic acetylcholine receptors (nAChRs) was available (Figure 3). The compounds were nicotinic agonists binding to a population of the nicotinic receptors in rat cortex labelled by [^{3}H]methylcarbamylcholine. This receptor population is generally believed to consist mainly (90%) of α4β2 nAChRs [15]. Hence, the conclusions appearing in this report are only to be considered relevant for agonists at the α4β2 nAChRs.

The purpose of the work presented here was to develop an improved nicotinic pharmacophore derived solely from agonists at the α4β2 nAChRs, as well as a new and improved CoMFA-model which is able to describe the 3D-QSAR for agonists at these receptors.

Methods

Computational studies

The protonated compounds were subjected to a Monte Carlo conformational search in MacroModel version 6.0 [19], using the MMFF force field [20–23] with a solvation model for water [24, 25]. The search was continued until all conformations had been found at least 5 times. The low energy conformations of the compounds were then imported to Sybyl 6.5 [26], where the remaining studies were conducted. Gasteiger-Hückel charges were assigned to the compounds, site points added (2.9 Å away from the corresponding nitrogen or oxygen atoms in the direction of the lone pairs), and centroids defined (the centre of a heteroaromatic ring or the centre of a C=O bond).

The high-affinity compound **1d** was found to possess only one energy minimum [13], therefore all low energy conformations of the compounds were aligned

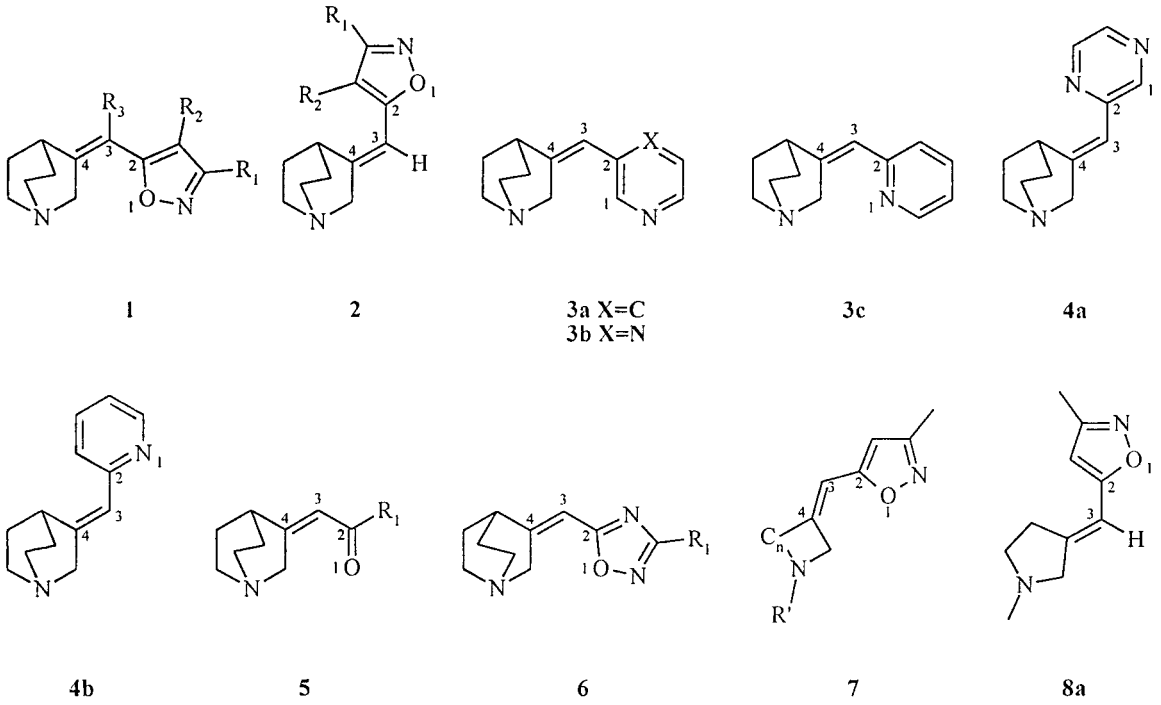


Figure 3. The numbered atoms were used for measuring the torsional angle.

with the global energy minimum conformation of **1d** using the pharmacophoric elements **a**, **b**, and **c** (Figure 5) (*vide infra*). Pharmacophore element **a**, is a site point in the direction of the N-H bond and 2.9 Å from the nitrogen atom. Pharmacophore element **b**, is a site point in the hydrogen bond accepting direction and 2.9 Å from the hydrogen bond accepting atom. Pharmacophore element **c**, is the centre of a heteroaromatic ring or a C=O bond. On the basis of a visual inspection the conformations with the best alignment were chosen for the further studies. For the CoMFA study the site points and centroids were deleted and the compounds were converted to their non-protonated equivalents by removal of the proton defining site point **a**, whereupon Gasteiger-Hückel charges were assigned. In the case of compounds with flexible alkyl-substituents (i.e., Et, *n*-Pr, *i*-Pr, MeOMe) the conformer of lowest energy for the alkyl-substituent was chosen throughout the series.

Some of the compounds were symmetrical and thus displayed two global energy minima. For these compounds the conformation, which obeyed the stereoselectivity demands previously discussed for anatoxin-a and UB-165, was chosen. Thus, the activity of a symmetrical compound was related to the sterically allowed conformer.

(*R*)- and (*S*)-epibatidine were used in the conformations with $\tau = \pm 20°$ (*vide infra*).

