Pergamon

Bioorganic & Medicinal Chemistry 7 (1999) 1437–1443

BIOORGANIC & MEDICINAL CHEMISTRY

Three-dimensional Pharmacophore Hypotheses for the Locust Neuronal Octopamine Receptor (OAR3). Part 2: Agonists

Akinori Hirashima,[a,*] Canping Pan,[b] Eiichi Kuwano,[a] Eiji Taniguchi[a] and Morifusa Eto[a]

[a]*Division of Bioresource and Bioenvironmental Sciences, Graduate School, Kyushu University, Fukuoka 812-8581, Japan*
[b]*Department of Applied Chemistry, Beijing Agricultural University, Beijing 100094, People's Republic of China*

Received 25 December 1998; accepted 19 February 1999

Abstract—Three-dimensional pharmacophore hypotheses were built from a set of 43 agonists against octopamine receptor class 3 (OAR3) in locust nervous tissue. Among the 10 chemical-featured models generated by program Catalyst/Hypo, a hypothesis including hydrogen-bond acceptor (HBA), hydrophobic (Hp), and hydrophobic aliphatic (HpAl) features was considered to be important and predictive in evaluating OAR3 agonists. While the ideal and null hypotheses had a cost of 156.40 and 239.20, respectively, the 10 resulting hypotheses possessed costs from 169.89 to 175.81. The best hypothesis that was confirmed to have a 95% chance of true correlation yielded a low RMS of 0.757 and high regression r of 0.933. Active agonists mapped well onto all the features of the hypothesis such as HBA, Hp, and HpAl. On the other hand, inactive compounds were shown to be difficult to achieve the energetically favorable conformation which is found in the active molecules in order to fit the 3-D chemical feature pharmacophore models.

Introduction

The biogenic monoamine octopamine (OA, α-aminomethyl-*p*-hydroxybenzyl alcohol), which has been found to present in high concentrations in various insect tissues, is the monohydroxylated analogue of the vertebrate hormone noradrenaline. OA receptors are perhaps the only non-peptide receptors whose occurrence is restricted to invertebrates. OA was first discovered in the salivary glands of octopus by Erspamer and Boretti in 1951.[1] It has been found that OA is present in a high concentration in various invertebrate tissues.[2] This multifunctional and naturally occurring biogenic amine has been well studied and established as (1) a neurotransmitter, controlling the firefly light organ and endocrine gland activity in other insects; (2) a neurohormone, inducing mobilization of lipids and carbohydrates; (3) a neuromodulator, acting peripherally on different muscles, fat body, and sensory organs such as corpora cardiaca and the corpora allata, and (4) a centrally acting neuromodulator, influencing motor patterns, habituation, and even memory in various invertebrate species.[3–6] The action of OA is mediated through various receptor classes which is coupled to G-proteins and is specifically linked to an adenylate cyclase. Thus, the physiological actions of OA have been shown to be associated with elevated levels of cyclic AMP.[7] Three different receptor classes OAR1, OAR2A, and OAR2B had been distinguished from non-neuronal tissues.[8] In the nervous system of locust, 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.[9–13]

Recently much attention has been directed at the octopaminergic system as a valid target in the development of safer and selective pesticides.[14] Structure–activity studies of various types of OA-agonists and antagonists were also reported using the nervous tissue of the migratory locust, *Locusta migratoria* L.[15] However, information on the structural requirements of these OA-agonists and antagonists for high OA-receptor ligands is still limited. In rational drug design process, it is common that the binding activity data of a set of compounds acting upon a particular protein is known, while information of the three-dimensional 3-D structure of the protein active site is absent. A 3-D pharmacophore hypothesis that is consistent with known data should be useful and predictive in evaluating new compounds and directing further synthesis. A pharmacophore model postulates that there is an essential 3-D arrangement of functional groups that a molecule must possess to be recognized by the receptor. It collects chemical features distributed in 3-D space that is intended to represent groups in a molecule that participates in important

Key words: *Locusta migratoria*; octopamine agonist; receptor hypothesis; catalyst; pharmacophore.
* 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)00057-7

binding interactions between ligands and their receptors. Hence, a pharmacophore model provides crucial information about how well the chemical features of a subject molecule overlap with the hypothesis model. It also informs the ability of molecules to adjust their conformations in order to fit a receptor with energetically reasonable conformations. Such characterized 3-D models convey important information in an intuitive manner and can provide predictive capability for evaluating new compounds.

Our interest in octopaminergic agonists was aroused by the results of quantitative stucture–activity relationship (QSAR) study using various physicochemical parameters as descriptors[15] or receptor surface model.[16] Furthermore, molecular modeling and conformational analysis were performed in Catalyst/Hypo to gain a better knowledge of the interactions between octopaminergic antagonists and OAR3 in order to understand the conformations required for binding activity.[17] The current work is aimed to generate 3-D chemical function-based hypotheses from a set of OA agonists.

