Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 12 (2004) 6277–6285
Bioorganic & Medicinal Chemistry

Bisquaternary caracurine V and *iso*-caracurine V salts as ligands for the muscle type of nicotinic acetylcholine receptors: SAR and QSAR studies

D. P. Zlotos,[a,*] D. Gündisch,[b] S. Ferraro,[b] M. C. Tilotta,[b] N. Stiefl[a] and K. Baumann[a]

[a]*Pharmaceutical Institute, University of Würzburg, Am Hubland, 97074 Würzburg, Germany*
[b]*Pharmaceutical Institute, University of Bonn, Kreuzbergweg 26, 53115 Bonn, Germany*

Received 23 January 2004; accepted 27 August 2004
Available online 28 September 2004

Abstract—The binding constants (K_i values) of 24 caracurine V and 6 *iso*-caracurine V analogues for the muscle type of nicotinic ACh receptors (nAChR) from *Torpedo californica* were determined in a binding assay using (±)-[^{3}H]epibatidine as a radioligand. The allyl alcohol group present in the *iso*-caracurine V ring system was found to be essential for high binding affinity. The most potent compounds are the dimethyl and di-(4-nitrobenzyl)-*iso*-caracurinium V salts **29** (18 nM), and **31** (79 nM), respectively. Compound **29** and the corresponding diallyl analogue **30** (350 nM) exhibited similar binding affinities as the equally substituted neuromuscular-blocking agents toxiferine I (14 nM) and alcuronium (234 nM), respectively. The SAR results were confirmed by QSAR studies, which additionally revealed that the presence of hydrogen-bond acceptor groups close to the quaternary nitrogen, is detrimental for the nicotinic binding affinity. The diallyl- and dimethylcaracurinium V salts **13** and **27**, respectively, which are known to be among the most potent allosteric modulators of M_2 receptors (EC_{50} = 10 and 8 nM, respectively), exhibited rather low nicotinic binding affinities for muscle type nAChR (K_i = 1.5 and 5.2 μM, respectively). Such a large difference in affinity suggests that it is possible to develop compounds with high muscarinic allosteric potency and low or negligible affinities for $(\alpha 1)_2\beta 1\gamma\delta$ nAChR. Additionally, the *iso*-caracurine V analogues with binding affinities comparable to those of (+)-tubocurarine and alcuronium could become a new class of neuromuscular-blocking agents.

1. Introduction

Bisquaternary ammonium salts of the *Strychnos* alkaloid caracurine V (**1**)[1,2] and the isomeric *iso*-caracurine V (**2**),[2] are among the most potent ligands for the allosteric site of muscarinic M_2 receptors. Owing to the close structural relationship of caracurine V and *iso*-caracurine V analogues to the muscle relaxants toxiferine I (**3**) and alcuronium (**4**) (see Fig. 1), they are likely to exhibit neuromuscular-blocking activity. This would limit their usefulness as muscarinic research tools, and make their therapeutical use impossible. However, the neuromuscular-blocking activity of caracurine V dimethochloride **27**, which is a double ring-closed product of toxiferine I **3**, was reported to be about 50-fold lower than that of toxiferine.[3]

In order to investigate the essential structural features responsible for the interaction with $(\alpha 1)_2\beta 1\gamma\delta$ nAChR, the binding affinities of various *N*-substituted caracurine V and *iso*-caracurine V analogues to the muscle type of the nicotinic acetylcholine receptors were determined. The obtained SARs were supplemented and validated by a 3D-QSAR study employing the MaP technique (mapping property distributions of molecular surfaces).[4,5] The findings reported here are important for the development of potent allosteric modulators of muscarinic M_2 receptors with negligible neuromusblocking activity, as well as for the identification of new lead structures with neuromuscular-blocking activity.

2. Results and discussion

2.1. Chemistry

The structures of the investigated compounds are given in Figure 1 and Tables 1 and 2. Toxiferine I (**3**) was prepared from the dimerization of Wieland–Gumlich aldehyde methochloride with pivalic acid according to the

* Corresponding author. Tel.: +49 931 8885489; fax: +49 931 8885494; e-mail: zlotos@pzlc.uni-wuerzburg.de

0968-0896/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2004.08.053

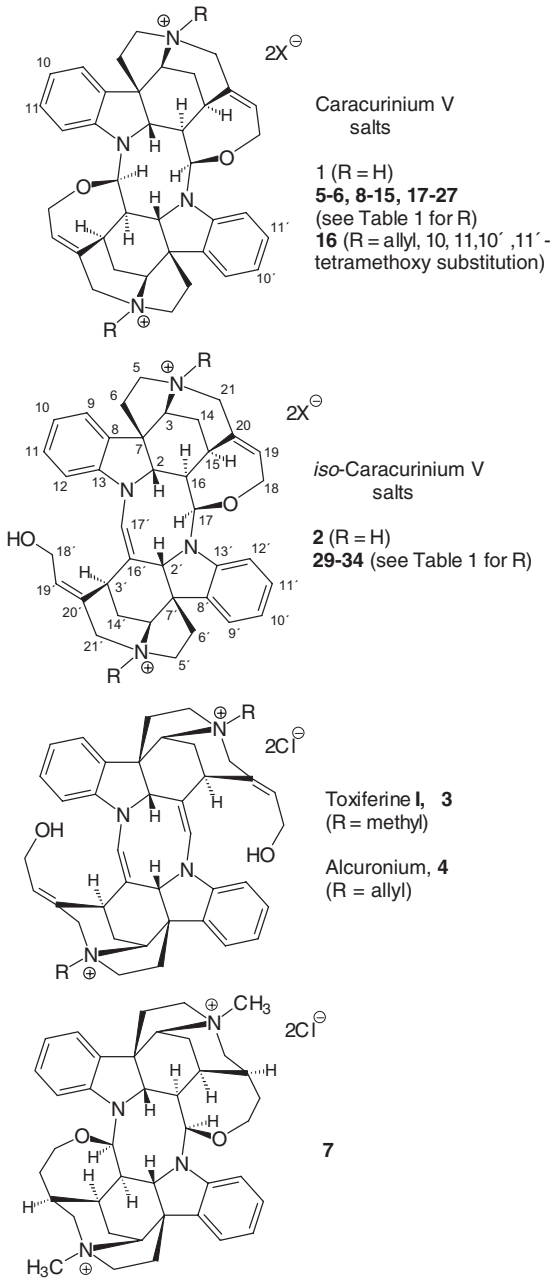
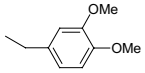
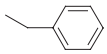
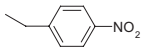
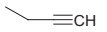
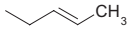
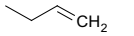
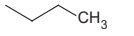
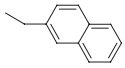
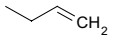
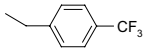
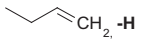
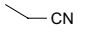
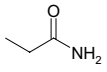
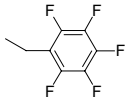
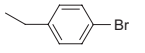
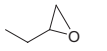
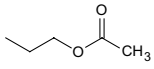
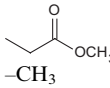