CoMFA calculations were performed using the QSAR module of Sybyl 6.5 [26] and with the following settings: The grid in which the molecules were embedded was regularly spaced (1 Å) with dimensions of 20 × 20 × 20 Å (the molecules did not extend 12 × 12 × 12 Å). Within the Tripos standard field steric and electrostatic interaction energies were calculated using a CH_3^+ probe. The dielectric constant was set to distance-dependant (1/r), and the standard setting of 30 kcal/mol as energetic cut off value with no electrostatic interactions at steric bad contacts was used. The transition was set to Abrupt.

Regression analyses were done using the Sybyl implementation of the PLS algorithm, initially with cross-validation (leave one out) and ten components. The optimum number of components to be used in the conventional PLS analyses was chosen on the basis of the highest cross-validated r^2 (r^2_{cv}) value and the minimum number of components. The steric and electrostatic field columns were weighted using the CoMFA standard scaling option. In this method a field is considered as a whole and every CoMFA variable is affected by the overall field mean and standard deviation. In order to reduce the computation time,

the SAMPLS algorithm was used. Therefore, only r^2_{cv}- and s_{cv}-values were provided from the cross-validation.

Final PLS (non-cross-validated) models were then derived using the optimal number of components identified, and to improve the signal-to-noise ratio, a 2.0 kcal/mol energy column filter was applied. In an attempt to reduce the noise, region focusing (standard deviation, weight 0.3) was used to derive the models designated B.

To test and select between the different models, the activities of a set of compounds not used in the model generation were predicted. The test set consisted of 13 compounds whose affinities spanned 2 orders of magnitude covering a range of 1.4 log units (from 5.63 to 7.04).

Binding studies

[^{3}H]-Methylcarbamylcholine binding in rat cerebral cortex for compounds **1**–**8**, **11**–**12**, (*R*/*S*)-epibatidine, (*S*)-nicotine, and ABT 418 was performed as previously described [13]. The results are all triplicates. The binding data for the literature compounds **9** [15], **10** [27], (*R*/*S*)-epiboxidine [28], and UB-165 [7] were transformed into IC_{50} values corresponding to the IC_{50} values obtained for (*S*)-nicotine (6.6 nM) or (*R*/*S*)-epibatidine (0.30 nM) in our own assay. The transformation was performed as shown in Equation 1.

$$IC_{50} = \frac{K_i(\text{Compound})}{K_i(\text{Nicotine})} \times 6.6\ \text{nM}$$
$$\text{or} \qquad (1)$$
$$IC_{50} = \frac{K_i(\text{Compound})}{K_i(\text{Epibatidine})} \times 0.30\ \text{nM}$$

In the calculation of the p(IC_{50}) values, the average of the two results (Table 1) was used. Epibatidine was tested as the racemate, but as the two enantiomers are equipotent when tested separately [9, 29], the activity of the racemate (0.30 nM) is assigned to both enantiomers.

Results and discussion

Pharmacophore study

From the collection of compounds in Figure 3 some compounds were left out due to a lowering in affinity that could not be ascribed to a change in low energy conformations. This was the case for compounds in series **1**, **2**, and **3** where an increase in the size of the R_1 substituent led to a lowered affinity. Likewise when the R_3 substituent in series **1** was a CH_3 or F group or an oxadiazole ring was present (series **6**) a lowered affinity was observed that could not be ascribed to a change in low energy conformations.

The remaining compounds were selected for the pharmacophore studies (Table 2). In addition to the compounds from our own series, a number of compounds from the literature were selected. These were (*R*)- and (*S*)-epibatidine, ABT 418, an anatoxin-a analogue **9** [15], epiboxidine [28], UB-165 [7], and a rigid nicotine analogue **10** (Figure 4) [27].

In order to be able to align the compounds for the CoMFA study by using elements relevant for the biological activity, a search for a third pharmacophoric element was initiated. Taking into account that the (*R*)- and (*S*)-enantiomers of epibatidine are equipotent [9, 29] the low energy conformations that superimposed with **1d** ($\tau = \pm 7°$) were manually fitted to each other by ensuring an overlap of the site points **a** and **b**, and then rotating around the single bond connecting the two ring systems (see Figure 4 for a definition of the torsional angle). The conformations with $\tau = \pm 20°$ ($\Delta E = 0.6$ kcal/mol) provided maximum overlap of the carbon skeletons, the two site points (**a**,**b**), plus the planes of the pyridine rings.

Following this observation, the distance (in the supposed bioactive conformation found by superimposing with **1d**) between site point **a** and the centroid **c** was measured along with the **a–b** distance (Table 2). Interestingly, both the **a–b** and the **a–c** distance had a narrow optimum range of 7.3–8.0 Å and 6.5–7.4 Å, respectively, which was only displayed by the high affinity compounds. The difference between the previously reported **a–b** distance [13] is probably caused by a change in the force field and electrostatics used for the conformational searches, as the earlier study was performed in Sybyl [26] using the Tripos force field and Gasteiger-Hückel charges.

The angle between **a–b** and **a–c** was then measured (Table 2) and again an optimum range was displayed by the high affinity compounds (Δ**bac** = 30.4–35.8°).

If the compounds with an IC_{50} value below 10 nM are used for defining the limits of the intervals, the pharmacophoric requirements for the interatomic distances and the angle are as illustrated in Figure 5.

If the two distance parameters (**a–b** and **a–c**) and the angle (Δ**bac**) interval are treated as three individual pharmacophoric requirements, a plot of affinity (expressed as p(IC_{50})) versus number of pharma-

Table 1. Transformation of literature binding data.

