Pergamon

Bioorganic & Medicinal Chemistry 7 (1999) 2621–2628

BIOORGANIC & MEDICINAL CHEMISTRY

Three-Dimensional Quantitative Structure–Activity Studies of Octopaminergic Agonists Responsible for the Inhibition of Sex-Pheromone Production in *Hercoverpa armigera*

Akinori Hirashima,[a,*] Ada Rafaeli,[b] Carina Gileadi[b] and Eiichi Kuwano[a]

[a]*Division of Bioresource and Bioenvironmental Sciences, Graduate School, Kyushu University, Fukuoka 812-8581, Japan*
[b]*Department of Stored Products, Pheromone Research Laboratory, Volcani Centre, PO Box 6, Bet Dagan 50250, Israel*

Received 28 April 1999; accepted 8 July 1999

Abstract—The quantitative structure–activity relationship (QSAR) of octopaminergic agonists responsible for the inhibition of sex-pheromone production in *Hercoverpa armigera*, was analyzed using physicochemical parameters, molecular shape analysis (MSA), molecular field analysis (MFA), and receptor surface model (RSM), respectively. The dose-response studies were performed in vitro analyzing the effect of these compounds on intracellular cAMP production in the presence of pheromone biosynthesis activating neuropeptide (PBAN) at 1 pmol/intersegment. Six active derivatives were identified in the order of decreasing pheromonostatic activity: 2-(2,6-dimethylanilino)imidazolide (**6**) > 2-(2-methyl-4-chloroanilino)oxazolidine (**1**) > clonidine (**5**) > 2-(2,6-diethylanilino)thiazolidine (**8**) > 2-(3,5-dichlorobenzylamino)-2-oxazoline (**4**) > tolazoline (**10**) which were all active in the nanomolar range in inhibition of cAMP production by 1 pmol PBAN/intersegment. Four other compounds were less active having K_i in the micromolar range. An MSA was tried to obtain QSAR equation that incorporates spatial molecular similarity data of those compounds. MFA on the training set of those compounds evaluated effectively the energy between a probe and a molecular model at a series of points defined by a rectangular or spherical grid. An RSM was generated using some subset of the most active structures. Three-dimensional energetics descriptors were calculated from RSM/ligand interaction and these three-dimensional descriptors were used in QSAR analysis. These results indicate that these derivatives could provide useful information in the characterization and differentiation of octopaminergic receptor types and subtypes.

Introduction

Octopamine (OA), which has been found to present in high concentrations in various insect tissues, is the monohydroxyllic analogue of the vertebrate hormone noradrenaline. It has been well established that OA acts as neurohormone, neurotransmitter, and neuromodulator in insects.[1] Three different receptor classes OAR1, OAR2A, and OAR2B had been distinguished from non-neuronal tissues.[2] In the nervous system of migratory locust *Locusta migratoria* L., a particular receptor class was characterized and established as a new class OAR3 by pharmacological investigations of the [^{3}H]OA binding site using various agonists and antagonists.[3–7] Structure–activity studies of various types of OA-agonists and antagonists were also reported using the nervous tissue of *L. migratoria*.[8–11]

Production of the pheromone blend is under the regulation of a neuropeptide termed pheromone biosynthesis activating neuropeptide (PBAN).[12–19] The direct action of PBAN has been demonstrated by studies in vitro[20–25] showing stimulation of pheromone production in the presence of synthetic peptide by isolated pheromone gland tissue. The exact tissue involved was delineated as the intersegmental membrane which is situated between the 8th and 9th abdominal segments.[26,27] In *Helicoverpa armigera*, we have shown that OA, tyramine, and clonidine significantly inhibit the pheromonotropic action due to PBAN in intact moths and decapitated moths, as well as pheromone gland incubations in vitro.[26–28] The inhibition was also reflected in a significant inhibitory effect on intracellular cAMP levels which were stimulated in the presence of PBAN. This inhibitory action is a result of a receptor (separate from the PBAN-receptor) which can be inhibited by pertussis toxin.[28] This provided evidence that this specific pheromonostatic–aminergic receptor is linked to a G-inhibitory protein. The pheromonostatic receptor, acting in a neuromodulatory role, represents a

Key words: *Hercoverpa armigera*; octopaminergic agonist; receptor surface model; molecular field analysis; molecular shape analysis.
* Corresponding author. Tel.: +81-92-642-2856; fax: +81-92-642-2864/642-2804; e-mail: ahirasim@agr.kyushu-u.ac.jp

0968-0896/99/$ - see front matter
PII: S0968-0896(99)00189-3

novel type of octopaminergic receptor which induces an inhibitory and not a stimulatory action on adenylate cyclase. It is therefore of critical importance to provide information on the pharmacological properties of this OA receptor types and subtypes. This report deals with the three-dimensional quantitative structure–activity relationship (QSAR) of octopaminergic agonists responsible for the inhibition of sex-pheromone production in *H. armigera*, using physicochemical parameters, molecular shape analysis (MSA), molecular field analysis (MFA), and receptor surface model (RSM), respectively.