Figure 1. Structures of the investigated compounds.

procedure by Battersby and Hodson.[6] Caracurine V (**1**),[7] *iso*-caracurine V (**2**),[8] the caracurinium V salts **5**–**27**, as well as the *iso*-caracurinium V salts **29**, **30** and **32** were prepared as recently described by Zlotos et al.[2] The new *iso*-caracurine V analogues **31**, **33** and **34** were synthesized by double quaternization of *iso*-caracurine V in a chloroform solution at room temperature using a 2.5-fold excess of 4-nitrobenzyl bromide, 3,4-dimethoxybenzyl chloride[9] and benzyl bromide, respectively. Caracurine V dioxide (**28**) was prepared by *N*-oxidation of **1** using hydrogen peroxide as previously described by Verpoorte and Baerheim Svendsen.[10] Alcuronium (**4**) was a generous donation from Roche, Basel.

Table 1. Binding constants of the caracurine V salts (**5**–**6**, **8**–**27**), dimethyltetrahydrocaracurinium V dichloride (**7**) and caracurine V dioxide (**28**) to the nicotinic ACh receptors from the *Torpedo californica* electric organ (for structures see Fig. 1)

	R	K_i (nM)
5		270 ± 59
6		370 ± 1
7[c]	–CH_3	390 ± 78
8		570 ± 2
9*		760 ± 12
10		820 ± 52
11*		1400 ± 82
12		1395 ± 22
13		1500 ± 23
14*		1600 ± 275
15		2000 ± 79
16[b]		2100 ± 280
17*		2300 ± 552
18		2400 ± 34
19[a]		2600 ± 101
20		3200 ± 297
21		3200 ± 326
22*		3600[d]
23		3600[d]
24		3800[d]
25		4300[d]
26		4700[d]
27	–CH_3	5200[d]
1	–H	>100,000
28	–O^-	>150,000

Data represent means ± SEM obtained from 3 to 6 experiments. Members of the test set in the QSAR analysis are marked with an asterisk (*).

[a] *N*-Allylcaracurinium V bromide.

[b] 10,11,10′,11′-Tetramethoxycaracurine V.

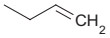
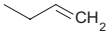
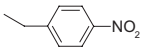
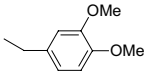
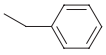
[c] 19,20,19′,20′-Tetrahydrocaracurine V ring system.

[d] Data from a single experiment.

2.2. Pharmacological studies

All compounds listed in Tables 1 and 2 were tested for their in vitro affinity for $(\alpha 1)_2\beta 1\gamma\delta$ nAChR (muscle type)

Table 2. Binding constants of the *iso*-caracurine V salts **29**–**34** and the muscle relaxants toxiferine I, alcuronium and (+)-tubocurarine to the nicotinic ACh receptors from the *Torpedo californica* electric organ (for structures see Fig. 1)

	R	K_i (nM)
(+)-Tubocurarine	—	16 ± 3
Toxiferine I (**3**)	$-CH_3$	14 ± 0.8
29*	$-CH_3$	18 ± 2
Alcuronium (**4**)		234 ± 19
30*		350 ± 39
31		79 ± 26
32		95 ± 2
33		180 ± 31
34		100 ± 3

Data represent means ± SEM obtained from 3 to 6 experiments. Members of the test in the QSAR analysis set are marked with an asterisk (*).

by radioligand binding assays. (±)-[^{3}H]Epibatidine bound to a single population of binding sites in membrane fractions of *Torpedo californica* electric organ, exhibiting a K_d value of 2.0 ± 0.3 nM. For (−)-epibatidine, competition assays yielded a K_i value of 2.5 nM, which is about three times lower than the K_i value obtained in a well established assay using [^{125}I]α-bungarotoxin as radioligand.[11]

2.3. SAR discussion

2.3.1. Caracurine V series. Except for caracurine V dioxide (**28**), which is the only compound with positive charges fully compensated by negatively charged oxygen atoms, all investigated compounds were able to displace (±)-[^{3}H]epibatidine from *Torpedo californica* membranes. Hence, it can be concluded that ammonium cations are important for the receptor–ligand interactions. The only bistertiary compound, caracurine V (**1**), exhibited a very poor binding affinity (125 μM). These findings are in agreement with the fact that all known neuromuscular-blocking agents posses at least one permanently quaternized nitrogen atom.[12]

As expected, double quaternization of **1** caused a dramatic increase of binding affinity (24–460-fold). The only monoquaternary compound in this series, *N*-allylcaracurinium V bromide **19** (2.6 μM) was nearly as potent as the corresponding double alkylated analogue **13** (1.5 μM), indicating that one permanently charged nitrogen atom is sufficient for the ligand–receptor interactions. Introduction of methoxy goups at C10, C11, C10′, C11′ at both aromatic rings of **13** to give the tetramethoxy analogue **16** generated only slightly reduced binding (2.1 μM). The K_i values in the caracurine V series ranged from 270 nM (**5**, R = cyclohexen-3-yl) to 5.3 μM (**27**, R = methyl). Inspection of the data given in Table 1 revealed no clear correlation between steric and electronic properties of the *N*-substituent and the nicotinic binding affinity. For instance, on the one hand, the most potent caracurinium V salts carry rather bulky *N*-substituents such as cyclohexen-3-yl (**5**, 270 nM), benzyl (**8**, 570 nM), 3,4-dimethoxybenzyl (**6**, 370 nM) and 4-nitrobenzyl (**10**, 820 nM). On the other hand, substitution with such bulky substituents as 4-bromobenzyl and pentafluorobenzyl (**23** and **22**, respectively, both 3.6 μM), 2-naphthyl (**15**, 2 μM) as well as trifluoromethylbenzyl (**18**, 2.4 μM) produced compounds with rather low binding affinities. The least potent compound in the bisquaternary caracurine V series is dimethylcaracurinium V dichloride **27** (5.2 μM). However, hydrogenation of both the double bonds of **27** to give 4,4′-dimethyl-19,20,19′,20′-tetrahydrocaracurinium V dichloride[2] **7** (390 nM), led to a 13-fold increase in affinity, which might be due to the different 3D structures of both ring systems.[13]