Compound	K_i (nM)	IC_{50} (nM) Using (S)-nicotine	IC_{50} (nM) Using (±)-epibatidine	$p(IC_{50})$[a]
9	4.6	29	20	7.61
10	18	52	110	7.09
(±)-Epiboxidine	0.6	3.9	3.5	8.43
UB-165	0.17	–	2.4	8.62

[a]The $p(IC_{50})$ values are calculated from the average of the two transformed IC_{50} values (Equation 1).

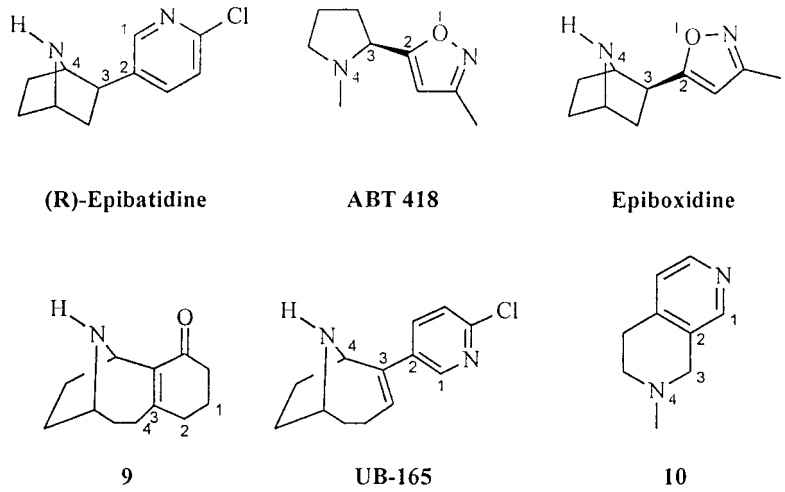
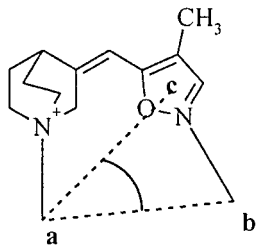


Figure 4. The numbered atoms were used for measuring the torsional angle.



a–b distance: 7.3 – 8.0 Å

a–c distance: 6.5 – 7.4 Å

Δbac: 30.4 – 35.8°

Figure 5. **a** is the site point corresponding to the protonated nitrogen atom, **b** is the site point corresponding to the electronegative atom capable of forming a hydrogen bond, **c** is the centre of a heteroaromatic ring or a C=O bond. Δ**bac** is the angle measured between the interatomic distance vectors **a–b** and **a–c**.

cophoric requirements met by the compounds (0, 1, 2, or 3) gives a nice separation of the compounds with different scores (Figure 6). As meeting a pharmacophoric requirement is a question of yes and no, some of the compounds that are very close to fulfilling the requirements have rather high affinity ($p(IC_{50}) = 7–8$). Yet, it is noteworthy that no compound with IC_{50} values above 10 nM fulfils all pharmacophoric requirements.

CoMFA-study

The conformations and alignment discussed in the previous section were used in a 3D-QSAR study.

In this study the steric and electrostatic interactions with a CH_3^+ probe were examined in order to obtain more information about the interactions between a set of nicotinic agonists and the α4β2 nAChRs. A training set covering 5 orders of magnitude of biological activity with $p(IC_{50})$ values ranging from 4.56 to 9.52 was used (Table 3).

Table 2. Compounds used in the pharmacophore study.

Compound	R_1	R_2	R_3	N-position	n	R′	Torsion[a] (°)	ΔE (kcal/mol)	a–b distance (Å)	a–c distance (Å)	Δ**bac** (°)	Pharmacophoric Score[b]	IC_{50} (nM)
1a	H	H	H	–	–	–	0	0	7.3	6.9	33.3	3	1.65
1b	Me	H	H	–	–	–	0	0	7.3	6.9	33.3	3	3.90
1d	H	Me	H	–	–	–	0	0	7.3	6.9	33.5	3	3.20
2a	Me	H	–	–	–	–	± 143	1.5	8.8	8.7	26.7	0	504
3a	–	–	–	meta	–	–	± 67	0.6	8.0	7.0	32.6	3	2.00
3b	–	–	–	ortho, meta	–	–	180	0	2.8	7.2	0.0	1	169
3c	–	–	–	ortho	–	–	0	0	2.8	7.1	1.1	1	1680
4a	–	–	–	ortho, meta	–	–	± 57	0.7	8.1	8.9	29.0	0	340
4b	–	–	–	ortho	–	–	± 126	0	8.0	8.9	28.9	1	458
5a	Me	–	–	–	–	–	0	0	7.6	6.1	24.7	1	13.0
7a	–	–	–	–	1	H	1	0	9.4	7.6	24.9	0	27200
7b	–	–	–	–	2	Me	1	0	8.6	6.7	27.8	1	783
8a	–	–	–	–	2	Me	−162	1.6	8.7	8.4	27.2	0	2330
ABT 418	–	–	–	–	–	–	129	2.7	9.3	6.1	19.3	0	56
(R)-Epibatidine	–	–	–	–	–	–	20	0.6	7.8	7.1	33.3	3	0.30
Epiboxidine	–	–	–	–	–	–	50	0	8.0	6.5	30.7	3	3.9[c]
UB-165	–	–	–	–	–	–	56	0.2	7.3	6.6	35.8	3	2.5[c]
9	–	–	–	–	–	–	159	0	7.7	5.4	33.9	2	30.0[c]
10	–	–	–	–	–	–	166	0	8.2	4.7	23.9	0	51.0[c]

[a]Only the supposed binding conformation is listed.
[b]The number of pharmacophoric requirements met; **a–b**: 7.3–8.0 Å, **a–c**: 6.5–7.4 Å, Δ**bac**: 30.4–35.8°.
[c]Transformed from literature values (see Methods).