Results and Discussion

Six active derivatives were identified in order of decreasing pheromonostatic activity: 2-(2,6-dimethylanilino)imidazolide (**6**) > 2-(2-methyl-4-chloroanilino)-oxazolidine (**1**) > clonidine (**5**) > 2-(2,6-diethylanilino)-thiazolidine (**8**) > 2-(3,5-dichlorobenzylamino)-2-oxazoline (**4**) > tolazoline (**10**) which were all active in the nanomolar range in OA-agonist activity responsible for the inhibition PBAN-stimulated cAMP production in *H. armigera* (Table 1). Pheromonostatic activity of these OA agonists showed a good correlation with an inhibition of PBAN-stimulated cAMP production (unpublished data). Four other compounds were less active having K_i in the micromolar range. K_i is the concentration of OA agonist necessary for half-maximal inhibition of PBAN binding at 1 pmol/intersegment. In order to quantitatively understand the dependence of biological activities on physicoparameters of OA agonists, regression analysis was applied to representative compounds listed in Figure 1a and Table 1, leading to eq (1). Preliminary tests indicate that if the genetic function algorithm (GFA) method is used, non-linear terms must be included. Hence, GFA was used with non-linear terms for all following QSAR calculations.

$$pK_i = 3.08501 - 0.081986LowEne - 14.8188SXZF + 8.81828JX + 0.004254(MR - 57.7808)^2 \quad (1)$$

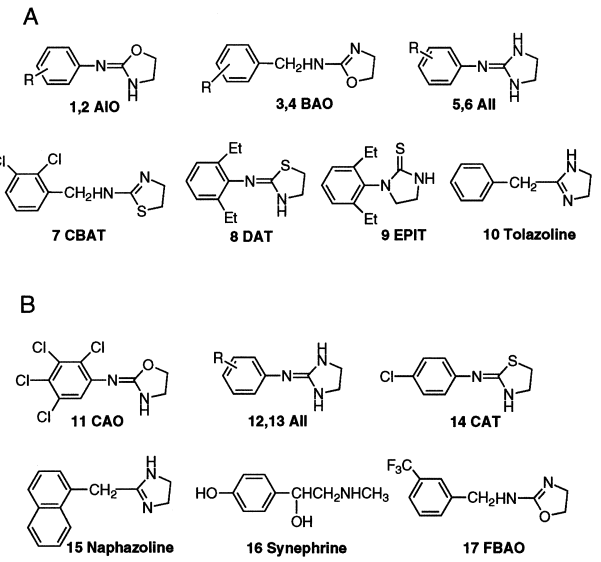

Figure 1. Structures of OA agonists used for regression analysis in (A) training and (B) test sets.

where $n = 10$, $r^2 = 1.000$, $F = 5681.367$, $CV{-}r^2 = 0.998$, and $Bsr^2 = 1.000 \pm 0.000$: *r*-squared (r^2), the square of the correlation coefficient; cross-validated r^2 ($CV{-}r^2$), a squared correlation coefficient generated during a validation procedure using the equation; bootstrap r^2 (Bsr^2), the average squared correlation coefficient calculated during the validation procedure; *n*, the number of data points; *r*, the correlation coefficient; *F*, the value of the *F*-test. The r^2 is used to describe the goodness of fit of the data in the study table to the QSAR model. A $CV{-}r^2$ is usually smaller than the overall r^2 for a QSAR equation. It is used as a diagnostic tool to evaluate the predictive power of an equation generated using the multiple linear regression or partial least squares method. A Bsr^2 is computed from the subset of variables used one-at-a-time for the validation procedure. It can be used more than one time in computing the r^2 statistic. Table 1 depicts structures of compounds, their experimental K_i values, their pK_i values, calculated pK_i

Table 1. Regression analysis of structure-OA agonist activity using conventional descriptors

Compound		K_i (nM)	pK_i			*LowEne*[b]	*SXZF*[c]	*JX*[d]	*MR*[e]
No.	*R*		Observed	Calculated[a]	Deviation				
1	AIO 2-CH_3,4-Cl	6.15	8.21	8.20	0.01	8.848	0.688	2.183	55.6
2	AIO 2-CH_3,6-CH_2CH_3	6150	5.21	5.23	−0.02	63.708	0.871	2.296	60.4
3	BAO 2-F	100,000	4.00	3.99	0.01	49.976	0.782	1.859	50.8
4	BAO 3,5-Cl_2	69.2	7.16	7.17	−0.01	−5.690	0.880	1.886	60.2
5	AII 2,6-Cl_2	8.97	8.05	8.00	0.05	52.652	0.714	2.247	57.3
6	AII 2,6-$(CH_3)_2$	3.85	8.41	8.43	−0.02	37.667	0.795	2.292	57.8
7	CBAT	76,900	4.11	4.11	0.00	73.526	0.689	1.921	66.6
8	DAT	61.5	7.21	7.25	−0.04	88.279	0.671	2.329	71.5
9	EPIT	5380	5.27	5.24	0.03	106.179	0.777	2.410	74.0
10	Tolazoline	76.9	7.11	7.11	0.00	13.907	0.861	1.997	49.1

[a] Calculated by eq (1).
[b] Conformational descriptor *LowEne* is the energy of the most stable conformation in the set of conformations belonging to each molecular model.
[c] The descriptor *SXZF* is the fraction of area of molecular shadow in the XZ plane over area of enclosing rectangle.
[d] The descriptor *JX* is a highly discriminating descriptor, whose values do not substantially increase with the molecule size and the number of rings present.
[e] Thermodynamic descriptor *MR* index of a substituent is a combined measure of its size and polarizability which was calculated using the atom typing scheme of Ghose et al.[30]

values using eq (1), difference between observed and calculated pK_i values, and descriptors used here to obtain eq (1). Values of pK_i, the logs of the reciprocals of K_i, were used as OA-agonist activity index. Conformational descriptor *LowEne* is the energy of the most stable conformation in the set of conformations belonging to each molecular model. Topological indices including Balaban index *JX* are 2-D descriptors based on graph theory concepts. These indices have been widely used in QSAR studies. They help to differentiate the molecules according mostly to their size, degree of branching, flexibility, and overall shape. Balaban index *JX* is a highly discriminating descriptor, whose values do not substantially increase with the molecule size and the number of rings present. The geometric and spatial descriptor of shadow index helps to characterize the shape of the molecules. The descriptors are calculated by projecting the molecular surface on three mutually perpendicular planes, XY, YZ, and XZ.[29] These descriptors depend not only on conformation but also on the orientation of the molecule. To calculate them, the molecules are first rotated to align the principal moments of inertia with the X, Y, and Z axes. Fraction of area of molecular shadow in the XZ plane over area of enclosing rectangle (*SXZF*). Thermodynamic descriptor molecular refractivity (*MR*) index of a substituent is a combined measure of its size and polarizability which was calculated using the atom typing scheme of Ghose et al.[30]