Results and Discussion

Assessment of 3-D-QSAR for binding activity

A set of 43 molecules, whose inhibitory activities were tested using the [^{3}H]-OA binding to OAR3 in the locust central nervous tissue, was selected from the reported data[12] as the target training set. Their molecular structures and experimental biological activities are listed in Figure 1 and Table 1. This set includes a variety of types of molecules and for this type of training set, the use of the hypothesis generation tool is appropriate. This tool builds hypotheses (overlays of chemical features) for which the fit of individual molecules to a hypothesis can be correlated with the molecule's affinity.

NC derivative substituted with 2,4,6-$(CH_2CH_3)_3$ (**30**) had the highest potency, followed by NC 2,4-Cl_2 (**20**), 2-CH_3,4-Cl (**21**), and AC-6 (**37**) in binding to OAR3 of locust nervous system (Table 1). Introduction of 3-Cl, 3-CH_3, 4-Cl, 4-CH_3, 2,3-Cl_2, 2,5-Cl_2, 2,5-Cl_2, 3,4-Cl_2, 3-Cl,4-F, and 2,5-Cl_2 to the phenyl moiety of BAT (**1**) increased the potency. Similarly, introduction of 2,4-Cl_2, 2-CH_3,4-Cl, 2,4-$(CH_3)_2$, 2,4,5-Cl_3, 2,6-Cl_2,4-N_3, 2,4,6-$(CH_3)_3$, 2,6-$(CH_2CH_3)_2$, 2,4,6-$(CH_2CH_3)_3$, and 2,6-$(CH_2CH_3)_2$,4-N_3 to the phenyl moiety of NC (**19**) increased the potency. The above data suggest that phenyl ring substitution requirements for BAT and NC derivatives active as octopaminergic agonists differ substantially from each other.

The 3-D-QSAR study was performed with the Catalyst (Version 3.1) package. The geometry of each compound was built with a visualizer and optimized by using the generalized CHARMm-like force field[18–21] implemented in the program. A preparative test was performed with hydrogen-bond acceptor (HBA), hydrogen bond donor (HBD), hydrophobic (Hp), hydrophobic aliphatic (HpAl), negative ionizable (NI), and positive ionizable (PI).[22] NI and PI were used rather than negative charge and positive charge in order to broaden the search for deprotonated and protonated atoms or groups at physiological pH. Furthermore, in order to emphasize the importance of an aromatic group corresponding to the phenol moiety of test compounds, ring aromatic (RA) which consists of directionality was chosen to be included in the subsequent run. It was found that hypotheses contain good correlation with HBA, Hp, and HpAl. The characteristics of the 10 lowest cost hypotheses that were obtained are listed in Table 2. The total fixed cost of the run is 156.40 and the cost of the null hypothesis is 239.20. The cost range between best hypothesis 1 and null hypothesis is 69.31, and the cost range over the 10 generated hypotheses is 5.92. Hypotheses 1, 2, 6, and 9 consist of the same chemical feature functions as a HBA, three Hp, and a HpAl features. The second group composes hypotheses 3, 4, 7, 8, and 10 which are characterized by a HBA, two Hp, and two HpAl features. Another hypothesis 5 is characterized by a HBA and four Hps. Comparison of the procedure and regression studies shows that hypotheses 1, 3, and 5 are the best models among the three groups and are selected for further evaluation.

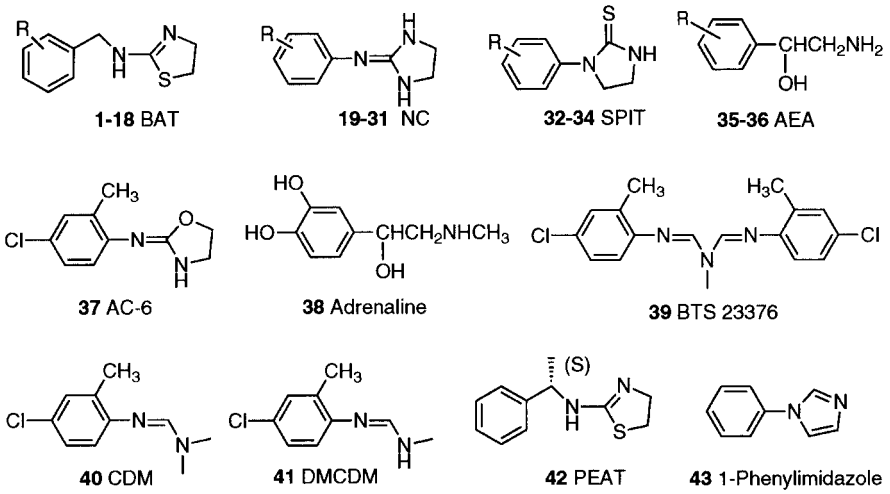

Figure 1. Structures of OA agonists in the training set.