2.3.2. *iso*-Caracurine V series. *iso*-Caracurine V (**2**) is a product of an intramolecular alcohol elimination from the central diazocane ring of caracurine V (**1**) and has therefore only one allyl alcohol side chain. A second alcohol elimination from the central diazocane ring of *iso*-caracurine V leads to the symmetrical ring skeleton of the muscle relaxants toxiferine I (**3**) and alcuronium (**4**), which are substituted with two allyl alcohol groups (see Fig. 1). In order to examine whether the side chains affect the nicotinic binding affinity, and to obtain more potent compounds for further QSAR studies, a number of bisquaternary *iso*-caracurinium V salts with the same substituents as those being present in the most potent caracurine V salts were prepared (see Table 2). Additionally, dimethyl and diallyl *iso*-caracurinium V analogues, respectively, were synthesized to compare their binding affinities with those of toxiferine I and alcuronium.

All *iso*-caracurine V salts **29**–**34** exhibited a higher affinity to the nicotinic ACh receptors than the equally substituted caracurine V analogues. The largest increase of binding affinity (294-fold) was observed for the dimethyl substituted compounds (caracurinium salt **27**, 5.2 μM; *iso*-caracurinium salt **29**, 14 nM). In the case of other substituents, opening of one tetrahydrooxepine ring of caracurine V produced only moderate increase in binding affinity (3-fold for cyclohexen-3-yl and dimethoxybenzyl, 4-fold for benzyl, 5-fold for allyl, 16-fold for 4-nitrobenzyl). Both ring systems adopt very similar 3D geometries in terms of the relative spatial arrangement of the aromatic rings and the distance between the positively charged nitrogen atoms (about 9.6 Å).[13] Hence, it can be concluded that the N^+–N^+ distance as well as the relative position of the aromatic rings are not essential for the ligand–receptor interactions. The different binding affinities of *iso*-caracurine V and caracurine V analogues are probably due to the presence or absence of conformationally flexible allyl alcohol side chains. While the oxygen atoms in the caracurine V ring system are incorporated in the rigid tetrahydrooxepine rings, and therefore, unable to change their relative positions to the cationic centres, the alcohol side chain of the

iso-caracurine V ring skeleton can adopt an optimal conformation to serve as a hydrogen-bond donor or acceptor for the receptor protein.

Interestingly, opening of the second tetrahydrooxepine ring of dimethyl (**29**) and diallyl-*iso*-caracurinium V (**30**) cations, resulting in toxiferine and alcuronium (two alcohol moieties), respectively, did not further improve the binding affinity. Apparently, one flexible hydroxy group is sufficient for high nicotinic potency.

Surprisingly, unlike in the caracurine V series, the most potent *iso*-caracurine V salt was the methyl analogue **29** (18 nM). Its affinity approached that of toxiferine I (14 nM). Both compounds are nearly as potent as the classical neuromuscular blocker (+)-tubocurarine (16 nM).

2.4. QSAR studies

To supplement and validate the SARs found and to obtain a model that can be used for the prediction of unsynthesized compounds, a QSAR study was performed. For outlier detection leave-one-out cross-validation (LOO-CV) residuals were visually inspected and caracurine V (**1**) was identified as an outlier for all techniques applied. This is not surprising, since it is structurally unique being the only compound without a permanent charge.

In order to identify the relevance of the SAR observations, the K_i-values of the structurally related series of *iso*-caracurine V and caracurine V analogues were correlated with a single variable encoding the existence of an allyl alcohol group (indicator variable). The result obtained with this simplistic approach was exceptionally good (leave-one-out cross-validated squared multiple correlation coefficient $R^2_{CV-1} = 0.61$, see also Table 3), indicating that the allyl alcohol group is the main structural feature relevant for the nicotinic potency.

However, since more structural features might influence the binding affinity, a CoMSIA[14] study was performed. Surprisingly, this 3D-QSAR technique did not yield much better models (best $R^2_{CV-1} = 0.63$). In summary, the best performing CoMSIA field, which was the H-bond acceptor field, also pinpoints the importance of the allyl alcohol moiety, but additionally indicates that acceptor groups close to the quarternary nitrogen reduce biological activity (contour plots are available on request). On the one hand, the rather low R^2_{CV-1}-value is understandable, since most of the variance in the data can be explained by a single variable. On the other hand, the alignment of the *N*-substituents is rather complicated and therefore, additional structural features might be well missed out by the CoMSIA model. As a consequence, the MaP technique[4] was applied since it is independent of the superimposition of the molecules (the descriptor is so-called translationally and rotationally invariant). Based on the knowledge about the significance of the allyl alcohol group, a differentiation between weak and strong H-bond donors was included in the model building procedure,[5] which is an extension of the original MaP descriptor. Since the full model using all MaP variables did not perform well (see Table 3) and since only a limited number of features control biological action, a variable selection procedure was applied. Using forward selection in combination with principal component regression (PCR) and leave-multiple-out cross-validation (LMO-CV) as the objective function, the following QSAR equation was obtained with the MaP descriptor:

$$\hat{pK}_i = 3.0387 + 0.1025 \cdot DsDs_3 - 0.0032 \cdot AA_9 - 0.0164 \cdot DsH_3 + 0.0202 \cdot DsDs_4$$

$$R^2_{CV-1} = 0.73, \quad RMSEP_{CV-1} = 0.39,$$
$$R^2_{CV-40\%} = 0.69, \quad RMSEP_{CV-40\%} = 0.42, \quad R^2 = 0.77,$$

$$RMSEC = 0.38, \quad m = 33, \quad LV = 3$$

where R^2_{CV-1}(also known as q^2) and $R^2_{CV-40\%}$ are the leave-one-out and leave-40%-out cross-validated squared multiple correlation coefficients, and $RMSEP_{CV-1}$ and $RMSEP_{CV-40\%}$ are the respective root-mean squared errors of prediction. R^2 is the coefficient of determination, RMSEC is the root mean squared error of calibration (also known as *s*), *m* is the