According to eq (1), the positive *JX*, and the negative *LowEne* and *SXZF* terms mean that the larger *JX* values, and the smaller *SXZF* and conformational energy *LowEne*, the greater the activity. The positive MR^2 term indicates that the OA-agonist activity is mimimun when *MR* value is 57.7808. The use of other parameters instead of or the addition of other parameters to eq (1) did not improve the correlation, e.g. fragment constant descriptors that relate the effect of substituents on a ''reaction center'' from one type of process to another such as, σm and σp (electronic effect sigma meta and sigma para, respectively), *F* and *R* [decompositions of sigma para constant into an inductive (polar) part (*F*) and a resonance part (*R*) for the case when the substituent is conjugated with the reaction center producing through-resonance effects], pi (hydrophobic character), HA (hydrogen bond acceptor), HB (hydrogen bond donor), Sterimol-L (steric length parameter), Sterimol-B1 through B4 (steric distances perpendicular to the bond axis), and Sterimol-B5 (the overall maximum steric distance perpendicular to the bond axis); conformational descriptors such as, Energy (the energy of the currently selected conformation in the study) and EPenalty (the difference between Energy and *LowEne*); electronic descriptors such as, Charge (sum of partial charges), Fcharge (sum of formal charges), Apol (sum of atomic polarizabilities), Dipole (dipole moment), HOMO (highest occupied molecular orbital energy), LUMO (lowest unoccupied molecular orbital energy), and Sr (superdelocalizability); topological descriptors (2-D descriptors based on graph theory concepts) such as, Zagreb Index (the sum of the squares of vertex valencies), Hosoya Index [the sum of all (nonzero) $p(k)$, the number of ways of choosing k non-adjacent edges from the graph], and the Molecular Connectivity Index; spatial descriptors such as, radius of gyration, Jurs charged partial surface area descriptors, molecular surface area, density, principal moment of inertia, and molecular volume; structural descriptors such as, MW (molecular weight), Rotlbonds (number of rotatable bonds), Hbond acceptor (number of hydrogen bond acceptors), Hbond donor (number of hydrogen bond donors); thermodynamic descriptors such as, AlogP (log of the partition coefficient), Fh2o (desolvation free energy for water), Foct (desolvation free energy for octanol), and Hf (heat of formation).

MSA

The goal of MSA[31,32] is to generate a QSAR equation that incorporates spatial molecular similarity data. The process has been described by Hopfinger and Burke.[33] The outcome of the MSA process is an optimized QSAR that can be used for activity estimation and ligand evaluation. The set of choices available for each task is employed to generate trial QSARs. The QSAR that corresponds to the best fit between observed activities and computed molecular descriptors defines the specific requirements for each MSA task. The ability to minimize a structure against a model allows one to flexibility fit a structure into the model. The minimization can also be used as a shape- (and not atom-)based alignment technique. Applying the fitting process to a set of molecules will force all the molecules to adopt a shape that is consistent with the model, leading to eq (2).

$$pK_i = 39.1401 - 0.960171(IC)^2 - 6.57924RG - 0.05538(152.765 - COSV)^2 \quad (2)$$

where $n = 10$, $r^2 = 0.984$, $F = 119.504$, $CV - r^2 = 0.961$, and $Bsr^2 = 0.984 \pm 0$. The descriptors generated for MSA, the activities of the compounds, and the predictions using their top model are shown in Table 2. Common overlap steric volume (*COSV*) is the common volume between each individual molecule and the molecule selected as the reference compound. This is a measure of how similar in steric shape the analogues are to the shape reference. Spatial descriptor radius of gyration (*RG*) is calculated using the following equation: $RG = \sqrt{\Sigma(xi + yi + zi)/N}$, where N is the number of atoms and x, y, and z are the atomic coordinates relative to the center of mass. In the approach of information-content (*IC*) descriptors, molecules are viewed as structures which can be partitioned into subsets of elements that are in some sense equivalent. The notion of equivalence depends on the particular descriptor. According to eq (2), the negative $(COSV)^2$ term indicates that the OA-agonist activity is maximun when *COSV* value is 152.765.

MFA

MFA evaluates the energy between a probe and a molecular model at a series of points defined by a rectangular

Table 2. MSA descriptors and prediction using top model

No.	IC[a]	RG[b]	$COSV$[c]	pK_i Observed	pK_i Calculated[d]	pK_i Deviation
1	3.236	3.212	152.765	8.21	7.95	0.26
2	3.507	3.340	170.173	5.21	5.36	−0.15
3	3.182	3.399	96.693	4.00	3.95	0.05
4	3.006	3.177	109.328	7.16	7.16	0.00
5	3.093	3.228	140.051	8.05	8.01	0.04
6	3.093	3.210	182.187	8.41	8.84	−0.43
7	3.323	3.374	116.155	4.11	4.31	−0.20
8	3.250	3.360	170.221	7.21	6.89	0.32
9	3.500	3.248	138.431	5.27	5.22	0.05
10	3.022	3.044	93.439	7.11	7.06	0.05