Table 1. Predicted activity from 10 best hypotheses against actual binding activity data for 43 agonists

Compound				Hypotheses									
No.	R	Exp.	Conf.	1	2	3	4	5	6	7	8	9	10
1	BAT H	280 ± 59[a]	120	240	440	260	500	430	300	320	560	500	830
2	BAT 2-F	447 ± 125[a]	212	640	600	830	520	700	680	880	560	930	740
3	BAT 2-CF_3	290 ± 203[a]	242	81	43	67	42	37	41	51	54	44	44
4	BAT 3-Cl	95 ± 66[a]	204	37	41	36	37	37	41	35	37	33	32
5	BAT 3-F	2509 ± 75[a]	206	700	680	920	580	850	990	880	580	930	830
6	BAT 3-CH_3	34 ± 26[a]	131	36	40	36	37	37	41	39	36	33	34
7	BAT 3-NO_2	185 ± 33[a]	200	230	150	240	650	99	170	320	240	270	580
8	BAT 4-Cl	89 ± 6[a]	121	38	26	35	24	32	38	38	28	33	33
9	BAT 4-F	460 ± 124[a]	123	280	470	320	410	600	370	380	510	530	520
10	BAT 4-CH_3	38 ± 16[a]	122	40	35	37	32	36	43	42	37	38	37
11	BAT 2,3-Cl_2	42 ± 19[a]	140	36	19	36	38	21	18	24	30	24	26
12	BAT 2,4-Cl_2	1.7 ± 1.1[a]	131	1.7	1.7	2.2	1.7	2.1	2.6	2.3	1.6	1.9	2.1
13	BAT 2-Cl,4-F	184 ± 90[a]	145	66	46	79	51	42	50	44	43	38	39
14	BAT 2,5-Cl_2	132 ± 83[a]	139	35	34	36	38	31	42	36	36	32	32
15	BAT 3,4-Cl_2	14 ± 4.2[a]	159	19	13	16	13	13	17	14	18	10	11
16	BAT 3-Cl,4-F	109 ± 59[a]	132	37	40	36	37	36	41	37	37	34	34
17	BAT 3,4-F_2	445 ± 196[a]	158	1100	610	990	520	740	780	880	650	930	770
18	BAT 3,5-Cl_2	1 ± 0[a]	142	1.5	2.4	2.4	2.3	2.8	3	3	2	2.5	2.6
19	NC H	23 ± 5[b]	15	50	67	44	61	52	53	62	70	71	64
20	NC 2,4-Cl_2	0.81 ± 0.18[b]	23	2.2	2	2.4	2	2	2.1	2	2.5	2.5	2.5
21	NC 2-CH_3,4-Cl	0.87 ± 0.32[b]	22	2	1.9	2.2	1.9	2.3	2	2.4	2.4	2.5	2.6
22	NC 2,4-$(CH_3)_2$	1.02 ± 0.42[b]	20	2.5	2.1	2.7	2.1	2.3	2.3	2.9	2.1	2.5	2.5
23	NC 2,6-Cl_2	47 ± 18[b]	24	42	53	43	48	42	49	48	41	42	41
24	NC 2,6-$(CH_3)_2$	20 ± 7[b]	26	37	41	36	38	30	32	37	37	32	32
25	NC 2,6-$[CH(CH_3)_2]_2$	132 ± 35.6[b]	16	58	52	49	47	47	44	17	23	22	22
26	NC 2,4,5-Cl_3	2.27 ± 0.89[b]	29	0.72	0.58	0.61	0.61	0.77	0.42	1.5	2.3	2.12	
27	NC 2,6-Cl_2,4-NH_2	58 ± 16[b]	25	39	50	41	49	43	47	26	37	34	34
28	NC 2,6-Cl_2,4-N_3	1.05 ± 0.47[b]	10	41	54	44	700	51	50	36	45	37	37
29	NC 2,4,6-$(CH_3)_3$	4.38 ± 1.30[b]	27	4.3	3.7	4.2	3.7	3.5	3.4	3.3	3.2	4.5	4.7
30	NC 2,4,6-$(CH_2CH_3)_3$	0.56 ± 0.14[b]	110	0.38	0.4	0.26	0.41	0.49	0.41	0.28	0.32	0.21	0.22
31	NC 2,6-$(CH_2CH_3)_2$,4-N_3	1.05 ± 0.47[b]	182	3	3.4	2.9	3.3	3.8	3.1	2.5	3.8	3	3
32	SPIT 4-Cl	280 ± 134[a]	4	240	410	160	360	960	190	220	400	340	310
33	SPIT 2-CH_3,4-Cl	20 ± 9[a]	4	44	49	46	49	41	48	49	43	41	41
34	SPIT 2,6-$(CH_2CH_3)_2$	170 ± 51[a]	45	67	73	62	67	80	38	43	26	19	15
35	AEA 3-OH	5050 ± 1860[b]	17	1200	1300	1000	1100	1200	1100	890	1400	970	860
36	AEA 3,4-$(OH)_2$	475 ± 42[b]	17	1200	1300	1000	1100	1100	1100	890	1300	960	850
37	AC-6	0.95 ± 0.24[b]	12	2.5	2.3	2.4	2.2	2.8	2.5	2.5	2.8	2.8	2.6
38	Adrenaline	416 ± 75[b]	47	1200	1300	990	1100	1100	1100	880	1200	940	830
39	BTS23376	8.9 ± 0.6[b]	14	1.6	3.7	2.1	3.5	5	2.8	4.8	7.8	9.7	7.8
40	CDM	137 ± 70[b]	11	37	44	45	40	42	52	45	12	38	37
41	DMCDM	137 ± 70[b]	16	6.7	5.5	5.5	5.4	2.6	7.1	4.1	4.8	6.9	8.6
42	PEAT	37 ± 23[a]	97	84	61	91	56	61	84	120	59	83	80
43	1-Phenylimidazole	813 ± 561[b]	4	1700	1600	2200	1800	1400	2600	3300	2500	1600	1400

Abbreviations: Exp. experimental data (K_i in nM); Confs. number of conformational models. Affinities of the agonists are expressed as their K_i values in nanomole and activities range over four orders of magnitude (min. 0.56 nM and max. 5050 nM). K_d for [^{3}H]OA was 7.90 ± 0.9 nM.[9]
[a]Personal communication (T. Roeder, Hamburg University).
[b]Cited from ref 12.