Table 3. The training set.

Compound	R_1	R_2	R_3	N-position	n	R′	IC_{50} (M)	p(IC_{50})
1a	H	H	H	–	–	–	1.65×10^{-9}	8.78
1b	Me	H	H	–	–	–	3.90×10^{-9}	8.41
1c	Et	H	H	–	–	–	1.00×10^{-7}	7.00
1d	H	Me	H	–	–	–	3.20×10^{-9}	8.49
1e	H	H	Me	–	–	–	7.15×10^{-7}	6.15
1f	Me	H	F	–	–	–	2.55×10^{-7}	6.59
2a	Me	H	–	–	–	–	5.04×10^{-7}	6.30
2b	Et	H	–	–	–	–	1.23×10^{-6}	5.91
2c	MeOMe	H	–	–	–	–	2.44×10^{-6}	5.61
2d	i-Pr	H	–	–	–	–	4.16×10^{-6}	5.38
3a	–	–	–	meta	–	–	2.00×10^{-9}	8.70
3c	–	–	–	ortho	–	–	1.68×10^{-6}	5.77
5a	Me	–	–	–	–	–	1.30×10^{-8}	7.89
5b	Et	–	–	–	–	–	1.20×10^{-8}	7.92
7a	–	–	–	–	1	H	2.72×10^{-5}	4.56
7b	–	–	–	–	2	Me	7.83×10^{-7}	6.11
ABT 418	–	–	–	–	–	–	5.60×10^{-8}	7.25
(R)-Epibatidine	–	–	–	–	–	–	3.00×10^{-10}	9.52
(S)-Epibatidine	–	–	–	–	–	–	3.00×10^{-10}	9.52

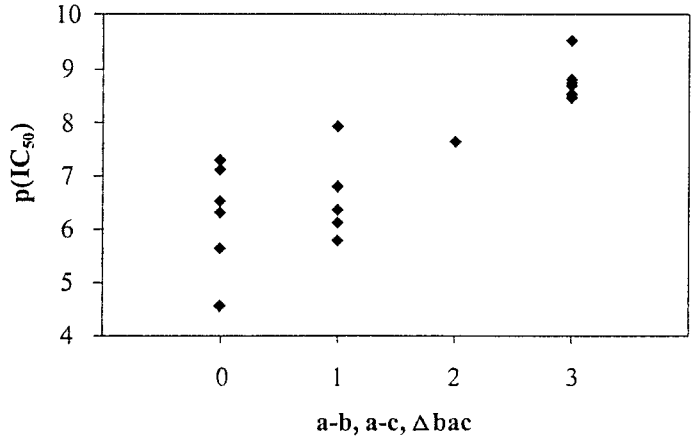


Figure 6. On the X-axis the number of fulfilled pharmacophoric requirements is given. p(IC_{50}) is calculated from IC_{50} values in M.

Table 4. Results from the PLS analyses.

Model[a]	r^2_{cv}	s_{cv}	Optimal number of components	s_{FIT}	r^2_{FIT}	Steric contribution	Electrostatic contribution
1A	0.289	1.481	4	0.216	0.983	56.7	43.3
1B	0.535	1.108	3	0.249	0.976	57.6	42.4
2A	0.368	1.434	7	0.072	0.998	55.7	44.3
2B	0.575	1.073	5	0.093	0.997	58.4	41.6
3A	0.512	1.154	5	0.172	0.989	60.5	39.5
3B	0.744	0.837	5	0.114	0.995	62.7	37.3
4A	0.407	1.353	5	0.197	0.987	54.7	45.3
4B	0.559	1.080	3	0.291	0.968	56.8	43.2

[a]In models 1A and 1B the full training set is used. In models 2A and 2B compound **7a** is left out. In models 3A and 3B compounds **7a** and **3c** are left out. In models 4A and 4B compound **3a** is left out.

In the following section the models designated A were derived without the use of region focusing, whereas region focusing was used for the models designated B. The numbers of the models (1, 2, 3, and 4) refer to different training sets.

First, model 1A was derived using the full training set (Table 3). The results from the cross-validation analysis showed a rather poor predictability within the training set, as r^2_{cv} was 0.289 for the optimum number of components (4) (Table 4). As an r^2_{cv} greater than 0.3 is usually considered acceptable [30] the r^2_{cv} obtained for model 1A is rather low.

The corresponding conventional PLS-analyses of model 1A showed very good correlation between the experimental and predicted activities within the training set ($r^2_{FIT} = 0.983$, $s_{FIT} = 0.216$). In an attempt to reduce the noise and thereby improve r^2_{cv}, variable selection was incorporated by the use of region focusing. As expected, the resulting model 1B did show significant improvements in the cross-validated r^2 ($r^2_{cv} = 0.535$). The steric and electrostatic contributions to the model did not change significantly (Table 4).