[a] In the approach of *IC* descriptors, molecules are viewed as structures which can be partitioned into subsets of elements that are in some sense equivalent.
[b] Spatial descriptor *RG* is calculated using the following equation: $RG = \sqrt{\Sigma(xi + yi + zi)/N}$, where N is the number of atoms and x, y, and z are the atomic coordinates relative to the center of mass.
[c] *COSV* is the common volume between each individual molecule and the molecule selected as the reference compound.
[d] Calculated by eq (2).

or spherical grid. These energies were added to the study table to form new columns headed according to the probe type. The new columns were used as independent X variables in the generation of QSAR, leading to eq (3).

$$\mathrm{p}K_\mathrm{i} = 7.20883 - 0.00301(H+/372)^2 + 0.062649(H+/298) + 0.02519CH3/164 \quad (3)$$

where $n = 10$, $r^2 = 0.995$, $F = 407.918$, $CV - r^2 = 0.986$, and $Bsr^2 = 0.995 \pm 0$. The descriptors generated for MFA, the activities of the compounds, and the predictions using their top model are shown in Table 3. The surface is generated from a "Shape Field". The atomic coordinates of the contributing models are used to compute field values on each point of a 3-D grid. The descriptors $H+/372$ and $H+/298$ are the energy between proton probe and the molecule at the rectangular points 372 and 298, respectively. The descriptor $CH3/164$ is the energy between methyl probe and the molecule at the rectangular point 164. According to eq (3), the negative $(H+/372)^2$ term indicates that the OA-agonist activity is maximum when $H+/372$ value is zero.

Table 3. MFA descriptors and prediction using top model

No.	$H+/372$[a]	$H+/298$[a]	$CH3/164$[b]	pK_i Observed	pK_i Calculated[c]	pK_i Deviation
1	−28.25	13.657	−0.908	8.21	8.02	0.19
2	30.000	−1.637	30.000	5.21	5.15	0.06
3	−30.000	−10.681	−0.492	4.00	3.82	0.18
4	2.939	0.895	−0.671	7.16	7.22	−0.06
5	−2.272	0.936	30.000	8.05	8.01	0.04
6	−14.115	30.000	−1.039	8.41	8.46	−0.05
7	−24.675	−19.321	2.269	4.11	4.22	−0.11
8	5.764	2.039	−0.614	7.21	7.22	−0.01
9	30.000	7.760	17.229	5.27	5.42	−0.15
10	−1.380	0.038	−0.482	7.11	7.19	−0.08

[a] The descriptors $H+/372$ and $H+/298$ are the energy between proton probe and the molecule at the rectangular points 372 and 298, respectively.
[b] The descriptor $CH3/164$ is the energy between methyl probe and the molecule at the rectangular point 164.
[c] Calculated by eq (3).

Figure 2 shows the most potential OA agonist **6** and the least potential OA agonist **3** embedded in a MFA generated from training set of OA agonists and the points used in the resulting QSAR model. Only points contributing to the equation are represented (a green sign stands for a positive contribution). Note here that all the points map in the variable part of the molecule **6**, where structural modifications have been tested. The dark blue methyl probe represents for highly active compound **6** depicted by light blue cylinder, while red methyl probe represents for the least active compound **3** depicted by dark red cylinder. The most active compound **6** has an imidazolidine ring overlapping on the methyl probe (blue region), which is favorable for OA-agonist activity. Meanwhile, the least active compound **3** has an oxazoline ring embedded in the methyl probe (red region), which is less favorable for OA-agonist activity.

RSM

The energies of interaction between the RSM and each molecular model were added to the study table as new columns, which were used for generating QSARs. Instead of one total number which is the sum of the interactions evaluated between each point on the surface and each molecular model, leading to one extra column in the study table, the energies at each surface point are available. Depending on the size of the drug

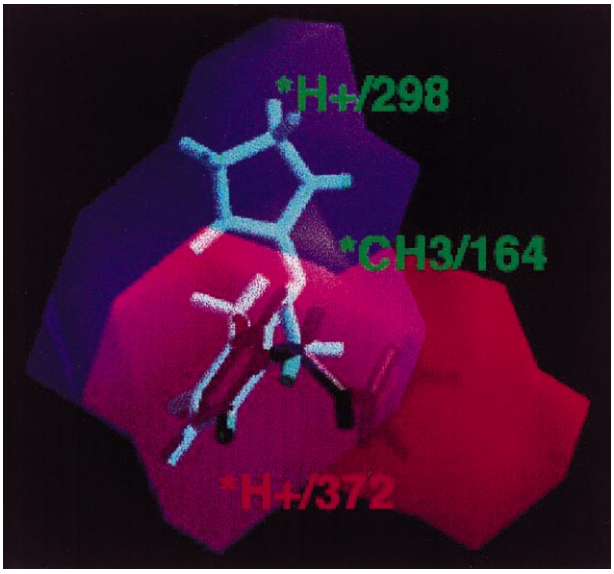

Figure 2. The most potential OA agonist **6** and the least potential OA agonist **3** embedded in a MFA generated from OA agonist training set and the points used in the resulting QSAR model. The blue methyl probe represents for highly active compound **6** depicted by light blue cylinder, while red methyl probe represents for the least active compound **3** depicted by dark red cylinder. Only points contributing to the equation are represented (a green sign stands for a positive contribution). Note here that all the points map in the variable part of the

molecules, this is potentially a great number of surface points. After adding the receptor surface point energies to the study table, a QSAR was calculated using the receptor surface energies and biological activities, leading to eq (4).

$$pK_i = 8.2137 + 2.7633ELE/137 - 1.52881 \times (ELE/546)^2 - 0.060721VDW/780 \quad (4)$$

where $n = 10$, $r^2 = 0.989$, $F = 184.362$, $CV - r^2 = 0.977$, and $Bsr^2 = 0.989 \pm 0$. Once the desired RSM has been constructed, all the structures in the training and test sets were evaluated against the model. The evaluation consists of computing several energetic descriptors that are based upon the interactions between ligand and model. The two functions used to calculate the field values are a van der Waals function and a Wyvill soft object function. For this work, three descriptors were generated. The descriptors *ELE*/137 and *ELE*/546 are the electrostatic interaction energy of the molecule with the receptor at points 137 and 546, respectively. The descriptor *VDW*/780 is the van der Waals interaction energy of the molecule with the receptor at point 780. According to eq (4), the negative $(ELE/546)^2$ term indicates that the OA-agonist activity is maximun when *ELE*/546 value is zero. By using receptor data to develop a QSAR model, the goodness of fit can be evaluated between a candidate structure and a postulated pseudo-receptor.