Validation of the hypothesis

The hypotheses are used to estimate the activities of the training set. The activities are derived from the conformers displaying the smallest root mean square (RMS) deviations when projected onto the hypothesis. Hypotheses 1, 3, and 5 share five common features located at almost exactly the same 3-D coordinates. The quality of the correlation among the data in the training set is given by the RMS score which was normalized by the log (uncertainty) and the regression constant *r*. All calculated binding activities from 10 best hypotheses and the number of generated conformations for each molecule are listed in Table 1. Even though hypothesis 1 has a slightly higher regression constant than those of hypotheses 3 and 5 (0.933, 0.929, and 0.928 for hypotheses 1, 3, and 5, respectively), its cost value and RMS are essentially the same as those of hypotheses 3 and 5. Roughly speaking, the greater the difference between the score of the generated hypothesis and the score of the null hypothesis, the more likely it is that the hypothesis reflects a chance correlation. The correlation between observed and estimated binding activities is satisfactory in the three hypotheses. The predicted activities are in the right order and parallel the values actually observed (Table 1).

The small cost range observed here may be due to two factors: molecules in the training set are fairly rigid and have a high degree of structural homology. Due to the relatively small range between the costs for an ideal versus null hypothesis and due moreover to the placement of the identified hypotheses within this range, special care was taken to test for chance correlation. Using the Fischer[23] method, the statistical significance of the hypotheses was tested by scrambling the

Table 2. Characteristics of 10 lowest cost hypotheses from 43 OAR3 agonists (cost of ideal hypothesis: 156.40, cost of null hypothesis: 239.20)

Hypotheses	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Cost	RMS	*r*
1	HBA	Hp 1	Hp 2	Hp 3	HpAl 1	169.89	0.757	0.933
2	HBA	Hp 1	Hp 2	Hp 3	HpAl 1	170.67	0.773	0.930
3	HBA	Hp 1	Hp 2	HpAl 1	HpAl 2	171.10	0.780	0.929
4	HBA	Hp 1	Hp 2	HpAl 1	HpAl 2	171.11	0.779	0.929
5	HBA	Hp 1	Hp 2	Hp 3	Hp 4	171.13	0.784	0.928
6	HBA	Hp 1	Hp 2	Hp 3	HpAl 1	172.99	0.828	0.919
7	HBA	Hp 1	Hp 2	HpAl 1	HpAl 2	173.65	0.843	0.916
8	HBA	Hp 1	Hp 2	HpAl 1	HpAl 2	173.87	0.859	0.913
9	HBA	Hp 1	Hp 2	Hp 3	HpAl 1	174.50	0.878	0.909
10	HBA	Hp 1	Hp 2	HpAl 1	HpAl 2	175.81	0.904	0.903

Abbreviations: HBA hydrogen-bond acceptor; Hp hydrophobic; HpAl hydrophobic aliphatic.

experimental activities in the training set and regenerating the set of hypotheses for each new trial. This was performed 19 times and among the 190 resulting hypotheses, none had a cost score lower than that of hypothesis 1. These tests indicate that the original hypothesis 1 represents true correlation in 95% chance. The hypothesis 1 was regressed using each molecule in its most chemically reasonable conformation (Fig. 2). Actually, the regression line for hypothesis 1 against the training set resulted in a good correlation across the entire collection of compounds. The regression line shown in Figure 2 indicates a good correlation between measured and estimated activities for the molecules of the training set. Hypothesis 1 had the highest statistical correlation ($r = 0.933$), although the range among 10 hypotheses is small. The smallest error value we obtained (RMS = 0.757) for hypothesis 1 indicates a reliable ability to predict activities within the training set. Thus, taken together, the hypothesis 1 turns out to be the most accurate measure for the whole training set as attested by the reasonably good correlation between observed and estimated activitities (cost = 169.89, cost of the null hypothesis = 239.20).