Table 3. QSAR results

Data set	$RMSEP_{CV-1}$	R^2_{CV-1}	$RMSEP_{CV-40\%}$	$R^2_{CV-40\%}$	RMSEC	R^2	$RMSEP_{Test}$	R^2_{Test}	m/m_{Test}	*N*	n_{sel}	LV
IND[a]	0.47	0.61	0.47	0.60	0.45	0.66	—	—	33/—	1	—	1
MaP-full[b]	0.57	0.43	—	—	0.45	0.69	—	—	33/—	293	—	4
MaP-selected[c]	0.39	0.73	0.42	0.69	0.38	0.77	—	—	33/—	293	4	3
MaP-selected[d]	0.45	0.66	0.47	0.62	0.43	0.73	0.34	0.76	25/8	293	3	3

Symbols: $RMSEP_{CV-1}$: leave-one-out cross-validated root mean squared error of prediction; R^2_{CV-1}: leave-one-out cross-validated coefficient of determination; $RMSEP_{CV-40\%}$ and $R^2_{CV-40\%}$ same as $RMSEP_{CV-1}$ and R^2_{CV-1} for leave-40%-out cross-validation; RMSEC: root mean squared error of calibration; R^2: coefficient of determination; $RMSEP_{Test}$, R^2_{Test} and m_{Test}: subscript 'Test' refers to test set prediction; *m*: number of objects; *N*: number of variables with nonzero variance; n_{sel}: number of variables selected by the search algorithm; LV: number of principal components used for PCR.

[a] Indicator variable for allyl alcohol.

[b] Full: no variable selection. All nonidentical variables were used.

[c] Selected: model with variable selection (see text).

[d] The entire data was split into a training and a test data set using the Kennard–Stone algorithm.

number of molecules and LV is the number of principal components.

The meaning of the variables is as follows: 'Ds', 'A' and 'H', describe surface points with strong H-bond donor, H-bond acceptor and hydrophilic properties, respectively. The abbreviations used in the QSAR equation describe the characteristics of the surface point pairs of the MaP descriptor as a triplet. The first letter encodes the property of the surface starting point, the second letter gives the property of the surface endpoint and the subscript denotes the mean distance separating the starting point and the endpoint. For example, the variable encoding a strong hydrogen-bond donor surface patch (Ds) in a distance of 3 Å to a strong hydrogen-bond donor surface patch (Ds) is abbreviated as $DsDs_3$. The other property–distance combinations are abbreviated accordingly.

Interpretation of the QSAR equation can be summarized as follows: The first ($DsDs_3$), third (DsH_3) and fourth ($DsDs_4$) variable need to be interpreted as an ensemble. $DsDs_3$ and $DsDs_4$ describe the presence or absence of the allyl alcohol group and DsH_3 discriminates between compounds with one and two allyl alcohol moieties. Since the latter compounds show very high entries for variables one and four (two alcohol groups), DsH_3 compensates for this (negative regression coefficient) since only compounds with two alcohol moieties show high counts for this variable. This is owing to the fact that the hydrophilic surface between the two alcohol groups is larger as compared to compounds with only one such functionality. The third variable AA_9, however, encodes the reduced binding affinity of compounds exhibiting an acceptor functionality close to the quaternary nitrogen (see Fig. 2). The relevance of this variable is supported by previous studies[2] and by the fact that the positive charge is necessary for biological activity. If an acceptor atom, which usually features a high electronegativity, is located close to this quarternary nitrogen, the magnitude of the positive charge is partially reduced. This in turn weakens the ligand–receptor interaction. The complete statistical characterization of the MaP models can be found in Table 3. It can be seen that the results obtained are only slightly better than those with the indicator variable only. This is not surprising since MaP uses three out of the four variables to distinguish between compounds with an alcohol group and those without. Nevertheless, MaP is able to highlight an additional feature of the caracurine V series that has a negative effect on biological activity (the acceptor moieties). Matter of factly, applying a significance test for the comparison of the predictive accuracy based on the leave-one-out cross-validated residuals,[15] reveals that the MaP model is significantly better than the model based on the indicator variable alone ($p = 0.02$, two sided).

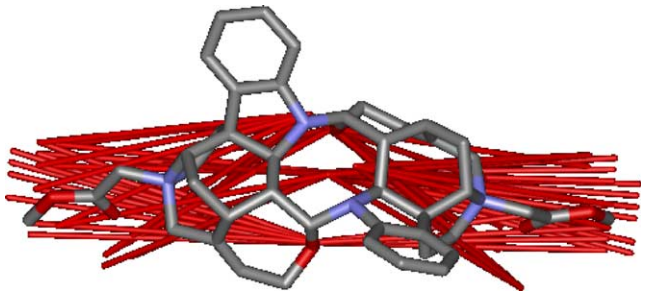


Figure 2. Variable AA_9 back-projected into the original molecular space of **26**. Red lines: starting and ending on acceptor areas. AA_9 describes the detrimental effect of H-bond acceptor functionalities close to the quaternary nitrogen atom, that is, compounds showing high counts for this variable will have a reduced biological activity.