The descriptors generated for RSM, the activities of the compounds, and the predictions using their top model are shown in Table 4. An RSM represents essential information about the hypothetical receptor site as a three-dimensional surface with associated properties mapped onto the surface model. The location and shape of the surface represent information about the steric nature of the receptor site: the associated properties represent other information of interest, such as hydrophobicity, partial charge, electrostatic potential, and hydrogen-bonding propensity. The isosurface procedure produces a surface that entirely encloses the molecules over which it is generated. The surface has no holes and is known as a closed model. RSMs are best constructed from a set of the most active analogues that are chosen to cover the variety of steric and electrostatic variations likely to appear in the test data. The approach we took was to automatically build a set of different RSMs from different combinations of the most active analogues, and then use a variable-selection technique such as GFA to discover the RSM whose descriptors yield the best QSARs of the full training set. GFA allows the discovery and use of nonlinear descriptors by using spline-based terms. A RSM was generated (Fig. 3) using some subset of the most active structures (**1**, **5**, and **6**). The rationale underlying this model is that the most active structures tend to explore the best spatial and electronic interactions with receptor, while the least active do not and tend to have unfavorable steric or electronic interactions. The best model generated using the descriptors from the closed RSM is given in eq (4). Figure 3 shows the top three molecules embedded in the RSM generated from them and the points used in the resulting QSAR model colored by electrostatic potential. Only points contributing to the equation are represented (a green sign stands for a positive contribution and a red sign stands for a negative one). Note here that all the points map in the variable part of the molecules — the part where structural modifications have been tested.

In drug discovery, it is common to have measured activity data for a set of compounds acting upon a particular protein but not to have knowledge of the three-dimensional structure of the protein active site. In the absence of such three-dimensional information, one can attempt to build a hypothetical model of the receptor site that can provide insight about receptor site characteristics. Such a model is known as a RSM, which

Table 4. RSM descriptors and prediction using top model

				pK_i		
No.	*ELE*/137[a]	*ELE*/546[a]	*VDW*/780[b]	Observed	Calculated[c]	Deviation
1	0.191	−0.525	−0.017	8.21	8.32	−0.11
2	−0.530	−1.239	−0.073	5.21	5.45	−0.24
3	−0.365	−1.427	2.116	4.00	3.96	0.04
4	1.018	−1.591	−0.239	7.16	7.17	−0.01
5	−0.132	−0.219	−0.017	8.05	7.78	−0.27
6	0.153	−0.483	−0.195	8.41	8.29	0.12
7	0.002	−1.673	−0.149	4.11	3.95	0.16
8	−0.142	−0.164	11.012	7.21	7.11	0.10
9	−0.643	−0.013	18.098	5.27	5.34	−0.07
10	0.000	0.741	−0.102	7.11	7.38	−0.27

[a] The descriptors *ELE*/137 and *ELE*/546 are the electrostatic interaction energy of the molecule with the receptor at points 137 and 546, respectively.
[b] The descriptor *VDW*/780 is the van der Waals interaction energy of the molecule with the receptor at point 780.
[c] Calculated by eq (4).

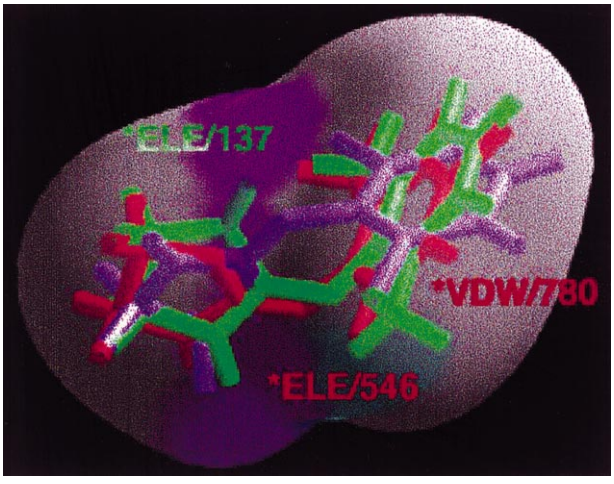

Figure 3. The three most active compounds **1** (green), **5** (purple), and **6** (red) embedded in a closed RSM generated from them and the points used in the resulting QSAR model colored by electrostatic potential. A purple sign stands for a negative contribution of electrostatic potential. Only points contributing to the equation are represented (a green sign stands for a positive contribution and a red sign stands for a negative one). Note here that all the points map in the variable part of the molecules—the part where structural modifications have been tested.

provides compact and quantitative descriptors which capture three-dimensional information about a putative receptor site. This model ought to be predictive and sufficiently reliable to guide the medicinal chemist in the design of novel compounds. These descriptors were used for predictive QSAR models. This approach is effective for the analysis of data sets where activity information is available but the structure of the receptor site is unknown. RSM attempts to postulate and represent the essential features of a receptor site itself, rather than the common features of the molecules that bind to it. RSMs differ from pharmacophore models in that the former tries to capture essential information about the receptor, while the latter captures information about the commonality of compounds that bind. Pharmacophore models tend to be geometrically underconstrained (while topologically overconstrained); this steric underconstraint leads to false positives, that is, compounds that are deemed active by the model but which are inactive when tested. They postulate a 3-D arrangement of atoms recognizable by the active site in terms of the similarity of functional groups common to the set of binding molecules. RSMs, on the other hand, tend to be geometrically overconstrained (and topologically neutral) since, in the absence of steric variation in a region, they assume the tightest steric surface which fits all training compounds. RSMs do not contain atoms, but try to directly represent the essential features of an active site by assuming complementarity between the shape and properties of the receptor site and the set of binding compounds. The RSM application uses 3-D surfaces that define the shape of the receptor site by enclosing the most active members (after appropriate alignment) of a series of compounds.