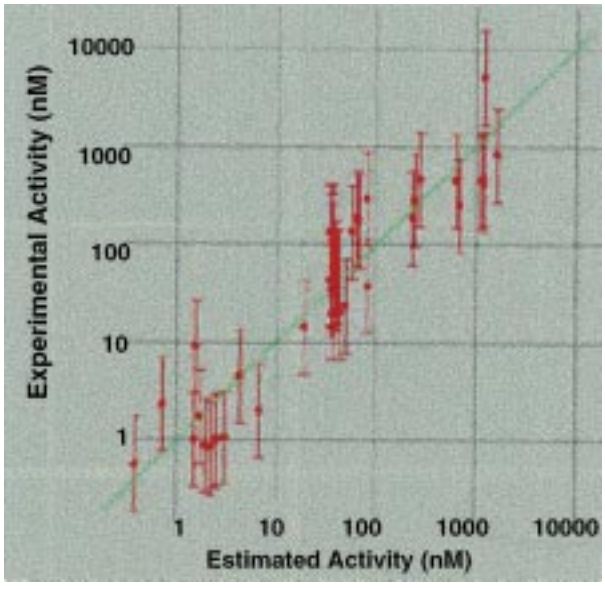

Figure 2. Plot of predicted binding affinity activities against experimental values (K_i in nM).

Receptor–drug interaction

QSAR modeling is an area of research pioneered by Hansch and Fujita.[24,25] The QSAR study assumes that the difference of the molecules in the structural properties experimentally measured accounts for the difference in their observed biological or chemical properties.[24–26] The result of QSAR usually reflects as a predictive formula and attempts to model the activity of a series of compounds using measured or computed properties of the compounds. More recently, QSAR has been extended by including the 3-D information. 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 3-D structure of the active site. In the absence of such 3-D information, one may attempt to build a hypothetical model of the receptor site that can provide insight on the nature of the receptor site. Such a model is known as a Hypo. Catalyst/Hypo is useful in building 3-D pharmacophore models from the binding activity data and conformational structure. It can be used as an alternative for QSAR methods, because of easy visualization and high prediction.

Figures 3–5 depict the most active compound **30** mapped onto the lowest cost hypothesis 1, on the third lowest cost hypothesis 3, and on the fifth lowest cost hypothesis 5, respectively. In general, hypotheses overestimates the binding affinity of **30**. This result may imply that Catalyst treats **30**-like structures reasonably in the process of calculating a hypothesis. The molecule maps well onto all five hypotheses features in a similar way (see Figs. 3–5) and therefore these hypotheses are considered to be equivalent. One of Hp of hypothesis 5 is replaced by an HpAl, leading to hypothesis 1 and an Hp of hypothesis 1 is replaced by an HpAl, leading to hypothesis 3. Roughly speaking, hypotheses 1, 3, and 5 have the good similarity in 3-D spatial shape. The phenyl moiety, 4-ethyl group, and imidazolidine ring of **30** overlaps with the three Hp features of hypothesis 1, whereas the bridge nitrogen atom, between a phenyl and imidazolidine rings, serves as a HBA (Fig. 3). The HpAl feature overlaps neatly with the 2-ethyl group and another Hp is disposed over the imidazolidine ring. The most active member, **30**, maps closely with the statistically most significant hypothesis 1, which is characterized by five features. The present studies on OA agonists demonstrate that a HBA site and four hydrophobic sites

located on the molecule seem to be essential for binding activity. Generally, more active molecules map onto all the features of the hypothesis, and compounds that are estimated to have low activity map poorly to the hypothesis (see Fig. 6). AEA **35** (*m*-isomer of OA) with low affinity (exp. 5050 nM) fits only two features of hypothesis 1.

The validity and the predictive character of this five-feature hypothesis was further assessed using molecules, whose structures are shown in Figure 7, outside of the training set. The best statistically significant hypothesis **1** was applied to access some OA antagonists, whose binding activity was tested against OAR3 on locust.[13] The predicted values and the bioassay activities of these

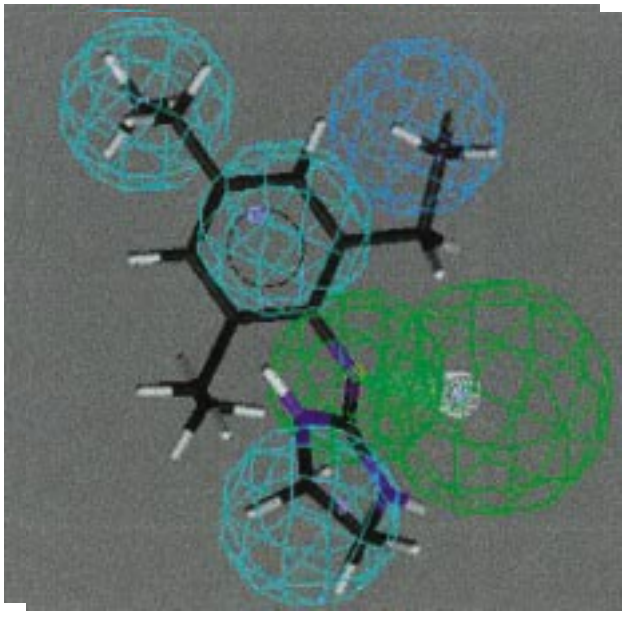

Figure 3. Mapping of **30** on to hypothesis 1, which contains a HBA (green), Hp 1, 2, 3 (light blue), and a HpAl (blue).