Since variable selection routines are susceptible towards chance correlations,[16] further validation steps were carried out to show the relevance of the derived MaP equation. Firstly, a permutation test was applied to assess the risk of chance correlation. The results of the permutation test for the MaP model are summarized as follows: The 50th- and 95th-percentile of the distribution of $R^2_{CV-40\%}$-values (referred to as $R^2_{CV-40\%,PT}$) for 500 permutations of the response vector are 0.06 and 0.28, respectively. This means that 95% of the data show a $R^2_{CV-40\%,PT}$ of less than 0.28. The maximal $R^2_{CV-40\%,PT}$ was 0.50. Hence, the probability of chance correlation is very low. Secondly, the entire data set was split into a training data set ($m = 25$) and a test set ($m_{Test} = 8$) using the Kennard–Stone algorithm.[17] The training set was then used to select the most relevant variables and the independent test was used to assess the validity of the derived equation. It can be seen in Table 3 that the statistical figures of merit for the reduced training data set as well as for test set prediction ($R^2_{Test} = 0.76$) are quite good, which further underlines the validity of the derived model. It should be noted that the selected variables for the reduced training set ($m = 25$) are almost identical to those that are selected when all molecules ($m = 33$) are used ($DsDs_3$, DsH_5 and AA_9 instead of $DsDs_3$, AA_9, DsH_3 and $DsDs_4$).

In summary, the QSAR studies substantiate the SARs described above. Using the allyl alcohol group alone (indicator variable) approximately 60% of the variance in the nicotinic binding affinity can be explained ($R^2_{CV-1} = 0.61$). In addition to this, the results obtained with MaP show that acceptor groups close to the quarternary nitrogen (e.g., **25**) lower the nicotinic potency to a small extent ($R^2_{CV-1} = 0.73$). This might be owing to the fact that the partial positive charge, which is necessary for biological activity is reduced by this kind of functionality. Interpretation as well as thorough validation of the MaP models gives evidence for their relevance. These findings suggest that no other structural features varied within this data set can be correlated to the respective biological activity.

3. Conclusion

In this study, we have determined the binding constants of 24 caracurine V and 6 *iso*-caracurine V analogues to the muscle type of nicotinic ACh receptors from *Torpedo californica* in a novel binding assay using

(±)-[^{3}H]epibatidine as a radioligand. All bisquaternary *iso*-caracurine V salts **29**–**34** exhibited a higher affinity to the nicotinic ACh receptors than the equally substituted caracurine V analogues, which is due to the presence of the flexible allyl alcohol side chain in the *iso*-caracurine V ring system. The most potent compounds are dimethyl and di-(4-nitrobenzyl)-*iso*-caracurinium V salts **29** (18 nM), and **31** (79 nM), respectively. Compound **29** and the corresponding diallyl analogue **30** (350 nM) exhibited similar binding affinities as the equally substituted neuromuscular-blocking agents toxiferine I **3** (14 nM) and alcuronium **4** (234 nM), respectively. The SAR results were confirmed and extended by QSAR studies. In addition to the prominent alcohol feature, MaP as well as CoMSIA were able to identify an H-bond acceptor moiety close to the quaternary nitrogen as detrimental for nicotinic potency.

All investigated compounds are also ligands for the allosteric binding site of muscarinic M_2 receptors. The diallyl- and dimethylcaracurine V analogues **13** and **27**, respectively, which are among the most potent allosteric modulators of M_2 receptors (EC_{50} = 10 and 8 nM, respectively), exhibited rather low nicotinic binding affinities (K_i = 1.5 and 5.3 μM, respectively). The difference in affinity of such magnitude suggests that it is possible to develop compounds with high muscarinic allosteric potency and negligible neuromuscular-blocking activity. On the other hand, the *iso*-caracurine V analogues with nicotinic binding affinities comparable to those of (+)-tubocurarine and alcuronium, could become a new class of neuromuscular-blocking agents.

4. Experimental

4.1. General procedures

Melting points were determined with a Dr. Tottoli melting point apparatus (Büchi, Switzerland) and were not corrected. ^{1}H and ^{13}C NMR spectra were recorded on a Bruker AV 400 instrument. TLC was carried out on silica gel 60 F_{254} aluminium sheets; Merck. Proton chemical shifts are referred to DMSO-d_6 (2.55 ppm); carbon chemical shifts are referred to DMSO-d_6 (39.50 ppm). IR spectra were obtained using a Biorad PharmalyzIR FT-IR spectrometer. Elemental analyses were performed by the microanalytical section of the Institute of Inorganic Chemistry, University of Würzburg. All reactions were carried out under an argon atmosphere.

4.1.1. General double quaternization procedure of *iso*-caracurine V. The solution of the respective halide (0.85 mmol) in chloroform (5 mL) was added to a solution of *iso*-caracurine V[8] (100 mg, 0.17 mmol) in chloroform (5 mL). After stirring at room temperature for 30 min, the crystallized ammonium salt was isolated by filtration. If no crystallization occurred, the product was precipitated by adding diethyl ether. The collected ammonium salt was washed with chloroform or with a chloroform/diethyl ether mixture 1:1, and dried in vacuum at 80 °C. No further purification was necessary as indicated by TLC (silica gel, mobile phase MeOH/2 M NH_3/2 M aqueous NH_4NO_3 84:24:12) and ^{1}H NMR spectra.

Melting points of all *iso*-caracurinium V salts are higher than 300 °C (dec).

^{1}H chemical shifts and coupling constants for hydrogen atoms belonging to the *iso*-caracurine V ring system of **33** and **34** coincide with the δ-values of the corresponding atoms of **31** within ±0.05 ppm and ±0.1 Hz, respectively. ^{13}C chemical shifts for carbon atoms belonging to the caracurine V ring system of **33** and **34** coincide with the δ-values of the corresponding atoms of **31** within ±0.5 ppm.