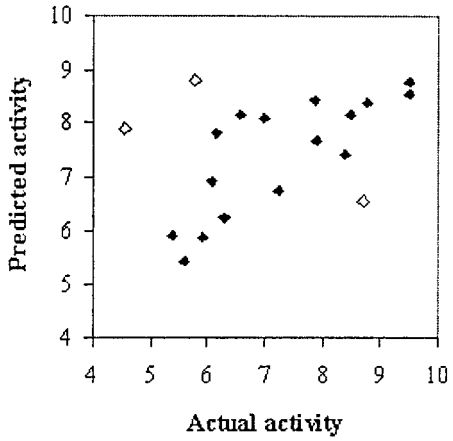


Figure 7. ◇: Compounds **3a**, **3c**, and **7a**. ♦: Remaining compounds.

Even though the results of the internal validation for model 1B were better than for model 1A, there was room for improvements. Thus, in an attempt to improve models 1A and 1B, the results from the cross-validation (leave one out) of model 1A were analysed (Figure 7), revealing that compounds **7a**, **3c**, and **3a** were badly predicted by the remaining compounds. Hence, new models were prepared by leaving out these compounds from the original training set. In models 2A and 2B compound **7a** is left out. In models 3A and 3B compounds **7a** and **3c** are left out. In models 4A and 4B compound **3a** is left out. A possible reason why compound **7a** and **3c** are not well predicted in the cross validation procedure can be that **7a** is one of few compounds with an N-H group and compound **3c** is the only compound with a nitrogen atom in ortho position.

The results in Table 4 show that slight improvements are observed when **7a** or **3a** are left out (models 2 and 4, $r^2_{CV} = 0.368$–0.559), whereas major improvements result when **7a** and **3c** are left out simultaneously (models 3A and 3B, $r^2_{CV} = 0.512$ and 0.744, respectively). Hence, based on internal validation the performance of the models decrease in the following order: 3B > 1B, 2B, 3A, 4B > 2A, 4A > 1A.

As internal validation can fail to detect overfitting in the models and thus give overoptimistic indicators of predictive performance, an external validation was performed on a test set. The molecules in the test set (Table 5) comprise structural features present in the training set along with some new features, as it would be interesting to test the models' predictability towards novel structures.

The actual and predicted p(IC_{50})-values for the test set are listed in Table 6. The average absolute error (AAE) was calculated for all models revealing that models 1A, 1B, 4A, and 4B, having AAE values of 0.36–0.48, are best at predicting the compounds in the test set.

Removal of compound **7a** (models 2A and 2B) resulted in significantly higher AAE values (0.76–0.77) reflecting a loss of predictability towards the pyrrolidines (**7c** and **8a**) and probably other azetidines than **7a** itself, if these were to be tested in the future. Further removal of compound **3c** (models 3A and 3B) resulted in even higher AAE values (0.86–0.94) due to a loss of predictability towards **7c** and **8a** as well as the (*Z*)-pyrazine compound **3b** indicating that models 3A and 3B cannot extrapolate to pyrazines and *ortho*-pyridines (as **3c**). Interestingly, model 3A performed well in predicting the (*E*)-pyrazine **4a**, which is badly predicted by the remaining models. The prediction of the test set did not reveal any significant changes upon removal of compound **3a** from the training set (models 4A and 4B), but this can be due to the fact that no other *meta*-pyridines are present in the test set. Thus, if some new substituted *meta*-pyridines were to be tested, models 4A and 4B might fail to predict their activities.

The performance of the models in the external validation decrease in the following order: 1A, 1B, 4A, 4B > 2A, 2B > 3A, 3B. It is remarkable that model 3B, which performed best in the internal validation was the worst in the external validation (Figure 8).

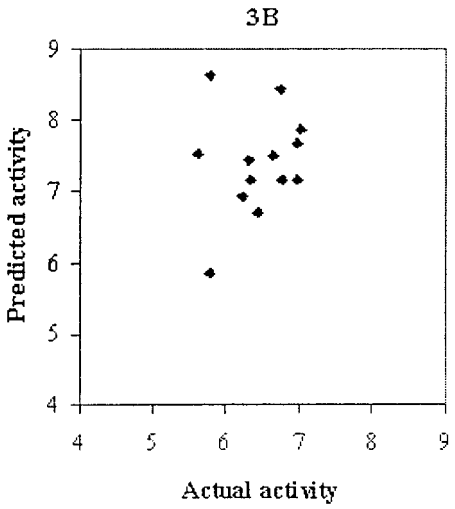


Figure 8. Predicted activity vs. Actual activity of the test set for model 3B.

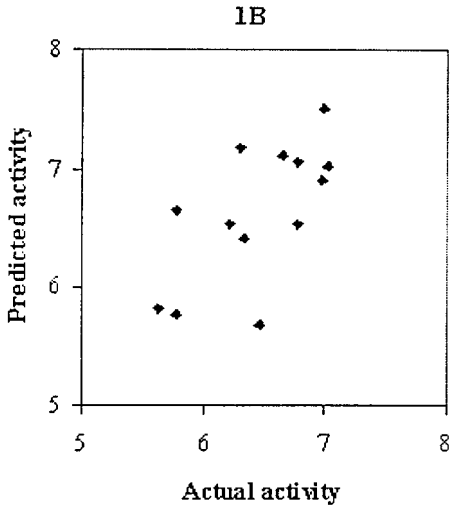


Figure 9. Predicted activity vs. Actual activity of the test set for model 1B.

Table 5. The test set.