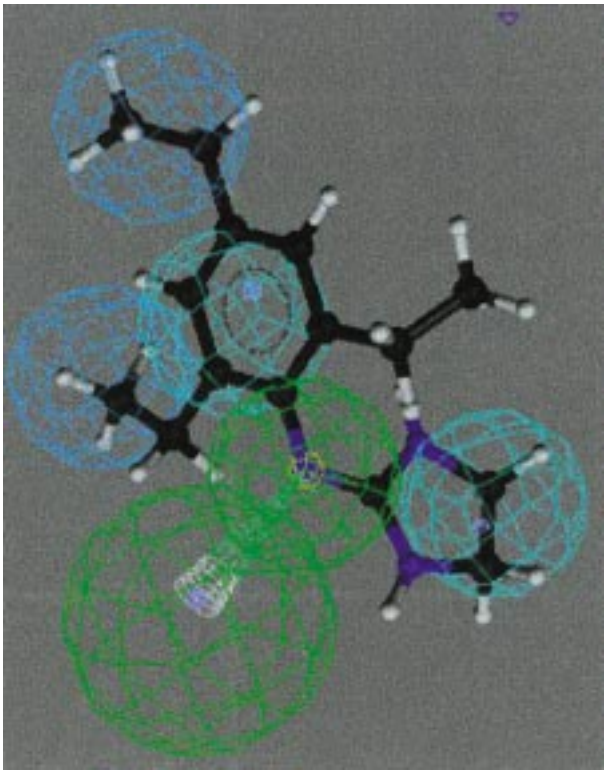

Figure 4. Mapping of **30** on to hypothesis 3, which consists of a HBA (green), Hp 1, 2 (light blue), HpAl 1, and 2 (blue).

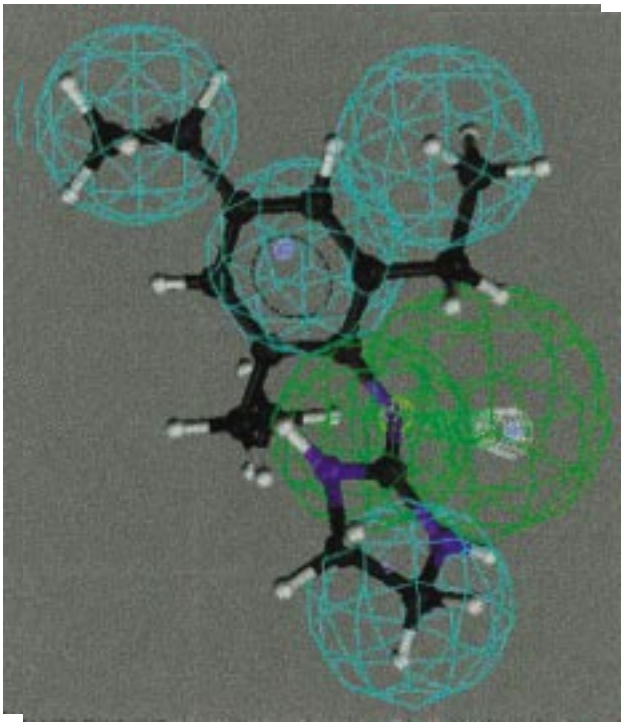

Figure 5. Mapping of **30** on to hypothesis 5, which consists of a HBA (green), Hp 1, 2, 3, and 4 (light blue).

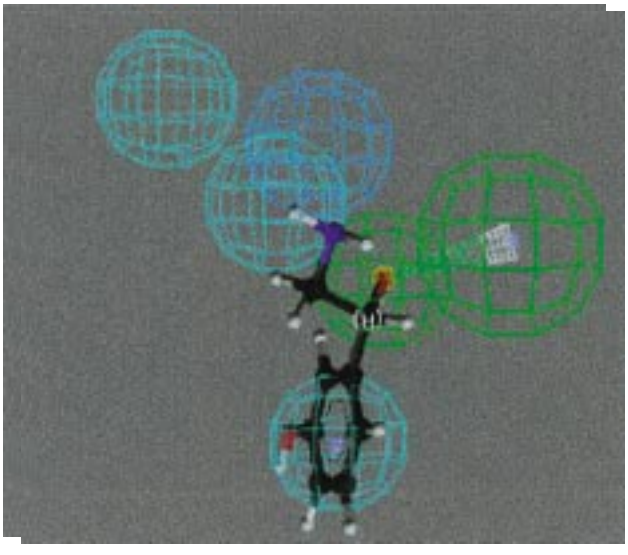

Figure 6. Mapping of **35** on to hypothesis 1, which contains a HBA (green), Hp 1, 2, 3 (light blue), and a HpAl (blue).

molecules are listed in Table 3. Although antagonists may not interact with the same part of the membrane with which the agonists interact, the presence of some common structural elements, such as the phenyl ring, suggests the binding sites may have some features in common. The ionophore, one component of the cell membrane, may be activated between a ligand–receptor interaction. Antagonists may act via the receptor or via the ionophore or a combination of both. Taken the part of the membrane with which the agonist interacts as the true receptor, the antagonist may well interact with an area surrounding the receptor including the ionophore. Hence, further research on the comparison of the 3-D hypotheses from agonists and antagonists as well as those generated from the corresponding data from various insect tissues might be interesting and stimulating to investigate further.