4.1.2. 4,4′-Di-(4-nitrobenzyl)-*iso*-caracurinium V dibromide (31). Obtained from *iso*-caracurine V and 4-nitrobenzyl bromide (184 mg) as a yellow solid; yield 145 mg (82%); $[\alpha]_D^{22}$ +40.4 (*c* 1.2, DMSO); IR (ATR) 3375 br, 2947, 1603, 1598, 1524, 1485, 1459, 1348, 851, 744 cm^{-1}; ^{1}H NMR (400 MHz, DMSO-d_6): δ 1.67 (d, br, *J* = 14.6 Hz, 1H, H-14a′), 1.82 (d, br, *J* = 14.6 Hz, 1H, H-14a), 2.09 (m, 2H, H-6a, 6a′), 2.27 (m, 2H, H-6b, 16), 2.60 (m, 3H, H-14b, 14b′, 6b′), 3.47 (m, br, 2H, H-15, 15′), 3.69 (d, *J* = 13.6 Hz, 1H, H-21a), 3.77 (d, *J* = 14.1 Hz, 1H, H-21a′), 3.82 (s, 1H, H-2), 3.95 (m, 2H, 5-CH_2), 4.05 (m, 2H, 5′-CH_2), 4.20 (m, 5H, 18′-CH_2, 18-CH_2, 21b′), 4.51 (d, *J* = 13.6 Hz, 1H, H-21b), 4.63 (s, br, 1H, H-3′), 4.73 (s, br, 1H, H-3), 4.80 (t, *J* = 5.1 Hz, 1H, –OH), 4.83 (s, 1H, H-2′), 5.01 (s, br, 1H, H-17), 5.06 (m, 4H, N–CH_2), 5.87 (t, *J* = 5.7 Hz, 1H, H-19′), 6.24 (s, br, 1H, H-17′), 6.40 (m, 1H, H-19), 6.65 (d, *J* = 7.6 Hz, 1H, H-12′), 6.78 (t, *J* = 7.6 Hz, 1H, H-12), 6.93 (t, *J* = 7.6 Hz, 1H, H-10′), 7.01 (t, *J* = 7.6 Hz, 1H, H-10), 7.22 (d, *J* = 7.6 Hz, 1H, H-11), 7.26 (t, *J* = 7.6 Hz, 1H, H-11′), 7.41 (d, *J* = 7.6 Hz, 1H, H-9′), 7.53 (d, *J* = 7.6 Hz, 1H, H-9), 8.22 (d, *J* = 8.8 Hz, 2H), 8.46 (d, *J* = 8.8 Hz, 2H), 8.47 (d, *J* = 8.8 Hz, 4H); ^{13}C NMR (100 MHz, DMSO-d_6): δ 23.2 (C-14, C-14′), 29.7 (C-15), 34.6 (C-15′), 37.9 (C-6′), 38.1 (C-6), 49.6 (C-16), 55.1 (C-7), 55.3 (C-7′), 56.9 (C-18′), 58.2 (C-21), 58.4 (C-5), 58.6 (C-5′), 59.0 (C-2′), 60.8 (C-21′), 63.9 (N^+-CH_2), 64.2 (C-2), 64.3 (N^+-CH_2), 65.9 (C-18), 73.5 (C-3′), 74.4 (C-3), 94.0 (C-17), 110.8 (C-12′), 111.6 (C-12), 119.8 (C-16′), 120.1 (C-10′), 120.6 (C-10), 122.2 (C-9′), 122.5 (C-9), 128.7 (C-17′), 129.4 (C-11′), 129.6 (C-11), 129.78 (C-8, C-8′), 131.0 (C-20′), 133.8 (C-20), 134.7 (C-19′), 134.8 (C-19), 147.8 (C-13′), 151.4 (C-13), 124.0, 135.2, 135.7, 136.7, 148.7, 148.8 (benzyl groups). Anal. Calcd for $C_{54}H_{52}N_6O_6Br_2 2H_2O$: C, 60.23; H, 5.24; N, 7.80. Found: C, 59.99; H, 5.54; N, 8.21.

4.1.3. 4,4′-Bis-(3,4-dimethoxybenzyl)-*iso*-caracurinium V dichloride (33). Obtained from *iso*-caracurine V and 3,4-dimethoxybenzyl chloride[9] (160 mg) as a white solid; yield 110 mg (67%); $[\alpha]_D^{22}$ +42.5 (*c* 1, DMSO); IR (ATR) 3350, br, 2947, 2877, 2837, 1658, 1599, 1518, 1487, 1460, 1268, 1149, 1109, 819, 749 cm^{-1}; ^{1}H NMR (400 MHz, DMSO-d_6): δ 3.86 (s, 6H, –CH_3), 3.89 (s, 6H, –CH_3), 4.77 (s, br 2H, N^+-CH_2), 4.82 (s, br 2H, N^+-CH_2), 7.19 (m, 4H), 7.46 (d, *J* = 1.3 Hz, 1H), 7.55

(d, $J = 1.3$ Hz, 1H); ^{13}C NMR (100 MHz, DMSO-d_6): δ 55.1 and 55.5 (CH_3), 67.3 (N^+-CH_2), 111.7, 116.3, 116.5, 119.9, 129.7, 148.7, 150.2, 150.4, Anal. Calcd for $C_{56}H_{62}N_4O_6Cl_2H_2O$: C, 68.81; H, 6.61; N, 5.74. Found: C, 68.40; H, 6.58; N, 5.84.

4.1.4. 4,4′-Dibenzyl-*iso*-caracurinium V dibromide (34). Obtained from *iso*-caracurine V and benzyl bromide (145 mg) as a white solid; yield 125 mg (78%); $[\alpha]_D^{22}$ −124 (*c* 0.5, DMSO); IR (ATR) 3380, br, 2952, 2880, 1654, 1597, 1488, 1459, 1034, 745, 708 cm^{-1}; ^{1}H NMR (400 MHz, DMSO-d_6): δ 4.66 (s, 2H, N^+-CH_2), 4.80 (s, 2H, N^+-CH_2), 7.59 (m, 6H), 7.80 (m, 2H), 7.86 (m, 2H), ^{13}C NMR (100 MHz, DMSO-d_6): δ 67.2, 68.0 (N^+-CH_2), 128.5, 128.9, 133.9, 134.5; Anal. Calcd for $C_{52}H_{54}N_4O_2Br_23H_2O$: C, 63.67; H, 6.17; N, 5.71. Found: C, 63.60; H, 5.82; N, 5.82.

4.2. Radioligand binding studies

(±)-[^{3}H]Epibatidine (33–66 Ci/mmol) was obtained from NEN Life Science Products (Cologne, Germany). All other chemicals were obtained from Sigma–Aldrich (Deisenhofen, Germany). Frozen *Torpedo californica* electric organ was purchased from Marinus Inc. (Long Beach, CA, USA).