Compound	R_1	R_2	R_3	N-position	R′	n	IC_{50} (M)
1g	MeOMe	H	H	–	–	–	2.18×10^{-7}
1h	n-Pr	H	H	–	–	–	1.67×10^{-7}
1i	i-Pr	H	H	–	–	–	1.07×10^{-7}
2e	n-Pr	H	–	–	–	–	1.64×10^{-6}
3b	–	–	–	ortho, meta	–	–	1.69×10^{-7}
4a	–	–	–	ortho, meta	–	–	3.40×10^{-7}
4b	–	–	–	ortho	–	–	4.58×10^{-7}
5c	n-Pr	–	–	–	–	–	1.03×10^{-7}
5d	OEt	–	–	–	–	–	4.90×10^{-7}
6a	Me	–	–	–	–	–	9.20×10^{-8}
6b	Et	–	–	–	–	–	5.90×10^{-7}
7c	–	–	–	–	H	2	1.65×10^{-6}
8a	–	–	–	–	Me	2	2.33×10^{-6}

By combining the internal and external validation 1B (Figure 9), 4A, and 4B turned out to be the most promising models.

Going into further detail with the external validation of these models shows that the compound providing the most diverging predictions is **4a**, which is the only (*E*)-pyrazine compound present in both sets. The large difference between the actual and predicted activity is therefore probably a result of its uniqueness. The models also have some difficulty in predicting the pyrrolidine compound **7c**. This may be caused by the fact that **7c** is the only unsubstituted pyrrolidine compound present. The remaining compounds with moderately diverging predictions all have long flexible substituents (**1g**, **1h**, **5c**, and **5d**). Hence, the diverging predictions can derive from the choice of substituent conformation used for the analyses, or the fact that no (*Z*)-compound with a R_1-substituent longer than *i*-Pr is present in the training set. Interestingly, the oxadiazole compounds **6a** and **6b** are predicted quite well, despite the fact that this class of compounds is not present in the training set. This indicates, that the models can be applied for prediction of compound classes outside the training set.

Taking all the information into account, model 1B is the most interesting model, as it is able to predict the biological activities of compounds belonging to compound classes present in the training set as well as novel structures.

CoMFA maps. The CoMFA electrostatic and steric fields derived from model 1B are shown as contour maps in Figure 10, Figure 11, and Figure 12 along with the structure of (*R*)-epibatidine and **1f**.

The contours of the electrostatic map are shown in red and blue (Figure 10), and are to be interpreted as follows: A higher activity is correlated with more positive charge near the blue contours and more negative charge near the red contours.

Red areas near the basic nitrogen atom and the electronegative atom, which is supposed to accept a hydrogen bond, correspond with the site points (**a**,**b**) and confirm that these interactions are definitely important for binding. Additional red areas near the chlorine atoms of (*R*)- and (*S*)-epibatidine reflect that no other compounds have halogen substituents in this region. The blue areas near the chlorine atom probably arise from the isoxazole compounds with small alkyl substituents. The additional blue area close to the heteroaromatic ring may explain the good predictions of the oxadiazole compounds **6a** and **6b** in the test set.

It has been described [14] that decreased electron density is favoured near the basic nitrogen atom. However, since the previous study used protonated compounds this does not contradict the findings in our study.

The contours of the steric map are shown in green and yellow (Figure 12), and are to be interpreted as follows: A higher activity is correlated with more steric bulk near the green contours and less steric bulk near the yellow contours.

Corresponding with earlier results [14], large green areas are present in the upper part of the aliphatic heterocycle showing that steric bulk in this region is

Table 6. The actual and predicted p(IC_{50})-values for the test set.

Compound	pIC_{50}	1A		1B		2A		2B		3A		3B		4A		4B	
		Pred.	Res.	Pred.	Res.	Pred.	Res.	Pred.	Res.	Pred.	Res.	Pred.	Res.	Pred.	Res.	Pred.	Res.
1g	6.66	7.28	0.62	7.11	0.45	7.26	0.60	7.31	0.65	7.32	0.66	7.50	0.84	7.29	0.63	7.11	0.45
1h	6.78	7.32	0.54	7.06	0.28	7.07	0.29	7.08	0.30	7.15	0.37	7.14	0.36	7.29	0.51	7.09	0.31
1i	6.97	7.17	0.20	6.91	−0.06	7.20	0.23	7.12	0.15	7.19	0.22	7.14	0.17	7.18	0.21	7.01	0.04
2e	5.79	5.62	−0.17	5.77	−0.02	5.84	0.05	5.81	0.02	5.78	−0.01	5.84	0.05	5.63	−0.16	5.86	0.07
3b	6.77	6.46	−0.31	6.53	−0.24	6.40	−0.37	6.46	−0.31	8.60	1.83	8.42	1.65	6.25	−0.52	6.72	−0.05
4a	6.47	5.24	−1.23	5.69	−0.78	5.55	−0.92	6.34	−0.13	6.28	−0.19	6.69	0.22	5.12	−1.35	5.54	−0.93
4b	6.34	6.23	−0.11	6.42	0.08	6.79	0.45	7.08	0.74	6.93	0.59	7.16	0.82	6.20	−0.14	6.32	−0.02
5c	6.99	7.65	0.66	7.51	0.52	7.76	0.77	7.68	0.69	7.57	0.58	7.66	0.67	7.59	0.60	7.47	0.48
5d	6.31	7.05	0.74	7.19	0.88	7.17	0.86	7.40	1.09	7.34	1.03	7.44	1.13	6.93	0.62	7.20	0.89
6a	7.04	7.29	0.25	7.02	−0.02	7.67	0.63	7.65	0.61	7.69	0.65	7.86	0.82	7.28	0.24	7.04	0.00
6b	6.23	6.78	0.55	6.53	0.30	6.53	0.30	6.62	0.39	6.88	0.65	6.92	0.69	6.65	0.42	6.55	0.32
7c	5.78	6.41	0.63	6.65	0.87	8.42	2.64	8.76	2.98	8.32	2.54	8.62	2.84	6.52	0.74	6.57	0.79
8a	5.63	5.69	0.06	5.82	0.19	7.49	1.86	7.46	1.83	7.47	1.84	7.53	1.90	5.79	0.16	5.98	0.35
Sum	–	–	6.07	–	4.69	–	9.97	–	9.89	–	11.16	–	12.16	–	6.30	–	4.70
AAE[a]	–	–	0.47	–	0.36	–	0.77	–	0.76	–	0.86	–	0.94	–	0.48	–	0.36