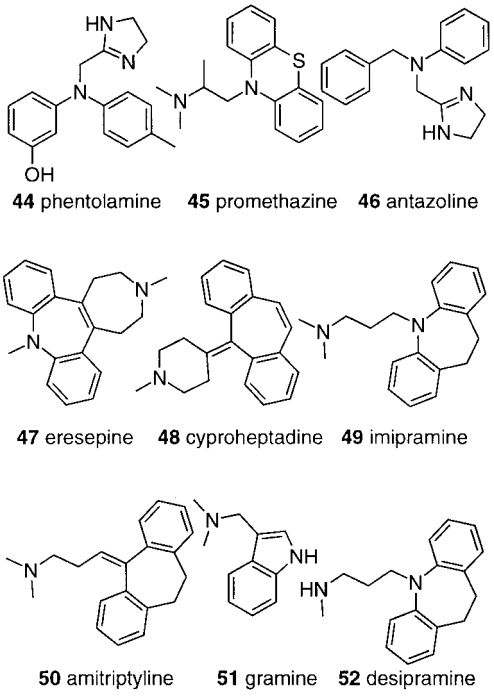

Figure 7. Structures of OA antagonists in the test set.

Table 3. Activity and predicted activity of OA antagonists from hypothesis 1

Compound no.	Confs.	Activity (nM)	Predicted activity (nM)	Error
44	101	19.1	6.4	−3
45	55	35.5	51	1.4
46	71	117.5	130	1.1
47	22	478.6	160	−2.9
48	14	851.1	1200	1.4
49	62	1445.4	190	−7.8
50	58	1584.9	300	−5.3
51	21	1819.7	1300	−1.4
52	89	3235.9	1200	−2.7

Confs means number of conformational models. Activity was cited from ref 13.

Conclusion

The present work shows how a set of binding activities of various OA agonists may be treated statistically to uncover the molecular characteristics which are essential for high activity. These characteristics are expressed as chemical features disposed in 3-D space and are collectively termed a hypothesis. Hypotheses were obtained and applied to map the active or inactive compounds. Important features such as HBA, Hp, and HpAl of the surface-assessable models were found. A HBA and four hydrophobic features are the minimum components of an effective OA agonistic binding hypothesis. It was found that more active agonists map well on to all the features of the hypotheses. For some inactive compounds, their lack of binding affinity is primarily due to their inability to achieve an energetically favorable conformation shared by the active compounds.

Several 3-D pharmacophore models for the OA agonists-OA3 receptor interactions have been proposed. A further comparison study of the 3-D hypothesis models of antagonists and agonists is in progress. Such work will surely help to elucidate the mechanisms of OAR3 and other types of OA receptor-ligand interactions.

Experimental Procedure

Hypothesis generation

All experiments were conducted on a Silicon Graphics O2, runnning under the IRIX 6.3 operating system. Hypotheses generation was applied against previously described data sets and their functionality is available as part of Molecular Simulations Incorporated's Catalyst/Hypo modelling environment (Burlington, MA). Molecules were edited using the Catalyst 2-D/3-D visualizer. The Catalyst model treated molecular structures as templates consisting of strategically positioned chemical functions that will bind effectively with complementary functions on receptors. The biologically most important binding functions were deduced from a small set of compounds that covered a broad range of activity. Catalyst automatically generated conformational models for each compound using the Poling Algorithm.[18–20] The number of conformations needed to produce a good representation of a compound's conformational space depends on the molecule. Conformation-generating algorithms were adjusted to produce a diverse set of conformations, avoiding repetitious groups of conformations all representing local minima. The conformations generated were used to align common molecular features and generate pharmacophoric hypotheses. Hypo used conformations generated to align chemically important functional groups common to the molecules in the study set. A pharmacophoric hypothesis was then generated from these aligned structures. The Hypo was also used to incorporate biological activity data into the hypothesis-generating process. Each hypothesis was tested by regression techniques to compare estimated activity with actual activity data. The software used the data from these tests

to select the hypotheses that do the best job predicting activity for the set of study molecules.

The models emphasized a conformational diversity under the constraint of 20 kcal/mol energy threshold above the estimated global minimum based on use of the CHARMm force field.[18–21] Molecular flexibility is taken into account by considering each compound as a collection of conformers representing a different area of conformational space accessible to the molecule within a given energy range. The molecules associated with their conformational models was submitted to Catalyst hypothesis generation. Hypotheses approximating the pharmacophore are described as a set of features distributed within a 3-D space. This process only considered surface accessible functions such as HBA, HBD, Hp, HpAl, negative charge, positive charge, and RA etc.[25] The hypothesis generator was restricted to select only five features due to the molecule's flexibility and functional complexity. For molecules larger than dipeptides, Catalyst often will find five-feature hypotheses automatically, but for smaller molecules, three- or four-feature hypotheses might be in the majority. Since hypotheses with more features are more likely to be stereospecific and generally more restrictive models, the total features min value was set to 5 in order to force Catalyst to search for 5-feature hypotheses.