Frozen samples of *Torpedo californica* electric organ were thawed for 30–60 min before membrane preparation. Total membrane fractions were isolated by homogenization with a polytron. The tissue was homogenized in ice-cold HEPES-salt solution (HSS) containing HEPES (pH 7.4, 15 mM), 120 mM NaCl, 5.4 mM KCl, 0.8 mM $MgCl_2$ and 1.8 mM $CaCl_2$ followed by centrifugation at 30,000*g* for 10 min at 4 °C. The pellets were collected, washed four times with HSS through rehomogenization and centrifugation at the same settings. The remaining pellets were collected, resuspended in HSS and stored in aliquots at −80 °C. For each assay the samples of membrane fractions were thawed, homogenized and centrifuged at 30,000*g* for 10 min at 4 °C. The resultant pellets were resuspended in HSS and used in binding assays.

Assays were carried out in HSS at 22 °C. Each assay was performed in duplicates. Nonspecific binding was determined in the presence of 300 μM (−)-nicotine. Membranes were incubated for 90 min in 0.5 mL HSS containing 0.5 nM (±)-[^{3}H]epibatidine and different concentrations of test compounds. The reaction was terminated by vacuum filtration through Whatman GF/B glass fibre filters, pre-soaked in 1% poly(ethylenimine) using a Brandel 48-channel cell harvester. The radioactivity was measured using a liquid scintillation counter (Tri-Carb 2100 TR; Packard, Dreieich, Germany).

Competition binding data were analyzed using nonlinear regression methods. K_i values were calculated by the Cheng–Prusoff equation ($K_i = IC_{50}/(1+F/K_d)$, where F is the used radioligand concentration) based on the measured IC_{50} values and $K_d = 2$ nM for binding of (±)-[^{3}H]epibatidine. The K_d values were obtained from five independent experiments performed on the same membrane preparations that were used for the competition assays.

4.3. Molecular modelling

All molecules were built and optimized according to Zlotos et al.[2,13] Briefly, template structures were chosen for each structural class and these were aligned at the corresponding quaternary nitrogen atoms and the centroids of the aromatic ring systems. Based on these templates the other structures were built with the backbones fixed and only the *N*-substituents were geometry optimized. Here, for the CoMSIA study substituents were built to give a high conformational similarity, whereas for MaP no such alignment step was used. For the latter procedure the SMILES strings of the substituents were automatically converted to 3D-geometries and then attached to the quaternary nitrogen atom within DS ViewerPro (DS ViewerPro 5.0, Accelrys Inc., San Diego, CA, USA). Afterwards the first local energy minimum obtained with the TRIPOS force field was used. After this minimization all substituents occupied a similar volume in space. Since the structural backbones of the compounds are very rigid, no conformational search on these backbones was performed. Afterwards AM1 charges were calculated for all structures (MOPAC keywords: 1SCF AM1 MMOK).

The binary variable describing the presence or absence of a free allyl alcohol group was generated by hand (0 = no free allyl alcohol group; 1 = at least one free allyl alcohol group). This variable was correlated to the biological activity using linear regression.

CoMSIA fields were generated using SYBYL version 6.7.[18] The CoMSIA region was automatically defined and all available fields (steric, electrostatic, acceptor, donor and hydrophobic) were calculated. Coefficient contour maps showing the product of standard deviation and regression coefficient ('StdDev * Coeff') were used for interpretation purposes. To identify appropriate contour levels, field value histograms for each feature were analyzed and levels that gave meaningful results were applied.

Here, only the core calculation of the MaP descriptor is given. A more detailed description of the procedure can be seen in Ref. 4. A three step procedure is used to compute the MaP descriptor (see Fig. 3).

First, an approximation to the molecular surface with equally distributed surface points is computed. With default settings, which were applied here, the surface points are 0.8 Å apart. This distance is referred to as grid spacing. Next, molecular properties (currently: H-bond donor (D), H-bond acceptor (A), hydrophilicity (H) and hydrophobicity (L)) are projected onto this surface. Finally, the distribution of surface properties is encoded into a translationally and rotationally invariant molecular descriptor. The algorithm generating the MaP descriptor is based on the idea that the occurrence of particular surface point pairs will be different in active and inactive substances. Surface point pairs are

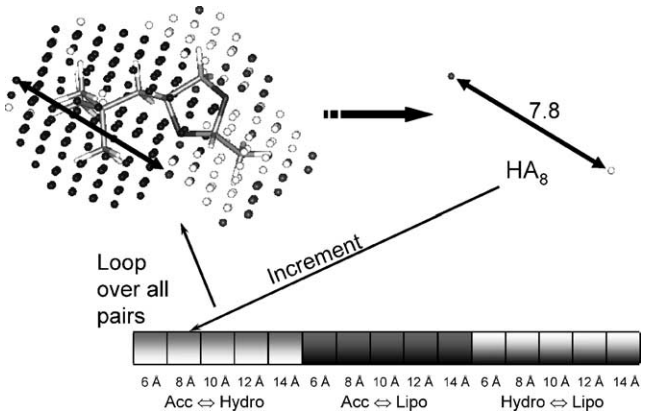


Figure 3. Work-flow of the MaP descriptor calculation. Starting from equally distributed surface points a vector for all possible property combinations and distances is generated. Next, a surface point pair is chosen and the respective property–property–distance combination is identified (here hydrophilic-acceptor-8, HA_8) and incremented. For this, a fuzzy incrementation method is applied. This is repeated for all surface point pairs.

characterized by the Euclidean distance separating the two points and the property combination of the two points. Hence, the MaP descriptor consists of a set of distance histograms for all possible property combinations (A ↔ A, A ↔ D, etc.). According to these features, surface point pairs are grouped into bins. A bin is characterized by a lower and an upper distance. The distance range depends on the user-defined resolution, which is set by default to 1 Å. The first bin ranges from 0.5 to 1.5 Å, the second from >1.5 to 2.5 Å, and so forth, up to the largest distance in the data set. A loop over all surface point pairs is initiated, where each pair is visited only once. If a point pair matches a particular bin (property–property–distance combination), the counter of the respective bin and the nearest neighbouring bin is incremented proportionally (fuzzy counts).[4]

MaP was applied using its default parameter settings (see Table 4).