[a]AAE = Average absolute error.

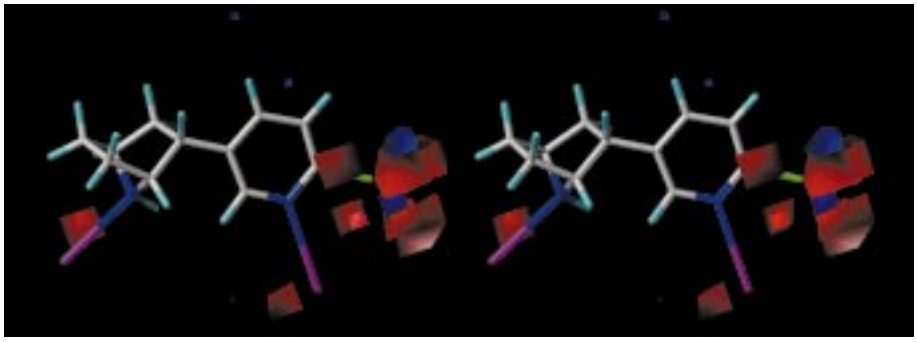


Figure 10. Stereoview of the electrostatic CoMFA-map with the structure of (*R*)-epibatidine (site points added). The CoMFA-maps shows the Std. Dev. coefficient at the levels 75:25.

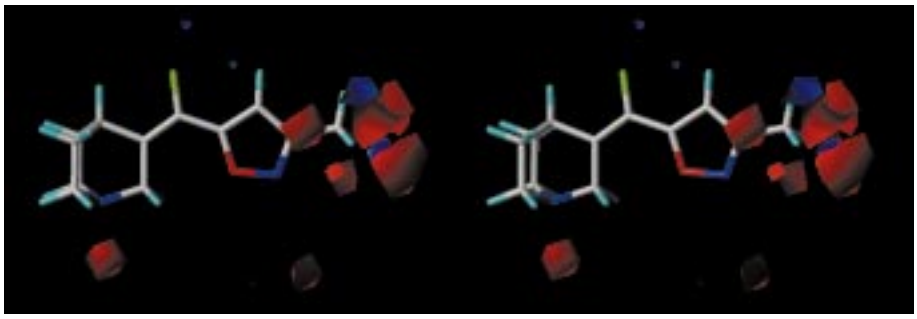


Figure 11. Stereoview of the electrostatic CoMFA-map with the structure of **1f.** The CoMFA-maps shows the Std. Dev. coefficient at the levels 75:25.

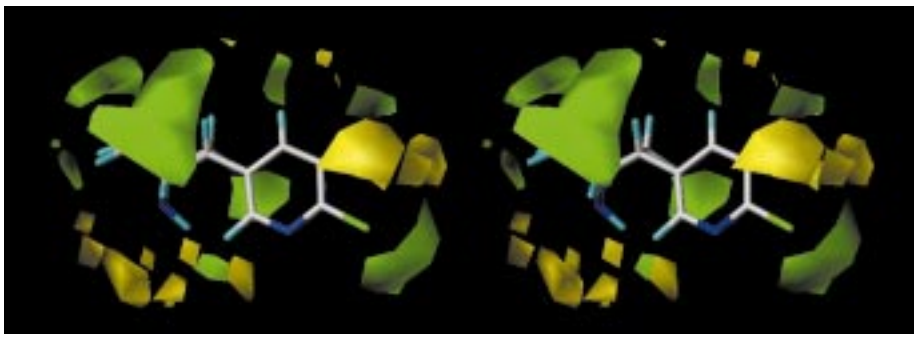


Figure 12. Stereoview of the steric CoMFA-map with the structure of (*R*)-epibatidine. The CoMFA-maps shows the Std. Dev. coefficient at the levels 75:25.

favourable. Near the H_5 and chlorine substituents on the pyridine ring a yellow area tells that steric bulk in this position is unfavourable. However, the exact location of this yellow area is uncertain due to the lack of information regarding the choice of conformation for the flexible substituents in that area. The green area close to the chlorine atoms of (*R*)- and (*S*)-epibatidine is again a result of the big influence these two very potent compounds have on the CoMFA model. The yellow areas below and to the left of the aliphatic heterocycle are very interesting, as they are situated exactly where a large portion of steric bulk from the aliphatic heterocycle of ent-5 is situated. This indicates, that model 1B displays the same stereoselectivity as the nAChRs do.