Validation of the hypothesis

Using the Fischer method,[23] the statistical significance of the hypothesis was accessed by randomizingly scrambling the bioassay activities in 19 new training sets and regenerating the hypotheses for each new trial. The statistical relevance of the various hypotheses obtained is assessed on the basis of their cost relative to the null hypothesis and their correlation coefficients r. During a hypothesis generation run, Catalyst considers and discards many thousands of models. It attempts to minimize a cost function consisting of two terms. One penalizes the deviation between the estimated activities of the training set molecules and their experimental values. The other penalizes the complexity of the hypothesis. The overall assumption used is based on Occam's razor, that between otherwise equivalent alternatives, the simplest model is best. Simplicity is defined using the minimum description length principle from information theory. The overall cost of a hypothesis is calculated by summing three cost factors: weight cost (a value that increases in a Gaussian form as the feature weight in a model deviates from an idealized value of 2.0), error cost (a major value that increases as the RMS difference between estimated and measured activities), and configuration cost (a fixed cost which depends on the complexity of the hypothesis, equal to entropy of the hypothesis space).

Besides providing a numerical score for each generated hypothesis, Catalyst provides two numbers to help the chemist assess the validity of a hypothesis. One is the cost of an ideal hypothesis, which is a lower bound on the cost of the simplest possible hypothesis that still fits the data perfectly. The other is the cost of the null hypothesis, which presumes that there is no statistically significant structure in the data, and that the experimental activities are normally distributed about their mean. Generally, the greater the difference between the two costs, the higher the probability for finding useful models. In terms of hypothesis significance, a generated hypothesis with a score that is substantially below that of the null hypothesis is likely to be statistically significant and bears visual inspection. It has been reported that a returned hypothesis which has a cost that differs from the null hypothesis by 40–60 bits might possess a 75–90% chance of representing a true correlation.

Acknowledgement

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. Erspamer, V.; Boretti, G. *Arch. Int. Pharmacodyn.* **1951**, *88*, 296.
2. Axelrod, J.; Saavedra, J. M. *Nature* **1977**, *265*, 501.
3. Evans, P. D. In *Comparative Molecular Neurobiology*; Heller, S. R., Ed.; Birkhäuser Verlag: Basel, 1993, pp 287–296.
4. Evans, P. D. In *Comprehensive Insect Physiology Biochemistry Pharmacology*; Kerkut, G. A., Gilbert, G., Eds.; Pergamon Press: Oxford, 1985; Vol 11, pp 499–530.
5. Saavedra, J. M. *Handb. Exp. Pharmacol.* **1989**, *90*, 181.
6. Orchard, I.; Ramirez, J. M.; Lange, A. B. *Ann. Rev. Entomology* **1993**, *38*, 227.
7. Nathanson, J. A. *Mol. Pharmac.* **1985**, *28*, 254.
8. Evans, P. D. *J. Physiol.* **1981**, *318*, 99.
9. Roeder, T.; Nathanson, J. A. *Neurochem. Res.* **1993**, *18*, 921.
10. Roeder, T.; Gewecke, M. *Biochem. Pharm.* **1990**, *39*, 1793.
11. Roeder, T. *Life Science* **1991**, *50*, 21.
12. Roeder, T. *Br. J. Pharmac.* **1995**, *114*, 210.
13. Roeder, T. *Eur. J. Pharmac.* **1990**, *191*, 221.
14. Harmar, A. J.; Horn, A. S. *Molec. Pharmac.* **1997**, *13*, 512.
15. Hirashima, A.; Shinkai, K.; Pan, C.; Kuwano, E.; Taniguchi, E.; Eto, M. *Pestic. Sci.* **1999**, *55*, 119.
16. Hirashima, A.; Pan, C.; Shinkai, K.; Tomita, J.; Kuwano, E.; Taniguchi, E.; Eto, M. *Bioorg. Med. Chem.* **1998**, *6*, 903.
17. Pan, C.; Hirashima, A.; Kuwano, E.; Eto, M. *J. Molec. Model.* **1997**, *3*, 455.
18. Smellie, A.; Teig, S. L.; Towbin, P. *J. Comp. Chem.* **1995**, *16*, 171.
19. Smellie, A.; Kahn, S. D.; Teig, S. L. *J. Chem. Inf. Comp. Sci.* **1995**, *35*, 285.
20. Smellie, A.; Kahn, S. D.; Teig, S. L. *J. Chem. Inf. Comp. Sci.* **1995**, *35*, 295.
21. Brooks, B. R.; Brucolleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
22. Greene, J.; Kahn, S.; Savoj, H.; Sprague, P.; Teig, S. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1297.
23. Fischer, R. In *The Design of Experiments*; Hafner Publishing: New York, 1996.
24. Hansch, C.; Leo, A. In *Exploring QSAR: Fundamentals and Applications in Chemistry and Biochemistry*; American Chemical Society: Washington, DC, 1995.
25. Hansch, C.; Fujita, T. *J. Am. Chem. Soc.* **1964**, *86*, 1616.
26. Golender, V. E.; Vorpagel, E. R. In *3D-QSAR in Drug Design: Theory, Methods, and Applications*; Kubinyi, H., Ed.; ESCOM Science Publishers: The Netherlands, 1993, p 137.