Proximity distance $d_{\text{cut-off}}$ and Δd control the distance dependent mapping of the atomic hydrophilic and hydrophobic features onto the surface points. The resulting value of a single surface point, which depends on all atomic properties and their respective distance to that particular surface point, is referred to as molecular surface lipophilic potential. Atomic hydrophilicity/hydrophobicity is computed according to the fragmental approach of Ghose et al.[19] The parameters of $d_{\text{cut-off}}$ and Δd are chosen in a way as to reflect local features of the molecule.[5] The two further cut-off values given in Table 4 denote the decision boundary for the molecular surface lipophilic potential of a surface point. Depending on its actual value the surface is point is categorized as hydrophilic (H), weakly hydrophobic (Lw) or strongly hydrophobic (Ls). H-bond donor and H-bond acceptor properties are assigned to surface points if the nearest atom to this point is considered to be a H-bond donor or a H-bond acceptor, respectively. To distinguish between different H-bond donor strengths, an additional cut-off value based on the partial charge of the hydrogen atom (+0.18) connected to the respective heteroatom was incorporated into the parameter set.[5]

Table 4. Default parameters of MaP

Parameter	Value
Default	
Grid spacing	0.8 Å
Resolution (res) of the radial distribution function	1 Å
Proximity distance $d_{\text{cut-off}}$	2 Å
Δd	2 Å
Hydrophilic/hydrophobic cut-off	0
Weak hydrophobic/strong hydrophobic cut-off	0.12

Principal component regression (PCR) is used to correlate the MaP encoded molecules with their respective biological activities. The descriptors were mean-centred prior to analysis. The number of latent variables for the full model is determined by leave-one-out cross-validation (LOO-CV). In order to identify important variables, forward selection in combination with PCR and leave-multiple-out cross-validation (LMO-CV) as the objective function is employed.[4,16,20] The leave-multiple-out cross-validated statistical figures of merit $R^2_{\text{CV}-k}$ (cross-validated coefficient of determination), and $\text{RMSEP}_{\text{CV}-k}$ (cross-validated root mean squared error of prediction) were averaged over $3m$ splits into a construction data set of size $m–k$ and a validation data set of size k, where m is the number of molecules in the data set. k was set to the nearest integer of $m \cdot 0.4$ for all experiments.[21] LMO-CV is used to simultaneously determine the variable subset and the number of latent variables for PCR.[20,21]

For variable selection a maximum number of four variables was allowed to be selected to reduce the risk of chance correlation.[5,16] In order to check for the risk of chance correlation a permutation test based on the repetitive randomization of the response vector was employed. In each cycle of the test the response vector is randomly rearranged, the entire selection procedure is carried out on the scrambled data (same settings as for the variable selection above), and the leave-k-out cross-validated squared multiple correlation coefficient $R^2_{\text{CV}-k}$ is recorded for each cycle. All computations were done from scratch after scrambling the responses since scrambling only the finally selected model yields far overoptimistic results.[16] If the majority of the $R^2_{\text{CV}-k}$ values of the scrambled data sets is lower than the $R^2_{\text{CV}-k}$ value of the original data set it is concluded that the derived model is relevant. To further validate the results of the variable selection routine the entire data set was split into a training set of size $m = 25$ and a test set of size $m_{\text{Test}} = 8$ using Kennard–Stone's CADEX algorithm[17] on the full descriptor data. The CADEX algorithm maximizes the minimal Euclidean distances between already selected objects and the remaining objects. The selection starts with the two most distant objects using the Euclidean distance. For each of the remaining objects the shortest Euclidean distance to the already selected objects is computed and stored in a distance list. Next, the object with maximum distance

in the distance list is selected. This procedure is repeated until enough objects are selected. The CADEX algorithm results in a balanced and representative split of the data. It was applied since it had performed well in training set selection in other studies.[5,21,22]

Acknowledgements

Thanks are due to the Fonds der Chemischen Industrie, Deutschland for financial support.

References and notes

1. Zlotos, D. P.; Buller, S.; Tränkle, C.; Mohr, K. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 2529.
2. Zlotos, D. P.; Buller, S.; Stiefl, N.; Baumann, K.; Mohr, K. *J. Med. Chem.* **2004**, *47*, 3561.
3. Berlage, F.; Bernauer, K.; von Philipsborn, W.; Waser, P.; Schmidt, H.; Karrer, P. *Helv. Chim. Acta* **1959**, *42*, 394.
4. Stiefl, N.; Baumann, K. *J. Med. Chem.* **2003**, *46*, 1390.
5. Stiefl, N.; Bringmann, G.; Rummey, C.; Baumann, K. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 347.
6. Battersby, A. R.; Hodson, H. F. *J. Chem. Soc.* **1960**, 736.
7. Zlotos, D. P. *J. Nat. Prod.* **2000**, *63*, 864.
8. Zlotos, D. P. *J. Nat. Prod.* **2003**, *66*, 119.
9. Zlotos, D. P.; Meise, W. *Heterocycles* **1997**, *45*, 2137.
10. Verpoorte, R.; Baerheim Svendsen, A. *J. Pharm. Sci.* **1978**, *67*, 171.
11. Mukhin, A. G.; Gündisch, D.; Horti, A. G.; Koren, A. O.; Tamagnan, G.; Kimes, A. S.; Chambers, J.; Vaupel, D. B.; King, S.; Picciotto, M. R.; Innis, R.; London, E. D. *Mol. Pharmacol.* **2000**, *57*, 642.
12. Lee, C. *Br. J. Anaesth.* **2001**, *87*, 755.
13. Zlotos, D. P. *Eur. J. Org. Chem.* **2004**, *11*, 2375.
14. Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130.
15. Van der Voet, H. *Chemom. Intell. Lab. Sys.* **1994**, *25*, 313.
16. Baumann, K. *Trends Anal. Chem. (TrAC)* **2003**, *22*, 395.
17. Kennard, R. W.; Stone, L. A. *Technometrics* **1969**, *11*, 137.
18. SYBYL® 6.7.1 Tripos Inc., 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA, 2000.
19. Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. *J. Phys. Chem. A* **1998**, *102*, 3762.
20. Baumann, K.; von Korff, M.; Albert, H. *J. Chemom.* **2002**, *16*, 339.
21. Baumann, K.; Albert, H.; von Korff, M. *J. Chemom.* **2002**, *16*, 350.
22. Wu, W.; Walczak, B.; Massart, D. L.; Heuerding, S.; Erni, F.; Last, I. R.; Prebble, K. A. *Chemom. Intell. Lab. Sys.* **1996**, *33*, 35.