The predictive character of the QSARs was further assessed using molecules, whose structures are shown in Figure 1b, outside of the training set. The best statistically significant eqs (1)–(4) were applied to access some OA agonists. The predicted values of these molecules are listed in Table 5. Some OA agonists of 2-(arylimino)oxazolidine (AIO) **11**, 2-(arylimino)imidazolidine (AII) **12** and **13**, 2-(4-chloroanilino)thiazolidine (CAT) **14**, naphazoline **15**, and synephrine **16** were active according to eqs (1)–(4) in inhibition of intracellular cAMP production by 1 pmol PBAN/intersegment. The most active compounds are found to be different from different equations, e.g. eq (1) predicts **13** and **16** as the most active, eq (2) **12** as the most active, eq (3) **11** and **14** as the most active, and eq (4) none so active. A distinguishing characteristic of the eqs (3) and (4) is that they overestimated the OA-agonist activity of **1** (experimental pK_i: less than 4, estimated pK_i: 7.217 and 7.552 for eqs (3) and (4), respectively). This result may imply that the process of calculating an MFA and RSM does not treat **17**-like structures reasonably. Thus, the above data suggest that phenyl ring substitution requirements for AIO and NC derivatives active as octopaminergic agonists differ substantially from each other and other various types of OA agonists could be potent inhibitors in sex-pheromone production of female moths, although the number of compounds tested here is still limited to draw any conclusions.

Table 5. Predicted activity of OA agonists from eqs (1)–(4) in the test set

Compound		Predicted activity (pK_i)			
No.	*R*	eq (1)	eq (2)	eq (3)	eq (4)
11	CAO	8.768	7.234	9.050	7.922
12	AII H	7.216	9.317	8.593	8.259
13	AII 2,4,5-Cl_3	9.383	8.254	8.616	8.012
14	CAT	7.925	7.319	9.039	8.261
15	Naphazoline	7.638	7.739	6.994	8.219
16	Synephrine	9.350	8.175	7.267	7.892
17[a]	FBAO	3.232	1.335	7.217	7.552

[a] The experimental activity (pK_i) of **17** was less than 4.000.

Conclusion

QSAR modeling is an area of research pioneered by Hansch and Fujita. QSAR attempts to model the activity of a series of compounds using measured or computed properties of the compounds as shown in eq (1). In this report, QSAR has been extended by including the analysis of three-dimensional information about the series by three-dimensional shape descriptors as illustrated by the MSA approach in eq (2), by field descriptors as illustrated by MFA approach in eq (3), and by receptor descriptors as illustrated by RSM approach in eq (4), respectively. Some models were statistically significant and were used to correctly predict the activities of a set of training molecules ranging over 4 orders of magnitude (max pK_i 8.41 and min pK_i 4), indicating that these models could be useful tools to design active OA agonists. This set included a variety of types of molecules and for this types of training set, the use of these methods was appropriate. These methods generate multiple models that can be checked easily for validity.

OA is not likely to penetrate either the cuticle or the central nervous system of insects effectively, since it is fully ionized at physiological pH. Derivatization of the polar groups would be one possible solution to this problem in trying to develop potential pest-control agents. The above QSAR, MSA, MFA, and RSM studies show that agonists with certain substituents can be potential ligands to OA receptors, although the number of compounds tested here is still limited to draw any conclusions. They may help to point the way towards developing extremely potent and relatively specific OA agonists, leading to potential inhibitors in sex-pheromone production of moths. In order to optimize the activities of these compounds as OA-agonists, more detailed experiments are in progress.

Experimental

Synthesis of OA agonists

1-(2,6-Diethylphenyl)imidazolidine-2-thione (EPIT) **9** was synthesized by the cyclization of monoethanolamine hydrogen sulfate with 2,6-diethylphenylisothiocyanate in the presence of sodium hydroxide as described in the

previous report.[34] AIOs **1** and **2**, 2-(2,3,4,5-tetrachloroanilino)oxazolidine (CAO) **11**, 2-(substituted benzylamino)-2-oxazolines (BAOs) **3** and **4**, and 2-(3-trifluoromethylbenzylamino)-2-oxazoline (FBAO) **1** were obtained by cyclodesulfurizing the corresponding thiourea with yellow mercuric oxide.[35] 2-(2,3-Dichlorobenzylamino)-2-thiazoline (CBAT) **7**, 2-(2,6-diethylanilino)thiazolidine (DAT) **8**, and CAT **14** were synthesized by cyclization of the corresponding thiourea with concd hydrogen chloride.[36] AII **6**, **12**, and **13** were prepared according to a reported method by refluxing the corresponding aniline and 1-acetyl-2-imidazolidone in phosphoryl chloride followed by hydrolysis.[37] The structures of the compounds were confirmed by ^{1}H and ^{13}C NMR measured with a JEOL JNM-EX400 spectrometer at 400 MHz, tetramethyl silane (TMS) being used as an internal standard for ^{1}H NMR and elemental analytical data. Tolazoline (**10**) and naphazoline (**15**) were obtained from Janssen Chimica (Beerse, Belgium) and Aldrich Chemical Company, Inc. (Milwaukee, WI), respectively; clonidine (**5**) and synephrine (**16**) from Sigma Chemical Company (St. Louis, MO).