Conclusion

An improved pharmacophore for nicotinic agonists binding to the α4β2 nAChRs has been developed using 19 compounds of varying molecular structure spanning a wide range of binding activities. The pharmacophore is composed of three elements **a**, **b**, and **c**; where **a** is the site point corresponding to the protonated nitrogen atom, **b** is the site point corresponding to the electronegative atom capable of forming a hydrogen bond, and **c** is the centre of a heteroaromatic ring or a C=O bond. Δ**bac** is the angle measured between the interatomic distance vectors **a–b** and **a–c**. The pharmacophoric elements where found to be related by the following parameters:

a–b distance :	7.3–8.0 Å,
a–c distance :	6.5–7.4 Å,
Δ**bac** :	30.4–35.8°.

In addition, a 3D-QSAR model was developed using the CoMFA methodology for a set of 19 nicotinic agonists also spanning a wide range of binding affinities. The model was able to predict the activity of a test set composed of 13 compounds covering structural elements present in the training set as well as novel elements. Finally, the model was able to display a stereoselectivity, which was not inherent in the training set, but has been described in the literature.

References

1. Beers, W.H. and Reich, E., Nature, 228 (1970) 917.
2. Sheridan, R.P., Nilakantan, R., Dixon, J.S. and Venkataraghavan, R., J. Med. Chem., 29 (1986) 899.
3. Manallack, D.T., Gallagher, T. and Livingstone, D.J., in Devillers, J. (Ed), Neural Networks in QSAR and Drug Design, Academic Press, London, 1996, pp. 177–208.
4. Holladay, M.W., Lebold, S.A. and Lin, N.-H., Drug Dev. Res., 35 (1995) 191.
5. Glennon, R.A. and Dukat, M., Med. Chem. Res. (1996) 465.
6. Hacksell, U. and Mellin, C., Prog. Brain Res., 79 (1989) 95.
7. Wright, E., Gallagher, T., Sharples, C.G.V. and Wonnacott, S., Bioorg. Med. Chem. Lett., 7 (1997) 2867.
8. Zhang, X., Stjernlöf, P., Adem, A. and Nordberg, A., Eur. J. Pharmacol., 135 (1987) 457.
9. Dukat, M., Damaj, M.I., Glassco, W., Dumas, D., May, E.L., Martin, B.R. and Glennon, R.A., Med. Chem. Res., 4 (1993) 131.
10. Abreo, M.A., Lin, N.-H., Garvey, D.S., Gunn, D.E., Hettinger, A.M., Wasicak, J.T., Pavlik, P.A., Martin, Y.C., Donnelly-Roberts, D.L., Anderson, D.J., Sullivan, J.P., Williams, M., Arneric, S.P. and Holladay, M.W., J. Med. Chem., 39 (1996) 817.
11. Koren, A.O., Horti, A.G., Mukhin, A.G., Gundisch, D., Kimes, A.S., Dannals, R.F. and London, E.D., J. Med. Chem., 41 (1998) 3690.
12. Cheng, Y.-X., Dukat, M., Dowd, M., Fiedler, W., Martin, B., Damaj, M.I. and Glennon, R.A., Eur. J. Med. Chem., 34 (1999) 177.
13. Tønder, J.E., Hansen, J.B., Begtrup, M., Pettersson, I., Rimvall, K., Christensen, B., Ehrbar, U. and Olesen, P.H., J. Med. Chem., 42 (1999) 4970.
14. Glennon, R.A., Herndon, J.L. and Dukat, M., Med. Chem. Res., 4 (1994) 461.

15. Holladay, M.W., Dart, M.J. and Lynch, J.K., J. Med. Chem., 40 (1997) 4169.
16. Okazawa, A., Akamatsu, M., Ohoka, A., Nishiwaki, H., Cho, W.-J., Nakagawa, Y., Nishimura, K. and Ueno, T., Pestic. Sci., 54 (1998) 134.
17. Sukekawa, M. and Nakayama, A., J. Pestic. Sci., 24 (1999) 38.
18. Olesen, P.H., Tønder, J.E., Hansen, J.B., Hansen, H.C. and Rimvall, K., Bioorganic and Medicinal Chemistry, submitted (2000).
19. Mohamadi, F., Richards, N.G.J., Guida, W.C., Liskamp, R., Lipton, M., Caufield, C., Chang, G., Hendrickson, T. and Still, W.C., J. Comput. Chem., 11 (1990) 440.
20. Halgren, T.A., J. Comput. Chem., 17 (1996) 490.
21. Halgren, T.A., J. Comput. Chem., 17 (1996) 520.
22. Halgren, T.A., J. Comput. Chem., 17 (1996) 553.
23. Halgren, T.A., J. Comput. Chem., 17 (1996) 587.
24. Hasel, W., Hendrickson, T.F. and Still, W.C., Tetrahedron Comput. Method., 1 (1988) 103.
25. Still, W.C., Tempczyk, A., Hawley, R.C. and Hendrickson, T., J. Am. Chem. Soc., 112 (1990) 6127.
26. Tripos Inc., 1699 South Hanley Road, St. Louis, Missouri, 63144, USA
27. Dukat, M., Fiedler, W., Dumas, D., Damaj, I., Martin, B.R., Rosecrans, J.A., James, J.R. and Glennon, R.A., Eur. J. Med. Chem., 31 (1996) 875.
28. Badio, B., Garraffo, H.M., Plummer, C.V., Padgett, W.L. and Daly, J.W., Eur. J. Pharmacol., 321 (1997) 189.
29. Badio, B. and Daly, J.W., Mol. Pharmacol., 45 (1994) 563.
30. Kim, K.H., Greco, G. and Novellino, E., Perspectives in Drug Discovery and Design, 12/13/14 (1998) 257.