Insect culture

The study was conducted on *H. armigera*. The larvae were raised on an artificial diet at a constant temperature of 25°C and 14:10 (light:dark) photoperiod as reported previously.[38] Pupae were sexed and males and females were allowed to emerge separately.

Intracellular cAMP levels

Compounds found to inhibit pheromone production in vivo by at least 50% were subsequently analyzed using the assay for cAMP production by intersegmental membranes.[39] Ovipositor tips, consisting of the 8th and 9th abdominal segments with the attached intersegmental membrane, were removed during the 10–12th h photophase from 2- or 3-day old females. The intersegmental membranes were isolated by dissection and washed twice in Pipes buffered physiological medium (21 mM KCl, 12 mM NaCl, 3 mM $CaCl_2$, 18 mM $MgCl_2$, 85 mM glucose, and 43 mM trehalose in 5 mM Pipes buffer brought to pH 6.6 using 0.1 N KOH). The intersegmental membranes were incubated in 10 μL medium containing 1 μg/μL isobutylmethylxanthine (Sigma, USA) in the presence of synthetic Hez-PBAN (1 pmol/intersegment) and the test compounds for 10 min. The reaction was stopped by transferring the intersegmental membranes into a 1:5 mixture of 20% perchloric acid and Hepes buffer (50 mM) on ice. After incubation on ice for 20 minutes the intersegmental membranes were homogenized, titrated to pH = 7 with 1.5 M KOH and centrifuged at 2000×*g* for 10 min. The supernatant was removed and used for radioimmunoassay (RIA). The RIA for cAMP was performed as described previously.[38]

QSAR calculations

All experiments were conducted with Molecular Simulations Incorporated's Cerius2 3.5 QSAR environment (Burlington, MA) on a Silicon Graphics O2, running under the IRIX 6.3 operating system. Due to the large number of points used as independent variables, GFA was used to derive the QSAR models. The GFA algorithm was initially conceived by taking inspiration from two seemingly disparate algorithms: Holland's genetic algorithm[40] and Friedman's[41] multivariate adaptive regression splines (MARS) algorithm. Genetic algorithms are derived from an analogy with the evolution of DNA. In this analogy, individuals are represented by a one-dimensional string of bits. An initial population of individuals is created, usually with random initial bits. A fitness function is used to estimate the quality of an individual, so that the best individuals receive the best fitness scores. Individuals with the best scores are more likely to be chosen to mate and to propagate their genetic material to offspring through the crossover operation, in which pieces of genetic material are taken from each parent and recombined to create the child. After many mating steps, the average fitness of the individuals in the population increases as good combinations of genes are discovered and spread through the population. Genetic algorithms are especially good at searching problem spaces with a large number of dimensions, as they conduct a very efficient directed sampling of the large space of possibilities. Friedman's MARS algorithm is a statistical technique for modeling data. It provided an error measure, called the lack of fit (LOF) score, that automatically penalized models with too many features. It automatically selects which features are to be used in the models. It is better at discovering combinations of features that take advantage of correlations between multiple features. It incorporates Friedman's LOF error measure, which estimates the most appropriate number of features, resists overfitting, and allows user control over the smoothness of fit. It also inspired the use of splines as a powerful tool for non-linear modeling. The GFA uses a genetic algorithm to perform a search over the space of possible QSAR models using the LOF score to estimate the fitness of each model. Such evolution of a population of randomly constructed models leads to the discovery of highly predictive QSARs. The GFA algorithm approach has a number of important advantages over other techniques: it builds multiple models rather than a single model; it can use a larger variety of equation term types in construction of its models (for example, splines, step functions, or high-order polynomials); it provides, through study of the evolving models, additional information not available from standard regression analysis (such as the preferred model length and useful partitions of the data set).

Molecular modeling

Once a reasonable RSM has been defined, a series of structures can be evaluated against the model. When a receptor model has been generated and the models have been aligned, a QSAR can be built using data from the receptor–structure interactions. The results of the minimization procedure were used as descriptors either to refine the model or to predict activity. Three-dimensional energetics descriptors were calculated from

RSM–ligand interaction. These three-dimensional descriptors were used in QSAR analysis. A RSM represents the global volume that can accommodate one or more molecules and can be seen as the shape of an active site built from the ligands that fit into it in their "active" conformation. The descriptors used in this study account for phenomena that occur at the contact surface between the ligands and the protein active site. This concept differs from that used in classical 3-D-QSAR in which fields of molecules are represented using grids in MFA. The technique resembles MFA but, instead of a rectangular grid, the points considered are taken from the receptor surface. Therefore, they are probably more chemically relevant than a rectangular grid because they exist on a surface that is shaped like a molecule, and even better, a surface constructed from a subset of active molecules. The MFA formalism calculates probe interaction energies on a rectangular grid around a bundle of active molecules. Receptor calculates molecule–receptor model interaction energies on a receptor surface, and, like MFA, these energies can serve as input for the calculation of a QSAR relationship. Just as each energy associated with an MFA grid point can be used, so can each point on the surface of a receptor model. Hopefully, the receptor surface is better able to sample the environment of the molecule than a rectangular grid, leading to better results.

Acknowledgements

This work was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, and Culture of Japan.

References

1. Evans, P. D. *Comp. Mol. Neurobiol.* **1993**, 287.
2. Evans, P. D. *J. Physiol.* **1981**, *318*, 99.
3. Roeder, T.; Nathanson, J. A. *Neurochem. Res.* **1993**, *18*, 921.
4. Roeder, T.; Gewecke, M. *Biochem. Pharm.* **1990**, *39*, 1793.
5. Roeder, T. *Life Science* **1991**, *50*, 21.
6. Roeder, T. *Br. J. Pharmac.* **1995**, *114*, 210.
7. Roeder, T. *Eur. J. Pharmac.* **1990**, *191*, 221.
8. Hirashima, A.; Shinkai, K.; Pan, C.; Kuwano, E.; Taniguchi, E.; Eto, M. *Pestic. Sci.* **1999**, *55*, 119.
9. Hirashima, A.; Pan, C.; Shinkai, K.; Tomita, J.; Kuwano, E.; Taniguchi, E.; Eto, M. *Bioorg. Med. Chem.* **1998**, *6*, 903.
10. Pan, C.; Hirashima, A.; Kuwano, E.; Eto, M. *J. Molec. Model.* **1997**, *3*, 455.
11. Hirashima, A.; Pan, C.; Kuwano, E.; Taniguchi, E.; Eto, M. *Bioorg. Med. Chem.* **1999**, *7*, 1437.
12. Raina, A. K. *Ann. Rev. Entomol.* **1993**, *38*, 320.
13. Ma, P. W. K.; Roelofs, W. *Insect Biochem. Molec. Biol.* **1995**, *25*, 467.
14. Fabrias, G.; Barrot, M.; Camps, F. *Insect Biochem. Molec. Biol.* **1995**, *25*, 655.
15. Zhu, J.; Millar, J.; Loefstedt, C. *Archs. Insect Biochem. Physiol.* **1995**, *30*, 41.
16. Fang, N.; Teal, P. E. A.; Tumlinson, H. *Archs. Insect Biochem. Physiol.* **1996**, *32*, 249.
17. Foster, S.; Roelofs, W. L. *Archs. Insect Biochem. Physiol.* **1996**, *32*, 135.
18. Rafaeli, A.; Gileadi, C. *Invertebrate Neuroscience* **1997**, *3*, 223.
19. Jurenka, R. A. *Arch. Insect Biochem. Physiol.* **1996**, *33*, 245.
20. Soroker, V.; Rafaeli, A. *Insect Biochem.* **1989**, *67*, 1.
21. Rafaeli, A.; Soroker, V.; Kamensky, B.; Raina, A. K. *J. Insect Physiol.* **1990**, *36*, 641.
22. Arima, R.; Takahara, K.; Kadoshima, T.; Numazaki, F.; Ando, T.; Uchiyama, M.; Nagasawa, H.; Kitamura, A.; Suzuki, A. *Appl. Entomol. Zool.* **1991**, *26*, 137.
23. Jurenka, R. A.; Jacquin, E.; Roelofs, W. L. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 8621.
24. Fonagy, A.; Matsumoto, S.; Schoofs, L.; Loof, A. De; Mitsui, T. *Biosci. Biotech. Biochem.* **1992**, *56*, 1692.
25. Matsumoto, S.; Ozawa, R.; Nagamine, T.; Kim, G.-H.; Uchiumi, K.; Shono, T.; Mitsui, T. *Biosci. Biotech. Biochem.* **1995**, *59*, 560.
26. Rafaeli, A.; Gileadi, C. *Insect Biochem. Molec. Biol.* **1995**, *25*, 827.
27. Rafaeli, A.; Gileadi, C. *Insect Biochem. Molec. Biol.* **1996**, *26*, 797.
28. Rafaeli, A.; Gileadi, C.; Fan, Y.; Meixun, C. *J. Insect Physiol.* **1997**, *43*, 261.
29. Rohrbaugh, R. H.; Jurs, P. C. *Analytica Chimica Acta* **1987**, *199*, 99.
30. Ghose, A. K.; Pritchett, A.; Crippen, G. M. *J. Comp. Chem.* **1988**, *9*, 80.
31. Burke, B. J.; Hopfinger, A. J. *J. Med. Chem.* **1990**, *33*, 274.
32. Burke, B. J.; Hopfinger, A. J. In *Molecular Similarity*; Johnson, M. A.; Maggiora, G. M., Eds.; John Wiley and Sons: New York, 1990; pp 11–73.
33. Hopfinger, A. J.; Burke, B. J. In *Concepts and Applications of Molecular Similarity*; Johnson, M. A.; Maggiora, G. M., Eds.; John Wiley and Sons: New York, 1990; p 173.
34. Hirashima, A.; Shinkai, K.; Kuwano, E.; Taniguchi, E.; Eto, M. *Biosci. Biotech. Biochem.* **1998**, *62*, 1179.
35. Hirashima, A.; Pan, C.; Katafuchi, Y.; Taniguchi, E.; Eto, M. *J. Pestic. Sci.* **1996**, *21*, 419.
36. Hirashima, A.; Tarui, H.; Eto, M. *Biosci. Biochem. Biotech.* **1994**, *58*, 1206.
37. Nathanson, J. A.; Kaugars, G. *J. Med. Chem.* **1989**, *32*, 441.
38. Rafaeli, A.; Soroker, V. *Mol. Cell. Endocrinol.* **1989**, *65*, 43.
39. Soroker, V.; Rafaeli, A. *Insect Biochem. Molec. Biol.* **1995**, *25*, 1.
40. Holland, J. In *Adaptation in Artificial and Natural Systems*; University of Michigan Press, 1975.
41. Friedman, J. In *Multivariate Adaptive Regression Splines*; Technical Report 102, Laboratory for Computational Statistics, Department of Statistics, Stanford University: Stanford, CA, 1988 (revised 1990).